Research Article

ETM-ANN Approach Application for Thiobenzamide and Quinolizidine Derivatives

M. Saracoglu,[1] F. Kandemirli,[2, 3] V. Kovalishyn,[4, 5] T. Arslan,[6] and E. E. Ebenso[7]

[1] *Faculty of Education, Erciyes University, 38039 Kayseri, Turkey*

[2] *Department of Chemistry, Kocaeli University, 41000 Izmit, Turkey*

[3] *Niğde University, Department of Chemistry, 41000 Niğde, Turkey*

[4] *Biomedical Department, Institute of Bioorganic Chemistry and Petrochemistry, Kyiv 02660, Ukraine*

[5] *Departamento de Quimica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa,*
  2829-516 Caparica, Portugal

[6] *Department of Chemistry, Eskişehir Osmangazi University, 26480 Eskişehir, Turkey*

[7] *Department of Chemistry, North West University (Mafikeng Campus), Private Bag X2046,*
  Mmabatho 2735, South Africa

Correspondence should be addressed to M. Saracoglu, muratsaracoglu@gmail.com

The structure anti-influenza activity relationships of thiobenzamide and quinolizidine derivatives, being influenza fusion inhibitors, have been investigated using the electronic-topological method (ETM) and artificial neural network (ANN) method. Molecular fragments specific for active compounds and breaks of activity were calculated for influenza fusion inhibitors by applying the ETM. QSAR descriptors such as molecular weight, E_{HOMO}, E_{LUMO}, ΔE, chemical potential, softness, electrophilicity index, dipole moment, and so forth were calculated, and it was found to give good statistical qualities (classified correctly 92%, or 48 compounds from 52 in training set, and 69% or 9 compounds from 13 in the external test set). By using multiple linear regression, several QSAR models were performed with the help of calculated descriptors and the compounds activity data. Among the obtained QSAR models, statistically the most significant one is the one of skeleton 1 with $R^2 = 0.999$.

1. Introduction

Influenza (also known as flu) virus is an etiologic cause of widespread epidemics of acute respiratory tract infections that occur annually during the winter season [1, 2]. Less frequent pandemic episodes of influenza are associated with greater mortality and morbidity, usually a consequence of genetic recombination of viral strains. Indeed, influenza is responsible for over 20000 deaths annually in the US, despite the annual seasonal vaccination campaigns that have become familiar [3, 4]. It is particularly dangerous to the very young, elderly, and to those who have suppressed immune systems. Vaccination has been the main preventive measure and has provided limited control, but vaccines must be reformulated each year in response to antigenic drift and may be ineffective against new variants of influenza viruses [5].

