Journal of Computational Methods in Molecular Design, 2014, 4 (3):92-105

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

QSAR and anticancer drug design of β-Carboline compounds utilizing computed molecular descriptors

Ruby Yadav and Sisir Nandi*

Division of Pharmaceutical Chemistry, Global Institute of Pharmaceutical Education and Research (GIPER), Affiliated to Uttarakhand Technical University, Kashipur, Uttarakhand, India

---

ABSTRACT

An attempt has been made for the development of quantitative structure-activity relationship (QSAR) models for a series of β-carboline derivatives having DYRK1A (dual specificity tyrosine phosphorylated and regulated kinase 1a) inhibitory activities as potent anticancer agents toward the activation of caspase-9 which leads to massive apoptosis in different human cancer cell types including glioma, esophageal cancer and non-small-cell lung cancers respectively. A number of highly descriptive and predictive QSAR models for these compounds were obtained by considering in vitro anticancer activities against glioma cell lines including U373 and Hs683 respectively with various sets of theoretical molecular descriptors including topological, constitutional, geometrical, functional groups and atom centered fragment indices calculated solely from the structures of 48 synthesized β-carboline derivatives using stepwise-multiple linear regressions methods. Model validation is performed by incorporating training and test sets approach and calculating R^2, Q_{Loo}^2, R_{pred}^2 and standard error of estimation (SEE) respectively. From these models a number of significant features of these congeners including X3Av, EEig13x, MATS5m and SP05 which are responsible for size, shape and weight of the molecules whereas EEig03d, MATS4p and QYYp which indicate dipole moment, polarizabilities, conjugation and aromaticity have been predicted for the design of more promising β-carboline anticancer compounds.

Keywords: β-carboline compounds, DYRK1A inhibitors, computed molecular descriptors, topological indices, Multiple linear regression analysis, QSAR, Anticancer drug design.

INTRODUCTION

Cancer is characterized by the malignant tumors and malignant neoplasm which may be defined as abnormal, excessive, uncoordinated, and autonomous proliferation of cells even after cessation of stimulus for growth which caused it.Malignancies is developing as the most common cause of increasing death rate in the worldand thus the demand an extensive research in cancer biology and chemotherapy, both in revealing pathobiology of the diseases and discovery of new leads [1]. The main toxicity of the cancer chemotherapeutics is to kill normal cells along with the cancer infected cells. Thus scientists are now more concentrating about the design and discovery of potential anticancer leads which may cause stimulation of caspase-mediated apoptosis process without toxicity of killing the normal cells [2-5]. The major mechanism is to achieve these less toxic anticancer activities by developing phyto active constituents. One of the important phytoconstituents useful for the purpose to overcome the intrinsic resistance of cancer cells to apoptotic stimuli is β-carbolines that are structurally related to harmine, which is a naturally occurring and previously isolated from plants, including the Middle Eastern grass harmal or Syrian rue

(Peganumharmala) and the South American vine ayahuasca (Banisteriopsiscaapi). A lot of attention has been paid for the development ofβ-carboline derivativesfor its potential anticancer activities. The biochemical mechanisms of these compounds are DNA intercalation and inhibition ofdual specificity tyrosine phosphorylated and regulated kinase 1a(DYRK1A), an enzyme involved in uncontrolled cell proliferation and cancer cell chemoresistance that is overexpressed in melanomas intrinsically resistant to apoptotic stimuli. DYRK1A is a dual-specificity protein kinase that auto phosphorylates a conserved tyrosine residue in the activation loop but phosphorylates exogenous substrates only at serine or threonine residues[6].DYRK1A inhibition induces the activation of caspase-9 which leads to massive apoptosis in different human cancer cell types including glioma, esophageal cancer and non-small-cell lung cancerswhich are among the cancers associated with the worstprognoses because of their ability to resist most if not all of thecurrent chemotherapies [7-9].The molecular mechanism of harmine compounds have been explored by Seifert et al. [10] who identifiedthat the cysteine aspartyl protease caspase-9, a critical component of the intrinsic apoptotic pathway, as a substrate of DYRK1A. As per their observation, depletion of DYRK1A from human cells by short interfering RNA inhibits the basal phosphorylation of caspase-9 at an inhibitory site, Thr125. DYRK1A dependent phosphorylation of Thr125 is also blocked by harmine, confirming the use of this β-carboline alkaloid as a potent inhibitor of DYRK1A in cells[11].Caoet al.synthesized several series of β-carboline derivatives considering the starting material as L-Tryptophan on the basis of harminenucleus.In-vitro cytotoxic activities of these compounds were investigated. The results were shown that on the introduction of benzyl substituent at position 2,anti-tumor activities of these compounds were increased, along with the acute toxicity. Substitution of ethoxy carbonyl amino at position 3 reduced the acute toxicity as well as anti-tumor activity. Toxicity was reduced after the introduction of appropriate substituent at position 3 and 9 [12-13].In the study of Ma et al.[14], harmine was identified in a screening program as a novel breast cancer resistance protein(BCRP) reversal agent. It inhibitedBCRP-mediated drug efflux and increased the cytotoxicity of anticancer drugs in a BRCP overexpressing breast cancer cell line MDA-MB-231.Ishida et al. [15] reported thatharmine and β-carboline analogues exhibited significantactivities against several human tumor cell lines includingthree drug-resistant KB sublines with various resistancemechanisms, and α-(4-nitrobenzylidine)-harmine had a broadcytotoxicity spectrum against 1A9, KB, SaOS-2, A549, SKMEL-2, U-87-MG and MCF-7 cell lines respectively.

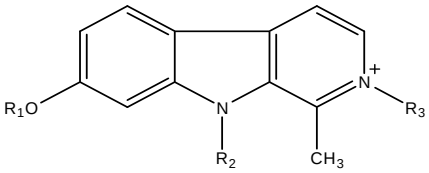
A number of 48 novel β-carbolines structurally related to harmine have been synthesized by Frederick et al. and evaluated the in vitro anticancer activities of these compounds against different types of glioma and esophageal cancer cell lines including Hs683 oligodendroglioma cell line (ATCC code HTB-138) and the U373 (ECACC code 89081403).As per their observations, these β-carbolines cause sensitization to the apoptotic stimuli and produce less intrinsic resistance to the cancer infected cells.Structure activity relationships showed that lipophilicity is one of the criteria for producing anticancer activity of these congeners[16].But there is hardly anyspecific QSAR modeling based on large number of computed structural descriptors calculated solely from the structures ofβ-carboline derivatives utilized as potent anti-cancer agents reported yet. Thus, an attempt has been made to perform the quantitative structure-activity relationship studies of these derivatives utilizing theoretical molecular descriptors computed from the structure of these compounds to explore the essential structural requirements to design more potent active β-carbolinecongeners having more effective treatment against various cancer cell lines.

MATERIALS AND METHODS

Biological activity data

A number of 48 β-carboline compounds having promising anticancer activities have been considered in the present study. These compounds were synthesized by Frederick et al. [16]. Table 1 contains structural substituent along with biological activities of 48 compounds. In vitro anticancer activities of these compounds have been measured in terms of IC_{50} against various glioma cell lines such as U373, and Hs683 respectively.Harmine structure contains β-carboline nucleus, also known as norharmane which is a nitrogen containing heterocycle. β-carboline consist of an indole ring fused with a pyridine ring having various aromatic and aliphatic substituents such as R_1,R_2 and R_3respectively to be substituted in the parent nucleus to produce a large number of compounds shown in the Table 1.These compounds have been considered in the present article for computation of molecular structure optimization and calculation of theoretical molecular descriptors including topological, constitutional, geometrical, functional group and atom fragment descriptors. The calculated molecular descriptors are then used to develop QSAR models derived from application of statistical tools correlating anticancer activities of β-carbolines and various structural invariants.

