Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 6137–6145
Bioorganic & Medicinal Chemistry

Quantitative structure–activity relationship study using refractotopological state atom index on some neonicotinoid insecticides

Bikash Debnath,[a] Shovanlal Gayen,[a] Anindya Basu,[a] Balaram Ghosh,[a] Kolluru Srikanth[b] and Tarun Jha[a,*]

[a]*Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, PO Box No 17020, Jadavpur University, Kolkata 700 032, India*
[b]*Department of Medicinal Chemistry, College of Pharmacy, 8-125 Weaver Densford Hall, 308, Harvard St SE, University of Minnesota, Minneapolis, MN 55455, USA*

Received 28 July 2004; accepted 2 September 2004
Available online 23 September 2004

Abstract—Importance of atom-level topological descriptors like electrotopological state atom (E-state) index in QSAR study is increasing. These descriptors help to relate structure and activity at atomic/fragmental level. In view of the earlier success of E-state index on some azidopyridinyl neonicotinoid insecticides, a relatively new atom-level topological descriptor; refractotopological state atom (R-state) index was used in this work. This was used to identify the important atoms/fragments related to dispersive/van der Waals interactions of neonicotinoids with the nicotinic acetylcholine receptor (nAChR). This study showed the structural requirements for the mammal $\alpha_4\beta_2$ and *Drosophila* nAChR agonistic activity. It also revealed that substituted imine, nitromethylene at X-position were selective to the insecticidal activity. Azido substitution at pyridine ring of neonicotinoids disfavored the binding with the receptors. This study confirmed the validity of the R-state index as a new tool for quantitative structure–activity relationships. It has the ability to find out the required structural features as well as to predict the activity of the neonicotinoids.

1. Introduction

Topological methods are applied to quantify the differences in structures by assigning some numerical value to the graph of the molecule. Thus, significant correlations of chemical structures with physical, chemical, and biological properties are possible.[1–3] The most successful approach to the quantitative evaluation of molecular structure is molecular connectivity indices of Kier and co-workers.[4] The nonempirical connectivity indices have been successfully used in quantitative structure–property relationship (QSPR) studies of organic molecules[5] and in quantitative structure–activity relationship (QSAR) studies of several bioactive compounds.[6] However, the information at the level of atoms/fragments and bonds are suppressed into these whole molecular indices. The generation of atom-level topological descriptors from the chemical graph of the molecule was first attempted by Kier and Hall. They developed electrotopological state atom (E-state) index[7–9] combining both electronic and topological description of atoms in the molecule. E-state indices were derived from electron density distributed over an atom according to its degree of bonds to nonhydrogen atoms. In view of the successful application of E-state index to specify the pharmacophoric/toxicophoric atoms in many QSAR studies, Carrasco et al. introduced refractotopological state atom (R-state) index[10] for modeling dispersive and hydrophobic interactions in an almost similar fashion of E-state formalism. The R-state indices are calculated from the atomic refractivity. These atomic refractivities are based on the assignment of atomic contributions as reported by Ghosh and coworkers.[11,12] The index does not only describe the representation of the atomic dispersive forces related to the molecular refractivity but also the influence of bonded

Keywords: Topology; Refractotopological state atom index; QSAR study; Azidopyridinyl neonicotinoids.
* Corresponding author. Tel./fax: +91 33 24146677; e-mail: tjjupharm@yahoo.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.09.003

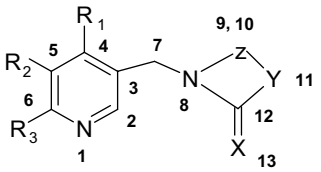


Figure 1. General structure of azidopyridinyl neonicotinoids with arbitrarily numbered common skeletal atoms.

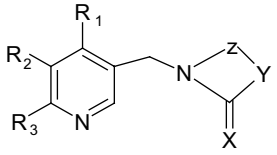
and nonbonded atoms as a measure of the distance effect of other groups in the molecule.

Neonicotinoids, the effective synthetic insecticides, have been increasingly used due to the remarkable selectivity toward the insect (*Drosophila*) nicotinic acetylcholine receptor (nAChR) over mammalian nAChR ($\alpha_4\beta_2$ subtype). This is a very useful class of compounds for the control of ectoparasites in animals.[13] In order to gain information about hydrophobic and dispersive interactions at the acetylcholine receptor surface as well as the required structural features for their selective binding to the insect nAChR receptor at atom/fragmental level, QSAR study was performed using R-state indices on some azidopyridinyl neonicotinoids reported by Zhang et al.[13] General structure of these neonicotinoids is shown in Figure 1. This work is a part of our composite program of rational drug design.[14–26] The same data set was used earlier[26] for identification of the structural requirements related to the electronic interactions of the neonicotinoids with receptor-binding sites using E-state indices. Here R-state index was tried for the comparison with that of E-state index in the same set of molecules as well as to find out the important common atoms for both the studies so that the useful pharmacophore can be suggested. E-state indices depend on topological connections as well as pi and lone pair electrons, which are more reactive and closely associated with long-range noncovalent intermolecular interactions such as drug receptor encounters.[7,8] Thus, E-state indices in QSAR studies are important to specify the atoms of ligands related with electronic interaction with the receptor. R-state indices are based on the atomic refractivities and the topological environment of the atom.[10] Sum of atomic refractivities, that is, molar refractivity is directly proportional to the polarizability of a substance, which determines London force/dispersive force between nonpolar molecules.[11,12] Thus, R-state indices is important for modeling the dispersive/van der Waals interactions with the receptor.

