Asian Journal of Chemistry Vol. 17, No. 2 (2005), 1031–1040

3D-QSAR Studies of Benzoxazinones: Analogs of Efavirenz

REENA RACHEL JACOB, SURENDRA KUMAR and MEENA TIWARI*

Department of Pharmacy, Shri Govindram Seksaria Institute of Technology and Science
23, Park Road, Indore-452 003, India
Email: reena_rachel_j@yahoo.co.in

In the present study, a set of 14 analogs of Efavirenz with human immunodeficiency virus-1 (HIV-1) reverse transcriptase (RT) inhibitory activity, were subjected to 3D-QSAR studies. Various combinations of thermodynanic, electronic and steric descriptors were used in order to understand the physicochemical properties desirable for interaction with the receptor. Multiple linear regression analysis was performed, using VALSTAT, to select the descriptors and to generate various models that relate the structural features to the biological activity. Among them, an informative and statistically significant model both in fitting and predictive ability ($r = 0.9354$ and $r^2_{cv} = 0.8059$) was selected. Cross-validation was performed using leave-one-out (LOO) and bootstrapping method. The significant model indicated that the thermodynamic descriptors, *viz.*, Henry's law constant and stretch bend energy play an important role in RT inhibitory activity. Consequently, the best QSAR model will be of major importance to aid the design of new HIV-1 reverse transcriptase inhibitor.

Key Words: HIV-1, Reverse transcriptase, 3D-QSAR.

INTRODUCTION

The treatment of acquired immunodeficiency syndrome (AIDS) is the most challenging worldwide medical problem. Reverse transcriptase (RT) is a key enzyme of the HIV, involved in the transcription phase of replication of the virus. The main function of this enzyme is to produce double stranded DNA from single stranded RNA so that the genetic information can be integrated into the host cell chromosome and the virus can replicate[1]. Human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT) is an important target for chemotherapeutic agents against the AIDS disease[2]. It contains multiple sites where inhibitors can efficiently bind.

Drugs[3–6] targeting RT can be grouped into two classes: nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs mechanistically act as DNA chain terminators and are equally active against HIV-1 and HIV-2 RT. They can also be incorporated into

cellular DNA by host DNA polymerase causing severe adverse reactions. On the other hand, NNRTIs bind in a non-competitive manner to a hydrophobic pocket close to, but distinct from, the RT active site in the p66 subunit[3], altering its ability to function and achieving a highly selective suppression of HIV-1 replication[7, 8]. They are less toxic than nucleoside analogs. Despite the favourable toxicity and adherence properties, resistance to NNRTIs develops rapidly. However, in view of increasing incidence of resistance to current drug regimens and frequency of adverse events, there is a need to develop potent, safe and selective anti-HIV agents.

The aim of the present study is to carry out QSAR analysis for benzoxazinone derivatives, in order to identify the necessary structural and physicochemical requirements for binding with reverse transcriptase enzyme. QSAR has been used in the discovery and development of new drugs to achieve objectives[9] like quantitative prediction of biological activity of a compound, classification of compounds into various classes, optimization of a lead compound and refinement of synthetic targets.

EXPERIMENTAL

The data of 14 benzoxazinone derivatives was taken from the reported work[10]. The structures along with their corresponding biological activities, expressed as IC_{50} (nM: concentration at which 50% of the enzyme is inhibited) are shown in Table-1. All the molecules studied have the same parent skeleton (Fig. 1). The compounds with numerically undefined biological activities were excluded from the series. The biological activity data (IC_{50}) was converted to negative logarithmic dose (pIC_{50}) for the QSAR analysis. All the molecular modeling studies were performed using C.S. Chem. Office Software Version 6.0 (Cambridge Soft)[11]. The energy minimization was carried out using MM2 force field. The minimization continued until the root mean square (RMS) gradient reaches a value smaller than 0.1 kcal/mol Å. Hamiltonian approximation Austin model-1 (AM1) method[12] was used for reoptimization until the RMS gradient value becomes smaller than 0.0001 kcal/mol Å.

