

Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers

Indrani Mitra · Achintya Saha · Kunal Roy

Received: 1 November 2009 / Accepted: 11 January 2010 / Published online: 1 March 2010
© Springer-Verlag 2010

Abstract Predictive pharmacophore models have been developed for a series of arylamino-substituted benzo[b]thiophenes exhibiting free radical scavenging activity. 3D pharmacophore models were generated using a set of 20 training set compounds and subsequently validated by mapping 6 test set compounds using Discovery Studio 2.1 software. Further model validation was performed by randomizing the data using Fischer's validation technique at the 95% confidence level. The most predictive pharmacophore model developed using the conformers obtained from the *BEST* method showed a correlation coefficient (r) of 0.942 and consisted of three features: hydrogen bond donor, hydrogen bond acceptor and aromatic ring. Acceptable values of external validation parameters, like R^2_{pred} (0.853) and $r^2_{m(\text{test})}$ (0.844), also implied that the external predictivity of the model was significant. The development of further pharmacophore models using conformers obtained from the *FAST* method yielded a few models with good predictivity, with the best one ($r=0.904$) consisting of two features: hydrogen bond donor and hydrogen bond acceptor. Significant values of external validation param-

eters, R^2_{pred} (0.913) and $r^2_{m(\text{test})}$ (0.821), also reflect the high predictive ability of the model. Again, Fischer validation results implied that the models developed were robust enough and their good results were not based on mere chance. These validation approaches indicate the reliability of the predictive abilities of the 3D pharmacophore models developed here, which may thus be further utilized as a 3D query tool in the virtual screening of new chemical entities with potent antioxidant activities.

Keywords QSAR · Antioxidant · Pharmacophore · Benzo[b]thiophene

Introduction

Free radical formation occurs continuously in cells as a consequence of both enzymatic and nonenzymatic reactions [1]. A certain amount of these free radicals is required for normal physiological processes, and this amount is kept to a minimum by a number of scavenging mechanisms that detoxify the free radicals. However, free radical production is accelerated by various stress conditions, including inflammation, infections and environmental stresses (toxic pollutants, smoke, etc.) [2]. Several xenobiotics (drugs and chemicals foreign to the human body) have also been implicated in the process of accelerating free radical production [3]. When free radical production exceeds clearance, oxidative damage results, and this condition is responsible for a series of deadly diseases, such as atherosclerosis, Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, cancer, etc. Large-scale production of these free radicals is now believed to be the reason for most human chronic diseases, including AIDS, chronic fatigue syndrome, psoriasis and asthma.

I. Mitra · K. Roy (✉)
Drug Theoretics and Cheminformatics Lab,
Division of Medicinal and Pharmaceutical Chemistry,
Department of Pharmaceutical Technology, Jadavpur University,
Kolkata 700 032, India
e-mail: kunalroy_in@yahoo.com
URL: <http://sites.google.com/site/kunalroyindia/>

A. Saha
Department of Chemical Technology,
University College of Science and Technology,
University of Calcutta,
92, A. P. C. Road,
Kolkata 700 009, India

Antioxidants serve as the primary method of controlling systemic free radical attack. Antioxidants are chemical entities that break the free radical chain reactions through being oxidized and chelating the metal ions that catalyze these free radical chain reactions [4]. Antioxidant activity is primarily based on three different molecular mechanisms: (a) hydrogen atom transfer (HAT); (b) single-electron transfer followed by proton transfer (SET-PT); and (c) sequential proton loss electron transfer (SPLET) [5–7]. Several ecological, case–control and cohort studies indicate that diets rich in plant food serve as a surplus source of effective antioxidants. The substantial role played by antioxidants in detoxifying free radicals, the cause of a series of deadly diseases, has led to a need to design and synthesize molecules with potent antioxidant activities.

The process involving the search and design of efficient antioxidant molecules can be simplified greatly by using quantitative structure–activity relationship (QSAR) methodology. QSAR is a computational tool that correlates the biological activities of a series of molecules with several numerical parameters called descriptors using various statistical methods [8, 9]. This technique is now used extensively as a reliable tool for searching for active antioxidant molecules. Using the QSAR technique, the antioxidant activities of various classes of chemicals have been modeled by several authors. Ray et al. [10] reported on the QSARs of the antiradical and antioxidant activities of flavonoids using electrotopological state (E-state) atom parameters. In another study [11], they also developed QSAR models for the lipid peroxidation inhibition potentials of flavonoids using structural and topological parameters. Roy et al. [12] performed molecular shape analysis to predict the antioxidant and squalene synthase inhibitory activities of aromatic tetrahydro-1,4-oxazine derivatives using a genetic algorithm as a chemometric tool. The antioxidant activities of wine polyphenols were modeled by Rastija et al. [13] using a QSAR technique, where the descriptors were calculated from 2D and 3D representations of the molecules. Gupta et al. [14] reported a QSAR analysis of phenolic antioxidants using MOLMAP (molecular maps of atom-level properties) descriptors for local properties of the molecules. However, very few 3D pharmacophore models have been developed to explore the essential features that contribute to the potency of antioxidant molecules. Advances in QSAR models of antioxidants have recently been reviewed [15].

The main aim of this study is to construct pharmacophore models based on key chemical features of a particular class of compounds with antioxidant activities (*vide infra*) covering a satisfactorily wide range of magnitude. A pharmacophore refers to an ensemble of steric and electronic features that is required to ensure optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response [16]. A

database of diverse chemical compounds can then be searched for more molecules with the ensemble corresponding to the pharmacophore (i.e., the same features separated by a similar distance to that seen in the pharmacophore). Here we have developed 3D pharmacophore models for predicting the radical scavenging activities of arylamino-substituted benzo[b]thiophenes [17] using the Discovery Studio 2.1 software [18].

The reducing properties of diarylamines afford them potential antioxidant activity, especially as radical scavengers [19]. These secondary amines react with the free radicals according to the HAT mechanism. This involves hydrogen transfer of the N–H bond to peroxy radicals, which leads to aminyl radicals ($RR'N_2$) that react again with RO_2 to give nitroxide radicals ($RR'NO\cdot$) in the second step [20]. In a study, Ferreira et al. [21] reported that the antioxidant properties of several substituted diarylamines in the benzo[b]thiophene series, aminated either on the benzene moiety or on the thiophene ring, seem to be dependent on the type and position of the substituents. Considering the efficient radical scavenging activity of diarylamines, 3D pharmacophore models for identifying such molecules with increased antioxidant activities have been developed in the present work.

