Journal of Molecular Graphics and Modelling 28 (2009) 12–19

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Predictive modeling of the total deactivation rate constant of singlet oxygen by heterocyclic compounds

Andrew G. Mercader [a,*], Pablo R. Duchowicz [a], Francisco M. Fernández [a], Eduardo A. Castro [a], Franco M. Cabrerizo [b,1], Andrés H. Thomas [a]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), División Química Teórica, Diag. 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina

[b] Centro de Investigaciones en Hidratos de Carbono (CIHIDECAR-CONICET), Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón 2, 3p, Ciudad Universitaria, 1428 Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form 4 March 2009
Accepted 7 March 2009
Available online 20 March 2009

Keywords:
QSPR
Enhanced replacement method
Genetic algorithm
Heterocycles
Singlet oxygen

ABSTRACT

We constructed a predictive model of the total deactivation rate constant (k_t) of singlet oxygen by heterocyclic compounds that are widespread in biological systems and participate in highly relevant biologic functions related with photochemical processes, by means of quantitative structure–property relationships (QSPR). The study of the reactivity of singlet oxygen with biomolecules provides their antioxidant capability, and the determination of the rate constants allows evaluation of the efficiencies of these processes. Our optimal linear model based on 41 molecular structures, which have not been used previously in a QSPR study, consists of six variables, selected from more than thousand geometrical, topological, quantum-mechanical and electronic types of molecular descriptors. Our recently developed strategy to determine the optimal number of descriptors in model is successfully applied. As a practical application of our QSPR model we estimated the unknown k_t of several heterocyclic compounds that are of particular interest for further experimental studies in our research group.

1. Introduction

Oxygen is ubiquitous, comprising nearly 50% of the Earth's crust and is an essential component in the metabolic pathways of all higher organisms [1]. With two singlet states lying close above its triplet ground state, the O_2 molecule possesses a very unique configuration, which gives rise to a very rich and easily accessible chemistry, and also to a number of important photophysical interactions [2]. The lowest electronic excited state of molecular oxygen, named singlet oxygen ($O_2(^1\Delta_g)$, denoted throughout as $^1O_2$), is an electrophilic molecule that has a high capacity to oxidize a variety of electron-rich organic compounds [3]. This active species has physical and chemical properties that have intrigued researchers in several areas of science for more than 70 years [4,5]; participating in reactions that comprise great interest in different fields of science: environmental chemistry, bromatology, biochemistry, biology, etc. Although $^1O_2$ can be generated by chemical, enzymatic and photochemical paths, photosensitization is primarily responsible for the production of $^1O_2$ *in vivo* [6]. Moreover, $^1O_2$ is one of the main activated species responsible for the damaging effects of light on biological systems (photodynamic effects) [5].

$^1O_2$ relaxes to ground state $^3O_2$ by both radiation-less and radiative pathways:

$$^1O_2 \xrightarrow{k_d} {}^3O_2 \quad (1)$$

$$^1O_2 \xrightarrow{k_e} {}^3O_2 + h\nu \quad (2)$$

It may also be deactivated by oxidation of an acceptor molecule Q

$$Q + {}^1O_2 \xrightarrow{k_r} QO_2 \quad (3)$$

and/or the interaction with a physical quencher

$$Q + {}^1O_2 \xrightarrow{k_q} Q + {}^3O_2 \quad (4)$$

Any biological compound that is able to deactivate $^1O_2$ may efficiently have a protective role against $^1O_2$ *in vivo* and very likely against other reactive oxygen species. Therefore, the study of the reactivity of $^1O_2$ with biomolecules would provide their antioxidant capability, and the determination of the rate constants of $^1O_2$ for the total (physical and chemical) quenching ($k_t = k_r + k_q$) allows evaluation of the efficiencies of these processes. However, determination of k_t values is experimentally difficult and requires

* Corresponding author. Tel.: +54 221 425 7430/7291; fax: +54 221 425 4642.
E-mail address: amercader@inifta.unlp.edu.ar (A.G. Mercader).
[1] Tel.: +54 11 45763346; fax: +54 11 45763352.

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.03.002

specialized equipment to detect the weak $^1O_2$ luminescence emission in the near infrared [7]. Therefore, despite the biological importance of the subject and the enormous diversity of compounds able to interact with $^1O_2$, the kinetic studies on reactivity of this reactive oxygen species are still scarce.

Heterocyclic compounds are widespread in biological systems and participate in highly relevant biologic functions. Important families of biomolecules, such as nucleobases, porphyrins, flavins and pterins are included in this group of compounds. In particular, reactivity of $^1O_2$ with pterins has been recently studied [8–10].

Clearly, it is of great interest the prediction of unknown deactivation constants of $^1O_2$ by a given set of compounds, as well as attempting to determine the molecular structural parameters which k_t depends on. A generally accepted remedy for overcoming the lack of experimental data in complex chemical phenomena is the analysis based on quantitative structure–property relationships (QSPR) [11]. The ultimate role of the different formulations of the QSPR theory is to suggest mathematical models for estimating relevant properties of interest, especially when they cannot be experimentally determined for some reason. These studies simply rely on the assumption that the physicochemical properties of a compound are determined solely by its molecular structure. The molecular structure is therefore translated into the so-called molecular descriptors through mathematical formulae obtained from several theories, such as Chemical Graph Theory, Information Theory, Quantum Mechanics, etc. [12,13]. Currently there are available in the literature thousands of theoretical descriptors, and one usually faces the problem of selecting those that are the most representative of the property under consideration.

The present study reports the predictions of k_t for 41 heterocyclic compounds that include different functional groups, whose experimental data are available in the literature [10,14,15]. Here, this set of molecules is used in a QSPR study for the first time. A great number of structural molecular descriptors including definitions of all classes are searched using the recently presented enhanced replacement method (ERM) [16] to perform the optimal variable subset selection. With the purpose of testing and comparison we also apply the well-established replacement method (RM) [17–20] and genetic algorithm (GA) [21]. As a practical application of our QSPR model we estimate the unknown k_t of several heterocyclic compounds that are of particular interest for further experimental studies in our research group.

2. Materials and methods

2.1. Data set

The training set of present study consists of 41 heterocyclic compounds (which have not been used for this purpose before, as far as we know) with known total rate constant k_t of $^1O_2$ quenching [10,14,15]. Among the available experimental data, we select only those ones measured in polar protic solvents at the same temperature (T = 298 K) and pressure (P = 1 atm). Average of multiple reported values of the same compound are used. According to the experimental data available in the literature [15] we assume that different polar protic solvents would not affect significantly the measure of the deactivation rate constant. This premise would be verified by the statistical analysis. Table 1 shows the experimental values of $\log(k_t)$ for the chosen heterocyclic compounds.

