Received: 18 October 2019 | Revised: 18 December 2019 | Accepted: 18 December 2019

DOI: 10.1002/jcc.26154

Check for updates

FULL PAPER

Journal of COMPUTATIONAL CHEMISTRY WILEY

A novel hybrid method named electron conformational genetic algorithm as a 4D QSAR investigation to calculate the biological activity of the tetrahydrodibenzazosines

Kader Sahin[1] | Emin Saripinar[2]

[1]Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey

[2]Science Faculty, Department of Chemistry, Erciyes University, Kayseri, Turkey

Correspondence
Kader Sahin, Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey.
Email: kader.sahin@med.bau.edu.tr

Funding information
Erciyes Üniversitesi

Abstract

To understand the structure–activity correlation of a group of tetrahydrodibenzazocines as inhibitors of 17β-hydroxysteroid dehydrogenase type 3, we have performed a combined genetic algorithm (GA) and four-dimensional quantitative structure–activity relationship (4D-QSAR) modeling study. The computed electronic and geometry structure descriptors were regulated as a matrix and named as electron-conformational matrix of contiguity (ECMC). A chemical property-based pharmacophore model was developed for series of tetrahydrodibenzazocines by EMRE software package. GA was employed to choose an optimal combination of parameters. A model has been developed for estimating anticancer activity quantitatively. All QSAR models were established with 40 compounds (training set), then they were considered for selective capability with additional nine compounds (test set). A statistically valid 4D-QSAR ($R^2_{training}$ = 0.856,R^2_{test} = 0.851 and q^2 = 0.650) with good external set prediction was obtained.

KEYWORDS

drug design, genetic algorithm, molecular modeling, pharmacophore, tetrahydrodibenzazocines

1 | INTRODUCTION

Hydroxysteroid dehydrogenases (HSDs) produce a group of aldo-keto reductase and dehydrogenase/reductase enzymes, and they catalyze carbonyl or alcohol stereospecific oxide reduction responses. 17β-HSDs play an important role in hormonal function and regulation in humans.

So these enzymes contain attractive therapeutic goals for diseases like prostate cancer, breast cancer, benign prostate hyperplasia, hair loss, and acne.[1] The regulation or effect of androgen biosynthesis on androgen receptor (AR) takes place in the treatment of prostate cancer (PCa).[2]

17β-HSDs were found to be elevated not only in testis and seminal vesicles but also in prostate tissue, indicating its possible involvement in both gonadal and nongonadal biosynthesis of testosterone.[3–5] 17b-HSD3 directs the production of nongonadal active hormone and converts androstenedione to testosterone.[6,7]

17β-HSDs are a desired molecular objective in the definition of prostate cancer inhibitors due to its role in the biosynthesis of testosterone. Although the achievement of clinical tests exist, new cures and agents have restrictions pertained to the development of obtained drug resistance and side effects.[8] There is a need for therapeutic agents that retain the original domain of action with fewer side effects.[9] The discovery of compounds that may act against cancer has always attracted the attention of medical chemists. In this respect, computer-aided drug design (CADD) modalities have raised as a promising tool in the seek for novel compounds.[10]

The traditional design of new drugs in drug development is an inconvenient, interdisciplinary, costly, and time-consuming process.[11] It was assumed that overall U.S. pharmaceutical industry spending on growth and exploration in the early 1990s amounted to over $20 billion a year, and this picture has been rising. The average cost of each new

 © 2020 Wiley Periodicals, Inc.

study of a laboratory drug is a startling amount: $400–$650 million[12–14] and the whole operation can take up to 14 years.[15]

Therefore, significant problem here is the early estimation of the activity-related features of drug candidates. So a suitable and effective method to estimate biological activity from structural data is an obligation. A drug design involves finding a molecule that has specific activity against a damaging organism.[16] There are many powerful techniques based on a quantitative connection between structure and activity (QSAR) that guarantee an estimate of activity.[11]

QSARs are useful in comprehending and clarifying the mechanism of drug effect and they enable the improvement and design of novel compounds ensuring intended biological features.[17] QSAR studies characterize compounds' biological activity based on their diverse structural characteristics and how biodiversity is dependent on variations in chemical composition.[18] Drug design endeavors to demonstrate a mathematical relation between molecular features and biological activity of a compound. The structural features of the molecules are mostly represented by a set of parameters. These parameters' values affect molecules' activity in some way.[19,20]

QSARs indicate mathematical relationships between chemical, biological, physical, or environmental considerations as well as computable or quantifiable variables such as electronic, physicochemical, topological, or stereo chemical indices.[21,22] There are many examples of successful use of SAR or QSAR models in the literature.[23–31] The thorough description of the characteristics of each molecular structure is assigned to 3D-QSAR. They are affected by the particular shape of a molecule.

In the design of quantitative 3D pharmacophore models for a ligand derivative sequence, 4D-QSAR is useful and accurate when the related receptor structure is identified.[32] 4D-QSAR models resemble to 3D QSAR models, but in 4D QSAR, the constructional knowledge is assumed for a group of conformers (the fourth dimension), in place of one conformation.[33,34] The bioactivity of flexible molecules depend on molecular stability and conformation constraints.[35]

In the 4D-QSAR research, we are able to form a lot of conformers covering separate spatial areas for a single molecule. In reality, the common 3D patterns being searched in a group of molecules during molecular dynamic simulations[36–38] is a possibility. The electron-conforming technique (ECM) is based on the premise that chemical compounds' electronic and conforming features influence their biological features.[39]

In the ECM outline, each molecule is symbolized by a group of electronic and conformational variables. The structural and physicochemical features of the molecules are generally symbolized by a group of parameters (descriptors). The notion of pharmacophore is commonly used in drug design and is generally described as a three-dimensional map of biological features common in all active conformations of a group of compounds with a specific activity.[40]

Many techniques for discovering pharmacophore and its practices have been improved.[41] Dimoglo et al.[42] proposed the technique of topological electron (ET). A QSAR method used to define a pharmacophore is the electron conformational (EC) method.[43] In addition, the EC technique proposed by Bersuker is used with a nonlinear mathematical model to estimate bioactivity.[44]

In drug design, Bersuker et al. offer the Pha finding method of EC and quantitative bioactivity estimation, which is essentially different from ETM.[45] This technique explores one molecule[46] conformer symbolized by matrices called electronic-topological matrices congruity.

In order to acquire a QSAR model, three kinds of products are crucial. First, there must be a quantified biological activity for a compound group. Second, chemical substances must be described by their topological indices or physicochemical characteristics.[47,48] Finally, a statistical method must be used for connecting the first two components.[47,49] Bersuker's EC technique is intended to search for guidelines for distinct activity estimation.[50]

QSAR studies can be approximately splitted into two types: classification and regression. QSAR models consist essentially of statistical methods for chemical data sets such as machine learning and statistical literature. There are a lot of methods/algorithms for selection and optimization of diverse parameters entailed in the model building procedure such as the Nelder–Mead simplex algorithm and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[51–54] or stochastic methods such as genetic algorithms (GAs),[55–57] simulated annealing,[58,59] and evolutionary algorithms.[60,61]

A set of parameters usually symbolize the physicochemical and structural characteristics of molecules, and the assumption is that the activity of the molecule involves the numerical values of these parameters in some way. Not all parameters for a particular molecule are crucial in QSAR studies. In many QSAR approaches, various methods such as primary component analysis (PCA),[62] partial least squares (PLS),[63] GA,[64–68] and neural networks (NNs)[69] were used to pick precious parameters and to represent predictive models.

GAs comprise a sort of search algorithms particularly suited to solving complicated optimization problems. We offer in this study the main properties of GAs to solve arduous design problems. We shortly introduce the main concepts of GAs such as genetic operators, representation, selection, and fitness evaluation.

It is probable to profit from excellent characteristics of each technique and decrease the mistakes by integrating distinct techniques into a hybrid method. In this research, the EC-GA technique is a mixture of GA and conformational electron method and was used to select the structural parameters linked to the binding receptor affinity of a group of tetrahydrodibenzazocines and to explain the pharmacophore (Pha).

This technique has been effectively implemented as a 4D-QSAR technique to define the pharmacophore group for benzotriazine, penicilin, triaminotriazine derivatives, pyrazole pyridine, ruthenium(II) arene complexes, and quantitative activity estimates in our past research.[70–75]

2 | MATERIALS AND METHODS

In our work, structure–anticancer activity relations have been studied for the group of tetrahydrodibenzazocines by using EC-GA method. The biological data were obtained from the literature.[76] In QSAR and

pharmacophore modeling, the logarithm of quantified IC_{50} (μM) values was used to correlate linear data with free energy change. The logarithmic transformation of IC_{50} values was used to minimize prospective mistakes.

The 3D structures of the molecules were constituted by using the Spartan software (version 0.6)[77] and after energy is minimized using MMFF molecular mechanics. Later the structures were completely optimized using the semiemperical method PM3 level theory. Quantum-chemical semiempirical PM3 calculations were carried out for the resulting in 49 molecules. The conformational manner of all compounds was searched using molecular mechanics. Conformational evaluation was conducted at PM3 level using the randomized search technique used by Monte Carlo.[78] After that, conformers with imaginary frequencies are withdrawn. The presence of an imaginary frequency leads the molecule to a state of transition and does not fully optimize the molecule.[79]

Then we calculated the optimized molecular structures in the software Spartan, the Spartan files were transferred into software EMRE to calculate all kinds of parameters. The software EMRE can extract and calculate topological, geometrical, constitutional, electrostatic, and quantum chemical parameters and has been used in a lot of QSAR studies. The compounds chosen for the 4D-QSAR investigations and their predicted biological activities are given in Table 1. Forty-nine compounds employed in the study, 40 compounds were used to set up model as training set, and 9 compounds were used as test set. The selection was carried out randomly for compounds of training and test set.

