

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227748378>

2D and 3D Quantitative Structure–Activity Relationship Studies on a Series of bis-Pyridinium Compounds as Choline Kinase Inhibitors

ARTICLE in QSAR & COMBINATORIAL SCIENCE · SEPTEMBER 2006

Impact Factor: 1.55 · DOI: 10.1002/qsar.200530199

CITATIONS

26

READS

69

3 AUTHORS:



Sridhara Janardhan

Indian Institute of Chemical Technology

11 PUBLICATIONS 94 CITATIONS

SEE PROFILE



Srivani Palukuri

8 PUBLICATIONS 204 CITATIONS

SEE PROFILE



G Narahari Sastry

Indian Institute of Chemical Technology

262 PUBLICATIONS 5,153 CITATIONS

SEE PROFILE

2D and 3D Quantitative Structure-Activity Relationship Studies on a Series of *bis*-Pyridinium Compounds as Choline Kinase Inhibitors

S. Janardhan, P. Srivani and G. Narahari Sastry*

Molecular Modeling Group, Organic Chemical Sciences, Indian Institute of Chemical Technology, Tarnaka, Hyderabad-500007, India, Tel/fax: +91-4027160512; e-mail: gnsastry@iict.res.in (Dr. G. N. Sastry), IICT communication number: 050826

Keywords: 2D-QSAR; CoMFA; CoMSIA; ChoK; *bis*-pyridinium derivatives

Received: November 28, 2005; Accepted: February 17, 2006

DOI: 10.1002/qsar.200530199

Abstract

Two-dimensional (2D) and three-dimensional (3D) quantitative structure activity relationship (QSAR) studies have been carried out on a series of 55 *bis*-pyridinium compounds to find out the structural requirements of choline kinase (ChoK) inhibitors. The best predictions were obtained from the model where 44 compounds were considered in the training set and the remaining 11 in the test set. The heuristic and BMLR methods resulted with r^2 and q^2 values of 0.86, 0.83 and 0.87, 0.84 respectively. The obtained Fisher and S^2 values for both the methods are 49.02 and 54.05, 0.053 and 0.0397 respectively. The best model for 3D-QSAR has been obtained with $r^2=0.97$, $q^2=0.58$ and $r^2_{\text{pred}}=0.68$ when CoMFA fields were used. The r^2 of 0.85, q^2 of 0.55 and $r^2_{\text{pred}}=0.66$ have been observed when CoMSIA fields were used. The results that are obtained from 2D and 3D-QSAR studies may provide useful insights into the roles of various substitution patterns on the *bis*-pyridinium skeleton and may also help to design more potent compounds.

1 Introduction

Choline kinase (ChoK) is a cytosolic enzyme that has involved in biosynthesis of a membrane glycerophospholipid, phosphatidylcholine (PC) through the Kennedy pathway [1, 2]. During this process, ChoK catalyzes the phosphorylation of choline to form phosphorylcholine (PCho) in the presence of ATP and magnesium [3–5]. Hence, ChoK appears to be crucial to initiate the PC biosynthesis as well as to release the essential lipid second messengers like PCho and diacylglycerol to transmit the mitogenic signals [6]. The generation of PCho is considered to be one of the crucial steps in regulating growth factor stimulated cell proliferation, malignant transformation, invasion and metastasis [7–9]. It has been reported that the deregulation of *ras* oncogene activity by mutation will drastically effect the functioning of intracellular kinases, nuclear transcription factors and growth factors, which in turn lead to tumorigenesis [2]. In relation to the above reports, an altered level of PCho, due to enhanced action of ChoK, has been observed in both *ras* dependent and *ras* independent cancers such as breast, lung, colon, prostate, and diverse murine tumours [10–12]. Furthermore, it has been demonstrated that the generation of PCho from choline by ChoK is an essential process for growth factors to exert

their mitogenic action. As a result, ChoK has been emerged as a novel target for anticancer therapy in recent times [13, 14].

The hemicholinium-3 (HC-3), a well-known choline uptake blocker, has been served as a potent ChoK inhibitor by inhibiting the effects of growth factors on mitogen induced DNA synthesis. However, it exhibits few drastic toxic effects on cholinergic nerve terminals like potent respiratory paralysis [15, 16]. Hence, to overcome these undesirable side effects new series of compounds such as *bis*-pyridinium, *tris*-pyridinium, *bis*-quinolinium, *bis*-isoquinolinium derivatives have been proposed. A set of 55 *bis*-pyridinium analogues, which were derived from HC-3, has been considered for the present computational study [17–24].

Developing quantitative structure activity relationships using various physicochemical parameters has been an important task in lead optimization. In the present study, we have employed 2D-QSAR tools along with two important 3D-QSAR tools, CoMFA and CoMSIA. The 3D-QSAR protocols have been selected not only because of the robustness of the model it produces but the ensuing steric and electrostatic maps may further aid to understand the nature of the ligand interactions with the active site of the enzyme. The CoMSIA method differs from CoMFA by

the way the molecular fields are calculated and also by including the molecular fields, such as lipophilic and hydrogen bond potentials. Hence, it has been chosen to get additional information about hydrophobicity and about hydrogen bond fields. Using these molecular modeling approaches the study further delves on the structural requirements of this class of compounds towards ChoK enzyme.

2 Computational Methods

2.1 Data set for analysis

Reported *ex-vivo* data of a set of 55 *bis*-pyridinium analogues was considered for the present study (Table 1) [17–24]. The inhibitor activities (IC_{50}) against human ChoK and anti-proliferative activities (AA) were converted in to pIC_{50} according to the formula $pIC_{50} = -\log (IC_{50})$.

2.2 Molecular modeling

The molecular modeling calculations were performed using SYBYL 6.9 version installed on a Linux operating system [25]. The energy minimizations were performed using MMFF94 force field and MMFF94 partial atomic charges with distance dependent dielectric constant and Powell's conjugate gradient method was used until a convergence criterion of 0.001 kcal/mol was reached. The most significant requisite for any 3D-QSAR studies was to align the dataset on a suitable conformational template, either by taking a reported crystal structure of a bioactive compound or by considering the most active compound. In the present study, since, we don't have any reported crystal structure we considered the most active compound as a template for the alignment. The *N*-benzyl pyridinium moiety of the bioactive molecule (**20**) was used as a substructure and the rest of the molecules were aligned to it using

Table 1. The series of *bis*-pyridinium analogs and their activities in pIC_{50} .

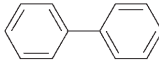
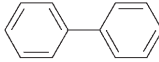
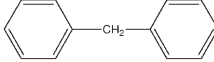
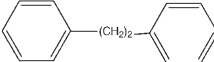
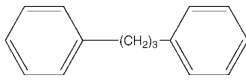





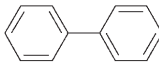
S. No.	R_4	Spacer isomer	Z	pIC_{50}	
				ChoK	AA
1.	–H	3', 3''		4.77	4.52
2.	–H	4', 4''		4.64	4.52
3.	–H	4', 4''		4.51	4.15
4.	–H	4', 4''		4.51	4.22
5.	–H	4', 4''		4.60	4.82
6.		3', 3''		4.07	4.30
7.		4', 4''		4.62	–
8.		3', 3''		5.24	5.58

Table 1. (cont.)