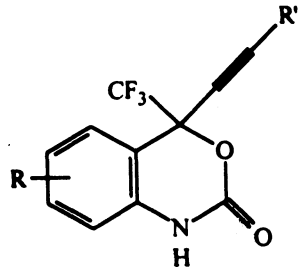

Fig. 1. General structure of benzoxazinone derivatives (R and R′: substituents)

TABLE-1
STRUCTURES AND RT INHIBITORY ACTIVITY OF BENZOXAZINONE DERIVATIVES

Compound No.	Substituents		– log IC_{50}^{*}
	R^1	R^2	
1	5–F	c-Propyl	–1.892
2	5–F	Ethyl	–2.103
3	5–F	*n*-Propyl	–2.193
4	5–F	Isopropyl	–2.000
5	6–NO_2	c-Propyl	–2.320
6	6–NO_2	Ethyl	–2.440
7	6–NO_2	*n*-Propyl	–2.482
8	6–NO_2	Isopropyl	–2.298
9	6–NH_2	c-Propyl	–2.904
10	6–NH_2	Ethyl	–3.277
11	6–NH_2	*n*-Propyl	–3.177
12	6–NH_2	Isopropyl	–2.952
13	6–N (H)CH_3	c-Propyl	–2.783
14	6–N (H)CH_3	Isopropyl	–2.674

*Negative logarithm of concentration of 50 per cent binding affinity in nM

The thermodynamic, electronic and steric descriptors (Table-2) were calculated using "compute properties module" of the software.

TABLE-2
LIST OF DESCRIPTORS USED IN THE STUDY

S.No.	Descriptors	Description
1	Boiling point (b.p.)	The boiling point for the structure at 1 atm.
2	Critical pressure (P_C)	The minimum pressure that must be applied to liquefy the structure at the critical temperature.
3	Critical temperature (T_C)	The temperature above which the gas form of the structure cannot be liquefied, no matter what the applied pressure.
4	Critical volume (V_C)	The volume occupied at the compound's critical temperature and pressure.
5	Heat of formation (ΔHf)	The heat of formation for the structure at 298.15 K and 1 atm.
6	Henry's law constant (H)	The logarithm of Henry's law constant.
7	Ideal gas thermal capacity (C_p)	The constant pressure (1 atm) molar heat capacity at 298.15 K for an ideal gas compound.
8	log P	The logarithm of the partition coefficient for *n*-octanol/water.

S.No.	Descriptors	Description
9	Melting point (m.p.)	The melting point for the structure at 1 atm.
10	Molar refractivity (m.r.)	The molar refraction index.
11	Standard Gibb's free energy (G)	The Gibb's free energy (ΔG) for the structure at 298.15 K and 1 atm.
12	Connolly solvent accessible surface (SAS)	The locus of the centre of a spherical probe as it is rolled over the molecular model.
13	Connolly molecular surface area (MS)	The contact surface created when a spherical probe sphere is rolled over the molecular model.
14	Connolly solvent-excluded volume (SEV)	The volume contained within the contact molecular surface.
15	Exact mass (EM)	The exact molecular mass of the molecule, where atomic masses of each atom are based on the most common isotope for the element.
16	Ovality	The ratio of the molecular surface area to the minimum surface area.
17	Principal moment of inertia (PMI: X, Y, Z)	The moments of inertia when the Cartesian coordinate axes are the principal axes of the molecule.
18	Dipole moment (D)	Molecular dipole moment.
19	Dipole-dipole energy (E_d)	The sum of the electrostatic energy resulting from interaction of two dipoles.
20	Electronic energy (E_e)	The total electronic energy.
21	HOMO energy	Energy of the highest occupied molecular orbit.
22	LUMO energy	Energy of the lowest unoccupied molecular orbit.
23	Repulsion energy (E_r)	Total core-core internuclear repulsion between the atoms.
24	Torsion energy (E_t)	The sum of the dihedral bond rotational energy term of the force field equation.
25	Total energy (E)	The sum of all terms of the force field equation.
26	Bend energy (E_b)	The sum of the angle-bending terms of the force field equation.
27	Stretch-bend energy (E_{sb})	The sum of the stretch-bend coupling terms of the force field equation.
28	van der Waals 1,4-energy	The sum of pairwise van der Waals interaction energy terms for atoms separated by exactly 3 chemical bonds.
29	Non-1,4 van der Waals energy	The sum of pairwise van der Waals interaction energy terms for atoms separated by more than 3 chemical bonds.

Multiple linear regression analysis method was used to obtain various QSAR equations using VALSTAT[13]. The QSAR models were cross-validated using leave-one-out (LOO) and bootstrapping method. Statistical measures used were: n: number of sample in the regression, r^2: squared correlation coefficient,

S: standard deviation, F-test (Fischer's value) for statistical significance, r^2_{cv}: cross-validated squared correlation coefficient, r^2_{bs}: boot strapped squared correlation coefficient, S_{PRESS}: standard deviation of prediction, S_{DEP}: standard deviation of error of preduction and intercorrelation matrix to show mutual correlation among the parameters.

