European Journal of Medicinal Chemistry 44 (2009) 1952–1958

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo[*b*]thiophenes

Rui M.V. Abreu [a,c], Isabel C.F.R. Ferreira [a,*], Maria João R.P. Queiroz [b]

[a] CIMO-ESAB, Instituto Politécnico de Bragança, Campus de Sta. Apolónia, Apartado 1172, 5301-855 Bragança, Portugal
[b] Centro de Química, Universidade do Minho, Campus de Gualtar 4710-057 Braga, Portugal
[c] Instituto de Biotecnologia e Bioengenharia, Centro de Genética e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro (IBB/CGB-UTAD), Apdo 1013, 5001-801, Vila Real, Portugal

ARTICLE INFO

Article history:
Received 13 August 2008
Received in revised form
18 November 2008
Accepted 20 November 2008
Available online 6 December 2008

Keywords:
Benzo[*b*]thiophenes
Di(hetero)arylamines
Antioxidants
QSAR
PLS

ABSTRACT

A QSAR study was developed in order to model the antioxidant activity, specifically the radical scavenger activity (RSA), of 26 di(hetero)arylamines' derivatives of benzo[*b*]thiophenes. The QSAR model was constructed, using the partial least squares projection of latent structures (PLS) method, and its robustness and predictability were verified by internal and external cross-validation methods. A total of 4 molecular descriptors, belonging to RDF (Radial Distribution Function) descriptors (RDF020e and RDF045e) and 2D-autocorrelation descriptors (GATS8p and MATS5e) were selected to build the QSAR model. RDF descriptors seem to relate the presence of electronegative atoms at the inner atmosphere of the compounds to increase RSA. 2D-Autocorrelation descriptors associate the presence of polarizable and electronegative pairs of atoms, at specific topological distance, with the RSA of the compounds. Finally this QSAR model proved to be a useful tool in the prediction of radical scavenger activity of congeneric compounds and will be used to guide the synthesis of new diarylamines in our laboratory.

1. Introduction

Free radicals play important roles in many physiological and pathological conditions [1]. In general, excess of free radicals caused by the imbalance between free radical generation and scavenging may contribute to disease development [2]. Free radical can damage membranes, proteins, enzymes and DNA [3], increasing the risk of diseases such as cancer, Alzheimer's, Parkinson's [4], angiocardiopathy [5], arthritis, asthma, diabetes, and degenerative eye disease [2]. Cells are equipped with several defence systems against free radical damage, including oxidative enzymes such as superoxide dismutase, catalase, glutathione peroxidase and glutathione reductase, or chemical compounds such as α-tocopherol, ascorbic acid, carotenoids, polyphenol compounds and glutathione [2,6]. Considering that preventable diseases make up approximately 70% of the burden of diseases and its associated costs, it is easy to understand the great importance of the knowledge about free radicals production and control [7]. This control can be achieved through an input of antioxidants and free radicals' scavengers. Synthetic products with antioxidant activity may help the endogenous defence system and therefore, it is important to obtain effective free radicals' scavengers for the treatment and prevention of several disorders [8,9].

The antioxidant properties of diarylamines have recently been reported [10]. Particularly, some structure–activity relationship (SAR) studies have been made on the antioxidant activity of diaryl and di-heteroarylamines derivatives of benzo[*b*]thiophenes [11,12]. According to these SAR studies, the antioxidant activity of di(hetero)arylamines' derivatives of benzo[*b*]thiophenes was related to the position of arylamination and the number and position of substitution groups on both benzene or thiophene rings [11,12]. Nevertheless, none of those reports were in fact quantitative.

Quantitative structure–activity relationships (QSAR) analysis has often been used to find correlations between biological activities and molecular descriptors of different classes of compounds [9]. A number of QSAR studies on the antioxidant activity of several classes of compounds have also been reported substantiating the applicability on this type of studies [13–15]. The QSAR model was obtained using the PLS (partial least squares projection of latent structures) statistical method [16]. This approach is one of most useful techniques for molecular modeling in drug design and it has been successfully applied to several QSAR studies [13–15,17]. Recently, many software tools were developed to calculate thousands of different molecular descriptors, which can be applied

* Corresponding author. Tel.: +351 273 303219; fax: +351 273 325405.
E-mail address: iferreira@ipb.pt (I.C.F.R. Ferreira).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.11.011

in QSAR studies, including the DRAGON software used in this study [18].

In the present study the main goal was to build a QSAR model for description and prediction of radical scavenging activity of di(heteroaryl)amines in the benzo[*b*]thiophene series, using the PLS method. The molecular descriptors used in the QSAR equation were examined to search for clues on the free radical scavenging activity mechanism. This QSAR model will guide the synthesis of potential new diarylamines derivatives of benzo[*b*]thiophene as radical scavengers.

