CODEN IEJMAT *Internet Electronic Journal of Molecular Design* **2008**, *7*, 55–67 ISSN 1538–6414
BioChem Press http://www.biochempress.com

Inter*net* Electronic Journal of Molecular Design

March 2008, Volume 7, Number 3, Pages 55–67

Editor: Ovidiu Ivanciuc

Computational Modeling of Substitution Effect on HIV–1 Non–Nucleoside Reverse Transcriptase Inhibitors with Kier–Hall Electrotopological State (E–state) Indices

Nitin S. Sapre,[1] Nilanjana Pancholi,[1] and Swagata Gupta [1]

[1] Department of Applied Chemistry SGS Institute of Technology and Science, Indore (MP) India 452003

Received: May 26, 2006; Revised: October 14, 2007; Accepted: February 3, 2008; Published: March 31, 2008

Citation of the article:
N. S. Sapre, N. Pancholi, and S. Gupta, Computational Modeling of Substitution Effect on HIV–1 Non–Nucleoside Reverse Transcriptase Inhibitors with Kier–Hall Electrotopological State (E–state) Indices, *Internet Electron. J. Mol. Des.* **2008**, *7*, 55–67, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Computational Modeling of Substitution Effect on HIV–1 Non–Nucleoside Reverse Transcriptase Inhibitors with Kier–Hall Electrotopological State (E–state) Indices

Nitin S. Sapre,[1,*] Nilanjana Pancholi,[1] and Swagata Gupta [1]

[1] Department of Applied Chemistry SGS Institute of Technology and Science, Indore (MP) India 452003

Received: May 26, 2006; Revised: October 14, 2007; Accepted: February 3, 2008; Published: March 31, 2008

Internet Electron. J. Mol. Des. **2008**, *7* (3), 55–67

Abstract

Motivation The application of Kier–Hall electrotopological state (E–state) indices has helped in deriving mathematical models, where the electronic attributes of an atom in a chemical graph could be encoded as an index. By changing a substituent in a molecule, a corresponding change in its E–state value as well as physicochemical/biological properties is observed. With the help of Kier–Hall electrotopological indices, the effect of substituent could be analyzed and subsequently used in modeling the inhibitory activity of HIV–1 non–nucleoside reverse transcriptase inhibitors (NNRTIs). The applicability of Kier–Hall E–state indices can provide a new insight to the quantitative structure–activity relationship studies of potential drug compounds.

Method. The multiple linear regression (MLR) method is used to generate a predictive model for the training set consisting of a series of 55 derivatives of 2–amino–6–arylsulfonylbenzonitrile (AASBN). The result is cross–validated with the help of a test set.

Results. A QSAR model for a series of 55 AASBN derivatives is generated to understand the effect of substitution on their inhibitory activity (n = 55, R = 0.843, R^2 = 0.711, R^2_{adj} = 0.638, R^2_{pre} = 0.468, SE = 0.513, SSE = 0.865, PRESS = 20.83, p–value = 0.00, F–ratio = 9.64). The functional groups with positive coefficient of E–state values are OCH_3, CH_3, CN (on ring B), Cl, Br, F and CN (on ring A) while with negative coefficient is NH_2. The E–state for the bridging atoms S, SO and SO_2 also showed different positive coefficients. The result of the test set too gave a good fit (n = 5, R = 0.897, R^2 = 0.804, R^2_{adj} = 0.738, R^2_{pre} = 0.239, SE = 0.301, SSE = 0.841, PRESS = 1.05, p–value = 0.0393, F–ratio = 12.30).

Conclusions. A high positive coefficient for the cyano group on ring B ($E_{CN–B}$) indicates that its presence will have a major contribution towards enhancing the inhibitory activity while a high negative coefficient for amino group (E_{NH2}) suggests that its presence will retard the inhibitory activity. Among the halo substituents, Cl seemed to be more effective as compared to Br and F. Also, S as a bridging atom will have a higher impact on enhancing the inhibitory activity, followed by SO and SO_2. Other substituents, owing to their smaller positive coefficients will have a lesser impact on the inhibitory activity. Thus, Kier–Hall E–state indices could be well applied in deriving QSAR models and analyzing the substitution effects.

Keywords. Non nucleoside reverse transcriptase inhibitors; QSAR; quantitative structure–activity relationships; Kier–Hall electrotopological state indices; multiple linear regression.

* Correspondence author; E–mail: sukusap@yahoo.com, nsapre@sgsits.ac.in.

Abbreviations and notations

AASBN, 2–amino–6–arylsulfonylbenzonitrile	MLR, multiple linear regression
AIDS, acquired immunodeficiency syndrome	NNRTIs, non nucleoside reverse transcriptase inhibitors
E–state, electrotopological state	QSAR, quantitative structure–activity relationships
HIV, human immunodeficiency virus	

1 INTRODUCTION

The human immunodeficiency virus–1 (HIV–1) is the most interesting virus in the history of biomedical research [1–10] causing acquired immunodeficiency syndrome (AIDS) which is a pandemic disease. In 1981, AIDS was first detected in the USA while in India it was detected in 1984. Currently, around 20 million people are affected by AIDS. The etiologic agent of AIDS, HIV–1, generally infects the CD–4 helper T–cells of immune system leading to destruction of host immunity. There are three types of viral enzyme related to HIV–1, namely: HIV–1 reverse transcriptase (RT), HIV–1 protease (PT) and HIV–1 integrase (IN). RT is an important target for the development of selective inhibitors and these reverse transcriptase inhibitors (RTIs) are mainly of two types, namely nucleoside reverse transcriptase inhibitors (NRTIs) and non–nucleoside reverse transcriptase inhibitors (NNRTIs). The safety, selectivity and high potency of NNRTIs have made them more crucial as compared to NRTIs [11–13].

Hydroxyethoxyphenylthiothymine(HEPT) [14], qinoxaline derivatives (efavirenz) [15], α–anilinophenylacetamide (α–APA) derivatives [16], 2',5'–bis(O–(tert–butyldimethylsilyl)–3'–spiro–5"–(4"amino–1",2"–oxathiole–2",2"–dioxide)pyrimidine (TSAO) derivatives [17], bis(heteroaryl)piperazine derivatives (BHAP) (Delavirdine) [18], dipyridodiazepinone (Nevirapine) [19], tetrahydroimidazobenzodiazepinone (TIBO) [20], phenylethylthioureathiazole (PETT) [21], are some of the structurally diverse NNRTIs and of these Nevirapine, Delavirdine and Efavirenz have been approved for the treatment of HIV–1 infection.

