Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 3571–3581

BIOORGANIC & MEDICINAL CHEMISTRY

Quantitative Structure–Activity Relationship Studies on 5-Phenyl-3-ureido-1,5-benzodiazepine as Cholecystokinin-A Receptor Antagonists

Vijay K. Agrawal,[a] Ruchi Sharma[a] and Padmakar V. Khadikar[b,*]

[a]*QSAR and Computer Chemical Laboratories, Department of Chemistry, A. P. S. University, Rewa-486 003, India*
[b]*Research Division, Laxmi Fumigation and Pest Control Pvt Ltd 3 Khatipura, Indore-452 007, India*

Received 4 March 2002; accepted 13 May 2002

Abstract—Quantitative structure–activity relationship (QSAR) studies on a series of 5-phenyl-3-ureido-1,5-benzodiazepine-2,4-diones has been carried out using a pool of distance-based topological indices. Step-wise regression analysis indicated that penta-parametric regression expression containing *Sz, B, Ip*$_1$, *Ip*$_2$ and *Ip*$_3$ is the most potent and selective for *CCK-A* affinity. The predictive potential of the model is discussed on the basis of cross-validation parameters as well as by estimating root mean square (*RMSR*) of the residuals.

Introduction

Cholecystokinin (CCK) is a gastrointestinal peptide harmone of 33 amino acids that was originally isolated from porcine gut.[1] In addition, in high concentration it is also reported in the brain. In the brain it is present as C-terminal octapeptide,[2,3] usually sulfated on tyrosine **7** (*CCK-8S*). *CCK* exhibits many of the characteristics of a neurotransmitter; it is synthesized in neurons, stored in synaptic vesicles, metabolized in brain and so on. Furthermore, *CCK* has specific binding sites associated with nerve terminals containing *CCK*.[4] Therefore, *CCK* can be considered to belong to a class of peptides that act both as gut harmones as well as central neurotransmitters. In the central nervous system (CNS), it has been found to induce a variety of physiological effects such as hypothermia, analgesia, hyperglycemia, stimulation of pituitary harmone release, and decrease in exploratory behavior.[5] The peripheral actions of *CCK* are mediated by derivatives (Fig. 1) designated as L-364, 718 was found to be a highly potent and selective antagonist of *CCK-A* with receptor affinity comparable to that of native peptide.[6–8]

Very recently, Ursini et al.[9] synthesized a series of aforementioned compounds (Fig. 1, Table 1) and tested them in vitro to determine their affinity for *CCK-A*. These affinities with different substituents were determined by radioligand binding studies using rat pancreatic membranes (*CCK-A* receptors). Such pK_i (*CCK-A*) values were adopted from the work of Ursini et al.[9] and listed in Table 2. It is interesting to record that no attempt is made till today in investigating the quantitative structure–activity relationship (QSAR) for these compounds using topological indices. Our earlier studies[10,11] have shown that distance-based topological indices are well suited for modelling pK_i. In this communication, therefore, we report QSAR studies on the modelling of pK_i (*CCK-A*) for a series of 1-alkyl-5-aryl-3-ureido-1,5-benzodiazepine-2,4-diones as synthesized and evaluated by Ursini et al.[9] (Table 1) using distance-based topological indices. The results, as discussed below, show that pK_i (*CCK-A*) can be modelled successfully by multiparametric regression upon intro-

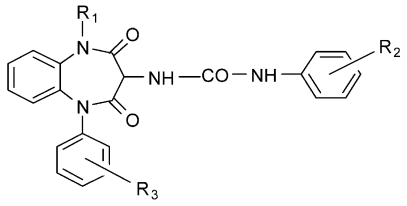

Figure 1. Chemical structure of 1-alkyl-5-aryl-3-ureeido-1,5-benzodiazepine-2,4-diones in *CCK-A* binding assays.

*Corresponding author. Tel.: +91-731-531906; fax: +91-731-531906; e-mail: vijay-agrawal@lycos.com

0968-0896/02/$ - see front matter
PII: S0968-0896(02)00167-0

duction of indicator parameters. For obtaining such statistically significant models we have used a pool of distance-based topological indices (Table 2).

Results and Discussion

A perusal of Table 1 shows that no degeneracy is present in pK_i (*CCK-A*) binding constants. However, low to high degeneracy is observed for the distance-based topological indices (Table 2) used in the present investigation. Also, pK_i (*CCK-A*) is found to be highest (7.71) for compound **23** and lowest (4.7) for compound **33**. No definite reasoning could be found for such a diversified exhibition of pK_i (*CCK-A*). Perhaps, it may be attributed to different sites of action. In case of topological indices the observed degeneracy (except for χ) may be attributed to the fact that they belong to first-generation topological indices.[12–14] According to Balaban[12] the first-generation topological indices in spite of their degeneracy are quite useful in developing QSAR/QSPR modes. This is found to be the case in the present study also.

The correlation matrix (Table 3) indicates that pK_i (*CCK-A*) can be modelled by monoparametric regressions using *W, J, Sz* and log *RB* significantly. However, such monoparametric models will be of fairly low quality. The data presented in Table 3 also shows that high collinearity exists between: (i) *W, B*; (ii) *W,* χ; (iii) *W, J*; (iv) *W,* log *RB*; (v) *B, Sz*; (vi) *B,* log *RB*; (vii) χ*, Sz*; (viii) χ, log *RB* and (ix) *Sz,* log *RB*. Comparatively lesser collinearity exists between: (i) *W, J*; (ii) *B, J*; (iii) χ*, J*; (iv) *Sz, J* and (v) log *RB, J*. This means that any multiparametric model containing any such combination of topological indices may suffer from the defect due to collinearity. This aspect is dealt with in detail in the following section.

Step-wise regressions (Table 4; mono- as well as polyparametric) resulted into five statistically significant mono-parametric models, out of which the model based on the Sz index was found to be most appropriate. This model is found as:

$$pK_i(CCK-A) = -1.6432 \times 10^{-4}(\pm 2.4720 \times 10^{-5})\ Sz + 8.0449 \quad (1)$$

$n = 40$, $Se = 0.4271$, $R = -0.7332$, $F = 44.184$, $p = 7.412 \times 10^{-8}$

Table 1. 1-Alkyl-5-aryl-3-ureido-1,5-benzodiazepine-2,4-diones used in the present study, their pK_i (*CCK-A*) values and indicator parameters