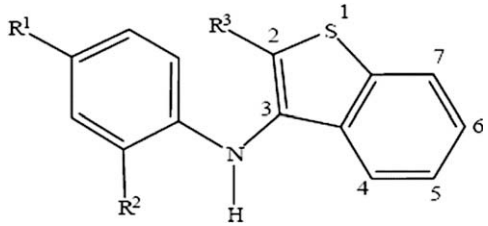
2. Results and discussion

To build the QSAR model for radical scavenging activity (RSA) of di(hetero)arylamines in the benzo[*b*]thiophene series, we used a total of 26 compounds, belonging to 3 different classes (**A**, **B** and **C**), that differ in terms of position of the (hetero)arylamino functionalization in the benzo[*b*]thiophene moiety (Tables 1–3). Classes **A**, **B** and **C** have the amino bond in the 3, 6 and 7 positions, respectively. All the compounds were obtained by Buchwald–Hartwig palladium-catalyzed C–N coupling reactions. The compounds of class **A** were obtained by coupling the ethyl 3-bromobenzo[*b*]thiophene-2-carboxylate with anilines and 5-aminoindole [19]. Compounds of class **B** were synthesized from 6-bromo or 6-aminobenzo[*b*]thiophenes coupled, respectively, with substituted anilines or phenylbromides [20,21]. Compounds of class **C** were synthesised by coupling 7-bromo or 7-amino-2,3-dimethylbenzo[*b*]thiophenes with methoxylated anilines and 3-aminopyridine or substituted bromobenzenes and 2-bromopyridine, respectively [12].

2.1. DPPH antioxidant activity assay

From the 26 compounds used in this study, the RSA of 12 were already reported by us, using the DDPH (2,2-diphenyl-1-picrylhydrazil) assay [11,12]. The remaining 14 were screened, using the same methodology, and the results are presented in Fig. 1. The RSA was measured as the percentage of DPPH radical inhibition. DPPH radical scavenging activity is a standard assay in antioxidant activity studies. The DPPH assay is very convenient for the screening of samples because it is rapid and independent of sample polarity [22]. As the antioxidant efficiencies of the 26 compounds used in this study differ in several orders of magnitude, the EC_{50} values (concentration required to inhibit DPPH radical formation by 50%) were transformed into pEC_{50} values.

Table 1
Structures and antioxidant activities (pEC_{50}) of 3-(arylamino)benzo[*b*]thiophenes **A** class.



Compound	R^1	R^2	R^3	Experimental	Predicted	Residual
A1	OCH_3	OCH_3	H	4.06	3.97	0.10
A2	OCH_3	OCH_3	COOH	3.61[b]	3.73	−0.12
A3[a]	OCH_3	OCH_3	$COOCH_2CH_3$	3.77[b]	3.73	0.04
A4	OH	OH	$COOCH_2CH_3$	3.84[b]	3.42	0.42
A5	H	OH	$COOCH_2CH_3$	3.91	3.70	0.21
A6[a]	H	F	H	3.81[b]	4.18	−0.37

[a] Test set.
[b] Ref. [11].

2.2. QSAR model

The 26 compounds were first divided in two groups: training and test sets. The training set, representing about two thirds of the total number of compounds (18 compounds), was used to build the QSAR model. The remaining one third (8 compounds) was assigned to the test set and used to validate the model. The division was made taking into consideration that both sets should: (a) represent all the benzo[*b*]thiophene classes used and (b) cover all the antioxidant activity scale [23,24]. Fig. 2A shows the number of compounds assigned to the three benzo[*b*]thiophene classes for both training and test sets. Compounds were also divided in three groups of antioxidant activity: 1–3, 3–4 and 4–5 [15], according to pEC_{50} values (Fig. 2B).

While constructing the model, great care was taken in order to avoid inclusion of highly collinear molecular descriptors. A pairwise correlation method was used and in the end only 4 descriptors were selected. The correlation matrix between the experimental pEC_{50} values and the molecular descriptors, and the molecular descriptors values used in this study are given in Tables 4 and 5, respectively. The molecular descriptors finally selected were:

RDF020e: Radial Distribution Function – 20/weighted by atomic Sanderson electronegativities.
RDF0245e: Radial Distribution Function – 45/weighted by atomic Sanderson electronegativities.
GATS8p: Geary autocorrelation – lag 8/weighted by atomic polarizabilities.
MATS5v: Moran autocorrelation – lag 5/weighted by atomic Sanderson electronegativities.

The QSAR model equation obtained and the statistical parameters were the following:

$$pEC_{50} = 0.3812 + 0.1690RDF020e + 0.0479RDF045e + 0.6129GATS8p + 1.3840MATS5e$$

$$N = 18; R^2 = 0.881; \rho < 10^{-7}; F = 42.72; S = 0.2731;$$

$$Q^2_{LOO} = 0.844; Q^2_{LMO\ (25\%)} = 0.817; Q^2_{LMO\ (50\%)} = 0.817;$$

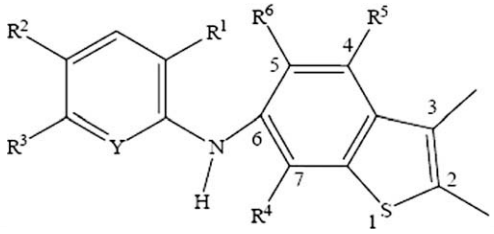
$$Q^2_{ext} = 0.843; RMSE_{(training\ set)} = 0.2816; RMSE_{(test\ set)} = 0.2216$$

were N is the number of compounds used, R^2 is the squared correlation coefficient, ρ is the significance of the model, F is the Fisher ratio, Q^2_{LOO}, $Q^2_{LMO\ (25\%)}$ and $Q^2_{LMO\ (50\%)}$ are the square of the leave-one-out and leave-many-out (25% and 50%) cross-validations, and $RMSE_{(training\ set)}$ and $RMSE_{(test\ set)}$ are Root Mean Squared Errors for the training and test sets, respectively.