Table 1: Biological activity data of β-caboline compounds



Compound number	R_1	R_2	R_3	-log IC_{50} U373	-log IC_{50} HS683
1	CH_3	H	-	-1.505	-1.568
2*	$CH_2CH=CH_2$	H	-	-1.505	-1.447
3*	$CH_2CH(CH_3)_2$	H	-	-1.380	-1.361
4	$(CH_2)_2$-OCH_3	H	-	-1.806	-1.724
5	$(CH_2)_2$-OH	H	-	-1.623	-1.612
6	(CH_2)-cyclohexyl	H	-	-0.875	-1.342
7*	CH_2-C_6H_5	H	-	-0.968	-1.230
8	$(CH_2)_2$-C_6H_5	H	-	-0.959	-1.322
9	CO-C_6H_5	H	-	-1.430	-1.724
10	CH_2-2′-pyridyl	H	-	-1.322	-1.301
11*	CH_2-3′-pyridyl	H	-	-1.380	-1.477
12	CH_2-4′-pyridyl	H	-	-0.397	-0.908
13	CH_2-napthyl	H	-	-0.869	-1.278
14*	H	CH_2-C_6H_5	-	-1.903	-1.897
15	CH_2-C_6H_5	CH_2-C_6H_5	-	-1.255	-1.518
16	3′-fluorobenzyl	3′-fluorobenzyl	-	0.397	-1.255
17*	4′-fluorobenzyl	4′-fluorobenzyl	-	-1.477	-1.230
18	CH_2-cyclohexyl	CH_2-cyclohexyl	-	-1.491	-1.477
19	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	-	-1.518	-1.477
20	CH_2-C_6H_5	$(CH_2)_2CH_3$	-	-1.419	-1.113
21*	CH_2-C_6H_5	H	CH_2-C_6H_5	-0.591	-0.556
22	CH_2-C_6H_5	CH_2-C_6H_5	CH_2-C_6H_5	0.356	0.301
23	CH_2-C_6H_5	CH_2-C_6H_5	2′-fluorobenzyl	0.301	0.356
24	CH_2-C_6H_5	CH_2-C_6H_5	4′florobenzyl	0.221	-0.204
25*	CH_2-C_6H_5	CH_2-C_6H_5	$(CH_2)_2$-C_6H_5	-0.301	-0.361
26	CH_2-C_6H_5	CH_2-C_6H_5	$(CH_2)_2CH_3$	-0.397	0.221
27	CH_2-C_6H_5	CH_2-C_6H_5	$(CH_2)_2CH_3$	-0.154	0.221
28*	CH_2-C_6H_5	CH_2-C_6H_5	$(CH_2)_2$-$CH(CH_3)_2$	-0.408	-0.4318
29	CH_2-C_6H_5	CH_2-C_6H_5	$(CH_2)_2OH$	-0.954	-1.491
30	3′-fluorobenzyl	3′-fluorobenzyl	CH_2-C_6H_5	0.096	-0.447
31	3′-fluorobenzyl	3′-fluorobenzyl	2′-fluorobenzyl	0.221	-0.342
32*	3′-fluorobenzyl	3′-fluorobenzyl	4′-fluorobenzyl	-0.045	-0.653
33	4′-fluorobenzyl	4′-fluorobenzyl	CH_2-C_6H_5	0.366	0.0457
34	4′-fluorobenzyl	4′-fluorobenzyl	2′-fluorobenzyl	0.301	-0.079
35	4′-fluorobenzyl	4′-fluorobenzyl	4′-fluorobenzyl	-0.462	-1.113
36*	CH_2-cyclohexyl	CH_2-cyclohexyl	CH_2-C_6H_5	0.431	-0.420
37	CH_2-cyclohexyl	CH_2-cyclohexyl	2′-fluorobenzyl	0.146	0.602
38*	CH_2-cyclohexyl	CH_2-cyclohexyl	4′-fluorobenzyl	1.397	-0.491
39	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2OH$	-0.397	-0.602
40*	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	0.657	0.585
41	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	CH_2-C_6H_5	-0.530	-0.490
42	CH_2-C_6H_5	$(CH_2)_2CH_3$	CH_2-C_6H_5	0.387	-0.414
43	CH_2-C_6H_5	$(CH_2)_2CH_3$	2′-fluorobenzyl	0.408	0.346
44	CH_2-C_6H_5	$(CH_2)_2CH_3$	4′-fluorobenzyl	0.055	-0.491
45*	$(CH_2)_2$-$CH(CH_3)_2$	$(CH_2)_2$-$CH(CH_3)_2$	CH_2-C_6H_5	0.420	0.327
46	CH_2-cyclohexyl	CH_2-cyclohexyl	$(CH_2)_2OH$	-0.230	-0.579
47*	CH_2-cyclohexyl	CH_2-cyclohexyl	$(CH_2)_2CH(CH_3)_2$	0.397	0.376
48	CH_2-cyclohexyl	CH_2-cyclohexyl	$(CH_2)_2$-C_6H_5	0.698	0.508

**Test set compounds*

Computation

Structure optimization: All the structures of 48 β-carboline compounds were drawn using 2D Chemdraw. The drawn structures were then converted into 3D modules and the geometries of all compounds were fully optimized using MM2 force field considering the default conversion procedure implemented in Chem3D Ultra [17].

Calculation of theoretical molecular descriptors: Theoretical molecular descriptors are the numerical representation of molecule, achieved by applying the principles of graph theory to molecular structure. It encodes molecular architecture and quantifies such aspects of molecular structure as size, shape, symmetry, complexity, branching, cyclicity, stereoelectronic character, etc.Structural descriptors can be categorized as physicochemical, constitutional and geometrical (3D), electrostatic, functional and atom-centered and topological respectively. The physicochemical descriptors include AlogP98 value, AMR value, buffer solubility, polarizability, vapor density, water solubility, solvation free energy, and so forth. The constitutional descriptors consist of molecular descriptors such as molecular mass, molecular formula, formal charges, fraction of rotatable bonds, and number of rigid bonds, rings, charged groups, and so forth. The three-dimensional or shape descriptors (3D) are more complex, encoding information about the three-dimensional aspects of molecular structure. The electrostatic descriptors constitute charged polarization, polarity parameter,local dipole index, maximum positive charged, maximum negative charged, total absolute atomic charge, total negative charge, total positive charge, and so forth. The functional group and atom centered descriptors represent the contribution of different functional groups and atoms upon biological activity of the compounds whereas topological descriptors are the largest set of molecular descriptors and may be subdivided into two classes: topostructural and topochemical descriptors. Topostructural descriptors encode information strictly on the neighborhood and connectivity of atoms within the molecule, while the topochemical descriptors encode information related to both the topology of the molecule and the chemical nature of atoms and bonds within it [18-22].

In the present work a total number of 898 topological descriptors, useful for our purpose, were calculated via DRAGON software [23], and before model development, these were reduced to 415. The reduction in the descriptors was due to keeping a constant value for, or nearly all, of the compounds, and for those that perfectly correlated (r = 1.0) with other descriptors. The reduced sets of descriptors were then treated by multiple linear regressions (MLR) algorithm for developing QSAR models. Table 2 represents different classes of molecular descriptors along with their symbols.