The NRTIs act at the catalytic site of HIV–1 RT by terminating the DNA synthesis [22], while NNRTIs inhibit the enzyme non–competitively to a site adjacent to the deoxyribonucleoside triphosphate binding site of an enzyme [23–25]. Due to the occurrence of drug–resistant mutations, the effectivity of NNRTIs is reduced, thus posing a serious limitation to its potential utilization [26]. NNRTIs of a new ring system containing 2–amino–6–arylsulfonylbenzonitriles are found to be effective in inhibition of the replication of a variety of HIV–1 strains at the site of reverse transcriptase [27].

To get an insight of the biological activity of an organic compound, it is required to have a precise understanding of the structure–activity relationship (SAR). Thus a quantitative structure–activity relationship (QSAR) study based on Kier–Hall electrotopological state (E–state) indices is performed on 2–amino–6–arylsulfonylbenzonitrile (AASBN) derivatives [27] to analyze the effect of substitution on their inhibitory activity.

A method of mathematical encoding the molecules according to their structural features is adopted herein. The conversion of a chemical structure into a mathematical number (numerical value) can be achieved in various ways [29–32]. E–state indices provide rich information while representing a molecular structure. Here it is derived from the counts of valence and sigma bonded electrons in a hydrogen–suppressed chemical graph representing a molecule. The index is formulated to encode information about the electronegativity, pi and lone pair electron content, topological status and the environment of an atom within the molecule. The environment of an atom mitigates the electronic and topological structure of that atom in such a way that a characteristic state is produced. If a structural change is introduced into a molecule, the environmental effect on a given atom produces some change in the electronic and the topological state of an atom [33–40].

2 MATERIALS AND METHODS

A series of 55 AASBN derivatives [27] is selected for performing molecular modeling studies. For understanding the effect of substitution, Kier–Hall E–state topological indices are used, namely; E_{NH2} (for NH_2 group on ring B), $S_S/S_{SO}/S_{SO2}$ (for bridging S or SO or SO_2 groups), $S_{CN–B}$ (for CN group present on ring B), and for substituents on ring A, S_{CH3} (for CH_3), S_{OCH3} (for OCH_3), $S_{CN–A}$ (for CN), S_{Cl} (for Cl), S_{Br} (for Br) and S_F (for F). Kier–Hall Electrotopological (E–state) indices for various functional groups are calculated using e–calc ver. 1.1 [28]. Multiple linear regression analysis is performed using Sagata ver. 1.0 [41].

2.1 Molecular Modeling

Transformation of a chemical structure into a mathematical graph makes it possible to express the chemical structure of these compounds by a single number. An E–state index is a structure descriptor for an atom within a covalently bonded molecule. The advantage of E–state indices is that they can be used directly as a single number molecular descriptor in QSAR. These relationships are mathematical models that enable the prediction of properties and/or activities from the structural parameters. Statistical modeling is done using Sagata ver. 1.0 [41]. Firstly, a correlation matrix is derived from the program and then regression analyses are performed. Results are summarized for further comparison.

2.2 Methodology

2.2.1 Kier–Hall Electrotopological state (E–state) indices

A molecule is represented with a hydrogen–suppressed graph (i.e. a chemical graph) in which the atoms are identified as elements with certain valence states. The sigma bonds are represented by dimensionless connection between the atoms. The ingredients in the chemical graph are: (a) presence of an atom (b) *valence* state of an atom and (c) degree of adjacency. From these

ingredients, we may create a parameter reflecting the electronic and topological state of an atom in the chemical graph. Quantitatively the topological attribute may be accomplished by using the degree of adjacency which is equal to the count of sigma electrons contributed by an atom in the chemical graph known as δ. To give primary atom the largest value of degree of adjacency, it is convenient to reciprocate the δ value (i.e. 1/δ). The electronic attribute of an atom in a chemical graph must also be encoded into an index with the help of information about the number of pi, sigma and lone–pair of electrons associated with each atom. In addition to these, attribute related to the interactions taking place between atoms within a molecule is interpreted in terms of electronegativity of an atom. The difference between the number of valence electrons and the number of sigma electrons in an atom is known as Kier–Hall electronegativity. The valence electron count on an atom in a chemical graph is designated by δ^v, which is equal to the count of valence electrons contributed by an atom in a molecule (Z^v) minus the count of hydrogen atoms on that atom (H), thus Z^v–H = δ^v. The Kier–Hall electronegativity is given by $\delta^v - \delta$, where: δ = sigma electrons; δ^v = sigma + pi + lone–pair electrons; and $\delta^v - \delta$ = pi + lone pair electrons. This expression provides a way of enumerating the relative structure of an atom in a chemical graph and renders information regarding the potential of both intramolecular as well as intermolecular phenomena. The intrinsic state of an atom in a chemical graph is computed with the formula:

$$I = \frac{\delta^v + 1}{\delta} \tag{1}$$

The intrinsic state of an atom in a chemical graph reflects its electronic and topological attributes in the absence of interaction with rest of the molecule. The E–state of an atom is explicitly given by the sum of its I–state value and the values of all the perturbing terms due to the remaining atoms in the molecule. The E–state for the i^{th} atom is given by the expression

$$S_i = I_i + \sum_j \Delta I_{ij} \tag{2}$$

where the summation is over all the other atoms in a molecule.

The sum of all E–state values in a molecule is equal to the sum of its I–state values:

$$\sum_i S_i = \sum_i I_i \tag{3}$$

2.2.2 Computational details

The E–state indices for NH_2 (on ring B), S/SO/SO_2 (Bridging groups), CN (on ring B), CH_3, OCH_3, CN, Cl, Br, F (on ring A) functional groups used in the present study are calculated using e–calc ver. 1.1 [28]. Multiple linear regression analysis for correlation of these E–state indices with the inhibitory activity (pIC_{50}) is performed using Sagata ver. 1.0 [41] software.