The basic steps for EC-GA study are shown in Figure 3. Select the training set of compounds with the acknowledged bioactivity. Carry out electronic structure computations and 3D geometry optimization for all of the molecules. The minimized design generated a conformational set with an energy limit of 1.5 kcal/mol. A maximum of 100 conformers per molecule was generated. The flexibilities of the gathered molecules were considered by counting each molecule as an accumulation of conformers standing for distinct areas of the conformational space. For each compound in the series, the $n \times n$ matrices named ECMC were constituted[80] from the output file, where n symbolizes the atom numbers in the compound. ECMC includes both geometric and electronic characteristics that are supposed to adequately represent the characteristics of the molecules under consideration.

a_{ii} is the diagonal component and one of the electronic atomic characteristics, for example, valence activities, atomic charges, atomic polarizabilities HOMO or LUMO in the ECMC, i is the ith atom in the compound. a_{ij} is off-diagonal component and it has two types. If i and j label bonded atoms, a_{ij} is electronic parameter such as wiberg index, bond energy (ionic covalent or total), bond order, and bond polarizability. If i and j label nonbonded atoms, a_{ij} is the distance between ith and jth atoms. ECMC shall be generated for each compound in series.[81]

To detect those matrix components that exist in all active compounds but not in inactive compounds,[82] select the ECMC as a model for one of the active compounds and compare it within certain tolerance limits with the other ECMCs. The identified submatrix is called the contiguity for activity (ECSA) EC submatrix. The pharmacophore called the activity property is thus determined as a submatrix containing both interatomic distances and structural electron descriptors.

After detecting the ECSA with the reduced number of matrix components, it is easy to detect the flexibility of each descriptor. If the identified submatrices are not sufficiently replaced, replace the tolerance limits until appropriate fragments are identified. If the obtained properties return compelling outcomes, the ECSA will be assessed conveniently for the assessment of the activity. If not, the deeper investigation of the deficiency of the activity feature should be undertaken and complete steps should be repeated. Figure 1 shows the ECMC of the smallest energy conformer in the tetrahydrodibenzazocine group of the most active compound (Compound 45).

PFB > 7.1 compounds were classified as high-activity compounds (25 compounds) and PFB < 7.1 molecules were identified as low-activity compounds (24 compounds) to begin the classification of pharmacophore atoms. To obtain pharmacophore groups, a template active compound and the rest of the compound series are contrasted as weighted graphs. For each compound taken as a model (low active or high active), its ECMC has been contrasted with the ECMCs of the other compounds.

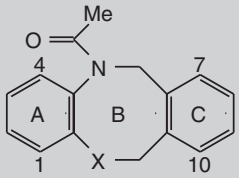
Restrictions on flexibility are important for the return of the Pha. For example, further growth of the upper limit of any submatrix components results in weakly active compounds with high activity characteristics.[50,83] Thereafter, peak tolerance values for all conformers of all compounds are calculated without any limitations on tolerance values and ECSAs are acquired. Through the improvement of the 4D-QSAR modality, one significant job is to get the best activity estimation function. Most of the active conformer is not the smallest energy conformer and consequently the compounds' flexibility is necessarily ignored (Figure 2).

Since tiny energy differences between conformers can lead to important diversity in the electronic structure, it is important to consider the Boltzmann populations and the dynamics of these conformational components of all compounds in the 4D-QSAR[32] strategy in order to comprehend the impact on biological activity of all energetically stable conformers.[84] In this research, the final activity formula (Eqn. (1)) to assess the bioactivity included not only the smallest energy conformation but also all plausible conformers provided in Table 1. The contributions of the compound's separate conformations are regarded through the distribution of Boltzmann. In Eqn. (1), Bersuker[50] used a general formula of activity.

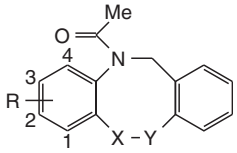
$$A_n = A_0 \frac{\sum_{i=1}^{m_e} \delta_{ni}[\mathrm{Pha}] e^{-S_{ni_e} - E_{ni}/kT}}{\sum_{i=1}^{m_e} e^{-E_{ni}/kT}} \quad (1)$$

where δ is the Kronecker delta. This is a function of two parameters that is 1 if the pharmacophore exists and 0 if not. A_n and A_l are the activities of the nth compound and the reference molecule, and m_n is the number of conformations of the nth molecule. E_{li} is the relative energy of the ith conformation of the reference compound, E_{ni} is the

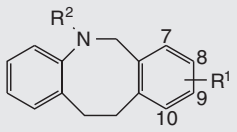
TABLE 1 Molecular structures, calculated and experimental activity values for tetrahydrodibenzazocines series



No	X	m_n	A^{exp}	A^{calc}
⋆1	S	3	6.398	7.027
2	CH_2	3	7.046	7.147
3	O	2	5.527	5.639
4	NH	3	5.324	5.212



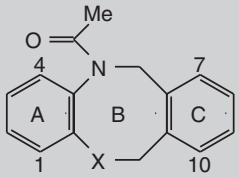
No	X-Y	R	m_n	A^{exp}	A^{calc}
⋆5	S—CH_2	2-cl	3	7.000	7.032
6	S—CH_2	2-cl	2	7.523	7.963
7	S—CH_2	2-F	3	7.046	7.505
⋆8	S—CH_2	2-Br	2	6.602	7.285
9	S—CH_2	2,3-diCl	2	8.000	7.402
⋆10	S—CH_2	4-cl	1	4.939	4.846
11	S—CH_2	1,2-diCl	3	6.119	6.491
12	S—CH_2	3-CF_3	2	5.551	5.537
13	S—CH_2	2-OMe	9	5.686	5.109
14	S—CH_2	2-OH	6	5.313	4.724
15	S—CH_2	2-SO_2Me	8	5.000	5.101
16	S—CH_2	2-Me	2	6.000	6.574
17	C_2H_2	2-Cl	2	6.620	6.096
18	C_2H_4	2-Cl	2	7.155	7.023
19	C_2H_4	3-Cl	3	7.301	7.522



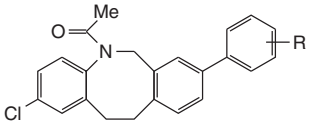
No	R^1	R^2	m_n	A^{exp}	A^{calc}
20	8-Br	C(O)Me	3	8.046	7.351
⋆21	9-Br	C(O)Me	3	8.155	7.718
22	9-Br	C(O)NH_2	2	6.553	5.901
⋆23	9-Br	Et	3	6.000	5.733
24	9-Br	C(O)c-Prop	4	6.495	6.838
25	9-NH_2	C(O)Me	6	6.310	6.033
⋆26	9-C(O)NH_2	C(O)Me	6	6.000	6.312

(Continues)

TABLE 1 (Continued)



No	X		m_n	A^{exp}	A^{calc}
27	9-NHBn	C(O)Me	25	8.097	7.994
28	9-NHC(O)Ph	C(O)Me	3	6.796	7.362
29	C(O)NHnBu	C(O)Me	42	7.456	7.886
30	8-Ph	C(O)Me	3	10.699	10.699



No	R	m_n	A^{exp}	A^{calc}
31	H	3	8.398	9.493
32	*o*-OMe	12	8.699	8.276
33	*m*-OMe	17	8.097	9.344
34	*p*-OMe	18	8.523	8.905
⋆35	*o*-CN	6	8.301	8.682
36	*m*-CN	4	8.523	8.567
37	*p*-CN	3	8.523	8.536
38	*o*-$NHSO_2Me$	15	8.699	8.092
39	*m*-$NHSO_2Me$	12	6.854	6.658
40	*m*-CH_2NH_2	11	6.538	6.537
41	3'-Piridil	9	7.398	8.870
42	4'-Piridil	2	7.398	7.516
43	*o*-CO_2Me	13	10.699	9.627
44	*m*-CO_2Me	12	8.854	9.374
⋆45	*o*-C(O)Me	2	10.699	9.111
46	*o*-CH_2OH	7	9.046	9.036
47	*o*-CH_2NH_2	3	6.622	7.499
48	*o*-C(O)NHMe	2	7.602	7.180
49	*o*-CO_2Me	13	9.155	9.636

Labeled with an'⋆' symbol correspond to the nine compounds of the test set m_n conformer number.

relative energy of the *i*th conformation of the *n*th compound (in kcal. Mol^{-1}), R (kcal. $Mol^{-1}.K^{-1}$) is the gas constant, and *T* is the temperature in Kelvin.