Table 2: Computed Theoretical Molecular descriptors used in this study

Descriptor classes	Descriptor names
Topological Descriptors	First Zagreb index (ZM1),first Zagreb index by valence vertex degrees(ZM1V), second Zagreb index(ZM2), second Zagreb index by valence vertex degrees(ZM2V), quadratic index(Qindex), Narumi simple topological index (log function)(SNar), Narumi harmonic topological index(HNar), Narumi geometric topological index (GNar), total structure connectivity index(Xt), Pogliani index (Dz), Pogliani index (Ram), polarity number(Pol), log of product of row sums (PRS), log of product of row sums (LPRS),(VDA),mean square distance index (MSD), Schultz Molecular Topological Index (SMTI), Schultz Molecular Topological Index by valence vertex degrees(SMTIV),Gutman Molecular Topological Index (GMTI),Gutman Molecular Topological Index by valence vertex degrees (GMTIV),Xu index(Xu),superpendentic index (SPI),W,WA,Har,Har2, quasi-Wiener index (Kirchhoff number) from Laplace matrix (QW), first Mohar index from Laplace matrix(TI1), second Mohar index from Laplace matrix(TI2), spanning tree number (log function) from Laplace matrix (STN),HyDp,RHyDp, Wiener-like index from topological distance matrix (w),ww,Rww, Wiener-like index from distance/detour matrix (D/D), all-path Wiener index (Wap),WhetZ,Whetv, Whete,Whetp,J,JhetZ,Jhetv,Jhete,Jhetp, maximal electrotopological negative variation (MAXDN), maximal electrotopological positive variation (MAXDP), molecular electrotopological variation (DELS), E-state topological parameter (TIE), Kier symmetry index (S0K), 1-path Kier alpha-modified shape index (S1K), 2-path Kier alpha-modified shape index (S2K), 3-path Kier alpha-modified shape index (S3K), Kier flexibility index (PHI), Kier benzene-likeliness index (BLI), path/walk 2 - Randic shape index(PW2), path/walk 3 - Randic shape index(PW3), path/walk 4 - Randic shape index(PW4), path/walk 5 - Randic shape index(PW5), 2D Petitjean shape index(PJI2), eccentric connectivity index (CSI), eccentricity (ECC), average eccentricity (AECC), eccentric (DECC), mean distance degree deviation (MDDD),unipolarity (UNIP), centralization(CENT), variation (VAR),Balaban centric index (BAC), lopping (Lop), radial centric information index (ICR),D/Dr06, sum of topological distances between N..O(T(N..O)),sum of topological distances between N..S(T(N..S)), sum of topological distances between O..O(T(O..O)), sum of topological distances between O..S(T(O..S)), molecular walk count of order 2 (MWC02), molecular walk count of order 3(MWC03), molecular walk count of order 4(MWC04), molecular walk count of order 5(MWC05), molecular walk count of order 6(MWC06), molecular walk count of order 7(MWC07), molecular walk count of order 8(MWC08), molecular walk count of order 10(MWC10), total walk coun(TWC), SRW01, self-returning walk count of order 2(SRW02), self-returning walk count of order 4(SRW04), self-returning walk count of order 6(SRW06), self-returning walk count of order 7(SRW07), self-returning walk count of order 8(SRW08), self-returning walk count of order 10(SRW10), molecular path count of order 2 (Gordon-Scantlebury index)(MPC02), molecular path

Descriptor classes	Descriptor names
	count of order 3(MPC03), molecular path count of order 4(MPC04), molecular path count of order 5(MPC05), molecular path count of order 6(MPC06), molecular path count of order 7(MPC07), molecular path count of order 8(MPC08), molecular path count of order 9(MPC09), molecular path count of order 10(MPC10), molecular multiple path count of order 1(piPC01), molecular multiple path count of order 2(piPC02), molecular multiple path count of order3 (piPC03), x molecular multiple path count of order 4(piPC04 molecular multiple path count of order 4), molecular multiple path count of order 5(piPC05), molecular multiple path count of order 6(piPC06), molecular multiple path count of order 7(piPC07), molecular multiple path count of order 8(piPC08), molecular multiple path count of order 9(piPC09), molecular multiple path count of order 10(piPC10), total path count(TPC), conventional bond order ID number(piID), ratio of multiple path count over path coun(PCR), difference between multiple path count and path count(PCD),Randic ID number(CID), Balaban ID number(BID), connectivity index of order 0(X0), connectivity index of order 1 (Randic connectivity index)(X1), connectivity index of order 2(X2), connectivity index of order 3(X3), connectivity index of order 4(X4), connectivity index of order 5(X5), average connectivity index of order 0(X0A), average connectivity index of order 1(X1A), average connectivity index of order 2(X2A), average connectivity index of order 3(X3A), valence connectivity index of order 0(X0v), valence connectivity index of order 1(X1v), valence connectivity index of order 2(X2v), valence connectivity index of order 3(X3v), valence connectivity index of order 4(X4v), valence connectivity index of order 5(X5v), average valence connectivity index of order 0(X0Av), average valence connectivity index of order 1(X1Av), average valence connectivity index of order 2(X2Av), average valence connectivity index of order 3(X3Av), average valence connectivity index of order 4(X4Av), solvation connectivity index of order 0(X0sol), solvation connectivity index of order 1(X1sol), solvation connectivity index of order 2(X2sol), solvation connectivity index of order 3(X3sol), solvation connectivity index of order 4(X4sol), solvation connectivity index of order 5(X5sol), modified Randic index(XMOD), reciprocal distance sum Randic-like index(RDCHI), reciprocal distance sum inverse Randic-like index(RDSQ), information index on molecular size(ISIZ), x total information index on atomic composition(IAC), mean information content on the distance equalit(IDE), x mean information content on the distance magnitude(IDM mean information content on the distance magnitude), mean information content on the distance degree equality(iDDE), mean information content on the distance degree magnitude(IDDM), total information content on the distance equality(IDET), total information content on the distance magnitude(IDMT), mean information content on the vertex degree equality(IVDE), mean information content on the vertex degree magnitude(IVDM), graph vertex complexity index(HVcpx), graph distance complexity index (log function)(HDcpx), Balaban U index(Uindex), Balaban V index(Vindex), Balaban X index(Xindex), Balaban Y index(Yindex), Information Content index (neighborhood symmetry of 0-order)(IC0), Total Information Content index (neighborhood symmetry of 0-order)(TIC0), Structural Information Content index (neighborhood symmetry of 0-order)(SIC0), Complementary Information Content index (neighborhood symmetry of 0-order)(CIC0), Bond Information Content index (neighborhood symmetry of 0-order)(BIC0), Information Content index (neighborhood symmetry of 1-order)(IC1), Total Information Content index (neighborhood symmetry of 1-order)(TIC1), Structural Information Content index (neighborhood symmetry of 1-order)(SIC1), Complementary Information Content index (neighborhood symmetry of 1-order)(CIC1), Bond Information Content index (neighborhood symmetry of 1-order)(BIC1), Information Content index (neighborhood symmetry of 2-order)(IC2), Total Information Content index (neighborhood symmetry of 2-order)(TIC2), Structural Information Content index (neighborhood symmetry of 2-order)(SIC2), Complementary Information Content index (neighborhood symmetry of 1-order)(CIC1), Bond Information Content index (neighborhood symmetry of 2-order)(BIC2), Information Content index (neighborhood symmetry of 3-order)(IC3), Total Information Content index (neighborhood symmetry of 4-order)(TIC4), Structural Information Content index (neighborhood symmetry of 3-order)(SIC3), Complementary Information Content index (neighborhood symmetry of 3-order)(CIC3), Bond Information Content index (neighborhood symmetry of 3-order)(BIC3), Information Content index (neighborhood symmetry of 4-order)(IC4), Total Information Content index (neighborhood symmetry of 4-order)(TIC4), Structural Information Content index (neighborhood symmetry of 4-order)(SIC4), Complementary Information Content index (neighborhood symmetry of 4-order)(CIC4), Bond Information Content index (neighborhood symmetry of 4-order)(BIC4), Information Content index (neighborhood symmetry of 5-order)(IC5), Total Information Content index (neighborhood symmetry of 5-order)(TIC5), Structural Information Content index (neighborhood symmetry of 5-order)(SIC5),Complementary Information Content index (neighborhood symmetry of 5-order)(CIC5), Bond Information Content index (neighborhood symmetry of 5-order)(BIC5), Autocorrelation.Eigenvalue 03 from edge adj. matrix weighted by dipole moments(EEig03d),Eigenvalue 13 from edge adj. matrix weighted by dipole moments(EEig13d),Eigenvalue 13 from edge adj. matrix weighted by edge degrees (EEig13x),shape profile no. 05(SP05),Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities(MATS3e),Moran autocorrelation - lag 4 / weighted by atomic polarizabilities(MATS4p),Moran autocorrelation of lag 5 weighted by mass (MATS5m)
Constitutional Descriptors	Molecular weight (MW), average molecular weight (AMW), sum of atomic van der Waals volumes (scaled on Carbon atom) (Sv), sum of atomic Sanderson electronegativities (scaled on Carbon atom)(Se), sum of atomic polarizabilities (scaled on Carbon atom) (Sp), sum of first ionization potentials (scaled on Carbon atom) (Si), mean atomic van der Waals volume (scaled on Carbon atom (Mv), mean atomic Sanderson electronegativity (scaled on Carbon atom) (Me), mean atomic polarizability (scaled on Carbon atom) (Mp), mean first ionization potential (scaled on Carbon atom) (Mi), number of atoms (nAT), number of non-H atoms (nSK), number of bonds (nBT), number of non-H bonds(nBO), number of multiple bonds(nBM), sum of conventional bond orders (H-depleted) (SCBO),number of rotatable bonds (RBN), rotatable bond fraction (RBF),number of double bonds (nDB), number of triple bonds (nTB),number of aromatic bonds(nAB),number of Hydrogen atoms(nH),number of Carbon atom (nC),number of Nitrogen atoms (nN), number of Oxygen atoms (nO),number of Phosphorous atoms (nP),number of Sulfur atoms (nS), number of Fluorine atoms (nF) , number of Chlorine atoms (nCL) ,number of Bromine atoms (nBR),number of Iodine atoms (nI), number of Boron atoms (nB), number of heavy atoms (nHM),number of heteroatoms (nHet),number of halogen atoms (nX),percentage of H atoms