Table 1. Anti–HIV–1 activity (inhibitory concentration pIC_{50}) and electrotopological state (E–state) indices of 2–amino–6–arylsulfonylbenzonitrile (AASBN) analogs (Training set) – see Scheme 1 for the general structure.

No.	R	R'	IC_{50}	pIC_{50}	S_{NH2}	S_S	S_{SO}	S_{SO2}	S_{CN-A}	S_{CH3}
1	H	S	8.70	5.061	5.750	1.550	0.000	0.000	11.172	0.000
2	2–OCH_3	S	2.70	5.569	5.789	1.477	0.000	0.000	11.264	0.000
3	3–OCH_3	S	1.50	5.824	5.781	1.500	0.000	0.000	11.246	0.000
4	4–OCH_3	S	17.0	4.770	5.775	1.514	0.000	0.000	11.232	0.000
5	3–CH_3	S	0.96	6.018	5.780	1.565	0.000	0.000	11.241	2.050
6	4–CH_3	S	5.70	5.244	5.774	1.559	0.000	0.000	11.228	2.051
7	2–Cl	S	7.20	5.143	5.755	1.445	0.000	0.000	11.185	0.000
8	3–Cl	S	16.00	4.796	5.754	1.480	0.000	0.000	11.182	0.000
9	4–Cl	S	12.00	4.921	5.753	1.500	0.000	0.000	11.181	0.000
10	2–Br	S	21.00	4.678	5.776	1.530	0.000	0.000	11.234	0.000
11	3–Br	S	15.00	4.824	5.770	1.535	0.000	0.000	11.220	0.000
12	3–F	S	12.00	4.921	5.706	1.325	0.000	0.000	11.074	0.000
13	2–CN	S	9.10	5.041	5.755	1.379	0.000	0.000	11.188	0.000
14	4–CN	S	19.00	4.721	5.752	1.460	0.000	0.000	11.182	0.000
15	3–CF_3	S	7.10	5.149	5.669	1.081	0.000	0.000	10.999	0.000
16	3–NH_2	S	23.00	4.638	5.755	1.485	0.000	0.000	11.185	0.000
17	2,5–Cl_2	S	3.50	5.456	5.759	1.375	0.000	0.000	11.195	0.000
18	3,5–$(CH_3)_2$	S	1.10	5.959	5.809	1.579	0.000	0.000	11.311	2.071
19	3–OCH_3, 5–CH_3	S	0.14	6.854	5.811	1.514	0.000	0.000	11.315	2.011
20	3–OCH_3, 5–CF_3	S	13.00	4.886	5.700	1.030	0.000	0.000	11.072	0.000
21	2–OCH_3	SO	12.00	4.921	5.730	0.000	11.075	0.000	11.119	0.000
22	3–OCH_3	SO	19.00	4.721	5.722	0.000	11.032	0.000	11.101	0.000
23	4–OCH_3	SO	20.00	4.699	5.716	0.000	11.005	0.000	11.087	0.000
24	2–CH_3	SO	21.00	4.678	5.729	0.000	11.121	0.000	11.116	1.894
25	3–CH_3	SO	10.00	5.000	5.721	0.000	11.054	0.000	11.096	1.936
26	4–CH_3	SO	21.00	4.678	5.715	0.000	11.015	0.000	11.083	1.966
27	2–Br	SO	19.00	4.721	5.717	0.000	11.044	0.000	11.089	0.000
28	3–Br	SO	4.80	5.319	5.711	0.000	11.003	0.000	11.075	0.000
29	4–Br	SO	21.00	4.678	5.707	0.000	10.979	0.000	11.066	0.000
30	2–CN	SO	9.90	5.004	5.696	0.000	10.906	0.000	11.043	0.000
31	3–CN	SO	15.00	4.824	5.695	0.000	10.907	0.000	11.039	0.000
32	4–CN	SO	23.00	4.638	5.693	0.000	10.909	0.000	11.036	0.000
33	3–CF_3	SO	22.00	4.658	5.610	0.000	10.478	0.000	10.854	0.000
34	3,5–$(CH_3)_2$	SO	0.50	6.301	5.751	0.000	11.196	0.000	11.166	1.957
35	2,5–Cl_2	SO	6.20	5.208	5.700	0.000	10.933	0.000	11.051	0.000
36	3–Cl, 5–CH_3	SO	0.52	6.284	5.724	0.000	11.023	0.000	11.105	1.898
37	3–OCH_3, 5–CF_3	SO	0.90	6.046	5.636	0.000	10.571	0.000	10.915	0.000
38	H	SO_2	6.90	5.161	5.620	0.000	0.000	21.012	10.851	0.000
39	2–OCH_3	SO_2	1.40	5.854	5.659	0.000	0.000	21.443	10.943	0.000
40	3–OCH_3	SO_2	0.60	6.222	5.651	0.000	0.000	21.324	10.924	0.000
41	3–CH_3	SO_2	0.20	6.699	5.650	0.000	0.000	21.293	10.920	1.809
42	4–CH_3	SO_2	7.30	5.137	5.644	0.000	0.000	21.217	10.906	1.873
43	2–Cl	SO_2	5.90	5.229	5.625	0.000	0.000	21.122	10.863	0.000
44	3–Cl	SO_2	0.40	6.398	5.624	0.000	0.000	21.092	10.861	0.000
45	4–Br	SO_2	14.00	4.854	5.636	0.000	0.000	21.166	10.889	0.000
46	2–F	SO_2	5.00	5.301	5.564	0.000	0.000	20.568	10.723	0.000
47	2–CN	SO_2	6.00	5.222	5.625	0.000	0.000	21.203	10.867	0.000
48	3–CN	SO_2	1.80	5.745	5.623	0.000	0.000	21.145	10.863	0.000
49	4–CN	SO_2	18.00	4.745	5.622	0.000	0.000	21.111	10.860	0.000
50	3–CF_3	SO_2	5.30	5.276	5.539	0.000	0.000	20.652	10.677	0.000
51	2,5–Cl_2	SO_2	0.30	6.523	5.629	0.000	0.000	21.202	10.873	0.000
52	3–Cl, 5–CH_3	SO_2	0.005	8.301	5.653	0.000	0.000	21.374	10.930	1.746
53	3–OCH_3, 5–CH_3	SO_2	0.01	8.000	5.681	0.000	0.000	21.606	10.993	1.770
54	3–OCH_3, 5–CF_3	SO_2	0.04	7.398	5.570	0.000	0.000	20.965	10.751	0.000
55	1–Naphthyl	SO_2	1.00	6.000	5.732	0.000	0.000	22.113	11.112	0.000