RESULTS AND DISCUSSION

All the calculated descriptors were considered as independent variable and biological activity (pIC_{50}) as dependent variable. Multiple Linear Regression (MLR) analysis was used to relate the RT inhibitory activity to the molecular descriptors, resulting in various statistical equations.

$$pIC_{50} = -0.0094\ (\pm\ 0.0027)*BP - 0.0022\ (\pm\ 0.0005)*T - 6.4628 \quad (1)$$

$n = 14$, $r = 0.9457$, $r^2 = 0.8944$, $S = 0.1537$, $F = 46.6215$

$r^2_{bs} = 0.9128$, chance < 0.001, $r^2_{cv} = 0.8337$, $S_{PRESS} = 0.1929$, $S_{DEP} = 0.1710$.

$$pIC_{50} = 0.3964\ (\pm\ 0.1157)*D + 0.2949\ (\pm\ 0.1150)*PC - 4.0302 \quad (2)$$

$n = 14$, $r = 0.9359$, $r^2 = 0.875945$, $S = 0.1666$, $F = 38.8353$.

$r^2_{bs} = 0.9065$, chance < 0.001, $r^2_{cv} = 0.7887$, $S_{PRESS} = 0.21748$, $S_{DEP} = 0.192775$

$$pIC_{50} = -0.2492\ (\pm\ 0.0675)*HLC - 0.2041\ (\pm\ 0.2618)*E_{sb} - 0.4896 \quad (3)$$

$n = 14$, $r = 0.9354$, $r^2 = 0.8751$, $S = 0.1671$, $F = 38.5559$.

$r^2_{bs} = 0.90530$, chance < 0.001, $r^2_{cv} = 0.8059$, $S_{PRESS} = 0.2084$, $S_{DEP} = 0.1847$.

$$pIC_{50} = -0.2632\ (\pm\ 0.0723)*HLC + 0.0045\ (\pm\ 0.0091)*SEV - 1.4314 \quad (4)$$

$n = 14$, $r = 0.9256$, $r^2 = 0.8567$, $S = 0.1790$, $F = 32.9072$

$r^2_{bs} = 0.88487$, chance < 0.001, $r^2_{cv} = 0.7571$, $S_{PRESS} = 0.2332$, $S_{DEP} = 0.2067$.

$$pIC_{50} = -0.2594\ (\pm\ 0.0714)*HLC + 0.0343\ (\pm\ 0.0683)*E_t - 0.2680 \quad (5)$$

$n = 14$, $r = 0.9258$, $r^2 = 0.8571$, $S = 0.1788$, $F = 33.0013$

$r^2_{bs} = 0.8886$, chance < 0.001, $r^2_{cv} = 0.7318$, $S_{PRESS} = 0.2450$, $S_{DEP} = 0.2171$.

The quality of the models, judged by r, S, F, r^2_{cv}, r^2_{bs}, S_{PRESS} and intercorrelation (Table-4) within the descriptors, was statistically tested. Consequently, eqns. (1) to (5) were evaluated. Accordingly, two models [eqns. (3) and (5)] were considered for further analysis.

$$pIC_{50} = -0.2492\ (\pm\ 0.0675)*HLC - 0.2041\ (\pm\ 0.2618)*E_{sb} - 0.4896 \quad (3)$$

Statistical and cross validated results

$n = 14$, $r = 0.9354$, $r^2 = 0.8751$, $STD = 0.1671$, $F = 38.5559$, $r^2_{bs} = 0.9053$, chance < 0.001, $r^2_{cv} = 0.8059$, $S_{PRESS} = 0.2084$, $S_{DEP} = 0.1847$.

Eqn. (3) has good correlation ($r = 0.9354$) between the descriptors, namely,

Henry's law constant (HLC), stretch bend energy (E_{sb}) and biological activity. The statistical measures, r^2, S and F determine the estimation power of the model and evaluate it internally. The model also explains 87.51% variance in activity (Fig. 3). The data shows better statistical significance (> 99.9%) with the F-value $F_{(2, 11)} = 38.5559$, against the value of 99.9% significance $F_{(2, 11, \alpha\, 0.001)} = 13.1$. The value of standard deviation is low (S = 0.1671), which is an absolute measure of quality of fit of the model. The values of the cross-validation parameters [eqn. (3)] r^2_{cv}, r^2_{bs}, S_{PRESS}, S_{DEP} and chance determine the validity and predictive power of the model.