Descriptor classes	Descriptor names
	(H%)percentage of C atoms (C%), percentage of N atoms (N%),percentage of O atoms (O%) , percentage of halogen atoms (X%), number of sp3 hybridized Carbon atoms (nCsp3), number of sp2 hybridized Carbon atoms(nCsp2), number of sp hybridized Carbon atoms (nCsp)
Geometrical descriptors	Gravitational index G1(G1), gravitational index G2 (bond-restricted)(G2), radius of gyration (mass weighted)(RGyr), span R(SPAN), average span R(SPAM), molecular eccentricity(MEcc), spherosity(SPH), asphericity(ASP), 3D Petitjean shape index(PJI3), length-to-breadth ratio by WHIM(L/Bw), Folding degree index (FDI), Harmonic Oscillator Model of Aromaticity index(HOMA), ring complexity index(RCI), aromaticity index(AROM), HOMA total(HOMT), displacement value / weighted by mass(DISPm), quadrupole x-component value / weighted by mass(QXXm), quadrupole y-component value / weighted by mass(QYYm), quadrupole z-component value / weighted by mass(QZZm), displacement value / weighted by van der Waals volume(DISPv), quadrupole x-component value / weighted by van der Waals volume(QXXv), quadrupole y-component value / weighted by van der Waals volume(QYYv), quadrupole z-component value / weighted by van der Waals volume(QZZv), displacement value / weighted by Sanderson electronegativity(DISPe), quadrupole x-component value / weighted by Sanderson electronegativity(QXXe), quadrupole y-component value / weighted by Sanderson electronegativity(QYYe), quadrupole z-component value / weighted by Sanderson electronegativity(QZZe), displacement value / weighted by polarizability(DISPp), quadrupole x-component value / weighted by polarizability(QXXp), quadrupole y-component value / weighted by polarizability(QYYp), quadrupole z-component value / weighted by polarizability(QZZp), sum of geometrical distances between N..N(G(N..N)), sum of geometrical distances between N..O(G(N..O)), sum of geometrical distances between N..S(G(N..S)), sum of geometrical distances between N..F(G(N..F)), sum of geometrical distances between N..Cl(G(N..Cl)), sum of geometrical distances between N..I((N..I)), sum of geometrical distances between O..O(G(O..O)), sum of geometrical distances between O..S(G(O..S)), sum of geometrical distances between O..F(G(O..F)), sum of geometrical distances between O..Cl(G(O..Cl)),sum of geometrical distances between O..I(G(O..I)), sum of geometrical distances between S..S(G(S..S)),sum of geometrical distances between S..F(G(S..F)), sum of geometrical distances between S..Cl(G(S..Cl)), sum of geometrical distances between S..I(G(S..I)),x sum of geometrical distances between Cl..Cl(G(Cl..Cl)
Functional Group and Atom centered Fragments descriptors	number of terminal primary C(sp3)(nCp), number of total secondary C(sp3)(nCs), number of total tertiary C(sp3)(nCt), number of ring secondary C(sp3)(nCrs), number of ring tertiary C(sp3)(nCrt), number of aromatic C(sp2)(nCar), number of unsubstituted benzene C(sp2)(nCbH), - number of substituted benzene C(sp2)(nCb), number of non-aromatic conjugated C(sp2)(nCconj), number of terminal primary C(sp2)(nR=Cp), number of aliphatic secondary C(sp2)(nR=Cs), number of aliphatic tertiary C(sp2)(nR=Ct), number of esters (aromatic)(nArCOOR), number of positively charged N(nN+), number of nitro groups (aromatic)(nArNO2), number of hydroxyl group(nROH), number of ethers (aliphatic)(nROR), number of ethers (aromatic)(nArOR), number of CH2RX(nCH2RX), number of X on aromatic ring(nArX), number of donor atoms for H-bonds (N and O)(nHDon), number of acceptor atoms for H-bonds (N,O,F)(nHAcc), CH3R / CH(C-001), CH2R2(C-002), CHR3(C-003), CH3X(C-005), CH2RX(v)(C-006), CHR2X(C-008), =CH2(C-015), =CHR(C-016), =CR2(C-017), R--CH—R(C-024), R--CR—R(C-025), R--CX—R(C-026), R-C(=X)-X / R-C#X / X=C=X(C-040), H attached to C0(sp3) no X attached to next C(H-046), H attached to C1(sp3)/C0(sp2)(H-047), H attached to C2(sp3)/C1(sp2)/C0(sp)(H-048), H attached to heteroatom(H-050), H attached to alpha-C(H-051), H attached to C0(sp3) with 1X attached to next C(H-052), alcohol(O-056), #NOME?(O-058), Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X(O-060), O—(O-061),Cl attached to C1(sp2)(Cl-089), I attached to C1(sp2)(I-099), R=S(S-108).