Compd	R_1	R_2	R_3	pK_i (*CCK-A*)	Ip_1	Ip_2	Ip_3
1	Butyl	H	H	6.260	1	0	0
2	3-Methyl,but-1-yl	H	H	6.490	1	0	0
3	3-Methyl,but-1-yl	H	H	6.740	1	0	0
4	3-Methyl,but-1-yl	H	F	6.710	1	1	0
5	1,3-Dimethyl,but-1-yl	H	H	6.940	1	0	1
6	2,3-Dimethyl,but-1-yl	H	H	6.910	1	0	1
7	3,3-Dimethyl,but-1-yl	H	H	6.950	1	0	1
8	3,3-Dimethyl,but-1-yl	H	F	7.150	1	1	1
9	1-Cyclopentyl,prop-2-yl	H	H	7.110	1	0	0
10	2-Cyclopentyl,ethyl	H	F	7.000	1	1	0
11	2-Phenyl,ethyl	H	H	6.570	1	0	0
12	Adament-1-yl,methyl	H	H	6.320	1	0	0
13	2-(Adament-1-yl),ethyl	H	H	6.660	1	0	0
14	Adament-2-yl	H	H	6.660	1	0	0
15	Bicyclo(2,2,1)hept-2-yl	H	H	6.600	1	0	0
16	3-Methyl,but-1-yl	3-*N*,*N*-Dimethyl,amino	F	6.900	0	1	0
17	3-Methyl,but-1-yl	3-*N*,*N*-Dimethyl,amino	F	6.700	0	1	0
18	3-Methyl,but-1-yl	3-*N*,*N*-Dimethyl,amino	F	7.140	0	1	0
19	3-Methyl,but-1-yl	3-Methyl,thio	F	7.440	0	1	0
20	3,3-Dimethyl,but-1-yl	3-Cyano	H	7.010	0	0	1
21	3,3-Dimethyl,but-1-yl	3-(Tetrazol-5-yl)	H	7.640	0	0	1
22	3,3-Dimethyl,but-1-yl	3-Trifluoro,methoxy	H	6.950	0	0	1
23	3,3-Dimethyl,but-1-yl	3-*N*,*N*-Dimethyl,amino	F	7.710	1	1	1
24	2-Cyclopentyl,prop-2-yl	3-(Tetrazol-5-yl)	H	7.330	1	0	1
25	2-Cyclopentyl,ethyl	3-*N*,*N*-Dimethyl,amino	F	7.320	0	1	0
26	2-Cyclopentyl,ethyl	3-Methyl,thio	F	7.650	0	1	0
27	Adament-1-yl,methyl	3-Methyl	H	6.600	0	0	0
28	Adament-1-yl,methyl	3-Nitro	H	6.520	0	0	0
29	Adament-1-yl,methyl	3-Bromo	H	6.720	0	0	0
30	Adament-1-yl,methyl	3-Ethoxy,carbonyl	H	6.090	0	0	0
31	Adament-1-yl,methyl	3-Carboxy	H	6.200	0	0	0
32	Adament-1-yl,methyl	3-*N*,*N*-Dimethyl,amino	H	6.350	0	0	0
33	Adament-1-yl,methyl	3-(Morpholin-4-yl)methyl	H	4.700	0	0	0
34	Adament-1-yl,methyl	(1,2-Dihydroxy propyl)amino	H	5.130	0	0	0
35	Adament-1-yl,methyl	3-Acetamido	H	5.820	0	0	0
36	Adament-1-yl,methyl	3-Formyl	H	5.950	0	0	0
37	Adament-1-yl,ethyl	3-*N*,*N*-Dimethyl,amino	H	6.530	0	0	0
38	Adament-2-yl,	3-*N*,*N*-Dimethyl,amino	H	6.850	0	0	0
39	Bicyclo(2,2,1)hept-2-yl	3-*N*,*N*-Dimethyl,amino	H	7.350	0	0	0
40	Bicyclo(2,2,1)hept-2-yl	3-Chloro	H	7.320	0	0	0

Here and hereafter, n is the number of compounds, Se is the standard error of estimation, R is correlation coefficient, F is the F ratio and p is the probability.

The aforementioned eq 1 shows that Sz is negatively linearly correlated with pK_i (*CCK-A*). That is, with decrease in the magnitude of the Sz index, pK_i (*CCK-A*) goes on increasing. It is worth mentioning that the corresponding mono-parametric model based on the W index is found to be inferior. This suggests that in spite of high collinearity between the W and Sz indices, the latter index has some hither to unknown information content.

Table 2. Topological indices calculated for the compounds used in the present study (refer Table 1)

Compd	*W*	χ	*B*	*J*	*Sz*	log *RB*
1	3332	16.9516	16.9516	1.4047	4846	928.025
2	3599	17.3074	17.3074	1.4099	5195	995.138
3	3599	17.3074	17.3074	1.4099	5195	995.138
4	3493	17.3074	17.3074	1.4961	5256	977.436
5	3447	17.2527	17.2527	1.5187	5181	969.665
6	3474	17.3242	17.3242	1.5050	5208	974.568
7	3503	17.1873	17.1873	1.4922	5237	978.977
8	3732	17.5980	17.5980	1.5137	5580	1041.413
9	3894	18.5020	18.5020	1.4065	5861	1092.535
10	3852	18.4127	18.4127	1.4193	5856	1085.896
11	3787	18.5365	18.5365	1.3135	7286	1071.979
12	4734	19.9491	19.9491	1.1957	8181	1306.445
13	4734	19.9491	19.9491	1.1957	8181	1306.445
14	4848	19.9187	19.9187	1.0417	8452	1324.390
15	4785	19.2906	19.2906	1.4523	6988	1281.335
16	4466	18.6119	18.6119	1.4626	6574	1202.750
17	4466	18.6119	18.6119	1.4626	6574	1202.750
18	4466	18.6119	18.6119	1.4626	6574	1202.750
19	4128	18.2392	18.2392	1.4724	6121	1125.325
20	4138	18.1192	18.1192	1.4692	6100	1126.865
21	6074	20.4750	20.4750	1.3986	8361	1555.235
22	4816	18.7925	18.7925	1.4545	7006	1282.408
23	4743	18.9025	18.9025	1.4808	6945	1274.342
24	6631	21.2638	21.2638	1.3740	9008	1663.332
25	4888	19.7273	19.7273	1.2592	7442	1324.737
26	4528	19.3546	19.3546	1.2714	6954	1242.510
27	5094	20.3429	20.3429	1.1864	8706	1390.094
28	5900	21.2536	21.2536	1.1631	10060	1570.024
29	5094	20.3429	20.3429	1.1863	8706	1390.094
30	6841	22.2916	22.2916	1.1341	11961	1768.264
31	6348	21.7916	21.7916	1.1498	11033	1666.332
32	5900	21.2536	21.2536	1.1631	10060	1570.024
33	8700	23.8513	23.8513	0.8848	16129	2126.192
34	8392	23.6510	23.6510	1.0973	15455	2083.133
35	7584	22.6475	22.6475	1.0893	13336	1893.211
36	6798	22.1463	22.1463	1.1396	11892	1763.330
37	5900	21.2536	21.2536	1.1637	10060	1570.024
38	5448	20.8006	20.8006	1.1866	9396	1469.228
39	4785	19.2906	19.2906	1.4524	6988	1281.335
40	5104	19.6365	19.6365	1.4572	7428	1357.862

Successive regression resulted into two bi-parametric models having better statistics than that involved in eq 1. However, the bi-parametric model containing Sz and B gave better results. This model is expressed by the following equation:

$$pK_i(CCK-A) = 0.6909(\pm 0.1093)B - 6.1111 \times 10^{-4} \left(\pm 7.2783 \times 10^{-5}\right)Sz - 1.9080 \quad (2)$$

$n = 40$, $Se = 0.3001$, $R = 0.8819$, $F = 64.723$, $p = 8.300 \times 10^{-13}$

Eq 2 shows that extent of branching in the set of compounds used has a positive and dominant role in the exhibition of pK_i (*CCK-A*) activity.

