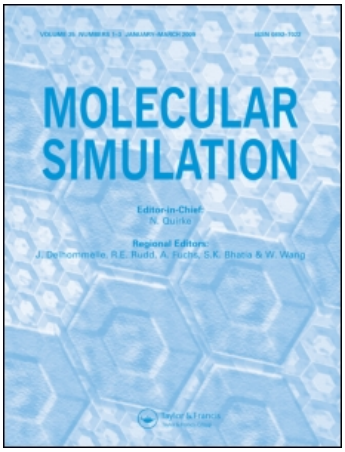



**Molecular Simulation**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

# A QSAR study for 2-(4-aminophenyl)benzothiazoles: using DFT optimisation of geometry of molecules

Rifaat Hilal[a]; Shabaan A. K. Elroby[b]
[a] Department of Chemistry, Faculty of Science, University of Cairo, Giza, Egypt [b] Department of Chemistry, Faculty of Science, King Abdul-Aziz University, Jeddah, Saudi Arabia

Online publication date: 28 January 2011

**To cite this Article** Hilal, Rifaat and Elroby, Shabaan A. K.(2011) 'A QSAR study for 2-(4-aminophenyl)benzothiazoles: using DFT optimisation of geometry of molecules', Molecular Simulation, 37: 1, 62 — 71
**To link to this Article: DOI:** 10.1080/08927022.2010.520133
**URL:** http://dx.doi.org/10.1080/08927022.2010.520133

*Molecular Simulation*
Vol. 37, No. 1, January 2011, 62–71

Taylor & Francis
Taylor & Francis Group

# A QSAR study for 2-(4-aminophenyl)benzothiazoles: using DFT optimisation of geometry of molecules

Rifaat Hilal[a] and Shabaan A.K. Elroby[b]*

[a]*Department of Chemistry, Faculty of Science, University of Cairo, Giza, Egypt;* [b]*Department of Chemistry, Faculty of Science, King Abdul-Aziz University, Jeddah, Saudi Arabia*

(*Received 29 May 2010; final version received 22 August 2010*)

Quantitative structure–activity relationships (QSARs) have been established for two sets of antitumour drugs 2-(4-aminophenyl)benzothiazoles (APBT). Constitutional, geometrical, topological, electronic descriptors (computed at the B3LYP/6-31G** level) and some empirical descriptors related to the hypophilicity were computed and analysed. Multiple regression analysis led to a set of equations that reflected the weight of each of the studied descriptors. The most relevant of these descriptors were grouped, and a new multiple regressions analysis was carried out and we arrived at the final QSAR models. A validation set of 11 APBT were selected, and their activities were computed using the proposed QSAR model. The correlation between the predicted and observed activities was excellent. The resulting best models exhibited good $q^2$ and $r^2$ values up to 0.867 and 0.954.

**Keywords:** QSAR; DFT; 2-(4-aminophenyl)benzothiazole (APBT); MlogP; $E_{HOMO} - E_{LUMO}$



## 1. Introduction

Quantitative structure–activity relationship (QSAR) methods show wide applications due to their well-established predictive power. Essentially, correlating the physico-chemical properties of a series of compounds with their respective biological activities is believed to provide a useful tool in designing new drugs. Since first introduced by Hansch [1], QSAR investigations of several series of drugs, enzymes and biologically active compounds have been performed, and the predictive power of QSAR equations is now well appreciated [2].

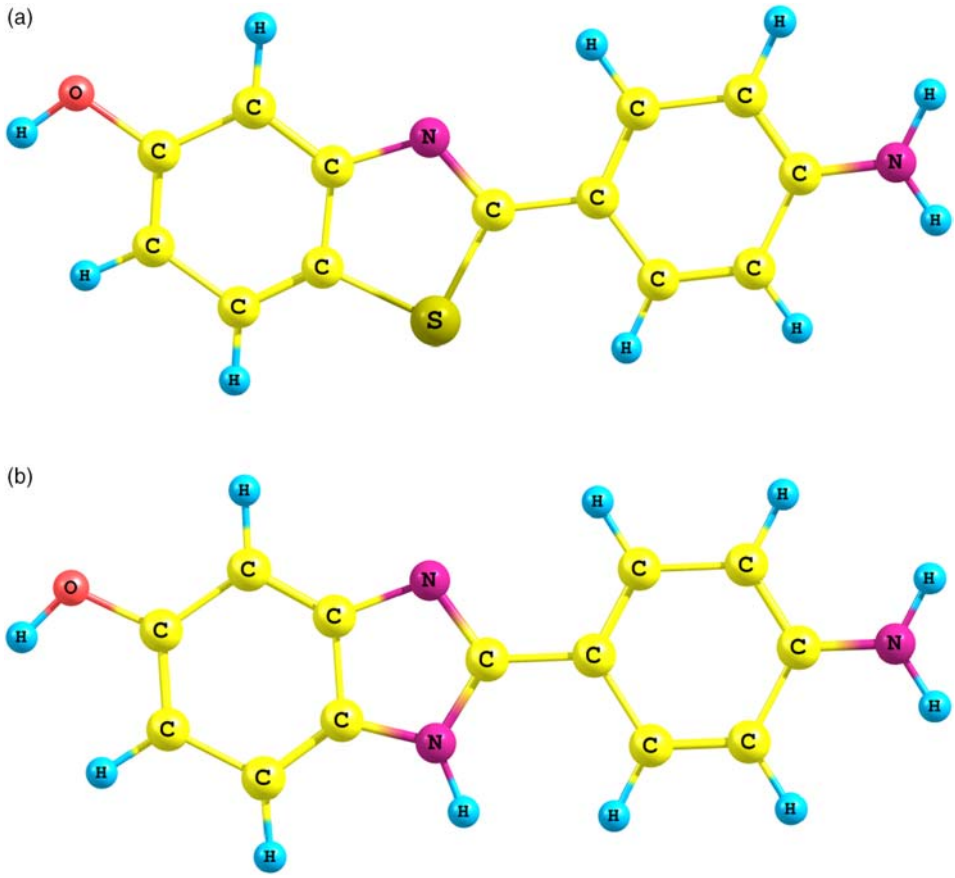
Novel 2-(4-aminophenyl)benzothiazoles (APBT) possess highly selective, potent antitumour properties *in vitro* and *in vivo*. Modification of the heterocyclic nucleus to generate benzoxazole or benzimidazole congeners APBT had a dyschemotherapeutic effect. Analysis of structure–activity relationships identified the benzothiazole nucleus as being essential for potent activity, and that substitution at position 3′ in the phenyl ring with a halogen atom or an alkyl group enhanced potency in the breast carcinoma panel and extended the *in vitro* spectrum of activity to include certain human ovarian, lung, renal and colon carcinoma cell lines [3] (Scheme 1). The nature of 3′-substituent in the arylamine fragment exerts a profound influence on the predominant metabolic process, the extent of metabolism and the bioactivity of N-acetyl metabolites [4]. Elucidation of the mechanism of action of this structurally simple class of compounds has occurred in parallel with the selection of a candidate clinical agent. Antitumour benzothiazoles induce and are biotransformed by cytochrome P-450 1A1 to putative active as well as inactive metabolites. Metabolic inactivation of the molecule has been thwarted by isosteric replacement of hydrogen with fluorine atoms at positions around the benzothiazole nucleus.