Methods and materials

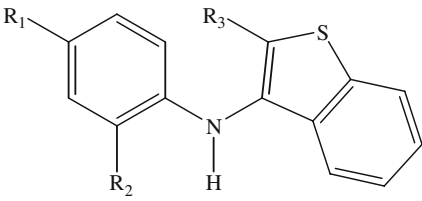
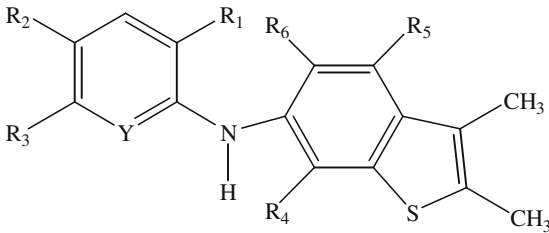
Dataset

The model dataset used in the present work was reported by Abreu et al. [17]. This dataset comprises of 26 arylamino-substituted benzo[b]thiophenes whose radical scavenging activities (RSAs) were screened using the 2,2-diphenyl-1-picrylhydrazil (DPPH) radical inhibition assay. The 50% inhibitory concentrations of these molecules were expressed in millimoles (mM) to develop the 3D pharmacophore models. The molecular structures of these compounds as well as their observed activity data are summarized in Table 1.

Development of the 3D pharmacophore model

The 3D pharmacophore model is a ligand-based approach that provides a unique tool for drug design [22]. A 3D pharmacophore is a collection of chemical features in space that are required for a desired biological activity. These include hydrophobic groups, charged/ionizable groups, hydrogen bond donors/acceptors, and others that are properly assembled in 3D space to reflect structural requirements. An interesting application of pharmacophore-based approaches is that the experimentally determined activities of a set of compounds can be used to drive the generation of pharmacophores. The advantage of this approach is that the pharmacophores, once validated, can be used to quantita-

Table 1 Structures and antioxidant activities [IC₅₀(mM)] of 26 benzo[b]thiophene derivatives belonging to three different classes

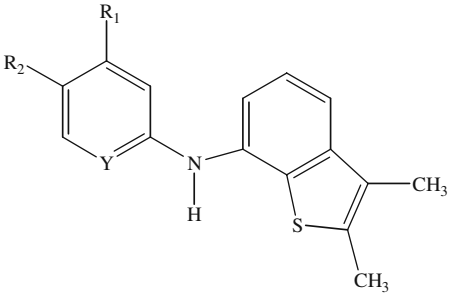
 <p style="text-align: center;">Class A: 3-(arylamino)benzo[b]thiophenes</p>										
Compound Nos.	R ₁	R ₂	R ₃	Observed activity [17]	Estimated activity ^b	Estimated activity ^c				
1	OCH ₃	OCH ₃	H	87.096	103.651	141.248				
2^a	OCH ₃	OCH ₃	COOH	245.471	73.462	114.402				
3	OCH ₃	OCH ₃	COOCH ₂ CH ₃	169.824	74.237	105.78				
4	OH	OH	COOCH ₂ CH ₃	144.544	82.664	107.22				
5	H	OH	COOCH ₂ CH ₃	123.027	124.336	220.601				
6	H	F	H	154.882	271.752	594.399				
 <p style="text-align: center;">Class B: 6-(heteroaryl amino) benzo[b]thiophenes</p>										
Compound Nos.	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	Y	Observed activity [17]	Estimated activity ^b	Estimated activity ^c
7	OCH ₃	OCH ₃	H	H	H	CH ₃	C	85.114	107.76	120.461
8	H	OCH ₃	OCH ₃	H	H	CH ₃	C	50.119	107.197	110.688
9^a	H	OCH ₃	H	CH ₃	CH ₃	H	C	128.825	102.869	123.432
10	H	OCH ₃	H	H	H	CH ₃	C	72.444	108.763	117.006
11^a	H	H	OCH ₃	H	H	CH ₃	C	416.869	359.663	328.371
12	H	CHO	H	H	H	CH ₃	C	4073.803	1492.6	1168.44
13^a	H	CN	H	CH ₃	CH ₃	H	C	5128.614	4978.63	2054.11
14	Br	OCH ₃	OCH ₃	H	H	CH ₃	C	154.882	108.969	97.35
15	Br	OCH ₃	H	CH ₃	CH ₃	H	C	104.713	103.895	145.796
16	Br	OCH ₃	H	H	H	CH ₃	C	204.174	114.339	130.3
17	Br	H	H	H	H	CH ₃	C	891.251	2225.28	1393.57
18	Br	H	H	CH ₃	CH ₃	H	C	3715.352	2734.41	1497.63

tively predict the activities of new compounds. Therefore, they constitute a powerful and fast tool for estimating the biological activities of new potential ligands in 3D databases of compounds [23–26].

Training set selection

The selection of the training set is a crucial aspect of the process of pharmacophore hypothesis generation. A molecule

Table 1 (continued)

19	I	H	H	H	H	CH ₃	C	1230.269	2216.03	1388.01
20	H	H	H	CH ₃	CH ₃	H	N	7943.282	2629.89	2109.13
 <p>Class C: 7-(heteroarylamino) benzo[b]thiophenes</p>										
Compound Nos.	R ₁	R ₂	Y	Z	Observed activity [17]	Estimated activity ^b	Estimated activity ^c			
21	OCH ₃	OCH ₃	C	C	54.954	83.457	106.603			
22	OCH ₃	H	C	C	1548.817	1475.24	331.273			
23	H	OCH ₃	C	C	50.119	85.346	107.478			
24^a	CN	H	C	C	14454.398	4595.18	7663.03			
25^a	H	H	C	N	3162.278	993.404	1468.07			
26	H	H	N	C	5495.409	5226.54	5128.61			

^aTest set compounds^bActivity estimated from hypothesis 1 developed with conformers obtained from the *BEST* method of conformer generation^cActivity estimated from hypothesis 3 developed with conformers obtained from the *FAST* method of conformer generation

that is structurally very similar to the training set molecules will be predicted well because the model has captured features that are common to the training set molecules and is able to find them in the new molecule [27]. When a compound is highly dissimilar to all compounds in the modeling set, reliable prediction of its activity is unlikely. Thus, selection of the training set was done using the cluster analysis method [28]. This method divides different objects into groups in such a way that the degree of association between two objects is maximized if they possess the same groups (otherwise it is minimized). Hierarchical and nonhierarchical are two types of clustering techniques. The hierarchical technique forms relationships within clusters in subsequent steps. In the nonhierarchical method, compounds are first classified into a defined number of clusters based on nearest neighbor distributions in chemical space. In our study, *k*-means clustering (one of the best-known nonhierarchical clustering techniques) [29] was applied to a standardized descriptor matrix comprising spatial, physicochemical and structural descriptors that were calculated using the Cerius 2 software [30]. Subsequently, the whole dataset was divided into a training set of 20 compounds and a test set of 6 compounds by selecting 25% of compounds from each of

the three clusters as the members of the test set. Table 2 shows the compounds selected as the members of the test set, together with their cluster membership.