2. Results and discussion

The biological activity (mammalian $\alpha_4\beta_2$ and *Drosophila* nAChR agonistic activity) data of the neonicotinoids are presented in Table 1 and the R-state indices as well as indicator parameters are listed in Table 2. Correlation

Table 1. Agonistic activity data of neonicotinoids for mammalian $\alpha_4\beta_2$ [$K_i(\alpha_4\beta_2)$] and *Drosophila* [K_i(*Drosophila*)] nicotinic receptors



Compd[a]	X	Y	Z	R_1	R_2	R_3	$K_i(\alpha_4\beta_2)$ (nM)[b]	p$K_i(\alpha_4\beta_2)$	K_i(*Drosophila*) (nM)[b]	pK_i(*Drosophila*)
1	NH	NH	CH_2CH_2	H	H	Cl	2.2	−0.342	800	−2.903
2	NH	NH	CH_2CH_2	H	N_3	Cl	2.6	−0.415	5700	−3.756
3	NH	S	CH=CH	H	N_3	H	8.3	−0.919	7200	−3.857
4	NH	S	CH=CH	N_3	H	H	820	−2.914	30,000	−4.477
5	NH	S	CH=CH	H	H	Cl	0.39	0.409	100	−2.000
6	NH	S	CH=CH	H	N_3	Cl	0.47	0.328	750	−2.875
7	NH	S	CH=CH	N_3	H	Cl	110	−2.041	3200	−3.505
8	NNO_2	NH	CH_2CH_2	H	H	Cl	720	−2.857	2.2	−0.342
9	NNO_2	NH	CH_2CH_2	H	N_3	Cl	340	−2.531	24	−1.380
10	NNO_2	NH	CH=CH	H	N_3	H	5400	−3.732	540	−2.732
11	NNO_2	NH	CH=CH	N_3	H	H	95,800	−4.981	3600	−3.556
12	NNO_2	NH	CH=CH	H	H	Cl	470	−2.672	0.85	0.071
13	NNO_2	NH	CH=CH	H	N_3	Cl	300	−2.477	13	−1.114
14	NNO_2	NH	CH=CH	N_3	H	Cl	16,900	−4.228	460	−2.663
15	NNO_2	S	CH=CH	H	H	Cl	170	−2.230	0.35	0.456
16	NNO_2	S	CH=CH	H	N_3	Cl	170	−2.230	3.9	−0.591
17	$CHNO_2$	NH	CH_2CH_2	H	H	Cl	60	−1.778	0.12	0.921
18	$CHNO_2$	NH	CH_2CH_2	H	N_3	Cl	40	−1.602	0.72	0.143
19	NCN	S	CH=CH	H	H	Cl	120	−2.079	1.4	−0.146
20	NCN	S	CH=CH	H	N_3	Cl	120	−2.079	28	−1.447

[a] Compound number.
[b] Taken from Ref. 8.

Table 2. R-state indices of different common atoms of azidopyridinyl neonicotinoids and indicator parameters

Compd[a]	$\Re_1$	$\Re_2$	$\Re_3$	$\Re_4$	$\Re_5$	$\Re_6$	$\Re_7$	$\Re_8$	$\Re_9$	$\Re_{10}$	$\Re_{11}$	$\Re_{12}$	$\Re_{13}$	$\Re_{av}$	I_1[b]	I_2[c]	I_3[d]
1	1.201	5.220	2.588	4.909	4.893	2.657	5.719	1.419	5.498	5.322	3.898	1.914	4.438	3.852	1	0	0
2	1.234	5.377	2.717	5.380	3.053	2.933	5.968	1.438	5.589	5.392	3.940	1.929	4.487	3.959	1	0	0
3	1.248	5.248	2.599	5.252	2.984	5.373	5.729	1.053	5.392	4.713	10.990	1.080	4.102	3.712	0	0	0
4	1.225	5.341	2.745	2.935	5.193	5.151	5.831	1.067	5.426	4.735	11.039	1.088	4.117	3.767	0	0	0
5	1.127	5.119	2.442	4.808	4.819	2.600	5.563	1.007	5.302	4.644	10.854	1.044	4.048	3.601	0	0	0
6	1.160	5.276	2.572	5.280	2.980	2.876	5.739	1.026	5.398	4.717	11.025	1.059	4.098	3.698	0	0	0
7	1.138	5.368	2.718	2.905	5.290	2.730	5.840	1.039	5.431	4.739	11.074	1.067	4.113	3.753	0	0	0
8	1.224	5.336	2.680	5.025	4.983	2.699	6.033	1.560	5.740	5.564	4.1690	2.215	2.927	4.020	0	1	0
9	1.257	5.492	2.810	5.496	3.106	2.974	6.210	1.578	5.831	5.634	4.211	2.230	2.942	4.116	0	0	0
10	1.339	5.458	2.827	5.462	3.106	5.498	6.183	1.571	6.021	5.813	4.183	2.229	2.945	4.173	0	0	0
11	1.317	5.550	2.973	3.098	5.352	5.275	6.285	1.584	6.054	5.835	4.196	2.237	2.951	4.228	0	0	0
12	1.219	5.329	2.670	5.018	4.978	2.695	6.017	1.525	5.931	5.743	4.134	2.193	2.915	4.062	0	1	0
13	1.252	5.486	2.800	5.489	3.101	2.971	6.193	1.543	6.026	5.817	4.176	2.208	2.930	4.160	0	0	0
14	1.230	5.578	2.946	3.068	5.449	2.825	6.295	1.557	6.059	5.839	4.189	2.216	2.936	4.215	0	0	0
15	1.150	5.235	2.535	4.924	4.909	2.642	5.805	1.148	5.554	4.895	11.608	1.345	2.538	3.760	0	1	0
16	1.183	5.391	2.664	5.395	3.032	2.918	5.981	1.166	5.649	4.968	11.780	1.360	2.553	3.857	0	0	0
17	1.172	5.263	2.572	4.952	4.931	2.659	5.855	1.204	5.561	5.386	3.812	3.716	6.284	3.859	0	0	1
18	1.205	5.420	2.702	5.423	3.054	2.935	6.031	1.222	5.652	5.455	3.855	3.767	6.358	3.955	0	0	1
19	1.123	5.175	2.474	4.864	4.861	2.613	5.690	1.046	5.436	4.777	11.334	1.176	2.204	3.674	0	1	0
20	1.156	5.331	2.604	5.335	2.996	2.889	5.866	1.064	5.531	4.851	11.506	1.191	2.219	3.772	0	0	0

[a] Compound number.
[b] Indicator variable for the presence of unsubstituted imidazolidine-2-imine ring system.
[c] Indicator variable for substituted imines with unsubstituted R_1 and R_2 positions.
[d] Indicator variable for the presence of $CHNO_2$ at X-position.

analysis was performed using mammalian $\alpha_4\beta_2$ and *Drosophila* nAChR agonistic activity data, R-state indices, and some indicator parameters. The result of this analysis is shown in Table 3.