Auxiliary group (AG) and antipharmacophore shielding (APS) affect their activity in all the molecular structures under account. This is the reason why some are more effective than others among the compounds that include a pharmacophore. It is then possible to define the APS and AG as molecular descriptors. The impact of these descriptors is the sum of all these factors by using the feature *S* as follows:

$$S_{ni} = \sum_{j=1}^{N} \kappa_j a_{ni}^{(j)} \quad (2)$$

where $a_{ni}^{(j)}$ is the parameter defining the *j*th type of impact in the *n*th molecule's *i*th conformation, and *N* is the amount of parameters chosen. The activity is exponentially dependent on $S(A\ e^S)$; thus, it is presumed that *S* considers the specific properties of the drug-receptor interaction that quantitatively detect the activity.

κ_j, the adjustable (variational) constants are calculated using the function S considering the Boltzmann population depends on the

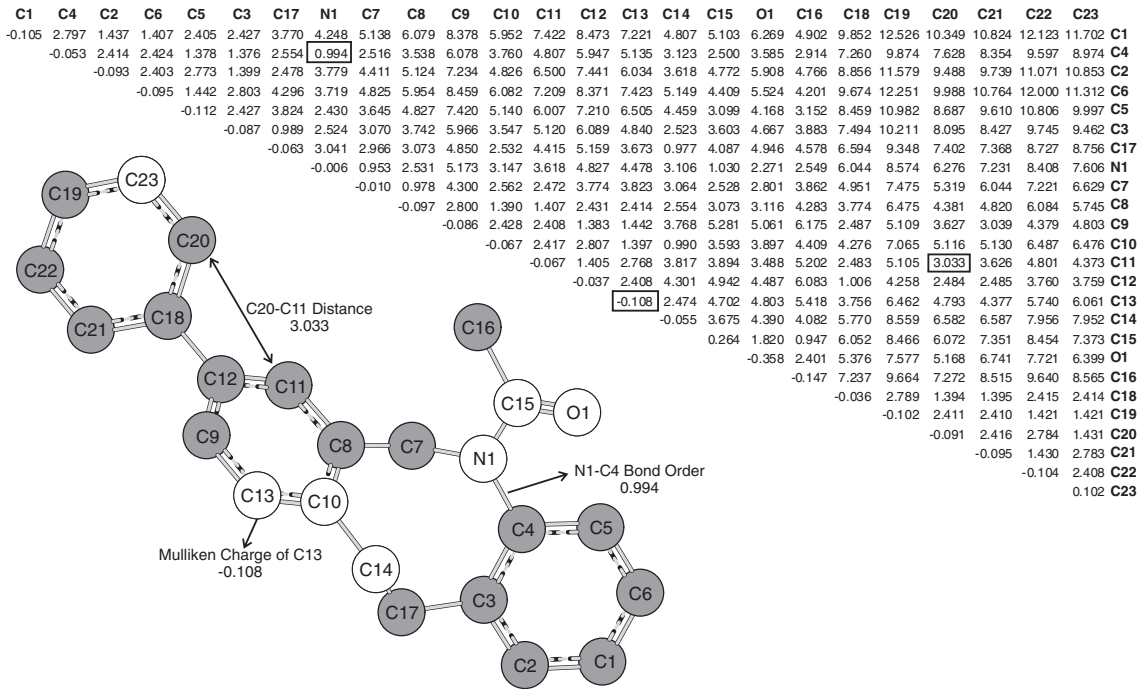


FIGURE 1 Model molecule's (compound 45) lowest energy conformer ECMC. ECMC, electron-conformational matrix of contiguity

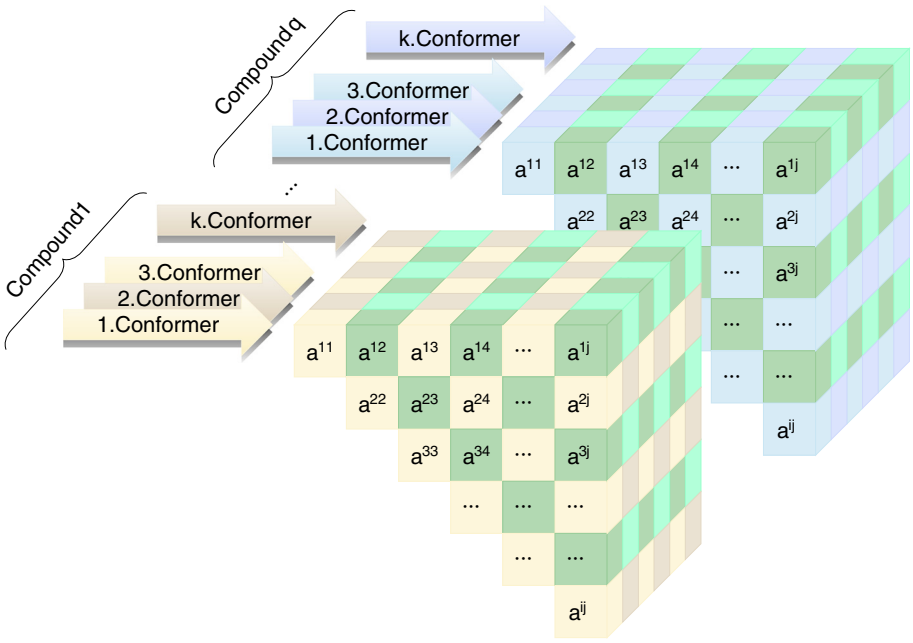


FIGURE 2 3D appearance of ECMC. ECMC, electron-conformational matrix of contiguity [Color figure can be viewed at wileyonlinelibrary.com]

energy and temperature of each conformation. Within the statistics toolbox in MATLAB,[85] the lsqnonlin function was used to obtain the values of the respective model parameters by numerically solving the scheme of differential equations for the highest parameter subset. Parameters capture certain molecular structure characteristics and are generally classified into parameter classes based on the data they represent[86] (Figure 3).

QSAR modeling's main assumption is that molecular structure can be associated with physical or biological characteristics. The main requirement is therefore some technique of coding various structural characteristics in a molecule. This needs to be met by molecular parameters. Parameters are numerical depictions (generally) of particular molecular characteristics. A total of 680 parameters in four categories were calculated: (a) electronic, (b) thermodynamic, (c) geometrical, and (d) quantum-chemical parameters.

In terms of its 3D coordinates, geometric parameters characterize the shape and size of the molecule. As a result, precise coordinates are engaged, so that before these parameters can be calculated, the

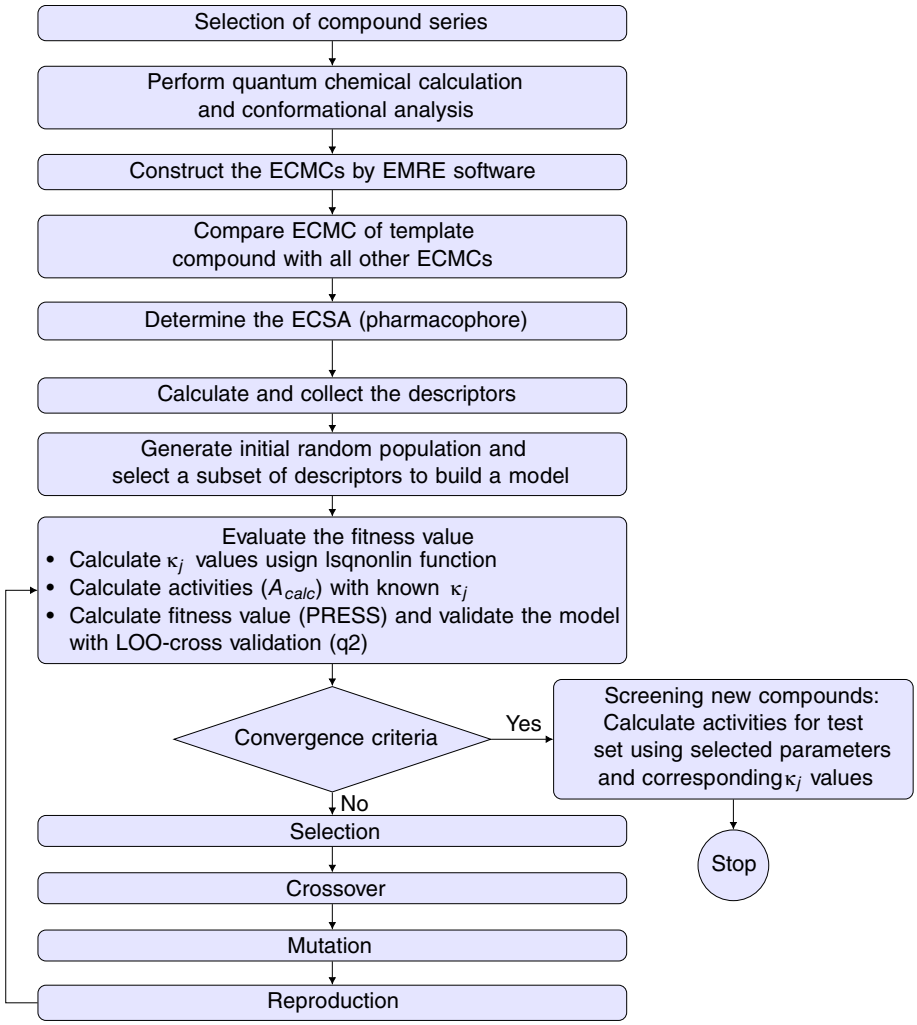


FIGURE 3 Schema the suggested hybrid EC-GA method. EC-GA, electron conformational genetic algorithm [Color figure can be viewed at wileyonlinelibrary.com]

structure needs to be optimized geometry. Examples include inertia moment,[87] molecular surface area and volumes,[88] and parameters of shadow.[89,90] Electronic parameters take the various characteristics of the electronic setting of the molecules under consideration. These comprise the energy of LUMO and HOMO, electronegativity, and various partial charge parameters centered on the atom. The quantum chemical parameters such as hardness, chemical potential, ionization potential, electron affinity, and electrophilicity are calculated direct from orbital energies of the optimized geometries.[91] Thermodynamic parameters characterize the state of the system. There are two basic categories: extensive parameters and intensive parameters. Extensive parameters are those that scale with the system; entropy, enthalpy, mass, volume, and internal energy. Intensive parameters do not scale with the system; pressure and temperature, chemical potential, and Gibbs free energy.