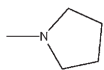
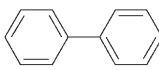
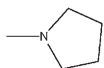
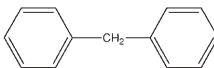
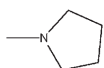
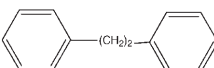
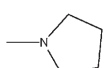
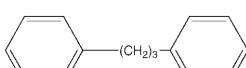
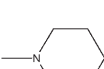

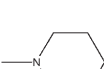

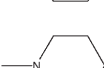
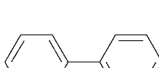
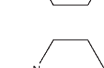
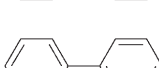
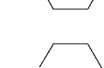

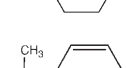

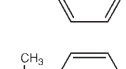

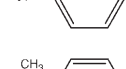
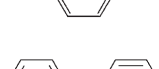
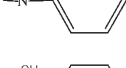
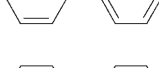
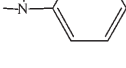
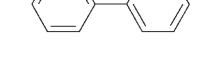
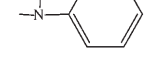

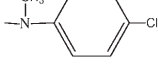

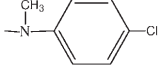
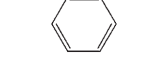
S. No.	R ₄	Spacer isomer	Z	pIC ₅₀	
				ChoK	AA
9.		4', 4''		4.95	5.22
10.		4', 4''		4.22	6.00
11.		4', 4''		4.61	6.00
12.		4', 4''		4.60	6.30
13.		3', 3''		4.28	4.37
14.		4', 4''		4.42	—
15.		3', 3''		5.72	5.80
16.		4', 4''		5.05	5.38
17.		4', 4''		5.02	6.40
18.		3', 3''		4.43	5.30
19.		4', 4''		4.80	5.16
20.		3', 3''		6.37	5.66
21.		4', 4''		5.37	5.70
22.		4', 4''		5.19	6.47
23.		3', 3''		5.14	5.48
24.		4', 4''		5.54	5.46
25.		3', 3''		5.82	5.72

Table 1. (cont.)

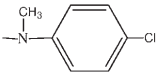
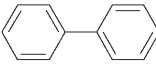
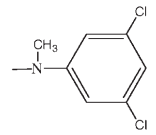

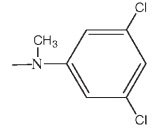

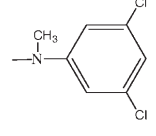
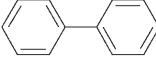
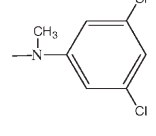
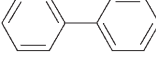
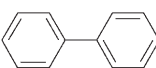
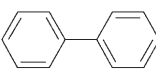
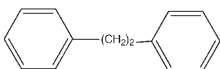
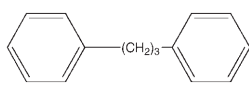
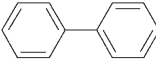
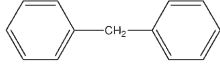
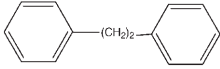
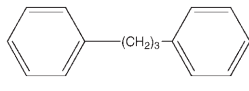
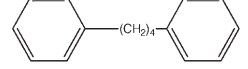
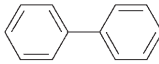
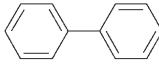
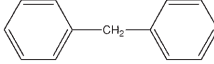
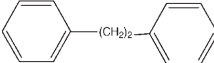
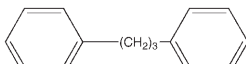
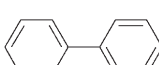
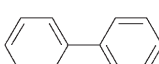
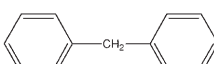
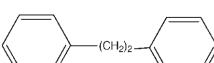
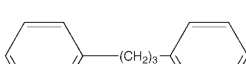
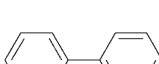




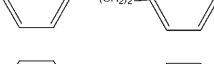
S. No.	R ₄	Spacer isomer	Z	pIC ₅₀	
				ChoK	AA
26.		4', 4''		5.74	6.30
27.		3', 3''		5.08	5.40
28.		4', 4''		5.01	5.54
29.		3', 3''		5.59	5.75
30.		4', 4''		5.47	6.19
31.	-NH ₂	3', 3''		4.51	—
32.	-NH ₂	4', 4''		4.41	4.40
33.	-NH ₂	4', 4''		4.64	4.40
34.	-NH ₂	4', 4''		4.00	5.15
35.	-N(CH ₃) ₂	4', 4''		4.22	4.70
36.	-N(CH ₃) ₂	4', 4''		4.16	5.30
37.	-N(CH ₃) ₂	4', 4''		4.77	5.70
38.	-N(CH ₃) ₂	4', 4''		4.66	5.60
39.	-N(CH ₃) ₂	4', 4''		4.66	6.22

Table 1. (cont.)

S. No.	R ₄	Spacer isomer	Z	pIC ₅₀	
				ChoK	AA
40.	–CH ₃	3', 3''		4.40	–
41.	–CH ₃	4', 4''		4.00	4.70
42.	–CH ₃	4', 4''		4.48	4.40
43.	–CH ₃	4', 4''		4.00	4.30
44.	–CH ₃	4', 4''		4.09	5.22
45.	–CH ₂ OH	3', 3''		4.00	4.00
46.	–CH ₂ OH	4', 4''		4.22	–
47.	–CH ₂ OH	4', 4''		4.05	–
48.	–CH ₂ OH	4', 4''		4.00	4.00
49.	–CH ₂ OH	4', 4''		4.46	–
50.	–COCH ₃	3', 3''		4.46	4.00
51.	–COCH ₃	4', 4''		4.19	–
52.	–CN	4', 4''		4.12	–
53.	–COOH	4', 4''		3.86	–
54.	–NC ₆ H ₁₂	4', 4''		4.82	6.40
55.	–N(allyl) ₂	4', 4''		4.77	6.26

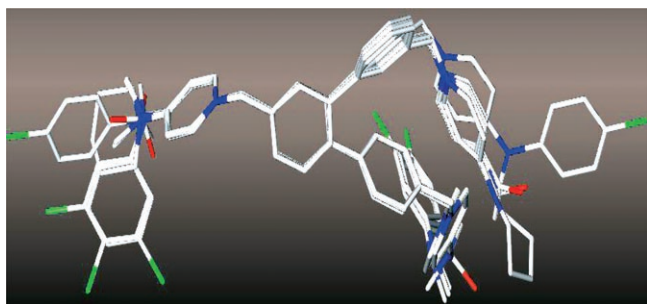


Figure 1. Database alignment of the dataset.

two alignment protocols first, the database alignment method second, atom-fit method i.e. searching the best fit between atom of the template to the corresponding atoms of the molecules to be aligned. The aligned data set molecules were shown in Figure 1. For 2D-QSAR the above constructed molecules were taken and optimized using AMPAC 5.0 at AM1 level and considered for further studies [26].