The different possible combinations of R-state indices and indicator parameters were used as the independent variables for the development of QSAR models on the basis of the autocorrelation study. Combination of $\Re_4$, $\Re_7$, and $\Re_{13}$ (R-state indices of atoms 4, 7, and 13, respectively) yielded a model for mammalian $\alpha_4\beta_2$ agonistic activity as shown in Eq. 1.

$$\begin{aligned} \mathrm{p}K_i(\alpha_4\beta_2) &= 20.215(\pm 4.782) + 0.564(\pm 0.168)\Re_4 \\ &\quad - 4.395(\pm 0.754)\Re_7 + 0.319(\pm 0.134)\Re_{13} \\ n &= 20;\ \ R = 0.898;\ \ \%\mathrm{EV} = 80.59; \\ R^2_A &= 0.770;\ \ F_{(3,16)} = 22.148;\ \ p < 0.0000; \\ s &= 0.672;\ \ \mathrm{PRESS} = 11.787;\ \ \mathrm{SSY} = 37.268; \\ R^2_{CV} &= 0.684;\ \ S_{DEP} = 0.768 \end{aligned} \quad (1)$$

where n is the number of data points, R is correlation coefficient, %EV, R^2_A, F, p, s, PRESS, SSY, R^2_{CV}, S_{DEP} are percentage of explained variance, adjusted R^2, ratio between the variances of observed and calculated activities, probability factor related to F-ratio, standard error of estimate, predicted residual sum of squares, total sum of squares, squared cross-validated correlation coefficient, and standard deviation error of prediction, respectively. Values within the parenthesis are standard error of corresponding parameters. Eq. 1 explains up to 80.59% of the variances in the activity data and suggests the importance of atoms 4, 7, and 13 in biological activity. Positive coefficients of $\Re_4$ and $\Re_{13}$ (R-state indices of atoms 4 and 13, respectively) imply that higher values of these indices correspond to higher mammalian receptor agonistic activity, that is, higher toxicity to mammals. Negative coefficient of $\Re_7$ signifies that higher value of this index is detrimental to the activity and thus, safe for mammals.

From the correlation study it was found that atoms 1, 2, 3, 7, 8, 9, and 10 were important for the mammalian receptor agonistic activity but these were autocorrelated. Thus, a composite R-state index, $\Re_{av}$ (average of $\Re_1$–$\Re_3$ and $\Re_7$–$\Re_{10}$) was considered for the QSAR study. When this composite index, $\Re_{av}$ along with $\Re_5$, $\Re_6$, and $\Re_{13}$ (R-state indices of atoms 5, 6, and 13, respectively) were modeled for mammalian nAChR agonistic activity the following QSAR equation was developed as:

$$\begin{aligned} \mathrm{p}K_i(\alpha_4\beta_2) &= 16.650(\pm 3.554) - 0.475(\pm 0.160)\Re_5 \\ &\quad - 0.392(\pm 0.162)\Re_6 + 0.352(\pm 0.142)\Re_{13} \\ &\quad - 4.278(\pm 0.873)\Re_{av} \\ n &= 20;\ \ R = 0.890;\ \ \%\mathrm{EV} = 79.16; \\ R^2_A &= 0.736;\ \ F_{(4,15)} = 14.245;\ \ p < 0.0000; \\ s &= 0.720;\ \ \mathrm{PRESS} = 13.868;\ \ \mathrm{SSY} = 37.268; \\ R^2_{CV} &= 0.628;\ \ S_{DEP} = 0.833 \end{aligned} \quad (2)$$

Eq. 2 explains up to 79.16% of the variances in the activity data and suggests the importance of atoms 1, 2, 3, 5, 6, 7, 8, 9, 10, and 13 to the mammalian nAChR binding activity.

Another QSAR model was developed with the combination of $\Re_{av}$, $\Re_5$, $\Re_{13}$, and an indicator parameter as shown below:

Table 3. Correlation matrix for the biological activity parameters, R-state indices and indicator parameters

	$\Re_1$	$\Re_2$	$\Re_3$	$\Re_4$	$\Re_5$	$\Re_6$	$\Re_7$	$\Re_8$	$\Re_9$	$\Re_{10}$	$\Re_{11}$	$\Re_{12}$	$\Re_{13}$	$\Re_{av}$	I_1	I_2	I_3	p$K_i(\alpha_4\beta_2)$	pK_i(*Drosophila*)
$\Re_1$	1.00																		
$\Re_2$		0.70	0.78	−0.02	−0.20	0.70	0.75	0.76	0.74	0.73	−0.62	0.35	−0.10	0.80	0.06	−0.25	−0.11	−0.57	−0.35
$\Re_3$		1.00	0.96	−0.26	−0.14	0.29	0.95	0.68	0.83	0.73	−0.53	0.39	−0.17	0.87	−0.14	−0.34	−0.02	−0.72	−0.19
$\Re_4$			1.00	−0.42	0.03	0.43	0.91	0.71	0.81	0.72	−0.55	0.33	−0.15	0.87	−0.07	−0.33	−0.11	−0.76	−0.36
$\Re_5$				1.00	−0.74	−0.25	−0.12	0.04	−0.06	0.05	−0.11	0.18	0.01	−0.07	0.15	0.11	0.16	0.46	0.47
$\Re_6$					1.00	−0.11	−0.09	0.04	0.01	0.03	−0.04	−0.02	−0.03	−0.02	−0.06	0.38	−0.05	−0.32	0.02
$\Re_7$						1.00	0.23	0.10	0.20	0.08	0.01	−0.10	−0.04	0.22	−0.16	−0.31	−0.16	−0.40	−0.59
$\Re_8$							1.00	0.83	0.94	0.87	−0.66	0.51	−0.24	0.96	−0.16	−0.13	0.00	−0.78	0.01
$\Re_9$								1.00	0.88	0.95	−0.87	0.52	−0.21	0.94	0.20	0.06	−0.11	−0.56	0.03
$\Re_{10}$									1.00	0.92	−0.70	0.49	−0.34	0.97	−0.15	0.02	−0.07	−0.77	0.11
$\Re_{11}$										1.00	−0.92	0.70	−0.06	0.96	0.09	0.00	0.13	−0.60	0.16
$\Re_{12}$											1.00	−0.81	−0.26	−0.81	−0.31	0.07	−0.32	0.32	−0.12
$\Re_{13}$												1.00	0.53	0.57	0.02	−0.08	0.79	−0.26	0.46
$\Re_{av}$													1.00	−0.19	0.25	−0.42	0.78	0.44	−0.03
I_1														1.00	−0.01	−0.08	−0.00	−0.72	0.01
I_2															1.00	−0.17	−0.11	0.41	−0.32
I_3																1.00	−0.17	−0.14	0.56
p$K_i(\alpha_4\beta_2)$																	1.00	0.09	0.48
pK_i(*Drosophila*)																		1.00	−0.03
																			1.00