For the choice of the most significant parameters, the GA technique was used. The objective of the parameter selection method is to select the minimum quantity of suitable structural parameters to identify the biological activity. Optimization in QSAR modeling is a main problem. This material contains two significant applications of optimization methods. The first includes optimizing the parameters of a model in a previously defined neural feed-forward network, such as biases and weights. The other region where optimization plays a significant role is in the selection process.

Here, the selection of compounds from a library or library design can be expressed from distinct products[92,93] or the choice of parameters to construct models can be expressed. A GA belongs to the group of algorithms known as evolutionary algorithms that benefit ideas of biological evolution to support effective solutions to optimization.[57] In the field of QSAR modeling,[94–96] cheminformatics,[97,98] and chemometrics,[99–102] GA's have been commonly used.

This study focuses on applying GAs to investigate many dimensional spaces as effective tools. In particular, one implementation of GA's in QSAR modeling is to investigate a parameter space to find optimal subsets of parameters that can be used to create selective models. GA is a simulation method that mimics some of the procedures observed in natural evolution and was efficiently conducted in QSAR analysis for the selection of characteristics.

In our study, data reduction was at first performed employing GA in MATLAB software (version7.0) with the following set-up: number of iteration, 500; number of generation, 500; population size, 500; probability of mutation, 0.01; probability of cross-over, 0.5.

QSAR models with 10 parameters (according to the 5:1 rule of thumb) were constituted. The physical meaning of each parameter is

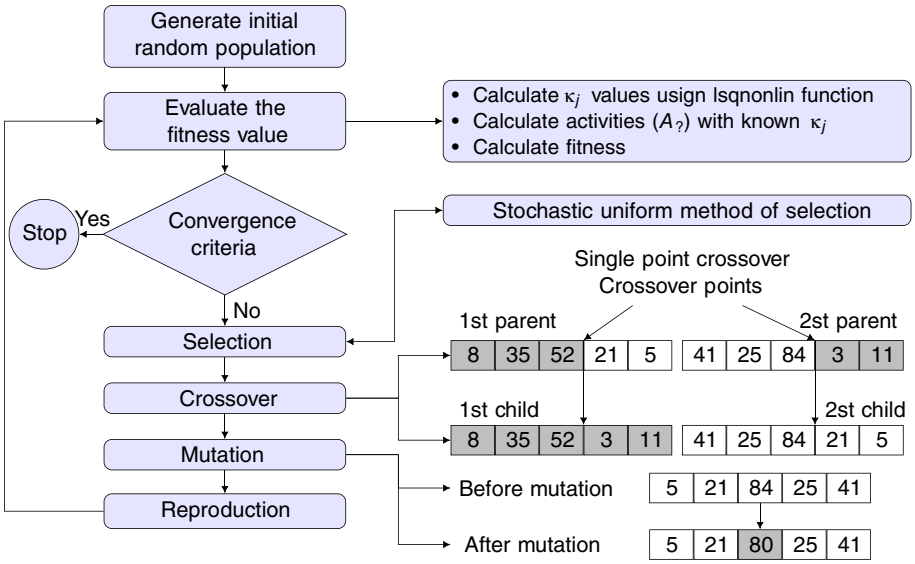


FIGURE 4 Schema of genetic algorithm [Color figure can be viewed at wileyonlinelibrary.com]

mentioned. The fitness function has a significant impact on a GA procedure's convergence velocity. We used as a fitness function the predictive remaining amount of squares (PRESS) (Eqn. (3)). In this research, the fitness value for each parameter was calculated by cross-validation leave-one-out (LOO-CV).[103] For many valid examples *n*, the PRESS is a standard index for assessing the accuracy of a modeling method based on the LOO cross-validation technique. PRESS is defined as the amount of the square difference between the experimental (exp) values and the expected (pred) values (Figure 4):

$$\mathrm{PRESS}_N = \sum_{n=1}^{N} \left| A_n^{\mathrm{exp}} - A_n^{\mathrm{calc}} \right|^2 \tag{3}$$

where A_n^{exp} is the experimental activity and A_n^{calc} is the activity foreseen in the LOO cross validation model, using *N* parameters (in this event *N* = 10). The cross-validated correlation coefficient q^2 can then be written as a measure of the model's predictive results.[104]

$$q^2 = 1 - \frac{\sum_{n=1}^{N} \left| A_n^{\mathrm{exp}} - A_n^{\mathrm{calc}} \right|^2}{\sum_{n=1}^{N} \left| A_n^{\mathrm{exp}} - \bar{A}_n^{\mathrm{exp}} \right|^2} \equiv 1 - \frac{\mathrm{PRESS}}{\mathrm{SSY}} \tag{4}$$

where *N* is the total number of training compounds in the entire data set; where SSY is the sum of the squares of deviations between the experimental values (A_n^{exp}) and their mean; and where A_n^{exp} and average zA_n^{exp} are the measured and averaged values of the dependent parameter (over the entire training set).

The small PRESS implies the model's excellent predictability. SSY is the sum of squares associated with the appropriate sources of change. A value of q^2 higher than 0.5 is often assessed as proof of the model's elevated predictive capacity (i.e., these criteria are used by CoMFA software marketed by Tripos as a possible gage of the predictive power of a model).[105]

If q^2 is low, it can be thought that the model has low predictive capability but the contrary might not be true because we have proved lately[106] that high q^2 do not mean that the model has important predictive power. The predictive power of a QSAR model can be commented by its capability to estimate correctly the target feature of molecules not used in model development. Therefore, to prove the predictive power of a model, one requirement is to separate the compounds of series into the training and test sets. After a model is developed for the training set, it is commented by its precision in predicting the target feature of the test set.[107]

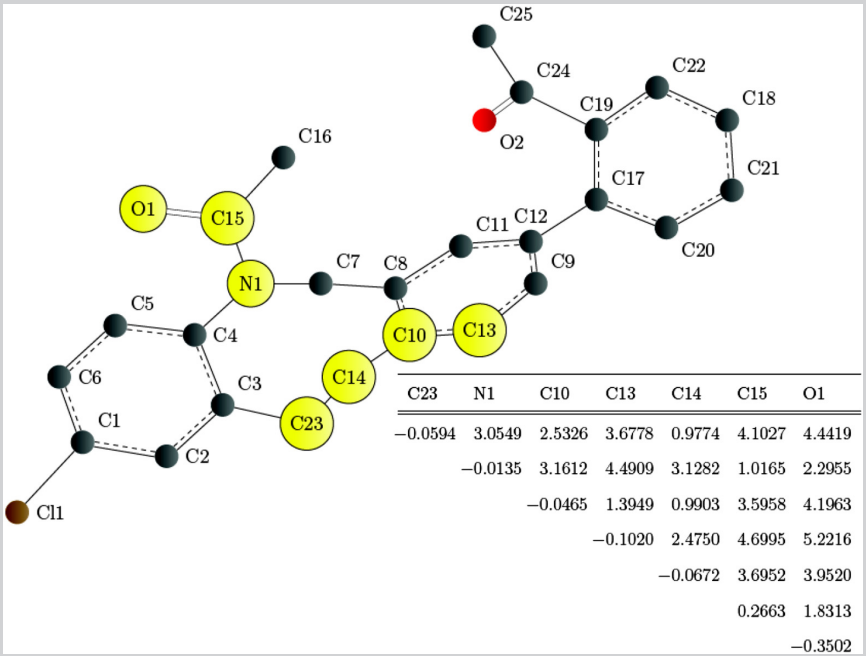
In this study, the EC-GA methods were developed by random combination for compounds of test and training. Several methods with R^2 (cross validation) and (R^2 non-cross validation) values were constituted. As indicated by Golbraikh and Tropsha,[106] high R^2 (cv) values are essential but are not adequate circumstances to accomplish whether a given model is credible or not. They proposed an external set to verify credibility and strength of a model. Therefore, the predictive capabilities of all the constituted models were appraised by calculating their R^2 employing their related test set compounds.[108] Two different equations that were discussed by Schuuremann et al.[104] used for the calculation of q^2 from an external evaluation set

$$q^2_{\mathrm{ext1}} = 1 - \frac{\sum_{n=1}^{N} \left| A_{n_{\mathrm{test}}}^{\mathrm{exp}} - A_{n_{\mathrm{test}}}^{\mathrm{calc}} \right|^2}{\sum_{n=1}^{N} \left| A_{n_{\mathrm{test}}}^{\mathrm{exp}} - \bar{A}_{tr}^{\mathrm{exp}} \right|^2} \tag{5}$$

$$q^2_{\mathrm{ext2}} = 1 - \frac{\sum_{n=1}^{N} \left| A_{n_{\mathrm{test}}}^{\mathrm{exp}} - A_{n_{\mathrm{test}}}^{\mathrm{calc}} \right|^2}{\sum_{n=1}^{N} \left| A_{n_{\mathrm{test}}}^{\mathrm{exp}} - \bar{A}_{n_{\mathrm{test}}}^{\mathrm{exp}} \right|^2} \tag{6}$$

where A_n^{exp} is the experimental activity, A_n^{calc} is the predicted activity, *N* is the number of tested molecules without employing the left-out