The model was validated for its robustness and predictive power by internal leave-one-out (LOO) and leave-many-out (LMO) cross-validations, as demonstrated by R^2 and Q^2 values, and by external validation as demonstrated by Q^2_{ex} value. Also RMSE' values, for both the training and test sets, validate the model by presenting low and similar values.

A plot of predicted pEC_{50} *versus* experimental pEC_{50} values, for both the training and test sets, is shown on Fig. 3A. The agreement observed between the predicted and experimental values

Table 2
Structures and antioxidant activities (pEC_{50}) of 6-(heteroarylamino)benzo[*b*]thiophenes **B** class.



Compound	R^1	R^2	R^3	R^4	R^5	R^6	Y	Experimental	Predicted	Residual
B1	OCH_3	OCH_3	H	H	H	CH_3	C	4.07	4.07	0.00
B2	H	OCH_3	OCH_3	H	H	CH_3	C	4.30[b]	4.50	−0.20
B3	H	OCH_3	H	CH_3	CH_3	H	C	3.89	3.88	0.01
B4[a]	H	OCH_3	H	H	H	CH_3	C	4.14[b]	3.86	0.27
B5	H	H	OCH_3	H	H	CH_3	C	3.38	3.53	−0.14
B6[a]	H	CHO	H	H	H	CH_3	C	2.39	2.68	−0.29
B7	H	CN	H	CH_3	CH_3	H	C	2.29	2.45	−0.16
B8	Br	OCH_3	OCH_3	H	H	CH_3	C	3.81	4.06	−0.25
B9	Br	OCH_3	H	CH_3	CH_3	H	C	3.98	3.63	0.35
B10[a]	Br	OCH_3	H	H	H	CH_3	C	3.69	3.61	0.08
B11	Br	H	H	H	H	CH_3	C	3.05	2.82	0.23
B12[a]	Br	H	H	CH_3	CH_3	H	C	2.43	2.57	−0.15
B13	I	H	H	H	H	CH_3	C	2.91	3.10	−0.19
B14	H	H	H	CH_3	CH_3	H	N	2.10	2.52	−0.42

[a] Test set.
[b] Ref. [11].

confirmed the efficiency of this QSAR model. A plot of the residuals (predicted pEC_{50}–experimental pEC_{50}) *versus* experimental pEC_{50}, for both the training and test sets, is also shown on Fig. 3B, and a random distribution of the residuals about zero was observed for both sets. Considering the 3 standard deviation limit line (3S) for spotting outliers, all data were retained in the model.

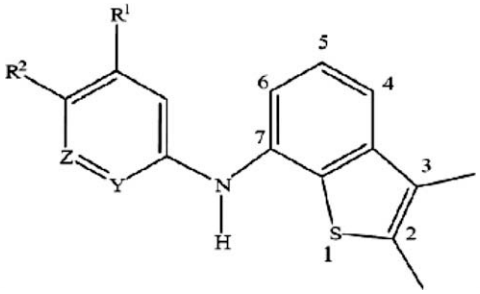
2.3. Model interpretation

Interpreting a QSAR model in terms of the specific contribution of substituents and other molecular features to the modeled activity is always a difficult task [25]. The standardized coefficients of each individual descriptor used in QSAR model presented similar values (Fig. 4), indicating that they all contribute to the variation of pEC_{50} values. This fact and the lack of a strong intercorrelation among the descriptors ($r < 0.75$) (Table 4), indicate that the examination of the molecular descriptors can lead to a better understanding of the relation between structure and antioxidant activity of the compounds [26].

RDF020e and RDF045e descriptors contribute more significantly to pEC_{50} variation as indicated by higher standardized coefficients values (Fig. 4). The radial distribution function (RDF) descriptors are based on the distance distribution of the compounds. The RDF descriptors of a molecule of *n* atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius *R* [27]. These 3D descriptors suggest the occurrence of some linear dependence between the RSA of the compounds and the 3D molecular distribution of electronegative atoms calculated at radius of 2.0 and 4.5 Å, from the geometrical center of each molecule. These radiuses correspond to the inner part of the compounds and can be roughly assigned to the first ring connected to the central amino group. The positive contribution of both RDF020e and RDF045e underlines the importance of the electrostatic environment surrounding the amino group on the overall antioxidant activity of the studied compounds. To highlight the meaning of the RDF020e and RDF045e descriptors, Fig. 5 shows an approximate representation of the descriptors for the most (Fig. 5A) and less (Fig. 5B) active compounds in our data set [28]. As it can be observed, the main difference is the presence of electronegative atoms like O and N in the inner regions of the compounds, specifically inside 2 and 4.5 Å radiuses. These observations are in good agreement with the SAR (structure–activity relationship) studies described on some of these diarylamines, in which the increase of the antioxidant activity is related to the presence of electron-donating substituents at the benzene rings [11,12].

The other two molecular descriptors that contribute to the QSAR model are GATS8p and MATS5e, and belong to the GATSd and MATSd families of 2D autocorrelation descriptors. The 2D-autocorrelation descriptors in general explain how the values of certain functions, at intervals equal to the lag d, are correlated. In the case of the descriptors used, lag is the topological distance, and the atomic properties are the functions correlated. These descriptors can be obtained by summing up the products of certain properties

Table 3
Structures and antioxidant activities (pEC_{50}) of 7-(hetero-arylamino)benzo[*b*]thiophenes **C** class.