$$\begin{aligned} p K_i(\alpha_4\beta_2) &= 12.059(\pm 3.072) + 0.552(\pm 0.152)\Re_4 \\ &\quad + 0.277(\pm 0.123)\Re_{13} - 4.574(\pm 0.719)\Re_{av} \\ &\quad + 1.353(\pm 0.471) I_1 \\ n &= 20;\ \ R = 0.924;\ \ \%EV = 85.30; \\ R^2_A &= 0.814;\ \ F_{(4,15)} = 21.760;\ \ p < 0.0000; \\ s &= 0.604;\ \ PRESS = 9.124;\ \ SSY = 37.268; \\ R^2_{CV} &= 0.755;\ \ S_{DEP} = 0.675 \end{aligned} \tag{3}$$

Eq. 3 explains up to 85.30% of the variances in the activity and shows the importance of atoms 1, 2, 3, 4, 6, 7, 8, 9, 10, and 13. Except atoms 4 and 13, all other atoms have negative contributions to the agonistic activity. I_1 was an indicator variable for the presence of unsubstituted imidazolidine-2-imine ring system. Positive coefficient of I_1 suggests that unsubstituted imidazolidine-2-imine ring system is conducive to the mammalian receptor agonistic activity.

The best QSAR model with mammalian receptor agonistic activity was developed, using the combination of $\Re_5$, $\Re_6$, $\Re_7$, and $\Re_{13}$ (R-state indices of atoms 5, 6, 7, and 13, respectively), as suggested by the statistical parameters. It can be presented as:

$$\begin{aligned} p K_i(\alpha_4\beta_2) &= 27.489(\pm 3.783) - 0.547(\pm 0.119)\Re_5 \\ &\quad - 0.364(\pm 0.121)\Re_6 - 4.566(\pm 0.615)\Re_7 \\ &\quad + 0.290(\pm 0.107)\Re_{13} \\ n &= 20;\ \ R = 0.940;\ \ \%EV = 88.40; \\ R^2_A &= 0.853;\ \ F_{(4,15)} = 28.586;\ \ p < 0.0000; \\ s &= 0.537;\ \ PRESS = 8.113;\ \ SSY = 37.268; \\ R^2_{CV} &= 0.782;\ \ S_{DEP} = 0.637 \end{aligned} \tag{4}$$

Eq. 4 explains up to 88.40% of the variances in the activity data. The observed, calculated, residual, LOO-predicted, and predicted residual activities of Eqs. 1–4 are shown in Table 4.

When the *Drosophila* nAChR agonistic activity was modeled, using different R-state indices and indicator parameters, another QSAR equation was developed in combination with $\Re_3$, $\Re_6$, $\Re_{12}$, and indicator parameter I_1. The equation can be represented as follows:

$$\begin{aligned} p K_i(\textit{Drosophila}) &= 10.667(\pm 4.117) - 4.493(\pm 1.714)\Re_3 \\ &\quad - 0.687(\pm 0.222)\Re_6 + 1.123(\pm 0.271)\Re_{12} \\ &\quad - 2.315(\pm 0.655) I_1 \\ n &= 20;\ \ R = 0.885;\ \ \%EV = 78.34; \\ R^2_A &= 0.726;\ \ F_{(4,15)} = 13.563; \\ p &< 0.0000;\ \ s = 0.867; \\ PRESS &= 17.668;\ \ SSY = 52.101; \\ R^2_{CV} &= 0.661;\ \ S_{DEP} = 0.940 \end{aligned} \tag{5}$$

Eq. 5 explains up to 78.34% of the variances in the activity data. It suggests the importance of atoms 3, 6, and 12

Table 4. Observed, calculated, residual, LOO-predicted, predicted residual activities of Eqs. 1–4