2.3 Calculating 2D-descriptors

CODESSA (COMprehensive DESCriptors for Structural and Statistical Analysis) version 2.0 was used for calculating 2D descriptors as well as for regression analysis [27]. It was supported by AMPAC 5.0 to provide the necessary data to calculate quantum mechanical descriptors. Initially, a total of 540 default descriptors were calculated and these descriptors were further classified into following groups viz. constitutional, topological, geometrical, electrostatic, quantum-chemical and thermodynamical descriptors.

The heuristic and best multi-linear regression (BMLR) methodologies were used to yield the QSAR equation and to validate the predictive ability of the model using test set [27]. Both heuristic and BMLR techniques belong to multilinear regression methods, and they differ in the selection criteria of descriptors and of one- and two- parameter correlations. The selection of descriptors is based on F-test's value and R_{\min}^2 of one-parameter correlation, and the parameter's t-value and intercorrelation among the descriptors are also been considered for the descriptor selection in the heuristic method. In contrast, in BMLR, all orthogonal pairs of descriptors are sorted out then the two-parameter regressions are carried out. The N_c (≤ 400) pairs with highest regression correlation coefficients are chosen for performing the higher-order regressions. A final model is obtained by iteratively increasing the F value, via a three parameter regressin model [27].

2.4 Generation of CoMFA and CoMSIA fields

The CoMFA steric and electrostatic potential fields were calculated at each lattice intersection of a regularly spaced

grid of 2.0 Å in all the three dimensions with in the defined region [28]. The van der Waals potential (Lennard-Jones 6–12) and columbic term, which represent steric and electrostatic fields respectively, were calculated using a sp^3 carbon atom with a radius of 1.52 Å and +1.0 charge. The steric and electrostatic contributions were truncated to ± 30 kcal/mol and electrostatic contributions were ignored at the lattice intersections with maximum steric interactions. CoMSIA fields were derived according to Klebe et al by employing a probe atom with a radius of 1.0 Å and +1 charge, hydrophobicity of +1 and hydrogen bond donor and acceptor properties of +1 [29]. The corresponding CoMSIA fields (A_F) for a molecule j with atoms i at a grid point q and at distance r was calculated by using the following equation.

$$A_{F,k}^q(j) = \sum_i w_{probe,k} w_{ik} e^{-\alpha r_{iq}^2}$$

Five CoMSIA fields (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were calculated using the probe atom. A Gaussian type distance dependence function was used between the grid point q and each atom i of the molecule. The default attenuation factor of 0.3 was used. In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from atomic partial charges, the hydrophobic fields are derived from atom-based parameters developed by Viswanadhan et al., and the hydrogen bond indices are obtained by a rule-based method derived from experimental values [30].

2.5 Partial Least Squares (PLS) analysis

PLS analysis was used to derive a linear correlation between the 3D-fields (independent variables) and the inhibitory activity values (dependent variables) [31]. The cross-validation analysis was performed using leave-one-out (LOO) method in which one compound was removed from the dataset and its activity was predicted using the model built from rest of the dataset. The cross-validation run returns the optimum number of components for which it has maximum cross-validated r^2 (q^2) and minimum standard error of prediction (SEP). So as to overcome the over-fitting problem, the regular run was performed using the same optimum number of components that resulted from the cross-validation run. Equal weights were assigned to steric and electrostatic fields using CoMFA-standard scaling option. To speed up the analysis and to reduce the noise, a minimum filter value (σ) of 2.0 kcal/mol was used. To validate the derived CoMFA and CoMSIA models, biological activities of a test set were predicted using the models derived from the training set. The predictive ability of the models was measured by means of predictive r^2 (r_{pred}^2) value. To further assess the statistical validity and robustness of the derived 3D-QSAR models, bootstrapping anal-

ysis (100 runs) and group cross-validation run for 50 times with 11 groups were performed.

3 Results and Discussion

In this section we start with a description of how various 2D and 3D-QSAR models are constructed followed by the presentation of 2D-QSAR and CoMFA and CoMSIA results and discussion. In 2D-QSAR two-regression methods, heuristic (Model I) and BMLR (Model II), were used to build the QSAR equation. As two approaches have their own significance, both the models were considered for further analysis. As specified previously, for 3D-QSAR, we have considered both the activities, inhibitory activity against ChoK and AA on HT-29 cultured cells. Since AA values are not available for compounds **31**, **40**, **46**, **47**, **51** and **52** and the molecules **7**, **14**, **49**, **53** are having uncertain values ($>100-1000$), all the 10 molecules were ignored and only 44 molecules, 36 in training set and remaining 8 in test set, were considered. The AA values are taken as a supplementary parameter to gauge the structural requirements in more detail. For CoMSIA, four models (Model IV–VII) were built by changing the combination of five fields (S, E, H, D, A). Model IV that includes all the five fields was considered for further analysis since it was providing the most descriptive information compared to other three models.

3.1 2D-QSAR

While the correlation between various descriptors with biological activity is the most important means of any structure-activity relationship study, the knack lies in deciding when to stop adding descriptor to the model. Thus, the best equation should use the minimum number of descriptors to obtain the best fit. To achieve this, a popular procedure is to find out the saturation point, a point beyond which there is no considerable improvement in regression coefficient (r^2 and q^2) values has observed. The plot of r^2 and q^2 vs number of descriptors that are obtained from the BMLR method represents the optimum number of descriptors (five) required for the best model (Figure S1). By interpreting the resulted 2D-QSAR descriptors, it is possible to gain some insight into factors that are likely to govern the inhibitory activity. The contribution of each descriptor can be validated by means of its t-value since it reflects the significance of the parameter within the model, i.e. part of the variance explained by that particular parameter.

The models I and II yielded r^2 and q^2 values of 0.86, 0.83 and 0.87, 0.84, respectively. The obtained Fisher and S^2 values for both models are 49.0 and 54.1, 0.053 and 0.0397 respectively. The predicted and actual activities of the training and test set molecules were given in Table 2 the graph was displayed in Figure 6 and descriptors, coefficients and

t-values for both the methods were provided in supplementary data.

The model I can be represented by the following Eq. 1:

$$\text{Activity} = 0.2231 * \text{RI} - 0.5214 * \text{MASEC} - 2.0812 * \text{YZSYZR} + 3.5242 * \text{MTICH} + 13.8150 * \text{MERIN} - 19.0357 \quad (1)$$

While, the model II can be represented by the following Eq. 2:

$$\text{Activity} = -0.0775 * \text{NC} - 3.2134 * \text{YZSYZR} + 4.6723 * \text{MTICH} - 6398.8 * \text{ESPHASA} + 1307.4 * \text{MERIC} - 54.7880 \quad (2)$$

The ten descriptors that are obtained from the two regression methods can be classified as follows: topological (RI- Randic index), geometrical (YZSYZR- YZ Shadow/YZ rectangle) constitutional (NC- Number of C atoms), and the remaining are quantum mechanical descriptors; MASEC (min atomic state energy for a C atom), MTICH (min total interaction for a C–H bond), MERIN (max electrophilic reactivity index for a N atom), ESPHASA (ESP-H-acceptors surface area) and MERIC (max electrophilic reactivity index for a C atom).

