

Molecular modelling studies on 2-amino 6-aryl-sulphonylbenzotrioles as non-nucleoside reverse transcriptase inhibitors of HIV-1: A QSPR approach

NITIN S SAPRE,^{1,*} NILANJANA PANCHOLI,¹ SWAGATA GUPTA,² ARUN SIKRWAR³ and NEELIMA SAPRE⁴

¹Department of Applied Chemistry, SGSITS, Indore 452 001

²Department of Chemistry, Government P.G. College, MHOW

³Department of Chemistry, Government Holkar Science College, Indore 452 001

⁴Department of Computer Applications, S.D. Bansal College of Technology, Indore 452 001
e-mail: sukusap@yahoo.com, nsapre@sgsits.ac.in

MS received 11 April 2007; revised 30 October 2007

Abstract. Lipophilicity or hydrophobicity is a crucial physico-chemical property of an oral drug compound. In the present study, we have analysed the structural parameters responsible for enhancing the lipophilicity expressed in terms of Octanol–Water partition coefficient, $\log P$, of 2-amino-6-arylsulphonylbenzotriole (AASBN) derivatives used as NNRTIs in AIDS therapy. Connectivity based Randic (χ) and Balaban (J) and atomistic Kier–Hall electrotopological state (E -state) indices have been used to develop Quantitative Structure–Property Relationship (QSPR) and to predict the effect of substitution on the $\log P$. Model has been developed using multiple linear regression analysis (MLR) for the training set (67 compounds) and the model was tested on a test set (7 compounds). Significant results were obtained for the training set ($R^2 = 0.948$, $R_{\text{adj}}^2 = 0.939$, $SE = 0.177$, F -ratio = 101.22). The results of the test set too implicated a good fit ($R^2 = 0.941$, $R_{\text{adj}}^2 = 0.929$, $SE = 0.157$, F -ratio = 80.05). Among the two connectivity based topological indices; Randic (χ) index showed better predictive ability than the Balaban (J) index. Kier–Hall E -state indices indicated that among the functional groups, methyl, bromo, chloro groups on ring A, with their positive coefficients enhanced the lipophilicity. Amino, cyano group on ring B and the bridging S, SO, SO₂ with their negative coefficients showed an adverse effect on the lipophilicity parameter. Thus, Kier–Hall E -state indices along with topological indices could be well applied for deriving QSPR models and analysing substitution effects of various functional groups. The training set, correlation matrix and observed and experimental $\log P$ values are available as supplementary material for this article.

Keywords. QSPR; AASBN; NNRTIs; randic; Balaban; Kier–Hall E -state indices; MLR.

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is a pandemic disease, whose etiologic agent is human immunodeficiency virus type-1 (HIV-1). The causative agent of AIDS, infects the cells of the immune system leading to destruction of the host immunity, generally the CD-4 helper T-cells. There are three types of viral enzymes related to HIV namely: HIV-protease, HIV-integrase and HIV-reverse transcriptase.^{1–3} The reverse transcriptase enzyme is an important target for the development of selective inhibitors. The HIV reverse transcriptase inhibitors are of two types: nucleoside reverse transcriptase inhibitors

(NRTIs) and non nucleoside reverse transcriptase inhibitors (NNRTIs). The safety, selectivity and high potency of NNRTIs have made them more important in the field of drug research as compared to NRTIs.⁴

Structurally diverse NNRTIs such as hydroxyethoxyphenylthiothymine (HEPT),⁵ quinoxaline derivatives (efavirenz),⁶ α -anilinophenylacetamide (α -APA) derivatives,⁷ 2',5'-bis(O-(tert-butyl dimethylsilyl)-3'-spiro-5''-(4''-amino-1'',2''-oxathiole-2'',2''-dioxide) pyrimidine (TSAO) derivatives,⁸ bis(heteroaryl)piperazine derivatives (BHAP) (Delavirdine),⁹ dipyridodiazepinone (Nevirapine),¹⁰ tetrahydroimidazobenzodiazepinone (TIBO),¹¹ phenylethylthiourea-thiazole (PETT),¹² have been reported in literature and out of these Nevirapine, Delavirdine and Efavirenz have been approved for the treatment of HIV-1 infection.