TABLE 2 (a) ECSA of reference compound (45) for tetrahydrodibenzazocines derivatives, (b) tolerance matrix of ECSA values for 25 compounds with high activity, (c) tolerance values for 24 compounds with low activity, (d) tolerance values for 325 conformations of 49 compounds. [Color table can be viewed at wileyonlinelibrary.com]

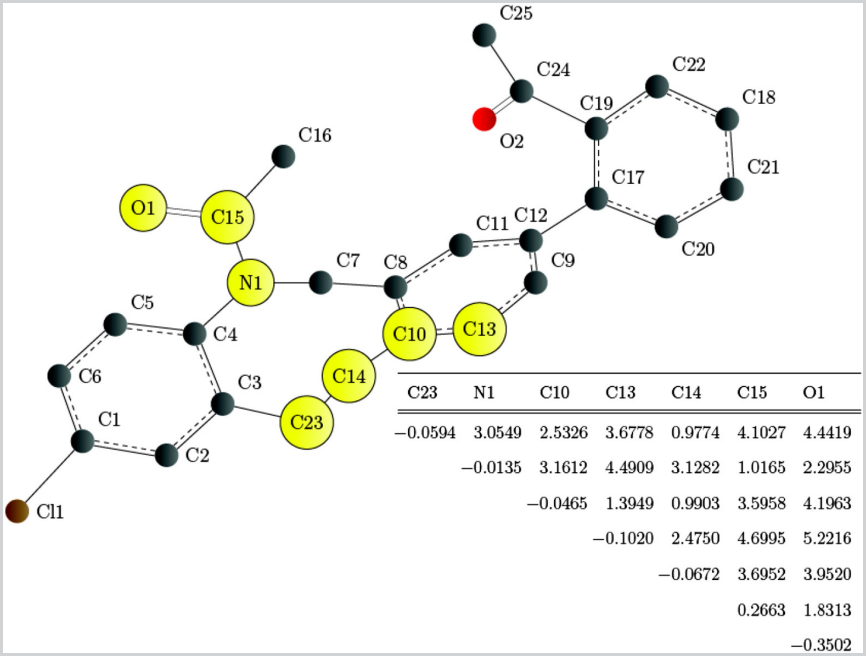


C23	N1	C10	C13	C14	C15	O1
−0.0594	3.0549	2.5326	3.6778	0.9774	4.1027	4.4419
	−0.0135	3.1612	4.4909	3.1282	1.0165	2.2955
		−0.0465	1.3949	0.9903	3.5958	4.1963
			−0.1020	2.4750	4.6995	5.2216
				−0.0672	3.6952	3.9520
					0.2663	1.8313
						−0.3502

C23	N1	C10	C13	C14	C15	O1	Pha atoms
(a) ECSA (Pha) of reference compound							
−0.060	+3.055	+2.533	+3.678	+0.977	+4.103	+4.442	C23
	−0.014	+3.161	+4.491	+3.128	+1.017	+2.296	N1
		−0.047	+1.395	+0.990	+3.596	+4.196	C10
			−0.102	+2.475	+4.610	+5.222	C13
				−0.067	+3.695	+3.952	C14
					+0.266	+1.831	C15
						−0.352	O1
(b) Tolerance values for 25 compounds with high activity							
±0.290	±0.189	±0.298	±0.201	±0.008	±0.451	±0.931	C23
	±0.292	±0.108	±0.084	±0.172	±0.021	±0.027	N1
		±0.029	±0.017	±0.008	±0.309	±0.497	C10
			±0.058	±0.012	±0.392	±0.623	C13
				±0.195	±0.223	±0.577	C14
					±0.043	±0.039	C15
						±0.017	O1
(c) Tolerance values for 24 compounds with low activity compounds							
±0.151	±0.205	±0.303	±0.233	±0.943	±0.449	±0.953	C23
	±0.019	±0.192	±0.109	±0.501	±0.052	±0.204	N1
		±0.042	±0.019	±0.018	±1.055	±0.880	C10
			±0.050	±0.032	±1.219	±1.216	C13
				±0.158	±1.381	±1.778	C14
					±0.344	±0.836	C15
						±0.232	O1

(Continues)

TABLE 2 (Continued)



C23	N1	C10	C13	C14	C15	O1
−0.0594	3.0549	2.5326	3.6778	0.9774	4.1027	4.4419
	−0.0135	3.1612	4.4909	3.1282	1.0165	2.2955
		−0.0465	1.3949	0.9903	3.5958	4.1963
			−0.1020	2.4750	4.6995	5.2216
				−0.0672	3.6952	3.9520
					0.2663	1.8313
						−0.3502

C23	N1	C10	C13	C14	C15	O1	Pha atoms
(d) Tolerance values for 325 conformations of 49 compounds							
±0.302	±0.205	±0.303	±0.233	±0.943	±0.481	±0.993	C23
	±0.292	±0.199	±0.139	±0.538	±0.052	±0.228	N1
		±0.051	±0.024	±0.018	±1.082	±1.238	C10
			±0.073	±0.032	±1.269	±1.514	C13
				±0.195	±1.381	±1.778	C14
					±0.346	±0.837	C15
						±0.232	O1

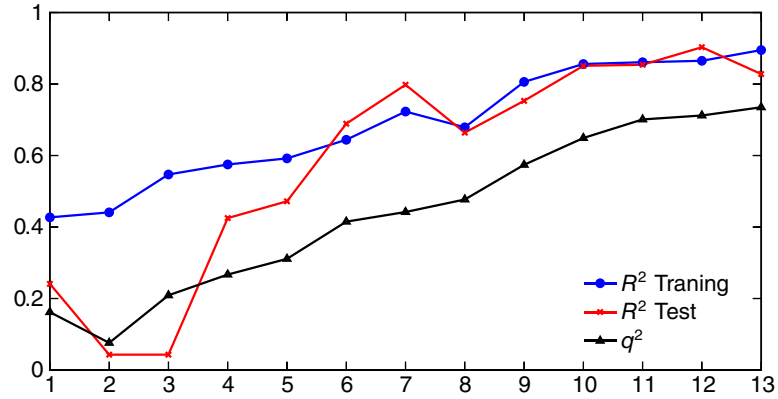
compound in the model construction, and $\bar{A}_n^{exp}$ is the average of experimental activities.

3 | RESULTS AND DISCUSSION

In our study, 49 biological data from the work of Fink et al[76] were employed. The activity values of these 49 compounds were indicated in this article. The molecules are indicated in Table 1. All quantum-chemical semiemprical calculations at PM3 level were performed with Spartan 0.6. ECMCs were constructed from the electronic structure calculation and conformational analysis data of each compound in the series by the EMRE software based on the Delphi program developed by our research group.[109] The carbon-hydrogen bond's geometric and electronic values have not been regarded because it has an equal impact in each compound (Figure 1). We selected the ECMC of the compound's lowest power conformer with the highest activity as a pattern (Compound 45) and referred it to the ECMCs of conformers with the lowest energy of other compounds within tolerances.[50]

By gradually changing the tolerance limits for nondiagonal components (bond orders or interatomic distances) and diagonal elements (charges), we ultimately obtained the tolerance limits set out in Table 2 of the ECSA, showing the best separation of active compounds from comparable small active or inactive compounds. Table 2 shows the first ECSA as a reference compound; the second ECSA shows the tolerance values for high-activity compounds; and the third submatrix shows tolerance values for low-activity compounds. Then, the peak tolerance values were calculated for all conformers of all compounds without any limitations on the tolerance values of the pharmacophore. The last ECSA shows the tolerance values of all compounds (49) for the conformers (325). After scanning the 49 active compounds within the initial ECSA tolerances not exceeding 0.25 for diagonal and 1.30 for off-diagonal elements, the ECMC submatrix common to all active molecules consists of seven atoms corresponding to C23, N1, C10, C13, C14, C15, and O1.