Diverse conformation generation

Before starting the pharmacophore generation process, conformational analysis of the molecules was performed using the poling algorithm [31]. The poling algorithm eliminates much of the redundancy in conformation generation and improves the coverage of conformational space. The number of conformers generated for each

Table 2 Selection of test set members using *k*-means clustering of the standardized descriptor matrix

Cluster no.	Number of test set compounds in each cluster	Test set compound serial nos. in each cluster
1	3	13, 24, 25
2	1	2
3	2	9, 11

compound was limited to a maximum of 255 with an energy range of 20 kcal mol^{-1} (i.e., the difference in energy values among different conformers of a particular compound was $<20 \text{ kcal mol}^{-1}$). In the present work, conformers were generated using both the *BEST* and the *FAST* methods of conformer generation. The *BEST* method provides complete and improved coverage of conformational space by performing a rigorous energy minimization and optimizing the conformations in both torsional and Cartesian space using the poling algorithm, while *FAST* generation only searches for conformations in torsion space and so takes less time [18]. The *BEST* routine performs three steps for conformation generation: conjugate-gradient minimization in torsion space followed by conjugate-gradient minimization in Cartesian space and quasi-Newton minimization in Cartesian space [18]. The *FAST* conformation generation method uses one of the three algorithms, depending on the size of the molecule [18]. The conformational space of small molecules is generated using an efficient systematic search. The conformational space is composed of discretized rotations about the rotatable bonds. The conformational space is systematically searched and conformations that have excessive van der Waals clashes are removed. A random search method with poling is used to generate conformations for medium-sized molecules. In this technique, the molecule is split into pieces and a systematic search is performed on each piece, before the pieces are finally randomly reconnected. If the number of rotatable bonds is greater than 30, only one conformation is created for every possible stereocenter.

Generation of the 3D pharmacophore

For the present work, pharmacophore models were developed using the *HypoGen* module implemented in Discovery Studio 2.1 [18] with the conformers generated for the molecules in the training set ($n_{\text{training}}=20$). Predictive pharmacophores were generated in three phases: a constructive phase, a subtractive phase, and an optimization phase. In the constructive phase, pharmacophores are generated that are common among the active molecules of the training set. *HypoGen* identifies all allowable pharmacophores consisting of up to five features among the two most active compounds and investigates the remaining active compounds in the list. The subtractive phase deals with the pharmacophores that were created in the constructive phase, and the program removes pharmacophores from the data structure that are not likely to be useful. Finally, the optimization is performed using the well-known simulated annealing algorithm. The algorithm applies small perturbations to the pharmacophores created in the constructive and subtractive phases in an attempt to improve the score. The algorithm accepts all improvements and some detrimental

steps based on a probability function. The highest-scoring unique pharmacophores are then exported.

Ten hypotheses were generated for each of the two sets of conformers (*BEST* and *FAST*) used. *HypoGen* allows a maximum of five features during pharmacophore generation. After eliminating the features that do not map the training set molecules, the following 5 features were selected for subsequent pharmacophore generation: hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), hydrophobic aliphatic (HY-ALI), hydrophobic aromatic (HY-ARO) and ring aromatic (RA). The hypotheses generated were analyzed in terms of their correlation coefficients and the cost function values. The *HypoGen* module performs a fixed-cost calculation that represents the simple model that fits all of the data and a null-cost calculation that assumes that there is no relationship in the dataset and that the experimental activities are normally distributed about their average value. A small range of total hypothesis costs obtained for each of the hypotheses indicates that the corresponding hypothesis is homogeneous and that the training set selected for the purpose of pharmacophore generation is adequate. Again, values of total cost close to those of fixed cost indicate that the hypotheses generated are statistically robust.

Pharmacophore model validation

Validation of a quantitative model was performed in order to determine whether the developed model is able to identify active structures and to forecast their activities precisely. Validation of the obtained pharmacophore models was done using two procedures: Fischer's validation (as available in the *HypoGen* module), and external validation using the test set prediction method.

Fischer's validation

The statistical significance of the structure–activity correlation was estimated using Fischer's randomization test [18]. This was done by scrambling the activity data of the training set molecules and assigning them new values, and then generating pharmacophore hypotheses using the same features and parameters as those used to develop the original pharmacophore hypothesis. The number of spreadsheets obtained using the randomization test depends on the level of statistical significance desired. At the 95% confidence level, 19 spreadsheets are generated. The original hypothesis is considered to be generated by mere chance if the randomized dataset results in the generation of a pharmacophore with better correlation compared to the original one. Since no guideline is given regarding how large the difference between the correlation coefficients of

the hypotheses obtained for the original and randomized datasets should be, we used the parameter R_p^2 [32–34] (threshold value=0.5). The parameter R_p^2 (given by Eq. 1) compares the values of the squared correlation coefficient (R^2) of the original dataset with the squared average correlation coefficient (R_r^2) obtained from the randomized dataset, and penalizes the model R^2 for small differences between the values of R^2 and R_r^2 .

