SAR and QSAR in Environmental Research

ISSN: 1062-936X (Print) 1029-046X (Online) Journal homepage: www.tandfonline.com/journals/gsar20

4D-QSAR investigation and pharmacophore identification of pyrrolo[2,1-c][1,4]benzodiazepines using electron conformational–genetic algorithm method

A. Özalp, S. Ç. Yavuz, N. Sabancı, F. Çopur, Z. Kökbudak & E. Sarıpınar

To cite this article: A. Özalp, S. Ç. Yavuz, N. Sabancı, F. Çopur, Z. Kökbudak & E. Sarıpınar (2016) 4D-QSAR investigation and pharmacophore identification of pyrrolo[2,1-c][1,4]benzodiazepines using electron conformational–genetic algorithm method, SAR and QSAR in Environmental Research, 27:4, 317-342, DOI: 10.1080/1062936X.2016.1174152

To link to this article: https://doi.org/10.1080/1062936X.2016.1174152

SAR AND QSAR IN ENVIRONMENTAL RESEARCH, 2016
VOL. 27, NO. 4, 317–342
http://dx.doi.org/10.1080/1062936X.2016.1174152

Taylor & Francis
Taylor & Francis Group

4D-QSAR investigation and pharmacophore identification of pyrrolo[2,1-c][1,4]benzodiazepines using electron conformational–genetic algorithm method

A. Özalp[a], S. Ç. Yavuz[a], N. Sabancı[b], F. Çopur[c], Z. Kökbudak[a] and E. Sarıpınar[a]

[a]Department of Chemistry, Science Faculty, Erciyes University, Kayseri, Turkey; [b]Department of Chemistry, Science and Arts Faculty, Siirt University, Siirt, Turkey; [c]Information Technology Department, Abdullah Gül University, Kayseri, Turkey

ABSTRACT

In this paper, we present the results of pharmacophore identification and bioactivity prediction for pyrrolo[2,1-c][1,4]benzodiazepine derivatives using the electron conformational–genetic algorithm (EC–GA) method as 4D-QSAR analysis. Using the data obtained from quantum chemical calculations at PM3/HF level, the electron conformational matrices of congruity (ECMC) were constructed by EMRE software. The ECMC of the lowest energy conformer of the compound with the highest activity was chosen as the template and compared with the ECMCs of the lowest energy conformer of the other compounds within given tolerances to reveal the electron conformational submatrix of activity (ECSA, i.e. pharmacophore) by ECSP software. A descriptor pool was generated taking into account the obtained pharmacophore. To predict the theoretical activity and select the best subset of variables affecting bioactivities, the nonlinear least square regression method and genetic algorithm were performed. For four types of activity including the GI_{50}, TGI, LC_{50} and IC_{50} of the pyrrolo[2,1-c][1,4] benzodiazepine series, the r^2_{train}, r^2_{test} and q^2 values were 0.858, 0.810, 0.771; 0.853, 0.848, 0.787; 0.703, 0.787, 0.600; and 0.776, 0.722, 0.687, respectively.

ARTICLE HISTORY
Received 22 November 2015
Accepted 30 March 2016

KEYWORDS
Electron conformational–genetic algorithm; pyrrolo[2,1-c][1,4] benzodiazepines; 4D-QSAR, pharmacophore; genetic algorithm; electron conformational method

Introduction

Many of the clinically potent anticancer agents directly target DNA to show their antitumour effects [1]. In recent years, there has been an increasing interest in DNA interactive ligands which can bind to DNA to achieve the required sequence selectivity. As a gene-targeted ligand, naturally occurring pyrrolo[2,1-c][1, 4]benzodiazepines (PBDs) showing antibiotic and antitumour effects are derived from the fermentation broth of various *Streptomyces* species, well-known members of which include anthramycin, tomaymycin, sibiromycin and DC-81 [2–4].

As the most reliable and cited approach, quantitative structure–activity relationships (QSARs) have been utilized to correlate the biological activities of a compound library and

CONTACT E. Sarıpınar emin@erciyes.edu.tr

its structural/molecular information in drug design [5]. These relationships form statistical models, which help in the development of new bioactive chemical compounds by predicting the biological activities as a function of molecular descriptors. In accordance with the similarity-property principle, since structurally analogous molecules have a tendency to produce similar biological activities [6], structural information is encoded by molecular descriptors including the physicochemical properties of a ligand molecule.

Starting from the traditional 2D-QSAR studies in which only physicochemical properties were handled to predict the biological activities of related compounds [7,8], over the last 50 years a number of QSAR methodologies indicating different dimensions of QSARs from 2D to nD and containing conformation-dependent 3D [9,10], 4D with conformational Boltzmann sampling [11], 5D with induced-fit hypotheses [12], 6D with multiple solvation models [13] and 7D (target-based receptor model data 7D) [14] have been developed to overcome previous limitations.

Although there have been several studies related to the design, synthesis and biological evaluation of pyrrolo[2,1-c][1, 4]benzodiazepines [15–19], structure–activity relationship studies are not available, except for that of Antonow and co-workers who presented a broad SAR examination of monomeric C2-aryl PBDs as antitumour agents [20]. In the mentioned paper, SAR studies were conducted on four different cytotoxicity parameters (GI_{50}, TGI, LC_{50} and IC_{50}) of 80 analogues containing a wide range of substituents at the C2-position of the PBD. The GI_{50} is the concentration that causes 50% decrease in the cell growth. The TGI is the concentration that totally inactivates the growth of cells. The LC_{50} is the median lethal concentration that is expected to kill 50% of organisms in a given population. The IC_{50} is the median concentration of a drug that causes 50% inhibition.

Based on the experimental biological activity data of Antonow et al. [20], and taking into account all possible molecular conformations subject to Boltzmann distribution, we applied a 4D-QSAR study to explore a better understanding of C2-aryl PBDs using the EC–GA method developed by Sarıpınar et al. [21].

Following the electron topological method (ET) as a pharmacophore identification process [22], the electron conformational (EC) method consisting of both pharmacophore identification and bioactivity prediction was developed by Bersuker [23]. This method, which has many applications in the literature, is based on a triangle ET matrix of geometric and electronic features for the pharmacophore generation procedure and a nonlinear equation for the prediction of bioactivity considering the most probable conformer of each compound. The most important limitation of the EC method is the dependence on only one conformer of each related compound, so that it overlooks all possible conformers owing to the complexity of the nonlinear equation. Detailed information about this method is available in Bersuker et al. [24].

Genetic algorithms (GAs), which have found many applications in QSAR analysis, offer a consistent, efficient and meaningful way to explore a large space and to construct predictive and robust models among a large number of descriptors. By specifically working with a large number of descriptors, the GA overcomes this complexity [25].

The EC–GA method, first introduced in 2010, was generated as an integrated method of the EC method and the GA [21]. In contrast to the EC method and many other methods in which model construction is based on physicochemical information for a single molecular conformation ignoring the highly populated conformational space (except the lowest energy conformer), this versatile EC–GA method considers all low-energy Boltzmann weighted

conformations knowing that a number of low-energy conformations are available at room temperature for a molecule and each low-energy conformer produces a considerable effect on biological activity and contributes to model power. In this method, the biological activity prediction and pharmacophore identification are performed as a function of physicochemical and structural descriptors for a set of low-energy conformers of each compound, instead of a single lowest energy conformation. To establish a meaningful and predictive QSAR model, it is crucial to select the best subset of molecular descriptors in the optimum number. Here the GA optimization technique is used for descriptor selection. The final model is cross-validated by the leave-one-out cross-validation (LOO–CV) method. As a promising 4D-QSAR approach the EC–GA method, which provides pharmacophore detection, variable selection and quantitative bioactivity prediction, was performed employing C2-aryl PBD derivatives for four types of biological activities.

Materials and methods

C2-aryl PBD derivatives were analysed by the EC–GA method to distinguish the pharmacophore group and to derive a relationship between biological activities and selected molecular parameters. Detailed information about the methodology can be found in the literature [26–30]. The GI_{50}, TGI and LC_{50} activity values for compound 17 and 38 are not given in Table 1 since they were not determined experimentally.

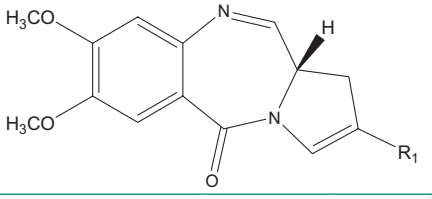
The structure of relevant compounds and their experimental biological activities including the GI_{50}, TGI, LC_{50} and IC_{50} values obtained from the literature are given in Table 1. The concentrations which are in μM were converted to a negative logarithmic scale which allows us to better handle the numbers.

Spartan 10 [31] software was used for the construction of the 3D structures of compounds, conformational analysis and quantum chemical calculations at Hartree Fock 3-21 G* level. Even though the more complicated basis sets give more accurate results, they expend a great deal of computation time. In case of a large number of compounds and conformations, as in this study, the required computation time increases due to much larger basis sets. Accordingly, we have considered the basis set 3-21G*, which is faster and sufficiently small without compromising the required level of accuracy. Water was used as solvent since it is the most similar solvent to biological systems. Following the conformational search of each molecule, conformers with Boltzmann distribution under 1/10000 were excluded, and higher ones were kept.

Mulliken charges and bond orders/interatomic distances were utilized to generate the electron conformational matrix of congruity (ECMCs) for individual conformations of the entire compound set and placed in diagonal and non-diagonal positions, respectively. Non-diagonal elements are of two types: bond orders for chemically bonded atom pairs and interatomic distances for non-bonded atom pairs [32]. An example of ECMC is illustrated for the lowest energy conformer of compound 63 as the template in Figure 1. For 87 analogues of C2-aryl 1, 4PBD derivatives, 997 ECMCs were created to be used in the comparison of the ECMCs by EMRE software [26–30] after eliminating the conformers which overlap and have lower Boltzmann distribution.

Of all the conformers of individual compounds, the lowest energy conformer of the most active one was chosen as template. The compounds were categorized as active and inactive by indicating a proper activity threshold value which is based on the data of the activity range for each type of activity. Up to a specified tolerance value, by adjusting the tolerance

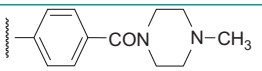
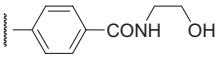
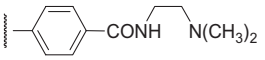
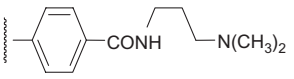
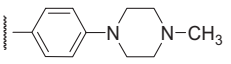
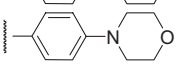
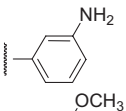
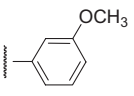
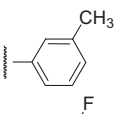
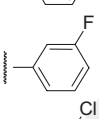
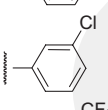
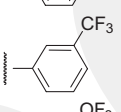
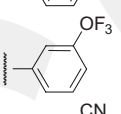
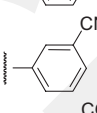
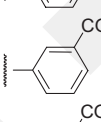
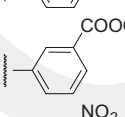
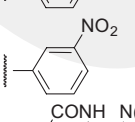
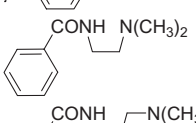
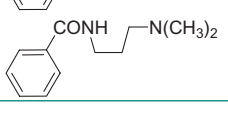
Table 1. Chemical structures, substituents and experimental pGI_{50}, pTGI, pLC_{50} and pIC_{50} activity values for C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine derivatives.