The topological descriptors represent the atomic connectivity of the molecular structure as graphs, atoms as vertices and covalent bonds as edges. According to Randic [32], Kier and Hall [33] these descriptors represent the additive value of the increments assigned to the specific structural fragments in terms of order and sub graph type such as path, cluster, path/cluster and chain. Randic index is calculated using the formula

$${}^n\chi = \sum_{(\text{Path of length } n)}^{N_{SB}} (\delta_{i1} \dots \delta_{in+1})^{-1/2}$$

Where δ_i and δ_j ($i \neq j$) correspond to the coordination numbers of atoms (Randic index). The Randic index (order-3) explains the amount of branching ring structures and their flexibility where the molecules have any paths of length 3, which would require three edges (three bonds). The t-value of 9.0455 infers that, the connectivity of heteroaromatic *bis*-pyridinium compounds with methylene linker as well as R_4 substituents like, N-alkyl amine connectivity and cycloalkyl amines connectivity are crucial for the activity.

The number and maximum electrophilic reactivity index of carbon atoms characterizes the constitution and electrophilic interactions of carbon atoms. The positive coefficient represents their importance towards the activity.

The descriptor YZ Shadow/YZ rectangle, a geometric descriptor, explains the fraction of area of molecular shadow in the YZ plane over area of enclosing rectangle (S_{yz} , f). These shadow indices gave an idea about size, shape and orientation of molecular surface in the space along the

Table 2. Experimental, predicted inhibitory activities (pIC_{50}) and residuals of training and test set (bold) for 2D and 3D QSAR models.

S. No.	Exp.	Predicted				Residuals			
		M I	M II	M III	M IV	M I	M II	M III	M IV
1.	4.77	4.97	4.98	4.77	4.71	0.20	0.21	0.00	0.06
2.	4.64	4.30	4.26	4.33	4.29	− 0.34	− 0.38	0.31	0.35
3.	4.51	4.57	4.42	4.63	4.15	0.06	−0.09	−0.13	0.36
4.	4.51	4.67	4.55	4.56	4.29	0.16	0.04	−0.05	0.22
5.	4.60	4.66	4.57	4.53	4.34	0.06	−0.03	0.07	0.26
6.	4.07	4.18	4.26	4.27	4.32	0.11	0.19	− 0.20	− 0.25
7.	4.62	4.18	4.18	4.60	4.57	− 0.44	− 0.44	0.01	0.05
8.	5.24	5.19	5.39	5.22	5.20	−0.05	0.15	0.02	0.04
9.	4.95	4.92	4.94	5.00	4.76	−0.03	−0.01	−0.04	0.19
10.	4.22	4.39	4.28	4.24	4.35	0.17	0.06	−0.02	−0.13
11.	4.61	4.86	4.89	4.66	4.76	0.25	0.28	−0.05	−0.15
12.	4.60	4.86	4.88	4.69	4.64	0.26	0.28	− 0.09	− 0.04
13.	4.28	4.49	4.74	4.30	4.44	0.21	0.46	−0.02	−0.16
14.	4.42	4.27	4.35	4.47	4.67	−0.15	−0.07	−0.05	−0.25
15.	5.72	5.24	5.48	5.82	5.44	−0.48	−0.24	−0.10	0.28
16.	5.05	5.03	5.08	5.11	4.84	−0.02	0.03	−0.05	0.22
17.	5.02	5.01	5.08	5.03	4.91	−0.01	0.06	−0.01	0.10
18.	4.43	4.71	4.74	4.42	4.61	0.28	0.31	0.01	−0.18
19.	4.80	4.72	4.67	4.94	5.06	−0.08	−0.13	−0.12	−0.23
20.	6.37	5.77	5.97	6.29	5.87	−0.60	−0.40	0.08	0.50
21.	5.37	5.31	5.22	5.40	5.42	−0.06	−0.15	−0.03	−0.50
22.	5.19	5.40	5.35	5.22	5.02	0.21	0.16	−0.03	0.17
23.	5.14	5.04	4.93	4.80	4.67	− 0.10	− 0.21	0.34	0.47
24.	5.54	5.26	5.21	5.36	5.29	−0.28	−0.33	0.18	0.24
25.	5.82	5.93	5.90	5.86	5.88	0.11	0.08	−0.04	−0.06
26.	5.74	5.56	5.32	5.59	5.56	−0.18	−0.42	0.16	0.19
27.	5.08	5.19	5.08	5.08	4.77	0.11	0.00	0.00	0.31
28.	5.01	5.11	4.90	5.22	5.29	0.10	−0.11	−0.21	−0.27
29.	5.59	5.96	5.88	5.54	5.87	0.37	0.29	0.05	−0.28
30.	5.47	5.73	5.48	5.49	5.62	0.26	0.01	−0.02	−0.15
31.	4.51	4.61	4.68	4.37	4.39	0.10	0.17	0.14	0.12
32.	4.41	4.55	4.49	4.38	4.45	0.14	0.08	0.13	−0.05
33.	4.64	4.33	4.22	4.45	4.81	−0.31	−0.42	0.18	−0.17
34.	4.00	4.35	4.23	4.15	4.16	0.35	0.23	−0.15	−0.16
35.	4.22	4.30	4.39	4.35	4.56	0.08	0.17	−0.13	−0.33
36.	4.16	4.21	4.37	4.15	4.29	0.05	0.21	0.00	−0.14
37.	4.77	4.40	4.56	4.77	4.59	−0.37	−0.21	0.00	0.18
38.	4.66	4.42	4.60	4.52	4.53	−0.24	−0.06	0.13	0.13
39.	4.66	4.58	4.80	4.59	4.63	−0.08	0.14	0.06	0.02
40.	4.40	4.31	4.36	4.34	4.76	−0.09	−0.04	0.06	−0.36
41.	4.00	3.90	3.82	4.08	4.36	−0.10	−0.18	−0.08	−0.36
42.	4.48	4.19	4.17	4.35	4.16	−0.29	−0.31	−0.14	0.32
43.	4.00	4.19	4.15	4.27	4.39	0.19	0.15	−0.27	−0.34
44.	4.09	4.27	4.26	4.20	4.35	0.18	0.17	−0.11	−0.26
45.	4.00	4.09	3.98	3.97	3.68	0.20	−0.02	0.03	0.32
46.	4.22	3.59	3.43	4.21	4.37	− 0.34	− 0.79	0.01	− 0.14
47.	4.05	4.06	3.96	4.15	4.14	0.06	−0.09	−0.10	−0.09
48.	4.00	4.11	4.01	4.36	4.05	0.16	0.01	− 0.36	− 0.05
49.	4.46	4.21	4.17	4.32	4.46	0.06	−0.29	0.13	−0.04
50.	4.46	4.78	4.90	4.70	4.42	0.11	0.44	− 0.24	0.04
51.	4.19	4.52	4.56	4.33	4.04	− 0.44	0.37	− 0.14	0.15
52.	4.12	4.35	4.21	4.01	4.12	−0.05	0.09	0.12	0.00
53.	3.86	3.85	3.88	3.77	3.74	−0.03	0.02	0.09	0.13
54.	4.82	4.96	4.99	5.01	5.10	0.17	0.17	− 0.18	− 0.28
55.	4.77	4.69	5.02	4.76	4.86	0.25	0.25	0.00	0.09

principle moments of inertia with Y and Z-axes. The t-value of -2.6963 (heuristic) and -4.5279 (BMLR) infers that the 3', 3'' spacer with biphenyl linker is preferred over 4', 4'' as the molecules with 3', 3'' spacer (**1**, **8**, **15**, **20**, **29**, **31**) are showing comparatively more activity than the 4', 4'' spacer (**2**, **9**, **16**, **21**, **30**, **32**).