An important step in predicting the effects of these chemicals is the estimation of their binding to the receptor. To date, however, the use of QSAR models to estimate binding affinity across multiple chemical classes has shown only modest success possibly due, in part, to a focus on minimum energy chemical structures as the active molecules. McKinney and co-workers [5,6] showed that chemicals which have greater ability to accept electron density through charge-transfer interaction should bind to the aryl hydrocarbon receptor (AhR) with greater affinity than those with lower electron-acceptor properties. This would suggest, therefore, that these stronger electron acceptors should have a lower energy unoccupied frontier orbital ($E_{LUMO}$), lower energy for the occupied frontier orbital ($E_{HOMO}$) and a lower energy difference in these frontier orbitals ($E_{HOMO} - E_{LUMO}$), which can be related to molecular reactivity [7,8]. The charge-transfer interaction should increase in conformers that are more planar because aligning the planes of the aromatic rings increases electron-acceptor properties and because the planar configuration permits closer proximity to the putative electron donor region of the receptor. The purpose of the present study is to

*Corresponding author. Email: skamel@kau.edu.sa

ISSN 0892-7022 print/ISSN 1029-0435 online

DOI: 10.1080/08927022.2010.520133
http://www.informaworld.com




Scheme 1. Structures of (a) compound **1** and (b) compound **3** in the first working set.



attempt to use the mechanistic concepts of AhR binding to develop a robust QSAR model applicable to multiple classes of chemicals.

A previous QSAR study of $^{125}$I-labelled APBT derivatives as imaging agents for β-amyloid in the brain with Alzheimer's disease was carried out by Wang et al. [9]. The study of QSAR indicated that the initial brain uptake was correlated with molecule volume (Vm) and dipole moment (Dp), but the Dp was the main factor.

In the present study, QSAR models were generated using two training sets. The first consists of 16 molecules (the biological activity was expressed in $IC_{50}$) and the second of 15 molecules (the biological activity was expressed in $GI_{50}$) (cf. Tables 1 and 2).

Our QSAR is intended to be comprehensive, in the sense that structural parameters such as constitutional, topological, geometrical and empirical descriptors will be thoroughly investigated.

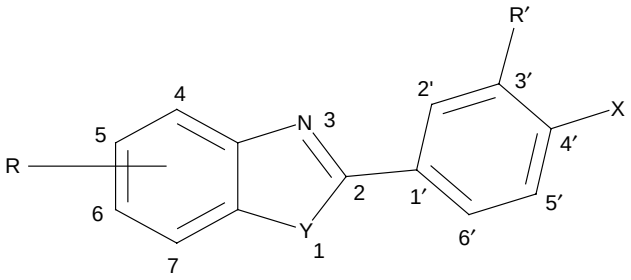
## 2. Materials and method

Full geometry optimisations were performed at the DFT level of theory. Equilibrium geometries were identified, and zero-point energies were computed for the two sets of APBT studied (cf. Tables 1 and 2) at the B3LYP/6-31G** level. All computations were carried out using the Gaussian 98 [10] computer software package. The electronic descriptors were obtained from a single-point calculation at the B3LYP/6-311++G** level.

The Dragon software package [11] is used to compute all constitutional, [12] topological, [13] geometrical [14] and empirical [15] descriptors [16]. Dragon generates over 800 descriptors, and no QSAR treatment can handle such a huge number of descriptors. Therefore, limitation of this number is essential from the statistical point of view. To select the set of descriptors that were most relevant to the bioactivity of APBT, a two-step computational strategy was adopted. Each of the four main groups of descriptors is considered independently in the first step. A nonlinear regression analysis of each group of descriptors against the $\text{Log}(1/IC_{50})$ and $\text{Log}(1/GI_{50})$ is performed. This is followed by a factor impact analysis to evaluate the weight of each descriptor. The second step is to collect all relevant descriptors emerged in step 1 in the final QSAR analysis. The descriptors used in the present work are presented and defined in Appendix 1.

Statistical treatment of the data and multiple regression analysis were carried out using the Origin (version 6.0) computer software package. It operates under Windows

Table 1. Structures of APBT in the first working set.




| Compound numbers | R | R′ | Y | X | Compound name | MW (g/mol) | Mean $IC_{50}$ (M)[a] |
|---|---|---|---|---|---|---|---|
| **1** | 5-OH | H | S | NH2 | IIS5OH | 242.32 | 100 [4] |
| **2** | 6-OH | H | S | NH2 | IIS6OH | 242.32 | 21.6 [4] |
| **3** | H | H | N | NH2 | IIN | 209.27 | 3.16 [16] |
| **4** | H | Cl | S | NH2 | IISCl | 305.21 | 0.001 [4] |
| **5** | H | Br | S | NH2 | IISBr | 260.76 | 0.001 [17] |
| **6** | 4-OH | Cl | S | NH2 | IISCl4OH | 276.76 | 61.4 [4] |
| **7** | 5-OH | Cl | S | NH2 | IISCl5OH | 276.76 | 60.4 [4] |
| **8** | 6-OH | Cl | S | NH2 | IISCl6OH | 276.76 | 54.5 [4] |
| **9** | H | I | S | NH2 | IISI | 268.36 | 0.001 [4] |
| **10** | H | Me | S | NH2 | IISMe | 240.35 | 0.001 [4] |
| **11** | 4-OH | Me | S | NH2 | IISMe4OH | 256.35 | 100 [4] |
| **12** | 5-OH | Me | S | NH2 | IISMe5OH | 256.35 | 78.7 [4] |
| **13** | 6-OH | Me | S | NH2 | IISMe6OH | 256.35 | 100 [4] |
| **14** | 7-OH | Me | S | NH2 | IISMe7OH | 256.35 | 90.8 [4] |
| **15** | H | H | O | NH2 | IIO | 210.25 | 0.03 [17] |
| **16** | 4-OH | H | S | NH2 | IIS4OH | 242.32 | 100 [4] |
| | | | | Validation set | | | |
| **17** | 4-OH | Cl | S | NHAc | IISAcCl4OH | 318.8 | 56.8 [4] |
| **18** | 5-OH | Cl | S | NHAc | IISAcCl5OH | 318.8 | 31 [4] |
| **19** | H | H | N | NHAc | IINAc | 251.31 | 10 [17] |
| **20** | H | H | O | NHAc | IIOAc | 252.29 | 10 [17] |
| **21** | 6-OH | Me | S | NHAc | IISAcMe6OH | 298.39 | 33 [4] |

[a] IC, inhibitory concentration: $IC_{50}$: the concentrations at which growth or activity is inhibited by 50% applies to ligand and growth inhibition. Log 10 scale is frequently used when $x$ values are a serial dilution. Better estimate of the standard error is obtained when a log 10 scale is used.



environment and employs a large number of regression models (both linear and nonlinear) as well as various interpolation schemes to represent the data in the most precise and convenient way. The cross-validation analysis was carried out using the leave-one-out method where one compound was removed from the data-set and its activity was predicted using the model derived from the rest of the data-set. The cross-validated $q^2$ and the optimum number of components were obtained.