Further attempts of obtaining still better model(s) resulted into four statistically significant tri-parameteric models containing: (i) Sz, Ip_2, Ip_3; (ii) Sz, B, Ip_4; (iii) Sz, Ip_2, Ip_3; and (iv) Sz, J, B. Out of these four tri-parametric models, the model based on Sz, Ip_2, I_{p3} was found to be inferior than the bi-parametric model discussed above. The remaining three models were found to be statistically of better quality than the model expressed by eq 2. All these models invariably contain the B index as one of the correlating parameters. This again establishes the positive and dominating role of branching in the exhibition of pK_i (*CCK-A*) activity. It is worth mentioning that among the tri-parametric model, the models based on Sz, B and Ip_3 gave better results. The model is found as:

$$pK_i(CCK-A) = 0.7841(\pm 0.0974)B - 6.5087 \times 10^{-4} \left(\pm 6.3611 \times 10^{-5}\right)Sz + 0.4177(\pm 0.1123)Ip_3 - 3.4990 \quad (3)$$

$n = 40$, $Se = 0.2585$, $R = 0.9162$, $F = 62.745$, $p = 4.000 \times 10^{-14}$

The improved quality of the model expressed by eq 3 is due to the addition of the Ip_3 term. Note that this is an indicator parameter for the presence of tri-substitution at R_1. Thus, tri-substitution at R_3 favors the exhibition of pK_i (*CCK-A*).

Table 3. Correlation matrix for the data presented in Tables 1 and 2

	pK_i (*CCK-A*)	*W*	*B* (=χ)	*J*	*Sz*	log *RB*	Ip_1	Ip_2	Ip_3
pK_i (*CCK-A*)	1.0000								
W	−0.6121	1.0000							
B (=χ)	−0.5951	0.9720	1.0000						
J	0.6991	−0.7615	−0.8478	1.0000					
Sz	−0.7332	0.9696	0.9711	−0.8579	1.0000				
log *RB*	−0.6064	0.9972	0.9858	−0.7887	0.9749	1.0000			
Ip_1	0.0168	−0.6267	−0.6026	0.3103	−0.5296	−0.6314	1.0000		
Ip_2	0.4218	−0.3135	−0.3376	0.3915	−0.3481	−0.3195	−0.0894	1.0000	
Ip_3	0.3535	−0.2847	−0.3835	0.4899	0.3354	−0.3056	0.1291	0.0000	1.0000

The aforementioned model (eq 3) indicated that in addition to the topological indices *Sz* and *B*, the indicator parameters like-wise have dominant role in developing statistically better models. We have, therefore, tried to obtain tetra- and penta-parametric models in that we have used in addition to *Sz* and *B*, different combinations of indicator parameters Ip_1, Ip_2, and Ip_3. Out of the tetra-parametric models attempted, the model based on the combination of *Sz*, *B*, Ip_1 and Ip_2 gave better results than the model expressed by eq 3. This model is found as:

$$pK_i(CCK-A) = 0.5453(\pm 0.1051)B - 5.4204 \times 10^{-14}\left(\pm 6.4815 \times 10^{-5}\right)Sz - 0.3521(\pm 0.1183)Ip_1 + 0.1444(\pm 0.1077)Ip_2 + 0.4867 \quad (4)$$

$n = 40$, $Se = 0.2555$, $R = 0.9206$, $F = 48.660$, $p = 8.000 \times 10^{-14}$

The eq 4, therefore, shows that the improved statistics is due to the presence of indictor parameters Ip_1 and Ip_2. Thus, presence of H at R_2 and F at R_3 favors the exhibition of pK_i (*CCK-A*).

Finally, successive regression analyses resulted into two penta-parametric models having better qualities than the model expressed by eq 4. These models were found to contain the following parameters: (i) *Sz, B, J,* Ip_1 and Ip_2 and (ii) *Sz, B,* Ip_1, Ip_2 and Ip_3; respectively. The latter model was found to be of highest quality. This, most significant model was found to contain:

$$pK_i(CCK-A) = 0.6843(\pm 0.0921)B - 5.9595 \times 10^{-4}\left(\pm 5.46060 \times 10^{-5}\right)Sz - 0.2346(\pm 0.1008)Ip_1 + 0.2349(\pm 0.0908)Ip_2 + 0.4061(\pm 0.0994)Ip_3 - 1.9522 \quad (5)$$

$n = 40$, $Se = 0.2094$, $R = 0.9490$, $F = 61.587$, $p = 0.000$

The aforementioned results indicate that all the proposed models contained *B* index whose coefficient ranges between 0.5453 and 0.7841, and is larger than the coefficients of other parameters involved in the proposed models. Also that, coefficients of *B* in each case were larger than the corresponding standard deviations. This shows the relative dominating role of branching in the exhibition of pK_i (*CCK-A*) activity.

The model expressed by eq 5 represents a highly significant correlation accounting for 90.06% of variance in the activity ($R^2 = 0.9006$). Such an excellent variance in activity resulted due to the appearance of indicator

Table 4. Regression parameters and qualities of the correlations attempted in the present study