The descriptor minimum atomic state energy for a C atom, a quantum mechanical energy related descriptor, illustrates magnitude of the perturbation experienced by C atom in the molecular environment in comparison to the isolated atom, which can be calculated as follows

$$E_{\text{tot}} = \sum \left| E_{\text{C}}(\text{C}) + E_{\text{exc}}(\text{C}) \right|$$

Where E_{C} is intramolecular electrostatic interaction energy and E_{exc} is two-center exchange energy of carbon [32]. The descriptor has a t-value of -5.7606 . It has a negative sign in the QSAR equation means decrease in E_{tot} of *para* and *meta* positions of the pyridinium ring may perhaps improves the inhibitory activity.

Another descriptor minimum total interaction for a C–H bond, a quantum mechanical energy-related descriptor, characterizes the total interaction energy between C and H atoms as defined below

$$E_{\text{tot}}(\text{CH}) = E_{\text{C}}(\text{CH}) + E_{\text{exc}}(\text{CH})$$

Where E_{C} is coulomb interaction energy between C and H and E_{exc} is electronic exchange energy between C and H, which reflects change in Fermi correlation energy between the two electrons localized on carbon and hydrogen, respectively and in turn determine the conformational changes and spin properties of the molecule. The descriptor is carrying a t-value of 5.3051 (heuristic) and 7.6559 (BMLR), the positive sign of the t-value suggests that the C–H interaction of the methylene moiety, linked between pyridinium ring and phenyl ring as well as $(\text{CH}_2)_{n=0-4}$ linker between two phenyl rings are playing an important role towards the inhibitory activity of ChoK.

The descriptor maximum electrophilic reactivity index for N atom ($t=4.4084$), a quantum-chemical descriptor, explains the electrophilic nature of the nitrogen atom. The positive sign of the t-value infers the importance of quaternary ammonium ion. Also, any group that can enhance the electrophilicity of the N atom may contribute more towards the inhibitory activity.

3.2 CoMFA

The best CoMFA model (model III) was obtained with 44 compounds in training set and remaining 11 compounds (**1**, **2**, **6**, **7**, **12**, **23**, **46**, **48**, **50**, **51**, **54**) in the test set. The model III has q^2 of 0.58 , r^2 of 0.97 and standard error of estimate (SEE) of 0.114 , which signifies a good correlation between observed and computed affinities of the training set.

Similarly, PRESS and r^2_{pred} values, 0.50 and 0.68 , of the model represents that the predictive ability of the model is considerably good. To further appraise the statistical validity and robustness of the derived 3D-QSAR models, bootstrapping analysis (100 runs) was performed and an r^2_{bs} of 0.99 was obtained. Further, to determine the true predictivity of the model, group cross-validation was performed with 11 groups for 50 times and the mean r^2_{cv} value of 0.50 reveals that the results were not based on chance correlation. The model that has been constructed with AA values results in r^2 of 0.99 and q^2 of 0.71 with seven components.

The results of PLS analysis are shown in Table 3. The contour maps were generated as scalar products of coefficients and standard deviation associated with each CoMFA column are shown in Figure 2. The CoMFA steric interactions are represented by green and yellow colored contours while electrostatic interactions are represented by red and blue colored contours. The bulky substituents are favored at green regions and unfavored at yellow regions. The increase in positive charge is favored in blue region while increase in negative charge is favored in red region. The experimental, predicted inhibitory activities and residuals for CoMFA training and test set were shown in Table 2 and graph of predicted vs. experimental was displayed in Figure 7.

A quick pursue of CoMFA maps reveals that the steric contribution is more than the electrostatic contribution. The green contour near the *para* position of the *bis*-pyridinium ring suggests bulky groups are favoring the activity. The electrostatic blue contour near the quaternary ammonium ion and at the *para* position of the *bis*-pyridinium ring infers that the positive charge at the corresponding positions is contributing more towards the inhibitory activity. Exactly the same trend has been observed for present dataset molecules. The series with hydrogen as substituents (**1–5**) at the *para* position is exhibiting lower activity compared to the rest of the series. Whilst, the series with electronegative groups like, CH_2OH (**45–49**), COCH_3 (**50**, **51**), COOH (**53**) and neutral groups like CH_3 (**40–43**) are exhibiting lower activities. However, altered activities have been observed among the compounds with substituted amines (**6–39**, **54**, **55**). As the bulkiness of the substituents on the amine N increases the activity also increases and vice versa. Also, any group, which can delocalize the positive charge of the pyridine nitrogen atom may possibly enhance the activity. For example, the compound **25** has (4-chloro-phenyl)-methylamine group as a substituent at R_4 position, when the 4-chloro-phenyl was substituted with phenyl group (**20**) the activity has drastically improved and the compound became the most active among the dataset molecules. Therefore, electron-donating groups like alkyl, substituted amines, methoxy etc, may well increases the activity, while electron withdrawing groups like $-\text{Br}$, $-\text{Cl}$, $-\text{CN}$, COOH may significantly decreases the activity. The encircling of *bis*-pyridinium ring with electronegative red contours further suggest that the

Table 3. General statistics of CoMFA and CoMSIA models. Bold for ChoK inhibitory activity and normal for antiproliferative activity.

Parameter	CoMFA	CoMSIA			
		M IV ^a	M V ^b	M VI ^c	M VII ^d
r^2	0.97	0.85	0.95	0.80	0.79
	0.99	0.98	0.97	0.97	0.95
q^2	0.58	0.55	0.53	0.50	0.53
	0.71	0.74	0.76	0.71	0.58
SEE ^e	0.114	0.238	0.149	0.278	0.282
	0.084	0.118	0.122	0.128	0.185
SEP ^f	0.417	0.419	0.433	0.439	0.427
	0.423	0.398	0.374	0.431	0.521
N ^g	7	4	7	4	4
	7	7	5	8	8
F ^h	161.36	56.47	91.83	38.82	37.45
	352.13	179.21	233.67	132.76	61.98
Steric	60.8	19.1	25.7	29.2	–
	64.9	30.7	32.9	45.6	–
Electrostatic	39.2	25.9	36.9	40.6	–
	35.1	33.1	33.6	48.5	–
Hydrophobic	–	30.0	37.4	–	75.2
	–	30.9	33.5	–	90.3
Donor	–	14.1	–	17.6	11.0
	–	3.4	–	3.5	5.3
Acceptor	–	11.1	–	12.6	13.8
	–	1.9	–	2.5	4.4
PRESS ⁱ	0.50	0.52	–	–	–
	1.44	0.41	–	–	–
r^2_{pred}	0.68	0.66	–	–	–
	0.81	0.95	–	–	–
Cross-validated r^2	0.50	–	–	–	–
	–	–	–	–	–
Bootstrapping r^2_{bs}	0.99	–	–	–	–
	–	–	–	–	–
SD ^j	0.002	–	–	–	–
	–	–	–	–	–