*For correspondence

The NRTIs act at the catalytic site of HIV-RT by terminating the DNA synthesis¹³ while NNRTIs inhibit the enzyme non-competitively to a site adjacent to deoxyribonucleoside triphosphate binding site of the enzyme.^{14–16} This adjacent site is approximately 10 Å away from the catalytic site. The rapid development of drug resistance and mutation occurs, thus reducing the effectivity of NNRTIs to a large extent.¹⁷ Also, it is well known that the side chain and the backbone of the residues surrounding the active site pocket, adjust to each bound drug in a common fashion. Thus, this viral protein is able to accommodate inhibitors of different chemical structures.

NNRTIs of a new ring system containing AASBN derivatives have been found to effectively inhibit the replication of a variety of HIV-1 strains at the reverse transcriptase step.¹⁸ Therefore, the present work has been performed with the aims of deriving the QSPR and understanding the effects of substitution of various functional groups of these AASBN derivatives with the help of topological indices and electro-topological indices on their lipophilicity. These relationships are mathematical models that enable the prediction of properties and/or activities from the structural parameters. It is worthwhile to mention that the graphs here consist of one and the same cycle; they differ in the acyclic part only.

Mathematical encoding of the molecules according to their structural features has been used here. Transformation of a chemical structure into a mathematical graph makes it possible to express the chemical structure of these compounds by a single number and can be achieved in different ways.^{19–22}

It appears that among many topological indices^{23–36} that have been proposed, the Randic connectivity index (χ)³¹ introduced by Randic and the Balaban index (J)³⁷ introduced by Balaban, are the most widely used indices in QSAR as well as QSPR studies. 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/QSPR studies.

The present study has been performed with the aim of obtaining information about the structural characteristics responsible for enhancing the octanol-water partition coefficient ($\log P$) of this class of compounds i.e. to understand the effect of substitution on their lipophilicity. Also the study is aimed at determining QSPR of HIV-1 NNRTIs for the class of AASBN derivatives.¹⁸

2. Materials and methods

In the present work, 74 compounds of 2-amino-6-arylsulfonylbenzotrioles¹⁸ are taken and Randic,³¹ Balaban³⁷ indices (topological indices) and for understanding the substitution effect electro-topological indices are correlated with lipophilicity $\log P$. Connectivity based Randic and Balaban indices were calculated using in-house built program. *E*-state indices for various functional groups were calculated using e-calc.⁴⁵ Multiple linear regression analysis was performed using systat regression software ver. 11.0.⁴⁴ Firstly, a correlation matrix was derived from the program and then regression analyses were performed.

2.1 Methodology

2.1a *Topological indices used*: Two connectivity based topological indices namely, Randic connectivity index (χ)³¹ and Balaban index (J)³⁷ are well presented in literature. Therefore they will be described here rather briefly.

The Randic connectivity index (χ).^{31–33} The connectivity index, $\chi = \chi(G)$ of G is defined as

$${}^1\chi_R = \sum [\delta_i \delta_j]^{-0.5} \quad (1)$$

where δ_i and δ_j are the valencies of vertices i and j , equal to the number of bonds connected to the atoms i and j in G , representing the graph of a compound.

The Balaban index (J)³⁷. The Balaban index $J = J(G)$ of G is defined as:

$$J = [b / (\mu + 1)] \sum_{i-k} (\delta_i \delta_k)^{-1/2}, \quad (2)$$

where b is number of bonds in G , μ is the cyclomatic number of G and δ_i and δ_k (i or $k = 1, 2, 3, N$; is the number of vertices in G) are the distance sums. The cyclomatic number $\mu = \mu(G)$ of a polycyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related to acyclic subgraph. In the case of monocyclic graph $\mu = 1$.

2.1b Kier–Hall electrotopological state (*E*-state) indices^{38–43}: A molecule is represented with a hydro-

gen-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. $1/\delta$). 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 $Z^v - H = \delta^v$. 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 electron.

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.

Intrinsic state of an atom in a chemical graph is shown by the formula:

$$I = \frac{\delta^v + 1}{\delta} \quad (3)$$

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.