No	R_1	pGI_{50}	*pTGI*	pLC_{50}	pIC_{50}
1	H	8.699	7.398	5.290	8.602
2	$N(CH_3)_2$	8.523	7.398	5.491	7.493
3	NH_2	8.699	7.523	5.320	7.854
4	NHAc	8.222	6.699	4.939	7.527
5	OH	8.699	7.097	5.470	7.987
6	OCH_3	9.000	6.824	4.721	8.310
7	OPh	8.523	7.523	5.900	7.292
8	$C(CH_3)_3$	8.699	7.301	5.712	7.321
9	CH_3	8.699	6.745	5.051	8.553
10	CH_2CH_3	9.000	7.155	5.380	7.100
11	$CH(CH_3)_2$	8.398	8.222	5.351	6.939
12	$CHCH_2$	8.301	8.000	6.208	7.161
13	Ph	8.097	8.046	5.440	7.708
14	SCH_3	8.301	7.699	5.842	7.721
15	F	8.699	6.921	4.951	8.796
16	Cl	8.699	7.155	5.350	7.504
17	Br	-	-	-	6.626
18	CHO	8.046	6.796	4.821	7.580
19	CF_3	8.398	7.523	5.780	7.614
20	CN	8.523	6.824	5.130	7.703
21	COOH	4.879	4.631	4.600	6.000
22	NO_2	8.398	7.155	5.160	7.883
23	$CONH_2$	8.046	6.337	4.860	7.807

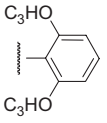
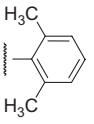
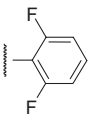
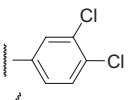
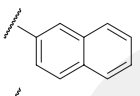
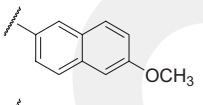
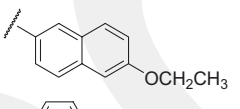
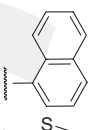
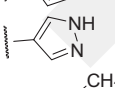
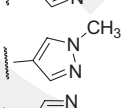
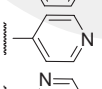
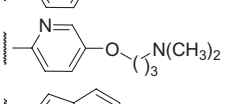
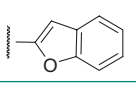
(*Continued*)

Table 1. (*Continued*)

No	R_1	pGI_{50}	*pTGI*	pLC_{50}	pIC_{50}
24		7.824	6.149	4.860	7.225
25		6.801	5.730	4.900	6.794
26		7.721	7.046	5.410	6.943
27		7.237	6.081	4.450	6.783
28		8.301	7.097	5.080	9.201
29		8.301	6.796	4.879	8.056
30		8.398	7.301	5.230	7.225
31		7.959	7.222	5.390	9.347
32		8.301	7.155	5.120	8.022
33		8.523	7.301	4.780	8.194
34		7.770	7.000	6.268	7.712
35		8.222	7.523	5.870	7.330
36		8.222	6.745	5.140	7.116
37		8.398	6.252	4.790	7.821
38		-	-	-	6.000
39		8.301	7.398	5.120	7.907
40		8.301	7.301	5.250	7.236
41		7.569	5.730	4.979	6.924
42		6.991	6.071	4.851	6.564

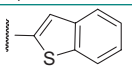
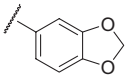
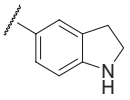
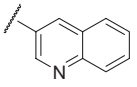
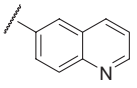
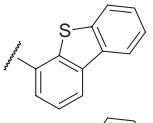
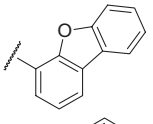
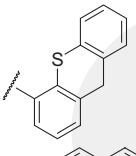
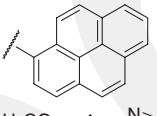
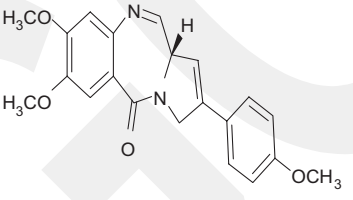
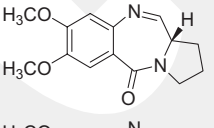
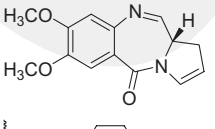
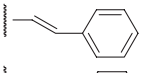
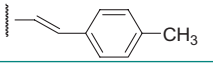
(*Continued*)

Table 1. (*Continued*)

No	*R_1*	*pGI_{50}*	*pTGI*	*pLC_{50}*	*pIC_{50}*
43		7.921	6.602	4.971	8.032
44		7.796	6.796	5.361	6.588
45		6.959	5.959	4.932	7.215
46		6.070	5.461	4.680	6.000
47		7.602	6.481	5.361	8.149
48		8.523	7.301	5.520	6.975
49		9.222	7.824	6.398	7.343
50		9.046	8.155	6.569	8.638
51		8.699	7.699	5.710	7.900
52		6.860	6.022	4.857	6.242
53		8.000	6.824	6.131	8.959
54		7.921	6.569	5.190	7.697
55		8.523	7.301	5.270	7.914
56		8.301	7.155	4.770	7.740
57		7.745	6.222	4.570	7.542
58		8.398	7.222	5.390	8.027
59		8.222	7.301	6.357	7.967

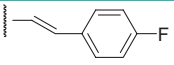
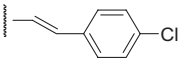
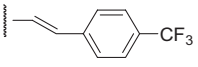
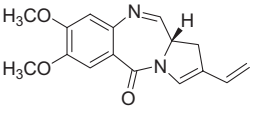
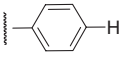
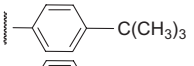
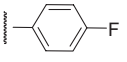
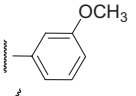
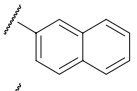
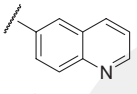
(*Continued*)

Table 1. (*Continued*)

No	R_1	pGI_{50}	*pTGI*	pLC_{50}	pIC_{50}
60		8.301	8.046	6.268	7.573
61		8.699	7.523	5.440	9.328
62		8.301	7.398	5.150	7.900
63		10.000	9.000	6.745	8.051
64		9.398	8.398	6.721	8.854
65		7.602	6.658	5.270	6.787
66		7.310	6.678	5.550	6.792
67		6.300	5.759	5.050	6.000
68		6.400	5.842	4.971	7.064
69		6.991	6.347	6.041	6.027
70		5.670	4.721	4.240	6.312
72		5.010	4.360	4.040	6.000
73		5.870	5.090	4.220	7.842
74		8.398	7.699	5.801	10.000

(*Continued*)

Table 1. (*Continued*)

No	R_1	pGI_{50}	*pTGI*	pLC_{50}	pIC_{50}
75		8.699	7.301	6.357	7.386
76		8.301	8.000	5.959	9.886
77		8.301	7.699	6.000	7.812
78		8.523	7.699	5.959	7.506
79		7.268	6.367	5.240	7.783
80		8.097	7.301	6.022	7.155
81		8.301	7.398	5.000	7.788
82		8.301	7.155	6.201	8.149
83		8.699	7.699	5.590	8.745
84		9.699	8.398	6.456	8.886
85	Anthramycin methyl ether	7.886	5.801	4.611	8.097
86	Sibiromycin (R)	7.481	6.071	4.7773	8.8539
87	Sibiromycin (S)	7.4815	6.071	4.7773	8.8539

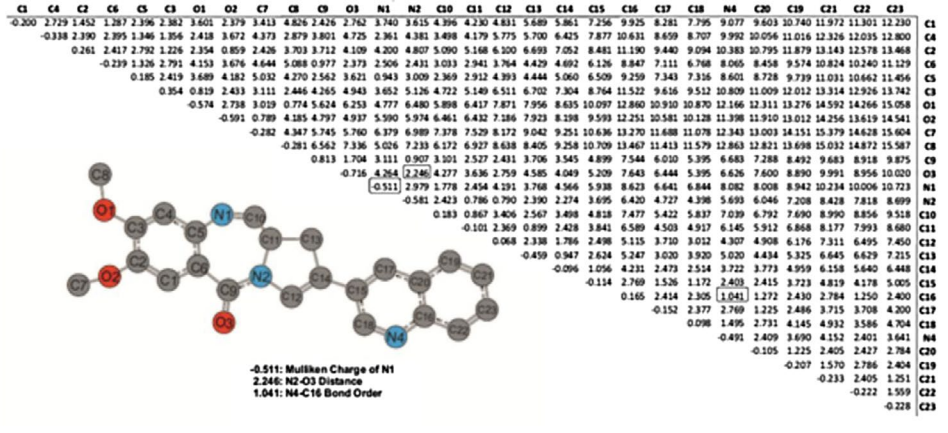


Figure 1. ECMC of the lowest energy conformer of the most active template molecule (compound 63) in the data set. The diagonal members correspond to the Mulliken charges whereas the non-diagonal elements refer to the bond orders for chemically bonded atom pairs and interatomic distances for non-bonded pairs. Hydrogen atoms attached to carbon atoms are omitted in the ECMC for clarity.

limit steadily all matrix elements of the ECMC of the template compound were compared with that of other ECMCs. Through the comparison of ECMCs, we obtained several electron conformational submatrices of activity (ECSA). Each ECSA was evaluated according to two commonly used criteria (P_α and α_a) given in Equations 1 and 2 below [33]:

$$P_\alpha = (n_1 + 1)/(n_1 + n_3 + 2) \tag{1}$$

$$\alpha_a = (n_1 \times n_4 - n_2 \times n_3)/(m_1 \times m_2 \times m_3 \times m_4)^{1/2} \tag{2}$$

where n_1 and n_2 are the numbers of molecules including and not including pharmacophore atoms (ECSA) in the class of highly active compounds, respectively, whereas n_3 and n_4 have similar meaning for weakly active compounds; m_1 and m_2 are the numbers of molecules in the class of highly active and weakly active compounds, respectively; $m3 = n1 + n3$; $m4 = n2 + n4$ [34]. Herein the first term P_α is related with the possibility of pharmacophore presence in active compounds while the second one is related with the possibility of pharmacophore presence in inactive/low active compounds.

To make clear how additional groups affect biological activity besides the pharmacophore, auxiliary groups (AG) and anti-pharmacophore shielding groups (APS) [23] were determined. AG and APS groups are distinguished by their opposite effects on biological activity. While the AG group promotes biological activity, the APS group shows a reducing effect. The out-of-pharmacophore groups are described by the following S function [35]:

$$Sn_i = \sum_{j=1}^{N} \kappa_j a_{ni}^{(j)} \tag{3}$$

where $a_{ni}^{(j)}$ is the parameter depicting the jth kind of feature in the jth conformation of the nth compound, N is the number of selected parameters and κ_j is the relative weight of different parameters. Each parameter has a different and constant κ_j value.