Compound	R^1	R^2	Y	Z	Experimental	Predicted	Residual
C1[a]	OCH_3	OCH_3	C	C	4.26[b]	4.18	0.08
C2	OCH_3	H	C	C	2.81[b]	3.06	−0.25
C3	H	OCH_3	C	C	4.30[b]	4.13	0.17
C4	CN	H	C	C	1.84[b]	2.15	−0.31
C5	H	H	C	N	2.50[b]	1.94	0.55
C6[a]	H	H	N	C	2.26[b]	2.02	0.25

[a] Test set.
[b] Ref. [12].

of the two atoms located at a given topological distance or spatial lag [29,30]. There are slight differences between the 2D-autocorrelation descriptors of type GATSd and MATSd [29,30]; but in general, they describe how the considered property is distributed along the topological structure. GATS8p indicates that the presence of polarizable atoms at topological distance equal to 8 contribute positively to the antioxidant activity. A possible polarizable atom pair at topological distance of 8 bonds is presented for a top antioxidant activity compound **C3** (Fig. 6A). It is important to note that GATS8p contribution to the antioxidant activity is always positive as its mathematical definition implies only positive values. The topological distance 8 bonds (Fig. 6A) is most likely to be achieved between the two rings and suggest that the presence of polarizable atoms on both rings may give a positive contribute to the antioxidant activity. Likewise the MAST5e indicates that electronegative atoms at a topological distance equal to 5 bonds contribute to the antioxidant activity. To better understand this molecular descriptor, a possible electronegative atom pair that positively contributes to the antioxidant activity of **C1** compound is presented on Fig. 6B.

3. Conclusions

In this work the radical scavenger activity (RSA) of 26 di(hetero)arylamines derivatives of benzo[*b*]thiophenes was successfully modeled through Partial Least Square Projection to Latent Structures (PLS), using 4 molecular descriptors belonging to the Radial Distribution Function (RDF020e and RDF045e) and 2D-autocorrelation (GATS8p and MATS5e) families.

The QSAR model obtained showed high correlation coefficients ($Q^2_{LOO} = 0.844$, $Q^2_{LMO\ (25\%)} = 0.817$, $Q^2_{LMO\ (50\%)} = 0.817$ and $Q^2_{ext} = 0.843$), and also low root mean squared errors ($RMSE_{(training\ set)} = 0.2816$ and $RMSE_{(test\ set)} = 0.2216$), for both internal and external validation, confirming the good predictive power of the model.

The RDF descriptors used relate the presence of electronegative atoms at the inner atmosphere of the compounds to increased antioxidant activity. The 2D-autocorrelation descriptors associate the presence of polarizable and electronegative pairs of atoms at specific topological distance with the compound radical scavenging activity. The presence of these electronegative and polarizable atoms possibly increases RSA of the compounds by facilitating the hydrogen radical abstraction from the diarylamino group as previously described [12,21].

Finally this QSAR model proposed can be used in the prediction of the antioxidant activity of congeneric compounds of the derivatives of benzo[*b*]thiophenes used in this work in order to guide the synthesis of new compounds in our laboratory.

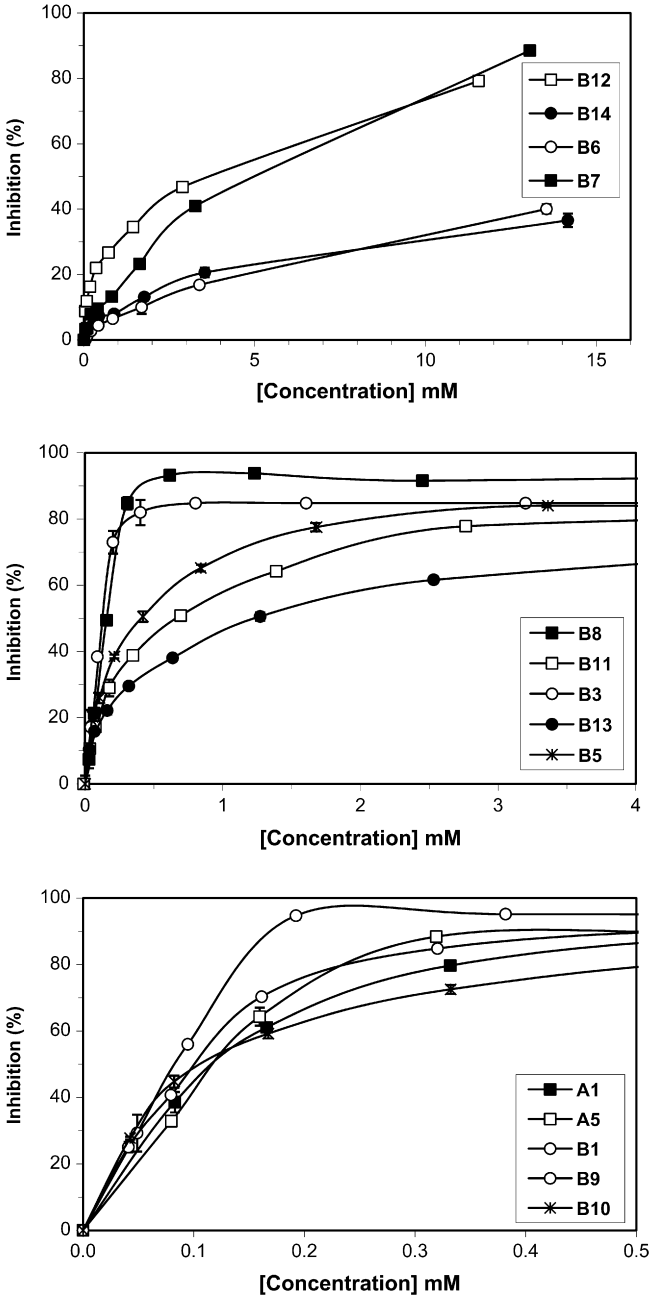


Fig. 1. Scavenging activity (%) on DPPH radicals (RSA) of di(hetero)arylamines' derivatives of benzo[*b*]thiophenes. Each value is expressed as mean ± standard deviation.