Mathematically E -state for the i th atom is given by the expression

$$S_i = I_i + \sum_j \Delta I_{ij}, \quad (4)$$

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

Sum of the molecule's E -state values is equal to the sum of its I -state values.

$$\sum_i S_i = \sum_i I_i \quad (5)$$

Hence the sum depends only on the number and the type of atoms in a molecule and not on their mutual interactions.

3. Results and discussion

All the 67 chemical structures, taken as the training set, of 2-amino-6-arylsulfonylbenzotrile derivatives are illustrated (see supplementary material 1). It also records the value of graph theoretical descriptors (J , χ) and Kier–Hall electrotopological state indices (E_{NH_2} , E_{S} , E_{SO} , E_{SO_2} , E_{CH_3} , E_{OCH_3} , E_{TCN} , E_{Cl} , E_{Br}) of AASBN analogs. The respective octanol-water partition coefficient, $\log P$ values, the position of the substituents (R and R') on the benzene ring where R' represents the S/SO/SO₂ group bridging the two aromatic rings and R represents various substituents attached to the benzene ring A, are also recorded in it.

The correlation matrix for the correlation of $\log P$ with the earlier mentioned structural descriptors (topological indices) and Electrotopological descriptors of AASBN derivatives are shown (see supplementary material 2).

A perusal of the correlation matrix showed that in the case of univariate correlation, among the topological indices, Balaban index (J) showed better correlation potential as compared to Randic index (χ). But among the E -state indices E_{S} showed best correlation potential while E_{CH_3} showed poorest correlation potential. The E_{CN} and E_{Br} were capable of predicting the $\log P$ value to similar extent, though E_{CN} gave slightly better result than E_{Br} . The order of correlation for the value of E -state indices and topological indices from univariate correlation matrix showed the following order:

$$E_{\text{S}} > J > E_{\text{SO}_2} > E_{\text{SO}} > E_{\text{Cl}} > \chi > \\ E_{\text{CN}} > E_{\text{Br}} > E_{\text{NH}_2} > E_{\text{CH}_3}.$$

A QSPR model was generated using the aforementioned parameters for understanding the effect of substitution as well as the connectivity and following equation was obtained:

$$\begin{aligned} \log P = & 0.370(\pm 0.047)\chi - 1.898 \\ & (\pm 0.353)J - 2.234(\pm 0.423)E_{\text{NH}_2} - \\ & 0.505(\pm 0.364)E_{\text{S}} - 0.202(\pm 0.045)E_{\text{SO}} - \\ & 0.109(\pm 0.022)E_{\text{SO}_2} - 0.830(\pm 0.220)E_{\text{TCN}} + \quad (6) \\ & 0.319(\pm 0.034)E_{\text{CH}_3} + 0.132(\pm 0.010)E_{\text{Cl}} + \\ & 0.303(\pm 0.021)E_{\text{Br}} + 26.422 \end{aligned}$$

$n = 66$, $R = 0.974$, $R^2 = 0.948$, $R_{\text{adj}}^2 = 0.939$, $\text{SE} = 0.177$, $F\text{-ratio} = 101.22$, outlier = Comp. 65.

The quality of the model was considered as statistically satisfactory on the basis of the above mentioned results. In the aforementioned equation, connectivity based Randic index showed a high positive coefficient where as Balaban index showed a high negative coefficient. But in the case of E -state indices a high positive coefficient for E_{CH_3} indicates that the presence of CH_3 group will have an enhancing effect on the $\log P$ value. Among the halo substituents, high positive coefficient for Br as compared to Cl shows that, a bulky and less electronegative halogen substituent will have an enhancing effect on the lipophilicity. A high negative coefficient for E_{NH_2} and E_{CN} indicates that the presence of nucleophilic groups such as NH_2 and CN on ring B will adversely affect the octanol–water partition coefficient. As far as bridging groups are concerned, the negative coefficients of E_{S} , E_{SO} and E_{SO_2} in decreasing order indicates that they all have an unfavourable impact on the $\log P$ values and also suggests that there should be a more bulky bridging group than SO_2 to have a better lipophilicity.

The $\log P$ values for the AASBN derivatives have been calculated using the best multivariate correlation. Such calculated $\log P$ values are recorded (see supplementary material 3). The observed (experimental) $\log P$ values are also given. The quality of correlations has been demonstrated by their residual values, i.e. the difference between the observed and the calculated $\log P$ values. The residuals obtained from the best correlation are also given therein.