Compd[a]	Obsd[b]	Eq. 1				Eq. 2				Eq. 3				Eq. 4			
		Calcd[c]	Res[d]	Pred[e]	Pres[f]	Calcd	Res	Pred	Pres	Calcd	Res	Pred	Pres	Calcd	Res	Pred	Pres
1	−0.342	−0.735	0.393	−0.788	0.446	−1.631	1.289	−1.801	0.732	−0.271	−0.071	−0.197	−0.145	−0.982	0.640	−1.092	0.750
2	−0.415	−1.548	1.133	−1.690	1.275	−1.305	0.890	−1.470	−0.595	−0.486	0.071	−0.560	0.145	−1.200	0.785	−1.338	0.923
3	−0.919	−0.693	−0.226	−0.664	−0.255	−1.305	0.386	−1.581	0.095	−0.884	−0.035	−0.881	−0.038	−1.071	0.152	−1.185	0.266
4	−2.914	−2.443	−0.471	−2.252	−0.662	−2.498	−0.416	−2.261	0.038	−2.411	−0.503	−2.199	−0.715	−2.659	−0.255	−2.530	−0.384
5	0.409	−0.231	0.640	−0.410	0.819	−0.633	1.042	−0.902	0.096	−0.635	1.044	−0.878	1.287	−0.322	0.731	−0.542	0.951
6	0.328	−0.722	1.050	−0.853	1.181	−0.269	0.597	−0.404	0.041	−0.808	1.136	−0.989	1.317	−0.206	0.534	−0.325	0.653
7	−2.041	−2.500	0.460	−2.690	0.649	−1.538	−0.503	−1.446	0.525	−2.366	0.325	−2.512	0.471	−1.872	−0.169	−1.846	−0.195
8	−2.857	−2.532	−0.325	−2.504	−0.353	−2.937	0.080	−2.952	0.350	−2.743	−0.114	−2.734	−0.122	−2.918	0.061	−2.927	0.070
9	−2.531	−3.040	0.509	−3.156	0.625	−2.562	0.031	−2.568	−0.263	−2.920	0.389	−2.995	0.464	−2.796	0.265	−2.865	0.334
10	−3.732	−2.939	−0.793	−2.781	−0.951	−3.794	0.062	−3.828	−0.390	−3.200	−0.532	−3.068	−0.664	−3.591	−0.141	−3.518	−0.214
11	−4.981	−4.718	−0.263	−4.592	−0.389	−5.006	0.025	−5.022	−0.846	−4.754	−0.227	−4.636	−0.345	−5.202	0.221	−5.366	0.385
12	−2.672	−2.469	−0.203	−2.453	−0.219	−3.119	0.447	−3.197	−1.188	−2.944	0.272	−2.975	0.303	−2.845	0.173	−2.869	0.197
13	−2.477	−2.973	0.496	−3.078	0.601	−2.749	0.272	−2.827	−1.682	−3.127	0.650	−3.281	0.804	−2.718	0.241	−2.778	0.301
14	−4.228	−4.784	0.556	−5.067	0.839	−4.040	−0.188	−3.965	−0.616	−4.714	0.486	−4.958	0.731	−4.413	0.185	−4.503	0.275
15	−2.230	−1.711	−0.519	−1.631	−0.599	−1.908	−0.322	−1.840	−1.089	−1.721	−0.509	−1.636	−0.594	−1.929	−0.301	−1.870	−0.360
16	−2.230	−2.214	−0.016	−2.212	−0.018	−1.535	−0.695	−1.384	0.732	−1.901	−0.329	−1.851	−0.379	−1.803	−0.427	−1.715	−0.515
17	−1.778	−0.719	−1.059	−0.216	−1.562	−1.027	−0.751	−0.590	−0.595	−1.117	−0.661	−0.728	−1.050	−1.090	−0.688	−0.686	−1.092
18	−1.602	−1.203	−0.399	−0.911	−0.691	−0.630	−0.972	0.080	0.095	−1.278	−0.324	−1.010	−0.592	−0.946	−0.656	−0.420	−1.182
19	−2.079	−1.346	−0.733	−1.109	−0.970	−1.624	−0.455	−1.463	0.038	−1.454	−0.625	−1.262	−0.817	−1.464	−0.615	−1.245	−0.834
20	−2.079	−1.849	−0.230	−1.805	−0.274	−1.259	−0.820	−0.990	0.096	−1.636	−0.443	−1.535	−0.544	−1.344	−0.735	−1.122	−0.957

[a] Compound number.
[b] Observed activity.
[c] Calculated activity.
[d] Residual activity.
[e] LOO-predicted activity.
[f] Predicted residual activity.

for insecticidal activity. $\Re_3$ and $\Re_6$ possess positive contributions while $\Re_{12}$ possesses negative contribution to the insect nAChR agonistic activity. Negative coefficient of I_1 implies that the presence of unsubstituted imidazolidine-2-imine ring system in those compounds is detrimental to the activity.

Combination of $\Re_1$ and $\Re_{12}$ as well as an indicator parameter I_2 for substituted imines with unsubstituted R_1 and R_2 positions yielded Eq. 6 as shown below:

$$\begin{aligned} \mathrm{p}K_i(\textit{Drosophila}) &= 10.762(\pm 4.686) - 12.788(\pm 3.979)\Re_1 \\ &\quad + 1.340(\pm 0.280)\Re_{12} + 2.003(\pm 0.536)I_2 \\ n &= 20;\ R = 0.858;\ \%\mathrm{EV} = 73.56;\ R^2_{\mathrm{A}} = 0.686; \\ F_{(3,16)} &= 14.842;\ p < 0.0000;\ s = 0.928; \\ \mathrm{PRESS} &= 18.175;\ \mathrm{SSY} = 52.101; \\ R^2_{\mathrm{CV}} &= 0.651;\ S_{\mathrm{DEP}} = 0.953 \end{aligned} \quad (6)$$

Eq. 6 shows the importance of atoms 1 and 12 as pharmacophore for *Drosophila* nAChR agonistic activity. Positive coefficient of I_2 implies that substituted imines are advantageous and presence of azido group at R_1 or R_2 position is detrimental to the activity. The Eq. 6 explains up to 73.56% of the variances in the insecticidal activity.

The best equation for the insect nAChR agonistic activity was modeled in combination with $\Re_4$, $\Re_6$, I_2, and another indicator variable I_3 for the presence of $CHNO_2$ at X-position. The equation of this model is shown as:

$$\begin{aligned} \mathrm{p}K_i(\textit{Drosophila}) &= -3.271(\pm 1.321) + 0.479(\pm 0.211)\Re_4 \\ &\quad - 0.445(\pm 0.196)\Re_6 + 2.091(\pm 0.493)I_2 \\ &\quad + 2.563(\pm 0.638)I_3 \\ n &= 20;\ R = 0.900;\ \%\mathrm{EV} = 80.93;\ R^2_{\mathrm{A}} = 0.758; \\ F_{(4,15)} &= 15.914;\ p < 0.0000;\ s = 0.814; \\ \mathrm{PRESS} &= 15.443;\ \mathrm{SSY} = 52.101; \\ R^2_{\mathrm{CV}} &= 0.704;\ S_{\mathrm{DEP}} = 0.879 \end{aligned} \quad (7)$$

Eq. 7 explains up to 80.93% of the variances in the activity data. It suggests the importance of atoms 4 and 6. The Eq. 7 also shows that presence of nitromethylene group at X-position is conducive to the activity in addition to the positive contribution of I_2.

These Eqs. 5–7 show the negative contributions of atoms 1, 3, 6 and positive contributions of atoms 4, 12 toward the insecticidal activity of the azidopyridinyl neonicotinoids. The observed, calculated, residual, LOO-predicted, and predicted residual activities of Eqs. 5–7 are shown in Table 5. All the coefficients of parameters and intercepts in all equations are of 95% confidence intervals as supported by their *t*- and *p*-values. These are shown in Table 6.