Statistical Methodology

Stepwise-Multiple Linear Regression Analysis: Quantitative structure–activity relationships are regression models having significant role in the biochemical sciences and engineering. QSAR regression models relate a set of predictor variables (X) calculated from the chemical structures to the potency of the response variable (Y) which is biological activity. It is necessary to consider a large number of physicochemical as well as other calculated molecular descriptors such as constitutional, geometrical, electrostatic, topological, functional group and atom centered fragments descriptors for the QSAR modeling of chemical compounds. In the present study, a large number of different types of topological as well as physicochemical descriptors have been taken into consideration to develop QSAR of β-carboline compounds. Multivariate regression analysis (MRA), one of the oldest data reduction methodologies, continues to be widely used in QSAR [23], as it does not impose any restriction on the type and number of graphical invariants used in structure–property–activity studies.The ultimate goal of QSAR-based drug design is to find out which structural properties confer the drug highest potency or lowest toxicity. The drug's potency is here a dependent variable, and the structural properties, also called molecular descriptors, are the independent variables. The experimental signal that measures the potency could be, for example, the binding affinity of a drug candidate to its target protein [24].Simple linear regression is the method of choice when the research question is to predict the value of a response (dependent) variable, denoted *Y*, from an explanatory (independent) variable X.The regression model is

$$Y = a + bX$$

The extension of simple regression to two or more independent variables is straightforward. For example, if four independent variables are being studied, the multiple regression models is

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

where X_1 is the first independent variable and b_1 is the regression coefficient associated with it, X_2 is the second independent variable and b_2 is the regression coefficient associated with it, and so on. This arithmetic equation is called a linear combination; thus, the response variable Y can be expressed as a (linear) combination of the explanatory variables. Note that a linear combination is really just a weighted average that gives a single number (or index) after the X's are multiplied by their associated b's and the bX products are added[25-26].

The selection of significant predictor variables is a crucialstep in any QSAR study. If the association between the parameter(s) selected and activity is strong, then activity predictions will be possible. If there is only weak association, knowing the value of the parameter(s) will not help in predicting activity. Thus, for a given study, parameters should be selected which are relevant to the activity for the series of molecules under investigation and these parameters should have values which are obtained in a consistent manner. In the present study, stepwise forward-backward based feature selection method incorporated in Minitab software[27] has been applied to predict the significant variables.The stepwise forward-backward based feature selection method begins with no candidate variables in the model. Predictor variables are then checked one at a time using the partial correlation coefficient (equivalently F to enter) as a measure of importance in predicting the dependent variable. At each stage the variable with the highest significant partial correlation coefficient (F to enter) is added to the model. Once this has been done the partial F statistic (F to remove) is computed for all variables present in the model to check if any of the variables previously added can now be deleted. This procedure is continued until no further variables can be added or deleted from the model. The partial correlation coefficient for a given variable is the correlation between the given variable and the response when the present independent variables in the equation are held fixed. It is also the correlation between the given variable and the residuals computed from fitting an equation with the present independent variables in the equation. After variable selection, multiple linear regression (MLR) method has been used to derive a number of training QSAR models using different types of descriptor such as topological, constitutional and geometrical, functional group and atom centered fragments indices respectively which are solely calculated from the structure of the β-caboline compounds [28].

Model Validation

The QSAR model based on the topological, constitutional, geometrical, functional group and atom centered fragments descriptors calculated solely from the structures of β-caboline compounds are validated prior to its application for prediction of biological activities of the newly generated compounds. For proper validation of the model, the total compound data set is divided into training and test sets. 69% of the 80 molecules are considered as training set to build QSAR models while remaining 31% is taken as test set. The division is done by random selection. Test set molecules are indicated by asterisk given in Table 1.The quality of each model is denoted by R^2 (R is the square root of multiple R-square for regression), Q^2(cross-validated r^2) values for the training set, an external validation was performed by calculating predictive R^2 (R_{pred}^2) and the standard error of estimation,SEE, represents standard deviation which is measured by the error mean square,which expresses the variation of the residuals or the variation about the regression line. Thusstandard deviation is an absolute measure of quality of fit and should have a low value for theregression to be significant.

$R^2$and $Q^2$of a model can be obtained from:

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - \bar{Y})^2}$$

R^2 is a measure of explained variance. Each additional X variableadded to a model increases R^2. $R^2$is a relativemeasure of fit by the regression equation. Correspondingly, it represents the part of the variation inthe observed data that is explained by the regression.

Calculation of Q^2(cross-validated r^2) is called as internal validation.

$$Q^2 = 1 - \frac{\sum(Y_{obs} - Y_{pred})^2}{\sum(Y_{obs} - \bar{Y})^2}$$

where, Y_{obs} and Y_{pred} indicate observed and predicted activity values respectively and $\overline{Y}$ indicates mean activity value. A model is considered acceptable when the value of Q^2 exceeds 0.5.

External validation or predictability of the models are performed by calculating predictive R^2 (R_{pred}^2).

$$R^2_{Pred} = 1 - \frac{\sum(Y_{pred(Test)} - Y_{(Test)})^2}{\sum(Y_{(Test)} - \bar{Y}_{training})^2}$$

where, $Y_{pred(test)}$ and $Y_{(test)}$ indicate predicted and observed activity values respectively of the test set compounds and $\overline{Y}_{training}$ indicates mean of observed activity values of the training set. For a predictive QSAR model, the value of R^2_{pred} should be more than 0.5 [29-30].

RESULTS AND DISCUSSION

QSAR modeling

A number of six QSAR models have been developed for β-carboline compoundsutilizingvarious sets of computed molecular descriptors. In QSAR modeling for these congeners, the predictors consist of various sets of physico-chemical properties or theoretical molecular descriptors of chemical compounds andanticancer activities of the chemicals against different glioma cancer cell lines including U373 and Hs683 which are considered as response-variable. The developed models are given in the following Table 3.

Table 3: Different QSAR models along with the statistical quality parameters

Glioma cancer Cell lines	Model number	Descriptor type	Model Equation	Statistical parameters related to quality of the model				
				R^2	SEE	PRESS	Q^2	R^2_{pred}
U373 (n=33)	1	Topological	-log (IC50) = - 24.81 + (7.19) x (EEig03d)+(43.4) x (X3Av)+(-2.18)x (XEEig13d)+(1.11) X (EEig13x)+(1.72) X (MATS5m)+ (-0.28)X (MAXDN)	0.917	0.243	2.25380	0.878	0.646
	2	Constitutional and Geometrical	-log (IC50) = -2.034+(0.00206) x (QYYm)	0.716	0.413	5.95002	0.679	0.613
	3	Functional Group	-log (IC50) = 0.05103+(-0.642) x (nHDon)	0.596	0.492	8.31865	0.552	0.61
HS683 (n=33)	4	Topological	-log (IC50)= -5.659 +(0.793)x(SP05)+ (3.50)x(MATS4p)+ (-0.730)x(TI2)+ (-2.06)x(MATS3e)	0.830	0.330	3.778	0.780	0.431
	5	Constitutional and Geometrical	-log (IC50) = 0.17184+(0.0028) x (QYYp)+ (-0.71) x (RGyr) +(-0.112) x (PC01-05)	0.802	0.350	4.484	0.750	0.697
	6	Functional Group	= - 0.1687 + (-0.656 x nHDon)	0.642	0.45	7.264	0.596	0.591

Where, R^2 (R is the square root of multiple R-square for regression), Q_{Loo}^2(Leave one out cross-validated r^2) values for the training set,R_{pred}^2 is the predictive R^2 for the test set, PRESS is predictive sum of squared deviation for the training set, SEE is the standard error of estimation.

From the QSAR models 1-3 developed for U373 glioma cancer cell line, it is cleared that topological descriptors can explain and predict 91.7% and 87.8% of variances of the in vitro glioma cancer cell inhibitory activities of the studied compounds. This model can also produce 64.6% external predictability.

Constitutional and geometrical descriptors can explain and predict 71.6% and 67.9% of variances of the activities of the studied compounds. This model can also produce 61.3% external predictability whereas functional group and atom centered descriptors can explain 59.6% of the variances and can produce 55.2% of the internal and 61% of the external predictability respectively. Then the training QSAR models 1-3 are used to predict anticancer activities against U373 cancer cell line for test compounds.