2.2. Molecular descriptors

The structures of the compounds are firstly pre-optimized with the molecular mechanics force field (MM+) procedure included in the package Hyperchem 6.03 [22], and the resulting geometries are further refined by means of the semi-empirical method PM3 (Parametric Method-3) using the Polak–Ribiere algorithm and a gradient norm limit of 0.01 kcal Å^{-1}. We compute the molecular descriptors using the software e-Dragon [23], including parameters of all types: Constitutional, Topological, Geometrical, Charge, GETAWAY (Geometry, Topology and Atoms-Weighted AssemblY), WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MoRSE (3D-Molecular Representation of Structure based on Electron diffraction), Molecular Walk Counts, BCUT descriptors, 2D-Autocorrelations, Aromaticity Indices, Randic Molecular Profiles, Radial Distribution Functions, Functional Groups, Atom-Centred Fragments, Empirical and Properties [24]. We add 20 constitutional descriptors and 4 quantum-chemical ones (molecular dipole moments, total energies, homo-lumo energies) not provided by the program e-Dragon to the pool. We thus end with a total of D = 1659 descriptors.

2.3. Model search

Our calculations are based on a suite of routines written in the computer system Matlab 5.0 [25]. It is our purpose to search the set $\mathbf{D}$ of D descriptors for an optimal subset $\mathbf{d}$ of $d \ll D$ ones with minimum standard deviation S according to multivariable linear regression (MLR):

$$S = \frac{1}{(N-d-1)} \sum_{i=1}^{N} res_i{}^2 \tag{5}$$

where N is the number of molecules in the training set, and res_i the residual for molecule i, the difference between the experimental property ($\mathbf{p}$) and the predicted one ($\mathbf{p_{pred}}$). More precisely, we want to obtain the global minimum of $S(\mathbf{d})$ where $\mathbf{d}$ is a point in a space of $D!/[d!(D-d)!]$ ones. A full search (FS) of optimal variables is impractical because it requires $D!/[d!(D-d)!]$ linear regressions. Some time ago we have proposed the replacement method (RM) [17–20] and more recently the enhanced replacement method (ERM) [16], both algorithms produce linear regression QSPR–QSAR models that are quite close the FS ones with much less computational work. These techniques are able to approach the minimum of S by judiciously taking into account the relative errors of the coefficients of the least-squares model given by a set of d descriptors $\mathbf{d} = \{X_1, X_2, \ldots, X_d\}$. The RM gives models with better statistical parameters than the Forward Stepwise Regression procedure [26] and similar ones to the more elaborated genetic algorithms [21], and the ERM is an improvement on the RM [16].

A GA is a search technique based on natural evolution where variables play the role of genes (in this case a set of descriptors) in an individual of the species. An initial group of random individuals (population) evolves according to a fitness function (in this case the standard deviation) that determines the survival of the individuals. The algorithm searches for those individuals that lead to better values of the fitness function through selection, mutation and crossover genetic operations. The selection operators guarantee the propagation of individuals with better fitness in future populations. The GAs explore the solution space combining genes from two individuals (parents) using the crossover operator to form two new individuals (children) and also by randomly mutating individuals using the mutation operator. The GAs offer a combination of hill-climbing ability (natural selection) and a stochastic method (crossover and mutation) and explore many solutions in parallel processing information in a very efficient manner. The practical application of GAs requires the tuning of some parameters such as population size, generation gap, crossover rate, and mutation rate. These parameters typically interact among themselves nonlinearly and cannot be optimized one at a time. There is considerable discussion about parameter settings

Table 1
Experimental and predicted (Eq. (7)) log (k_t), and residuals.

Number	Name	log (k_t) exp.	log (k_t) pred.	Residual
1	7,8-Dihydrofolic acid	8.74	8.59	0.15
2	7,8-Dihydrobiopterin	8.57	8.28	0.29
3	7,8-Dihydroneopterin	8.66	8.71	−0.05
4	6-Formyl-7,8-dihydropterin	8.32	8.19	0.13
5	Sepiapterin	8.28	8.28	0.00
6	7,8-Dihydroxantopterin	8.83	9.09	−0.26
7	Pterin	6.46	6.92	−0.46
8	6-Methylpterin	6.90	6.91	−0.01
9	6,7-Dimethylpterin	7.60	7.25	0.35
10	6-(Hydroxymethyl)pterin	6.49	6.21	0.28
11	6-Formylpterin	6.15	6.23	−0.08
12	6-Carboxypterin	6.15	6.22	−0.07
13	Biopterin	6.38	6.35	0.03
14	Neopterin	6.36	6.36	0.00
15	Folic acid	7.48	7.52	−0.04
16	Histamine	8.06	8.21	−0.15
17	Imidazole	7.46	7.31	0.15
18	4-Methyl-imidazole	8.11	7.89	0.22
19	Indole	7.65	8.00	−0.35
20	2,3-Dimethy-indole	8.76	8.65	0.11
21	3-Methyl-indol	8.20	8.33	−0.13
22	Indole 3 acetic acid	8.83	8.49	0.34
23	Indole-3-propianamide	7.89	8.07	−0.18
24	Indole-3-propionic acid	7.91	8.31	−0.40
25	2,5-Diphenyl-oxazole	8.20	8.05	0.15
26	2,5-Diphenyl-4-methyl-oxazole	7.53	7.68	−0.15
27	4-Methyl-2-(3-chlorophenyl)-5-phenyl oxazole	7.23	7.23	0.00
28	4-Methyl-2-(4-chlorophenyl)-5-phenyl oxazole	7.28	7.43	−0.15
29	4-Methyl-2-(4-methoxyphenyl)-5-phenyl oxazole	7.72	7.59	0.13
30	4-Methyl-2-(4-methylphenyl)-5-phenyl oxazole	7.57	7.67	−0.10
31	4-Methyl-2-(4-nitrophenyl)-5-phenyl oxazole	7.08	7.01	0.07
32	2,3-Dihydro-1-methyl-4-phenyl-pyridinium	6.23	6.09	0.15
33	1-Methyl-pyridinium	5.81	6.14	−0.33
34	1-Methyl-4-phenyl-pyridinium	5.95	5.83	0.12
35	cis(−)-2,3,4,4a,5,9b-Hexahydro-2,8-dimethyl-pyrido[4,3-b]indole	8.11	8.37	−0.26
36	1,2,3,4-Tetrahydro-2,8-dimethyl-pyrido[4,3-b]indole	8.23	8.46	−0.23
37	1-(1,1-dimethylethyl)-pyrrole	8.08	7.89	0.19
38	2-(1,1-dimethylethyl)-pyrrole	8.18	8.28	−0.10
39	3-(1,1-dimethylethyl)-pyrrole	8.26	8.21	0.05
40	Quinoline	9.00	8.51	0.49
41	1,2-Dihydro-2,2,4-trimethyl-quinoline, homopolymer (Permanax 45)	8.98	8.86	0.12

and approaches to parameter adaptation in the evolutionary computation literature; however there does not seem to be conclusive results on which may be the best [27].