## 3. Results and discussion

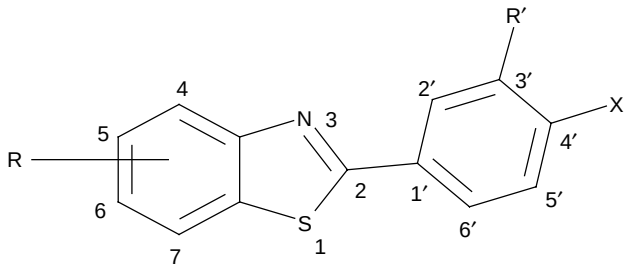
### *3.1 Structural correlations*

Table 3 presents the nine constitutional, molecular properties and empirical descriptors computed for the representative sets of APBT (cf. Tables 1 and 2). Attempts were made to correlate the individual constitutional descriptors with both the $Log1/IC_{50}$ and $Log(1/GI_{50})$. Multiple regression analysis on all nine constitutional descriptors vs. the $Log1/IC_{50}$ results in an excellent correlation ($r^2 = 0.888$), which is given in Table 8. Out of the nine descriptors involved in the regression, $Log(1/IC_{50})$ seems to be governed by only four descriptors, namely the Moriguchi octanol–water partition coefficient (MlogP), mean atomic van der Waals volume (Mv), mean atomic Sanderson electronegativity (Me) and mean atomic polarisability (Mp). This correlation suggests that the biological activity of the first set ($Log(1/IC_{50})$) is governed to a large extent by the molecular properties and the constitutional factors. On the other hand, a less satisfactory correlation is obtained for the biological activity of the second set ($Log(1/GI_{50})$) vs. the molecular properties and the constitutional descriptors. The QSAR equations and the corresponding statistical data are given in Table 9.

Tables 4 and 5, present the geometrical and topological descriptors, respectively computed for the first set of benzothiazoles studied in the present work.

Multiple regression analysis on the activity of the first set vs. topological descriptors failed to arrive at a good

Table 2. Structures of APBT in the second working set.




| Compound numbers | R | R′ | X | Compound name | MW (g/mol) | Mean $GI_{50}$ (M)[a] |
|---|---|---|---|---|---|---|
| **1** | H | H | NH2 | IIS | 226.32 | 16.3 [18] |
| **2** | 5-F | H | NH2 | IIS5F | 244.31 | 23.4 [18] |
| **3** | 6-F | H | NH2 | IIS6F | 244.31 | 30.2 [18] |
| **4** | 5,6 di-F | Me | NH2 | IISMe5,6F | 276.33 | 67.6 [18] |
| **5** | 4-F | Me | NH2 | IISMe4F | 258.34 | 34.6 [18] |
| **6** | H | Me | NH2 | IISMe | 240.35 | 12.9 [17] |
| **7** | H | I | NHAc | IISAcI | 394.25 | 19.9 [17] |
| **8** | H | Br | NH2 | IISBr | 305.21 | 21.4 [17] |
| **9** | 5-F | Br | NH2 | IISBr5F | 323.2 | 5 [18] |
| **10** | H | Cl | NH2 | IISCl | 260.76 | 17.7 [17] |
| **11** | 6-F | Me | NH2 | IISMe6F | 258.34 | 47.8 [18] |
| **12** | 5-F | Cl | NH2 | IISCl5F | 278.75 | 5.8 [18] |
| **13** | 6-F | Cl | NH2 | IISCl6F | 278.75 | 22.3 [18] |
| **14** | 5-F | Me | NH2 | IISMe5F | 258.34 | 4.4 [18] |
| **15** | H | I | NH2 | IISI | 268.36 | 17.7 [17] |
| **16** | 5-F | I | NH2 | IISI5F | 370.2 | 7.5 [18] |
| | | | Validation set | | | |
| **17** | H | Me | NHAc | IISAcMe | 282.39 | 43.6 [17] |
| **18** | H | Cl | NHAc | IISAcCl | 302.8 | 28.8 [17] |
| **19** | H | Br | NHAc | IISAcBr | 347.25 | 30.2 [17] |
| **20** | H | F | NH2 | IISF | 244.31 | 14.4 [19] |
| **21** | H | CF3 | NH2 | IISCF3 | 294.32 | 14.1 [19] |

[a] GI, growth inhibition: $GI_{50}$: the drug concentration giving a 50% reduction in the net protein increase.



Table 3. Constitutional, empirical and molecular properties descriptors computed for the first working set of APBT.

| Compound numbers | MW (g/mol) | AMW | Mv | Me | Mp | Ms | MR ($cm^3$/mol) | PSA ($A^2$) | MlogP |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 242.32 | 8.97 | 0.70 | 1.01 | 0.75 | 2.28 | 62.16 | 28.24 | 2.01 |
| **2** | 242.32 | 8.97 | 0.70 | 1.01 | 0.75 | 2.28 | 62.16 | 28.24 | 2.01 |
| **3** | 209.27 | 7.75 | 0.68 | 1.00 | 0.71 | 2.11 | 64.06 | 15.79 | 2.42 |
| **4** | 260.76 | 10.03 | 0.74 | 1.01 | 0.79 | 2.17 | 65.27 | 28.24 | 3.11 |
| **5** | 305.21 | 11.74 | 0.75 | 1.00 | 0.81 | 2.09 | 68.09 | 28.24 | 3.24 |
| **6** | 276.76 | 10.25 | 0.73 | 1.02 | 0.78 | 2.37 | 66.96 | 28.24 | 2.28 |
| **7** | 276.76 | 10.25 | 0.73 | 1.02 | 0.78 | 2.37 | 66.96 | 28.24 | 2.28 |
| **8** | 276.76 | 10.25 | 0.73 | 1.02 | 0.78 | 2.37 | 66.96 | 28.24 | 2.28 |
| **9** | 268.36 | 8.66 | 0.70 | 1.00 | 0.74 | 2.23 | 68.85 | 45.31 | 2.86 |
| **10** | 240.35 | 8.29 | 0.69 | 0.99 | 0.74 | 2.05 | 65.50 | 28.24 | 3.11 |
| **11** | 256.35 | 8.54 | 0.69 | 1.00 | 0.73 | 2.25 | 67.20 | 28.24 | 2.28 |
| **12** | 256.35 | 8.54 | 0.69 | 1.00 | 0.73 | 2.25 | 67.20 | 28.24 | 2.28 |
| **13** | 256.35 | 8.54 | 0.69 | 1.00 | 0.73 | 2.25 | 67.20 | 28.24 | 2.28 |
| **14** | 256.35 | 8.54 | 0.69 | 1.00 | 0.73 | 2.25 | 67.20 | 28.24 | 2.28 |
| **15** | 210.25 | 8.09 | 0.69 | 1.00 | 0.71 | 2.18 | 54.02 | 13.14 | 2.42 |
| **16** | 242.32 | 8.97 | 0.70 | 1.01 | 0.75 | 2.28 | 62.16 | 28.24 | 2.01 |