The plot of observed versus predicted activities for the test compounds is represented in figures1-3. It is evident that the predicted activities of all the compounds in the test set using models 1-3 are in good agreement with their corresponding observed activities and optimal fit is obtained generated by the QSARs utilizing different set of topological, constitutional and geometrical, functional group and atom centered fragmentsdescriptors

respectively.The square correlation coefficients (r2) between experimental vs predicted activities of the test set molecules calculated using QSAR models 1-3 are 0.682, 0.659 and 0.656 respectively.

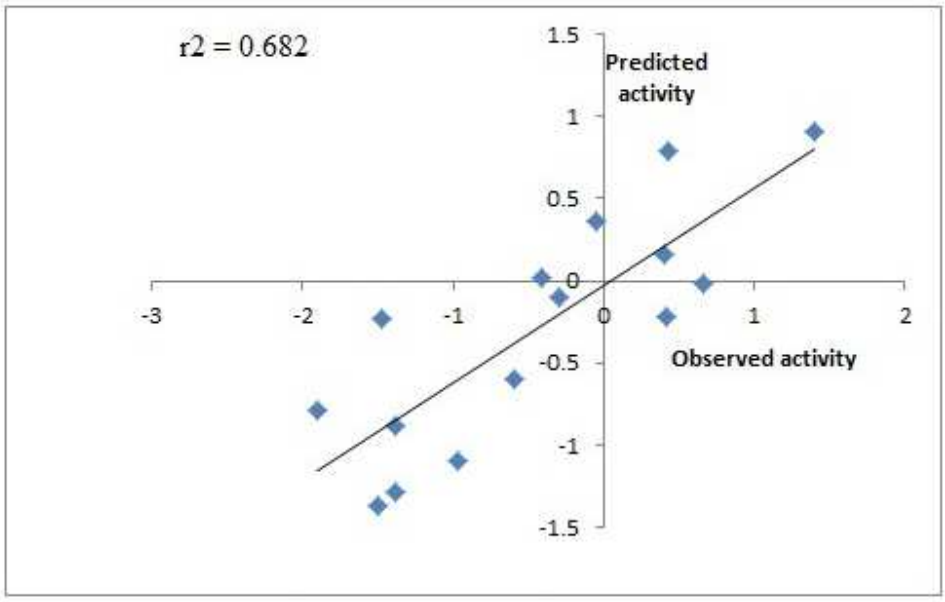


Figure 1: Observed activity vs. predicted activity of the test molecules (Using model 1 based on topological indices)

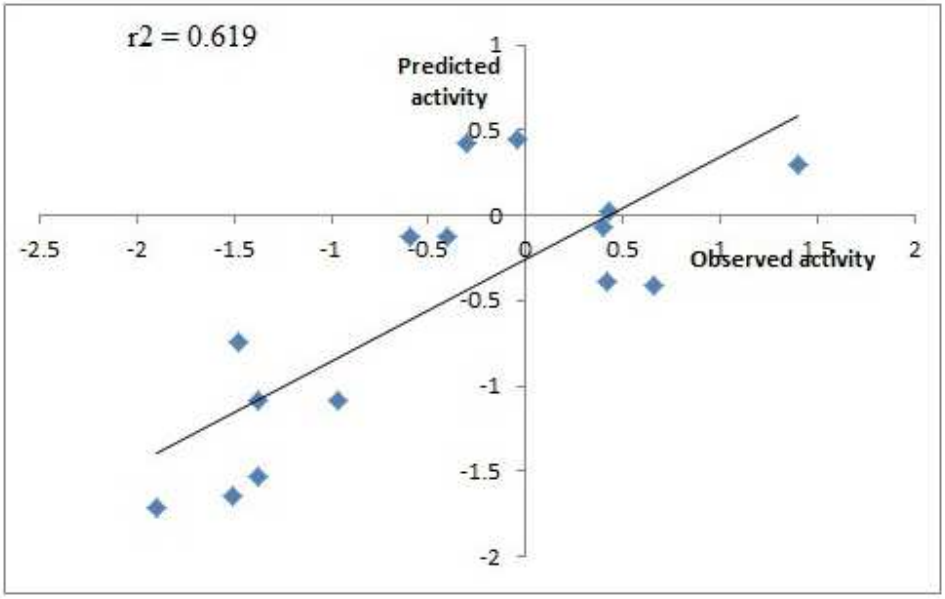


Figure 2: Observed activity vs. predicted activity of the test molecules (Using model 2 based on constitutional and geometrical descriptors)

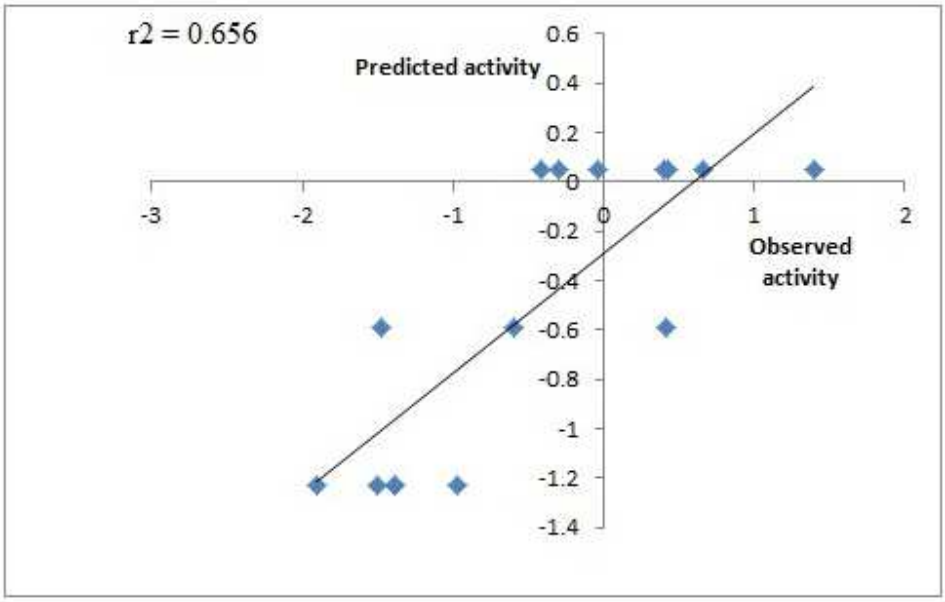


Figure 3: Observed activity vs. predicted activity of the test molecules
(Using model 3 based on atom centered fragment descriptors)

For the inhibition of U373 glioma cancer cell, topological indices contributethe highest significant impact on the biological activity. Topological indices such as average valence connectivity index of order 3 (X3Av) and Eigenvalue 03 from edge adjacency matrix weighted by dipole moments (EEig03d) with higher positive regression coefficients values produce higher level of significant inhibition of the cancer cells followed by Eigenvalue 13 from edge adj. matrix weighted by edge degrees (EEig13x) and Moran autocorrelation of lag 5 weighted by mass (MATS5m) respectively. The descriptors including Eigenvalue 13 from edge adj. matrix weighted by dipole moments (EEig13d) and Maximal electrotopological negative variation (MAXDN) can negatively contribute on the biological activities.Constitutional and geometrical description based model describe the positive effect of only one significant parameter as depicted byQyy COMMA2 value / weighted by atomic masses (QYYm).