3. Conclusion

The E-state/R-state indices, contributing significantly to the QSAR models, carry electronic/steric and topological information arising from the change in substitution pattern. These help to know the structural changes of

Table 5. Observed, calculated, residual, LOO-predicted, predicted residual activities of Eqs. 5–7

Compd[a]	Obsd[b]	Eq. 5				Eq. 6				Eq. 7			
		Calcd[c]	Res[d]	Pred[e]	Pres[f]	Calcd	Res	Pred	Pres	Calcd	Res	Pred	Pres
1	−2.903	−2.953	0.050	−3.006	0.103	−2.031	−0.872	−1.968	−0.935	−2.102	−0.801	−1.997	−0.906
2	−3.756	−3.706	−0.050	−3.653	−0.103	−2.432	−1.324	−2.335	−1.421	−2.000	−1.756	−1.760	−1.996
3	−3.857	−3.491	−0.366	−3.255	−0.602	−3.749	−0.108	−3.728	−0.129	−3.146	−0.711	−2.815	−1.042
4	−4.477	−3.985	−0.492	−3.824	−0.653	−3.445	−1.032	−3.29	−1.187	−4.157	−0.320	−3.987	−0.490
5	−2.000	−0.920	−1.080	−0.610	−1.390	−2.25	0.25	−2.315	0.315	−2.125	0.125	−2.143	0.143
6	−2.875	−1.677	−1.198	−1.504	−1.371	−2.652	−0.223	−2.616	−0.259	−2.022	−0.853	−1.911	−0.964
7	−3.505	−2.223	−1.282	−1.921	−1.584	−2.36	−1.145	−2.113	−1.392	−3.095	−0.410	−2.880	−0.625
8	−0.342	−0.743	0.401	−0.780	0.438	0.081	−0.423	0.254	−0.596	0.026	−0.368	0.148	−0.490
9	−1.380	−1.499	0.119	−1.517	0.137	−2.323	0.943	−2.419	1.039	−1.962	0.582	−2.048	0.668
10	−2.732	−3.311	0.579	−3.567	0.835	−3.373	0.641	−3.671	0.939	−3.101	0.369	−3.325	0.593
11	−3.556	−3.805	0.249	−3.931	0.375	−3.081	−0.475	−2.937	−0.619	−4.134	0.578	−4.428	0.872
12	0.071	−0.720	0.791	−0.791	0.862	0.116	−0.045	0.134	−0.062	0.024	0.047	0.008	0.063
13	−1.114	−1.477	0.363	−1.528	0.414	−2.289	1.175	−2.399	1.285	−1.964	0.850	−2.089	0.975
14	−2.663	−2.023	−0.640	−1.658	−1.005	−1.997	−0.666	−1.945	−0.718	−3.059	0.396	−3.222	0.559
15	0.456	−1.029	1.485	−1.243	1.699	−0.138	0.594	−0.357	0.813	0.003	0.453	−0.149	0.605
16	−0.591	−1.781	1.190	−1.904	1.313	−2.543	1.952	−2.742	2.151	−1.986	1.395	−2.180	1.589
17	0.921	1.455	−0.534	1.958	−1.037	0.756	0.165	0.61	0.311	0.481	0.440	0.030	0.891
18	0.143	0.7389	−0.596	1.095	−0.952	0.402	−0.259	0.575	−0.431	0.583	−0.440	1.034	−0.891
19	−0.146	−0.924	0.778	−1.097	0.951	−0.02	−0.126	0.036	−0.182	−0.013	−0.133	0.031	−0.177
20	−1.447	−1.681	0.234	−1.709	0.262	−2.424	0.977	−2.577	1.13	−2.002	0.555	−2.076	0.629

[a] Compound number.
[b] Observed activity.
[c] Calculated activity.
[d] Residual activity.
[e] LOO-predicted activity.
[f] Predicted residual activity.

Table 6. *t*-Values and *p*-values of QSAR equations

Eq.	Intercepts/parameters	*t*-Value	*p*-Value
1	Intercept	4.228	0.001
	$\Re_4$	3.357	0.004
	$\Re_7$	−5.832	0.000
	$\Re_{13}$	2.382	0.030
2	Intercept	4.685	0.000
	$\Re_5$	−2.974	0.009
	$\Re_6$	−2.419	0.029
	$\Re_{13}$	2.484	0.025
	$\Re_{av}$	−4.902	0.000
3	Intercept	3.925	0.001
	$\Re_4$	3.631	0.002
	$\Re_{13}$	2.254	0.040
	$\Re_{av}$	−6.361	0.000
	I_1	2.874	0.012
4	Intercept	7.266	0.000
	$\Re_5$	−4.576	0.000
	$\Re_6$	−3.009	0.009
	$\Re_7$	−7.425	0.000
	$\Re_{13}$	2.704	0.016
5	Intercept	2.591	0.020
	$\Re_3$	−2.622	0.019
	$\Re_6$	−3.101	0.007
	$\Re_{12}$	4.150	0.001
	I_1	−3.533	0.003
6	Intercept	2.297	0.035
	$\Re_1$	−3.214	0.005
	$\Re_{12}$	4.793	0.000
	I_2	3.736	0.002
7	Intercept	−2.476	0.026
	$\Re_4$	2.273	0.038
	$\Re_6$	−2.270	0.038
	I_2	4.241	0.001
	I_3	4.017	0.001

particular position governing a specific type of biological activity. Thus, it can be said that atoms associated with those indices useful in QSAR may form pharmacophore for electronic/steric interaction at the receptor site. This QSAR study shows the different structural fragments required for both activities. Except atoms 11 and 12 all other atoms of the general structure of these azidopyridinyl neonicotinoid insecticides possess significant contribution to the mammalian $\alpha_4\beta_2$ receptor agonistic activity. Atoms 1, 3, 4, 6, and 12 possess important contributions to the *Drosophila* nAChR agonistic activity. In view of the selectivity of these neonicotinoids, atom number 12 is of greatest importance as it is the only uncommon pharmacophoric atom of insect nAChR agonistic activity over mammalian receptor agonistic activity. Substituted imines with unsubstituted R_1 and R_2 positions have better insecticidal activity with minimum effect to the mammalian receptor. Nitromethylene derivatives are also selective to the insecticidal activity. Unsubstituted imidazolidine-2-imine ring system is not favorable for the *Drosophila* nAChR agonistic activity and is also toxic for the mammalian system.