Both the RM and ERM yield the optimal QSPR model with a given number d of descriptors. We also have to determine the optimal value d_{opt} of the number of variables of our model. The Kubinyi function (*FIT*) [28,29] is a statistical parameter that closely relates to the Fisher ratio (F), but avoids the main disadvantage of the latter that is too sensitive to changes in small d values and poorly sensitive to changes in large d values. The *FIT*($\mathbf{d}$) criterion has a low sensitivity to changes in small d values and a substantially increasing sensitivity for large d values. The greater the *FIT* value the better the linear equation.

Commonly, we expect a plot of *FIT* vs. d to present a maximum from which it is possible to calculate the optimal number of molecular descriptors (d_{opt}) to be included in the linear regression model. There are some occasions when the maximum is not reached after adding a reasonable number of descriptors in the model. For this reason we have recently proposed a variable *FIT* equation or *VFIT* which depends on a semi-empirical constant v that gives more weight to d in the *FIT* equation [30]. It reads:

$$VFIT = \frac{R^2(N - -1)}{(N + d^2)(1 - R^2)} \tag{6}$$

where R is the correlation coefficient for a model with d descriptors. By means of this equation, we obtain d_{opt} as the number of descriptors that yields the maximum value of *VFIT* (d_{max}) in the plot of *VFIT* vs. d. The semi-empirical constant v is determined by taking incremental values of 0.5 until that maximum complies with the rule of thumb that at least five data points should be present for each fitting parameter [31].

As a theoretical validation of all the models we choose the well-known Leave-One-Out (*loo*) and the Leave-More-Out Cross-Validation procedures (*l-n%-o*) [32], where n% is the number percent of molecules removed from the training set. We generate 5,000,000 cases of random data removal for *l-n%-o*, where n% = 30% (12 heterocyclic compounds).

2.4. Orthogonalization procedure

We employ the orthogonalization procedure introduced several years ago by Milan Randic [33,34] as a way of improving the statistical interpretation of the QSPR model that is built by using interrelated (overlapped) molecular descriptors. From our point of view, the co-linearity of the molecular descriptors should be as low as possible, because the interrelatedness among the different descriptors can lead to highly unstable regression coefficients, which makes it impossible to know the relative importance of an index and underestimates the utility of the regression coefficients of the model. As it is known, the crucial step of an orthogonalization process involves the choice of an appropriate order of orthogonalization. Therefore, we consider plausible to select the orthogonalization order in such a way that maximises the

correlation coefficient between each calculated orthogonal descriptor and the observed experimental property values (in decreasing order).

3. Results and discussion

By means of the ERM we search the total pool of $D = 1659$ descriptors and obtain optimal models with $d = 1, 2, \ldots, 15$ parameters linking the molecular structure of the heterocyclic compounds with their total rate constant k_t. On increasing ν in *VFIT* as indicated above we find four maxima at $d = 13$ ($\nu = 2$), $d = 9$ ($\nu = 3$), $d = 8$ ($\nu = 3.5$), and $d = 6$ ($\nu = 4$). The last one is consistent with the rule of thumb that states that in this case the number of fitting parameters should be less than 8. Fig. 1 shows the resulting *VFIT* with $\nu = 4$ that exhibits a maximum at $d = d_{max} = 6$. We assume that this is the optimal value of descriptors in the model. Fig. 1 also shows that *FIT* does not present a maximum in the interval $1 \le d \le 15$. Table 2 shows that $d_{max} = 6$ remains stable under additional increments of ν supporting the fact that this is actually the optimal number of model parameters.

According to the above mentioned results, the optimal QSPR model according to ERM is

$$\begin{aligned}\log(k_t) = {} & 7.8611(\pm 0.4) - 1.2329(\pm 0.2)\, GATS1e \\ & - 1.1219(\pm 0.1)\, Mor16u + 5.5727(\pm 0.9)\, E1\nu \\ & - 32.9421(\pm 3.8)\, R8m^{+} - 1.4465(\pm 0.1)\, nN^{+} \\ & - 1.1799(\pm 0.1) nArOH \end{aligned} \quad (7)$$

$$N = 41, \quad R = 0.9727, \quad S = 0.2323, \quad FIT = 7.7689,$$
$$p < 10^{-5}, \quad R_{loo} = 0.9609, \quad S_{loo} = 0.2778,$$
$$R_{l\text{-}30\%\text{-}o} = 0.8793, \quad S_{l\text{-}30\%\text{-}o} = 0.4892$$

where the absolute errors of the regression coefficients are given in parentheses, p is the significance of the model, *FIT* the Kubinyi function, and *loo* and *l-30%-o* stand for the Leave-One-Out and Leave-More-Out Cross-Validation techniques, respectively.

Following the same strategy the RM [17–20] yields the following optimal set of $d = 6$ descriptors:

$$\begin{aligned}\log(k_t) = {} & 6.8842(\pm 0.9) + 0.7754(\pm 0.1)\, nR10 \\ & - 0.803(\pm 0.2)\, GATS1\, p + 42687(\pm 1.4)\, E1e \\ & - 23.779(\pm 4.3)\, R8m^{+} - 1.5004(\pm 0.2)\, nN^{+} \\ & - 0.4449(\pm 0.03)\, N - 075 \end{aligned} \quad (8)$$

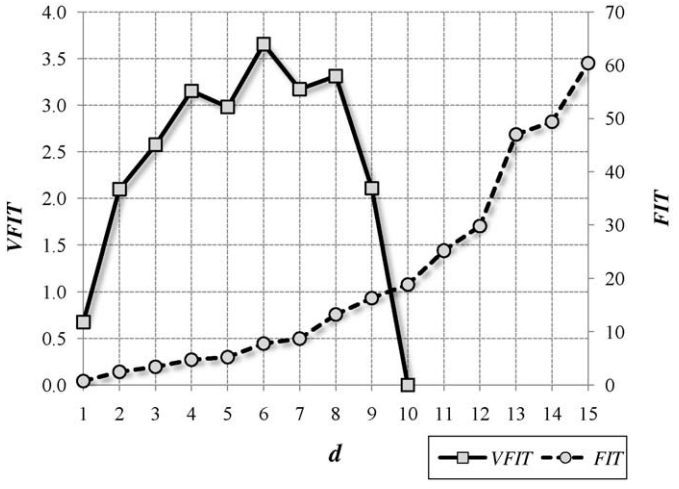

Fig. 1. *VFIT* (squares, left axe) and *FIT* (circles, right axe) in terms of the number of descriptors for the training set.