From the QSAR models 4-6 developed for Hs683 cell line, it is cleared that topological descriptors can explain and predict 83.0% and 78.0% of variances of the in vitro glioma cancer cell inhibitory activities of the studied compounds. This model can also produce 43.1% external predictability.Constitutional and geometrical descriptors can explain and predict 80.2% and 75.0% of variances of the activities of the studied compounds. This model can produce 69.7% external predictability whereas functional group and atom centered descriptors can explain 64.2% of the variances and can produce 59.6% of the internal and 59.1% of the external predictability respectively. Again the training QSAR models 4-6 are used to predict anticancer activities against Hs683 cancer cell line for the same test set compounds.

The predicted activities for the test compounds utilizing models 4-6 are plotted against observed activities. Here also it is marked that the predicted activities of all the compounds in the test set are almost aligned with their corresponding observed activitiesas the square correlation coefficients (r2) between experimental vs predicted activities of the test set molecules using QSAR models 4-6 shows 0.483, 0.720 and 0.597 respectively.

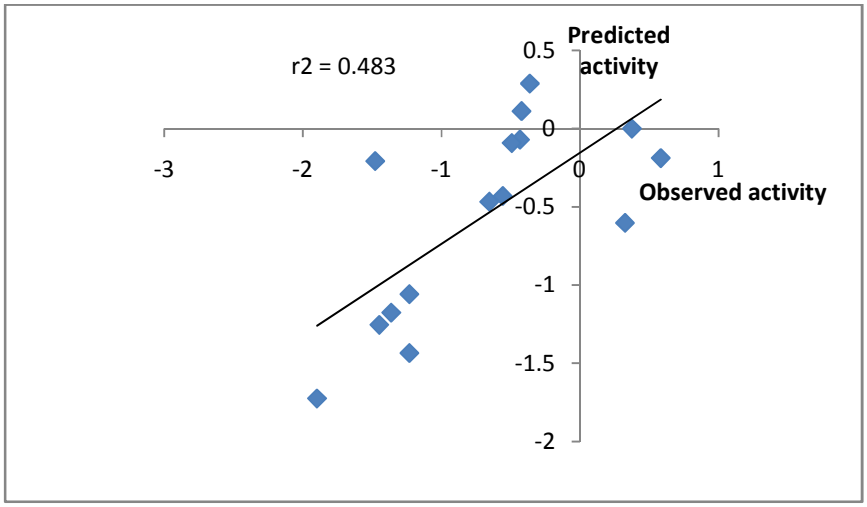


Figure 4: Observed activity vs. predicted activity of the test molecules (Using model 4 based on topological indices)

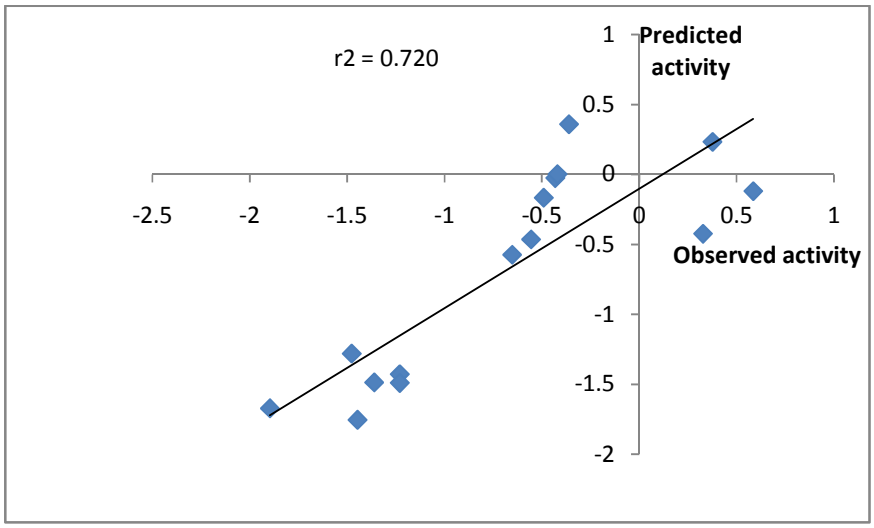


Figure 5:Observed activity vs. predicted activity of the test molecules (Using model 5 based on constitutional and geometrical descriptors)

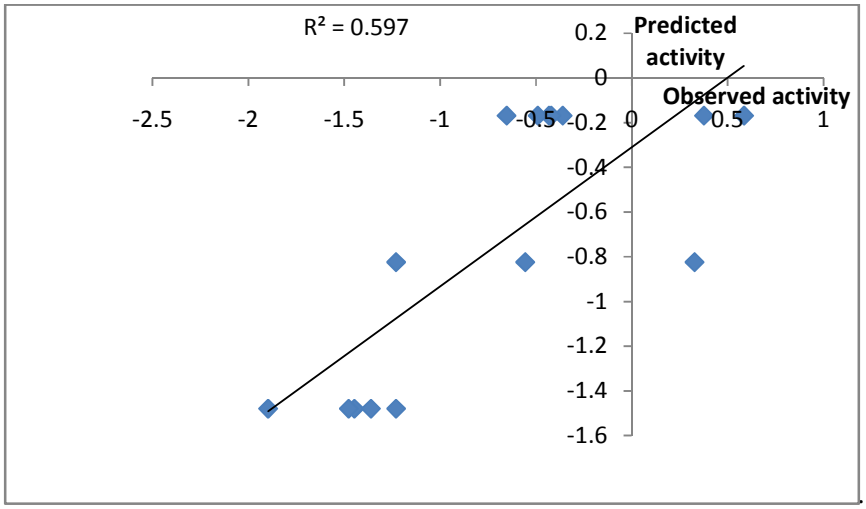


Figure 6: Observed activity vs. predicted activity of the test molecules
(Using model 6 based on functional and atom centered fragment descriptors)

Predicted activities for the test compounds generated by the variousQSARs utilizing different set of topological, constitutional and geometrical, functional group and atom centered fragmentsdescriptors respectively are given in Table 4.

Table 4: Predicted activities for test set compounds using the developed models

Test compounds	U373 Glioma cancer cell line				Hs683 Glioma cancer cell line			
	Observed activity	Predicted activity (using model 1)	Predicted activity (using model 2)	Predicted activity (using model 3)	Observed activity	Predicted activity (using model 4)	Predicted activity (using model 5)	Predicted activity (using model 6)
2	-1.505	-1.368	-1.647	-1.232	-1.447	-1.252	-1.755	-1.480
3	-1.380	-0.880	-1.534	-1.232	-1.361	-1.175	-1.487	-1.480
7	-0.968	-1.100	-1.088	-1.232	-1.230	-1.057	-1.429	-1.480
11	-1.380	-1.286	-1.086	-1.232	-1.477	-0.207	-1.281	-1.480
14	-1.903	-0.794	-1.713	-1.232	-1.897	-1.723	-1.673	-1.480
17	-1.477	-0.232	-0.739	-0.590	-1.230	-1.433	-1.491	-0.824
21	-0.591	-0.594	-0.120	-0.590	-0.556	-0.430	-0.466	-0.824
25	-0.301	-0.098	0.429	0.051	-0.361	0.288	0.356	-0.168
28	-0.408	0.014	-0.125	0.051	-0.431	-0.070	-0.027	-0.168
32	-0.045	0.363	0.442	0.051	-0.653	-0.466	-0.575	-0.168
36	0.431	0.794	0.025	0.051	-0.420	0.111	0.0003	-0.168
38	1.397	0.909	0.292	0.051	-0.491	-0.091	-0.168	-0.168
40	0.657	-0.019	-0.410	0.051	0.585	-0.187	-0.121	-0.168
45	0.420	-0.213	-0.385	-0.590	0.327	-0.602	-0.425	-0.824
47	0.397	0.156	-0.062	0.051	0.376	0.0001	0.231	-0.168

For the inhibition of Hs683 glioma cancer cell line, topological indices such asMoran autocorrelation - lag 4 / weighted by atomic polarizabilities (MATS4p) and shape profile number 05 (SP05) are responsible for producing positive impact on the cancer inhibiting activities whereas second Mohar index TI2 (TI2) and Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities (MATS3e) can negatively influence on the biological activities. Two parametric constitutional and geometrical based description model capture Qyy COMMA2 value / weighted by atomic polarizabilities(QYYp) as positive effect and radius of gyration(mass weighted) (RGyr) as negative effect on the anticancer activities.