S.N.	Parameters used	A_i $i = 1, 2, 3, 4, 5$	*B*	Standard error of est. (*Se*)	Corr. coeff. (*R*)	F-ratio	$Q = R/Se$	Prob.
1.	*W*	$-2.8064\times10^{-4}(\pm 5.8808\times10^{-5})$	8.1282	0.4966	−0.6121	22.773	1.2325	2.699×10^{-5}
2.	*B*	−0.2003(±0.0439)	10.6502	0.5047	−0.5951	20.837	1.1791	5.118×10^{-5}
3.	*J*	2.6338(±0.4371)	3.2484	0.4491	0.6990	36.315	1.5564	5.240×10^{-7}
4.	*Sz*	$-1.6432\times10^{-4}(\pm 2.4720\times10^{-5})$	8.0449	0.4271	−0.7332	44.184	1.7167	7.410×10^{-8}
5.	log *RB*	$-0.0012(\pm 2.60412\times10^{-4})$	8.3598	0.4494	−0.6064	22.094	1.3494	3.369×10^{-5}
6.	*B*	0.6909(±0.1093)	−1.9080	0.3001	0.8819	64.723	2.9386	8.300×10^{-13}
	Sz	$-6.1111\times10^{-4}(\pm 7.2783\times10^{-5})$						
7.	*Sz*	$-1.5521\times10^{-4}(\pm 2.6214\times10^{-5})$	7.9346	0.4266	0.7921	22.672	1.8567	3.737×10^{-7}
	Ip_3	0.1855(±0.1790)						
8.	*Sz*	$-1.3681\times10^{-4}(\pm 2.7368\times10^{-5})$	7.7044	0.4139	0.7671	17.162	1.8533	4.366×10^{-7}
	IP_2	0.2960(±0.1626)						
	IP_3	0.2276(±0.1752)						
9.	*B*	0.7841(±0.0974)	−3.4990	0.2585	0.9162	62.745	3.5442	4.000×10^{-14}
	Sz	$-6.5087\times10^{-4}(\pm 6.3611\times10^{-5})$						
	IP_3	0.4177(±0.1123)						
10.	*W*	$3.80365\times10^{-4}(\pm 1.6042\times10^{-4})$	−0.2094	0.2829	0.6987	50.413	3.1767	4.606×10^{-13}
	B	0.5415(±0.1208)						
	Sz	$-6.9476\times10^{-4}(\pm 7.7158\times10^{-5})$						
11.	*B*	0.7244(±0.1014)	−4.9022	0.2765	0.9035	53.339	3.2676	2.100×10^{-13}
	J	1.4517 (±0.5275)						
	Sz	$-5.5873\times10^{-4}(\pm 6.9718\times10^{-5})$						
12.	*B*	0.5453(±0.1051)	0.4867	0.2555	0.9206	48.660	3.6031	8.000×10^{-14}
	SZ	$-5.4204\times10^{-4}(\pm 6.4815\times10^{-5})$						
	Ip_1	−0.3521(±0.1183)						
	Ip_2	0.1444(±0.1077)						
13.	*B*	0.6363(±0.0995)	−1.3822	0.2283	0.9390	50.720	4.1130	6.000×10^{-14}
	J	0.2489(±0.5164)						
	Sz	$-5.7736\times10^{-4}(\pm 6.1232\times10^{-5})$						
	Ip_1	−0.3242(±0.1082)						
	Ip_3	0.3295(±0.1084)						
14.	*B*	0.6843(±0.0921)	−1.9522	0.2094	0.9490	61.587	4.5391	0.000
	Sz	$-5.999510^{-4}(\pm 5.4606\times10^{-5})$						
	Ip_1	−0.2346(±0.1008)						
	Ip_2	0.2349(±0.0908)						
	Ip_3	0.4061(±0.0994)						

parameters Ip_1, Ip_2 and Ip_3 in the model. This eq 5 shows that the coefficient of Ip_1 term is negative while those of Ip_2 and Ip_3 are positive. Thus, the obtained variance is favored by substitutions at R_2 and R_3 respectively while tri-substitution at R_1 has a negative role in the exhibition of pK_i (*CCK-A*) activity.

In order to confirm our findings we have calculated the quality factor Q for the models expressed by eqs 1–5. This quality factor is defined[15,16] as the ratio of correlation coefficient R to the standard error of estimation (Se), that is $Q = R/Se$ and thus accounts for the role of R and Se simultaneously in obtaining statistically significant models. In our case Q values are found as: 1.7167, 2.9387, 3.5443, 3.6031, and 4.5320, respectively, for the models expressed by eqs 1–5. This shows that the quality of models goes on increasing as we pass on from bi- to penta-parametric models and attains the optimum value for the penta-parametric model (eq 5).

The most appropriate model (eq 5) does not have any serious outlier. However, compounds **11** and **17** gave comparatively slightly higher residues. On deleting these two compounds from the regression (eq 5) a better quality regression expression is obtained:

$$pK_i(CCK-A) = 0.6966(\pm 0.0770)B - 5.9906 \times 10^{-4}(\pm 4.9212 \times 10^{-5})Sz - 0.2331(\pm 0.0821)Ip_1 + 0.3375(\pm 0.0829)Ip_2 + 0.4411(\pm 0.0811)Ip_3 - 2.0211 \quad (6)$$

$n = 38$, $Se = 0.1974$, $R = 0.9574$, $F = 70.380$, $Q = 4.8501$, $p = 2.000 \times 10^{-1}$

Now this modified model (eq 6) accounts for 91.66% ($R^2 = 0.9166$) variance in the activity. However, both these models (eqs 5 and 6) suggests that *CCK-A* binding is largely governed by branching.

It is worth mentioning that one more penta-parametric model of the better quality than the tetra-parametric model (eq 4) was also obtained (Tables 4 and 5). This model involves three topological indices Sz, B, and J along with indicator parameters Ip_1 and Ip_3, but suffers from the defect that coefficient of J term in this model was smaller than its standard deviation. Such models are not statistically allowed and thus discarded.

All the attempted (statistically significant) models are presented in Tables 4 and 5.

In order to confirm our findings we have estimated pK_i (*CCK-A*) values using each of the models expressed by eqs 4–6 and compared them with the observed (experimental) values of pK_i (*CCK-A*). Such a comparison is shown in Table 6. The data presented in this Table 6 show that pK_i (*CCK-A*) estimated from the model expressed by eq 6 are very close to the observed values. Furthermore, we have also estimated residue, that is the difference between observed (experimental) and calculated (estimated) pK_i *(CCK-A)* values. Such residues are also shown in Table 6 suggesting again that the model expressed by eq 6 is the most appropriate model for modeling the pK_i (*CCK-A*) activity.

Now, in order to infer about the predictive potential of the most appropriate models (eqs 4–6) we have estimated cross-validation[17–19] parameters (Table 7).

PRESS is a good estimate of the real prediction error of the models. If *PRESS* is smaller than the sum of the squares of the response value (*SSY*), the model predicts better than chance and can be considered statistically significant also. In our case all the three models (eqs 4–6) have $PRESS << SSY$ (Table 7) confirming that these models predict better than chance and are statistically significant. Furthermore, to be a reasonable QSAR model, $PRESS/SSY$ should be smaller than 0.4, and a value of this ratio smaller than 0.1 indicates an excellent model. The $PRESS/SSY$ values in the present case are in favor of this finding. Also, $PRESS/SSY$ was found to be 0.0909, that is smaller than 0.1 for the model expressed by eq 6, thereby indicating it to be an excellent model.