Table 2
Values of ν and d corresponding to maxima in *VFIT*.

V	d (max.)	ν	d (max.)	ν	D (max.)
1	–	6.5	3	12	2
1.5	–	7	3	12.5	2
2	13	7.5	2	13	2
2.5	13	8	2	13.5	2
3	9	8.5	2	14	2
3.5	8	9	2	14.5	2
4	6	9.5	2	15	2
4.5	6	10	2	15.5	2
5	4	10.5	2	16	2
5.5	4	11	2	16.5	2
6	4	11.5	2	17	1

$$N = 41, \quad R = 0.9603, \quad S = 0.2797, \quad FIT = 5.2249,$$
$$p < 10^{-3}, \quad R_{loo} = 0.9385, \quad S_{loo} = 0.3474, \quad R_{l\text{-}30\%\text{-}o} = 0.7993,$$
$$S_{l\text{-}30\%\text{-}o} = 0.6075$$

For comparison we also derive an optimal model with d_{opt} descriptors by means of GA. After several runs to optimize the GA parameters we find: number of individuals = 250; generation gap = 0.9; single point crossover probability = 0.6; mutation probability = $0.7/d$. The algorithm is stopped when a single individual occupied more than 90% of the population or when the number of generations reach 2500. We conclude that the best GA model is

$$\begin{aligned}\log(k_t) = {} & 7.1903(\pm 0.5) - 1.533(\pm 0.1) C - 032 \\ & - 23.733(\pm 3.9) R8m^{+} - 1.3538(\pm 0.2) nN^{+} \\ & + 5.4508(\pm 1.2) De - 1.1976(\pm 0.2) GATS1\, p \\ & + 0.8661(\pm 0.1) nR10 \end{aligned} \quad (9)$$

$$N = 41, \quad R = 0.968, \quad S = 0.2515, \quad FIT = 6.5671,$$
$$p < 10^{-4}, \quad R_{loo} = 0.9524, \quad S_{loo} = 0.306, \quad R_{l\text{-}30\%\text{-}o} = 0.4464,$$
$$S_{l\text{-}30\%\text{-}o} = 7.6249$$

Present results suggest that the ERM is preferable to the GA and RM for the search of a large number of descriptors. Table 3 shows a summary of the linear models with 1 to $d_{\text{opt}} + 1$ parameters for ERM and d_{opt} parameters for RM and GA, including the Leave-30%-Out Cross-Validation result. That the predictive power of the linear model is satisfactory is revealed by its stability upon the inclusion or exclusion of compounds. The resulting values $R_{loo} = 0.9609$ and $l\text{-}n\%\text{-}oR_{l\text{-}30\%\text{-}o} = 0.8793$ are in the range of a validated model: according to the literature $R_{l\text{-}n\%\text{-}o}$ must be greater than 0.71 [35]. The details of the molecular descriptors of Table 3 are presented in Table 4. The correlation matrix in Table 5 reveals that the descriptors of the linear model are not seriously inter-correlated ($R_{ij} < 0.5334$), and this fact substantiates the presence of all the parameters in the equation.

With the purpose of demonstrating that Eq. (7) does not result from happenstance we resort to a widely used approach to establish a model robustness: the so-called *y*-randomization [36] that consists of scrambling the experimental property **p** in such a way that activities do not correspond to the respective compounds. After analyzing 5,000,000 cases of *y*-randomization, the smallest S value that is obtained in this way $S = 0.5691$ is considerably greater than the one coming from the true calibration ($S = 0.2323$). We thus verify the robustness of the model and show that the calibration is not a fortuitous correlation but a reliable structure–activity relationship. This result together with the training and test statistical parameters are in good agreement with the premise that different polar protic solvents would not affect significantly the measure of k_t.

Table 3
QSPR models derived from the complete training set of $N = 41$ compounds. The best relationships found appear in boldface.

Model	Descriptors used	R	S	$R_{l\text{-}30\%\text{-}o}$	$S_{l\text{-}30\%\text{-}o}$
M1	*C-027*	0.664	0.699	0.195	1.099
M2	*nN*$^+$, *N-075*	0.864	0.477	0.628	0.762
M3	*nR05*, *nR10*, *N-075*	0.906	0.406	0.350	1.299
M4	*nR10*, *SRW05*, *MATS5p*, *N-075*	0.939	0.334	0.227	3.258
M5	*GATS8m*, *GATS1p*, *Mor16v*, *nN*$^+$, *nArOH*	0.953	0.300	0.789	0.610
M6	*GATS1e*, *Mor16u*, *E1v*, *R8m*$^+$, *nN*$^+$, *nArOH* (Eq. (7))	0.973	0.232	0.879	0.489
M7	*nR05*, *nR10*, *ATS8m*, *Mor29e*, *E1v*, *H5u*, *nArNH2*	0.980	0.204	0.509	2.709
M6B	*nR10*, *GATS1p*, *E1e*, *R8m*$^+$, *nN*$^+$, *N-075* (Eq. (8))	0.960	0.280	0.799	0.608
M6C	*C-032*, *R8m*$^+$, *nN*$^+$, *De*, *GATS1p*, *nR10* (Eq. (9))	0.968	0.251	0.446	7.625

Table 4
Symbols for molecular descriptors appearing in the different models.

Molecular descriptor	Type	Description
C-027	Atom-centred fragments	*C-027* corresponds to: R–CH–X
nN$^+$	Functional group counts	Number of ammonium groups (aliphatic)
N-075	Atom-centred fragments	*N-075* corresponds to: R–N–R/R–N–X
nR05	Constitutional descriptors	Number of 5-membered rings
nR10	Constitutional descriptors	Number of 10-membered rings
SRW05	Molecular walk counts	Self-returning walk count of
MATS5p	2D Autocorrelations	Moran autocorrelation – lag 5/weighted by atomic polarizabilities
GATS8m	2D Autocorrelations	Geary autocorrelation – lag 8/weighted by atomic masses
GATS1p	2D Autocorrelations	Geary autocorrelation – lag 1/weighted by atomic polarizabilities
Mor16v	3D-MoRSE	3D-MoRSE – signal 16/weighted by atomic van der Waals volumes
nArOH	Functional group counts	Number of aromatic hydroxyls
GATS1e	2D Autocorrelations	Geary autocorrelation – lag 1/weighted by atomic Sanderson electronegativities
Mor16u	3D-MoRSE	3D-MoRSE – signal 16/unweighted
E1v	WHIM	1st component accessibility directional WHIM index/weighted by atomic van der Waals volumes
R8m$^+$	GETAWAY	R maximal autocorrelation of lag 8/weighted by atomic masses
ATS8m	2D Autocorrelations	Broto-Moreau autocorrelation of a topological structure – lag 8/weighted by atomic masses
Mor29e	3D-MoRSE	3D-MoRSE – signal 29/weighted by atomic Sanderson electronegativities
H5u	GETAWAY	H autocorrelation of lag 5/unweighted
nArNH2	Functional Group Counts	Number of primary amines (aromatic)
E1e	WHIM	1st component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities
C-032	Atom-centred fragments	*C-032* corresponds to: X–CX–X
De	WHIM	D total accessibility index/weighted by atomic Sanderson electronegativities

For atom-centred fragments: R represents any group linked through carbon; X represents any electronegative atom (O, N, S, P, Se, halogens); – represents an aromatic bond as in benzene or delocalized bonds such as the N–O bond in a nitro group.