^a using S, E, H, D, A CoMSIA fields; ^b using S, E, H CoMSIA fields; ^c using S, E, D, A CoMSIA fields; ^d using H, D, A CoMSIA fields; ^e standard error of estimate; ^f standard error of prediction; ^g number of components; ^h Fisher test value; ⁱ predicted residual sum of squares for test set; ^j standard deviation.

phenyl group or any other groups, which can participate in stacking interactions with the receptor, are crucial for biological activity. The yellow contours are been observed near the linker region infer that the long alkyl chains (CH₂)_{n=1–3} (**10–12**, **17**, **22**, **34**, **36**, **43**, **47**, **48**, **51–53**), as linker are not favorable for the activity.

3.3 CoMSIA

The CoMSIA was obtained by using the same structural alignment, same training set and test set that were used for CoMFA. As mentioned previously, four different CoMSIA models (model IV–VII) have been built by changing the

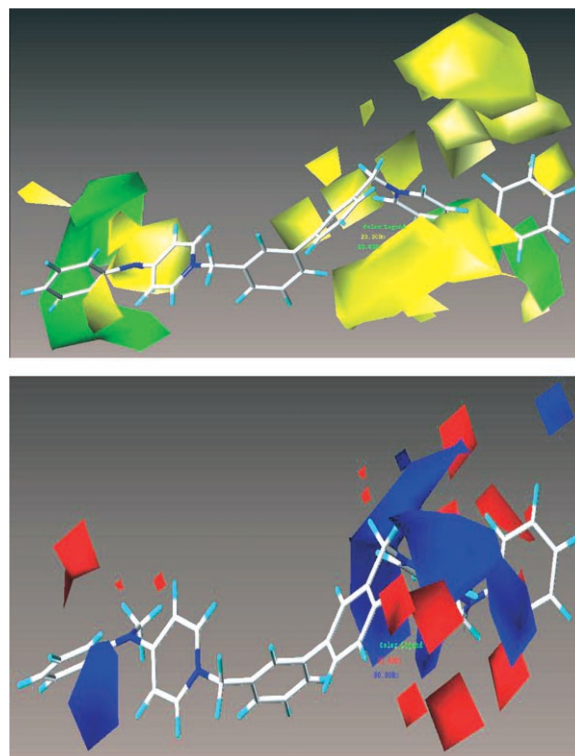


Figure 2. CoMFA steric and electrostatic contour maps. The most active compound (**20**) is displayed in the figure. Green contour indicates where bulky group increases activity, whereas yellow color contour indicates region where bulky group decreases activity. Blue contour indicates region where positive charge increases activity, whereas red contour indicates region where negative charge increases activity.

combination of five, (steric, electrostatic, hydrophobic and hydrogen bond donor and acceptor fields) CoMSIA fields. Model IV, which includes all the CoMSIA fields, has been taken for further analysis. The model gave statistically significant results with r^2 value of 0.850 and q^2 value of 0.55. The summary of CoMSIA calculations is shown in Table 3. The 3D-QSAR contour maps revealing the contribution of CoMSIA fields are shown in Figures 3–5. In the steric contours of CoMSIA, green indicates sterically favorable regions while yellow contours indicate sterically unfavorable regions. In the electrostatic contours, the introduction of electronegative substituents in red regions may increase the affinity while in blue regions decrease the affinity. In hydrophobic contours, orange region favors hydrophobic groups while white region favors hydrophilic groups. In hydrogen bond donor contours the cyan color favorable for biological activity whereas, purple color unfavorable for biological activity. In hydrogen bond acceptor contours the white color favorable for biological activity and magenta color unfavorable for biological activity. The experimental, predicted inhibitory activities and residuals for

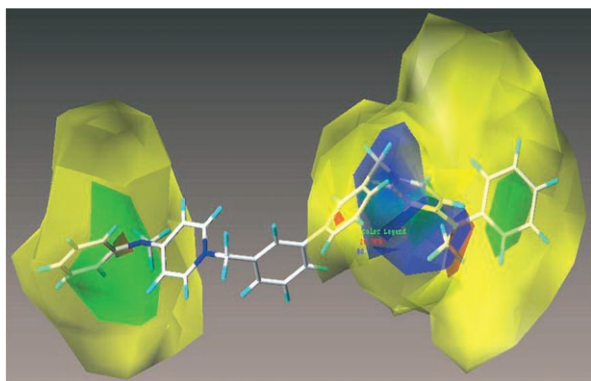


Figure 3. CoMSIA steric and electrostatic contour maps. The most active compound (**20**) is displayed in the figure.

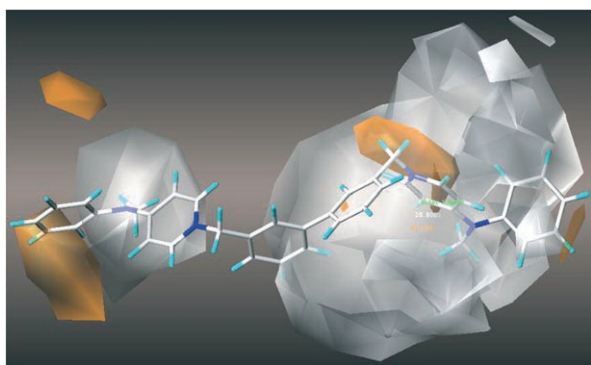


Figure 4. CoMSIA hydrophobic contour maps. The most active compound (**20**) is displayed in the figure. The orange color represents for favorable hydrophobic region whereas white represents for unfavorable region.

CoMFA training and test set are shown in Table 2 and graph of predicted vs. experimental was shown in Figure 7.

The steric maps that are generated by CoMSIA are quite different from CoMFA steric maps, instead of small yellow regions that are surrounding the linker region a continuous yellow region has been noticed. The blue electropositive region of CoMSIA is comparable with CoMFA blue contour. The white regions, disfavor for hydrophobic groups, are in good complementary with CoMFA and CoMSIA yellow regions. The overall model is favorable for hydrophilic groups rather than hydrophobic groups. The orange contour on *ortho*, *meta* and *para* positions of phenyl ring favors for hydrophobic groups and remaining part of the molecule has shown white color which is unfavored for hydrophobic groups. The cyan and white contours near the R_4 position suggest that either hydrogen bond donor or acceptor groups may contribute for the biological activity.