$$R_p^2 = R^2 * \sqrt{(R^2 - R_r^2)} \quad (1)$$

Prediction with test set molecules

The purpose of the pharmacophore hypothesis generation was not only to predict the activities of the training set compounds, but also to predict the activities of external molecules. In order to verify whether the pharmacophore was able to accurately predict the activities of test set molecules (i.e., to produce values in agreement with the experimentally determined values), the activities of the test set molecules were estimated using the developed pharmacophore model. Note that log-transformed activity values were used to calculate metrics of external predictivity. The conformers generated for the test set molecules ($n_{test}=6$) using both the *FAST* and the *BEST* methods were selected and mapped using the corresponding pharmacophore models developed with the training set compounds. External validation provides the ultimate proof of the true predictability of the model, and the predictive capacity of the model is judged based on the predictive R^2 values (R_{pred}^2 with a threshold value of 0.5) [35] calculated according to Eq. 2:

$$R_{pred}^2 = 1 - \frac{\sum (Y_{observed} - Y_{predicted})^2}{\sum (Y_{observed} - \bar{Y}_{training})^2} \quad (2)$$

In the above equation, $Y_{observed}$ and $Y_{predicted}$ are the observed and predicted activity data (log scale) for the test set compounds, while $\bar{Y}_{training}$ denotes the mean observed activity (log scale) of the training set molecules. The above equation shows that the value of R_{pred}^2 depends on the mean observed activity of the training set compounds. Thus, high values of this parameter may be obtained for compounds with a wide range of activity data, but this may not indicate that the predicted activity values are very close to those observed. Though a good overall correlation is maintained, there may be a considerable numerical difference between the two values. To better indicate the predictive ability of the model, modified r^2 [$r_{m(test)}^2$] [33, 36, 37] values were calculated according to Eq. 3:

$$r_{m(test)}^2 = r^2 * \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (3)$$

Here, r^2 and r_0^2 denote the squared correlation coefficient values of the observed and predicted activity values (log scale) of the test set compounds with and without the intercept, respectively. In the case of an external prediction, the r^2 values will be close to that of r_0^2 , resulting in an increase in the value of the $r_{m(test)}^2$ parameter (threshold value=0.5).

Results and discussion

Pharmacophore development with conformers generated using the *BEST* method of conformer search

A set of 9 pharmacophore hypotheses were generated with the conformers generated from the *BEST* method using the 20 training set compounds listed in Table 1. The results for all of the hypotheses together with the pharmacophore features, cost functions and correlation values are listed in Table 3. The total costs, expressed in bits, of the 9 best hypotheses varied from 107.248 to 112.435, and such a small range (covering only 4 bits) suggested that the generated hypotheses were homogeneous. For each of the 9 hypotheses, the total cost was close to the fixed cost value (104.206), indicating good hypotheses. Moreover, the values of the coefficients of correlation between the observed and calculated activities occurred within the range 0.826–0.940, indicating that the developed hypotheses were statistically significant. Among the 9 hypotheses, 6 had three features in common (HBD, HBA and RA), 2 had two features in common (HBA and RA), while the remaining hypothesis had a set of three features (HBA, HBD and HY-ARO) that differed from the other 8 hypotheses. The best hypothesis was selected based on the RMSD (root mean square difference) value indicating the quality of training set prediction as well as the correlation coefficient between the observed and estimated activity values. In this case, hypothesis 1, which had an RMSD value of 0.518 and a correlation coefficient of 0.942, exhibited good predictivity as well as correlation. Thus, hypothesis 1 was chosen as the best-ranking pharmacophore.

Hypothesis 1 possessed three features: HBD, HBA and RA. The distance between the hydrogen bond acceptor and aromatic ring features was 6.7636 Å, while that between the hydrogen bond donor and acceptor was 5.3556 Å. Moreover, the angle between the aromatic ring, hydrogen bond donor and hydrogen bond acceptor features was found to be 100.983 Å for hypothesis 1 (Fig. 1). Mapping this pharmacophore (Fig. 2) onto the most active compound (compound no. 8) revealed that the amino nitrogen (–NHRR') behaves as a hydrogen bond donor, while the methoxy oxygen (–OCH₃) bears the hydrogen bond acceptor feature. Since the amino nitrogen is secondary

Table 3 Results for the 9 pharmacophore hypotheses generated using conformers developed from the *BEST* method of conformer search

Hypothesis no.	Total cost	Error cost	RMS	Correlation (<i>R</i>)	Features	R^2_{pred}	$r^2_{m(\text{test})}$
1	108.221	69.954	0.518	0.942	HBA, HBD, RA	0.853	0.844
2	109.129	70.513	0.569	0.928	HBA, HBD, RA	0.763	0.717
3	109.562	71.313	0.636	0.914	HBA, HBD, RA	0.650	0.567
4	110.676	72.214	0.703	0.888	HBA, HBD, RA	0.462	0.418
5	111.547	73.221	0.771	0.863	HBA, HBD, RA	0.639	0.664
6	112.241	73.053	0.760	0.878	HBA, HBD, RA	0.398	0.441
7	112.635	73.981	0.819	0.853	HBA, RA	0.517	0.551
8	112.688	73.853	0.811	0.861	HBA, RA	0.531	0.535
9	112.84	74.195	0.832	0.846	HBA, HBD, HY-ARO	0.716	0.610

Fixed cost: 104.909

Null cost: 90.6016

and surrounded by bulky aromatic groups, it efficiently contributes to hydrogen bond formation by donating its partially positively charged hydrogen atom to the neighboring electron-rich free radical. This indicates that a hydrogen bond donor group positioned within the inner core of the substituted benzothiophene was crucial to the increased antioxidant activities of these compounds. This observation was in agreement with the results of the QSAR models developed by Abreu et al. [17] using RDF (radial distribution function) descriptors. On the other hand, the positive inductive (electron-releasing) effect of the methyl ($-\text{CH}_3$) group increases the electron density of the methoxy oxygen, and hence the methoxy oxygen behaves as a

nucleophilic center. It then contributes to the antioxidant mechanism of action by transferring a single electron and then deprotonating [5]. This suggests that electronegative atoms present at specific positions on the substituted benzothiophene moiety contribute to the increased antioxidant activities of these compounds, which again satisfies the observations of Abreu et al. [17]. The third feature indicates the importance of ring aromaticity at a distance of 6.7636 Å from the HBA feature to the antioxidant activities of these molecules. Since the aromatic ring is hydrophobic in nature, an area of transient electron deficiency develops [38], which in turn may interact with another transient electron-rich area on a nearby free radical. This implies the significance of the benzothiophene moiety in modulating the antioxidant activities of these compounds.