Table 5
Correlation matrix for the descriptors in Eq. (7) ($N = 41$).

	GATS1e	*Mor16u*	*E1v*	*R8m*$^+$	*nN*$^+$	nArOH
GATS1e	1	0.1651	0.3444	0.3685	0.3021	0.0160
Mor16u		1	0.5334	0.0566	0.2129	0.2327
E1v			1	0.2439	0.0863	0.3577
R8m$^+$				1	0.1114	0.2957
nN$^+$					1	0.1744
nArOH						1

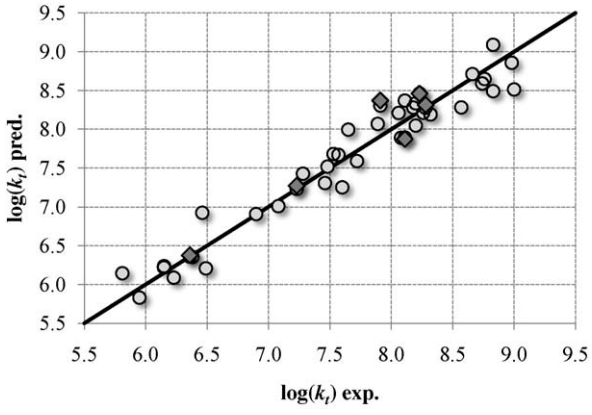

Fig. 2. Predicted versus experimental log (k_t). Results from Eq. (7) (circles) and from Eq. (10) for the test set (rhombus).

The plot of predicted vs. experimental log (k_t) shown in Fig. 2 suggests that the 41 heterocyclic compounds approximately follow a straight line. Table 1 also includes log (k_t) predicted by Eq. (7) for the set of molecules, and the corresponding residuals. Fig. 3 shows that the behavior of the residuals in terms of the predictions follows a normal distribution. No molecule in the set exhibited a residual larger than 2.5S that could be considered as outlier.

As an additional test of the predictive power of our method we remove six molecules (with varied values of the property) from the training set and calculate their log (k_t). To this end, we derive the following model with the remaining 35 molecules:

$$\begin{aligned}\log(k_t) = {} & 7.9238(\pm 0.5) - 1.3282(\pm 0.2)GATS1e \\ & - 1.1456(\pm 0.1)Mor16u + 5.7009(\pm 1)E1v \\ & - 32.2832(\pm 4.3)R8m^+ - 1.4369(\pm 0.1)nN^+ \\ & - 1.1991(\pm 0.1)nArOH \end{aligned} \quad (10)$$

$$N = 35, \quad R = 0.9749, \quad S = 0.2352,$$
$$FIT = 7.5683, \quad p < 10^{-5} \quad RMSE_{TestSet} = 0.2325$$

Its parameters are similar to those in Eq. (7), and $RMSE_{TestSet}$ stands for root mean squared errors of the calculated values of the property for the test set of 6 molecules. Figs. 2 and 3 show that the plot of predicted vs. experimental log (k_t) for the 6 heterocyclic

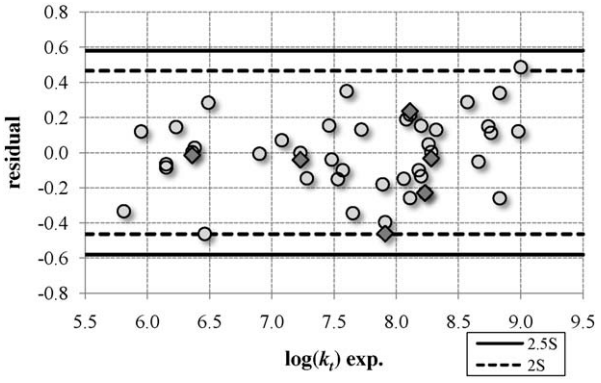

Fig. 3. Dispersion plot of the residuals for the training set (Eq. (7), circles) and test set (Eq. (10), rhombus).

compounds approximately follows a straight line. Besides, no molecule in the test set can be considered as outlier.

With the main purpose of assessing whether the selected test set of compounds can be used as a true external validation group for Eq. (10), we carry out a new ERM variable search for the training set of 35 molecules defined previously. The best model found in this case is almost identical to Eq. (7) (based on the complete set of 41 molecules) and with similar statistical quality for both the training and test sets. The only difference is that the descriptor *GATS1v* replaces *GATS1e*. However, these two descriptors have the same nature, are calculated with the same length of path connecting the atoms and have similar values. In consequence, as an increment of the number of data points enables to obtain more trustful QSPR models, we recommend employing Eq. (7) for any further prediction of the investigated property because it encodes experimental information from all the 41 molecules.

The molecular descriptors appearing in the linear Eq. (7) combine several two- and three-dimensional aspects of the molecular structure, and can be classified as follows: (i) a 2D Autocorrelation: *GATS1e*, Geary autocorrelation – lag 1/weighted by atomic Sanderson electronegativities; (ii) a 3D-MoRSE descriptor: *Mor16u*, 3D-MoRSE – signal 16/unweighted; (iii) a WHIM descriptor: *E1v*, 1st component accessibility directional WHIM index/weighted by atomic van der Waals volumes; (iv) a GETAWAY descriptor: *R8m*$^+$, R maximal autocorrelation of lag 8/ weighted by atomic masses; finally, two Functional Group Counts: *nN*$^+$, number of ammonium groups (aliphatic) and *nArOH*, number of aromatic hydroxyls.

The different structural variables introduced by Broto, Moreau, and Geary [37,38] correspond to bi-dimensional autocorrelations between pairs of atoms in the molecule, and have been defined in order to reflect the contribution of a considered atomic property to the experimental observations under investigation. The atomic properties that can be adopted to differentiate the nature of atoms are the mass (m), polarizability (p), electronegativity (e) or the volume (v). These indices can be readily calculated, i.e.: by summing products of atomic weights (employing atomic properties such as atomic polarizabilities, molecular volumes, etc.) of the terminal atoms of all the paths of a prescribed length. For the case of *GATS1e*, the path connecting a pair of atoms has length 1 and involves the atomic Sanderson electronegativities as weighting scheme to distinguish their nature.