Table 1. (continued)

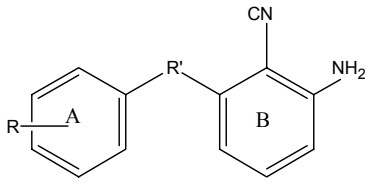
No.	R	R'	IC_{50}	pIC_{50}	S_{OCH3}	S_{CN-A}	S_{Cl}	S_{Br}	S_F
1	H	S	8.70	5.061	0.000	0.000	0.000	0.000	0.000
2	2–OCH_3	S	2.70	5.569	6.915	0.000	0.000	0.000	0.000
3	3–OCH_3	S	1.50	5.824	6.800	0.000	0.000	0.000	0.000
4	4–OCH_3	S	17.0	4.770	6.735	0.000	0.000	0.000	0.000
5	3–CH_3	S	0.96	6.018	0.000	0.000	0.000	0.000	0.000
6	4–CH_3	S	5.70	5.244	0.000	0.000	0.000	0.000	0.000
7	2–Cl	S	7.20	5.143	0.000	0.000	6.074	0.000	0.000
8	3–Cl	S	16.00	4.796	0.000	0.000	5.913	0.000	0.000
9	4–Cl	S	12.00	4.921	0.000	0.000	5.817	0.000	0.000
10	2–Br	S	21.00	4.678	0.000	0.000	0.000	3.484	0.000
11	3–Br	S	15.00	4.824	0.000	0.000	0.000	3.423	0.000
12	3–F	S	12.00	4.921	0.000	0.000	0.000	0.000	13.029
13	2–CN	S	9.10	5.041	0.000	11.157	0.000	0.000	0.000
14	4–CN	S	19.00	4.721	0.000	10.788	0.000	0.000	0.000
15	3–CF_3	S	7.10	5.149	0.000	0.000	0.000	0.000	0.000
16	3–NH_2	S	23.00	4.638	0.000	0.000	0.000	0.000	0.000
17	2,5–Cl_2	S	3.50	5.456	0.000	0.000	6.088	0.000	0.000
18	3,5–$(CH_3)_2$	S	1.10	5.959	0.000	0.000	0.000	0.000	0.000
19	3–OCH_3, 5–CH_3	S	0.14	6.854	8.005	0.000	0.000	0.000	0.000
20	3–OCH_3, 5–CF_3	S	13.00	4.886	6.894	0.000	0.000	0.000	0.000
21	2–OCH_3	SO	12.00	4.921	6.202	0.000	0.000	0.000	0.000
22	3–OCH_3	SO	19.00	4.721	6.699	0.000	0.000	0.000	0.000
23	4–OCH_3	SO	20.00	4.699	6.642	0.000	0.000	0.000	0.000
24	2–CH_3	SO	21.00	4.678	0.000	0.000	0.000	0.000	0.000
25	3–CH_3	SO	10.00	5.000	0.000	0.000	0.000	0.000	0.000
26	4–CH_3	SO	21.00	4.678	0.000	0.000	0.000	0.000	0.000
27	2–Br	SO	19.00	4.721	0.000	0.000	0.000	3.352	0.000
28	3–Br	SO	4.80	5.319	0.000	0.000	0.000	3.329	0.000
29	4–Br	SO	21.00	4.678	0.000	0.000	0.000	3.319	0.000
30	2–CN	SO	9.90	5.004	0.000	11.012	0.000	0.000	0.000
31	3–CN	SO	15.00	4.824	0.000	10.828	0.000	0.000	0.000
32	4–CN	SO	23.00	4.638	0.000	10.711	0.000	0.000	0.000
33	3–CF_3	SO	22.00	4.658	0.000	0.000	0.000	0.000	0.000
34	3,5–$(CH_3)_2$	SO	0.50	6.301	0.000	0.000	0.000	0.000	0.000
35	2,5–Cl_2	SO	6.20	5.208	0.000	0.000	5.832	0.000	0.000
36	3–Cl, 5–CH_3	SO	0.52	6.284	0.000	0.000	6.023	0.000	0.000
37	3–OCH_3, 5–CF_3	SO	0.90	6.046	6.016	0.000	0.000	0.000	0.000
38	H	SO_2	6.90	5.161	0.000	0.000	0.000	0.000	0.000
39	2–OCH_3	SO_2	1.40	5.854	6.451	0.000	0.000	0.000	0.000
40	3–OCH_3	SO_2	0.60	6.222	6.463	0.000	0.000	0.000	0.000
41	3–CH_3	SO_2	0.20	6.699	0.000	0.000	0.000	0.000	0.000
42	4–CH_3	SO_2	7.30	5.137	0.000	0.000	0.000	0.000	0.000
43	2–Cl	SO_2	5.90	5.229	0.000	0.000	5.900	0.000	0.000
44	3–Cl	SO_2	0.40	6.398	0.000	0.000	5.789	0.000	0.000
45	4–Br	SO_2	14.00	4.854	0.000	0.000	0.000	3.241	0.000
46	2–F	SO_2	5.00	5.301	0.000	0.000	0.000	0.000	13.618
47	2–CN	SO_2	6.00	5.222	0.000	10.836	0.000	0.000	0.000
48	3–CN	SO_2	1.80	5.745	0.000	10.703	0.000	0.000	0.000
49	4–CN	SO_2	18.00	4.745	0.000	10.618	0.000	0.000	0.000
50	3–CF_3	SO_2	5.30	5.276	0.000	0.000	0.000	0.000	0.000
51	2,5–Cl_2	SO_2	0.30	6.523	0.000	0.000	5.914	0.000	0.000
52	3–Cl, 5–CH_3	SO_2	0.005	8.301	0.000	0.000	5.894	0.000	0.000
53	3–OCH_3, 5–CH_3	SO_2	0.01	8.000	6.558	0.000	0.000	0.000	0.000
54	3–OCH_3, 5–CF_3	SO_2	0.04	7.398	5.866	0.000	0.000	0.000	0.000
55	1–Naphthyl	SO_2	1.00	6.000	0.000	0.000	0.000	0.000	0.000

2.3 Computer Software

2.3.1 Chemistry software

E–calc ver. 1.1 [28] and Sagata ver. 1.0 [41] are used to perform calculation of the E–state indices and multiple linear regression analysis respectively. WinIDAMS ver. 1.2 [42] is used for plotting the graphs for establishing a relationship between the observed and the calculated IC_{50} values.