4. Experimental

4.1. Data set

A total of 26 compounds belonging to three different chemical classes: 3-di-heteroarylamine derivatives of benzo[*b*]thiophene

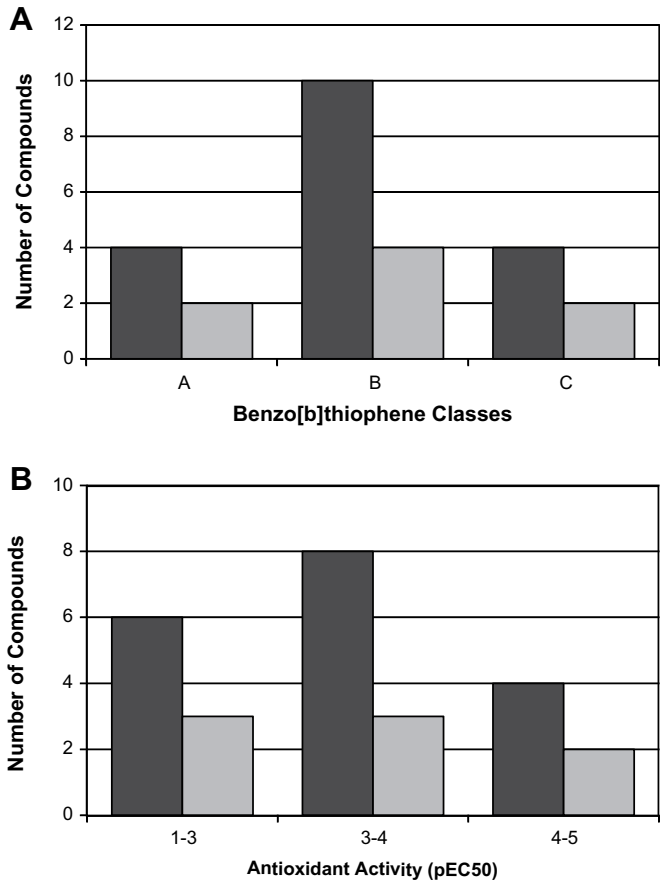


Fig. 2. Distribution of chemical classes (A) and biological activities (pEC_{50}) (B) *versus* number of di(hetero)arylamines' derivatives of benzo[*b*]thiophenes for the training set (black) and test set (grey) of the QSAR model.

(Class **A**: 6 compounds), 6-diarylamine derivatives of benzo[*b*]-thiophene (Class **B**: 14 compounds) and 7-di-heteroarylamine derivatives of benzo[*b*]thiophene (Class **C**: 6 compounds), were used in this study (Tables 1–3).

The Radical Scavenging Activity (RSA) of 12 of these compounds was previously reported by us [11,12]. The RSA of the remaining 14 compounds was monitored according to the method of Hatano et al. [31], with small modifications. Various concentrations of methanolic compounds' solutions (0.1 mL) were mixed with methanolic solution containing DPPH (2,2-diphenyl-1-picrylhydrazil) radicals (6×10^{-5} mol/L, 0.9 mL). The mixture was shaken vigorously and left to stand in the dark until stable absorption values were obtained (usually 60 min). The reduction of the DPPH radical was determined by measuring the absorption at 517 nm. The RSA was calculated as a percentage of DPPH discolouration using the equation: $\%RSA = [(A_{DPPH} - A_S)/A_{DPPH}] \times 100$, where A_s is the absorbance of the solution when the compound has been added at a particular concentration and A_{DPPH} is the absorbance of the DPPH solution. Mean values from three independent samples were calculated for each compound and standard deviations were also obtained.

Table 4
Correlation matrix between the experimental pEC_{50} values and the different molecular descriptors used to calculate the QSAR model.

	pEC_{50}	RDF020e	RDF045e	GATS8p	MATS5e
pEC_{50}	1	0.894	0.845	0.829	0.822
RDF020e		1	0.747	0.742	0.731
RDF045e			1	0.672	0.528
GATS8p				1	0.563
MATS5e					1

Table 5
Values of the 4 molecular descriptors used to calculate the QSAR model.