The 3D-MoRSE (3D Molecule Representation of Structure based on Electron diffraction) descriptors [39,40] provide 3D information from the three-dimensional structure of a molecule using a molecular transform derived from an equation used in electron diffraction studies. Various atomic properties can be taken into account giving high flexibility to this representation of a molecule. The simplified form of the transform is

$$I(s) = \sum_{i=2}^{N}\sum_{j=1}^{i-1} A_i A_j \frac{\sin sr_{ij}}{sr_{ij}} \qquad (11)$$
$$s = 0, \ldots, 31.0\,\text{Å}^{-1}$$

where N is the number of atoms; r_{ij} is the distance between atoms i and j; A_i can be any atomic property of atom i such as atomic number, mass, partial atomic charge, or atomic polarizability; s is a reciprocal distance. The value of s is considered only at discrete positions within a certain range. Normally 32 equidistant values between 0 and 31 Å^{-1} are chosen. The choice of the range of s and the number of values to be considered determines the resolution of the code for representing the 3D structure. For the case of *Mor16u*, an unweighting scheme is used and s is equal to 15 Å^{-1}.

WHIM (Weighted Holistic Invariant Molecular Descriptors) descriptors [41] are based on statistical indices calculated on the projections of atoms along principal axes. The aim is to capture 3D information regarding size, shape, symmetry and atom distributions with respect to invariant reference frames. To calculate them, a weighted covariance matrix is obtained from different weighting schemes for the atoms: the unweighted case, atomic mass, van der Waals volume, Sanderson atomic electronegativity, atomic polarizability and electrotopological state indices. Depending on the weighting scheme different covariances matrices and hence different principal axes are obtained. Essentially the WHIM descriptors provide a variety of principal axes with respect to a defined atomic property. For each weighting scheme, a set of statistical indices is calculated on the atoms projected onto the principal axes (i.e. principal components). Descriptor *E1v* is a first component accessibility directional WHIM descriptor that involves the van der Waals volume as weighting scheme. These types of descriptors are univariate statistical indices calculated on the scores of the individual principal components.

The GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) type of descriptors [42] have been designed with the main purpose of matching the 3D-molecular geometry. These numerical variables are derived from the elements h_{ij} of the Molecular Influence matrix (**H**), obtained through the values of atomic Cartesian coordinates. The diagonal elements of **H** (h_{ii}) are called leverages, and are considered to represent the influence of each atom on the shape of the molecule. For instance, the mantle atoms always have higher h_{ii} values than atoms near the molecule center, while each off-diagonal element h_{ij} represents the degree of accessibility of the jth atom to interactions with the ith atom. The Influence/Distance matrix (**R**) involves a combination of the elements of **H** matrix with those of the Geometric Matrix (**G**). Descriptor *R8m*$^+$ involved in Eq. (7) is of the R-GETAWAY type, and represents an **R** index of maximal contribution to the autocorrelation in lag 8 (topological distance) and involves the atomic masses as weighting scheme to distinguish their nature.

Functional Group Counts are, as their name indicates, molecular descriptors based on the number of chemical functional groups. They are normally relevant descriptors since the number of a certain functional group will evidently affect the properties of the molecule.

Now, by means of a proper orthogonalization of Eq. (7) and subsequent standardization [26] of the orthogonal regression coefficients, it is feasible to assign a greater importance to the molecular descriptors that exhibit larger (absolute) standardized coefficients. As mentioned previously, the orthogonalization order is chosen in such a way that maximises the correlation coefficient between each calculated orthogonal descriptor and the experimental deactivation rate constants (in decreasing order). In this way, one expects to construct a model based on hierarchical and independent contributions of structural variables. The calculated

average importance of the involved structural descriptors of Eq. (7) is the next one:

$$nN^{+}(0.61) > nArOH(0.57) > R8m^{+}(0.27) > Mor16u(0.26) > E1v(0.24) > GATS1e(0.21) \quad (12)$$

where the standardized regression coefficients are shown in parentheses. Although the inequality of (12) is true, it is statistically derived. For this reason, one can deduce that the resulting effect of the 6 involved descriptors on the property being analyzed depends upon the combination of all of them. However, it is possible to deduce tendencies for individual descriptors, according to the order of importance as expressed by inequality (12). The ranking of contributions given by this inequality suggests that the functional group counts, the number of aliphatic ammonium groups and the number of aromatic hydroxyls are the most relevant parameters for the chosen set of compounds, due to their standardized coefficients of 0.61 and 0.57, respectively. Such kind of descriptors have a quite direct interpretation on the physical property being analyzed, since it is expected that compounds which do not have ammonium groups nor aromatic hydroxyls would tend to display higher experimental total deactivation rate constant of singlet oxygen.

It has to be noticed that all the molecular descriptors appearing in inequality (12) are positive numerical variables. Therefore, considering the sign of the regression coefficients in inequality (12) (all negative with exception for E1v), it is expected that molecular structures displaying higher positive values for *E*1v and lower positive values for descriptors nN^{+}, *nArOH*, $R8m^{+}$, *Mor*16*u* and *GATS*1*e* would elicit higher predicted log (k_t) values. The fact that lower values of nN^{+} and *nArOH* lead to greater values of log (k_t) is in agreement with the experimental observation. This tendency about the effect of numerical values of descriptors on the predicted property can be demonstrated, for instance, by the compounds having the lowest and highest log (k_t) values in the training set, which are 33 (log (k_t) = 5.81) and 40 (log (k_t) = 9.00) (for further comparisons among compounds, refer to Table 6 which includes the numerical values for the mentioned theoretical descriptors).

Regarding the effect of the remaining descriptors appearing in Eq. (7), it can be concluded that these numerical variables has a secondary importance and smaller contribution on determining the variation of log (k_t) values, owing to their smaller standardized orthogonal coefficients. Their role in the QSAR is to help improving the prediction of the property in a somewhat better extent. Despite of this fact, we also provide an interpretation for the effect of these descriptors on the deactivation rate constants. Lower values for $R8m^{+}$, the R maximal autocorrelation of lag 8/weighted by atomic masses, means that the deactivation constants would tend to increase for compounds having lower molecular weights, as this particular geometry based descriptor considers the distribution of atomic masses located at topological distances of length 8. Lower values for *Mor16u*, the 3D-MoRSE – signal 16/unweighted, means that a lower molecular size of compounds would tend to increase log (k_t) values because of the dependence of this unweighted descriptor on the magnitude of interatomic distances r_{ij}. The WHIM descriptor *E*1*v*, 1st component accessibility directional WHIM index/weighted by atomic van der Waals volumes, is an index that contemplates mixed 3D information regarding size, shape, symmetry and atomic distributions. As previously stated, higher positive values for *E*1*v* would lead to higher values of the predicted property. Finally, the Geary autocorrelation – lag 1/weighted by atomic Sanderson electronegativities *GATS1e* describes the distribution of atomic electronegativities along paths connecting atom pairs of length 1. This descriptor characterizes the importance of charge in electrostatic interactions during the singlet oxygen deactivation process. Lower positive values of this charge distribution measure would tend to lead to higher predicted log (k_t).