3 RESULTS AND DISCUSSION



Scheme 1. General formula of the inhibitors; see Table 1 for substituents and IC_{50}.

Table 1 record the compounds of the series of fifty five AASBN derivatives taken as training set along with the position of substituents (R and R'), where R represents substituent attached to the benzene ring A and R' represents the S/SO/SO_2 group bridging the two aromatic rings. Table 1 also records the respective inhibitory activity IC_{50} and pIC_{50} and the E–state indices (S_{NH2}, S_S, S_{SO}, S_{SO2}, $S_{CN–B}$, S_{CH3}, S_{OCH3}, $S_{CN–A}$, S_{Cl}, S_{Br} and S_F) of AASBN analogs. The correlation matrix for the correlation of E–state indices with pIC_{50} for these aforementioned AASBN derivatives is shown in Table 2.

Table 2 The correlation matrix for AASBN derivatives.

	pIC_{50}	S_{NH2}	S_S	S_{SO}	S_{SO2}	$S_{CN–B}$	S_{CH3}
pIC_{50}	1.000						
S_{NH2}	–0.186	1.000					
S_S	–0.177	0.715	1.000				
S_{SO}	–0.280	0.052	–0.502	1.000			
S_{SO2}	0.474	–0.743	–0.523	–0.466	1.000		
$S_{CN–B}$	–0.202	0.999	0.733	0.053	–0.768	1.000	
S_{CH3}	0.409	0.249	–0.013	0.097	–0.044	0.232	1.000
S_{OCH3}	0.256	0.116	0.042	–0.016	–0.037	0.119	–0.091
$S_{CN–A}$	–0.216	–0.106	–0.098	0.058	0.038	–0.104	–0.229
S_{Cl}	0.217	–0.014	0.040	–0.111	0.070	–0.019	–0.051
S_{Br}	–0.245	0.113	0.000	0.144	–0.124	0.109	–0.194
S_F	–0.074	–0.200	0.034	–0.130	0.070	–0.195	–0.108

	pIC_{50}	S_{OCH3}	$S_{CN–A}$	S_{Cl}	S_{Br}	S_F
S_{OCH3}	0.256	1.000				
$S_{CN–A}$	–0.216	–0.229	1.000			
S_{Cl}	0.217	–0.261	0.194	1.000		
S_{Br}	–0.245	–0.194	–0.144	–0.165	1.000	
S_F	–0.074	–0.108	–0.080	–0.092	–0.068	1.000

A perusal of the correlation matrix given in Table 2 indicates that in univariate correlation E_{SO2} has the highest correlation potential while E_F has the lowest one. The order of univariate correlations of E–state values for various functional groups attached to aromatic ring A and B, with the inhibitory activity follow the order:

$$S_{SO2} > S_{CH3} > S_{SO} > S_{OCH3} > S_{Br} > S_{Cl} > S_{CN-A} > S_{CN-B} > S_{NH2} > S_S > S_F.$$

The substitution effect is highlighted in Eq. (4):

$$\begin{gathered} \text{pIC50} = -277.95(\pm 72.27)\, S_{NH2} + 1.83(\pm 1.08)\, S_S + 0.36(\pm 0.16)\, S_{SO} + 0.33(\pm 0.10)\, S_{SO2} \\ + 123.54(\pm 32.28)\, S_{CN-B} + 0.56(\pm 0.11)\, S_{CH3} + 0.10(\pm 0.03)\, S_{OCH3} + 0.02(\pm 0.02)\, S_{CN-A} + \\ 0.17(\pm 0.04)\, S_{Cl} + 0.10(\pm 0.08)\, S_{Br} + 0.07(\pm 0.03)\, S_F + 219.45(\pm 55.97) \\ n = 55\ R = 0.843\ R^2 = 0.711\ R^2_{adj} = 0.638\ R^2_{pre} = 0.468\ SE = 0.513\ PRESS = 20.83\ SSE = \\ 0.865\ p\text{–}value = 0.000\ F\text{–}ratio = 9.64 \end{gathered} \tag{4}$$

In Eq. (4) a high positive coefficient for E_{CN-B} indicates that presence of CN group on ring B will have an activity enhancing effect while a high negative coefficient for S_{NH2} indicates that presence of NH_2 group will adversely affect the inhibitory activity. As far as bridging groups are concerned, the high positive coefficient for S_S indicates that sulphur as a bridging group has a favorable effect on the inhibitory activity. Although the coefficients of S_{Br}, S_{Cl} and S_F are positive, the halogens have comparatively lower coefficients. Even then their extent of activity enhancing effect is just better as compared to S_{CN-A} (for cyano (CN) group on ring A).

A set of 55 compounds is taken as the training set to develop a QSAR model. Table 3 records the observed and calculated (from Eq. 4) values of pIC_{50} for this training set of AASBN analogs. The quality of correlation is demonstrated by their residual values i.e. the difference between observed and calculated pIC_{50}. The residual values thus obtained are also given in Table 3.

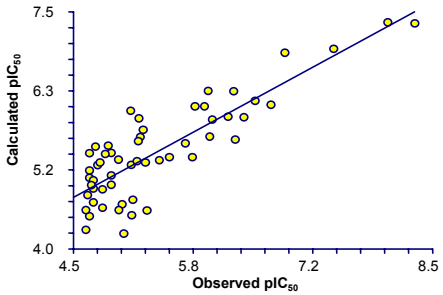


Figure 1. A graph of observed and calculated pIC_{50} values for the training set of AASBN derivatives.

A plot of observed and calculated pIC_{50} values for the training set of AASBN analogs, as plotted by WinIDAMS ver. 1.2 [42] is presented in Figure (1).