This flexibility within the Pha geometry and its impact on the operation is the specific property of the EC technique. We note that for greater active compounds, the tolerance matrix for less active



R^2 Traning	0.427	0.441	0.547	0.575	0.592	0.644	0.723	0.679	0.806	0.856	0.861	0.865	0.895
R^2 Test	0.241	0.043	0.043	0.425	0.472	0.489	0.798	0.664	0.753	0.851	0.854	0.903	0.828
q^2	0.162	0.076	0.209	0.267	0.311	0.415	0.442	0.477	0.574	0.649	0.701	0.712	0.735

FIGURE 5 Dependence of regression coefficients in calibration (R^2) and cross-validation correlation coefficients (q^2) on number of semiempirical chemical descriptors used in model [Color figure can be viewed at wileyonlinelibrary.com]

compounds does not match the tolerance matrix (Table 2). Tolerance values in high-activity compounds are usually smaller than low-activity compounds (Table 2). Two probabilistic estimates may be used to identify whether a property is good or not[110]:

$$P_\alpha = \frac{h_1 + 1}{h_1 + h_2 + 2} \tag{7}$$

$$\alpha_a = \frac{h_1 h_4 - h_2 h_3}{k_1 k_2 k_3 k_4} \tag{8}$$

where h_1 and h_2 are the numbers of molecules containing or not containing, respectively, the activity characteristics selected by the ECM in the active compound class and h_3 and h_4 are the associated values for small activity molecules, k_1 and k_2 are the complete amount of active and low activity molecules, respectively; and $k_3 = h_1 + h_3$ and $k_4 = h_2 + h_4$.[81] Where the parameters explaining the probability of property fulfillment are good.

$$P_\alpha = 0.9200 \text{ and } \alpha_a = 0.7831$$

In order to improve the method and estimate the anticancer activity, 680 parameters such as electronic, topological, structural, and physicochemical constants were evaluated as feasible input candidates in the model. Among these parameters, 10 parameters were chosen using the elimination selection method (GA). The information set must be approximately five times bigger than the amount of parameters selected to achieve excellent results, according to the thumb rule.[111] The physical significance of each parameter is discussed. The GA codes were written in Matlab in this study and used on multiprocessor and multicore parallel computers. In Eqn. (2), the adjustable constant, κ_j was mathematically optimized employing function lsqnonlin in the Matlab toolbox.[85] As illustrated in Figure 5, the best subset of the parameters is a mixture of 10 parameters that make the most of the activity and describe the impact of the flexibility of the ECSA parameter. As shown in Figure 5, after 10 parameters, the model reaches a stable state and any fresh parameters added after that are not necessary. We should have no more than 10 parameters in each model for a training set consisting of 40 compounds to abstain from overfitting the information.

TABLE 3 Optimum 10 molecular parameters chosen with GA and their κ_j values used in activity calculation for model

$a_{ni}^{(j)}$	Molecular parameters	κ_j values
$a^{(1)}$	Orthogonal distance from the C2 atoms to C23–C14–C10 plane	−0.202
$a^{(2)}$	Orthogonal distance from the C2 atoms to C23–N1–C10 plane+van der Waals radius of C2 atom	−0.255
$a^{(3)}$	Orthogonal distance from the C12 atoms to C14–N1–C10 plane+van der Waals radius of C12 atom	0.135
$a^{(4)}$	The angle (degree) between C15–N1–C4 atoms	−0.130
$a^{(5)}$	Fukui Lumo of the C5 atom	4.386
$a^{(6)}$	Fukui Lumo of the C13 atom	−5.912
$a^{(7)}$	Mulliken charge of C9 atom	−5.434
$a^{(8)}$	E (kcal/Mol)	0.002
$a^{(9)}$	Rel Eaq (kcal/Mol)	0.003
$a^{(10)}$	PSA (u)	0.008

The best set of parameters ($a^{[j]}$) in Table 3 was chosen after a lot of GA test calculations to describe the impact of ECSA parameter flexibility and nuclear groupings outside the pharmacophore. $a^{(1)}$ is the orthogonal distance, and $a^{(2)}$ and $a^{(3)}$ are the orthogonal distances plus the van der Waals radius, $a^{(4)}$ is the angle (in degree) between atoms, $a^{(5)}$ and $a^{(6)}$ are Fukui atomic electrophilic reactivity indexes of C5 and C13 atoms, $a^{(7)}$ is Mulliken charge of C9 atom, $a^{(8)}$ is E (kcal/mol), $a^{(9)}$ is rel Eaq (kcal/mol), and $a^{(10)}$ is PSA (u).

The precision of the suggested EC-GA method is explained using evaluation techniques: R^2 is the square of the correlation coefficient, q^2 is the cross-validated R^2, validation through an external test set and s is the standard error. The leave-one-out (LOO) cross-validated R^2 was acquired by using the LOO method of Cramer et al.[112] Previous reports indicated that models with q^2 greater than 0.5 were evaluated to have predictive ability better than estimate, and with higher q^2,

TABLE 4 E, q^2, $R^2_{training}$, and R^2_{test} values displaying the contribution of each parameter to the performance of the model

	$a^{(1)}$	$a^{(2)}$	$a^{(3)}$	$a^{(4)}$	$a^{(5)}$	$a^{(6)}$	$a^{(7)}$	$a^{(8)}$	$a^{(9)}$	$a^{(10)}$
E	0.787	0.885	0.164	0.746	0.822	0.814	0.489	0.713	0.626	0.310
$R^2_{traning}$	0.826	0.808	0.617	0.751	0.806	0.823	0.768	0.696	0.742	0.688
R^2_{test}	0.914	0.797	0.476	0.125	0.753	0.727	0.813	0.650	0.866	0.496
q^2	0.555	0.604	−1.128	0.530	0.573	0.569	0.284	0.508	0.440	−0.127

higher credibility could be obtained.[106] Therefore, the model was further refined to enhance q^2 and in turn its predictive power.

An extra statistical implement (E statistic),[113] which has significant convenient value, is shown in Table 4. Later, using the GA, the best subset of descriptors from the extended set is selected, E statistics are calculated as a measure to assess how the correct estimation of bioactivity affects each parameter. The following formulas have been used[114]:

$$E = \frac{\mathrm{PRESS}_N}{\mathrm{PRESS}_{N-1}} \tag{9}$$

in which

$$\mathrm{PRESS}_{N-1} = \sum_{n=1}^{N-1} \left| A_n^{\mathrm{exp}} - A_n^{\mathrm{calc}} \right|^2 \tag{10}$$

where A_n^{calc} represents the predicted and A_n^{exp} is the experimental activities in the LOO cross validation model, with N parameters being used (in this case N = 10).

The list of E values for each of the overlooked descriptors and the R^2 value obtained when the descriptor is removed are provided in Table 4, and it can be shown that the E values for $a^{(3)}$, $a^{(10)}$, $a^{(7)}$, $a^{(9)}$ and $a^{(8)}$ are the lowest; therefore, these five descriptors are the most effective. This is supported by the comparatively large decrease in the R^2 value from 0.856 to 0.617, respectively, 0.688, 0.768, 0.742 and 0.696. Ultimately, $a^{(2)}$ can be excluded without any significant loss of precision, resulting in a decrease in q^2 of only 0.650 to 0.604, and the other parameters are more essential because their predictive powers are significantly higher. The elevated values of $a^{(1)}$, $a^{(5)}$, and $a^{(6)}$ parameters are represented in Table 4. Thus, the subset's accuracy with 10 parameters indicated that the GA is compatible with the outcomes using E statistics.

Model commentary is a very significant item of QSAR; without it, there is only a restricted ability to regulate molecules. Nonlinear models are usually thought to be difficult to comment, and these models are occasionally referred to as blackbox models.[115,116] To discover novel potentially active compounds employing the EC-GA method, structural data about the new compound are a requirement. In the EC-GA method, after carrying out quantum-chemical calculations to acquire geometric, electronic, and thermodynamic properties of the structure of the new molecule, its bioactivity is estimated employing the chosen parameters and their κ coefficients in Eqn. (1) for the involved molecule.

4 | CONCLUSIONS

In short, this study's primary aim was to create a characteristic relationship between structural characteristics and inhibitory activity. We have improved a new EC-GA 4D-QSAR method that combines the EC and GA methods. This method was carried out on 49 chemically various molecules belonging to substituted tetrahydrodibenzazocines. The basic aim of a quantitative model is to define active structures and to predict their anticancer activity exactly. To confirm whether the hypothesis can also predict the activity of compounds that are structurally distinct from those contained in the training set, we performed a test set of nine molecules by different activity classes and of different structural knowledge. The pharmacophore model of tetrahydrodibenzazocine derivatives was constructed using EMRE software package. A few pharmacophore models were acquired and the best pharmacophore model inquired seven atoms: C23, N1, C10, C13, C14, C15, and O1. For each compound, 680 parameters were computed and gathered employing EMRE software. The choice of the most convenient structural parameters was succeeded by a GA. There are no certain rules that conduct the choice so it is difficult to select sufficient parameters for nonlinear QSAR work. This study indicated that GA is capable of choosing parameters that ensure a good result. A selection of 40 compounds has been used as the training set for QSAR modeling. The remaining nine molecules (about 20% of the data set) were used as an internal test subset to check the QSAR models. The findings of this research show that anticancer activity can be effectively modeled using physicochemical constants and structural parameters. The validation techniques used in our research explained the precision and strength of the QSAR model, which is produced not only by calculating its fitness on training sets but also by testing the model's predictive capacity. Given the elevated predictive capacity, the suggested technique offers a convenient solution to costly and time-consuming studies to detect tetrahydrodibenzazocin anticancer activity.

ACKNOWLEDGMENTS

This project was financially supported by the Scientific Technical Research Council of Turkey (TUBITAK, Grant No. 107T385) and the Research Fund of the Erciyes Üniversitesi (BAP, Project Number: FBD–09–924).

ORCID

Kader Sahin https://orcid.org/0000-0002-9056-9000
Emin Saripinar https://orcid.org/0000-0002-0116-0239

REFERENCES

[1] D. Poirier, *Curr. Med. Chem.* **2003**, *10*, 453. https://doi.org/10.2174/0929867033368222.