Compound	RDF020e	RDF045e	GATS8p	MATS5e
A1	8.961	22.084	1.635	0.009
A2	8.061	26.192	1.546	−0.155
A3	8.834	26.278	1.278	−0.135
A4	6.159	27.227	1.346	−0.098
A5	6.339	30.733	1.396	−0.056
A6	4.530	15.091	3.037	0.323
B1	9.304	28.040	1.196	0.030
B2	9.212	30.120	1.482	0.149
B3	7.351	24.758	1.008	0.327
B4	7.224	24.916	1.024	0.319
B5	7.231	20.544	1.671	−0.060
B6	6.209	17.863	0.792	−0.068
B7	4.873	20.050	0.657	−0.085
B8	9.033	22.721	1.269	0.205
B9	7.356	22.483	1.053	0.207
B10	6.727	24.650	1.064	0.190
B11	5.201	20.825	1.095	−0.078
B12	5.104	16.675	1.130	−0.117
B13	5.693	19.066	1.554	−0.079
B14	4.867	20.627	0.864	−0.147
C1	8.767	22.035	1.727	0.146
C2	7.186	18.211	1.123	−0.071
C3	7.161	19.728	1.872	0.326
C4	4.699	14.307	0.792	−0.145
C5	5.335	12.209	0.488	−0.162
C6	4.650	17.479	0.000	0.009

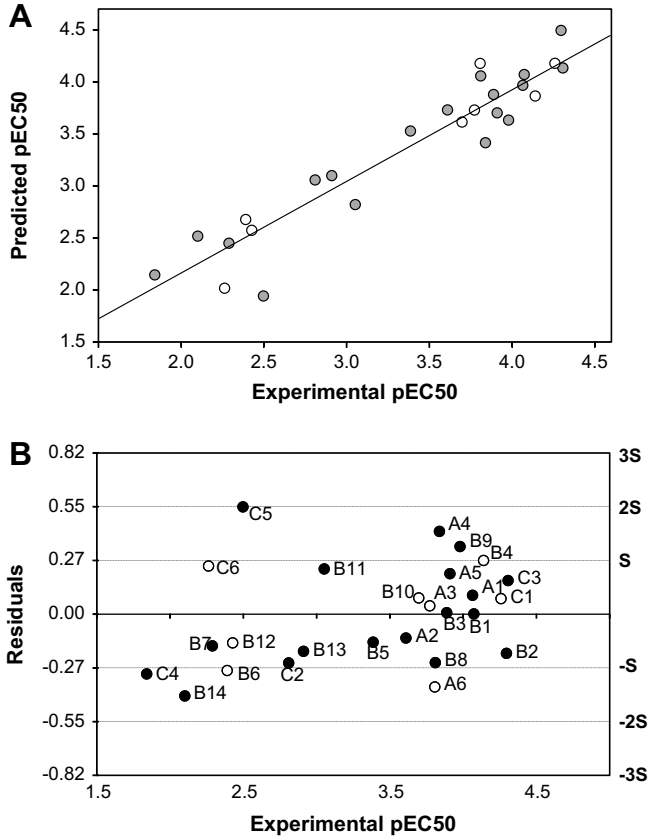


Fig. 3. Predicted *versus* experimental pEC_{50} (A) and residuals *versus* experimental pEC_{50} (B) for the training (•) and test sets (○) of di(hetero)arylamines' derivatives of benzo[*b*]thiophenes used in the QSAR model.

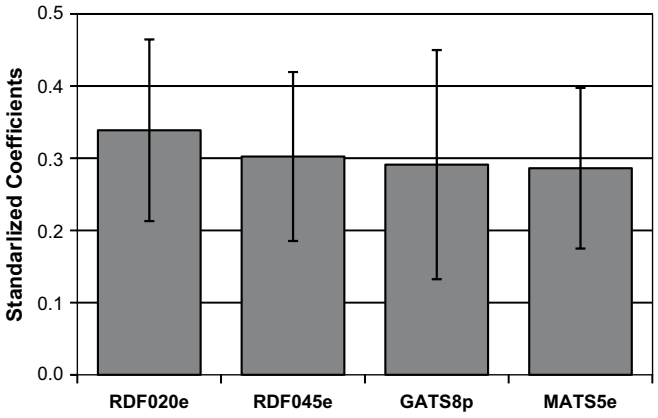


Fig. 4. Standardized coefficients and standard deviation of the descriptors used in the QSAR model.

To make the RSA data homogenous and directly comparable, all RSA activity was reported as EC_{50}, (expressed mol/L) required to inhibit the DPPH radical by 50%. Since the RSA varied by orders of magnitude, to guarantee the linear distribution of the dependent variable, EC_{50} values were transformed to log values ($pEC_{50} = \log 1/EC_{50}$).