If the *PRESS* value is transformed in to a dimensionless term by relating it to the initial sum of squares one obtains R^2_{CV}, that is the component of the fraction of

Table 5. Various regression expressions attempted

Model no.	Regression expression
1.	pK_i (*CCK-A*) $= -2.8064\times10^{-4}(\pm 5.88080\times10^{-5})W + 8.1282$.
2.	pK_i (*CCK-A*) $= -0.2003(\pm 0.0439)B + 10.6502$.
3.	pK_i (*CCK-A*) $= 2.6338(\pm 0.4371)J + 3.2484$.
4.	pK_i (*CCK-A*) $= -1.6432\times10^{-4}$ $(\pm 2.47199\times10^{-5})Sz + 8.0449$
5.	pK_i (*CCK-A*) $= -0.0012(\pm 2.60412\times10^{-4})\log Rb + 8.3598$.
6.	pK_i (*CCK-A*) $= 0.6909(\pm 0.1093)B - 6.1111\times10^{-4}$ $(\pm 7.278310^{-5})Sz - 1.9080$.
7.	pK_i (*CCK-A*) $= -1.5521\times10^{-4}(\pm 2.62138\times10^{-5})Sz + 0.1855(\pm 0.1790)Ip_3 + 7.9346$.
8.	pK_i (*CCK-A*) $= -1.3681\times10^{-4}(\pm 2.73678\times10^{-5})Sz + 0.2960(\pm 0.1626)$ $Ip_2 + 0.2276(\pm 0.1752)$ $Ip_3 + 7.7044$.
9.	pK_i (*CCK-A*) $= 0.7841(\pm 0.0974)B - 6.5087\times10^{-4}(\pm 6.3611\times10^{-5})Sz + 0.4177(\pm 0.1123)$ $Ip_3 - 3.4990$.
10.	pK_i (*CCK-A*) $= 3.80365\times10^{-4}(\pm 1.60415\times10^{-4})W + 0.5415(\pm 0.1208)B - 6.9476\times10^{4}(\pm 7.7158\times10^{-5})Sz - 0.2094$.
11.	pK_i (*CCK-A*) $= 0.7244(\pm 0.1014)B + 1.4517$ $(\pm 0.5275)J - 5.5873\times10^{-4}(\pm 6.9718\times10^{-5})Sz - 4.9022$.
12.	pK_i (*CCK-A*) $= 0.5453(\pm 0.1051)B - 5.4204\times10^{-4}(\pm 6.4815\times10^{-5})Sz - 0.3521(\pm 0.1183)Ip_1 + 0.1444(\pm 0.1077)$ $Ip_2 + 0.4867$.
13.	pK_i (*CCK-A*) $= 0.6363(\pm 0.0995)B + 0.2489(\pm 0.5164)J - 5.7736\times10^{-4}(\pm 6.12315\times10^{-5})Sz - 0.3242(\pm 0.1082)$ $Ip_1 + 0.3295(\pm 0.1084)$ $Ip_3 - 1.3822$.
14.	pK_i (*CCK-A*) $= 0.6843(\pm 0.0921)B - 5.9595\times10^{-4}(\pm 5.46060\times10^{-5}) - 0.2346(\pm 0.1008)$ $Ip_1 + 0.2349(\pm 0.0908)$ $Ip_2 + 0.4061(\pm 0.0994)$ $Ip_3 - 1.9522$.

the unexplained variance over the total variance. Such values as calculated from the following relationships are recorded in Table 7. Table 7 shows that R^2_{CV} is highest for the model expressed by eq 6, thus, confirming again our earlier findings.

$$R^2_{CV} = 1 - \frac{PRESS}{SSY} \quad (7)$$

It is interesting to record that uncertainty of the prediction (S_{PRESS}),[17–20] except for the model expressed by eq 5 are found to be the same as that of their standard error of estimation (Se). This shows that S_{PRESS} cannot be used as a parameter in deciding predictive potential of the proposed models. In view of this we have calculated predictive squared error (PSE)[17] for the proposed models. The predictive squared error (PSE) generally found to be more directly related to the uncertainty of the predictions. However, in our case the PSE values recorded in Table 7 show that except for the model expressed by eq 4, PSE is the same as that of S_{PRESS}. This shows that both S_{PRESS} and PSE are useless in the present case for deciding predictive potential of the proposed models.

Table 6. Estimated pK_i (*CCK-A*) and its comparison with the observed values

Compd	Obsd pK_i (*CCK-A*)	Estimated pK_i (*CCK-A*) from					
		Eq (4)		Eq (5)		Eq (6)	
		Est.	Res.	Est.	Res.	Est.	Res.
1	6.260	6.752	−0.4918	6.525	−0.2650	6.480	−0.2205
2	6.490	6.757	−0.2666	6.560	−0.0705	6.517	−0.0269
3	6.740	6.757	−0.0166	6.560	0.1795	6.517	0.2231
4	6.710	6.868	−0.1580	6.759	−0.0491	6.765	−0.0555
5	6.940	6.734	0.2056	6.937	0.0025	6.899	0.0412
6	6.910	6.759	0.1513	6.970	−0.0603	6.932	−0.0223
7	6.950	6.668	0.2816	6.859	0.0907	6.819	0.1309
8	7.150	6.851	0.2992	7.171	−0.0209	7.183	−0.0332
9	7.110	7.047	0.0629	6.981	0.1290	6.946	0.1636
10	7.000	7.145	−0.1455	7.158	−0.1578	7.173	−0.1727
11	6.570	6.293	0.2765	6.155	0.4146	Outlier	
12	6.320	6.579	−0.2586	6.589	−0.2687	6.547	−0.2272
13	6.660	6.579	0.0814	6.589	0.0713	6.547	0.1128
14	6.660	6.415	0.2448	6.406	0.2537	6.361	0.2987
15	6.600	6.866	−0.2662	6.849	−0.2490	6.812	−0.2122
16	6.900	7.217	−0.3170	7.101	−0.2008	7.159	−0.2592
17	6.700	7.217	−0.5170	7.107	−0.4008	Outlier	
18	7.140	7.217	−0.0770	7.101	0.0392	7.159	−0.0192
19	7.440	7.259	0.1807	7.116	0.3242	7.174	0.2658
20	7.010	7.061	−0.0509	7.217	−0.2073	7.229	−0.2191
21	7.640	7.120	0.5200	7.482	0.1581	7.500	0.1397
22	6.950	6.937	0.0131	7.138	−0.1881	7.149	−0.1988
23	7.710	7.174	0.5356	7.485	0.2254	7.548	0.1616
24	7.330	7.199	0.1306	7.230	0.1000	7.246	0.0836
25	7.320	7.355	−0.0348	7.347	−0.0268	7.411	−0.0907
26	7.650	7.416	0.2339	7.383	0.2674	7.447	0.2031
27	6.600	6.861	−0.2609	6.780	−0.1798	6.787	−0.1868
28	6.520	6.624	−0.1036	6.596	−0.0761	6.600	−0.0800
29	6.720	6.861	−0.1409	6.780	−0.0598	6.787	−0.0668
30	6.090	6.159	−0.0692	6.173	−0.0835	6.170	−0.0798
31	6.200	6.390	−0.1896	6.384	−0.1844	6.385	−0.1845
32	6.350	6.624	−0.2736	6.596	−0.2461	6.600	−0.2500
33	4.700	4.751	−0.0505	4.757	−0.0568	4.726	−0.0263
34	5.130	5.007	0.1234	5.021	0.1086	4.996	0.1340
35	5.820	5.608	0.2120	5.598	0.2224	5.583	0.2372
36	5.950	6.117	−0.1674	6.115	−0.1652	6.110	−0.1602
37	6.530	6.624	−0.0936	6.596	−0.0661	6.600	−0.0700
38	6.850	6.736	0.1135	6.682	0.1682	6.687	0.1628
39	7.350	7.218	0.1317	7.084	0.2664	7.096	0.2541
40	7.320	7.168	0.1516	7.058	0.2619	7.070	0.2499