In the equation below [23], biological activity is expressed as a function of molecular descriptors, its energy and temperature considering the Boltzmann weighting of the individual conformations of each compound as follows:

$$An = Al \frac{\sum_{i=1}^{m_i} e^{-El_i/RT} \sum_{i=1}^{m_n} \delta n_i [Pha] e^{-Sn_i} e^{-En_i/RT}}{\sum_{i=1}^{m_n} e^{-En_i/RT} \sum_{i=1}^{m_i} \delta l_i [Pha] e^{-Sl_i} e^{-El_i/RT}} \tag{4}$$

where An and Al are the activity values of the nth compound and the reference compound, respectively. El_i is the relative energy of the ith conformation of the reference compound (in kcal mol^{-1}). En_i is the relative energy of the ith conformation of the nth compound (in kcal mol^{-1}), R (kcal mol^{-1} K^{-1}) is the gas constant and T is the temperature in Kelvin. δ is a kind of Dirac δ function which takes two values based on pharmacophore presence. The value equals 1 if the pharmacophore is present and 0 if not. The same equation was also used to calculate the κ_j variational constants. To implement and solve the weighted least squares fitting problem for the κ_j values of the parameters, the lsqnonlin function of the optimization toolbox in Matlab [36] is used. The weighted nonlinear least-squares analysis combined with GA can

be efficiently utilized with parameter selection and any kind of nonlinear optimizations. In addition, a GA and the method including iteration of the lsqnonlin function combined with initial values generated stochastically within a wide parameter range are employed to explore the best parameter subset. The numbers "κ_j" = 1, 2,…, *N*, obtained in this way characterize the weights of each kind of the *ani* (j) parameters in the overall APS/AG influence [23].

Another significant point is the preparation and the selection of descriptors. Hereby, 1331 molecular descriptors based upon four main classes (quantum chemical, thermodynamic, electrostatic and geometrical) regarding the pharmacophore group were generated for each conformer of PBD derivatives by EMRE software [21,26–30]. To eliminate the irrelevant and unnecessary descriptors and to increase model accuracy, the descriptor pool was reduced to a small subset of parameters. For this purpose, the most important parameters, $a_{ni}^{(j)}$ in Equation 4, were selected by the GA technique [37,38] since it is a fast and efficient method. The GA procedure starts with a randomly generated initial population comprising *N* individuals, each of which corresponds to a different parameter subset randomly selected from the descriptor pool. The populations are mainly composed of integer units defining model parameters (κ_j indices) as genetic codes. To calculate the κ_j values of the model parameters, each parent is subjected to the lsqnonlin function. The initial selected population according to the fitness values is subjected to genetic operators named selection, mutation and crossover to yield the new generation. Thus, some part of the next generation is constituted from the mutation procedure and the other part from crossover. Repeating this procedure, a number of models giving different parameter subsets are obtained until they converge or the prespecified size of generation is reached. Here, we run the GA with the following parameters: number of generation: 400; population size: 400; number of iterations: 150; crossover fraction: 85%; mutation rate: 1.5%.

Through LOO–CV, the fitness value of each chromosome was calculated by the predictive residual sum of squares (PRESS) as the fitness function. The formula of PRESS which measures the distribution of the calculations obtained from LOO-cross-validated values is given by:

$$PRESS_N = \sum_{n=1}^{N} \left| A_n^{\exp} - A_n^{pred} \right|^2 \tag{5}$$

where A_n^{exp} is the experimental activity of the *n*th molecule in the experimental activity data, A_n^{pred} is the predicted value of activity of the *n*th molecule in the training set by LOO–CV, and *N* is the total number of compounds in the training set.

In this study, the quality of the each of the obtained models was assessed internally by the LOO–CV method and externally by an analogous test set. In the internal validation of the models, only the training set compounds were considered. Each compound is precluded one by one to determine the biological activity with remaining compounds. Therefore, the contribution of each molecule to the robustness of the model is evaluated. For internal validation of the models, the value of q^2 was found by the following formula:

$$q^2 = 1 - \frac{\sum_{n=1}^{N} \left| A_n^{\exp} - A_n^{pred} \right|^2}{\sum_{n=1}^{N} \left| A_n^{\exp} - \bar{A}_n^{\exp} \right|^2} \equiv 1 - \frac{PRESS}{SSY} \tag{6}$$

where N indicates the total number of compounds in the training set. $\bar{A}_n^{exp}$ is the mean value of experimental activity of all the molecules in the training set. A_n^{exp} is the experimental activity of the nth molecule in the training set. SYY expresses the sum of squared deviations of experimental activity from the mean ($\bar{A}_n^{exp}$). So as to verify the reliability and predictivity of the models on the new compounds which are not used in the model development, the data set is split into training and test sets. The model developed by training compounds is applied to the test compounds to confirm the prediction power. In order to calculate the q^2, two expressions of external validation were proposed by Schüürmann et al. [39] and are based on the average values involving the training set and test set means in the denominator and the sum of squares of the external set in the numerator. These equations are given by the following formulas [39]:

$$q^2_{ext1} = 1 - \frac{\sum_{n=1}^{N} \left| A^{\exp}_{n_{test}} - A^{pred}_{n_{test}} \right|^2}{\sum_{n=1}^{N} \left| A^{\exp}_{n_{test}} - \bar{A}^{\exp}_{training} \right|^2} \tag{7}$$

$$q^2_{ext2} = 1 - \frac{\sum_{n=1}^{N} \left| A^{\exp}_{n_{test}} - A^{pred}_{n_{test}} \right|^2}{\sum_{n=1}^{N} \left| A^{\exp}_{n_{test}} - \bar{A}^{\exp}_{n_{test}} \right|^2} \tag{8}$$

where N is the number of molecules to be tested. $A^{exp}_{n_{test}}$ and $A^{pred}_{n_{test}}$ are the experimental and the predicted activities of the nth compound in the test set. $\bar{A}^{exp}_{n_{training}}$ and $\bar{A}^{exp}_{n_{test}}$ are the arithmetic means of the experimental activities of the training and test sets, respectively.

Another external validation measure called Q^2_{F3} was introduced by Consonni et al. [40] for the purpose of discussing the predictive ability of QSAR models with external assessment described in Schüürmann et al.'s study [40]. The external prediction capability given by Consonni is calculated by the following equation:

$$q^2_{ext3} = 1 - \frac{\left[\sum_{n=1}^{N_{test}} \left| A^{exp}_{n_{test}} - A^{pred}_{n_{test}} \right|^2 \right] \Big/ N_{test}}{\left[\sum_{n=1}^{N_{training}} \left| A^{exp}_{ntraining} - \bar{A}^{exp}_{training} \right|^2 \right] \Big/ N_{training}} \tag{9}$$

where N_{test} and $N_{training}$ are the number of test and training molecules, respectively. Whereas $A^{exp}_{n_{test}}$ and $A^{Pred}_{n_{test}}$ refer to the experimental and the predicted activity values of the nth test compound, $A^{exp}_{n_{training}}$ is the experimental activity of the nth compound in the training set. $\bar{A}^{exp}_{n_{training}}$ is equal to the mean of the experimental activities of the training compounds. In Equation 9, the sum of squares in the denominator is related with the training set while that in the numerator is related with the external prediction set.

In addition to the external evaluation criteria given above, Chirico and Gramatica proposed a different and simpler alternative which gives more cautious and restrictive results

in proportion to other compared measures. The rearranged version of the concordance correlation coefficient (CCC) is given by following equation [41]:

$$CCC = \underset{training}{\hat{\rho}} = \frac{2\sum_{i=1}^{n_{training}} (A_i^{pred} - \bar{A}^{pred})(A_i^{exp} - \bar{A}^{exp})}{\sum_{i=1}^{n_{training}} (A_i^{pred} - \bar{A}^{pred})^2 + \sum_{i=1}^{n_{training}} (A_i^{exp} - \bar{A}^{exp})^2 + n_{training}(A_i^{exp} - \bar{A}^{exp})^2} \tag{10}$$

$$CCC = \underset{test}{\hat{\rho}} = \frac{2\sum_{i=1}^{n_{test}} (A_i^{pred} - \bar{A}^{pred})(A_i^{exp} - \bar{A}^{exp})}{\sum_{i=1}^{n_{test}} (A_i^{pred} - \bar{A}^{pred})^2 + \sum_{i=1}^{n_{test}} (A_i^{exp} - \bar{A}^{exp})^2 + n_{test}(\bar{A}^{pred} - \bar{A}^{exp})^2} \tag{11}$$

$$CCC = \underset{all}{\hat{\rho}} = \frac{2\sum_{i=1}^{n_{all}} (A_i^{pred} - \bar{A}^{pred})(A_i^{exp} - \bar{A}^{exp})}{\sum_{i=1}^{n_{all}} (A_i^{pred} - \bar{A}^{pred})^2 + \sum_{i=1}^{n_{all}} (A_i^{exp} - \bar{A}^{exp})^2 + n_{all}(\bar{A}^{pred} - \bar{A}^{exp})^2} \tag{12}$$

where A_i^{exp} and A_i^{pred} correspond to the experimental and predicted values of the activity, respectively. Similarly, $\bar{A}^{exp}$ and $\bar{A}^{pred}$ correspond to the averages of the experimental and predicted activity values. In the formula, by using both training and test sets, the reliability of the model was developed. The CCC, which has a value greater than 0.85, confirms the excellent precision and accuracy of the model.

For QSAR model development, different external evaluation functions which have advantages or drawbacks with regard to each other were introduced by different researchers. Among those expressions, Equations 7–9 were used to appraise model consistency in previous papers by us. We also made use of the last two external validation formulas (Equations 10–12) aforementioned for the first time in this 4D-QSAR EC–GA study.

At the end of the model development stage, evaluating the prediction abilities of all the models considering the r^2, q^2, q^2_{ext1}, q^2_{ext2}, q^2_{ext3} and CCC criteria by the LOO–CV technique, the best parameter subset and related best model were determined. Using the best parameter subset and corresponding κ_j values, we calculated the activity values of the compounds with unknown activity with Equation 4.

In the best parameter subset, one or several parameters make more contribution to the biological activity. To estimate which parameter/parameters in the subset is predominant, the *E*-statistics technique is used [42]. The statistical *E* value is calculated by the following formula as the ratio of the predictive sum of squares:

$$E = PRESS_N / PRESS_{N-1} \tag{13}$$

$$PRESS_{N-1} = \sum_{n=1}^{N-1} \left| A_n^{exp} - A_n^{pred} \right|^2 \tag{14}$$

where A_n^{exp} and A_n^{pred} refer to the experimental and predicted activity in the LOO–CV procedure. In this situation only a small number of parameters ($N = 9–11$ in this study) were used to construct the model. The value of the E defines the impact of the parameters. The greatest the increase in the E value, the lowest the contribution made by the parameter. In parallel with the high value of E, omission of the parameter reduces the model's performance.

Results and discussion

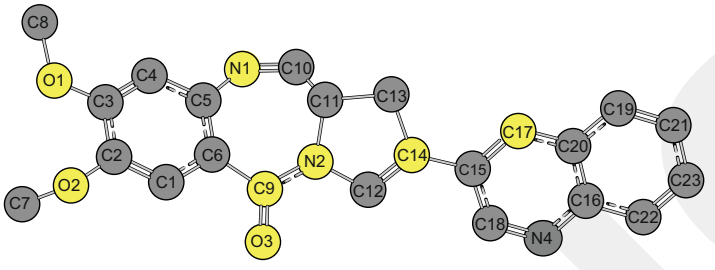
The chemical structures of the C2-aryl PBD derivatives with substituents and experimental pGI_{50}, pTGI, pLC_{50} and pIC_{50} values are given in Table 1 in the previous section. The data comprising atomic charges, Cartesian coordinates, bond orders and interatomic distances from the conformational analysis and quantum chemical calculations at Hartree Fock 3-21G* level were assigned to build the ECMCs of the 997 conformers of 87 compounds by the EMRE programme (see Figure 1 for sample matrix of the lowest energy conformer of reference compound). To describe the pharmacophore for GI_{50} activity, the value $pGI_{50} = 8.3010$ was regarded as the activity threshold. In total, 46 compounds with $pGI_{50} \geq 8.3010$ were categorized as high-activity compounds, 37 were classed as low-activity compounds and four compounds had unknown activity.