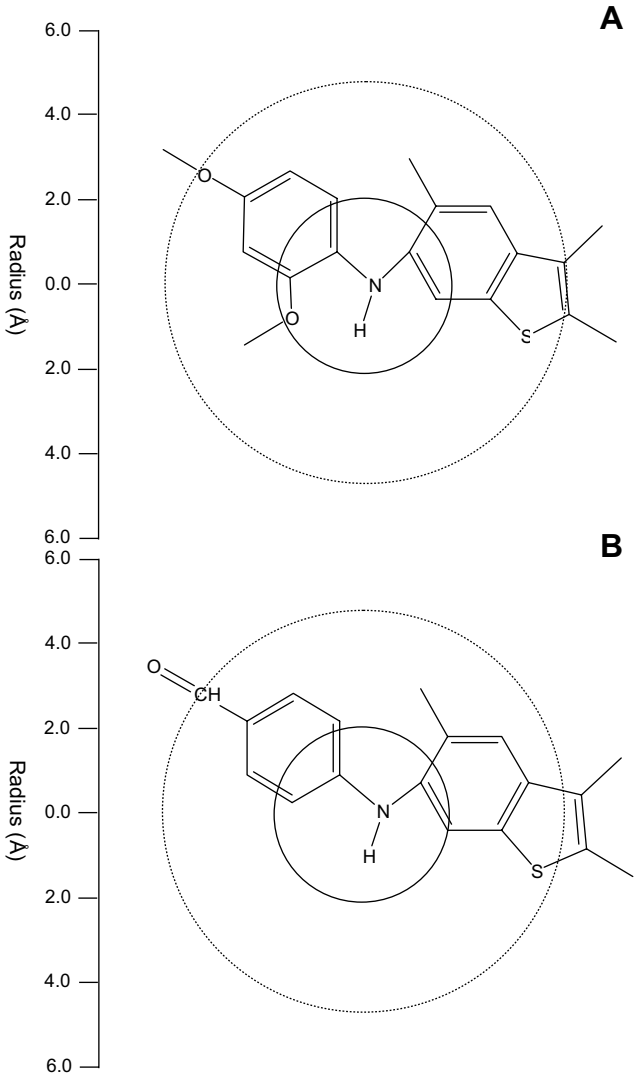


Fig. 5. Graphical representation of RDF020e and RDF045e descriptors for B1 (A) and B6 (B), respectively, the compounds with the highest and lowest antioxidant activities.

4.2. Selection of the molecular descriptors

The 3D structure models of the 26 compounds studied were generated using Ghemical molecular modeling software package [32] and then subject to geometry optimization, using semi-empirical quantum-chemical method AMI [33] implemented in MOPAC 7.0 computer software [34].

A total of 1664 molecular descriptors, belonging to different descriptor families, were calculated, for the 26 compounds, using the Dragon v5.3 computer software [18].

Descriptors with constant values were discarded. The correlations of the descriptors with each other and with the experimental pEC_{50} of the compounds were examined. Only the descriptors with a linear correlation coefficient to experimental pEC_{50} above 0.80 ($r > 0.80$) were retained, with 43 molecular descriptors meeting this criteria. For the remaining descriptors, a pairwise correlation analysis was performed consisting on the following steps: (1) starting from the descriptor with the highest correlation coefficient to experimental pEC_{50}, in this case RDF020e (Table 4), all the remaining molecular descriptors with a correlation coefficient with RDF020e above 0.75 ($r > 0.75$) were classified as collinear and were not included in the model; (2) next the same procedure was performed on the molecular descriptor with the highest correlation to experimental pEC_{50} still remaining on the list, and the process was continued until reaching the end of the list [14,35]. All this procedure was done using Dragon v5.3 software; only 4 molecular descriptors met the criteria and were finally selected and used to build the QSAR model (Table 4).

4.3. Statistical methods

The selected molecular descriptors obtained by DRAGON software were used to build a QSAR model using the Partial Least Square (PLS) [16] method implemented in SIMCA-P+ v12 statistics software [36].

The goodness of fit of the model was evaluated using the following statistical parameters: squared correlation coefficient (R^2), standard deviation of regression (S), significance of the model (ρ) and Fisher ratio value (F).

4.4. Model validation

The predictive stability and robustness of the model was first verified by internal cross-validation calculating the following parameters: Q^2_{LOO} (“Leave-One-Out”; 1-PRESS/TSS were PRESS is the Predictive Error Sum of Squares and TSS the Total Sum of Squares), Q^2_{LMO} (“Leave-Many-Out”) and $RMSE_{(training\ set)}$ (Root Mean Squared Errors for the training set). In this cross-validation method, one or many data points (in this case 25 and 50%) are removed from the set and the regression is recalculated, the predicted values for these values are then compared to their actual value. This is repeated until each data or data group have been omitted once [37,38].

Using the test set, the model was further checked by external cross-validation by calculating parameters: Q^2_{ext} (external, 1-PRESS/SD) and $RMSE_{(test\ set)}$ (Root Mean Squared Errors for the test set). PRESS is defined as the sum of the squared difference between the observed value and the predicted value for each compound in the test set, and SD is defined as the sum of the squared deviation between the observed value and the mean measured value of the training test [37].

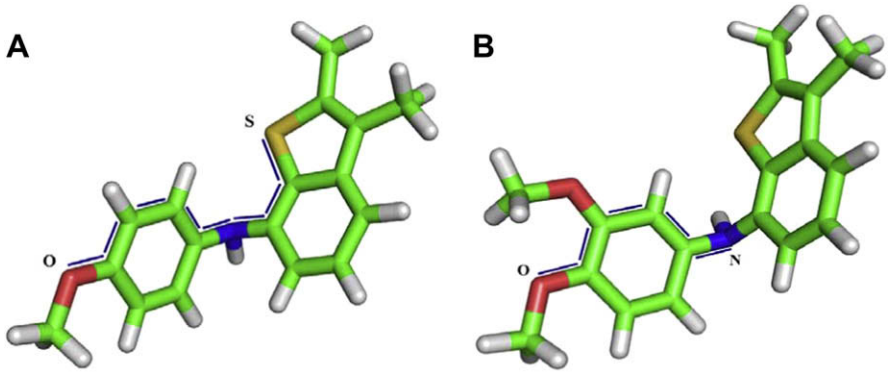


Fig. 6. Representation of possible polarizable atom pairs for compound **C3** at topological distance 8 (A) and for compound **C1** at topological distance 5 (B).

Acknowledgements

The authors are grateful to the Foundation for Science and Technology (Portugal) and FEDER for financial support to the research centres, research project POCI/QUI/59407/2004 and R.M.V.A. PhD grant (SFRH/BD/27430/2006).