Table 4. Geometrical descriptors computed for the first working set of APBT.

| Compound numbers | G1 | G2 | SPAM | DISPm | QXXm | DISPv | QXXv | QXXe | G(N···N) |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 23.527 | 11.763 | 0.478 | 6.984 | 28.571 | 5.581 | 29.382 | 56.465 | 6.43 |
| **2** | 23.606 | 11.767 | 0.478 | 8.934 | 26.548 | 5.763 | 29.247 | 55.789 | 6.44 |
| **3** | 20.504 | 10.759 | 0.467 | 2.237 | 18.138 | 4.644 | 25.501 | 52.173 | 12.88 |
| **4** | 24.99 | 11.878 | 0.479 | 11.583 | 44.604 | 4.868 | 33.469 | 56.358 | 6.44 |
| **5** | 27.843 | 12.434 | 0.477 | 20.882 | 86.023 | 3.907 | 39.429 | 58.569 | 6.43 |
| **6** | 27.221 | 12.648 | 0.472 | 9.041 | 54.879 | 5.599 | 38.223 | 67.898 | 6.44 |
| **7** | 26.922 | 12.648 | 0.476 | 9.377 | 48.608 | 6.538 | 36.577 | 64.063 | 6.44 |
| **8** | 26.784 | 12.622 | 0.484 | 8.662 | 44.663 | 3.717 | 33.362 | 57.712 | 6.44 |
| **9** | 26.99 | 13.111 | 0.478 | 9.109 | 32.495 | 14.304 | 33.515 | 63.438 | 6.43 |
| **10** | 23.387 | 11.579 | 0.474 | 8.901 | 31.52 | 10.293 | 37.3 | 71.929 | 6.55 |
| **11** | 25.589 | 12.34 | 0.458 | 9.841 | 40.593 | 10.237 | 41.729 | 82.828 | 6.44 |
| **12** | 25.297 | 12.34 | 0.457 | 10.966 | 35.587 | 9.314 | 40.082 | 77.964 | 6.44 |
| **13** | 25.244 | 12.319 | 0.475 | 10.76 | 31.919 | 7.615 | 38.48 | 75.246 | |
| **14** | 25.448 | 12.315 | 0.459 | 8.735 | 37.848 | 7.458 | 40.038 | 81.195 | 6.44 |
| **15** | 20.663 | 10.757 | 0.471 | 3.424 | 20.263 | 6.146 | 26.265 | 52.312 | 6.41 |
| **16** | 23.827 | 11.763 | 0.473 | 5.827 | 34.335 | 6.696 | 31.041 | 60.741 | 6.43 |

correlation ($r^2 = 0.566$). However, topological descriptors correlate with a better extent ($r^2 = 0.995$) with the second set.

Multiple regression analysis on the activity vs. geometrical descriptors for the second set results in excellent correlations ($r^2 = 0.870$ and 0.895), which are given in Tables 8 and 9. Combination between the geometrical and topological descriptors has a much better correlation with the activity of the first set.

In summary, the activity of the second set seems geometrically dependent, whereas the activity of the first set is much more involving and depends on the geometrical, constitutional and topological properties of the drug (Table 6). Furthermore, the activities of both the first and the second sets are highly dependent on the partition coefficient of the drug as measured by the MlogP properties descriptor.

### *3.2 Energy correlations*

Table 7 presents some energy descriptors for benzothiazole derivatives computed at the B3LYP/6-311++G** level of theory.

An important observation in the electronic QSAR Equation (1) is the occurrence of $E_{LUMO}$, $E_{HOMO}$ and $\Delta E_{gap}$ as common statistically significant descriptors for this equation (Table 8).

The regression equation is obtained for only nine descriptors. An important observation in this equation is the occurrence of $E_{HOMO}$, $E_{LUMO}$ and $\Delta E_{gap}$ as common descriptors. The QSAR equation and the corresponding statistical data are given in Table 8.

In summary, the activities of the first and second sets are highly dependent on the $\Delta E_{gap}$ of the drug as measured by the electronic descriptor, i.e. $\Delta E_{gap}$ is an important stability index. A large $\Delta E_{gap}$ implies high

Table 5. Topological descriptors computed for the first working set of APBT.

| Compound numbers | AAC | HNar | DELS | TIE | IC1 | SIC1 | SIC2 | CIC2 | IC3 |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 1.669 | 2 | 13.071 | 72.408 | 3.151 | 0.663 | 0.8 | 0.95 | 4.533 |
| **2** | 1.669 | 2 | 13.158 | 71.524 | 3.151 | 0.663 | 0.779 | 1.052 | 4.431 |
| **3** | 1.388 | 2.087 | 7.924 | 52.338 | 2.625 | 0.552 | 0.704 | 1.407 | 4.162 |
| **4** | 1.676 | 2 | 10.512 | 66.162 | 2.958 | 0.629 | 0.79 | 0.987 | 4.47 |
| **5** | 1.676 | 2 | 8.472 | 60.442 | 3.035 | 0.646 | 0.79 | 0.987 | 4.47 |
| **6** | 1.842 | 1.929 | 17.425 | 99.395 | 3.384 | 0.712 | 0.864 | 0.649 | 4.681 |
| **7** | 1.842 | 1.929 | 16.919 | 91.345 | 3.384 | 0.712 | 0.864 | 0.649 | 4.681 |
| **8** | 1.842 | 1.929 | 17.006 | 90.243 | 3.384 | 0.712 | 0.848 | 0.723 | 4.607 |
| **9** | 1.612 | 2 | 11.855 | 62.083 | 2.888 | 0.583 | 0.756 | 1.21 | 4.518 |
| **10** | 1.468 | 2 | 7.575 | 58.654 | 2.828 | 0.582 | 0.784 | 1.048 | 4.487 |
| **11** | 1.629 | 1.929 | 14.125 | 87.329 | 3.239 | 0.66 | 0.835 | 0.809 | 4.615 |
| **12** | 1.629 | 1.929 | 13.635 | 80.993 | 3.239 | 0.66 | 0.835 | 0.809 | 4.615 |
| **13** | 1.629 | 1.929 | 13.721 | 80.07 | 3.239 | 0.66 | 0.822 | 0.875 | 4.548 |
| **14** | 1.629 | 1.929 | 14.28 | 84.036 | 3.239 | 0.66 | 0.822 | 0.875 | 4.548 |
| **15** | 1.496 | 2.087 | 10.352 | 56.413 | 2.7 | 0.575 | 0.716 | 1.337 | 4.287 |
| **16** | 1.669 | 2 | 13.573 | 78.434 | 3.151 | 0.663 | 0.8 | 0.95 | 4.533 |