Table 6
Values for descriptors appearing in the QSPR designed together with experimental log (k_t) values.

Number	log (k_t) exp.	nN^{+}	*nArOH*	$R8m^{+}$	*Mor*16*u*	*E*1v	*GATS*1*e*
1	8.74	0	0	0.013	0.530	0.502	0.847
2	8.57	0	0	0.023	−0.086	0.401	0.934
3	8.66	0	0	0.033	−0.803	0.397	0.953
4	8.32	0	0	0.023	−0.180	0.382	1.008
5	8.28	0	0	0.019	−0.058	0.382	0.934
6	8.83	0	0	0.004	−0.398	0.399	1.062
7	6.46	0	1	0	0.179	0.324	1.105
8	6.90	0	1	0.003	0.168	0.325	1.053
9	7.60	0	1	0.003	0.184	0.381	1.013
10	6.49	0	1	0.025	0.260	0.338	1.008
11	6.15	0	1	0.028	0.132	0.335	1.008
12	6.15	0	1	0.025	0.113	0.312	1.016
13	6.38	0	1	0.029	0.218	0.363	0.934
14	6.36	0	1	0.032	0.149	0.372	0.953
15	7.48	0	1	0.013	0.389	0.493	0.847
16	8.06	0	0	0	0.399	0.401	1.167
17	7.46	0	0	0	0.435	0.283	1.333
18	8.11	0	0	0	0.400	0.363	1.250
19	7.65	0	0	0	0.575	0.339	0.900
20	8.76	0	0	0	0.208	0.386	0.917
21	8.20	0	0	0	0.377	0.362	0.909
22	8.83	0	0	0.004	0.670	0.405	0.603
23	7.89	0	0	0.019	0.510	0.390	0.621
24	7.91	0	0	0.022	0.416	0.426	0.598
25	8.20	0	0	0.010	0.920	0.486	0.941
26	7.53	0	0	0.010	1.148	0.467	0.944
27	7.23	0	0	0.032	1.158	0.493	0.830
28	7.28	0	0	0.034	1.139	0.536	0.830
29	7.72	0	0	0.011	0.919	0.423	1.002
30	7.57	0	0	0.009	1.091	0.448	0.946
31	7.08	1	0	0.011	0.893	0.466	0.517
32	6.23	1	0	0.009	0.590	0.421	1.393
33	5.81	1	0	0	0.351	0.354	1.500
34	5.95	1	0	0.009	0.742	0.406	1.393
35	8.11	0	0	0.002	0.105	0.387	1.188
36	8.23	0	0	0.002	0.228	0.428	1.188
37	8.08	0	0	0	0.337	0.405	1.500
38	8.18	0	0	0	0.519	0.401	1.000
39	8.26	0	0	0	0.425	0.370	1.000
40	9.00	0	0	0	0.172	0.353	0.909
41	8.98	0	0	0	0.007	0.386	0.929

By means of Eq. (7) we estimate the rate constant k_t for a group of heterocyclic compounds named β-carbolines (Fig. 4). Despite the fact that they have been suggested as possible antioxidants [43], to the best of our knowledge there is no experimental (k_t) data for them. In aqueous solution, β-carbolines show an acid–base equilibrium with pK_a around 7. Due to the fact that β-carbolines are present in a great number of living systems, with a physiological pH around 7, we decide to calculate the k_t for each acid–base form. The results are shown in Table 7. Our calculation suggests that k_t increases with the electronic activation of the ring of the β-carboline moieties. Therefore, the electrophilic attack should be higher on the neutral form of each $^1O_2$ β-carboline rather than on its cationic form.

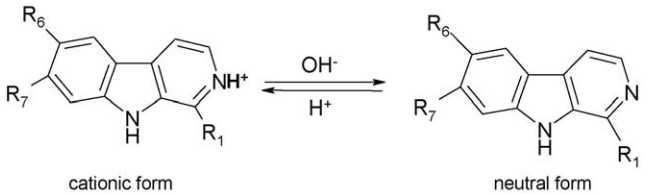

Fig. 4. Structure of β-carbolines acid–base equilibrium in aqueous solution.

Table 7
Log (k_t) predicted by Eq. (7) for the group of molecules with unknown experimental values. The plus sign on the compound name indicates that it is a cation in the acid form.

	Name	log (k_t) pred.	k_t (L mol^{-1} s^{-1}).
42	Norharmane	8.69	4.90E+08
43	Norharmane$^+$	7.01	1.03E+07
44	Harmane	8.65	4.50E+08
45	Harmane$^+$	7.12	1.32E+07
46	Harmine	8.30	1.99E+08
47	Harmine$^+$	6.69	4.89E+06
48	Harmaline	8.17	1.47E+08
49	Harmaline$^+$	6.46	2.90E+06
50	Harmol	7.65	4.42E+07
51	Harmol$^+$	6.16	1.43E+06

4. Conclusions

By means of our searching algorithm ERM we construct a QSPR model for the prediction of the rate constant of deactivation of singlet oxygen by heterocyclic compounds. In this study we choose 41 such compounds and find six molecular descriptors that take into account some 2D- and 3D-aspects of the molecular structure. Our results suggest that the ERM is preferable to the RM and GA. The recently developed strategy to determine the optimal number of descriptors is used successfully. Using the QSPR model we estimate the unknown k_t for a group of heterocyclic compounds, which have been suggested as antioxidants, named β-carbolines. The results imply that k_t increases with the electronic activation of the ring of the β-carboline moieties.

Acknowledgement

This work is supported by the National Council for Scientific and Technological Research (CONICET).