The pharmacophore obtained in hypothesis 1 could effectively map onto the most active compounds of the model dataset. Moreover, the activities estimated from hypothesis 1 for compounds **1**, **7**, **8**, **14**, **15**, **21** and **23** were remarkably close to the corresponding observed activity data. Since

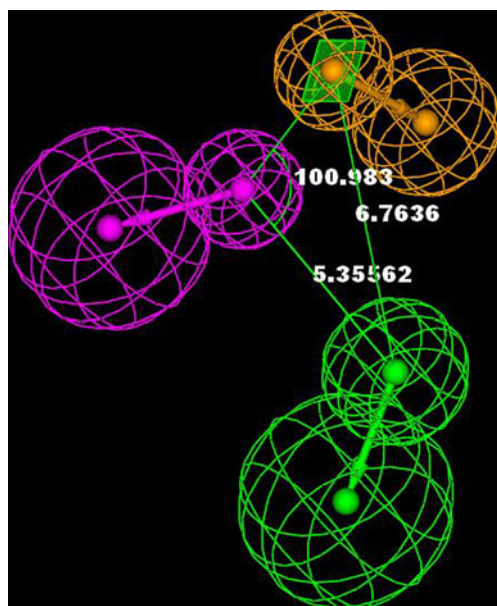


Fig. 1 Pharmacophore obtained from hypothesis 1 using the training set conformers developed from the *BEST* method of conformer generation (Shown are ring aromatic sphere (orange), hydrogen bond donor (magenta) and hydrogen bond acceptor (green) features with vectors in the direction of putative hydrogen bonds)

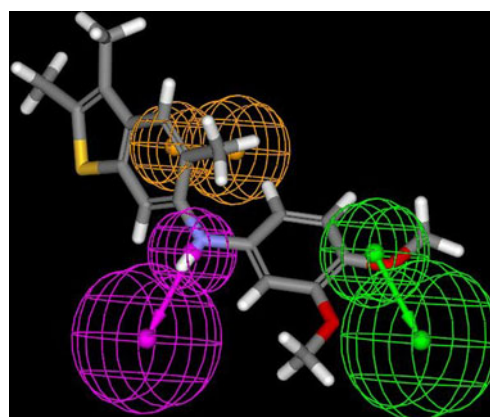


Fig. 2 Mapping of the pharmacophore obtained from hypothesis 1 (developed with conformers from the *BEST* method) onto the most active molecule (**8**)

these compounds, which had methoxy groups substituted at different positions on the aryl moiety, had relatively high activities, the electronegative hydrogen bond acceptor group appears to play an important role in interactions of these molecules with nocive free radicals. On the other hand, compounds **6**, **12** and **20**, which lack the aforementioned substitutions, were poorly mapped, and their estimated activity data showed large deviations from those observed.

Validation of the developed pharmacophore

To check the fitness of the developed pharmacophore model, Fischer's randomization test was performed at the 95% confidence level. The experimental activities of the compounds in the training set were permuted 19 times, and spreadsheets (Table 4) were obtained with the randomized activity data. The data indicated that none of the values generated after randomization produced hypotheses that exhibited predictive powers similar to that of hypothesis 1. The average correlation coefficient for the 19 trials was only 0.525, and none of these coefficients were greater than that of hypothesis 1. Moreover, in order to determine whether the difference between the randomized and average nonrandomized correlation coefficients was significant or not, the value of R_p^2 [32–34] was calculated. The parameter

R_p^2 penalizes the model R^2 for small differences between R^2 and R_r^2 . For hypothesis 1, the calculated value of R_p^2 was 0.693, which is well above the recommended value of 0.5. Thus, it can be inferred that hypothesis 1 is robust enough and its predictive power does not derive from mere chance.

The pharmacophore obtained from hypothesis 1 was further validated by assessing the predictive ability of the pharmacophore in relation to the test set compounds. The validity of the pharmacophore model was ascertained by applying the model to the test set to find out how accurately the model predicts the activities of the test set compounds. We validated the selected pharmacophore using the 6 test set compounds. The highly active compounds were efficiently mapped and were well predicted to have high activities, while the less active compounds were poorly mapped and were consequently predicted to have low activities. The correlation between the observed and predicted activities of the test set compounds (0.853), as given by the value of R_{pred}^2 , was also found to be much higher than the acceptable value of 0.5. Similar analyses performed with the remaining 8 hypotheses yielded results with much lower values of R_{pred}^2 (Table 3). For a better determination of the predictive abilities of the models, the values of $r_{m(test)}^2$ were also calculated (Table 3). The value of this parameter determines whether the predicted activity values are close to the corresponding observed ones, since a high value of R_{pred}^2 may not always indicate a low residual between the observed and predicted activity data. Among all of the hypotheses developed, the largest value of $r_{m(test)}^2$ (0.844) was observed for hypothesis 1, indicating that this model has acceptable predictive potential. Thus, the results suggest that the selected pharmacophore can correctly predict the activities of new compounds.

Pharmacophore generation with conformers obtained from the FAST method of conformer search

The conformers generated with the training set compounds using the FAST method of conformational search were also used for pharmacophore generation. As done for the BEST method, 9 hypotheses (Table 5) were obtained with 3 bearing the features HBD, HBA and RA, 2 possessing HBD and HBA, 3 with HBA and RA, while the remaining hypothesis had the features HBD, HBA and HY-ARO. Among these 9 hypotheses, hypothesis 1 showed a maximum correlation of 0.940 and a minimum root mean square deviation (RMSD) of 0.521. However, validation of hypothesis 1 using Fischer's randomization and external validation techniques yielded poor results. Fischer's validation performed at the 95% confidence level showed that the average correlation coefficient obtained from the 19 scrambled data series was much lower for hypothesis 3 than for hypothesis 1. This indicated that hypothesis 3 was more

Table 4 Results from Fischer's randomization test (BEST method of conformer search)

Validation	Correlation coefficient
random1	0.734
random2	0.742
random3	0.563
random4	0.322
random5	0.712
random6	0.353
random7	-0.061
random8	0.873
random9	0.604
random10	0.536
random11	0.277
random12	0.140
random13	0.515
random14	0.603
random15	0.730
random16	0.432
random17	0.553
random18	0.727
random19	0.630

Average (R_r): 0.525

Correlation coefficient of the nonrandomized model (R): 0.942

R_p^2 : 0.693

Table 5 Results for the 9 pharmacophore hypotheses generated using conformers developed from the *FAST* method of conformer search