The CoMFA and CoMSIA models that were built with anti-proliferative activity are exhibiting comparable electrostatic contours whereas the steric fields are differing considerably from aforesaid models. The contribu-

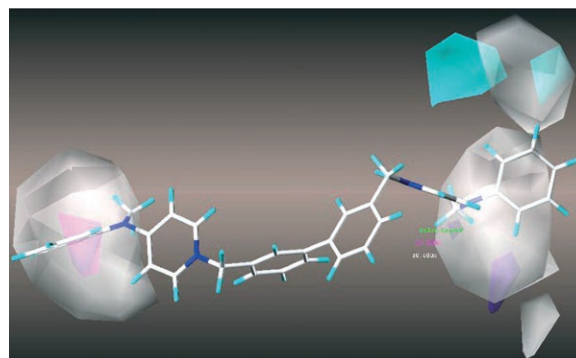


Figure 5. CoMSIA hydrogen bond donor and acceptor contour maps. The most active compound (**20**) is displayed in the figure. The cyan color represents for favorable hydrogen bond donor and purple color represents the unfavorable donor region whereas, white color represents for favorable acceptor region and magenta represents for the unfavorable acceptor region.

tions of steric unfavorable yellow region and hydrophobic regions are more when compared to the previous models (Table 3). Probably, the lipophilic property is playing a significant role for anti-proliferative activity.

4 Conclusions

2D and 3D QSAR analysis of ChoK inhibitors are performed to understand the structural requirements of *bis*-pyridinium analogues for ChoK inhibitor affinity. Both CoMFA and CoMSIA models have shown good correlative and predictive ability. Both the steric and electrostatic fields are significantly contributing to the biological activity whereas the addition of hydrogen bond donor and acceptor fields does not much improve the correlation coefficient (model IV-VII). The QSAR analysis suggests that the positive charge on 'N' of pyridinium group is crucial for ChoK inhibitory activity and groups which can delocalize the positive charge may favor for biological activity. In both CoMFA and CoMSIA models the steric unfavorable yellow region and electropositive blue region are observed. Hence, these compounds are not suitable for further bulky substituents on *bis*-pyridinium ring whereas the electropositive groups may significantly contribute to the activity. Randic index and YZ Shadow/YZ rectangle descriptors infer that 3', 3'' spacer with biphenyl ring is showing more inhibitory activity over the 4', 4'' spacer. Also, lengthening of linker moiety with $(CH_2)_{n=1-3}$ may not improve the activity.

Acknowledgements

PS thanks CSIR New Delhi for a senior research fellowship. We thank Dr. J. S. Yadav, Director, IICT for encouragement and support.

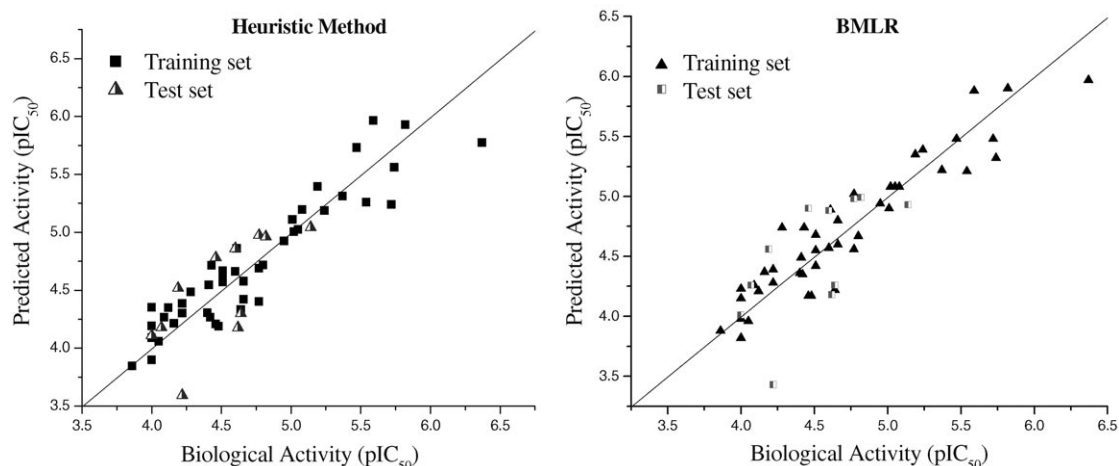


Figure 6. Graph of actual vs. predicted biological activities of the training and test sets for heuristic and BMLR methods.

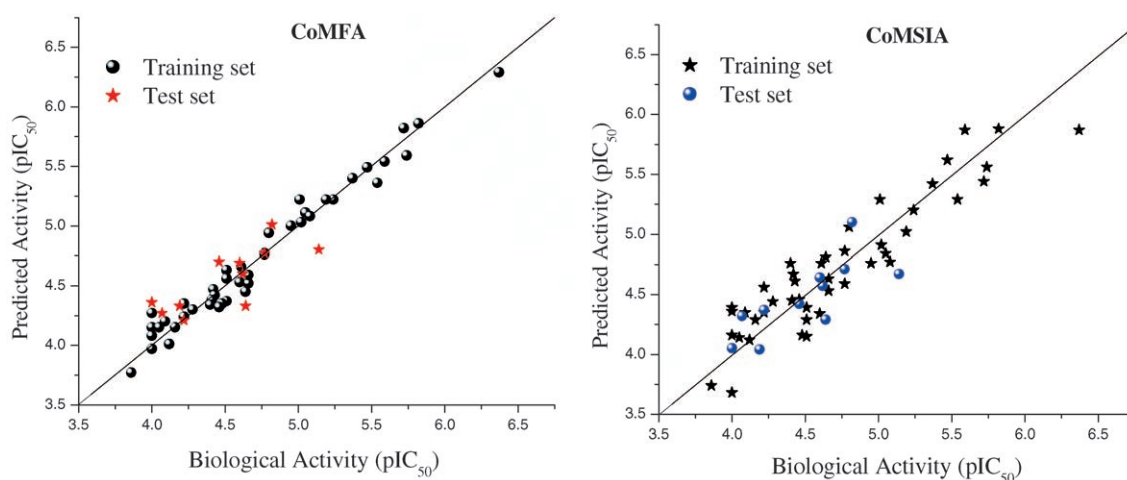


Figure 7. Graphs of actual vs. predicted biological activities of the training and test sets for CoMFA and CoMSIA fields.