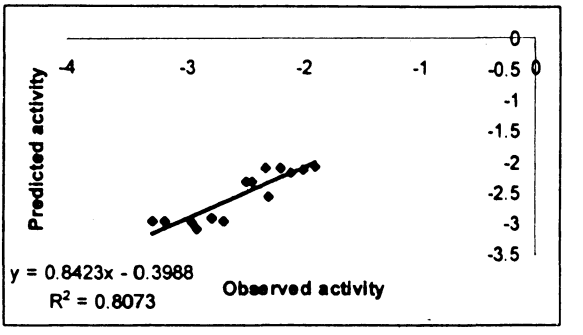

Fig. 2. Scatter plot showing observed *vs.* predicted activity (eq-3)

TABLE-3
SIGNIFICANT DESCRIPTORS IN Eqn. (3) AND Eqn. (5)

S. No.	$-\log IC^*_{50}$ [Obs.]	Descriptors		
		HLC	E_{sb}	E_t
1	–1.892	6.343	–0.447	–3.913
2	–2.1038	6.111	0.631	–5.953
3	–2.1931	5.988	0.661	–6.031
4	–2.0086	5.988	0.561	–6.454
5	–2.3201	7.295	–0.527	–6.170
6	–2.4409	7.063	0.380	–8.578
7	–2.4828	6.939	0.546	–8.227
8	–2.2988	7.658	0.612	–8.222
9	–2.9041	9.862	0.448	–6.514
10	–3.2773	9.630	0.632	–7.961
11	–3.1778	9.506	0.605	–3.658
12	–2.9523	9.506	0.530	–7.743
13	–2.7839	9.520	0.063	–5.588
14	–2.6748	9.164	0.638	–4.494

*Observed negative logarithm of concentration of 50% binding affinity in nM.

The observed *vs.* predicted activity obtained from eqn. (3) is shown in Table-5 and Fig. 2. The intercorrelation matrix (Table-4) shows that the mutual correlation

among the descriptors is less than 0.2. HLC and E_{sb} both are thermodynamic descriptors contributing negatively to the activity. The relative contribution of HLC and E_{sb} is 96.19 and 3.80% respectively. It is evident that the contribution of HLC is high, giving information about the distribution of drug in the phases. When HLC is increased, the drugs have more lipophilic character increasing the distribution of drug in the body. However, Eqn. (3) shows negative influence of HLC on the biological activity. This suggests that the substituents with less lipophilic character will be highly significant for the increase in anti-HIV activity.

$$pIC_{50} = -0.2594(\pm 0.0714)*HLC + 0.0343(\pm 0.0683)*E_t - 0.2680 \quad (5)$$

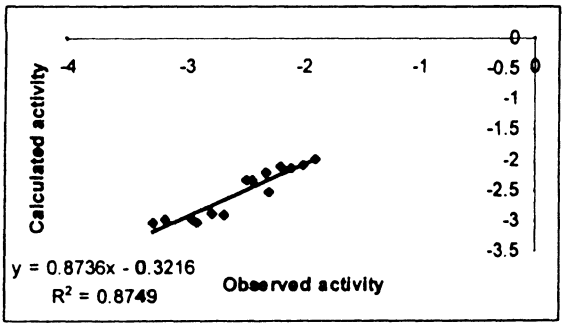

Fig. 3. Scatter plot showing observed *vs.* calculated activity (eq-3)

TABLE-4
INTERCORRELATION BETWEEN THE DESCRIPTORS

	BP	T	D	PC	HLC	E_{sb}	SEV	E_t
BP	1.0000							
T	0.6866	1.0000						
D	0.4889	0.8695	1.0000					
PC	0.8724	0.1800	0.1800	1.0000				
HLC	0.1192	0.5648	0.7673	0.4536	1.0000			
E_{sb}	0.1900	0.2860	0.3547	0.1888	0.1582	1.0000		
SEV	0.4632	0.5095	0.0580	0.1643	0.1576	0.0607	1.0000	
E_t	0.5268	0.3370	0.4244	0.5171	0.0508	0.3021	0.0079	1.0000

Statistical and corss validated results

n = 14, r = 0.9258, r^2 = 0.8571, S = 0.1788, F = 33.0013, r^2_{bs} = 0.8886, chance < 0.001, r^2_{cv} = 0.7318, S_{PRESS} = 0.2450, S_{DEP} = 0.2171.