Earlier work of us,[26] of the same data set, using the electrotopological state atom (E-state) indices showed that atoms 1, 2, 3, 4, 6, 7, 8, 12, and 13 are of great importance as they are associated with the electronic interactions of neonicotinoids with the mammalian receptor. The present study shows the importance of atoms 5, 9, 10 along with 1, 2, 3, 4, 6, 7, 8, 12, and 13. These atoms are associated with dispersive/van der Waals interactions of these insecticides with the mammalian receptor. Thus, common atoms of these two studies, that is, 1, 2, 3, 4, 6, 7, 8, 12, 13 are of great importance as these atoms are associated with both electronic and dispersive/van der Waals interactions. In case of *Drosophila* nAChR receptor agonistic activity previous study[26] showed that atoms 3, 4, 6, and 13 have major contributions for electronic interaction with the receptor-binding site. The present QSAR study shows the contributions of atoms 1, 3, 4, 6, and 12 as pharmacophore related to the dispersive/van der Waals interactions with the insect nAChR. However, atoms 4, 6, and 12 may not be exposed to the surface area very much, but may be sufficient enough to have some interaction with the receptor through the dispersive force. The common atoms of these two studies, that is, atoms 3, 4, and 6 are most important due to electronic as well as dispersive/van der Waals interactions with the *Drosophila* nAChR. Electron density and steric environments of these atoms are such as that these may be easily bind with the nicotinic acetylcholine receptor surface through electronic, van der Waals as well as steric interaction. The important atoms associated with dispersive/van der Waals as well as electronic interactions for the mammalian and insect nAChRs agonistic activities are presented in Figure 2.

The above QSAR study throws some light on the important atoms/atomic fragments for the selective insecticidal activity of azidopyridinyl neonicotinoids over mammalian toxicity. However, considering less number

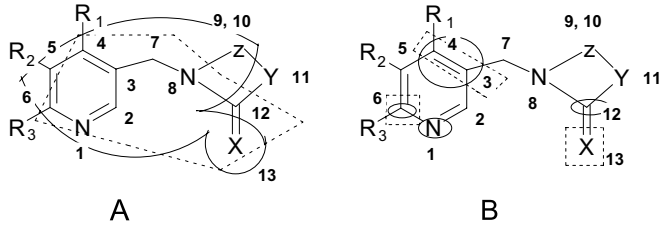


Figure 2. Atoms enclosed by discontinuous line represent the pharmacophore related to electronic interactions and atoms bounded by continuous line represent the pharmacophore related to dispersive/van der Waals interactions, required for the mammalian nAChR agonistic activity (A) and *Drosophila* nAChR agonistic activity (B).

of data points lack of diversity in the substitution pattern, more compounds need to be incorporated in the data set for further extensive QSAR analysis to derive robust QSAR models for such compounds.

4. Experimental

4.1. Dataset and parameters

4.1.1. Biological activity. Mammalian and insect nAChR agonistic activity data ($\alpha_4\beta_2$ subtype and *Drosophilla*, respectively) of some azidopyridinyl neonicotinoids as reported by Zhang et al.[13] were considered as biological activity parameters. The general structure of these neonicotinoids is shown in Figure 1 and their activity data are listed in Table 1.

4.1.2. R-state index. Molar refractivity is an important physicochemical property for modeling the dispersive/van der Waals interactions due to its high correlation with lipophilicity, molar volume, and steric bulk. It is also an important property of a ligand for its steric fit with the receptor cavity and hence it is useful in modeling repulsive nonbonded interactions. Ghosh and coworkers[11,12] reported the atomic contributions of molar refractivity. The calculation of molar refractivity is based on the idea that it is additive in nature. The sum of the atomic refractivity (a_i) is the molar refractivity, which can be represented as:

$$\mathrm{MR_{Calcd}} = \sum n_i a_i \tag{8}$$

where $\mathrm{MR_{Calcd}}$ = calculated molar refractivity; n_i = number (n) of a particular atom type (i) in the molecule.

In view of the importance of the molar refractivity in QSAR studies, Carrasco et al.[10] introduced the refracto-topological state atom (R-state) index in the same way of E-state index as defined by Kier and co-workers.[7–9] The nonempirical R-state indices encode the atomic refractivity and the influences of other atoms in the chemical graph as well as its topological state. The R-state index correlates not only with the molecular size but also with the lipophilicity of the molecule. The same initial consideration of E-state formalism was taken for R-state index. The idea was that every atom in a molecule is different from the other atoms in that molecule except where atoms mapped on to each other through a symmetry operation. This is due to the difference in refractive and topological environment. Atom-wise R-state indices were calculated using the computer program 'mouse'.[27] In the program, molecular connection table in specified format and the intrinsic values of different atoms are given as inputs and output result is the R-state indices of atoms. Before the calculation, atoms of the molecules were numbered arbitrarily keeping the serial number of atoms same in all molecules. The R-state index ($\Re_i$) of an atom (i) in a molecule is composed of an intrinsic refractivity (AR_i) and the perturbation effect ($\Delta\mathrm{AR}_i$). The intrinsic refractivities for atoms are shown in Table 7. These refractivities of the bonded hydrogens have been included to the nonhydrogen atoms connected to it as was used to consider their contribution to the molar refractivity.

Table 7. Atomic refractivity values (AR) of common atoms

Atom type	Atomic refractivity
–F	1.0632
–Cl	5.6105
–Br	8.6782
–I	13.8741
–O–	1.6351
=O	1.7956
O^*=N	2.1407
C_{sp^3}	2.8158
C_{sp^2}	3.8278
C_{sp}	3.8974
C (Ar^b)	3.5090
>C^*=X^a	3.0887
H	0.9155
N_{sp^3}	3.0100
N_{sp^2}, N_{sp}	3.2009
N^*O_2	3.5054
Ar–N^*=X^a	3.8095
N (Ar^b)	2.7662
S_{sp^3}	7.3190
S_{sp^2}	9.1680
R–S^*O_2–R	5.3321
R–S^*O–R	6.0762

*Atomic refractivity of the corresponding atom.
[a] Hetero atom.
[b] Aromatic.