Hypothesis no.	Total cost	Error cost	RMS	Correlation (<i>R</i>)	Features	R^2_{pred}	$r^2_{m(\text{test})}$
1	107.248	69.989	0.521	0.940	HBA, HBD, RA	0.360	0.165
2	108.605	71.177	0.625	0.922	HBA, HBD, HY-ARO	0.487	0.374
3	109.633	72.052	0.691	0.904	HBA, HBD	0.913	0.821
4	109.765	71.915	0.681	0.913	HBA, HBD	0.911	0.804
5	110.444	72.548	0.726	0.880	HBA, HBD, RA	0.396	0.390
6	111.707	73.646	0.798	0.859	HBA, RA	0.555	0.556
7	112.181	74.204	0.833	0.852	HBA, RA	0.285	0.371
8	112.292	74.738	0.864	0.826	HBA, HBD, HY-ARO	0.868	0.710
9	112.435	73.824	0.809	0.863	HBA, RA	0.585	0.563

Fixed cost: 104.206

Null cost: 90.6016

statistically more robust than hypothesis 1. In addition to the randomization results, the greatest correlation was obtained between the observed and predicted activity data when the 6 test set compounds were mapped using hypothesis 3. This suggested that hypothesis 3 could efficiently map the test set compounds, and hence this hypothesis can be satisfactorily employed to predict the activities of new compounds.

The total cost (107.248) for hypothesis 3 was close to the fixed cost (104.206), which is required for a good hypothesis. Besides this, hypothesis 3 also possessed a relatively low RMSD (0.691) and a statistically significant value of the correlation coefficient (0.904). The selected pharmacophore (Fig. 3) possessed two features, namely HBD and HBA, which maintained a distance of 5.491 Å from each other. The pharmacophore was analyzed by mapping the most active compound (compound no. 8) (Fig. 4), which yielded results similar to those obtained

using the previous set of conformers (obtained with the *BEST* method). Since the amino nitrogen (HBD) and the methoxy oxygen (HBA) readily form hydrogen bonds, interactions of this series of substituted benzothiophenes with nearby free radicals are facilitated. Thus, compounds **1**, **7**, **8**, **10**, **21** and **23** that have a methoxy substituent at various positions were mapped efficiently by the selected pharmacophore. The activities estimated for these compounds using the selected pharmacophore were also quite close to the corresponding observed data. Compounds **18**, **19**, **20** and **26**, which were reported to have low activities, lacked the necessary structural attributes and hence were poorly mapped. These findings once again imply the need to have hydrogen bond donor and acceptor groups separated by a specific topological distance in order to attain arylamino-substituted benzothiophene derivatives with optimal antioxidant activities.

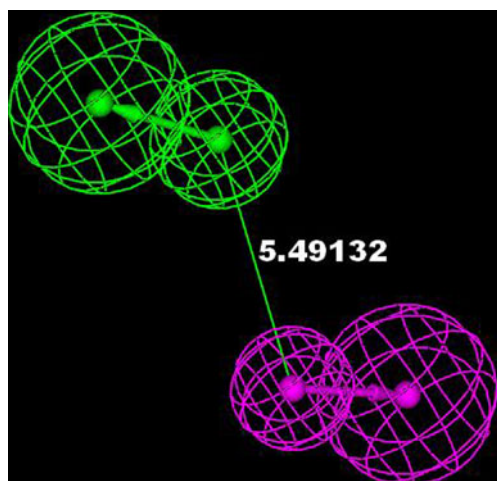


Fig. 3 Pharmacophore obtained from hypothesis 3 using the training set conformers developed via the *FAST* method of conformer generation (Shown are hydrogen bond donor (magenta) and hydrogen bond acceptor (green) features with vectors in the direction of putative hydrogen bonds)

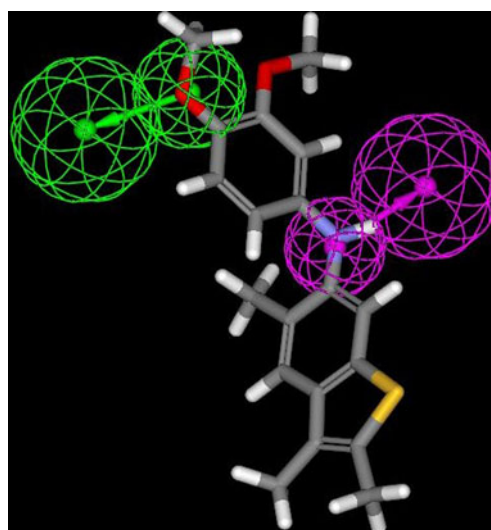


Fig. 4 Mapping of the pharmacophore obtained from hypothesis 3 (developed with conformers from the *FAST* method) onto the most active molecule (**8**)

Table 6 Results from Fischer's randomization test (*FAST* method of conformer search)

Validation	Correlation coefficient
random1	0.701
random2	0.446
random3	0.486
random4	0.258
random5	0.634
random6	0.282
random7	0.179
random8	0.786
random9	0.320
random10	0.700
random11	0.404
random12	0.052
random13	0.496
random14	0.558
random15	0.656
random16	0.487
random17	0.406
random18	0.768
random19	0.484

Average (R_r): 0.479Correlation coefficient of nonrandomized model (R): 0.904 R_p^2 : 0.627

Validation of the developed pharmacophore

The quality of the pharmacophore was assessed using Fischer's validation technique at the 95% confidence level (Table 6). The pharmacophore models obtained using the randomized datasets yielded results with inferior correlation coefficient values compared to those obtained for hypothesis 3. The significance of the difference between the correlation coefficient values of the randomized and nonrandomized models was judged using the R_p^2 parameter. The value of R_p^2 obtained for hypothesis 3 was 0.627, which was higher than the stipulated value of 0.5. The results provided confidence in the robustness of the pharmacophore generated from the training set molecules. Moreover, the external predictive power of the pharmacophore was determined by mapping 6

test set compounds. The values of R_{pred}^2 thus calculated (Table 5) revealed that hypothesis 3 could efficiently map all 6 compounds of the test set and discriminate the active molecules from the inactive ones. The low activity data of compound **24** derives from the absence of a highly electronegative hydrogen bond acceptor group around the aryl moiety. Values of the $r_{m(test)}^2$ parameter (Table 5) were also calculated to determine the extent of variation between the observed and corresponding predicted activity data. The largest values of R_{pred}^2 (0.913) and $r_{m(test)}^2$ (0.821) were obtained for hypothesis 3, implying that the observed activity range was predicted most efficiently by the selected pharmacophore among all of those developed.