Table 7. Cross-validation parameters and root mean square error of residues ($RMSR$) for the statistically most significant models

Model	Parameters involved	$PRESS$	SSY	$\frac{PRESS}{SSY}$	R^2_A	R^2_{CV}	S_{PRESS}	PSE	$RMSR$
1 [eq (4)]	Sz, B, Ip_1, Ip_2	2.2845	12.7045	0.1798	0.8302	0.8202	0.2555	0.2390	0.2390
2 [eq (5)]	Sz, B, Ip_1, Ip_2, Ip_3	1.4904	13.4986	0.1104	0.8859	0.8895	0.1930	0.1930	0.1930
3 [eq (6)][a]	Sz, B, Ip_1, Ip_2, Ip_3	1.2473	13.7163	0.0909	0.9036	0.9091	0.1812	0.1812	0.1712

[a]Considering compounds **11** and **17** as outliers.

The aforementioned failure prompted us to estimate root-mean-square error of the residue (*RMSR*) which is another method of deciding predictive potential of the models.[17] The *RMSR* values are estimated from the following relationship:

$$RMSR = \sqrt{\frac{\Sigma\{\text{observed } pK_i(CCK\text{-}A) - \text{estimated } pK_i(CCK-A)\}^2}{n}} \quad (8)$$

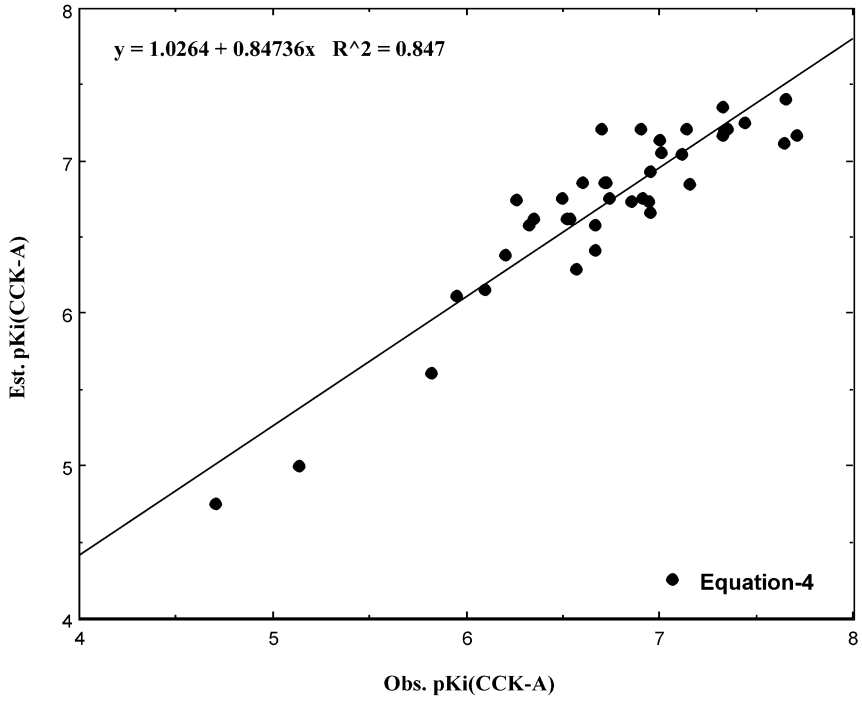

Figure 2. Correlation of observed pK_i (*CCK-A*) with their estimated values using eq 4.

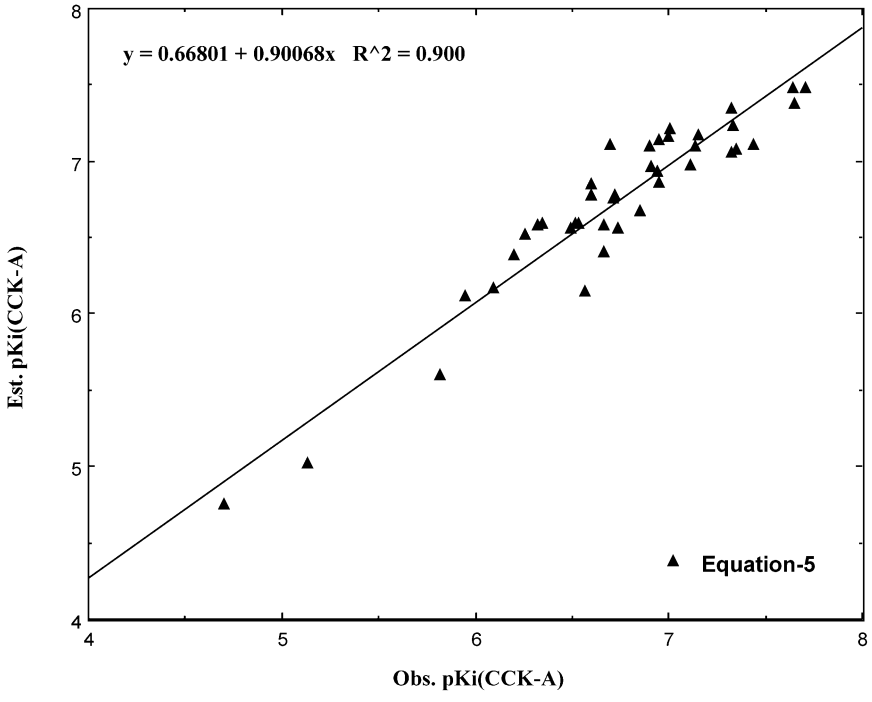

Figure 3. Correlation of observed pK_i (*CCK-A*) with their estimated values using eq 5.

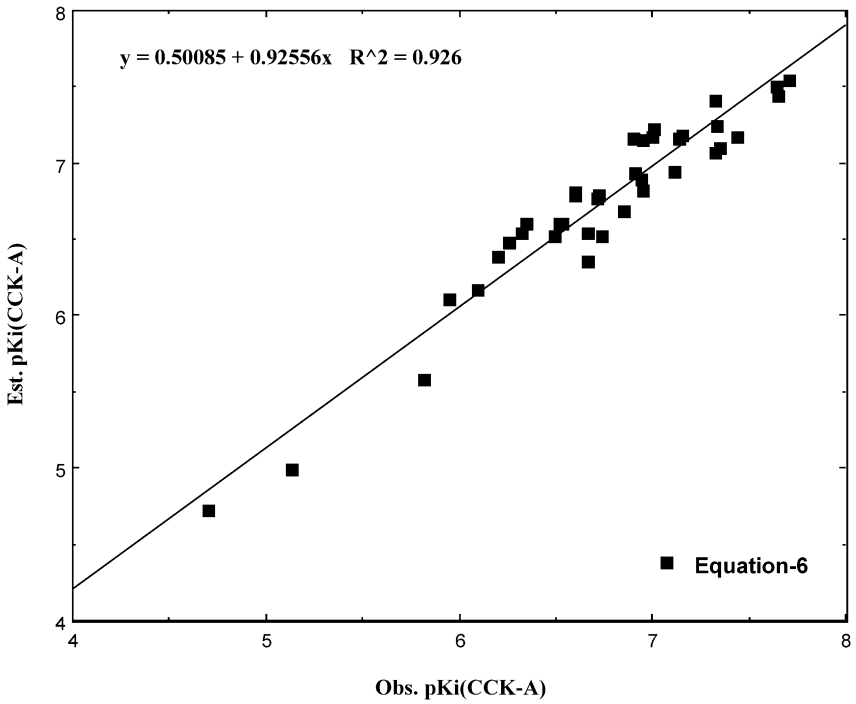

Figure 4. Correlation of observed pK_i (*CCK-A*) with their estimated values using eq 6.