Table 6. Final set of descriptors relevant to the activity computed for the first working data-set of APBT.

| Compound numbers | Log1/$IC_{50}$ (M) | SIC2 | IC3 | IC1 | TIE | G(N···N) | $Q_S$ | $\Delta E_{gap}$ | MlogP |
|---|---|---|---|---|---|---|---|---|---|
| **1** | −2.00 | 0.8 | 4.533 | 3.151 | 72.408 | 6.43 | −0.61926 | −0.1445 | 2.014 |
| **2** | −1.33 | 0.779 | 4.431 | 3.151 | 71.524 | 6.44 | −0.61437 | −0.1466 | 2.014 |
| **3** | −0.50 | 0.704 | 4.162 | 2.625 | 52.338 | 12.88 | −0.10119 | −0.1573 | 2.42 |
| **4** | 3.00 | 0.79 | 4.47 | 2.958 | 66.162 | 6.44 | −0.5690 | −0.1494 | 3.114 |
| **5** | 3.00 | 0.79 | 4.47 | 3.035 | 60.442 | 6.43 | −0.5910 | −0.1482 | 3.243 |
| **6** | −1.79 | 0.864 | 4.681 | 3.384 | 99.395 | 6.44 | −0.60014 | −0.1446 | |
| **7** | −1.78 | 0.864 | 4.681 | 3.384 | 91.345 | 6.44 | −0.61809 | −0.1444 | 2.278 |
| **8** | −1.74 | 0.848 | 4.607 | 3.384 | 90.243 | 6.44 | −0.63991 | −0.1452 | 2.278 |
| **9** | 3.00 | 0.756 | 4.518 | 2.888 | 62.083 | 6.43 | −0.57177 | −0.1479 | |
| **10** | 3.00 | 0.784 | 4.487 | 2.828 | 58.654 | 6.55 | −0.57305 | −0.1490 | 3.114 |
| **11** | −2.00 | 0.835 | 4.615 | 3.239 | 87.329 | 6.44 | −0.5934 | −0.1440 | 2.278 |
| **12** | −1.90 | 0.835 | 4.615 | 3.239 | 80.993 | 6.44 | −0.59778 | −0.1452 | 2.278 |
| **13** | −2.00 | 0.822 | 4.548 | 3.239 | 80.07 | 6.56 | −0.59778 | −0.1464 | 2.278 |
| **14** | −1.96 | 0.822 | 4.548 | 3.239 | 84.036 | 6.44 | −0.36691 | −0.1492 | 2.278 |
| **15** | 1.52 | 0.716 | 4.287 | 2.7 | 56.413 | 6.41 | −0.07909 | −0.1554 | 2.42 |
| **16** | −2.00 | 0.8 | 4.533 | 3.151 | 78.434 | 6.43 | −0.5984 | −0.145 | 2.014 |

stability for the molecule in the sense of its lower sensitivity in the chemical reaction. However, these correlations suggest that the studied set involves a charge-transfer interaction where the $E_{LUMO}$ of the drug plays a pronounced role.

### 3.3 The QSAR model

A good correlation between structure and activity should possess high correlation coefficient, *R*, low standard deviation, SD, and least number of variables. To this end, objective feature selection was done to weed out those descriptors that provide minimal or redundant information.

The present study presents a comprehensive QSAR analysis for APBT as an antitumour drug. Constitutional, geometrical, topological, electronic descriptors and some empirical descriptors related to the hypophilicity were computed and analysed. Multiple regression analysis led to a set of equations that reflected the weight of each of the studied descriptors. The most relevant of these descriptors were grouped, and a new multiple regression analysis was carried out and we arrived at the final QSAR equation for the activity of the first set. Thus, the QSAR equation can be written as

$$\mathrm{Log}(1/\mathrm{IC}_{50}) = -8.08 - 28.29\,\mathrm{SIC2} - 137.18\,\Delta E_{gap} + 4.014\,\mathrm{MlogP}, \quad (1)$$

$n = 16$, $r^2 = 0.943$, $q^2 = 0.854$, $F = 15.012$, SD = 0.546 and $P < 0.0001$.

From a statistical point of view, it is recommended that the best acceptable QSAR equation is the one which is characterised by an *n*/*m* ratio $\geq 5$, where *n* is the number of molecules in the set and *m* is the number of descriptors. For the present working set of 16 molecules, three or even two

Table 7. Electronic descriptors computed for the first working set of APBT at the B3LYP/6-311++G** level.

| Compound numbers | DM | $E_{HOMO}$ | $E_{LUMO}$ | $\Delta E_{gap}$ | C4′—N7 | $Q_{N7}$ | $Q_{N3}$ | $Q_S$ |
|---|---|---|---|---|---|---|---|---|
| **1** | 4.5052 | −0.2032 | −0.0586 | −0.1446 | 1.3720 | −0.3728 | 0.2081 | −0.6193 |
| **2** | 3.6023 | −0.2031 | −0.0565 | −0.1467 | 1.3883 | −0.2913 | 0.2445 | −0.6144 |
| **3** | 4.5193 | −0.2038 | −0.0465 | −0.1573 | 1.3903 | −0.2701 | 0.0144 | −0.1012 |
| **4** | 2.9498 | −0.2161 | −0.0666 | −0.1495 | 1.3800 | −0.2550 | 0.2430 | −0.5690 |
| **5** | 3.1498 | −0.2127 | −0.0644 | −0.1483 | 1.3600 | −0.3130 | 0.2480 | −0.5910 |
| **6** | 3.6900 | −0.2129 | −0.0683 | −0.1446 | 1.3646 | −0.3278 | 0.1995 | −0.6001 |
| **7** | 3.5117 | −0.2101 | −0.0656 | −0.1444 | 1.3650 | −0.3294 | 0.2306 | −0.6181 |
| **8** | 2.2913 | −0.2060 | −0.0607 | −0.1453 | 1.3656 | −0.3305 | 0.2480 | −0.6399 |
| **9** | 3.0466 | −0.2113 | −0.0633 | −0.1480 | 1.3660 | −0.3010 | 0.2393 | −0.5718 |
| **10** | 3.1631 | −0.2075 | −0.0585 | −0.1490 | 1.3893 | −0.2860 | 0.2632 | −0.5731 |
| **11** | 5.4850 | −0.2036 | −0.0595 | −0.1440 | 1.3743 | −0.3566 | 0.1999 | −0.5934 |
| **12** | 4.0731 | −0.2052 | −0.0599 | −0.1452 | 1.3895 | −0.2836 | 0.2353 | −0.5978 |
| **13** | 3.6040 | −0.2013 | −0.0549 | −0.1464 | 1.3904 | −0.2860 | 0.2664 | −0.5978 |
| **14** | 2.1108 | −0.2068 | −0.0575 | −0.1493 | 1.3899 | −0.2828 | 0.2624 | −0.3669 |
| **15** | 3.2880 | −0.2104 | −0.0549 | −0.1555 | 1.3868 | −0.2921 | 0.0724 | −0.0791 |
| **16** | 5.2588 | −0.2063 | −0.0612 | −0.1451 | 1.3715 | −0.3712 | 0.1775 | −0.5984 |