Eqn. (5) shows good correlation (r = 0.9258) between the descriptors, namely, Henry's law constant (HLC), torsion energy (E_t) and biological activity. The r^2 accounts for 85.71% variance in activity (Fig. 5). In addition to this, eqn. (5) has good statistical significance (>99.9%) with the F-value $F_{(2, 11)}$ = 33.0013, against the value of 99.9% significance $F_{(2, 11, \alpha 0.001)}$ = 13.1. The standard deviation is below 0.2. The satisfactory results of cross-validated squared correlation coefficient (r^2_{cv}) = 0.7318, boot strapping squared correlation coefficient (r^2_{bs}) = 0.8886

and chance correlation < 0.001 revealed that the results were not based on chance correlation.

The leave-one-out (LOO) predicted and calculated activities of eqn. (5) are shown in Table-5. In order to confirm the results we have estimated the values of pIC_{50} using LOO and correlated them with the observed value of IC_{50}. This correlation has been shown in Fig. 4. The intercorrelation matrix shows (Table-4) that the parameters HLC and E_t are not much correlated (0.06) as desired in the QSAR analysis. Eqn. (5) also shows that HLC and E_t have positive effect on the biological activity. Thus an increase in the value of both the descriptors will increase the anti-HIV activity. The relative contributions of thermodynamic descriptors were 9.67 and 90.32% respectively. The value of only those descriptors, which are significant in eqns. (3) and (5) is shown in Table-3.

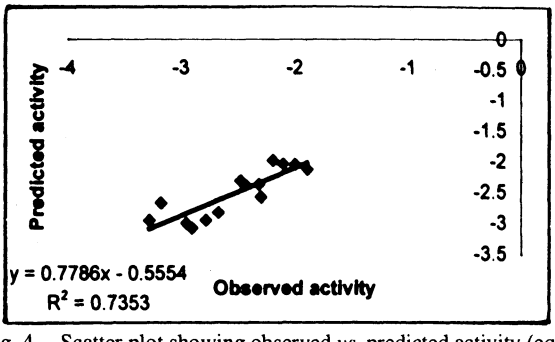

Fig. 4. Scatter plot showing observed *vs.* predicted activity (eq-5)

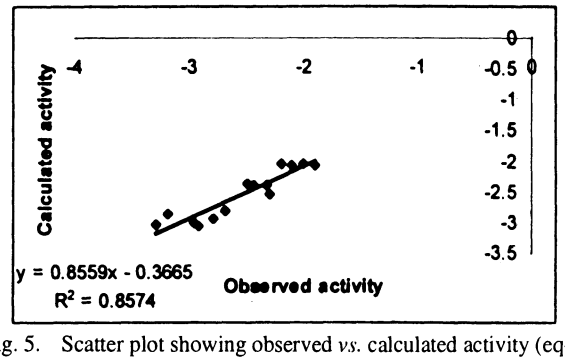

Fig. 5. Scatter plot showing observed *vs.* calculated activity (eq-5)

It is interesting to note that both the models are statistically significant when comparing the r, r^2, S, F-test (Fischer's value). The difference in the correlation coefficient (r) and standard deviation (S) of eqns. (3) and (5) is rather small. The F values of eqn. (3) (38.55) is better than that of eqn. (5) (33.00). However, the intercorrelation within the descriptor is more (0.2) in eqn. (3) when compared with that of eqn. (5) (0.06). Finally, when comparing the quality of the models, eqn. (3) and eqn. (5) it is evident that eqn. (3) shows higher predictive ability (r^2_{cv} = 0.8059). Consequently, eqn. (3) was considered to be the best QSAR model in the present study, both in statistical significance and predictive ability. It helps to understand the relationship of physicochemical descriptors with structure of

drug molecule and biological activity. It can also be used for theoretical prediction of RT inhibitory activity of the new active drug molecules.