A graph showing correlation of observed and calculated $\log P$ values for training set is presented in figure 1.

A test set of 7 compounds as given in table 1 has been taken to verify the validity of model obtained from aforementioned calculations.

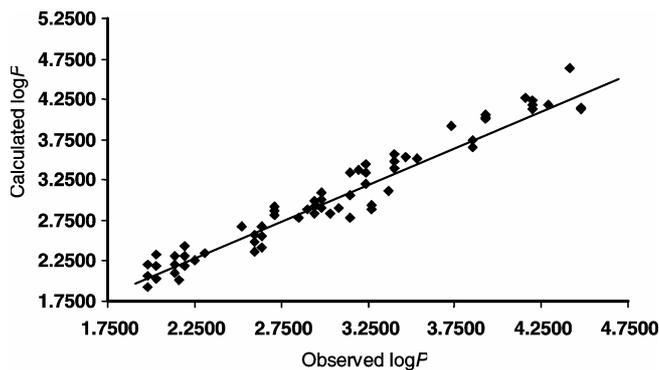


Figure 1. A plot of observed and predicted $\log P$ values for AASBN derivatives for training set.

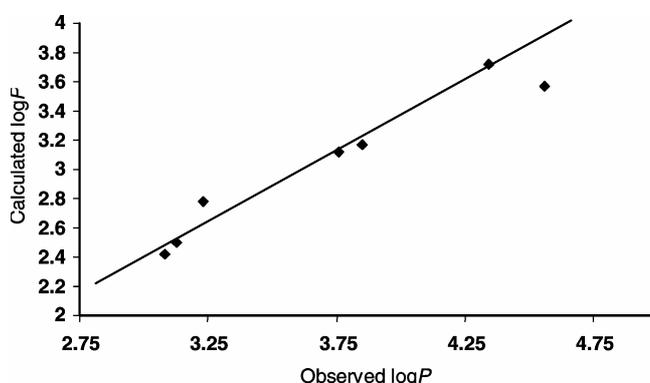


Figure 2. A plot of observed and predicted $\log P$ values for AASBN derivatives for test set.

Table 2 presents the observed and calculated $\log P$ values for the test set. The regression thus obtained for the observed and calculated $\log P$ values shows a high degree of relatedness.

A graph showing correlation of observed and calculated $\log P$ values for test set is presented in figure 2.

The above results of test set are self indicative of fitness of correlation in predicting the connectivity as well as substitution effect in AASBN derivatives.

4. Conclusions

The present study related to 2-amino-6-arylsulfonyl-benzonitrile derivatives leads us to make the following conclusions:

The results have indicated that among the two topological indices; Randic (χ) index showed a better predictive capability as compared to the Balaban (J) index. Among the functional groups with positive coefficient of E -state values on ring A; CH_3 group showed the highest positive value, while among the

Table 1. Octanol/water partition coefficient ($\log P$), topological indices (χ , J) and E -state indices (E_{NH_2} , E_{S} , E_{SO} , E_{SO_2} , E_{CH_3} , E_{OCH_3} , E_{TCN} , E_{Cl} , E_{Br}) of AASBN analogues (test set).

R	R'	$\log P$	χ	J	E_{NH_2}	E_{S}	E_{SO}	E_{SO_2}	E_{TCN}	E_{CH_3}	E_{Cl}	E_{Br}
4-CH ₃	S	3.850	9.113	1.907	5.774	1.559	0.000	0.000	11.228	0.000	0.000	0.0
3-NH ₂	S	4.56	10.024	1.934	5.755	1.485	0.000	0.000	11.185	0.000	0.000	0.0
3,5-(CH ₃) ₂	S	4.340	9.507	1.957	5.809	1.579	0.000	0.000	11.311	2.071	0.000	0.0
3,5-(CH ₃) ₂	SO	3.080	9.935	2.067	5.751	0.000	11.195	0.000	11.165	1.957	0.000	0.0
2,5-Cl ₂	SO	3.230	9.952	2.098	5.700	0.000	10.933	0.000	11.051	0.000	5.832	0.0
3,5-(CH ₃) ₂	SO ₂	3.130	10.268	2.183	5.679	0.000	0.000	21.575	10.989	1.830	0.000	0.0
3-O(CH ₂) ₃ CH ₃ ,5-CH ₃	SO ₂	3.760	12.306	2.100	5.736	0.000	0.000	22.014	11.118	1.802	0.000	0.0

Table 2. Calculated and observed $\log P$ values of 2-amino-6-arylsulfonylbenzotriazoles analogues (test set) derived from the regression (7).