For GI_{50} activity, the comparison procedure of the ECMCs defined in the materials and methods section resulted in a pharmacophore group comprising the O1, O2, C9, O3, N1, N2, C14 and C17 atoms with an optimum $P_\alpha = 0.9737$ and $\alpha_a = 0.7849$ values (i.e. with the highest P_α and α_a values). The final ECSA and relevant tolerance values for both active and inactive compounds including the compounds with unknown activity are reported in Table 2, in which pharmacophore atoms are shown in yellow. Table 2 contains six submatrices. The first submatrix corresponds to pharmacophore atoms for the lowest energy conformer of the template compound. The second and third ones are the tolerance submatrices for 46 compounds with high activity and 37 compounds with low activity, respectively. The fourth submatrix represents tolerance values for the overall conformers (997) of 87 compounds without tolerance limitation. As seen in (b) and (c) of Table 2, the atomic charge tolerances of the O1 atom are ±0.024 and ±0.093 and the tolerances of the distance between the N1 and N2 atoms are ±0.028 and ±0.189 for high and low active compounds, respectively. Table 2 proves that, in general, compounds with high activity possess lower tolerance values than those with low activity.

After careful analysis of the pharmacophore atoms, the O3, N1, O1 and O2 atoms present in the benzodiazephine ring are identified among the key pharmacophoric elements as hydrogen-bond acceptors. The C14 and C17 atoms located in the imidazole and quinoline ring, respectively, comprise the hydrophobic regions. Most of the pharmacophore atoms are placed on a rigid plane since the structure contains condensed heterocyclic units showing very little conformational flexibility. The O1, O2, N1, O3 and N2 atoms are defined as negatively charged atoms while the C9 atom is positively charged. The C14 and C17 atoms show lower negative charges than the others. The highest tolerance value of interatomic distances for high-activity compounds pertains to the C17–O2 distance which shows the flexibility of the position whereas the N2–O3 distance has minimum tolerance due to a rigid plane.

In the first step of bioactivity prediction, four data sets associated with pGI_{50}, pTGI, pLC_{50} and pIC_{50} values were randomly divided into three data sets: the training set, test set and unknown set. The compounds in the training, test and unknown set were randomly selected

Table 2. (a) Pharmacophore (ECSA) of reference compound (63) for pGI_{50} activity values of C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine derivatives; (b) Tolerance matrix of ECSA for 46 compounds with high activity; (c) Tolerance matrix of ECSA for 37 compounds with low activity; (d) Tolerance values for 997 conformers of 87 compounds; (e) Tolerance matrix of ECSA for the lowest energy conformer of 4 compounds with unknown activity; f) Tolerance matrix of 64 conformers of 4 compounds with unknown activity. Pharmacophore atoms are shown in yellow. The optimum P_α and α_a values found are 0.9737 and 0.7849, respectively.



a) ECSA of reference compound (63)

O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
-0.574	2.738	5.624	6.253	4.777	6.480	8.635	10.910	O1
	-0.591	4.797	4.937	5.590	5.974	8.198	10.581	O2
		0.813	1.704	3.111	0.907	3.545	6.010	C9
			-0.716	4.264	2.246	4.049	6.444	O3
				-0.511	2.979	4.566	6.641	N1
					-0.581	2.274	4.727	N2
						-0.096	2.473	C14
							-0.152	C17

b) Tolerance values for 46 compounds with high activity

O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
±0.024	±0.077	±0.021	±0.034	±0.037	±0.009	±0.042	±0.071	O1
	±0.008	±0.013	±0.046	±0.008	±0.032	±0.024	±0.110	O2
		±0.005	±0.046	±0.030	±0.025	±0.009	±0.017	C9
			±0.015	±0.019	±0.005	±0.054	±0.046	O3
				±0.226	±0.028	±0.095	±0.092	N1
					±0.012	±0.007	±0.020	N2
						±0.034	±0.021	C14
							±0.074	C17

c) Tolerance values for 37 compounds with low activity

O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
±0.093	±0.078	±0.151	±0.251	±0.118	±0.049	±0.055	±0.249	O1
	±0.132	±0.108	±0.994	±2.856	±0.801	±1.202	±1.467	O2
		±0.008	±0.038	±0.156	±0.048	±0.074	±0.505	C9
			±0.022	±0.195	±0.014	±0.128	±0.607	O3
				±0.288	±0.189	±0.223	±0.383	N1
					±0.018	±0.091	±0.426	N2
						±0.176	±0.068	C14
							±0.603	C17

d) Tolerance values for 997 conformers of 87 compounds

O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
±0.932	±1.836	±1.298	±1.676	±0.731	±1.762	±1.353	±1.735	O1
	±0.617	±1.046	±1.160	±2.856	±1.588	±1.539	±1.923	O2
		±1.395	±0.760	±1.758	±1.600	±1.232	±1.737	C9
			±0.901	±1.760	±0.915	±1.630	±1.621	O3
				±1.325	±1.746	±1.685	±1.617	N1
					±0.879	±1.471	±1.851	N2
						±0.908	±1.596	C14
							±0.923	C17

(*Continued*)

Table 2. (*Continued*)

O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
e) Tolerance matrix of ECSA for the lowest energy conformer of 4 compounds with unknown activity								
O1	O2	C9	O3	N1	N2	C14	C17	Pha Atoms
±0.097	±0.076	±0.147	±0.133	±0.080	±0.187	±0.223	±0.290	O1
	±0.035	±0.041	±0.038	±0.032	±0.077	±0.075	±0.123	O2
		±0.004	±0.032	±0.037	±0.011	±0.007	±0.010	C9
			±0.012	±0.075	±0.006	±0.046	±0.074	O3
				±0.244	±0.043	±0.136	±0.163	N1
					±0.024	±0.008	±0.011	N2
						±0.009	±0.011	C14
							±0.064	C17
f) Tolerance matrix of 64 conformers of 4 compounds with unknown activity								
O1	*O2*	*C9*	*O3*	*N1*	*N2*	*C14*	*C17*	*Pha Atoms*
±0.513	±0.333	±0.647	±0.731	±0.681	±0.524	±0.375	±0.331	O1
	±0.148	±0.490	±0.361	±0.632	±0.492	±0.544	±0.569	O2
		±0.874	±0.721	±0.622	±0.557	±0.297	±0.505	C9
			±0.666	±0.668	±0.685	±0.615	±0.837	O3
				±1.321	±0.732	±0.591	±0.687	N1
					±0.379	±0.339	±0.336	N2
						±0.908	±0.051	C14
							±0.540	C17

from the entire data set by GA. For each activity type, the generated models were evaluated both internally and externally.

These subsets for GI_{50} activity included 55 training, 27 test and five unknown compounds. Likewise, the pTGI, pLC_{50} and pIC_{50} datasets were classified as training, test and unknown sets (55, 27, 5; 55, 27, 5; and 48, 24, 15, respectively).

The main goal of descriptor selection is to develop a robust model by employing the minimum number of variables. As the optimal number of parameters is not known formerly, it is essential to run a number of models to explore the relationship between prediction power (q^2) and the number of parameters in the subset. First the compounds were randomly selected; then they were kept stable and we scanned the number of parameters from 1 to 15 to detect the optimum number of parameters. The number of parameters was plotted versus r^2 (for training and test set), q^2, q^2_{ext1}, q^2_{ext2}, q^2_{ext3} and CCC of pGI_{50} activity as shown in Figure 2. As seen in Figure 2, even if increasing the number of parameters causes a rise in r^2 and q^2 up to 11 descriptors, after 11 descriptors the model gains stability and a higher number of descriptors does not enhance the model performance very much. As a general rule, the ratio of the number of parameters to the number of compounds in the model should not be higher than 1:5 to avoid potential overfitting risk [43].

The plots showing the optimum number of parameters for pTGI, pLC_{50} and pIC_{50} values are also given in Figures S1–S3 as supporting information (available via the Supplementary Content tab on the article's online page). The pTGI activity values of C2-aryl PBD derivatives resulted in an optimum of 11 parameters for 55 training and 27 test compounds. In Figure S1, the statistical parameters exhibit an increase until 11 parameters. At 11 parameters, the model reaches a steady state and does not need any extra parameters. Thus, the model for pTGI was found as a function of the best 11 parameters. In the same way, the optimum numbers of parameters for pLC_{50} and pIC_{50} activities are determined in Figures S2 and S3 as 11 and 9, respectively.

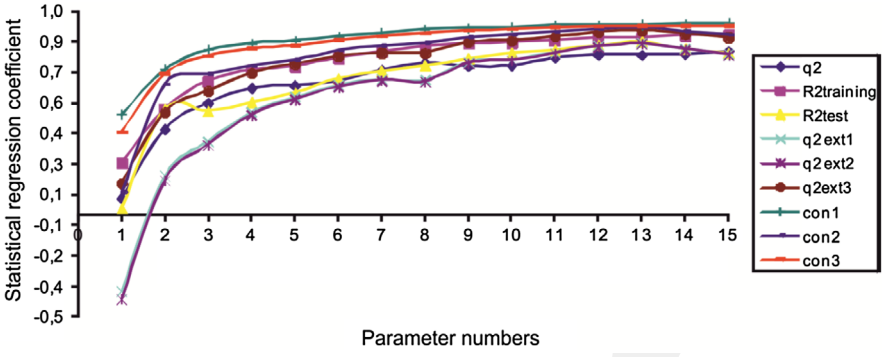


Figure 2. Plot of the correlation between number of parameters and r^2, q^2, q^2_{ext1}, q^2_{ext2}, q^2_{ext3} and CCC for pGI_{50} activity values.

Table 3. Optimal 11 descriptors chosen by GA and κ_j values for pGI_{50} activity values of C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine derivatives.

$a_{ni}^{(j)}$	*Molecular parameters*	κ_j
$a^{(1)}$	Orthogonal distance from C8 atom to the O1 N1 O3 plane (Å)	0.102
$a^{(2)}$	Orthogonal distance from O3 atom to the N1 N2 C14 plane (Å)	-0.128
$a^{(3)}$	Orthogonal distance from C4 atom to the N1 N2 C14 plane (Å) + van der Waals radius (Å)	0.297
$a^{(4)}$	Orthogonal distance from C8 atom to the C17 C14 N1 plane (Å) + van der Waals radius (Å)	-0.061
$a^{(5)}$	Orthogonal distance from C15 atom to the O1 O2 C17 plane (Å)	-0.141
$a^{(6)}$	Orthogonal distance from C11 atom to the N4 C12 O3 plane (Å)	0.064
$a^{(7)}$	Angle between O3 C9 N2 plane and the line of C14-C23	0.103
$a^{(8)}$	Electrostatic charge of N2 atom	-0.498
$a^{(9)}$	Nucleophilic atomic frontier electron density of O3 atom	-2.193
$a^{(10)}$	Nucleophilic atomic frontier electron density of N2 atom	-1.801
$a^{(11)}$	Fukui atomic electrophilic reactivity index of C17 atom	-30.654

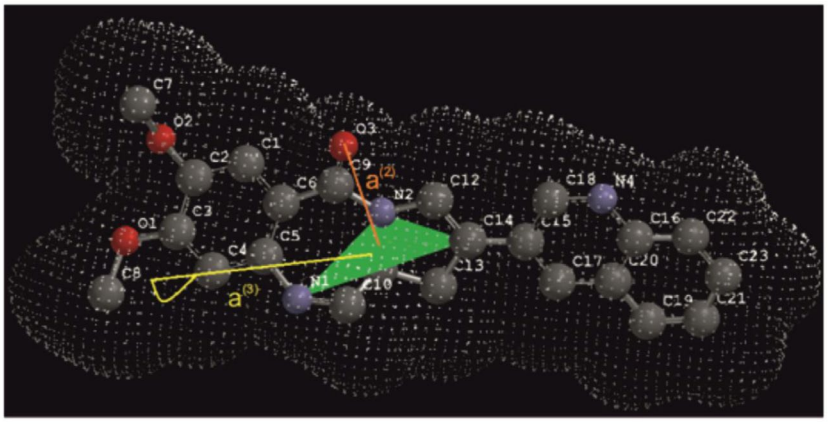


Figure 3. Presentation of orthogonal distance related parameters $a^{(2)}$ and $a^{(3)}$ for pGI_{50} activity values.