The values of *RMSR* thus calculated are found as: 0.2390, 0.1930, 0.1712, respectively, for the models expressed by eqs 4–6 (Table 7). This shows that *RMSR* is the most appropriate parameter for deciding predictive potential of the model. The lowest value of *RMSR* for the model expressed by eq 6 suggests it to have the highest predictive potential.

In order to further confirm our findings we have obtained relationships between observed and estimated pK_i (*CCK-A*) values. Such relationships are shown in Figures 2–4. The perusal of Figures 2–4 suggests that the predictive correlation coefficient ($R^2_{\text{predictive}}$) is the highest for the model expressed by eq 6 confirming again that it has got the highest predictive potential.

Before concluding the discussion, it is necessary to remark on the involvement of topological indices *Sz* and *B* in the aforementioned models. As stated earlier, both these indices, *Sz* and *B*, are highly linearly correlated and thus, any model containing them simultaneously may suffer from the defect of collinearity. However, in our earlier communiacations,[21,22] we have stated that if a variable is added that does not contribute its fair share then R^2_A will actually decline. However, here in the proposed models R^2_A goes on increasing from bi- to poly-parametric models indicating that collinearity defect is not serious. Furthermore, the coefficients of *Sz* and *B* terms in all the proposed models are found to be considerably higher than their respective standard deviation. Such parameters and the resulting models are considered statistically significant. Thus, the high collinearity between *Sz* and *B* indices does not make the correlation involving them as statistically insignificant correlation. Randic[23] has further clarified and offered justification in support of models containing highly correlated topological indices. According to Randic, in spite of collinearity, occurrence of highly correlated topological indices in the model is due to their different information contents. This is true in the present case because *Sz* and *B* indices have different information contents making their occurrence a statistically allowed model.

Conclusion

From the aforementioned results and discussions we conclude that:

1. The novel class of 1-alkyl-5-aryl-3-ureido-1,5-benzodiazepine-2,4-diones can be modelled successfully as *CCK-A* receptor antagonists using distance-based topological indices;
2. Out of the pool of topological indices used only *Sz* and *B* are found to be most suitable for modelling *CCK-A* activity;
3. Introduction of indicator parameters related to substituents at R_1, R_2, and R_3 enhances quality of correlations as well as predictive potential of the models; and
4. The substitution at the N-1 position with a suitable side chain has dramatic effect both on the activity as well as magnitude of topological indices used.

Experimental

Pharmaceutical activity

All the compounds used were tested by Ursini et al.[9] in vitro according to the method described in the litrature.[36,37] The affinities expressed as pK_i (*CCK-A*) were determined by radio-ligand binding studies using cerebral cortex membranes from guinea pig and rat pancreatic membranes. These values are adopted in the present study.

Topological indices

All the topological indices referred below are calculated from the carbon–hydrogen suppressed molecular graphs of the compounds used in the present study (Table 1) and are presented in Table 2.

Wiener index (W).[24] The Wiener index (W) is the oldest and widely used topological index. It is based on the vertex distances of the respective molecular graph.

Let us denote a molecular graph by G and having v_1, v_2, $v_3, \ldots, v_n$ as its vertices. Let $d(v_i, v_j|G)$ stand for the shortest distance between the vertices v_i and v_j. Then the Wiener index is defined as:

$$W = W(G) = 1/2\sum_{i=1}^{n}\sum_{j=1}^{n} d(v_i, v_j|G) \tag{9}$$

Szeged index (Sz).[25–28] Let e be an edge of the molecular graph G. Let $n_1(e|G)$ be the number of vertices of G lying closer to one end of e; let $n_2(e|G)$ be the number of vertices of G lying closer to the other end of e. Then the Szeged index (Sz) is defined[25,26] as:

$$Sz(G) = Sz = \sum_{e} n_1(e|G)n_2(e|G) \tag{10}$$

with the summation giving over all edges of G.

In cyclic graphs, there are edges equidistant from both the ends of edge e; by definition of Sz such edges are not taken into account.

The molecular connectivity indices.[29–31] The connectivity index $\chi = \chi(G)$ of a graph G is defined by Randic[29] as under

$$\chi = \chi(G) = \sum_{ij}\left[\delta_i\delta_j\right]^{-0.5} \tag{11}$$

where δ_i and δ_j are the valence of a vertex i and j, equal to the number of bonds connected to the atoms i and j, in G.

In the case of hetero-systems the connectivity is given in terms of valence delta values δ_i^v and δ_j^v of atoms i and j and is denoted by χ^v. This version of the connectivity index is called the valence connectivity index and is defined as[30,31] under:

$$\chi^v = \chi^v(G) = \sum_{ij}\left[\delta_i^v\delta_j^v\right]^{-0.5} \tag{12}$$

where the sum is taken over all bonds i–j of the molecule. Valence delta values are given by the following expression :

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_j - 1} \tag{13}$$

where Z_i is the atomic number of atom i, Z_i^v is the number of valence electron of the atom i and H_i is the number of hydrogen atoms attached to atom i. The δ_i^v values are available in the books of Kier and Hall.[30,31]

Recall that now-a-days the connectivity and the valence connectivity indices expressed by eqs 12 and 13 are termed as first-order connectivity and first-order valence connectivity indices, respectively. Lower or higher order connectivity indices are also possible which are defined analogously.