Log P (obsd.)	log P (calc.)	Residual = log P (obsd.) – log P (calc.)
3.850	3.170	0.6800
4.560	3.569	0.9910
4.340	3.723	0.6170
3.080	2.423	0.6570
3.230	2.777	0.4530
3.130	2.501	0.6290
3.760	3.122	0.6380

$$\log P(\text{calc.}) = 1.145(\pm 0.128)\log P(\text{obs.}) + 0.224 \quad (7)$$

($n = 7$, $R = 0.970$, $R^2 = 0.941$, $R_{\text{adj}}^2 = 0.929$, $\text{SE} = 0.157$, $F\text{-ratio} = 80.05$)

halo substituents, Br seemed to be better as compared to Cl indicating that, a bulky and a less electronegative halogen group would enhance the lipophilicity. The functional groups with negative coefficient were NH₂ and CN (on ring B) indicating that they would have an adverse effect on compound's lipophilicity, suggesting that, the presence of electrophilic groups on ring B may help in enhancing the magnitude of octanol/water partition coefficient. The E -state values for the bridging atoms also showed different negative coefficients in the order, $\text{S} > \text{SO} > \text{SO}_2$. This indicates that there should be a more bulky group than SO₂ in the referred position, which would help in enhancing the $\log P$ value. The test set too showed a good linearity between the observed and the calculated $\log P$ values, thus validating the aforementioned inferences.

Thus, the results suggest that the methyl and bulky halogens can play a major role in deciding the lipophilic character of these compounds. Also, new and more potent compounds with enhanced lipophilicity could be designed and synthesized, utilizing the

above information regarding electrophilicity of a group attached to it. On the basis of above results, these structure based indices ascertain their applicability in choosing the substituents in a chemical entity, so as to obtain a potent and novel molecule.

Acknowledgements

The authors thank Dr R C Sarswat, Director, and Dr P K Sen, Dean (R&D), SGSITS, Indore, for providing computing facilities.

References

- Kubinyi H 1994 *Quant. Struct.-Act. Relat.* **13** 285
- De Clercq E 1996 *Rev. Med. Virol.* **6** 97
- De Clercq E 1999 *Farmaco* **54** 26
- Campiani G, Ramunno A, Maga G, Nacci V, Fattorusso C, Catalanotti B, Morelli E and Novellino E 2002 *Curr. Pharm. Des.* **8** 615
- Baba M, Tanakas H, De Clercq E, Pauwels R, Balzarini J, Schols D, Nakashima H, Perno C F, Walker R T and Miyasaka T 1989 *Biochem. Biophys. Res. Commun.* **165** 1375
- Kleim J P, Bender R, Billhardt U M, Meichsner C, Riess G, Rosner M, Winkler I and Paessens A 1993 *Antimicrobial. Agents Chemotherapy* **37** 1659
- Pauwels R, Andries K, Debyser Z, Van Daele P, Schols D, Vandamme A M, Janssen C G M, Anne J, Cauwenbergh G, Desmyter J, Heykants J, Janssen M A C, De Clercq E and Janssen P A J 1993 *Proc. Natl. Acad. Sci. (USA)* **90** 1711
- Balzarini J, Pérez-Pérez M J, San-Felix A, Schols, Perno D C F, Vandamme A, Camarasa M J and De Clercq E 1992 *Proc. Natl. Acad. Sci. (USA)* **89** 4392
- Romero D L, Busso M, Tan C K, Reusser F, Palmer J R, Poppe S M, Aristoff P F, Downey K M, So A G, Resnick L and Tarpley W G 1991 *Proc. Nat. Acad. Sci.* **88** 8806
- Merluzzi V J, Hargrave K D, Labadia M, Grozinger K, Skoog M, Wu J C, Shih C K, Eckner K, Hattox S and Adams J 1990 *Science* **250** 1411