Comparison with the previously published model

Abreu et al. [17] reported a QSAR model that was developed by the PLS technique with RDF (radial distribution function) molecular descriptors and 2D autocorrelation descriptors using this series of arylamino-substituted benzo[b]thiophenes. The RDF descriptors and 2D-autocorrelation descriptors signify that the presence of polarizable and electronegative atoms enhances the radical scavenging activities of these molecules. In the present study, 3D pharmacophore models have been developed that present a qualitative picture of the geometries of the active molecules by identifying the requisite features and spatial arrangement needed to produce benzo[b]thiophene derivatives with effective antioxidant activities. Both of the significant pharmacophores reported in the present work indicate the importance of the presence of electronegative as well as electropositive centers to the potential antioxidant activities of these molecules. A detailed comparison of the statistical quality of each of these pharmacophore models with the model developed by Abreu et al. [17] is shown in Table 7.

Overview

This report describes the development of a 3D pharmacophore with quantitative predictive ability starting from a set of antioxidant molecules with reported activity data. The pharmacophore of a drug defines the important functional

Table 7 Comparative features of the different models developed to explore the antioxidant activities of benzo[b]thiophene derivatives

Reference	Model development technique	$n_{training}$	R^2	Q^2	n_{test}	R_{pred}^2
Abreu et al. [17]	Partial least squares projection of latent structures method	18	0.881	0.844	8	0.843
Present work	3D pharmacophore developed with conformers generated from the <i>BEST</i> method of conformer generation	20	0.887	–	6	0.853
	3D pharmacophore developed with conformers generated from the <i>FAST</i> method of conformer generation	20	0.817	–	6	0.913

groups that are required for its activity, and their relative positions in space. Once a pharmacophore has been identified, it can be used in computerized searches of compound databanks to see whether known structures contain the same pharmacophore. Such structures may then be tested to see if they show activity, in which case they can be used as new lead compounds. In the present work, a set of arylamino-substituted benzothiophene derivatives have been used to model the 3D pharmacophore and to predict the activities of other compounds that were not utilized to develop that pharmacophore. The statistically significant pharmacophore thus generated with the conformers of the training set molecules revealed the structures required to imbue these molecules with potent antioxidant activities.

Upon validating the developed pharmacophores using various validation tools, two pharmacophores were found to be acceptable in this work. The first one was developed using the conformers of the training set compounds generated from the *BEST* method of conformational search, while the latter was developed with conformers of the same training set compounds, but which had been generated from the *FAST* method of conformer search. Both of the pharmacophores highlighted the importance of HBA and HBD features to the potent antioxidant activities of these molecules. The presence of the secondary amino hydrogen donor group and the electronegative oxygen atom of the methoxy substituent are the prime structural attributes associated with an increased activity profile in this series of substituted benzothiophenes. Besides these features, the pharmacophore developed with conformers generated from the *BEST* method also indicated the influential role of the benzothiophene moiety in modulating the antioxidant activities of these molecules, as implied by the presence of a ring aromatic (RA) feature in the selected pharmacophore. Both of the selected pharmacophores (hypothesis 1 from the former method and hypothesis 3 from the latter) were statistically significant in terms of their RMS deviations (0.518, 0.691) and correlation coefficient values (0.942, 0.904). The total costs of both hypotheses were close to the fixed costs of the corresponding hypotheses, indicating the good quality of the selected hypotheses. However, in both cases, the difference between the null cost and the total cost was somewhat smaller than that recommended in the software. On the other hand, lower cost differences than those recommended in the software have already been reported by the software developer in one of their case studies published on their website [39], as well as by other authors [40]. Such small cost differences can be explained by two facts: (1) the rigidity of the molecules constituting the training set; (2) the structural homology of the training set molecules. Both of the selected pharmacophores were validated using Fischer's randomization test (prediction of compounds with permuted

activity data) and external validation (mapping and activity prediction of test set compounds). The robustness of the selected pharmacophores was established on the basis of the inferior quality of models developed with scrambled activity data. Moreover, the acceptable values of the R^2_{pred} (0.853, 0.913) and $r^2_{m(\text{test})}$ (0.844, 0.821) parameters obtained suggest that the developed pharmacophores can be satisfactorily used to predict the activities of other arylamino-substituted benzothiophene derivatives.

Conclusions

A pharmacophore can be rightly used as a search tool to identify new chemical entities in chemical databases with potential activity, as well as a predictive tool for estimating the biological activities of virtual compounds designed on the basis of structure–activity analyses. The study presented here clearly indicates that the selected pharmacophores successfully fulfill the desired criteria. Both of the significant pharmacophores reported in the present work revealed the importance of the hydrogen bond donor and hydrogen bond acceptor features to the optimal antioxidant activities of the arylamino-substituted benzo[b]thiophene derivatives. Thus, for these molecules, the presence of substituents that facilitate hydrogen bond formation with nearby free radicals may result in enhanced antioxidant activity. The 3D pharmacophores also suggested that the hydrophobicity of the benzothiophene moiety is a feature that is crucial to the efficient antioxidant activities of these molecules. Hence, the 3D-QSAR approach used in this study defines the structural requirements of the antioxidant molecules (arylamino-substituted benzo[b]thiophene derivatives) for their effective interaction with nearby free radicals. Thus, the models may be utilized to estimate the potential antioxidant activities of virtual libraries of newly designed antioxidant molecules of this class prior to synthesis or biological testing.

Acknowledgments Financial assistance from the All India Council for Technical Education (AICTE), New Delhi in the form of a Research Promotion Scheme is gratefully acknowledged.