References

- [1] K. Ishidate, *Biochim. Biophys. Acta*. **1997**, 1348, 70–78.
- [2] S. Janardhan, P. Srivani, G. N. Sastry *Curr. Med. Chem.* **2006**, in press.
- [3] J. Wittenberg, A. Kornberg *J. Biol. Chem.* **1953**, 202, 431–444.
- [4] C. Aoyama, H. Liao, K. Ishidate, *Prog. Lipid Res.* **2004**, 43, 266–281.
- [5] T. J. Porter, C. Kent, *J. Biol. Chem.* **1990**, 265, 414–422.
- [6] J. C. Lacal, *Idrugs*, **2001**, 4, 419–426.
- [7] R. A. Gonzalez, R. A. Molina, B. J. Razal, J. C. Lacal, *Prog. Cell Cycle Res.* **2003**, 5, 191–201.
- [8] S. R. Midgley, D. J. Kerr, *Crit. Rev. Oncol/Hematol.* **2002**, 44, 109–120.
- [9] R. A. Molina, R. A. González, J. C. Lacal, *Cancer. Lett.* **2004**, 206, 137–148.
- [10] R. A. Molina, R. A. González, V. Penalva, L. Lucas, J. C. Lacal, *Biochem. Biophys. Res. Commun.* **2001**, 285, 873–879.
- [11] R. A. Molina, V. Penalva, L. Lucas, J. C. Lacal, *Oncogene*, **2002**, 21, 937–946.
- [12] K. Nakagami, T. Uchida, S. Ohwada, Y. Koibuchi, Y. Suda, T. Sekine, Y. Morishita, *Jpn. J. Cancer. Res.* **1999**, 90, 419–424.
- [13] K. S. Crilly, M. Tomono, Z. Kiss, *Arch. Biochem. Biophys.* **1998**, 352, 137–143.
- [14] A. Cuadrado, A. Carnero, F. Dolfi, B. Jimenez, J. C. Lacal, *Oncogene* **1993**, 8, 2959–2968.
- [15] H. R. Alcoceba, F. Fernandez, J. C. Lacal, *Cancer Res.* **1999**, 59, 3112–3118.
- [16] H. R. Alcoceba, L. Saniger, J. Campos, M. C. Nunez, F. Khaless, M. A. Gallo, A. Espinosa, J. C. Lacal, *Oncogene* **1997**, 15, 2289–2301.
- [17] J. Campos, M. C. Núñez, V. Rodríguez, M. A. Gallo, A. Espinosa, *Bioorg. Med. Chem. Lett.* **2000**, 10, 767–770.
- [18] C. A. Garcia, J. Campos, S. R. M. Martin, M. A. Gallo, A. Espinosa, *J. Med. Chem.* **2003**, 46, 3754–3757.
- [19] C. A. Garcia, B. M. Coronel, S. R. M. Martin, R. A. Gonzalez, A. Ramos, R. A. Molina, A. Espinosa, M. A. Gallo, J. M. Campos, J. C. Lacal, *J. Med. Chem.* **2004**, 47, 5433–5440.
- [20] C. A. Garcia, A. Entrena, J. M. Campos, S. R. M. Martin, M. A. Gallo, A. Espinosa, *Eur. J. Med. Chem.* **2005**, 40, 315–319.

- [21] J. Campos, M. C. Núñez, S. R. M. Martin, G. J. A. Vidal, R. A. Gonzalez, M. Banez, M. A. Gallo, J. C. Lacal, A. Espinosa, *Bioorg. Med. Chem.* **2002**, *10*, 2215–2231.
- [22] J. Campos, M. C. Nunez, C. A. Garcia, S. R. M. Martin, H. R. Alcoceba, R. A. Gonzalez, J. C. Lacal, M. A. A. Gallo, A. Espinosa, *Curr. Med. Chem.* **2003**, *10*, 1095–1112.
- [23] C. A. Garcia, J. Campos, R. M. Sanchez, R. A. Gonzalez, J. C. Lacal, M. A. Gallo, A. Espinosa, *Eur. J. Med. Chem.* **2003**, *38*, 109–116.
- [24] J. Campos, C. Nunez, J. Diaz. M. Sanchez. M. A. Gallo, A. Espinosa, *Il Farmaco.* **2003**, *58*, 221–229.
- [25] SYBYL 6.9. Molecular Modeling Software, Tripos Inc., 1699 Hanley Road, St. Louis, MO 63144.
- [26] AMPAC 5.0, © 1994 Semichem, 7128 Summit, Shawnee, KS 66216.
- [27] a) CODESSA version 2.0 Semichem, 7204 Mullen, Shawnee, KS 66216 USA. b) M. Karelson, V. S. Lobanov, A. R. Kairitzky, *Chem. Rev.* **1996**, *96*, 1027–1044.
- [28] R. D. Crammer, D. E. Patterson, J. D. Bunce, *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- [29] G. Klebe, U. Abraham, T. Mietzner, *J. Med. Chem.* **1994**, *37*, 4130–4146.
- [30] V. N. Viswanadhan, A. K. Ghose, R. Revenkar, R. N. Robins, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.
- [31] M. Clark, R. D. Cramer, V. N. Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982–1012.
- [32] a) M. Randic, *J. Am. Chem. Soc.*, **1975**, *97*, 6609–6615. b) O. Strouf, *Chemical Pattern Recognition*, Wiley, New York **1986**.
- [33] L. B. Kier, L. H. Hall, *Molecular connectivity in chemistry and drug research*; Academic Press: New York, **1976**.

Supplementary Data

Table S1. Descriptors, coefficients (X), error (DX) and t-values for Model I

S. No	X	DX	t-test	Descriptor
0	−19.0357	12.9660	−6.9686	Intercept
1	0.2231	0.0246	9.0455	Randic index (order 3)
2	−0.5214	0.0905	−5.7606	Min atomic state energy for a C atom
3	−2.0812	0.7719	−2.6963	YZ Shadow/YZ Rectangle
4	3.5242	0.6643	5.3051	Min total interaction for a C–H bond
5	13.8150	3.1339	4.4084	Max electroph. react. index for a N atom

Activity = (−19.0357 + 0.2231 * RI − 0.5214 * MASEC − 2.0812 * YZSYZR + 3.5242 * MTICH + 13.8150 * MERIN)

Table S2. Descriptors, coefficients (X), error (DX) and t-values for Model II

S. No	X	DX	t-test	Descriptor
0	−54.7880	7.7934	−7.0301	Intercept
1	−0.0775	0.00942	8.2321	Number of Carbon atoms
2	−3.2134	0.7097	−4.5279	YZ Shadow/YZ Rectangle
3	4.6723	0.6103	7.6559	Min. total interaction for a C–H bond
4	−0.006398	0.00126	−5.0973	ESP-HASA (H-acceptor surface area)
5	13.0740	3.1877	4.1014	Max electroph. react. index for a C atom

Activity = (−54.7880 − 0.0775 * NC − 3.2134 * YZSYZR + 4.6723 * MTICH − 6398.8 * ESPHASA + 1307.4 * MERIC).

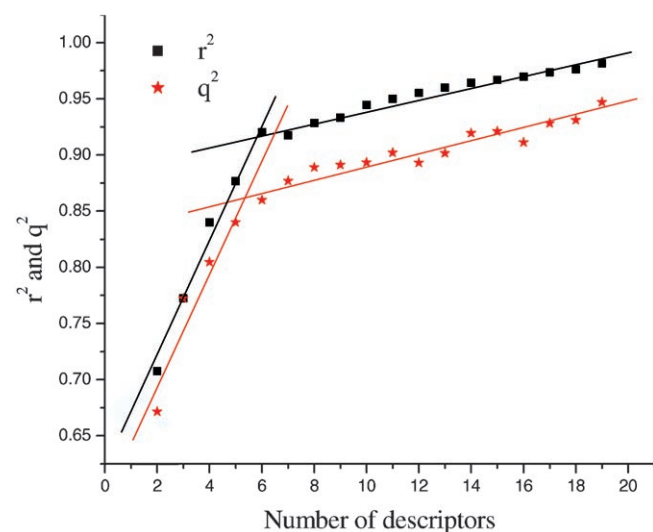


Figure S1. Correlation coefficient, r^2 and q^2 , vs. number of descriptors of Model II