There are three types of influenza viruses: A, B, and C. The first two types are responsible for annual epidemics and pandemic outbreaks; while influenza C is endermic vaccination provides limited protection based on the ability to predict the exact strains of influenza that will predominate a year in advance of the influenza season [4]. Millions of people each year become infected with the influenza A virus. Influenza A is characterized by the abrupt onset of constitutional and respiratory signs and symptoms (e.g., fever, myalgia, headache, severe malaise, nonproductive cough, sore throat, and rhinitis) [3]. Currently, control of influenza infection relies mainly on a preventive strategy dependent on the annual identification, production, and distribution of a multivalent vaccine designed to predict the predominant epidemic viral strains. Until recently, anti-influenza drugs were restricted to the M2 channel blockers
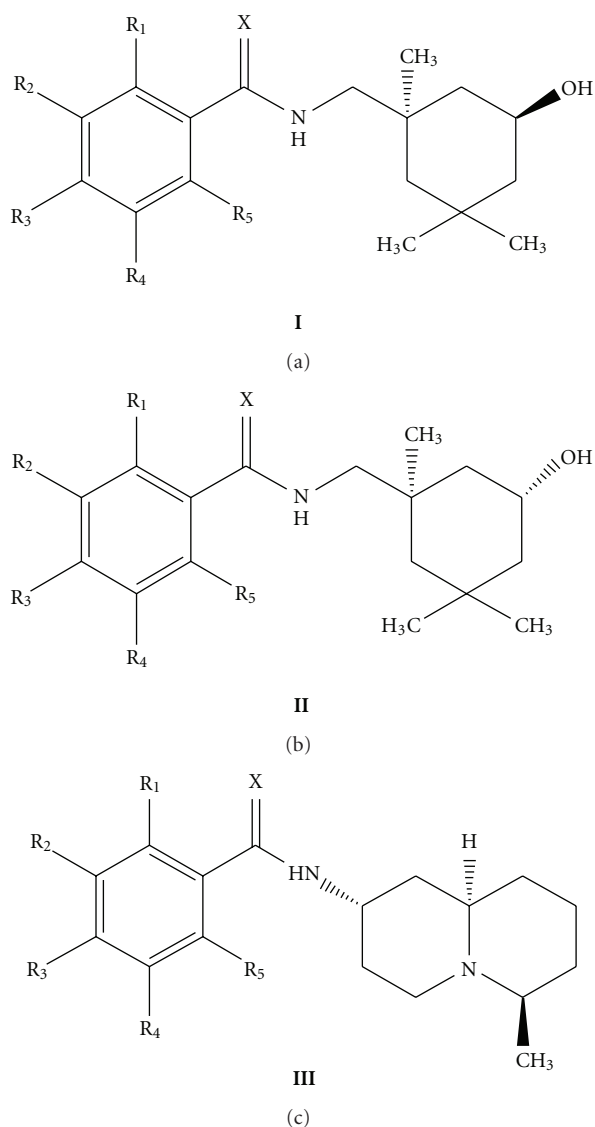
I

(a)

II

(b)

III

(c)

FIGURE 1: Common molecular skeletons.

FIGURE 2: Common scheme of the ETM.

amantadine and rimantadine, both of which are approved for the prophylactic and therapeutic treatment of the influenza viruses within the A family. However, the licensing of the neuraminidase inhibitors, zanamavir, and oseltamivir has provided the clinician with antiviral agents that effectively prevent and treat both type A and B influenza infections [1]. A new viral target has recently been exploited to develop clinically useful anti-influenza drugs [5].

In this study, we investigate the structure anti-influenza relationships by using the electronic-topological method (ETM) [6–20] for a class of thiobenzamide influenza fusion inhibitors derived from 1,3,3-trimethyl-5-hydroxy-cyclohexylmethylamine and quinolizidine derivatives [1, 2]. The series under investigation is presented in Table 1, and their common skeletons are shown in Figure 1.

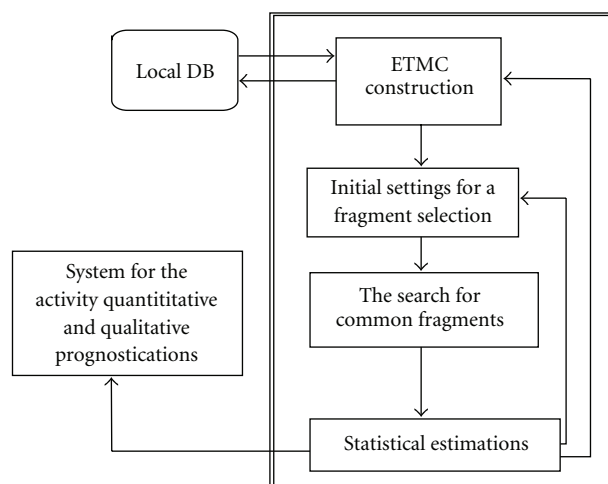All conformational and quantum-chemical data were obtained by means of the MM2 (molecular mechanic) method of the molecular mechanics and a semiempirical quantum-chemistry method known as AM1. Activity features' selection has been carried out by means of the ETM-software. To have more stable activity features, every active molecule was used as a template for comparison with the rest of molecules. As a result of this comparison, activity features (pharmacophores) P1, P2, and P3 were revealed. To decide which of pharmacophores is better, each inactive molecule was used as a template for comparison with the rest of molecules. So, inactivity features (antipharmacophores) AP1, AP2, and AP3 were revealed also to complete the system for the thiobenzamide and quinolizidine derivatives inhibitors activity prediction.