[2] H. I. Scher, G. Buchanan, W. Gerald, L. M. Butler, W. D. Tilley, *Endocr. Relat. Cancer* **2004**, *11*, 459.

[3] J. A. Sha, K. Dudley, W. R. Rajapaksha, P. J. O'Shaughnessy, *J. Steroid Biochem. Mol. Biol.* **1997**, *60*, 19. https://doi.org/10.1016/s0960-0760(96)00165-3.

[4] E. Koh, T. Noda, J. Kanaya, M. Namiki, *Prostate* **2002**, *53*, 154. https://doi.org/10.1002/pros.10139.

[5] T. E. Spires, B. E. Fink, E. K. Kick, D. You, C. A. Rizzo, I. Takenaka, R. M. Lawrence, Z. Ruan, M. E. Salvati, G. D. Vite, R. Weinmann, R. M. Attar, M. M. Gottardis, M. V. Lorenzi, *Prostate* **2005**, *65*, 159. https://doi.org/10.1002/pros.20279.

[6] W. M. Geissler, D. L. Davis, L. Wu, K. D. Bradshaw, S. Patel, B. B. Mendonca, K. O. Elliston, J. D. Wilson, D. W. Russell, S. Andersson, *Nat. Genet.* **1994**, *7*, 34. https://doi.org/10.1038/ng0594-34.

[7] V. Luu-The, Y. Zhang, D. Poirier, F. Labrie, *J. Steroid Biochem. Mol. Biol.* **1995**, *55*, 581. https://doi.org/10.1016/0960-0760(95)00209-x.

[8] J. Robert, C. Jarry, *J. Med. Chem.* **2003**, *46*, 4805. https://doi.org/10.1021/jm030183a.

[9] S. Hassan, S. Dhar, M. Sandstrom, D. Arsenau, M. Budnikova, I. Lokot, N. Lobanov, M. O. Karlsson, R. Larsson, E. Lindhagen, *Cancer Chemother. Pharmacol.* **2005**, *55*, 47. https://doi.org/10.1007/s00280-004-0855-5.

[10] S. Jin, M. Li, C. Zhu, V. Tran, B. Wang, *ChemBioChem* **2008**, *9*, 1431.

[11] A. Itai, N. Tomioka, Y. Kato, in *Qsar and Drug Design: New Developments and Applications*, Vol. 23 (Ed: T. Fujita), Pharmacochemistry Library, Elsevier, Netherlands, **1995**, p. 3.

[12] M. A. Collins, I. Shaw, D. C. Billington, *Drug Des. Discov* **1999**, *16*, 181.

[13] J. Drews, S. Ryser, *Nat. Biotechnol.* **1997**, *15*, 1318. https://doi.org/10.1038/nbt1297-1318.

[14] J. Drews, S. Ryser, *Drug Discov. Today* **1997**, *2*, 365.

[15] J. A. Dimasi, *Clin. Pharmacol. Ther.* **2001**, *69*, 286. https://doi.org/10.1067/mcp.2001.115132.

[16] C. A. Taft, V. B. Da Silva, C. H. Da Silva, *J. Pharm. Sci.* **2008**, *97*, 1089. https://doi.org/10.1002/jps.21293.

[17] J. P. Martins, E. G. Barbosa, K. F. Pasqualoto, M. M. Ferreira, *J. Chem. Inf. Model* **2009**, *49*, 1428. https://doi.org/10.1021/ci900014f.

[18] G. Thomas, *Fundamentals of medicinal chemistry*, Wiley-Blackwell, England, **2003**.

[19] M. Karelson, *Molecular descriptors in QSAR/QSPR including CD-ROM*, Wiley-Interscience, USA, **2000**.

[20] L. Bernazzani, C. Duce, A. Micheli, V. Mollica, A. Sperduti, A. Starita, M. R. Tiné, *J. Chem. Inform. Model.* **2006**, *46*, 2030.

[21] G. Melagraki, A. Afantitis, K. Makridima, H. Sarimveis, O. Igglessi-Markopoulou, *J. Mol. Model.* **2006**, *12*, 297.

[22] G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, C. T. Supuran, *Bioorg. Med. Chem.* **2006**, *14*, 1108.

[23] H. Gao, J. Bajorath, *Mol. Diversity* **1998**, *4*, 115.

[24] S. Oloff, R. B. Mailman, A. Tropsha, *J. Med. Chem.* **2005**, *48*, 7322.

[25] A. Meneses-Marcel, Y. Marrero-Ponce, Y. Machado-Tugores, A. Montero-Torres, D. M. Pereira, J. A. Escario, J. J. Nogal-Ruiz, C. Ochoa, V. J. Aran, A. R. Martinez-Fernandez, R. N. Garcia Sanchez, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3838. https://doi.org/10.1016/j.bmcl.2005.05.124.

[26] L. Santana, E. Uriarte, H. Gonzalez-Diaz, G. Zagotto, R. Soto-Otero, E. Mendez-Alvarez, *J. Med. Chem.* **2006**, *49*, 1149. https://doi.org/10.1021/jm0509849.

[27] G. F. Yang, X. Huang, *Curr. Pharm. Des.* **2006**, *12*, 4601. https://doi.org/10.2174/138161206779010431.

[28] J. Ferguson, Z. De Los Santos, N. Devi, E. Van Meir, S. Zingales, B. Wang, *Bioorg. Med. Chem. Lett.* **2017**, *27*, 1731.

[29] L. Shang, H. Fang, H. Zhu, X. Wang, Q. Wang, J. Mu, B. Wang, S. Kishioka, W. Xu, *Bioorg. Med. Chem.* **2009**, *17*, 2775. https://doi.org/10.1016/j.bmc.2009.02.034.

[30] S. Zheng, G. Kaur, H. Wang, M. Li, M. Macnaughtan, X. Yang, S. Reid, J. Prestegard, B. Wang, H. Ke, *J. Med. Chem.* **2008**, *51*, 7673.

[31] Q. Wang, M. Chen, H. Zhu, J. Zhang, H. Fang, B. Wang, W. Xu, *Bioorg. Med. Chem.* **2008**, *16*, 5473. https://doi.org/10.1016/j.bmc.2008.04.012.

[32] A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* **1997**, *119*, 10509.

[33] J. S. Duca, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1367.

[34] C. H. Andrade, K. F. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *Molecules* **2010**, *15*, 3281.

[35] O. M. Becker, Y. Levy, O. Ravitz, *J. Phys. Chem. B* **2000**, *104*, 2123.

[36] P. Venkatarangan, A. J. Hopfinger, *J. Chem. Inform. Comput. Sci.* **1999**, *39*, 1141.

[37] M. D. Krasowski, X. Hong, A. J. Hopfinger, N. L. Harrison, *J. Med. Chem.* **2002**, *45*, 3210.

[38] X. Hong, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 324. https://doi.org/10.1021/ci0200321.

[39] E. Rosines, I. E. Bersuker, J. Boggs, *Quant. Struct-Act. Rel.* **2001**, *20*, 327.

[40] T. Langer, *Pharmacophores and Pharmacophore Searches*, Wiley-VCH John Wiley, Weinheim, Chichester **2006**.

[41] O. F. Guner, *Curr. Top. Med. Chem.* **2002**, *2*, 1321. https://doi.org/10.2174/1568026023392940.

[42] A. Dimoglo, P. Vlad, N. Shvets, M. Coltsa, Y. Guzel, M. Saracoglu, E. Saripinar, S. Patat, *New J. Chem.* **1995**, *19*, 1217.

[43] I. B. Bersuker, S. Bahceci, J. E. Boggs, R. S. Pearlman, *SAR QSAR Environ. Res.* **1999**, *10*, 157.

[44] I. B. Bersuker, *J. Comput. Aided Mol. Des.* **2008**, *22*, 423. https://doi.org/10.1007/s10822-008-9191-x.

[45] N. Karali, A. Gursoy, F. Kandemirli, N. Shvets, F. B. Kaynak, S. Ozbey, V. Kovalishyn, A. Dimoglo, *Bioorg. Med. Chem.* **2007**, *15*, 5888. https://doi.org/10.1016/j.bmc.2007.05.063.

[46] K.-K. Bedia, O. Elçin, U. Seda, K. Fatma, S. Nathaly, R. Sevim, A. Dimoglo, *Eur. J. Med. Chem.* **2006**, *41*, 1253.

[47] W. Karcher, J. Devillers Eds., *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology (Eurocourses: Chemical and Environmental Science Series, Volume 1)*, Kluwer Academic, Italy, **1990**.

[48] Taylor & Francis Online, *Polycycl. Aromat. Compds.* **2000**, *18*, 243.

[49] J. Devillers, W. Karcher Eds., *Applied Multivariate Analysis in SAR and Environmental Studies (Eurocourses: Chemical and Environmental Science)*, Springer, Luxembourg, **1991**.

[50] I. B. Bersuker, *Curr. Pharm. Des.* **2003**, *9*, 1575. https://doi.org/10.2174/1381612033454586.

[51] C. G. Broyden, *IMA J. Appl. Math.* **1970**, *6*, 76.

[52] R. Fletcher, *Comput. J.* **1970**, *13*, 317.