Table 3. Observed ($pIC_{50\ exp}$), calculated with Eq. (4) ($pIC_{50\ calc}$), and residuals ($pIC_{50\ exp} - pIC_{50\ calc}$) for AASBN derivatives (training set)

No.	$pIC_{50\ exp}$	$pIC_{50\ calc}$	Residuals
1	5.0605	4.2371	0.8239
2	5.5686	5.3372	0.2318
3	5.8239	5.3674	0.4566
4	4.7696	5.3245	–0.5545
5	6.0177	5.6015	0.4165
6	5.2441	5.6528	–0.4088
7	5.1427	5.2758	–0.1328
8	4.7959	5.2203	–0.4243
9	4.9208	5.3952	–0.4742
10	4.6778	4.9645	–0.2865
11	4.8239	4.9060	–0.0820
12	4.9208	4.8619	0.0591
13	5.0410	4.6799	0.3611
14	4.7212	4.9151	–0.1941
15	5.1487	4.5214	0.6276
16	4.6383	4.3345	0.3035
17	5.4559	5.2738	0.1822
18	5.9586	6.2260	–0.2670
19	6.8539	6.8314	0.0226
20	4.8861	5.5359	–0.6499
21	4.9208	5.0761	–0.1551
22	4.7212	5.1113	–0.3903
23	4.6990	5.0338	–0.3348
24	4.6778	5.4288	–0.7508
25	5.0000	5.1808	–0.1808
26	4.6778	5.2452	–0.5672
27	4.7212	4.6557	0.0653
28	5.3188	4.5767	0.7423
29	4.6778	4.5670	0.1110
30	5.0044	4.6074	0.3966
31	4.8239	4.3888	0.4352
32	4.6383	4.5730	0.0650
33	4.6576	4.8403	–0.1823
34	6.3010	5.5534	0.7476
35	5.2076	5.3013	–0.0933
36	6.2840	6.4321	–0.1481
37	6.0458	5.7993	0.2467
38	5.1612	4.7970	0.3640
39	5.8539	6.1249	–0.2709
40	6.2218	5.9634	0.2586
41	6.6990	6.0911	0.6079
42	5.1367	6.0402	–0.9032
43	5.2291	5.9114	–0.6824
44	6.3979	5.9138	0.4842
45	4.8539	5.4031	–0.5491
46	5.3010	5.3576	–0.0566
47	5.2218	5.6105	–0.3885
48	5.7447	5.6512	0.0938
49	4.7447	5.5460	–0.8010
50	5.2757	5.6968	–0.4208
51	6.5229	6.0636	0.4594
52	8.3010	7.4683	0.8327
53	8.0000	7.2456	0.7544
54	7.3979	6.9259	0.4721
55	6.0000	6.2725	–0.2725

Table 4. Anti–HIV–1 activity (inhibitory concentration pIC_{50}) and electrotopological state (E–state) indices of 2–amino–6–arylsulfonylbenzonitrile (AASBN) analogs (Test set)

No.	R	R'	IC_{50}	PIC_{50}	S_{NH2}	S_S	S_{SO}	S_{SO2}	$S_{CN–B}$	S_{CH3}
1	3–Cl, 5–CH3	S	1.7	5.770	5.783	1.494	0.0	0.000	11.252	1.987
2	4–OCH_3	SO_2	13	4.886	5.645	0.000	0.0	21.249	10.911	0.000
3	2–CH_3	SO_2	4.5	5.347	5.658	0.000	0.0	21.423	10.939	1.714
4	2–Br	SO_2	12	4.921	5.646	0.000	0.0	21.316	10.912	0.000
5	3–$O(CH_2)_3CH_3$, 5–CH_3	SO_2	0.4	6.398	5.736	0.000	0.0	22.014	11.118	1.802

No.	R	R'	IC_{50}	pIC_{50}	S_{OCH3}	$S_{CN–A}$	S_{Cl}	S_{Br}	S_F
1	3–Cl, 5–CH_3	S	1.7	5.770	0.000	0.0	6.018	0.000	0.0
2	4–OCH_3	SO_2	13	4.886	6.480	0.0	0.000	0.000	0.0
3	2–CH_3	SO_2	4.5	5.347	0.000	0.0	00.00	0.000	0.0
4	2–Br	SO_2	12	4.921	0.000	0.0	0.000	3.201	0.0
5	3–$O(CH_2)_3CH_3$, 5–CH_3	SO_2	0.4	6.398	0.000	0.0	0.000	0.000	0.0

A test set of 5 compounds as given in Table 4, has been selected to verify the validity of the predictions obtained from the QSAR model. Table 5 presents the observed and the calculated pIC_{50} values for the test set. The regression thus obtained for the observed and calculated pIC_{50} values show a high degree of relatedness.

Table 5. Observed and calculated (Eq. 4) values for the test set of AASBN derivatives

No.	R	pIC_{50} (Observed)	pIC_{50}(Calculated)
1	3–Cl, 5–CH_3	5.77	6.91
2	4–OCH_3	4.886	6.02
3	2–CH_3	5.347	6.23
4	2–Br	4.921	5.56
5	3–$O(CH_2)_3CH_3$, 5–CH_3	6.398	6.92

$$pIC_{50}\ (\text{Calculated}) = +0.83(\pm 0.24)\ pIC_{50}\ (\text{Observed}) + 1.78(\pm 1.30)$$
$$n = 5\ R = 0.899\ R^2 = 0.804\ R^2_{adj} = 0.738\ R^2_{pre} = 0.239\ SE = 0.301\ PRESS = 1.05\ SSE = 0.841 \quad (5)$$
$$p\text{–}value = 0.039\ F\text{–}ratio = 12.30$$

The results obtained for the test set are self–indicative of fitness of correlation in predicting the substitution effect in AASBN derivatives. A plot of observed and calculated pIC_{50} values is presented in Figure 2.

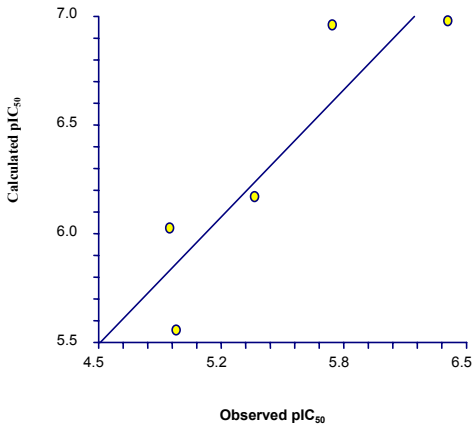


Figure 2. A graph of observed and calculated pIC_{50} values for the test set of AASBN derivatives.