2. Materials and Methods

In this study, the structure anti-influenza relationships of thiobenzamide and quinolizidine derivatives being influenza fusion inhibitors have been studied by means of the ETM. An ETMC is an $n \times n$ matrix, where n is the number of atoms in the corresponding molecule. Diagonal elements are values of an atomic property such as atomic charge. An off-diagonal element can be of one of two kinds. It is bond property, if the bond exists, and distance for the corresponding pair of atoms, otherwise [19]. Detailed descriptions of the ETM can be found elsewhere [6–20].

The matrices are easily understandable and extremely convenient for computer handling. In this way, the structures of the molecules under study get a unified description that is not bound to the atoms' identity. This circumstance is of primary importance when searching for pharmacophores in quite diverse structures of biologically active molecules.

The main steps of the ETM procedure are shown in Figure 2 [7, 9, 10, 18]:

The descriptors, such as E_{HOMO}, E_{LUMO}, ΔE, chemical potential, softness, electrophilicity index, dipole moment, and Mulliken charges of some atoms on studied molecules were calculated by the DFT method using B3LYP exchange correlation functionals [21, 22] with the 6-31G(d,p) basis set by using Gaussian Program [23]. The Fukui functions

TABLE 1: A list of chemical molecules under investigation.

No	X	R_1	R_2	R_3	R_4	R_5	Activity, EC_{50} (μg/mL)
				Skeleton I			
1	O	OH	Cl	H	H	H	0.06
2	S	H	Cl	H	H	H	NA
3	S	H	H	H	H	H	NA
4	S	Me	H	H	H	H	NA
5	S	H	Me	H	H	H	NA
6	S	H	F	H	H	H	NA
7	S	H	CF_3	H	H	H	NA
8	S	H	H	Me	H	H	NA
9	S	Me	H	Me	H	H	NA
10	S	Me	H	H	Me	H	NA
11	S	H	Me	H	Me	H	NA
12	S	OMe	H	H	H	H	NA
13	O	OH	H	H	H	H	0.05
14	S	OH	H	H	H	H	4.5
15	O	OH	H	H	Me	H	0.03
16	S	OH	H	H	Me	H	0.4
17	S	OH	H	H	Cl	H	0.9
				Skeleton II			
18	O	OH	H	H	Cl	H	0.2
19	S	H	Cl	H	H	H	0.02
20	S	H	H	H	H	H	0.05
21	S	Me	H	H	H	H	0.065 ± 0.035
22	S	H	Me	H	H	H	0.015 ± 0.007
23	S	H	F	H	H	H	0.038
24	S	H	CF_3	H	H	H	0.09
25	S	H	H	Me	H	H	NA
26	S	Me	H	Me	H	H	0.15
27	S	Me	H	H	Me	H	0.023 ± 0.011
28	S	H	Me	H	Me	H	0.03 ± 0.07
29	S	OMe	H	H	H	H	1.0
30	O	OH	H	H	H	H	25
31	S	OH	H	H	H	H	2
32	O	OH	H	H	Me	H	0.15 ± 0.071
33	S	OH	H	H	Me	H	0.5
34	S	OH	H	H	Cl	H	1.8
35	S	H	CN	H	H	H	2.5
36	S	H	NO_2	H	H	H	NA
37	S	H	CO_2Me	H	H	H	NA
				Skeleton III			
38	O	OH	H	NH_2	Cl	H	1–3
39	O	OMe	H	NH_2	Cl	H	NA
40	O	H	H	NH_2	Cl	H	NA
41	O	OH	H	H	H	H	1.6–3.1
42	O	OH	H	NH_2	H	H	NA
43	O	OH	H	H	F	H	0.4–0.8
44	O	OH	H	H	Cl	H	0.8
45	O	OH	H	H	Br	H	0.8
46	O	OH	H	H	Me	H	0.2-0.3
47	O	OH	H	H	OMe	H	0.6

TABLE 1: Continued.