[53] D. Goldfarb, *Math. Comput.* **1970**, *24*, 23.

[54] D. F. Shanno, *Math. Comput.* **1970**, *24*, 647.

[55] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Professional, Boston, MA, USA, **1989**.

[56] J. H. Holland, *Sci Am.* **1992**, *267*, 66.

[57] S. Forrest, *Science* **1993**, *261*, 872.

[58] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **1983**, *220*, 671. https://doi.org/10.1126/science.220.4598.671.

[59] J. M. Sutter, S. L. Dixon, P. C. Jurs, *J. Chem. Inform.Comput. Sci.* **1995**, *35*, 77.

[60] H. Kubinyi, *Quant. Struct.-Act. Relat.* **1994**, *13*, 285.

[61] H. Kubinyi, *Quant. Struct.-Act. Relat.* **1994**, *13*, 393.

[62] L. Eriksson, P. L. Andersson, E. Johansson, M. Tysklind, *Mol. Divers.* **2006**, *10*, 169. https://doi.org/10.1007/s11030-006-9024-6.

[63] A. Z. Dudek, T. Arodz, J. Galvez, *Comb. Chem. High Throughput Screen.* **2006**, *9*, 213.

[64] L. Terfloth, J. Gasteiger, *Drug Discov. Today* **2001**, *6*, 102.
[65] A. Verma, X. Llorà, S. Venkataraman, D. Goldberg, R, Campbell Scaling eCGA model building via data-intensive computing. Presented at 2010 IEEE World Congress on Computational Intelligence, WCCI 2010–2010 IEEE Congress on Evolutionary Computation, CEC **2010**.
[66] A. S. Reddy, S. Kumar, R. Garg, *J. Mol. Graph. Model* **2010**, *28*, 852.
[67] M. Fernandez, J. Caballero, L. Fernandez, A. Sarai, *Mol. Divers.* **2011**, *15*, 269. https://doi.org/10.1007/s11030-010-9234-9.
[68] B. Jin, Y. Q. Zhang, B. Wang, *Int. J. Data Min. Bioinform* **2007**, *1*, 270.
[69] R. Guha, P. C. Jurs, *J. Chem. Inf. Model* **2005**, *45*, 800. https://doi.org/10.1021/ci050022a.
[70] E. Saripinar, N. Gecen, K. Sahin, E. Yanmaz, *Eur. J. Med. Chem.* **2010**, *45*, 4157. https://doi.org/10.1016/j.ejmech.2010.06.007.
[71] N. Gecen, E. Saripinar, E. Yanmaz, K. Sahin, *J. Mol. Model* **2012**, *18*, 65. https://doi.org/10.1007/s00894-011-1024-5.
[72] K. Şahin, E. Sarıpınar, E. Yanmaz, N. Geçen, *SAR QSAR Environ. Res.* **2011**, *22*, 217.
[73] E. Yanmaz, E. Saripinar, K. Sahin, N. Gecen, F. Copur, *Bioorg. Med. Chem.* **2011**, *19*, 2199. https://doi.org/10.1016/j.bmc.2011.02.035.
[74] N. S. E. S. Burak Tuzun, S. C. Yavuz, *Curr. Comput. Aided Drug Des.* **2018**, *14*, 370.
[75] S. Yavuz, S. E, N. Sabanci, *Curr. Comput. Aided Drug Des.* **2018**, *14*, 79.
[76] B. E. Fink, A. V. Gavai, J. S. Tokarski, B. Goyal, R. Misra, H. Y. Xiao, S. D. Kimball, W. C. Han, D. Norris, T. E. Spires, D. You, M. M. Gottardis, M. V. Lorenzi, G. D. Vite, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1532. https://doi.org/10.1016/j.bmcl.2005.12.039.
[77] SPARTAN, *Spartan'06 for Windows and Linux*, Wavefunction, Inc., Irvine **2006**.
[78] W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc., Irvine, CA **2003**.
[79] V. Librando, A. Alparone, *J. Hazard. Mater.* **2009**, *161*, 1338. https://doi.org/10.1016/j.jhazmat.2008.04.095.
[80] I. B. Bersuker, S. Bahceci, J. E. Boggs, R. S. Pearlman, *J. Comput. Aided Mol. Des.* **1999**, *13*, 419.
[81] E. Saripinar, Y. Guzel, S. Patat, I. Yildirim, Y. Akcamur, A. S. Dimoglo, *Arzneimittelforschung* **1996**, *46*, 824.
[82] A. V. Marenich, P. H. Yong, I. B. Bersuker, J. E. Boggs, *J. Chem. Inf. Model* **2008**, *48*, 556. https://doi.org/10.1021/ci700401p.
[83] A. Altun, K. Golcuk, M. Kumru, A. F. Jalbout, *Bioorg. Med. Chem.* **2003**, *11*, 3861. https://doi.org/10.1016/s0968-0896(03)00437-1.
[84] T. Pavlov, M. Todorov, G. Stoyanova, P. Schmieder, H. Aladjov, R. Serafimova, O. Mekenyan, *J. Chem. Inf. Model* **2007**, *47*, 851. https://doi.org/10.1021/ci700014h.
[85] MATLAB and Statistics Toolbox Release, *The MathWorks, Inc., Natick*, United States, Massachusetts **2012**.
[86] R. Todeschini, V. Consonni, *Handbook of molecular descriptors*, Vol. 11, Wiley-VCH, Weinheim, Germany, **2000**.
[87] H. Goldstein, C. P. Poole, J. L. Safko, *Classical Mechanics*, 3rd ed., Pearson, England, **2001**.
[88] B. Zhang, D. Kilburg, P. Eastman, V. S. Pande, E. Gallicchio, *J. Comput. Chem.* **2017**, *38*, 740.
[89] T. R. Stouch, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 4.
[90] R. H. Rohrbaugh, P. C. Jurs, *Anal. Chem.* **1987**, *59*, 1048. https://doi.org/10.1021/ac00134a025.
[91] R. Parthasarathi, M. Elango, P. Jaganathan, V. Subramanian, D. Roy, U. Sarkar, P. Chattaraj, *IJC-A* **2006**, *45*, 111.
[92] V. J. Gillet, W. Khatib, P. Willett, P. J. Fleming, D. V. Green, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 375.
[93] D. K. Agrafiotis, D. N. Rassokhin, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 798.
[94] B. E. Mattioni, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 232.
[95] R. Guha, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2179. https://doi.org/10.1021/ci049849f.
[96] R. Guha, P. C. Jurs, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1440. https://doi.org/10.1021/ci0499469.
[97] V. Venkatraman, A. R. Dalby, Z. R. Yang, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1686. https://doi.org/10.1021/ci049933v.
[98] T. Wright, V. J. Gillet, D. V. Green, S. D. Pickett, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 381. https://doi.org/10.1021/ci0255836.
[99] H. C. Goicoechea, A. C. Olivieri, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1146.
[100] C. Hervas, M. Silva, J. M. Serrano, E. Orejuela, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1576. https://doi.org/10.1021/ci049948t.
[101] R. Leardi, *J. Chemometr.* **2001**, *15*, 559.
[102] B. K. Lavine, A. J. Moores, *Anal. Lett.* **1999**, *32*, 433.
[103] V. Consonni, D. Ballabio, R. Todeschini, *J. Chem. Inf. Model.* **2009**, *49*, 1669. https://doi.org/10.1021/ci900115y.
[104] G. Schuurmann, R. U. Ebert, J. Chen, B. Wang, R. Kuhne, *J. Chem. Inf. Model.* **2008**, *48*, 2140. https://doi.org/10.1021/ci800253u.
[105] SYBYL, *Theory Manual*, *version 6.7*, Tripos Associates Inc., St. Louis, MO **2000**.
[106] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269.
[107] P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1794. https://doi.org/10.1021/ci049923u.
[108] S. Van Damme, P. Bultinck, *J. Comput. Chem.* **2007**, *28*, 1924. https://doi.org/10.1002/jcc.20664.
[109] http://www.borland.com/delphi **2019**
[110] A. Dimoglo, N. Shvets, I. Tetko, D. Livingstone, *Quant. Struct.-Act. Relat.* **2001**, *20*, 31.
[111] J. G. Topliss, R. P. Edwards, *J. Med. Chem.* **1979**, *22*, 1238. https://doi.org/10.1021/jm00196a017.
[112] R. D. Cramer III., J. D. Bunce, D. E. Patterson, I. E. Frank, *Quant. Struct.-Act. Rel.* **1988**, *7*, 18.
[113] S. Wold, *Technometrics* **1978**, *20*, 397.
[114] A. H. Makkouk, I. B. Bersuker, J. E. Boggs, *Int. J. Pharm. Med.* **2004**, *18*, 81.
[115] T. Fox, J. M. Kriegl, *Curr. Top. Med. Chem.* **2006**, *6*, 1579. https://doi.org/10.2174/156802606778108915.
[116] C. Helma, J. Kazius, *Curr. Comput.-Aided Drug Des.* **2006**, *2*, 123.

How to cite this article: Sahin K, Saripinar E. A novel hybrid method named electron conformational genetic algorithm as a 4D QSAR investigation to calculate the biological activity of the tetrahydrodibenzazosines. *J Comput Chem.* 2020;41: 1091–1104. https://doi.org/10.1002/jcc.26154