4 CONCLUSIONS

The presence of an electronegative CN group has a great inhibitory activity enhancing impact, while the NH_2 group is adversely affecting the inhibitory activity of these AASBN derivatives to a larger extent. As far as the bridging groups are concerned, the results show that presence of S atom is more effective than the presence of SO or SO_2 group. Other substituents do have a role to play in deciding the inhibitory activity but have a very low contribution as observed from the coefficients of E–states of CH_3, OCH_3, Cl, Br and F groups. Among the halogens, Cl shows better result as compared to Br or F suggesting a moderately electronegative group would be more helpful in enhancing the activity. When E–State values of CH_3 and OCH_3 groups are compared it can be concluded that even though both have enhancing effect on the inhibitory activity, CH_3 appears to be a better substituent than OCH_3. The results of test set show a good linearity between the observed and the calculated values of pIC_{50}.

Thus the results suggests that, of the combination of the referred substituents, amino (NH_2) and cyano (CN) groups have a major role to play in deciding the inhibitory activity of these AASBN derivatives. On the basis of above results applicability of Kier–Hall electrotopological state indices is again proved and they can be used in choosing substituents in a compound to get more potent molecules and enriching the drug discovery process.

5 REFERENCES

[1] W. A. Haseltine, Molecular biology of the human immunodeficiency virus type 1, *FASEB J*. **1991**, *5*, 2349–2360.
[2] A. S. Fauci, The human immunodeficiency virus: infectivity and mechanisms of pathogenesis, *Science*, **1988**, *239*, 617–622.
[3] W. Blattner, R. C. Gallo, H. M. Temin, HIV causes AIDS, *Science*, **1988**, *241*, 515–516.
[4] O. Bagasra, S. P. Hauptman, H. W. Lischner, M. Sachs, R. J. Pomerantz, Detection of human immunodeficiency virus type 1 provirus in mononuclear cells by in situ polymerase chain reaction, *N. Engl. J. Med*. **1992**, *326*, 1385–1391.
[5] A. S. Fauci, Immunopathogenic mechanisms in human immunodeficiency virus (HIV) infection, *Ann. Intern. Med*. **1991**, *114*, 678–693.
[6] P. A. Volberding, S. W. Lagakos, M. A. Koch, et al., Zidovudine in asymptomatic human immunodeficiency virus infection: a controlled trial in persons with fewer than 500 CD4–positive cells per cubic millimetre, *N. Engl. J. Med*. **1990**, *322*, 941–949.
[7] M. V. Ragni, L. A. Kingsley, S. J. Zhou, The effect of antiviral therapy on the natural history of human immunodeficiency virus infection in a cohort of hemophiliacs, *J. Acquir. Immune. Defic. Syndr*. **1992**, *5*, 120–126.
[8] N. M. H Graham, S. L Zeger, L. P. Park, et al., Effect of zidovudine and *Pneumocystis carinii* pneumonia prophylaxis on progression of HIV–1 infection to AIDS, *Lancet* **1991**, *338*, 265–269.
[9] G. F Lemp, S. F Payne, D. Neal, T. Temelso, G. W. Rutherford, Survival trends for patients with AIDS, *JAMA*, **1990**, *263*, 402–406.
[10] B. Tindall, D. A. Cooper, Primary HIV infection: host responses and intervention strategies, *AIDS* **1991**, *5*, 1–14.
[11] E. De Clercq, What can be expected from non–nucleoside reverse transcriptase inhibitors in the treatment of human immunodeficiency virus type 1 infection? *Rev. Med. Virol*. **1996**, 97–117.
[12] E. De Clercq,Perspectives of non–nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV–1 infection, *Farmaco* **1999**, *54*, 26–45.
[13] G. Campiani, A. Ramunno, G. Maga, V. Nacci, C. Fattorusso, B. Catalanotti, B. Morelli, E. Novellino, Non–nucleoside HIV–1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives, *Curr. Pharm. Des*. **2002**, *8*, 615–657.