Table 8. QSAR equations and the corresponding statistical data ($n = 16$) computed for the first working set.

| Descriptors | QSAR equation | $r$ | $F$ statistical | $P$ |
|---|---|---|---|---|
| Geometrical | $\text{Log1/IC}_{50} = 82.77 + 7.41\,\text{G1} - 21.47\,\text{G2} + 2.43\,\text{SPAM} + 0.31\,\text{DISPm} - 0.23\,\text{QXXm} + 1.06\,\text{DISPv} - 0.18\,\text{QXXv} - 0.09\,\text{QXXe} + 0.20\,\text{G(N}\cdots\text{N)}$ | 87.390 | 4.622 | 0.03806 |
| Topological | $\text{Log1/IC}_{50} = 85.66 - 7.40\,\text{AAC} - 160.46\,\text{HNar} - 0.23\,\text{DELS} + 0.01\,\text{TIE} - 101.14\,\text{IC1} + 407.20\,\text{SIC1} + 108.17\,\text{SIC2} + 59.67\,\text{CIC2} + 33.75\,\text{IC3}$ | 99.520 | 138.884 | 0.0001 |
| Constitutional and empirical | $\text{Log1/IC}_{50} = 49.56475 - 0.09478\,\text{MW} + 3.05275\,\text{AMW} - 28.6614\,\text{Mv} - 29.6018\,\text{Me} + 46.4686\,\text{Mp} + 7.74025\,\text{Ms} - 0.12519\,\text{MR} + 0.16611\,\text{PSA} + 6.85129\,\text{MlogP}$ | 88.680 | 5.226 | 0.0285 |
| Electronic | $\text{Log1/IC}_{50} = -142.03 - 890.13\,E_{\text{HOMO}} + 719.45\,E_{\text{LUMO}} - 148.63\,E_{\text{g}} - 7.75\,(\text{C4—N7}) + 3.39\,Q_{\text{N3}} - 16.023\,Q_{\text{S}} - 3.985\,A_0$ | 90.866 | 11.369 | 0.0014 |
| Equation (1) | $\text{Log1/IC}_{50} = -8.08 - 28.29\,\text{SIC2} - 137.187\,\Delta E_{\text{gap}} + 4.014\,\text{MlogP}$ | 98.380 | 53.118 | 0.0001 |
| Equation (2) | $\text{Log1/IC}_{50} = -41.01 - 0.36\,\text{G(N}\cdots\text{N)} - 221.41\,\Delta E_{\text{gap}} + 4.26\,\text{MlogP}$ | 91.340 | 42.210 | 0.0001 |
| Equation (3) | $\text{Log1/IC}_{50} = -28.04 - 11.17\,\Delta E_{\text{gap}} + 4.61\,\text{MlogP}$ | 87.670 | 46.2 | 0.0001 |

Table 9. QSAR equations and the corresponding statistical data ($n = 15$) computed for the second working set.

| Descriptors | QSAR equation | $r$ | $F$ statistical | $P$ |
|---|---|---|---|---|
| Geometrical | $\text{Log1/GI}_{50} = -114.38659 - 34.4193\,\text{J3D} - 0.03787\,\text{G1} + 0.01306\,\text{QZZm} + 0.0922\,\text{QXXe} + 0.1248\,\text{QYYe} + 0.14138\,\text{DISPp} - 0.1429\,\text{QYYp} + 19.06681\,\text{G(N}\cdots\text{N)} + 5.24214\,\text{G(N}\cdots\text{S)}$ | 89.506 | 4.738 | 0.050 |
| Topological | $\text{Log1/GI}_{50} = 128.78383 + 48.8586\,\text{DES} - 0.03761\,\text{BAC} + 0.02374\,\text{SPI} - 65.237\,\text{LPI}$ | 56.613 | 2.348 | 0.125 |
| Constitutional and empirical | $\text{Log1/GI}_{50} = -116.19183 + 0.00108\,\text{AMW} + 6.328\,\text{Sv} - 0.7956\,\text{Sp} - 1.07099\,\text{Ss} - 1.20451\,\text{Mv} + 81.25643\,\text{Mp} + 24.00783\,\text{Ms} - 1.14427\,\text{Mr} + 3.1036\,\text{MlogP}$ | 40.094 | 0.3718 | 0.906 |
| Electronic | $\text{Log1/GI}_{50} = -267.027 + 1.84411\,\text{ZPE} + 1695.822\,E_{\text{HOMO}} - 1701.94\,E_{\text{LUMO}} - 1762.51\,E_{\text{g}} + 211.3343\,(\text{C2—N3}) - 22.3393\,(\text{C4}'\text{—N7}) + 0.40552\,Q_{\text{S}} + 4.29206\,Q_{\text{N3}} + 2.89\,A_0$ | 75.907 | 1.670 | 0.290 |
| Final equation (4) | $\text{Log1/GI}_{50} = 72.50 + 4.79\,\text{ZPE} - 154.44\,E_{\text{LUMO}} - 2.28\,\text{MlogP}$ | 95.310 | 9.032 | 0.024 |