For pGI_{50}, a brief definition of the best 11 descriptors selected with GA and the related κ_j values are listed in Table 3. The analysis of Table 3 shows that geometrical and electronic parameters have more impact on the GI_{50} activity of C2-aryl PBD derivatives. $a^{(1)}$, $a^{(2)}$, $a^{(3)}$, $a^{(4)}$,

$a^{(5)}$, $a^{(6)}$ and $a^{(7)}$ are the geometrical parameters involving mostly pharmacophore atoms. The parameters $a^{(1)}$, $a^{(2)}$, $a^{(5)}$ and $a^{(6)}$ are orthogonal distances. $a^{(3)}$ and $a^{(4)}$ are the orthogonal distances plus van der Waals radius (Å). The remaining four parameters represent the electronic features of the pharmacophoric atoms. $a^{(8)}$ is the electrostatic charge of the N2 atom placed in the imidazole ring. $a^{(9)}$ and $a^{(10)}$ are the nucleophilic atomic frontier electron density index values [44] of the O3 and N2 atoms, respectively. The last parameter, $a^{(11)}$, in Table 3 is the Fukui atomic electrophilic reactivity index value [45] of the C17 atom. The presentation of parameter $a^{(2)}$ and $a^{(3)}$ is shown in Figure 3.

The best descriptors and related κ_j values corresponding to pTGI, pLC_{50} and pIC_{50} values are given in Tables S1–S3 (available online). In Table S1 for TGI activity, it is seen that the first eight parameters ($a^{(1)}$)–$a^{(8)}$) are geometrical parameters including the orthogonal distance, orthogonal distance + van der Waals radius and the angle between the line and plane of atoms, whereas $a^{(9)}$ and $a^{(10)}$ symbolize the Fukui atomic electrophilic reactivity index values of the O1 and C17 atoms. $a^{(11)}$ is log P, which is the partition coefficient related with the compound's hydrophobicity. A similar situation is seen in Table S2 and Table S3 for LC_{50} and IC_{50} activities. For both types of activity, geometrical parameters are predominant. The parameter list of LC_{50} activity gave $a^{(1)}$–$a^{(7)}$ as geometrical parameters which are mainly composed of orthogonal distance and orthogonal distance + van der Waals radius. The other four parameters ($a^{(8)}$)–$a^{(11)}$) are the nucleophilic atomic frontier electron density index value of the O3 atom [46], the Fukui atomic electrophilic reactivity index value of the C17 atom, the HOMO and log P. The best parameters' list for pIC_{50} values (see Table S3) includes nine parameters of which $a^{(1)}$ is the orthogonal distance + van der Waals radius, $a^{(2)}$ is the orthogonal distance, $a^{(3)}$ is the angle between the C16 C17 C20 plane and the C14–C18 line, $a^{(4)}$ and $a^{(5)}$ are the electrophilic atomic frontier electron density index values of the C17 and C16 atoms [46] and $a^{(6)}$–$a^{(9)}$ are the dihedral angles.

To determine the AG and APS groups which contribute positively or negatively to the activity, the product of κ_j and the parameter value was taken into account. If the result of the product is positive then the related parameter is regarded as an AG parameter, otherwise it is an APS parameter. Accordingly, within the 11 optimal parameters in Table 3 for GI_{50} activity $a^{(2)}$, $a^{(4)}$, $a^{(5)}$, $a^{(9)}$, $a^{(10)}$ and $a^{(11)}$ are AG parameters while $a^{(1)}$, $a^{(3)}$, $a^{(6)}$, $a^{(7)}$ and $a^{(8)}$ are APS parameters. In the same way for TGI activity, $a^{(1)}$, $a^{(3)}$, $a^{(6)}$, $a^{(9)}$, $a^{(10)}$ and $a^{(11)}$ were determined as AG parameters and $a^{(2)}$, $a^{(4)}$, $a^{(5)}$, $a^{(7)}$ and $a^{(8)}$ as APS parameters. Among the parameters in Table S2 of LC_{50} activity, $a^{(2)}$, $a^{(4)}$, $a^{(6)}$, $a^{(8)}$, $a^{(9)}$ and $a^{(11)}$ are AG parameters while $a^{(1)}$, $a^{(3)}$, $a^{(5)}$, $a^{(7)}$ and $a^{(10)}$ are APS parameters. Finally $a^{(4)}$, $a^{(6)}$, $a^{(8)}$ and $a^{(9)}$ are AG parameters and $a^{(1)}$, $a^{(2)}$, $a^{(3)}$, $a^{(5)}$ and $a^{(7)}$ are APS parameters for IC_{50} activity.

In consideration of previous explanations, among the several models for pGI_{50}, pTGI, pLC_{50} and pIC_{50} activity values, the experimental and predicted activity values, r^2, standard error and both internal and external q^2 values for the best models of each activity type are listed in Table 4. As seen in Table 4, the data set of pGI_{50} was divided into a training set of 55 compounds and a test set of 27 compounds in order to get an exact robust model through a validation procedure with test compounds. The compounds marked with "a" correspond to test compounds while those marked with an asterisk are unknown compounds. The number of training, test and unknown sets for pTGI, pLC_{50} and pIC_{50} datasets are 55, 27, 5; 55, 27, 5 and 48, 24, 15, respectively.

As a general rule, if the q^2 values of the cross-validated models are higher than 0.5, the predictive ability of the model should be acceptable [47]. Based on internal validation, the

Table 4. Experimental and predicted activity values with statistical results of pGI_{50}, pTGI, pLC_{50} and pIC_{50} for C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine derivatives.

	pGI_{50}			pTGI			pLC_{50}			pIC_{50}	
Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}
1[a]	8.699	8.766	1[a]	7.398	6.985	1[a]	5.290	5.299	1	8.602	8.070
2	8.523	8.605	2[a]	7.398	7.099	2[a]	5.491	5.351	2	7.494	7.368
3[a]	8.699	8.606	3[a]	7.523	6.831	3[a]	5.320	5.139	3	7.854	7.948
4	8.222	8.416	4[a]	6.699	6.400	4	4.939	5.121	4	7.527	7.965
5[a]	8.699	8.426	5	7.097	6.920	5	5.470	5.029	5	7.987	7.935
6[a]	9.000	8.389	6	6.824	7.030	6	4.721	5.043	6	8.310	7.578
7	8.523	8.008	7	7.523	7.686	7[a]	5.900	5.557	7	7.292	7.518
8	8.699	8.563	8	7.301	7.814	8	5.712	5.522	8	7.321	7.351
9[a]	8.699	8.660	9[a]	6.745	7.032	9	5.051	5.399	9	8.553	8.056
10	9.000	8.554	10[a]	7.155	7.051	10	5.380	5.441	10[a]	7.100	7.157
11[a]	8.398	8.289	11	8.222	7.802	11[a]	5.351	5.549	11	6.939	7.318
12[a]	8.301	8.228	12[a]	8.000	7.094	12[a]	6.208	5.641	12	7.161	7.453
13	8.097	8.741	13[a]	8.046	7.304	13	5.440	5.642	13[a]	7.708	7.940
14[a]	8.301	8.373	14	7.699	7.024	14	5.842	5.195	14	7.721	7.709
15	8.699	8.837	15	6.921	6.953	15[a]	4.951	5.341	15*	-	8.030
16	8.699	8.817	16	7.155	7.121	16	5.350	5.060	16	7.505	7.447
17*	-	8.764	17*	-	7.148	17*	-	5.551	17 *	-	7.973
18[a]	8.046	7.668	18	6.796	6.578	18[a]	4.821	4.977	18	7.580	7.726
19	8.398	8.156	19[a]	7.523	6.979	19	5.780	5.732	19	7.614	7.897
20	8.523	7.988	20[a]	6.824	6.564	20	5.130	5.446	20[a]	7.703	8.042
21*	-	7.819	21[a]	4.631	6.025	21*	-	5.095	21*	-	7.807
22*	-	6.684	22	7.155	6.531	22*	-	4.623	22[a]	7.883	7.616
23[a]	8.046	8.228	23	6.337	6.789	23	4.860	5.151	23	7.807	7.330
24	7.824	7.895	24[a]	6.149	6.088	24	4.860	5.129	24	7.226	7.525
25*	-	7.625	25[a]	5.731	5.827	25*	-	5.255	25	6.794	7.144
26	7.721	7.786	26	7.046	6.690	26	5.410	5.162	26	6.943	7.308
27[a]	7.237	8.147	27[a]	6.081	6.367	27[a]	4.450	5.224	27[a]	6.783	7.606
28[a]	8.301	8.541	28[a]	7.097	6.726	28	5.080	5.289	28*	-	7.730
29[a]	8.301	8.485	29	6.796	6.913	29	4.879	5.127	29	8.056	7.774
30	8.398	8.021	30	7.301	6.670	30	5.230	5.028	30	7.226	7.180
31	7.959	8.426	31	7.222	7.300	31[a]	5.390	5.286	31*	-	7.095
32[a]	8.301	8.275	32	7.155	7.354	32	5.120	5.208	32	8.022	7.192
33	8.523	8.499	33	7.301	7.226	33	4.780	5.463	33	8.194	7.210
34	7.770	8.442	34	7.000	7.157	34	6.268	5.343	34	7.712	7.185
35	8.222	8.101	35	7.523	7.476	35	5.870	5.407	35[a]	7.330	7.372
36	8.222	8.497	36	6.745	7.035	36	5.140	5.082	36	7.116	7.249
37	8.398	7.857	37[a]	6.252	6.354	37[a]	4.791	5.020	37	7.821	7.230
38*	-	7.490	38*	-	6.115	38*	-	4.838	38*	-	7.230
39	8.301	7.755	39	7.398	6.914	39[a]	5.120	4.917	39	7.907	7.274
40	8.301	8.086	40	7.301	7.736	40	5.250	5.107	40	7.236	7.244
41	7.569	8.199	41[a]	5.731	6.307	41	4.979	4.942	41	6.925	7.693
42	6.991	7.680	42	6.071	6.245	42	4.851	5.097	42	6.564	7.241
43	7.921	7.524	43	6.602	6.422	43	4.971	4.929	43	8.032	8.044
44	7.796	7.612	44	6.796	6.541	44	5.361	4.869	44[a]	6.588	7.621
45	6.959	7.209	45[a]	5.959	6.046	45[a]	4.932	4.989	45	7.215	7.149
46	6.070	6.508	46[a]	5.461	5.659	46	4.680	4.796	46[a]	6.000	6.255
47[a]	7.602	7.956	47	6.482	7.147	47	5.361	5.260	47	8.149	7.918
48	8.523	8.391	48	7.301	7.288	48	5.520	5.537	48*	-	7.940
49	9.222	8.900	49	7.824	8.024	49[a]	6.398	6.480	49[a]	7.343	7.226
50	9.046	8.861	50	8.155	8.169	50	6.569	6.033	50[a]	8.638	8.221
51[a]	8.699	8.712	51[a]	7.699	7.502	51	5.710	6.019	51[a]	7.900	7.750
52	6.860	6.977	52[a]	6.022	6.122	52[a]	4.857	5.207	52[a]	6.242	7.155
53[a]	8.000	8.266	53	6.824	7.620	53	6.131	5.772	53	8.959	8.852
54	7.921	8.139	54	6.569	6.742	54	5.190	4.743	54	7.697	8.122
55	8.523	8.093	55	7.301	6.873	55[a]	5.270	4.857	55	7.914	7.539
56	8.301	8.290	56	7.155	6.959	56	4.770	4.774	56[a]	7.740	8.338
57	7.745	7.732	57	6.222	6.241	57	4.570	4.526	57*	-	4.447
58[a]	8.398	7.799	58	7.222	6.735	58	5.390	5.701	58	8.027	7.695