TABLE-5
CALCULATED (CAL.) ACTIVITY, PREDICTED (PRED.) ACTIVITY AND Z-VALUE DATA

Compound No.	Eqn. (3)			Eqn. (5)		
	Cal.[a]	Z-value	Pred.[b]	Cal.[c]	Z-value	Pred.[d]
1	–1.979	0.569	–2.041	–2.048	0.948	–2.129
2	–2.142	0.248	–2.153	–2.057	–0.280	–2.047
3	–2.117	–0.491	–2.092	–2.028	–1.001	–1.988
4	–2.097	0.575	–2.122	–2.042	0.208	–2.050
5	–2.200	–0.776	–2.094	–2.372	0.317	–2.377
6	–2.328	–0.733	–2.316	–2.394	–0.281	–2.380
7	–2.331	–0.985	–2.310	–2.350	-0.803	–2.318
8	–2.523	1.462	–2.549	–2.536	1.446	–2.585
9	–3.039	0.881	–3.072	–3.050	0.887	–3.085
10	–3.019	–1.677	–2.961	–3.039	–1.445	–2.960
11	–2.983	–1.265	–2.944	–2.859	–1.932	–2.680
12	–2.967	0.101	–2.970	–3.000	0.290	–3.013
13	–2.875	0.597	–2.903	–2.929	0.885	–2.959
14	–2.904	1.493	–2.943	–2.048	0.759	–2.835

[a]Calculated negative logarithm of concentration of 50 per cent binding affinity (nM) in eqn. (3)
[b]Predicted negative logarithm of concentration of 50 per cent binding affinity (nM) in eqn. (3)
[c]Calculated negative logarithm of concentration of 50 per cent binding affinity (nM) in eqn. (5)
[d]Predicted negative logarithm of concentration of 50 per cent binding affinity (nM) in eqn. (5).

Conclusion

The evaluation of the best QSAR model reveals that the thermodynamic descriptor plays an important role in the RT inhibitory activity. On the basis of the above study, it can be concluded that reverse transcription inhibition by benzoxazinone derivatives is strongly influenced by the thermodynamic nature of the substituents, hence it increases the probability of improving the activity. So, this study may be helpful in the development and optimization of existing reverse transcriptase inhibitors.

ACKNOWLEDGEMENTS

The authors are thankful to, Director, SGSITS, Indore for providing necessary facilities for carrying out this work. Authors RRJ and SK are thankful to AICTE, New Delhi for providing scholarship.

REFERENCES

1. A. Yadav and S.K. Singh, *Bioorg. Med. Chem.*, **11**, 1801 (2003).
2. L. Douali, D. Villemin and D. Cherqaoui, *Int. J. Mol. Sci.*, **5**, 48 (2004).
3. A. Ranise, A. Spallarossa, S. Schenone, O. Bruno, F. Bondavalli, L. Vargiu, T. Marceddu, M. Mura, P.L. Colla and A. Pani, *J. Med. Chem.*, **46**, 768 (2003).
4. (a) R.A. Katz and A.M. Shalka, *Annu. Rev. Biochem.*, **63**, 133 (1994).
   (b) Y.N. Vaishnav and F. Wong-staal, *Annu. Rev. Biochem.*, **60**, 577 (1991).
5. H. Jonckheere, J. Anne and E. De Clercq,*Med. Res. Rev.*, **20**, 129 (2000).
6. R.G. Nanni, J. Ding, A. Jacobo-Molina, S.H. Hughes and E. Arnold, *Drug Discovery Des.*, **1**, 129 (1993).
7. M.L. Barreca, J. Balzarini, A. Chimirri, E.D. Clercq, L.D. Luca, H.D. Holtje, M. Holtje, A.M. Monforte, P. Monforte, C. Pannecouque, A. Rao and M. Zappala, *J. Med. Chem.*, **45** (2002).
8. G. Hajos, S. Riedi, J. Molnar and D. Szabo, *Drugs Fut.*, **25**, 47 (2000).
9. C. Hansch, Comprehensive Medicinal Chemistry, Pergamon Press, New York, p. 9 (1990).
10. M. Patel, R.J. McHugh (Jr.), B.C. Cordova, R.M. Klabe, S. Erickson-Viitanem, G.L. Trainor and S.S. Ko., *Bioorg. Med. Chem. Lett.*, **9**, 3221 (1999).
11. C.S. Chem. Office, Version 6.0, Cambridge Soft Corporation, Software Publishers Association, 1730M Street, Nw, Suite 700, Washington D.C. 20036 (202) As2-1600, USA.
12. L.B. Kier, Molecular Orbital Theory in Drug Research, Academic Press, New York, 62 (1971).
13. A.K. Gupta, B.M. Arockia and S.G. Kaskhedikar, Indian Pharmaceutical Congress, BP-7, 192 (2003).

(*Received*: 2 July 2004; *Accepted:* 3 January 2005) AJC-4045