References

[1] E.L. Clennana, A. Paceb, Advances in singlet oxygen chemistry, Tetrahedron 61 (2005) 6665–6691.
[2] C. Schweitzer, R. Schmidt, Physical mechanisms of generation and deactivation of singlet oxygen, Chem. Rev. 103 (2003) 1685–1757.
[3] C.S. Foote, E.L. Clennan, Properties and Reactions of Singlet Dioxygen in Active Oxygen in Chemistry, Chapman & Hall, New York, 1995.
[4] M.C. DeRosa, R.J. Crutchley, Photosensitized singlet oxygen and its applications, Coord. Chem. Rev. 233–234 (2002) 351–371.
[5] K. Briviba, L.O. Klotz, H. Sies, Toxic and signaling effects of photochemically or chemically generated singlet oxygen in biological systems, Biol. Chem. 378 (1997) 1259–1265.
[6] E. Cadenas, Biochemistry of oxygen toxicity, Annu. Rev. Biochem. 58 (1989) 79–110.
[7] A.U. Kahn, Direct spectroscopic observation of 1.27 mm and 1.58 mm emission of singlet ($^1\Delta_g$) molecular oxygen in chemically generated and dye-photosensitized liquid solutions at room temperature, Chem. Phys. Lett. 72 (1980) 112–114.
[8] A.H. Thomas, C. Lorente, A.L. Capparelli, C.G. Martínez, A.M. Braun, E. Oliveros, Singlet oxygen ($^1\Delta_g$) production by pterin derivatives in aqueous solutions, Photochem. Photobiol. Sci. 2 (2003) 245–250.
[9] C. Lorente, A.H. Thomas, Photophysics and photochemistry of pterins in aqueous solution, Acc. Chem. Res. 39 (2006) 395–402.
[10] F.M. Cabrerizo, M.L. Dántola, G. Petroselli, A.L. Capparelli, A.H. Thomas, A.M. Braun, C. Lorente, E. Oliveros, Reactivity of conjugated and unconjugated pterins with singlet oxygen ($O_2(^1\Delta_g)$): physical quenching and chemical reaction, Photochem. Photobiol. 83 (2007) 526–534.
[11] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, D.C., 1995.
[12] N. Trinajstic, Chemical Graph Theory, CRC Press, Boca Raton, FL, 1992.
[13] A.R. Katritzky, V.S. Lobanov, M. Karelson, Chem. Soc. Rev. 24 (1995) 279–287.
[14] M.L. Dantola, A.H. Thomas, A.M. Braun, E. Oliveros, C. Lorente, Singlet oxygen ($O_2(^1\Delta_g)$) quenching by dihydropterins, J. Phys. Chem. A 111 (2007) 4280–4288.
[15] F. Wilkinson, P.W. Helman, A.B. Ross, Rate constants for the decay and reactions of the lowest electronically excited singlet state of molecular oxygen in solution. An expanded and revised compilation, J. Phys. Chem. Ref. Data 24 (1995) 663–677.
[16] A.G. Mercader, P.R. Duchowicz, F.M. Fernandez, E.A. Castro, Modified and enhanced replacement method for the selection of molecular descriptors in QSAR and QSPR theories, Chemom. Intell. Lab Syst. 92 (2008) 138–144.
[17] P.R. Duchowicz, E.A. Castro, F.M. Fernández, M.P. González, A new search algorithm of QSPR/QSAR theories: normal boiling points of some organic molecules, Chem. Phys. Lett. 412 (2005) 376–380.
[18] P.R. Duchowicz, E.A. Castro, F.M. Fernández, Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies, MATCH Commun. Math. Comput. Chem. 55 (2006) 179–192.
[19] P.R. Duchowicz, M. Fernández, J. Caballero, E.A. Castro, F.M. Fernández, QSAR of non-nucleoside inhibitors of hiv-1 reverse transcriptase, Bioorg. Med. Chem. 14 (2006) 5876–5889.
[20] A.M. Helguera, P.R. Duchowicz, M.A.C. Pérez, E.A. Castro, M.N.D.S. Cordeiro, M.P. González, Application of the replacement method as novel variable selection strategy in QSAR 1. Carcinogenic potential, Chemometr. Intell. Lab. 81 (2006) 180–187.
[21] S.S. So, M. Karplus, Evolutionary optimization in quantitative structure–activity relationship: an application of genetic neural networks, J. Med. Chem. 39 (1996) 1521–1530.
[22] HYPERCHEM, 6.03 (Hypercube) http://www.hyper.com.
[23] e-Dragon, Electronic remote version of Dragon 5.4, <http://www.vcclab.org/lab/edragon/>.
[24] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley, VCH, Weinheim, Germany, 2000.
[25] Matlab, 5.0 The MathWorks Inc. http://www.mathworks.com/.
[26] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, New York, 1981.
[27] M. Melanie, An Introduction to Genetic Algorithms Cambridge, Massachusetts, A Bradford Book The MIT Press, London, England, 1998.
[28] H. Kubinyi, Variable selection in QSAR studies. I. An evolutionary algorithm, Quant. Struct. Act Relat. 13 (1994) 285–294.
[29] H. Kubinyi, Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution, Quant. Struct. Act Relat. 13 (1994) 393–401.
[30] A.G. Mercader, P.R. Duchowicz, F.M. Fernández, E.A. Castro, E. Wolcan, QSPR study of solvent quenching of the $^5D_0 \rightarrow {}^7F_2$ emission of Eu(6,6,7,7,8,8,8-heptafluoro-2,2-dimethyl-3, 5-octanedionate)$_3$, Chem. Phys. Lett. 462 (2008) 352–357.
[31] C. Hansch, Comprehensive Drug Design, Pergamon Press, New York, 1990.
[32] D.M. Hawkins, S.C. Basak, D. Mills, Assessing model fit by cross validation, J. Chem. Inf. Model. 43 (2003) 579–586.
[33] M. Randic, J. Chem. Inf. Model. 31 (1991) 311–320.
[34] M. Randic, New J. Chem. 15 (1991) 517–525.
[35] A. Golbraikh, A. Tropsha, Beware of q2! J. Mol. Graphics Model. 20 (2002) 269–276.
[36] S. Wold, L. Eriksson, Chemometrics Methods in Molecular Design, VCH, Weinheim, 1995.
[37] G. Moreau, P. Broto, Nouv. J. Chim. 4 (1980) 757–764.
[38] G. Moreau, P. Broto, Nouv. J. Chim. 4 (1980) 359–360.
[39] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, Chemical information in 3d space, J. Chem. Inf. Comput. Sci. 36 (1996) 1030–1037.
[40] J.H. Schuur, P. Selzer, J. Gasteiger, The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure–spectra correlations and studies of biological activity, J. Chem. Inf. Comput. Sci. 36 (1996) 334–344.
[41] R. Todeschini, P. Gramatica, Sd-modelling and prediction by whim descriptors. Part 5. Theory development and chemical meaning of whim descriptors, Quant. Struct.–Act. Relat. 16 (1997) 113–119.
[42] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by getaway descriptors. 2. Application of the novel 3d molecular descriptors to QSAR/QSPR studies, J. Chem. Inf. Model. 42 (2002) 693.
[43] K.C. Pari, S. Sundari, S. Chandani, D. Balasubramanian, β-carbolines that accumulate in human tissues may serve a protective role against oxidative stress*, J. Biol. Chem. 275 (2000) 2455–2462.