Table 4. (*Continued*)

	pGI_{50}			pTGI			pLC_{50}			pIC_{50}	
Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}	Comp.	A_{exp}	A_{pred}
59[a]	8.222	8.044	59[a]	7.301	6.946	59[a]	6.357	6.131	59[a]	7.967	9.071
60	8.301	8.407	60	8.046	7.969	60	6.268	6.532	60*	-	9.023
61[a]	8.699	8.313	61	7.523	7.343	61	5.440	5.356	61*	-	7.829
62[a]	8.301	8.508	62	7.398	7.491	62	5.150	5.246	62[a]	7.900	8.151
63	10.000	10.000	63	9.000	9.000	63	6.745	6.745	63	8.051	8.070
64	9.398	9.298	64	8.398	8.469	64[a]	6.721	6.341	64[a]	8.854	8.360
65	7.602	7.131	65	6.658	6.748	65[a]	5.270	5.213	65[a]	6.787	7.400
66	7.310	7.142	66	6.678	6.672	66	5.550	5.070	66	6.792	7.259
67[a]	6.300	6.307	67[a]	5.760	5.766	67	5.050	5.066	67[a]	6.000	6.570
68	6.400	6.553	68[a]	5.842	5.720	68	4.971	4.847	68[a]	7.065	7.805
69	6.991	6.633	69	6.347	6.265	69	6.041	6.005	69	6.027	6.015
70	5.670	5.621	70	4.721	4.641	70[a]	4.240	4.050	70	6.312	6.412
71	5.010	5.073	71	4.360	4.465	71	4.040	4.473	71[a]	6.000	5.673
72[a]	5.870	5.431	72	5.090	5.036	72	4.220	4.335	72*	-	6.646
73	8.398	8.583	73[a]	7.699	7.211	73[a]	5.801	5.959	73	10.000	9.973
74[a]	8.699	8.481	74	7.301	7.910	74[a]	6.357	6.017	74*	-	9.738
75	8.301	8.574	75[a]	8.000	7.279	75[a]	5.959	6.035	75	9.886	9.886
76[a]	8.301	8.556	76	7.699	8.057	76	6.000	6.101	76*	7.813	9.710
77	8.523	8.396	77[a]	7.699	7.368	77	5.959	6.054	77*	-	9.053
78	7.108	7.140	78	5.910	6.019	78	4.520	4.589	78	8.569	8.677
79	7.268	7.649	79	6.367	6.799	79[a]	5.240	5.221	79	7.783	8.202
80[a]	8.097	7.775	80	7.301	6.928	80	6.022	5.335	80	7.155	7.756
81	8.301	7.698	81	7.398	6.897	81	5.000	5.200	81[a]	7.788	8.114
82	8.301	7.815	82	7.155	7.360	82[a]	6.201	5.796	82	8.149	7.458
83	8.699	9.430	83	7.699	7.788	83	5.590	6.247	83	8.745	8.426
84	9.699	10.092	84[a]	8.398	7.634	84	6.456	6.441	84[a]	8.886	8.288
85	7.886	7.971	85	5.801	6.007	85	4.611	4.581	85[a]	8.097	8.238
86[a]	7.482	7.326	86	6.071	6.027	86	4.777	4.643	86[a]	8.854	8.775
87	7.482	7.583	87[a]	6.071	6.114	87[a]	4.777	4.648	87*	-	8.200
Training											
r^2	0.858		r^2	0.853		r^2	0.703		r^2	0.776	
se	0.052		se	0.053		se	0.075		se	0.070	
q^2	0.771		q^2	0.787		q^2	0.600		q^2	0.687	
Test											
r^2	0.810		r^2	0.848		r^2	0.787		r^2	0.722	
se	0.074		se	0.074		se	0.092		se	0.112	
q^2_{ext1}	0.797		q^2_{ext1}	0.743		q^2_{ext1}	0.787		q^2_{ext1}	0.597	
q^2_{ext2}	0.791		q^2_{ext2}	0.731		q^2_{ext2}	0.781		q^2_{ext2}	0.564	

[a]Test compounds; *Compounds with unknown activity.

developed model gave excellent internal accuracy with a non-cross-validated r^2 value of 0.858 and cross-validated q^2 value of 0.771 for the training set, and an r^2 value of 0.810 for the test set. This model was used to predict the antitumour GI_{50} activity of the compounds in the external test set and also of unknown compounds. The model for external validation resulted in satisfactory external q^2 values (q^2_{ext1} = 0.797 and q^2_{ext2}=0.791). In addition, the difference between the experimental and predicted activity values is less than 1. The usefulness of the obtained models for future activity prediction of new PBD analogues can be seen from the high quality of the statistical results of the models. It is seen that the TGI activity results also showed very good predictive capability with internal and external validation criteria. For the best model of TGI activity with an optimum 11 of parameters, the r^2 and q^2 values of the training set were found as 0.848 and 0.787. In addition, the external validation results of the test set (r^2 = 0.848, q^2_{ext1} = 0.743 and q^2_{ext2} = 0.731), which is the real indicator of the prediction capacity of a model, are also highly predictive and acceptable. The models

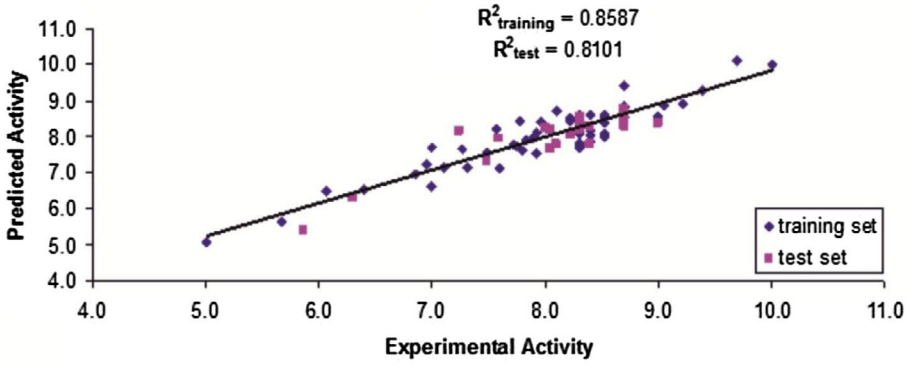


Figure 4. Plot of experimental vs. predicted pGI_{50} activity values of training and test sets obtained with 11 descriptors.

developed for LC_{50} and IC_{50} activities are quite good but had slightly lower r^2 (0.703 and 0.776) and q^2 (0.600 and 0.687) values in the training set than for the GI_{50} and TGI activities. Considering the external test set results, both LC_{50} and IC_{50} show r^2 values over 0.700 while q^2_{ext1} and q^2_{ext2} values for only LC_{50} activity are higher than 0.700. The lower q^2_{ext1} and q^2_{ext2} values indicate that the model is less capable of correctly predicting.

The plot of experimental vs. predicted pGI_{50} values of training and test sets obtained by 11 descriptors is shown in Figure 4. Consequently, taking into account all the conformers of the 87 compounds, both the training and test sets gave acceptable statistical results with an optimal 11 descriptors. The model generated with the EC–GA method produced a good prediction power (see Table 4, Figure 4, Figures S4–S6 (available online)). The pTGI, pLC_{50} and pIC_{50} corresponding plots are given in Figures S4–S6.

All calculations related to bioactivity prediction and statistical analysis were carried out in two ways: the first examined all the conformers and the second examined only the lowest energy conformer for each compound. The statistical results for pGI_{50} regarding both only one conformer and all conformers are presented in Figure 5. Regarding only the lowest energy conformer of each compound, we obtained the q^2, $r^2_{training}$, r^2_{test}, q^2_{ext1}, q^2_{ext2}, q^2_{ext3}, con1, con2 and con3 values as 0.573, 0.777, 0.764, 0.747, 0.740, 0.844, 0.867, 0.846 and 0.862, respectively. Accordingly, as shown in Figure 5, we obtained better statistical outcomes considering all conformers energetically reasonable for EC–GA model development. The other three activity types showed a similar trend. Comparisons of the statistical results for pTGI, pLC_{50} and pIC_{50} considering one conformer and all conformers are given in Figures S7–S9 (available online).

When we considered only the lowest energy conformer we achieved the following results: for TGI activity $q^2 = 0.720$, $r^2_{training} = 0.816$, $r^2_{test} = 0.660$, $q^2_{ext1} = 0.404$, $q^2_{ext2} = 0.378$, $q^2_{ext3} = 0.221$, con1 = 0.902, con2 = 0.570, con3 = 0.785; for pLC_{50}, $q^2 = 0.541$, $r^2_{training} = 0.681$, $r^2_{test} = 0.753$, $q^2_{ext1} = 0.743$, $q^2_{ext2} = 0.736$, $q^2_{ext3} = 0.685$, con1 = 0.813, con2 = 0.836, con3 = 0.822; for pIC_{50}, $q^2 = 0.490$, $r^2_{training} = 0.684$, $r^2_{test} = 0.729$, $q^2_{ext1} = 0.568$, $q^2_{ext2} = 0.532$, $q^2_{ext3} = 0.387$, con1 = 0.823, con2 = 0.757, con3 = 0.793. With all the statistical results for four data types, it was seen that taking into account all reasonable conformers gave higher internal and external validation values.

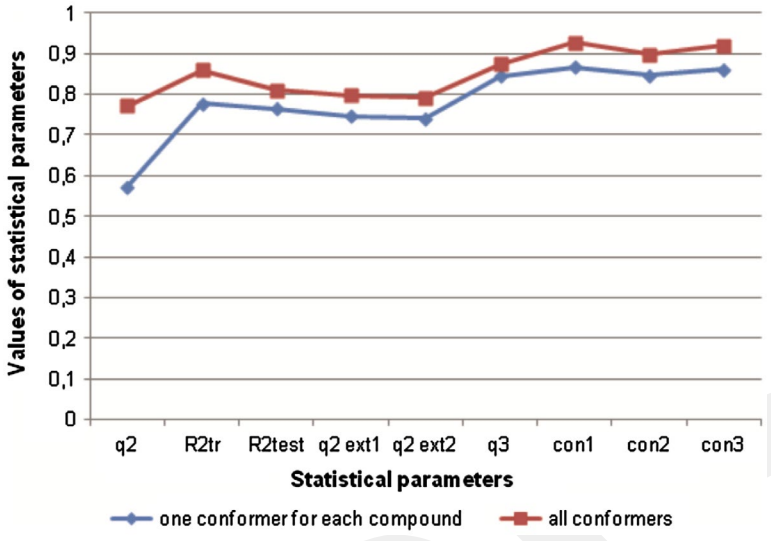


Figure 5. Comparison of statistical results of pGI_{50} activity values for C2-aryl pyrrolo[2,1-c][1,4] benzodiazepine derivatives using both only the lowest energy conformer of each compound and all conformer via optimum 11 parameters.