References

[1] G. Vendemiale, I. Grattagliano, E. Altomare, Int. J. Clin. Lab. Res. 29 (1999) 49–55.
[2] M. Valko, D. Leibfritz, J. Moncol, M.T. Cronin, M. Mazur, J. Telser, Int. J. Biochem. Cell. Biol. 39 (2007) 44–84.
[3] M.S. Cooke, M.D. Evans, M. Dizdaroglu, J. Lunec, FASEB J. 17 (2003) 1195–1214.
[4] B. Halliwell, Drugs Aging 18 (2001) 685–716.
[5] P. Abrescia, P. Golino, Expert Rev. Cardiovasc. Ther. 3 (2005) 159–171.
[6] E. Cadenas, Biofactors 6 (1997) 391–397.
[7] J.F. Fries, C.E. Koop, J. Sokolov, C.E. Beadle, D. Wright, Health Aff. (Millwood) 17 (1998) 70–84.
[8] A.A. Geronikaki, A.M. Gavalas, Comb. Chem. High Throughput. Screen. 9 (2006) 425–442.
[9] A.C. Kontogiorgis, A.E. Pontiki, D. Hadjipavlou-Litina, Mini-Rev. Med. Chem. 5 (2005) 563–574.
[10] M.A. Esteves, N. Narender, M.J. Marcelo-Curto, B. Gigante, J. Nat. Prod. 64 (2001) 761–766.
[11] I.C. Ferreira, M.J. Queiroz, M. Vilas-Boas, L.M. Estevinho, A. Begouin, G. Kirsch, Bioorg. Med. Chem. Lett. 16 (2006) 1384–1387.
[12] M.J. Queiroz, I.C. Ferreira, R.C. Calhelha, L.M. Estevinho, Bioorg. Med. Chem. 15 (2007) 1788–1794.
[13] D. Amic, D. Davidovic-Amic, D. Beslo, V. Rastija, B. Lucic, N. Trinajstic, Curr. Med. Chem. 14 (2007) 827–845.
[14] A.B. Camargo, E. Marchevsky, J.M. Luco, J. Agric. Food Chem. 55 (2007) 3096–3103.
[15] A.C. Durand, A. Farce, P. Carato, S. Dilly, S. Yous, P. Berthelot, P. Chavatte, J. Enzyme Inhib. Med. Chem. 22 (2007) 556–562.
[16] S. Wold, M. Sjostrom, L. Eriksson, Chemometr. Intell. Lab. Syst. 58 (2001) 109–130.
[17] L. Rackova, V. Snirc, M. Majekova, P. Majek, M. Stefek, J. Med. Chem. 49 (2006) 2543–2548.
[18] DRAGON for Windows (software for molecular descriptor calculations). Version 5.3, Talete Srl, 2005.http://www.talete.mi.it.
[19] M.J.R.P. Queiroz, A. Begouin, I.C.F.R. Ferreira, G. Kirsch, R.C. Calhelha, S. Barbosa, L.M. Estevinho, Eur. J. Org. Chem. (2004) 3679–3685.
[20] I.C.F.R. Ferreira, M.J.R.P. Queiroz, G. Kirsch, Tetrahedron 59 (2003) 3737–3743.
[21] I.C.F.R. Ferreira, M.J.R.P. Queiroz, G. Kirsch, Tetrahedron 59 (2003) 975–981.
[22] Koleva T.A. van Beek II, J.P. Linssen, A. de Groot, L.N. Evstatieva, Phytochem. Anal. 13 (2002) 8–17.
[23] L. Saiz-Urra, M.P. Gonzalez, M. Teijeira, Bioorg. Med. Chem. 15 (2007) 3565–3571.
[24] O. Farkas, J. Jakus, K. Heberger, Molecules 9 (2004) 1079–1088.
[25] M.P. Gonzalez, J. Caballero, A.M. Helguera, M. Garriga, G. Gonzalez, M. Fernandez, Bull. Math. Biol. 68 (2006) 735–751.
[26] J. Ghasemi, S. Asadpour, A. Abdolmaleki, Anal. Chim. Acta. 588 (2007) 200–206.
[27] M.C. Hemmer, V. Steinhauer, J. Gasteiger, Vib. Spectrosc. 19 (1999) 151–164.
[28] M.P. Gonzalez, J. Caballero, A. Tundidor-Camba, A.M. Helguera, M. Fernandez, Bioorg. Med. Chem. 14 (2006) 200–213.
[29] P. Broto, G. Moreau, C. Vandycke, Eur. J. Med. Chem. 19 (1984) 79–84.
[30] G. Moreau, P. Broto, Nouveau Journal De Chimie – New J. Chem. 4 (1980) 359–360.
[31] T. Hatano, H. Kagawa, T. Yasuhara, T. Okuda, Chem. Pharm. Bull. 36 (1988) 2090–2097.
[32] Ghemical. In: http://www.bioinformatics.org/ghemical/, ed.
[33] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, J. Am. Chem. Soc. 107 (1985) 3902–3909.
[34] J.J.P. Stewart, MOPAC2007, Stewart Computational Chemistry, Colorado Springs, CO, USA, 2007.
[35] M. Fernandez, J. Caballero, A.M. Helguera, E.A. Castro, M.P. Gonzalez, Bioorg. Med. Chem. 13 (2005) 3269–3277.
[36] SIMCA P+ v12. Umetrics AB, 2008.
[37] P. Gramatica, QSAR Comb. Sci. 26 (2007) 694–701.
[38] P. Gramatica, E. Papa, QSAR Comb. Sci. 24 (2005) 953–960.