[14] M. Baba, H. Tanakas, E. De Clercq, R. Pauwels, J. Balzarini, D. Schols, H. Nakashima, C. F. Perno, R. T. Walker, T. Miyasaka, Highly specific inhibition of human immunodeficiency virus type–1 by a novel 6–substituted acyclouridine derivative, *Biochem Biophys*. *Res*. *Commun*. **1989**, *165*, 1375–1381.
[15] C. F. Venezia, K. J. Howard, M. E. Ignatov, L. A. Holladay, M. D. Barkley, Effects of Efavirenz Binding on the Subunit Equilibria of HIV–1 Reverse Transcriptase, *Biochemistry* **2006**, *45*, 2779–2789.
[16] R. Pauwels, K. Andries, Z. Debyser, P. Van Daele, D. Schols, A. M. Vandamme, C. G. M. Janssen, J. Anne, G. Cauwenbergh, J. Desmyter, J. Heykants, M. A. C. Janssen, E. De Clercq, P. A. J. Janssen, Potent and Highly Selective Human Immunodeficiency Virus Type 1 (HIV–1) Inhibition by a Series of α–Anilinophenylacetamide derivatives targeted at HIV–1 Reverse Transcriptase, *Proc*. *Natl*. *Acad*. *Sci*. *USA* **1993**, *90*, 1711–1715.
[17] J. Balzarini, M. J. Pérez–Pérez, A. San–Felix, D. C. F. Schols, Perno, A. Vandamme, M. J. Camarasa, E. De Clercq, 2',5'–Bis–O–(tert–butyldimethylsilyl)–3'–spiro–5"–(4"–amino–1",2"–oxathiole–2",2'–dioxide) pyrimidine (TSAO) nucleoside analogues: highly selective inhibitors of human immunodeficiency virus type 1 that targeted at the viral reverse transcriptase, *Proc*. *Natl*. *Acad*. *Sci*. *USA* **1992**, *89*, 4392–4396.
[18] D. L. Romero, M. Busso, C. K. Tan, F. Reusser, J. R. Palmer, S. M. Poppe, P. F. Aristoff, K. M. Downey, A. G. So, L. Resnick; W. G. Tarpley, Nonnucleoside Reverse Transcriptase Inhibitors that Potently and Specifically Block Human Immunodeficiency Virus Type 1 replication, *Proc*. *Natl*. *Acad*. *Sci*. *USA* **1991**, *88*, 8806–8810.
[19] V.J. Merluzzi, K.D. Hargrave, M. Labadia, K. Grozinger, M. Skoog, J.C. Wu, C.K. Shih, K. Eckner, S. Hattox, J. Adams, Inhibition of HIV–1 replication by a nonnucleoside reverse transcriptase inhibitor, *Science*, **1990**, *250*, 1411–1413.
[20] R. Pauwels, K. Andries, J. Desmyter, D. Schols, M. J. Kukla, H. J. Breslin, A. Raeymaeckers, J. Van Gelder, R. Woestenborghs, J. Heykants, Potent and selective inhibition of HIV–1 replication in vitro by a novel series of TIBO derivative, *Nature*, **1990**, *343*, 470–474.
[21] C. Ahgren, K. Backro, F. W. Bell, A. S. Cantrell, M. Clemens, J. M. Colacino, J. B. Deeter, J. A. Engelhardt, M. Hogberg, S. R. Jaskunas, N. G. Johanssons, The PETT series, a new class of potent non–nucleoside inhibitors of human immunodeficiency virus type–1 reverse transcriptase, *Antimicrob*. *Agents Chemother*. **1995**, *39*, 1329–1335.
[22] To date six nucleosides have been approved by the FDA: AZT, 3TC, d4T, ddI, ddC, and abacavir.
[23] P. P. Mager, A check on rational drug design: Molecular simulation of the allosteric inhibition of HIV–1 reverse transcriptase, *Med*. *Res*. *Rev*. **1997**, *17*, 235–276.
[24] E. Arnold, K. Das, J. Ding, P.N. Yadav, Y. Hsiou, P.L. Boyer, S. H. Hughes, Targeting HIV reverse transcriptase for anti–AIDS drug design: structural and biological considerations for chemotherapeutic strategies, *Drug*. *Des*. *Discov*. **1996**, *13*, 29–47.
[25] H. Jonckheere, J. Anne, E. De Clercq, The HIV–1 reverse transcription (RT) process as target for RT inhibitors, *Med*. *Res*. *Rev*. **2000**, *20*, 129–154.
[26] D. Richman, C. K. Shih, I. Lowy, J. Rose, P. Prodanovich, S. Goff, and Griffin, Human Immunodeficiency Virus Type 1 Mutants Resistant to Nonnucleoside Inhibitors of Reverse Transcriptase arises in Tissue Culture, *Proc*. *Natl*. *Acad*. *Sci*. *USA* **1991**, *88*, 11241–11245.
[27] H. Chan. Joseph, 2–Amino–6–arylsulfonylbenzonitriles as Non–nucleoside Reverse Transcriptase Inhibitors of HIV–1, *J*. *Med*. *Chem*. **2001**, *44*, 1866–1882.
[28] Kier–Hall, E–calc version 1.1, **1999**.
[29] M. Randic, P. G. Seybold, Molecular Shape as a Critical Factor in Structure–Property–Activity Studies, *SAR QSAR Environ*. *Res*. **1993**, *1*, 77–85.
[30] M. Randić, In Search of Structural Invariants, *J*. *Math*. *Chem*. **1992**, *9*, 97–146.
[31] N. Trinajstić, *Chemical Graph Theory*; CRC Press: Boca Raton, FL, **1983**, Vols. 1 and 2.
[32] P. G. Seybold, M. May, U. A. Bagal, Molecular Structure–Property Relationships, *J*. *Chem*. *Educ*. **1987**, *64*, 575–581.
[33] L. B. Kier, L. H. Hall, *Molecular Structure Description: The Electrotopological State*, Academic Press, San Diego, **1999**.
[34] L. B. Kier, L. H. Hall, Design of molecules from quantitative structure–activity relationship models. III. Role of higher order path counts: Path Three, *J*. *Chem*. *Inf*. *Comput*. *Sci*. **1993**, *33*, 598–603.
[35] L. B. Kier, L. H. Hall, B. K. Mohney, The electro–topological state: An atom index for QSAR, *Quant. Struct.-Act. Relat.* **1991**, *10*, 43–51.
[36] B. Testa, L. B. Kier, The concept of structure in structure–activity relations by studies and drug design, *Med*. *Res*. *Rev*, **1991**, *11*, 35–48.
[37] L. B. Kier, L. H. Hall, The generation of molecular structures for a graph based equations, *Quant. Struct.-Act. Relat.* **1994**, *12*, 383–388.
[38] L. B. Kier, L. H. Hall, Quantitative information analysis: The new center of gravity in medicinal chemistry, *Med*. *Chem*. *Res*. **1997**, *7*,335–339.
[39] L. B. Kier, L. H. Hall, J. W. Frazer, Design of molecules from quantitative structure–activity relationship models.

I. Information transfer between path and vertex degree counts, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 143–147.
[40] C. Hansch, A quantitative approach to biochemical structure–activity relationships, *Acc. Chem. Res.* **1969**, *2*, 232–239.
[41] Sagata Regression ver. 1.0, Sagata Regression ver. 1.0, http://www.sagata.com.
[42] WinIDAMS 1.2, http://www.unesco.org/idams, July 2004.

Biographies

First Author is a Professor in the Department of Chemistry at the Shri. GS Institute of Technology and Science, Indore. After obtaining a Ph.D. degree in chemistry from the University of Indore, Dr. Nitin S. Sapre undertook postdoctoral research with Professor Susan Tucker at the University of California Davis and then with Prof. Mitchell Avery at the University of Mississippi.

Second Author is a Ph.D. Student having focused her work on computational chemistry and is working in the field of molecular modeling wherein she is using graph theoretical indices for study. Also, major part of her research covers virtual screening and its application in drug design.

Third Author is an Assistant Professor of Chemistry working for her Ph.D. with the first author. Her main focus of the work is on virtual screening and its application in drug design especially with reference to anti HIV–1 compounds.