11. Pauwels R, Andries K, Desmyter J, Schols D, Kukla M J, Breslin H J, Raeymaeckers A, Van Gelder J, Woestenborghs R and Heykants J 1990 *Nature (London)* **343** 470
12. Ahgren C, Backro K, Bell F W, Cantrell A S, Clemens M, Colacino J M, Deeter J B, Engelhardt J A, Hogberg M, Jaskunas S R and Johanssons N G 1995 *Antimicrobial Agents and Chemotherapy* **39** 1329
13. To date six nucleosides have been approved by the FDA: AZT, 3TC, d4T, ddI, ddC, and abacavir
14. Mager P P 1997 *Med. Res. Rev.* **17** 235
15. Arnold E, Das K, Ding J, Yadav P N, Hsiou Y, Boyer P L and Hughes S H 1996 *Drug Des. Discov.* **13** 29
16. Jonckheere H, Anne J and De Clercq E 2000 *Med. Res. Rev.* **20** 129
17. Richman D, Shih C K, Lowy I, Rose J, Prodanovich P, Goff S and Griffin 1991 *J. Proc. Natl. Acad. Sci. (USA)* **88** 11241
18. Chan Joseph H 2001 *J. Med. Chem.* **44** 1866
19. Lucic B, Nikolic S, Trinajstic N, Juric A and Mihalic Z 1995 *Croat. Chem. Acta* **68** 417
20. Randic M and Seybold P G 1993 *SAR-QSAR Environ. Res.* **1** 77
21. Randic M 1992 *J. Math. Chem.* **9** 97
22. Trinajstic N 1983 *Chemical graph theory* (Florida: CRC Press) vol 2, pp 3–70
23. Seybold P G, May M and Bagal U A 1987 *J. Chem. Educ.* **64** 575
24. Gutman I and Polansky P E 1989 *Mathematical concepts in organic chemistry* (Berlin: Springer-Verlag) pp 23–120
25. King R B 1987 *Chemical applications of topology and graph theory* (Amsterdam: Elsevier) pp 1–47
26. Merrifield R E and Simmons H E 1989 *Topological methods in chemistry* (New York: Wiley) pp 67–90
27. Trinajstic N, Sabijic A and Nikolic S 1986 *Acta. Pharm. Jugosel.* **36** 1
28. King R B and Rouvray D H 1987 *Graph theory and topology in chemistry* (Amsterdam: Elsevier) pp 1–50
29. Trinajstic N, Nicolic S and Carter S 1989 *QSAR: Theory Appl.* **38** 469
30. Randic M 1975 *J. Am. Chem. Soc.* **97** 6609
31. Randic M 1991 *J. Chem. Inf. Comput. Sci.* **31** 311
32. Randic M 1991 *New J. Chem.* **18** 517
33. Mihalic Z, Nikolic S, Trinajstic N 1992 *J. Chem. Inf. Comput. Sci.* **32** 28
34. Bonchev D 1983 *Information theoretic indices for characterization of chemical structures* (Chichester: Research Studies Press) pp 23–57
35. Mandloi M et al 2000 *J. Chem. Inf. Comput. Sci.* **40** 57
36. Dobrynin Gutman I 1994 *Pib. De L'inst. Math.* **56** 18
37. Balaban A T 1982 *Chem. Phys. Lett.* **80** 399
38. Kier L B and Hall L H 1999 *Molecular structure description: The electro-topological state* (San Diego: Academic Press) pp 20–58
39. Kier L B and Hall L H 1993 *J. Chem. Inf. Comput. Sci.* **33** 598
40. Kier L B, Hall L H and Mohny B K 1991 *Quant. Structure-Activity Relationships* **10** 43
41. Testa B and Kier L B 1991 *Med. Res. Rev.* **11** 35
42. Kier L B and Hall L H 1994 *Quant. Structure-Activity Relationships* **12** 383
43. Kier L B and Hall L H 1997 *Med. Chem. Res.* **7** 335
44. SYSTAT 10.0 SYSTAT Software Inc. USA
45. Kier-Hall, E-calc version 1.1, 1999