The general expression for the perturbation effect is as follows:

$$\Delta\mathrm{AR}_i = \sum(\mathrm{AR}_i - \mathrm{AR}_j)/r_{ij}^2 \tag{9}$$

where r_{ij} is the topological distance in the shortest path between ith and jth atoms. Thus, the R-state index is calculated as

$$\Re_i = \mathrm{AR}_i + \Delta\mathrm{AR}_i \tag{10}$$

4.1.3. Indicator parameter. Indicator parameters were used along with the R-state indices in order to find out the role of specific substituents/substituent patterns at specific positions toward the biological activity.

4.2. Statistical analysis

All the statistical analyses were carried out by a computer program 'Multi Regress'[28] developed in our laboratory.

4.2.1. Correlation analysis. Correlation analysis[29] of R-state indices, indicator parameters, and biological activities was performed. The auto-correlated parameters were eliminated depending on their individual correlation with the biological activity. All possible combinations of parameters were considered for the QSAR study.

4.2.2. Multiple linear regression analysis. Multiple regression analysis[30] was carried out using nAChR agonistic activities as the dependent variables and

R-state indices as well as indicator parameters as independent variables in all possible combinations. The statistical quality of the regression equations were justified by parameters like correlation coefficient (R), percentage of explained variance (%EV), adjusted R^2 (R^2_A), variance ratio (F), probability factor related to F-ratio (p), standard error of estimate (s). All the final equations have regression coefficients, intercepts, and variance ratio (F) significant to more than 95% level. Use of more than one variable in the multivariate equation was justified by autocorrelation study.

4.2.3. Validation of the QSAR model. The predictive powers of the equations were validated by Leave-One-Out (LOO-) cross-validation method.[31,32] Predicted residual sum of square (PRESS), total sum of squares (SSY), cross-validated R^2 (R^2_{CV}), and standard deviation error of prediction (S_{DEP}) were considered for the validation of these models.

Acknowledgments

Authors are thankful to University Grants Commission (U.G.C.), New Delhi and All India Council for Technical Education (A.I.C.T.E.), New Delhi for awarding research projects.

References and notes

1. Duart, M. J.; Anton-Fos, G. M.; de Julian-Ortiz, J. V.; Gozalbes, R.; Galvez, J.; Garcia-Domenech, R. *Int. J. Pharm.* **2002**, *246*, 111.
2. Golbraikh, A.; Bonchev, D.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 769.
3. Golbraikh, A.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 144.
4. Hall, L. M.; Hall, L. H.; Kier, L. B. *J. Comput. Aided Mol. Des.* **2003**, *17*, 103.
5. Kier, L. B.; Hall, L. H. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 792.
6. Kier, L. B.; Hall, L. H. *J. Mol. Graph. Mod.* **2001**, *20*, 76.
7. Rose, K.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651.
8. Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
9. de Gregorio, C.; Kier, L. B.; Hall, L. H. *J. Comput. Aided Mol. Des.* **1998**, *12*, 557.
10. Carrasco, R.; Padron, J. A.; Galvez, J. *J. Pharm. Pharmaceut. Sci.* **2004**, *7*, 19.
11. Ghose, A. K.; Pritchett, A.; Crippen, G. M. *J. Comput. Chem.* **1988**, *9*, 80.
12. Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21.
13. Zhang, N.; Tomizawa, M.; Casida, J. E. *J. Med. Chem.* **2002**, *45*, 2832.
14. Srikanth, K.; Kumar, C. A.; Goswami, D.; De, A. U.; Jha, T. *Ind. J. Biochem. Biophys.* **2001**, *38*, 12.
15. Srikanth, K.; Debnath, B.; Jha, T. *Bioorg. Med. Chem.* **2002**, *10*, 1841.
16. Srikanth, K.; Debnath, B.; Jha, T. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 899.
17. Srikanth, K.; Kumar, C. A.; Ghosh, B.; Jha, T. *Bioorg. Med. Chem.* **2002**, *10*, 2119.
18. Debnath, B.; Srikanth, K.; Banerjee, S.; Jha, T. *Internet Electron. J. Mol. Des.* **2002**, *1*, 488.
19. Debnath, B.; Vishnoi, S. P.; Sa, B.; Jha, T. *Internet Electron. J. Mol. Des.* **2003**, *1*, 128.
20. Debnath, B.; Samanta, S.; Naskar, S. K.; Roy, K.; Jha, T. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2837.
21. Jha, T.; Debnath, B.; Samanta, S.; De, A. U. *Internet Electron. J. Mol. Des.* **2003**, *2*, 539.
22. Debnath, B.; Gayen, S.; Samanta, S.; Jha, T. *Bioorg. Med. Chem.* **2003**, *11*, 5493.
23. Samanta, S.; Srikanth, K.; Banarjee, S.; Debnath, B.; Gayen, S.; Jha, T. *Bioorg. Med. Chem.* **2004**, *12*, 1413.
24. Gayen, S.; Debnath, B.; Samanta, S.; Jha, T. *Bioorg. Med. Chem.* **2004**, *12*, 1493.
25. Debnath, B.; Samanta, S.; Roy, K.; Jha, T. *Bioorg. Med. Chem.* **2003**, *11*, 1615.
26. Debnath, B.; Gayen, S.; Naskar, S. K.; Roy, K.; Jha, T. *Drug Des. Discov.* **2003**, *18*, 81.
27. Mouse, a computer program written in C^{++} language by S. K. Naskar and B. Debnath, authors of this article.
28. Multi Regress, a computer program written in C^{++} language by S. K. Naskar and B. Debnath, authors of this article.
29. Snedecor, G. W.; Cochran, W. G. *Statistical Methods*; Oxford & IBH: New Delhi, 1967; pp 381–418.
30. Lien, E. J. In *Drug Design*; Ariens, E. J., Ed.; Academic: New York, 1975; Vol. V, pp 84–94.
31. Schaper, K. J. *Quant. Struct. Act. Relat.* **1999**, *18*, 354.
32. Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407.