Balaban index (*J*).[14,32,33] The Balaban index, J (the average distance sum connectivity index) is defined[32] by:

$$J = J(G) = \frac{M}{\mu + 1}\sum_{\text{bonds}}\left(d_i d_j\right)^{-1/2} \tag{14}$$

where M is the number of bonds in a graph G, μ is the cyclomatic number of G and d_i's ($i = 1, 2, 3, \ldots, N$) are the distance sums (distance degrees) of atoms in G such that

$$d_i = \sum_{j=1}^{N}(D)_{ij} \tag{15}$$

The cyclomatic number μ of G indicates the number of independent cycles in G and is equal to the minimum number of cuts (removal of bonds) necessary to convert a polycylic structure into an acyclic structure:

$$\mu = M - N + 1 \tag{16}$$

For hetero-systems, the elements of distance matrix can be modified as below for calculating Balaban index J:
i: The diagonal elements:

$$(D)_{ij} = 1 - (Z_c/Z_i) \quad (17)$$

where $Z_c = 6$ and Z_i = atomic number of the given element.
ii. The off-diagonal elements:

$$(D)_{ij} d_i = \sum_r k_r \quad (18)$$

where the summation is over all bonds. The bond parameter k_r is given by:

$$k_r = 1/b_r\left(Z_c^2/Z_i Z_j\right) \quad (19)$$

where b_r is the bond weight with values: 1 for single bond, 2 for double bond, 1.5 for aromatic bond and 3 for triple bond. The values of $(D)_{ij}$ for various hetero-bonds and k_r for various types of hetero-bonds are given in the literature.[34,35]

Branching index (*B* and log *RB*).[34,35] The branching index *B* as well as log *RB* have been calculated by the method as described in the literature.[34,35]

Indicator parameters (Ip_1, Ip_2, Ip_3). Indicator variables (parameters), sometimes called dummy variables or de novo constants, are used in multiple linear regression analysis to account for certain features which cannot be described by continuous variables. In QSAR equations they normally describe a certain structural element, be it a substituent or another molecular fragment. Thus, Free Wilson analysis may be interpreted as a regression analysis approach using only indicator variables.

The indicator parameters (variables) take on only two values, usually zero and one. The two values signify that the observation belongs in one of the two possible categories. The numerical values of the dummy variables are not intended to reflect a quantitative ordering of categories, but only serve to identify category or class membership. Therefore, they show the significance of a particular group or a substituent in a given series of drug. They account for the abrupt increase or decrease of a given pharmacological activity at any specific site in the drug molecule. If the coefficient of indicator parameter carries a negative sign in the regression expression, this makes it very clear that the compound having this particular group at a particular position will have considerable lower potency.

In the present case, the indicator parameter Ip_1 was taken as unity when H was present at R_2 and in the absence of such substitution, Ip_1 was zero. Similarly, Ip_2 is used as unity when F is present at the substituent R_3, otherwise its value is taken as zero. Finally, Ip_3 is used as unity for tri-substitution at R_1.

Regression analysis. We have used the maximum R^2 improvement method to identify prediction models.[17–20] This method finds the 'best' one-variable model, the 'best' two variable model and so forth for the prediction of property/activity. Several models (combinations of variables) were examined to identify combinations of variables with good prediction capabilities. In all regression models developed, we have examined a variety of statistics associated with residues, that is the Wilks–Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results.[17–20] Finally, results are discussed on the basis of cross-validation parameters.

Multiple regression analyses for correlating antimalarial activities of the present set of compounds with the aforementioned molecular descriptors were carried out using *Regress-1* software as supplied by Professor I. Lukovits, Hungarian Academy of Sciences, Budapest, Hungary. Several multiple regressions were attempted using correlation matrix from this program and the best results are considered and discussed in developing QSAR and, hence, for modeling the antimalarial activities of the compounds in the present study.

Computations. All the computations were carried out in Power Macintosh 9600/233.

The authors are thankful to Professor Istavan Lukovits for providing software to carryout calculations of topological indices and for making regression analysis and to Prof. Ivan Gutman for introducing P.V.K. to this fascinating field of chemical graph theory and topology.

References and Notes

1. Mutt, V.; Jorpes, J. E.; Magnusson, S. *Eu. J. Biochem.* **1970**, *15*, 513.
2. Vanderhaeghen, J. J.; Signeau, J. C.; Gepts, W. *Nature (London)* **1975**, *257*, 604.
3. Dockray, G. J. *Nature (London)* **1976**, *264*, 586.
4. Dockray, G. J. *Br. Med. Bull.* **1982**, *38*, 253.
5. Crawley, J. N. *Ann. N. Y. Acad. Sci.* **1985**, *448*, 1.
6. Evans, B. E.; Bock, M. G.; Rittle, K. E.; Di Pardo, R. M.; Whitler, W. L.; Veber, D. R.; Anderson, P. S.; Freidinger, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4918.
7. Chang, R. S. L.; Lotti, V. J. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4923.
8. Evans, B. E.; Rittle, K. E.; Bock, M. G.; Di Pardo, R. M.; Freidinger, R. M.; Whitler, W. L.; Luden, G. F.; Veber, D. R.; Anderson, P. S.; Chang, R. S. L.; Lotti, V. J.; Crino, D. J.; Chen, T. B.; Kling, P. J.; Kunkel, K. A.; Springer, J. P.; Hirshfield, J. J. *J. Med. Chem.* **1988**, *31*, 2235.
9. Ursini, A.; Capelli, A. M.; Carr, R. A. E.; Cassara, P.; Corsi, M.; Curcuruto, O.; Curotto, G.; Cin, M. D.; Davalli, S.; Donati, D.; Feriani, A.; Finch, H.; Finizia, G.; Gaviraghi, G.; Marien, M.; Pentassuglia, G.; Polinelli, S.; Ratti, E.; Reggiani, A.; Tarzia, G.; Tedesco, G.; Tranquillini, M. E.; Trist, D. G.; Van Amsterdam, F. T. M. *J. Med. Chem.* **2000**, *43*, 3596.
10. Agrawal, V. K.; Karmarkar, S.; Khadikar, P. V. *SAR-QSAR Environ. Res.* **2001**, *12*, 529.
11. Karmarkar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K.; Bano, S. *Indian J. Chem.* **2001**, *40A*, 12.
12. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
13. Balaban, A. T. *Math. Chem.* **1986**, *21*, 115.
14. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chem. (Belg.)* **1997**, *106*, 767.
15. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
16. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.

17. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
18. Groebner, D. F.; Shanon, P. W. Business Statistics: A Decision Making Approach, Merrill Pub. Co., NY, USA.
19. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121.
20. Lucic, B.; Trinajstic, N.; Slid, S.; Karelson, M.; Katritzky, A. R. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610.
21. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3035.
22. Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. *Acta Pharm.* **2001**, *51*, 117.
23. Randic, M.; Basak, S. C. *J. Chem. Inf. Comput. Scic.* **2001**, *41*, 614.
24. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
25. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
26. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
27. Mandloi, M.; Sikarwar, A.; Sapre, N.; Karmarkar, S.; Khadikar, P. V. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57.
28. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *Commu. Math. Chem. (MATCH)* **2001**, *43*, 7.
29. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
30. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry; Drug Research*; Academic: New York, 1976.
31. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Relationship*; Academic: NewYork, 1986.
32. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
33. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
34. Todeschini, R.; Cosonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
35. Diudea, M. V., Ed. *QSPR/QSAR Studies by Molecular Descriptors*; Babes-Bolyai University: Cluj, Romania, 2000.
36. Innis, R. B.; Snyder, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 6917.