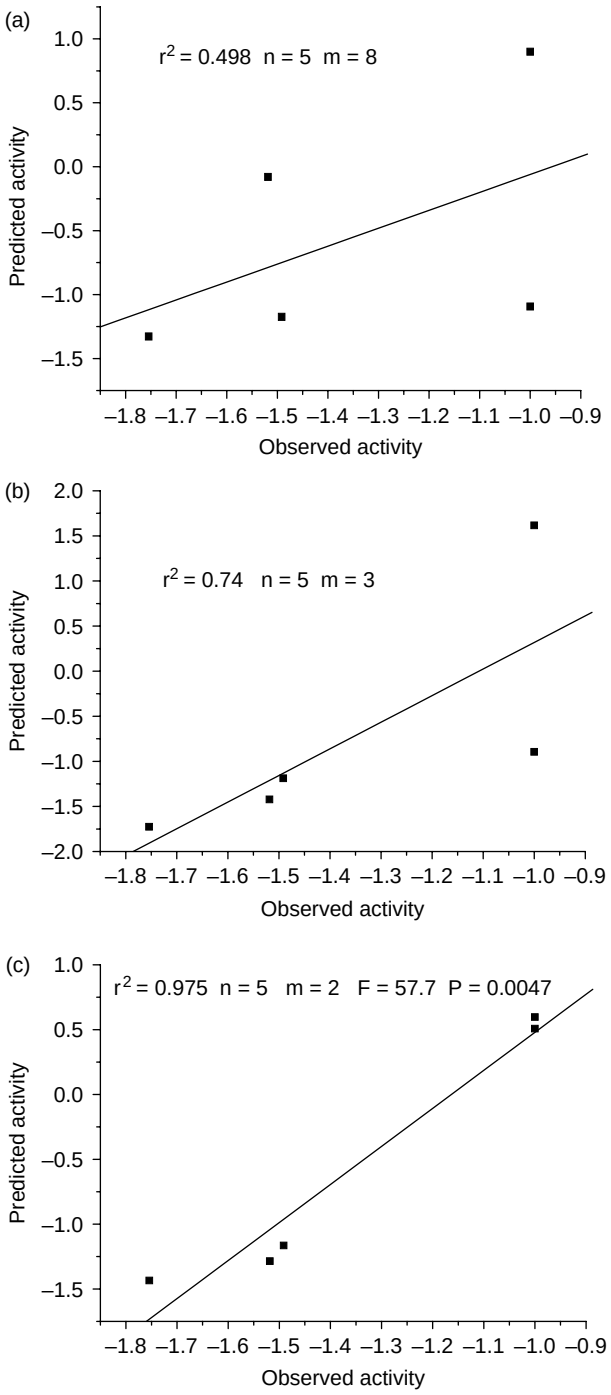



Figure 1. Correlation between observed and predicted activity using Equations (1)–(3) (*m*, number of descriptors; *n*, number of the validation set). (a) $m = 8$, (b) $m = 3$ and (c) $m = 2$.

descriptors should thus characterise the most statistically acceptable equation. The QSAR models (2) and (3) meet these requirements:

$$\mathrm{Log1/IC_{50}} = -41.01 - 0.36\,\mathrm{G(N\cdots N)} - 221.41\,\Delta E_{\mathrm{gap}} + 4.26\,\mathrm{MlogP}, \quad (2)$$

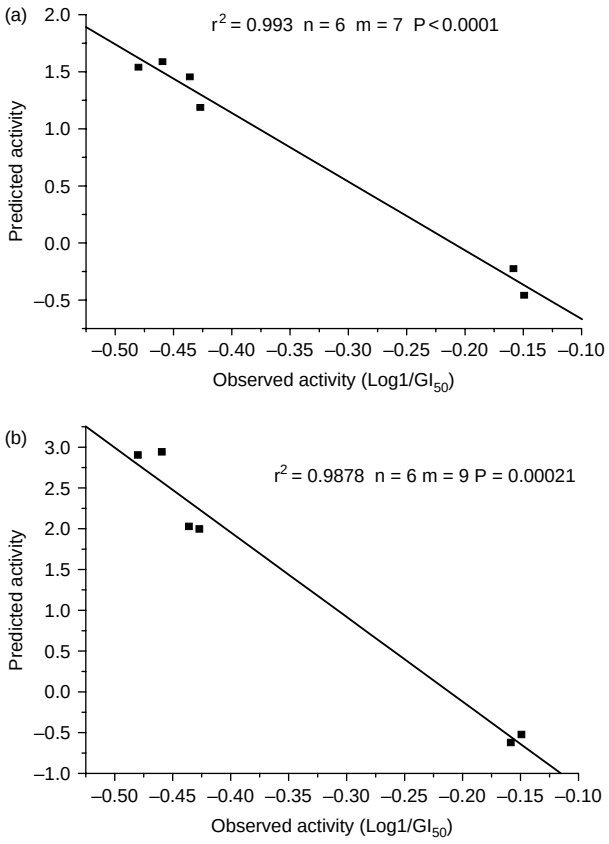



Figure 2. Correlation between observed and predicted activity using Equation (4) (*m*, number of descriptors; *n*, number of the validation set). (a) $m = 7$ and (b) $m = 9$.

$n = 16,\ r^2 = 0.913,\ q^2 = 0.843,\ F = 42.21,\ \mathrm{SD} = 0.717$ and $P < 0.0001$.

$$\mathrm{Log1/IC_{50}} = -28.045 - 111.18\,\Delta E_{\mathrm{gap}} + 4.61\,\mathrm{MlogP}, \quad (3)$$

$n = 16,\ r^2 = 0.876,\ q^2 = 0.818,\ F = 46.20,\ \mathrm{SD} = 0.820$ and $P < 0.0001$.

Similarly, for the second working set, the main QSAR equation and the best statistically acceptable one (Equation (4)) may be written as

$$\mathrm{Log1/GI_{50}} = 72.50 + 4.79\,\mathrm{ZPE} - 154.44\,E_{\mathrm{LUMO}} - 2.28\,\mathrm{MlogP}, \quad (4)$$

$n = 15,\ r^2 = 0.950,\ q^2 = 0.86,\ F = 9.03$ and $P = 0.024$.

### *3.4 Validation of the final QSAR equations*

The predictive power of the final QSAR equations (1)–(4) has been examined by computing the bioactivities of the

drugs in the validation set (cf. Tables 8 and 9). The correlation between the theoretically computed and experimentally observed activities for the validation set of APBT is presented in Figures 1 and 2. The best correlation is found using Equations (3) and (4) for the two activity scales, respectively.

## 4. Conclusions

The present work demonstrates that the relative affinity of APBT for the AhR can be predicted from molecular descriptors reflecting the electron-acceptor capability of the frontier orbitals and lateral substituents, as well as the hydrophobicity. It is considered that $\log P$ played an important role on the biological activity of the molecule that might be concerned with drugs distribution. That is, drugs should be solved in the water environment to penetrate into cells. Generally, $\log P$ value is inversely proportional to solubility in the water environment. It is of particular importance in drug design not only because it is correlated with the biological data but also because it encodes a wealth of structural information [16].

Finally, the QSARs presented herein have shown subtle preferences among the steric and the electron-acceptor parameters, most of which are sensitive to the chemicals selected for the study. Thus, we also plan to extend this QSAR analysis to include AhR binding data generated for other chemicals.



## References


[1] C. Hansch, *Quantitative approach to biochemical structure–activity relationships*, Acc. Chem. Res. 2(8) (1969), pp. 232–239.

[2] C. Hansch, J. McClaris, T. Klein, and R. Langridge, *A quantitative structure–activity relationship and molecular graphics study of carbonic anhydrase inhibitors,* Mol. Pharmacol. 27 (1985) pp. 493–498.

[3] D.-F. Shi, T.D. Bradshaw, S. Wrigley, C.J. McCall, P. Lelieveld, I. Fichtner, and M.F.G. Stevens, *Antitumor benzothiazoles. 3.1 Synthesis of 2-(4-aminophenyl)benzothiazoles and evaluation of their activities against breast cancer cell lines* in vitro *and* in vivo, J. Med. Chem. 39 (1996), pp. 3375–3384.