The statistical results of TGI, LC_{50} and IC_{50} activities containing the experimental and predicted activity values, r^2, standard error and both internal and external q^2 values for the best model obtained by the optimum number of descriptors are given in Table 4.

The best parameter subsets including 9–11 parameters which yielded the best models for the pGI_{50}, pTGI, pLC_{50} and pIC_{50} of C2-aryl PBD derivatives are the parameters suggested as contributing most to the activity. However, the contribution of each parameter is not equal. The *E*-statistic technique was used to analyse the individual effect of each parameter on the biological activity. In turn, each parameter was excluded and the model was established with other parameters. Consequently, neglecting the related parameter, the differentiation in the model performance was observed over the *E*, $r^2_{training}$, $se_{training}$, r^2_{test}, se_{test}, q^2, q^2_{ext1}, q^2_{ext2}, q^2_{ext3}, con1, con2 and con3 values that are represented in Table 5 for GI_{50} activity.

Table 5. E-statistic results for pGI_{50} activity values of C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine derivatives demonstrating how $r^2_{training}$, $se_{training}$, r^2_{test}, se_{test}, q^2, q^2_{ext1}, q^2_{ext2}, q^2_{ext3}, con1, con2 and con3 values were affected by each descriptor.

Parameter	*E*	r^2_{tr}	se_{tr}	r^2_{test}	se_{test}	q^2	$q2_{ext1}$	$q2_{ext2}$	q^2_{ext3}	*con1*	*con2*	*con3*
$a_{fit}^{(j)}$												
$a^{(1)}$	0.587	0.817	0.059	0.742	0.102	0.609	0.741	0.734	0.841	0.903	0.856	0.893
$a^{(2)}$	0.854	0.845	0.054	0.790	0.092	0.732	0.764	0.757	0.855	0.919	0.885	0.911
$a^{(3)}$	0.754	0.844	0.054	0.781	0.094	0.696	0.786	0.781	0.869	0.919	0.845	0.910
$a^{(4)}$	0.539	0.826	0.057	0.691	0.111	0.575	0.690	0.682	0.810	0.909	0.795	0.888
$a^{(5)}$	0.749	0.833	0.056	0.797	0.090	0.694	0.791	0.785	0.872	0.912	0.891	0.908
$a^{(6)}$	0.737	0.826	0.057	0.784	0.093	0.689	0.708	0.700	0.820	0.909	0.870	0.899
$a^{(7)}$	0.903	0.841	0.055	0.789	0.092	0.746	0.789	0.783	0.870	0.917	0.882	0.909
$a^{(8)}$	0.887	0.832	0.056	0.749	0.100	0.742	0.640	0.630	0.778	0.912	0.847	0.895
$a^{(9)}$	0.531	0.801	0.061	0.769	0.096	0.568	0.685	0.676	0.806	0.894	0.862	0.886
$a^{(10)}$	0.814	0.844	0.054	0.770	0.096	0.718	0.760	0.754	0.853	0.919	0.856	0.909
$a^{(11)}$	0.070	0.512	0.096	0.331	0.164	-2.279	-0.189	-0.222	0.269	0.676	0.508	0.643

Omission of a parameter in the *E*-static technique causes a decrease or increase in r^2, se and q^2 values depending on its influence on model performance.

The best model for pGI_{50} activity values generated by 11 descriptors had high $r^2_{training}$ (0.858) and q^2 (0.771) value in the training set. Upon analysis of Table 5, a remarkable decline in the r^2_{tr}, $q^{2,}$ q^2_{ext1} and q^2_{ext2} values from 0.858, 0.771, 0.797 and 0.791 to 0.512, –2.279, –0.189 and –0.222, respectively, proves that the $a^{(11)}$ parameter, which corresponds to the Fukui atomic electrophilic reactivity index value of the C17 atom, has maximal impact on the activity. The negative correlation between pGI_{50} activity values and the Fukui atomic electrophilic reactivity index also has the lowest *E* value. Hence omission of $a^{(11)}$ leads to a deterioration in the model performance. The angle between the O3 C9 N2 plane and the line of C14-C23, $a^{(7)}$, which has the highest *E* value, does not much affect the model's performance. $a^{(9)}$, $a^{(4)}$ and $a^{(1)}$ are the most potent second, third and fourth parameters; ignoring them gives a reasonable *E* value and noticeably low q^2 values compared with $a^{(11)}$. Considering the statistical values in Table 5, the contribution of parameters to the model quality is, respectively, as follows: $a^{(11)}$, $a^{(9)}$, $a^{(4)}$, $a^{(1)}$, $a^{(6)}$, $a^{(5)}$, $a^{(3)}$, $a^{(10)}$, $a^{(2)}$, $a^{(8)}$ and $a^{(7)}$.

The *E*-statistic results to determine which parameters contribute most to the pTGI, pLC_{50} and pIC_{50} activity values are listed in Tables S4–S6 (available online). Whereas the q^2 and $r^2_{training}$ values of the model with the optimum 11 descriptors based on pTGI activity values are 0.853 and 0.787, respectively, it is clearly seen neglecting the $a^{(10)}$ parameter, which is the Fukui atomic electrophilic reactivity index value (eV) of the C17 atom, obviously results in decreased q^2 (–0.245) and $r^2_{training}$ (0.638) values (see Table S4 online). In addition, remarkable negative q^2 (–0.245), q^2_{ext1} (–0.369), q^2_{ext2} (–0.429) and q^2_{ext3} (–0.790) values and the lowest *E* value (0.171) reveal how influential the $a^{(10)}$ parameter is on the activity and how essential it is for the model development as the most important contributor. The $a^{(3)}$ parameter (orthogonal distance from C6 atom to the C10 N2 O3 plane (Å)) whose *E* value (0.994) is the highest has very little effect on the model. This means that omitting the effect of the $a^{(3)}$ parameter on the activity gives an acceptable model without any loss of model performance. The orthogonal distance from the C14 atom to the N2 C9 O3 plane (Å), $a^{(5)}$, is the second most potent parameter as a geometrical parameter. Neglecting $a^{(5)}$ also gives negative q^2_{ext1}, q^2_{ext2} and q^2_{ext3} values, which affirm its impact on the activity. The descending contribution of the parameters to the biological activity is as follows: $a^{(10)}$, $a^{(5)}$, $a^{(2)}$, $a^{(9)}$, $a^{(11)}$, $a^{(8)}$, $a^{(4)}$, $a^{(6)}$, $a^{(1)}$, $a^{(7)}$ and $a^{(3)}$.

In consideration of pLC_{50} activity values, the q^2 value of the developed model with 11 parameters is 0.600. As seen from Table S5 (available online), the two most influential parameters with the lowest *E* and q^2 values are the Fukui atomic electrophilic reactivity index value of the C17 atom ($a^{(9)}$) and the nucleophilic atomic frontier electron density of the O3 atom ($a^{(8)}$). Exclusion of the $a^{(9)}$ parameter decreases the q^2 value from 0.600 to –17.770. With the lowest value of *E* (0.021), $a^{(9)}$ has the maximal impact. Moreover $r^2_{training}$, r^2_{test}, q^2_{ext1}, q^2_{ext2}, q^2_{ext3}, con1, con2 and con3 exhibit the lowest values for the situation of $a^{(9)}$. Neglecting $a^{(1)}$, the orthogonal distance from the C17 atom to the O1 O2 O3 plane +van der Waals radius (Å), we obtained relatively high statistical values of $r^2_{training}$, r^2_{test}, q^2_{ext1}, q^2_{ext2}, q^2_{ext3}, con1, con2 and con3, which indicates that it can be ignored. The $a^{(9)}$, $a^{(8)}$, $a^{(2)}$, $a^{(6)}$, $a^{(11)}$, $a^{(4)}$, $a^{(7)}$, $a^{(5)}$, $a^{(3)}$, $a^{(10)}$ and $a^{(1)}$ parameters show their contribution to activity in the given order.

For pIC_{50} activity values (Table S6, available online), the best nine parameters were taken into account. According to *E*–statistic results, the importance of the variables can be given as follows: $a^{(2)}$, $a^{(6)}$, $a^{(1)}$, $a^{(3)}$, $a^{(7)}$, $a^{(8)}$, $a^{(9)}$, $a^{(4)}$ and $a^{(5)}$. The accuracy of the model was influenced

by $a^{(2)}$ more than by the others. The orthogonal distance from the C11 atom to the N4 C12 O3 plane displays its effect by lowering all the statistical values, especially q^2_{ext1}, q^2_{ext2} and q^2_{ext3} negatively. We cannot eliminate this parameter without loss of accuracy. The variables whose effects are most negligible are $a^{(4)}$ and $a^{(5)}$. Their effects are equal to each other.

As a result, considering four types of activity it was seen that the Fukui atomic electrophilic reactivity index value (eV) of the C17 atom is the most important and essential parameter for GI_{50}, TGI and LC_{50} activities. For IC_{50} activity, the orthogonal distance is the dominant parameter.

Conclusion

In this study a mathematical model was developed for pharmacophore identification and antitumour activity prediction of 87 C2-aryl PBD derivatives by the extensive 4D-QSAR EC–GA method. For both stages of the study, a conformational ensemble of the compounds presenting molecular flexibility was used related to Boltzmann distribution. The defined pharmacophore, which is mainly located in benzodiazepine and imidazole rings, consists of eight atoms, namely the O1, O2, N1, O3, N2, C9, C14 and C17 atoms. By dividing the original data set into training and test sets, the generated QSAR models with LOO-cross-validated r^2 and q^2 values varying between 0.56 and 0.80 showed high internal and external accuracy for four types of activity and proved their robustness. The models were also applied and tested on the compounds with unknown activity to guide the employment of new bioactive benzodiazepines.

The final models and their validation results for all GI_{50}, TGI, LC_{50} and IC_{50} activities indicate that the geometrical and electrostatic descriptors used in this study are influential on the biological activity. The resulting EC–GA models and their internal and external validation for all of the dataset of pGI_{50}, pTGI, pLC_{50} and pIC_{50} activity values showed that the goodness of fit between experimental and predicted activities was over 0.700. The prediction power represented by q^2, q^2_{ext1} and q^2_{ext2} values for both training and test sets was greater than 0.6. Only for pIC_{50} activity values, the q^2_{ext1} and q^2_{ext2} values were lower than 0.6. Thus, the QSAR model of C2-aryl PBD derivatives created by the EC–GA method is a promising tool for the future design of novel benzodiazepine derivatives as antitumour agents.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Research Fund of Erciyes University under [grant number FBD-10-2980]; and the Scientific Technical Research Council of Turkey (TUBITAK) under [grant number 105T396] and [grant number 107T385].