Single parametric model is given by functional group and atom centered descriptors where no. of hydrogen bond donor (nHDon) is the main feature for inhibition of both U373 and Hs683 glioma cell lines. Decreasing the value of Hydrogen bond donor may increase the inhibition. The different significant modeled parameters have been interpreted in Table 5.

Table 5: Interpretation of the significant modeled predictors

Descriptor type	Abbreviation of the modeled descriptor	Modeled predictors
Topological	EEig03d	Eigenvalue 03 from edge adj. matrix weighted by dipole moments
	X3Av	Average valence connectivity index of order 3
	EEig13d	Eigenvalue 13 from edge adj. matrix weighted by dipole moments
	EEig13x	Eigenvalue 13 from edge adj. matrix weighted by edge degrees
	MATS5m	Moran autocorrelation of lag 5 weighted by mass
	MAXDN	Maximal electrotopological negative variation
	SP05	shape profile no. 05
	MATS4p	Moran autocorrelation - lag 4 / weighted by atomic polarizabilities
	TI2	second Mohar index TI2
	MATS3e	Moran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities
Constitutional and Geometrical	QYYm	Qyy COMMA2 value / weighted by atomic masses
	QYYp	Qyy COMMA2 value / weighted by atomic polarizabilities
	RGyr	radius of gyration (mass weighted)
Functional Group and atom centered descriptors	nHDon	Number of hydrogen bond donor

CONCLUSION

QSAR modelling for 48 β-caboline compounds having anticancer activities by the inhibition of DYRK1A mediated caspase-9 activation is performed in the present article considering a large number of theoretical molecular descriptor including topological, constitutional and geometrical,functional group and atom centered fragment descriptors respectively. From our point of view it is a novel attempt to predict the significant predictor properties responsible for producing anticancer activities of these congeners against different glioma cancer cell lines including U373 and Hs683. QSAR modeled parameters have been interpretated and it is represented that the significant variables including X3Av, EEig13x, MATS5m and SP05 correspond size, shape and weight of the molecules whereas EEig03d, MATS4p and QYYp are correlated with the dipole moment, polarizabilities, conjugation and aromaticity of the molecules. Increasing the values of the above parameters may enhance the cancer inhibitory activities of the β-carboline compounds.The QSAR model obtained by using functional group and atom centered descriptors has predicted hydrogen bond donor that represents electrostatic interaction between the ligand and receptor molecules.This is to convey that there is no specific theoretical modeling for these compounds so far as done, therefore, studies in this direction for exploration of essential structural features of the β-carboline congeners under the frame work of computed molecular descriptors may help to design more potent and active lead like β-carboline derivative which will be developed for the use of different cancers.

Acknowledgement
SN is sincerely thankful to National Institute of Chemistry, Slovenia for availing DRAGON Professional version 5.4-2006 software for the calculation of theoretical molecular descriptors used in the present work. Authors are sincerely thankful to GIPER for providing necessary research facilities. Ruby shows deep sense of gratitude to her supervisor Dr. Nandi.

REFERENCES

[1] PB Jensen; T. Hunter,*Nature*, **2001**, 411, 355–365.
[2]LA Allan; N Morrice; S Brady; G. Magee; SPathak;PR Clarke, *Nat. Cell. Biol*., **2003**, 5, 647–654.
[3]P Savage; J Stebbing; M Bower; T Crook,*Nat. Clin. Pract.Oncol*, **2009**, 6, 43−52.
[4]TR Wilson; PG Johnston; DB Longley, *Curr.Cancer Drug Target*, **2009**, 9, 307−319.
[5]RS Wong, *J. Exp. Clin. Cancer Res*., **2011**, 30, 87-100.
[6]Y Ma; M Wink, *Phytother. Res*.,**2010**, 24, 146−149.
[7]A Seifert; LA Allan; PR Clarke,*FEBS J.*,**2008**, 275, 6268−6280.
[8]A Laguna; S Aranda; MJ Barallobre; R Barhoum; E Fernandez; V Fotaki; JM Delabar; S de la Lu; P de la Villa; MLArbones,*Dev. Cell*., **2008**, 15, 841−853.
[9]NRD Wit; HJ Burtscher; UH Weidle; DJ Ruiter; GN van Muijen,*Melanoma Res*., **2002**, 12, 57−69.
[10]A Seifert; LA Allan;PR Clarke, *FEBS J.*,**2008**, 275, 6268–6280.
[11]NGockler; G. Jofre; C Papadopoulos; U Soppa; FJTejedor;W Becker, *FEBS J.*,**2009**, 276, 6324–6337.
[12]R Cao; H Chen; WPeng; YMa; X Hou; H Guan; X Liu;AXu,*Eur. J. Med. Chem*., **2005**, 40, 991–1001.

[13]RCao;WYi;QWu;XGuan;MFeng;CMa; ZChen; HSong;WPeng,*Bioorg.Med. Chem.Lett*, **2008**, 18, 6558–656.
[14]Y Ma, M Wink, *Phytother. Res.*, **2010**,24, 146–149.
[15]J Ishida; HK Wang; KF Bastow; CQ Hu; KH Lee,*Bioorg. Med. Chem. Lett.*,**1999,** 9, 3319–3324.
[16]R Frederick; C Bruyere; C Vancraeynest; J Reniers; C Meinguet; L Pochet; A Backlund; B Masereel; R Kiss;JWouters,*J. Med. Chem.*,**2012**, 55, 6489-501.
[17]N Mills, ChemDraw Ultra 10.0.,**2006**, 128, 1349-13650.
[18]EEstrada In: J Devillers, AT Balaban (eds) Gordon and Breach, Amsterdam, **1999**, pp 403–453.
[19]M. Randic, Novel shape descriptors for molecular graphs,*J. Chem. Inf.Comput. Sci*., **2001**, 41, 607–613
[20]MRandic, *J. Am. Chem. Soc.*, **1975**, 79, 6609–6615
[21] SNandi;MCBagchi,*Med. Chem. Res*., **2007**, 15, 393–406.
[22]SCBasak, *Curr. Comput. Aided Drug Des.*,**2013**, 4, 449-462.
[23]RTodeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Revised and Enlarged Edition. 2nd edition.Weinheim: Wiley-VCH, **2009**.
[24]ARKatritzky; R Petrukhin; D Tatham; S Basak; EBenfenati; M Karelson; U. Maran,*J. Chem. Inf.Comput. Sci*.,**2001**, 41, 679–685.
[25]A Tropsha; P Gramatica; VJ Gombar,*QSAR Comb. Sci.*,**2003**, 22, 69–77.
[26]NRDraper; H Smith, 3rd Ed., John Wiley & Sons, New York, **1998**.
[27] Minitab® Statistical Software: Minitab. ; **2010**. www.minitab.com (accessed October 29th 2013).
[28]S Nandi; MCBagchi, *Mol.Simul.*,**2011**, 37, 196 — 209.
[29]PP Roy; K. Roy, *Eur. J. Med. Chem*., **2009**, 44, 2913-2922.
[30]K Roy; PChakraborty; I Mitra; PK Ojha; SKar; RN Das, *J.Comput. Chem*., **2013**, 34, 1071-1082.