References

1. Genestra M (2007) Oxyl radicals, redox-sensitive signalling cascades and antioxidants. *Cell Signal* 19:1807–1819
2. Dröge W (2002) Free radicals in the physiological control of cell function. *Physiol Rev* 82:47–95
3. Bayol-Denizot C, Daval J-L, Netter P, Minn A (2000) Xenobiotic-mediated production of superoxide by primary cultures of rat cerebral endothelial cells, astrocytes, and neurons. *Biochim Biophys Acta-Mol Cell Res* 1497:115–126
4. Gutteridge JMC, Halliwell B (1994) Antioxidants in nutrition, health and disease. Oxford University Press, Oxford

5. Wright JS, Johnson ER, DiLabio GA (2001) Predicting the activity of phenolic antioxidants: theoretical method, analysis of substituent effects, and application to major families of antioxidants. *J Am Chem Soc* 123:1173–1183
6. Vafiadis AP, Bakalbassis EG (2005) A DFT study on the deprotonation antioxidant mechanistic step of ortho-substituted phenolic cation radicals. *Chem Phys* 316:195–204
7. Musialik M, Litwinienko G (2005) Scavenging of dpph• radicals by vitamin E is accelerated by its partial ionization: the role of sequential proton loss electron transfer. *Org Lett* 7:4951–4954
8. Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN (2008) Applications of 2D descriptors in drug design: a DRAGON tale. *Curr Top Med Chem* 8:1628–1655
9. Gonzalez MP, Teran C, Saiz-Urra L, Teijeira M (2008) Variable selection methods in QSAR: an overview. *Curr Top Med Chem* 8:1606–1627
10. Ray S, Sengupta C, Roy K (2007) QSAR modeling of antiradical and antioxidant activities of flavonoids using electrotopological state (E-State) atom parameters. *Cent Eur J Chem* 5:1094–1113
11. Ray S, Sengupta C, Roy K (2008) QSAR modeling for lipid peroxidation inhibition of flavonoids using topological and structural parameters. *Cent Eur J Chem* 6:267–276
12. Roy K, Mitra I, Saha A (2009) Molecular shape analysis of antioxidant and squalene synthase inhibitory activities of aromatic tetrahydro-1,4-oxazine derivatives. *Chem Biol Drug Des* 74:507–516
13. Rastija V, Medic-Saric M (2009) QSAR study of antioxidant activity of wine polyphenols. *Eur J Med Chem* 44:400–408
14. Gupta S, Matthew S, Abreu PM, Aires-de-Sousa J (2006) QSAR analysis of phenolic antioxidants using MOLMAP descriptors of local properties. *Bioorgan Med Chem* 14:1199–1206
15. Roy K, Mitra I (2009) Advances in quantitative structure–activity relationship models of antioxidants. *Expert Opin Drug Discov* 4:1157–1175
16. Wermuth CG, Langer T (2000) Pharmacophore identification. In: Kubinyi H (ed) *3D QSAR in drug design: theory, methods and applications*. Kluwer, Dordrecht, pp 117–149
17. Abreu RMV, Ferreira ICFR, Queiroz MRP (2009) QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo[b]thiophenes. *Eur J Med Chem* 44:1952–1958
18. Accelrys Inc. (2010) *Discovery Studio 2.1*. Accelrys Inc., San Diego
19. Esteves MA, Narender N, Marcelo-Curto MJ, Gigante B (2001) Synthetic derivatives of abietic acid with radical scavenging ability. *J Nat Prod* 64:761–766
20. Thomas JR (1960) The identification of radical products from the oxidation of diphenylamine. *J Am Chem Soc* 82:5955–5956
21. Ferreira ICFR, Queiroz MJRP, Vilas-Boas M, Estevinho LM, Begouinb A, Kirsch G (2006) Evaluation of the antioxidant properties of diarylamines in the benzo[b]thiophene series by free radical scavenging activity and reducing power. *Bioorg Med Chem Lett* 16:1384–1387
22. Kurogi Y, Guner OF (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr Med Chem* 8:1035–1055
23. Debnath AK (2002) Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of mycobacterium avium complex dihydrofolate reductase. *J Med Chem* 45:41–53
24. Kahnberg P, Howard MH, Liljefors T, Nielsen M, Nielsen EO, Sterner O, Pettersson I (2004) The use of a pharmacophore model for identification of novel ligands for the benzodiazepine binding site of the GABAA receptor. *J Mol Graph Model* 23:253–261
25. Faragalla J, Bremner J, Brown D, Griffith R, Heaton A (2003) Comparative pharmacophore development for inhibitors of human and rat 5- α -reductase. *J Mol Graph Model* 22:83–92
26. Ekins S, Bravi G, Wikel JH, Wrighton SA (1999) Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *J Pharmacol Exp Ther* 291:424–433
27. Leonard JT, Roy K (2006) On selection of training and test sets for the development of predictive QSAR models. *QSAR Comb Sci* 25:235–251
28. Everitt B, Landau S, Leese M (2001) *Cluster analysis*. Arnold, London
29. Dougherty ER, Barrera J, Brun M, Kim S, Cesar RM, Chen Y, Bittner M, Trent JM (2002) Inference from clustering with application to gene-expression microarrays. *J Comput Biol* 9:105–126
30. Accelrys Inc. (2010) *Cerius 2, v.4.10*. Accelrys Inc, San Diego
31. Smellie A, Teig SL, Towbin P (1995) Poling: promoting conformational variation. *J Comput Chem* 16:171–187
32. Roy K, Paul S (2009) Exploring 2D and 3D QSARs of 2,4-diphenyl-1,3-oxazolines for ovicidal activity against *Tetranychus urticae*. *QSAR Comb Sci* 28:406–425
33. Roy PP, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models. *Molecules* 14:1660–1701
34. Mitra I, Saha A, Roy K (2009) Quantitative structure-activity relationship modeling of antioxidant activities of hydroxybenzylacetones using quantum chemical, physicochemical and spatial descriptors. *Chem Biol Drug Des* 73:526–536
35. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Mod* 20:269–276
36. Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 27:302–313
37. Roy PP, Roy K (2008) Comparative QSAR studies of CYP1A2 inhibitor flavonoids using 2D and 3D descriptors. *Chem Biol Drug Des* 72:370–382
38. Patrick GL (2009) *An introduction to medicinal chemistry*. Oxford University Press, New York
39. Scott RK, Hoffmann RD Pharmacophoric design and validation for selective D₁ agonists. http://accelrys.com/references/case-studies/archive/studies/D1Agonists_full.html. Accessed on February 12, 2010
40. Hirashima A, Shigeta Y, Eiraku T, Kuwano E (2003) Inhibitors of calling behavior of *Plodia interpunctella*. *J Insect Sci* 3:4–12