[4] E. Kashiyama, I. Hutchinson, M.-S. Chua, S.F. Stinson, L.R. Phillips, G. Kaur, E.A. Sausville, T.D. Bradshaw, A.D. Westwell, and M.F.G. Stevens, *Antitumor benzothiazoles. 8.1 Synthesis, metabolic formation, and biological properties of the C- and N-oxidation products of antitumor 2-(4-aminophenyl)benzothiazoles,* J. Med. Chem. 42 (1999), pp. 4172–4184.

[5] C.L. Waller and J.D. McKinney, *Three-dimensional quantitative structure–activity relationships of dioxins and dioxin-like compounds: Model validation and Ah receptor characterization*, Chem. Res. Toxicol. 8(6) (1995), pp. 847–858.

[6] J.D. McKinney and L.G. Pedersen, *Biological activity of polychlorinated biphenyls related to conformational structure,* J. Biochem. 240 (1986), pp. 621–622.

[7] R.G. Parr and Z. Zhou, *Absolute hardness: Unifying concept for identifying shells and subshells in nuclei, atoms, molecules, and metallic clusters,* Acc. Chem. Res. 26(5) (1993), pp. 256–258.

[8] R.G. Pearson, *The principle of maximum hardness,* Acc. Chem. Res. 26(5) (1993), pp. 250–255.

[9] W. Wang, J. Zhang, and B. Liu, *QSAR study of 125I-labeled 2-(4-aminophenyl)benzothiazole derivatives as imaging agents for β-amyloid in the brain with Alzheimer's disease,* J. Radioanal. Nucl. Chem. 266 (2005), pp. 107–111.

[10] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, Jr, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, N. Rega, P. Salvador, J.J. Dannenberg, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, *Gaussian 98*, Gaussian, Inc., Pittsburgh, PA, 1998.

[11] R. Todeschini and V. Consonni, *DRAGON Software Version 1.11*, Milano Chemometrics and QSAR Research Group, Department of Environmental Sciences, Milano, Italy, 2001.

[12] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley, Weinheim, 2000.

[13] M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley, New York, 2000.

[14] E.V. Kostantinova, *The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems*, J. Chem. Inf. Comput. Sci. 36 (1997), pp. 54–57.

[15] F.R. Burden, *A chemically intuitive molecular index based on the eigenvalues of a modified adjacency matrix*, Quant. Struct. Act. Relat. 16 (1997), pp. 309–314.

[16] G. Burden and C. Ruckr, *Counts of all walks as atomic and molecular descriptors,* J. Chem. Inf. Comput. Sci. 33 (1993), pp. 683–695.

[17] C. Hansch, in *Drug Design*, E.J. Ariens, ed., Academic Press, New York, 1971.

[18] M.-S. Chua, D.-F. Shi, W. Samantha, T.D. Bradshaw, I. Hutchinson, S.P. Nicholas, D.A. Barrett, A.S. Lesley, and M.F.G. Stevens, *Antitumor benzothiazoles. 7. Synthesis of 2-(4-acylaminophenyl)-benzothiazoles and investigations into the role of acetylation in the antitumor activities of the parent amines*, J. Med. Chem. 42 (1999), pp. 381–392.

[19] I. Hutchinson, M.-S. Chua, H.L. Browne, V. Trapani, T.D. Bradshaw, A.D. Westwell, and M.F.G. Stevens, *Antitumor benzothiazoles. 14.1 Synthesis and* in vitro *biological properties of fluorinated 2-(4-aminophenyl)benzothiazoles*, J. Med. Chem. 44 (2001), pp. 1446–1455.

## Appendix 1

Table A1. List of descriptors used in the present study.

| Abbreviation | Definition |
|---|---|
| 1. Constitutional descriptors | |
| AMW | Average molecular weight |
| Sv | Sum of atomic van der Waals volumes (scaled on carbon atom) |
| Sp | Sum of atomic polarisabilities (scaled on carbon atom) |
| Ss | Sum of Kier–Hall electrotopological states |
| Mv | Mean atomic van der Waals volume (scaled on carbon atom) |
| Me | Mean atomic Sanderson electronegativity (scaled on carbon atom) |
| Mp | Mean atomic polarisability (scaled on carbon atom) |
| Ms | Mean electrotopological state |
| nAT | Number of atoms |
| 2. Molecular properties | |
| Hy | Hydrophilic factor |
| MR | Ghose–Crippen molar refractivity |
| PSA | Fragment-based polar surface area |
| MlogP | Moriguchi octanol–water partition coefficient ($\log P$) |
| 3. Geometrical descriptors | |
| J3D | 3D-Balaban index |
| G1 | Gravitational index G1 |
| QZZm | Qzz COMMA2 value/weighted by atomic masses |
| QXXe | Qxx COMMA2 value/weighted by atomic Sanderson electronegativities |
| QYYe | Qyy COMMA2 value/weighted by atomic Sanderson electronegativities |
| DISPp | d COMMA2 value/weighted by atomic polarisabilities |
| QYYp | Qyy COMMA2 value/weighted by atomic polarisabilities |
| G(N···N) | Sum of geometrical distances between N···N |
| G(N···S) | Sum of geometrical distances between N···S |
| G2 | Gravitational index G2 (bond-restricted) |
| MAXDN | Maximal electrotopological negative variation |
| MAXDP | Maximal electrotopological positive variation |
| DISPm | dMMA2 value/weighted by atomic masses |
| QXXm | Qxx COMMA2 value/weighted by atomic masses |
| DISPv | d COMMA2 value/weighted by atomic van der Waals volumes |
| QXXv | Qxx COMMA2 value/weighted by atomic van der Waals volumes |
| SPAM | Average span R |
| 4. Topological descriptors | |
| DELS | Molecular electrotopological variation |
| TIE | E-state topological parameter |
| HNar | Narumi harmonic topological index |
| IC3 | Information content index (neighbourhood symmetry of third order) |
| SIC2 | Structural information content (neighbourhood symmetry of second order) |
| IC1 | Information content index (neighbourhood symmetry of first order) |
| AAC | Mean information index on atomic composition |
| SIC1 | Structural information content (neighbourhood symmetry of first order) |
| CIC2 | Complementary information content (neighbourhood symmetry of two order) |
| 5. Electronic descriptors | |
| $Q_A$ | Net atomic charge on atom A |
| $E_{HOMO}$ (au) | Energy of the highest occupied molecular orbital |
| $E_{LUMO}$ (au) | Energy of the lowest unoccupied molecular orbital |
| DM | Dipole moment (Debye) |
| $\Delta E_{gap}$ (au) | $E_{HOMO}$ and $E_{LUMO}$ orbital energy difference |
| ZPE (au) | Zero-point energy (au) |