References

[1] L.H. Hurley, *DNA and associated targets for drug design*, J. Med. Chem. 32 (1989), pp. 2027–2033.
[2] D.E. Thurston, *Advances in the study of pyrrolo[2,1-c][1,4]benzodiazepine (PBD) antitumour antibiotics in Molecular Aspects of Anticancer Drug–DNA Interactions*, S. Neidle and MJ Waring, eds., MacMillan Press, London, 1993, pp. 54–88.

[3] M.D. Tendler and S. Korman, *"Refuin": A non-cytotoxic carcinostatic compound proliferated by a thermophilic actinomycete*, Nature 199 (1963), p. 501.
[4] L.H. Hurley and R.L. Petrusek, *Proposed structure of the anthramycin–DNA adduct*, Nature. 282 (1979), pp. 529–531.
[5] D.J. Abraham, *The history of quantitative structure activity relationships*, in *Burger's Medicinal Chemistry and Drug Discovery*, C.D. Selassie, ed., John Wiley and Sons Publishers, New York, NY, 2003, pp. 1–48.
[6] E.X. Esposito, A.J. Hopfinger, and J.D. Madura, *Methods for applying the quantitative structure-activity relationship paradigm*, Meth. Mol. Biol. 275 (2004), pp. 131–213.
[7] S.J. Free and J. Wilson, *A mathematical contribution to structure activity studies*, J. Med. Chem. 7 (1964), pp. 395–399.
[8] C. Hansch and T. Fujita, *ρ-σ-π Analysis: A method for the correlation of biological activity and chemical structure*, J. Am. Chem. Soc. 86 (1964), pp. 1616–1626.
[9] R. Cramer, D. Patterson, and J. Bunce, *Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins*, J. Am. Chem. Soc. 110 (1988), pp. 5959–5967.
[10] G. Klebe, *Comparative molecular similarity indices analysis: CoMSIA*, in *3D QSAR Drug Design*, Vol. 3, H. Kubinyi, G. Folkers, and Y.C. Martin, eds., Kluwer Academic Publishers, Newyork, 1998, pp. 87–104.
[11] A.J. Hopfinger, S. Wang, J.S. Tokarski, B. Jin, M. Albuquerque, P.J. Madhav, and C. Duraiswami, *Construction of 3D-QSAR models using the 4D-QSAR analysis formalism*, J. Am. Chem. Soc. 119 (1997), pp. 10509–10524.
[12] A. Vedani and M. Dobler, *5D-QSAR: The key for simulating induced fit?*, J. Med. Chem. 45 (2002), pp. 2139–2149.
[13] A. Vedani, M. Dobler, and M.A. Lill, *Combining protein modeling and 6D-QSAR-Simulating the binding of structurally diverse ligands to the estrogen receptor*, J. Med. Chem. 48 (2005), pp. 3700–3703.
[14] J. Polanski, *Receptor dependent multidimensional QSAR for modeling drug-receptor interactions*, Curr. Med. Chem. 16 (2009), pp. 3243–3257.
[15] A. Kamal, E.V. Bharathi, M.J. Ramaiah, D. Dastagiri, J.S. Reddy, A. Viswanath, and F. Sultana, S.N.C.V.L. Pushpavalli, M. Pal-Bhadra, H.K. Srivastava, G.N. Sastry, A. Juvekar, S. Sen, and S. Zingde, *Quinazolinone linked pyrrolo[2,1-c][1,4]benzodiazepine (PBD) conjugates: Design, synthesis and biological evaluation as potential anticancer agents*, Bioorg. Med. Chem. 18 (2010), pp. 526–542.
[16] D. Antonow and D.E. Thurston, *Synthesis of DNA-interactive pyrrolo[2,1-c][1,4]benzodiazepines (PBDs)*, Chem. Rev. 111 (2011), pp. 2815–2864.
[17] L. Cipolla, A.C. Araújo, C. Airoldi, and D. Bini, *Pyrrolo[2,1-c][1,4]benzodiazepine as a scaffold for the design and synthesis of anti-tumour drugs*, Anticancer Agents Med. Chem. 9 (2009), pp. 1–31.
[18] Y. Ohtake, A. Naito, H. Hasegawa, K. Kawano, D. Morizono, M. Taniguchi, Y. Tanaka, H. Matsukawa, K. Naito, T. Oguma, Y. Ezure, and Y. Tsuriya, *Novel vasopressin V2 receptor-selective antagonists, pyrrolo[2,1-a]quinoxaline and pyrrolo[2,1-c][1,4]benzodiazepine derivatives*, Bioorg. Med. Chem. 7 (1999), pp. 1247–1254.
[19] C. Paulussen, K. de Wit, G. Boulet, P. Cos, L. Meerpoel, and L. Maes, *Pyrrolo[1,2-a][1,4]benzodiazepines show potent in vitro antifungal activity and significant in vivo efficacy in a* Microsporum canis *dermatitis model in guinea pigs*, J. Antimicrob. Chemother. 69 (2014), pp. 1608–1610.
[20] D. Antonow, M. Kaliszczak, G.D. Kang, M. Coffils, A.C. Tiberghien, N. Cooper, T. Barata, S. Heidelberger, C.H. James, M. Zloh, T.C. Jenkins, A.P. Reszka, S. Neidle, S.M. Guichard, D.I. Jodrell, J.A. Hartley, P.W. Howard, and D.E. Thurston, *Structure-activity relationships of monomeric C2-aryl pyrrolo[2,1-c][1,4]benzodiazepine (PBD) antitumor agents*, J. Med. Chem. 53 (2010), pp. 2927–2941.
[21] E. Sarıpınar, N. Geçen, K. Şahin, and E. Yanmaz, *Pharmacophore identification and bioactivity prediction for triaminotriazine derivatives by electron conformational-genetic algorithm QSAR method*, Eur. J. Med. Chem. 45 (2010), pp. 4157–4168.
[22] I.B. Bersuker and A.S. Dimoglo, *The electron-topological approach to the QSAR problem*, in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd, eds., John Wiley and Sons Publisher Inc, New Jersey, 1991, pp. 423–460.

[23] I.B. Bersuker, *Pharmacophore identification and quantitative bioactivity prediction using the electron-conformational method*, Curr. Pharm. Des. 9 (2003), pp. 1575–1606.
[24] I.B. Bersuker, S. Bahçeci, and J.E. Boggs, *Improved electron-conformational method of pharmacophore identification and bioactivity prediction. Application to angiotensin converting enzyme inhibitors*, J. Chem. Inf. Comput. Sci. 40 (2000), pp. 1363–1376.
[25] N. Sukumar, G. Prabhu, and P. Saha, *Applications of genetic algorithms in QSAR/QSPR Modeling*, in *Applications of Metaheuristics in Process Engineering*, J. Valadi and P. Siarry, eds., Springer International Publishing, Switzerland, 2014, pp. 315–324.
[26] E. Yanmaz, E. Sarıpınar, K. Şahin, N. Geçen, and F. Çopur, *4D-QSAR analysis and pharmacophore modeling: Electron conformational-genetic algorithm approach for penicillins*, Bioorg. Med. Chem. 19 (2011), pp. 2199–2210.
[27] K. Şahin, E. Sarıpınar, E. Yanmaz, and N. Geçen, *Quantitative bioactivity prediction and pharmacophore identification for benzotriazines derivatives by electron conformational-genetic algorithm QSAR method*, SAR QSAR Environ. Res. 22 (2011), pp. 217–238.
[28] N. Geçen, E. Sarıpınar, E. Yanmaz, and K. Sahin, *Application of electron conformational–genetic algorithm approach to 1,4-dihydropyridines as calcium channel antagonists: Pharmacophore identification and bioactivity prediction*, J. Mol. Model. 18 (2012), pp. 65–82.
[29] L. Akyüz, E. Sarıpınar, E. Kaya, and E. Yanmaz, *4D-QSAR study of HEPT derivatives by electron conformational-genetic algorithm method*, SAR QSAR Environ Res. 23 (2012), pp. 409–433.
[30] L. Akyüz and E. Sarıpınar, *Conformation depends on 4D-QSAR analysis using EC–GA method: Pharmacophore identification and bioactivity prediction of TIBOs as non-nucleoside reverse transcriptase inhibitors*, J. Enzyme Inhib. Med. Chem. 28 (2013), pp. 776–791.
[31] Spartan'10; Wavefunction, Inc.: Irvine, CA, 2011.
[32] I.B. Bersuker, *QSAR without arbitrary descriptors: The electron-conformational method*, J. Comput. Aided. Mol. Des. 22 (2008), pp. 423–430.
[33] A.S. Dimoglo, P.F. Vlad, N.M. Shvets, and M.N. Coltsa, *Electronic-topolocigal investigations of the relationship between chemical structure and ambergris odor*, New J. Chem. 19 (1995), pp. 1217–1226.
[34] E. Sarıpınar, Y. Güzel, Ş. Patat, İ. Yıldırım, Y. Akçamur, and A.S. Dimoglo, *Electron-topological investigation of structure-antitubercular activity relationship of thiosemicarbazone derivatives*, Arzneimittelforschung 46 (1996), pp. 824–828.
[35] I.B. Bersuker, S. Bahçeci, J.E. Boggs, and R.S. Pearlman, *An electron conformational method of identification of pharmacophore and anti-pharmacophore shielding: Application to rice blast activity*, J. Comput. Aided. Mol. Des. 13 (1999), pp. 419–434.
[36] MATLAB (ver 7.0), The MathWorks Inc, 3 Apple Hill Drive, Natick, MA 01760-2098.
[37] J.H. Holland, *Adaptation in Artificial and Natural Systems*, University of Michigan Press, Michigan, Ann Arbor, 1975.
[38] J. Devillers, *Principles of QSAR and Drug Design: Genetic Algorithms in Molecular Modeling*, Academic Press, Lyon, 1996.
[39] G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, and R. Kuhne, *External validation and prediction employing the predictive squared correlation coefficient - test set activity mean vs training set activity mean*, J. Chem. Inf. Model. 48 (2008), pp. 2140–2145.
[40] V. Consonni, D. Ballabio, and R. Todeschini, *Comments on the definition of the Q2 parameter for QSAR validation*, J. Chem. Inf. Model. 49 (2009), pp. 1669–1678.
[41] N. Chirico and P. Gramatica, *Real external predictivity of QSAR models: How to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient*, J. Chem. Inf. Model. 51 (2011), pp. 2320–2335.
[42] Al H. Makkouk and I.B. Bersuker, and J.E. Boggs, *Quantitative drug activity prediction for inhibitors of human breast carcinoma*, Int. J Pharm. Med. 18 (2004), pp. 81–89.
[43] J.G. Topliss and R.P. Edwards, *Chance factors in studies of quantitative structure-activity relationships*, J. Med. Chem. 22 (1979), pp. 1238–1244.
[44] J. Wan, L. Zhang, and G. Yang, *Quantitative structure–activity relationships for phenyl triazolinones of protoporphyrinogen oxidase inhibitors: A density functional theory study*, J. Comp. Chem. 25 (2004), pp. 1827–1832.

[45] W. Long and P. Liu, *Quantitative structure activity relationship modeling for predicting radiosensitization effectiveness of nitroimidazole compounds*, J. Radiat. Res. 51 (2010), pp. 563–572.
[46] R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics*, Vol. 41, Wiley-VCH, Weinheim, 2009, pp. 625–626.
[47] A. Golbraikh and A. Tropsha, *Beware of q2!*, J. Mol. Graph. Model. 20 (2002), pp. 269–276.