No	X	R_1	R_2	R_3	R_4	R_5	Activity, EC_{50} ($\mu g/mL$)
				Skeleton III			
48	O	OH	H	H	OH	H	NA
49	O	OH	H	H	CF_3	H	NA
50	O	OH	H	H	$CH=CH_2$	H	2
51	O	OH	H	H	N_3	H	0.5
52	O	OH	H	H	tert-Bu	H	NA
53	O	OH	H	H	Ph	H	NA
54	O	OH	H	H	PhCO	H	NA
55	O	OH	H	CH_3	H	H	4
56	O	OH	H	N_3	Cl	H	NA
57	O	OH	H	–CH=CH–CH=CH–		H	NA
58	O	OH	H	H	–CH=CH–CH=CH–		NA
59	O	OH	H	$–(Me)_2–CCH_2CH_2C–(Me)_2$		H	NA
60	O	OH	CH_3	H	H	H	8
61	O	OH	Cl	H	Cl	H	10–20
62	O	OH	OH	H	H	H	NA
63	O	OH	H	H	H	F	10
64	O	NH_2	H	H	H	H	5–10
65	O	CH_3CONH	H	H	H	H	NA

NA, not active.

3. Results and Discussion

In this study, the optimized geometry data and electronic characteristics were used to form ETMCs for all molecules in a series of thiobenzamide and quinolizidine derivatives including 65 molecules [1, 2]. The effective charge on atoms are shown on diagonal elements, while bond characteristics and optimized distances are represented on non diagonal elements.

According to the activity level, molecules under study (65 in all) were classified and divided into 3 classes, namely,

(1) active molecules (30 mol. with $EC_{50} \leq 2.5$),

(2) low active molecules (7 mol. EC_{50} between 4 and 25),

(3) inactive molecules (28 mol. with EC_{50} being nonactive).

The parameters responsible for the activity form a matrix called electron-topological submatrix of activity (ETSA), calculated from an ETMC that represents one of the most active molecules ("a template" for comparison).

3.1. Pharmacophores Identification. For each template molecule, its ETMC was compared with the ETMCs of the rest of molecules in both series. The comparison resulted in a few common structural fragments for the two cases. The fragments were found as submatrices of the corresponding template ETMCs (i.e. electron-topological submatrices of contiguity (ETSCs)).

Molecule **19** taken as template molecule from active ones was accepted to be the active one and from the template molecule **19,** an activity feature **2** (or pharmacophore, **P2**) was found and presented in Figure 3, alone with the corresponding ETSCs, which describe electronic-topological characteristics of the fragments (see Figure 3).

In the matrices, the effective charges on atoms (local atomic characteristics, Q_i) were chosen for diagonal elements, while in place of nondiagonal elements representing interatomic characteristics there are either Wiberg's indices [21], (W_{ij}, for bonds) or optimized distances (R_{ij}, in Å, for chemically non-bonded pairs of atoms). The distance between C_{19} and O_{22} atom is 8.80 Å.

Submatrices given in Figure 3 are found after setting some allowable limits for the matrix elements comparison. For both series, the limits are $\delta_1 = 0.04$ for diagonal elements of the ETMCs and $\delta_2 = 0.14$ for their off-diagonal elements comparison. The pharmacophores found from the ETM-calculations are realized in *all* active molecules (30). Statistical estimates for the pharmacophores are given in Table 2.

An activity feature **P2** is shown along with its ETSA of the order 6x6 (see Figure 3). Only its upper triangle is given because of the symmetry of bounds. The pharmacophore was found in 21 of 30 active molecules, n_A, and it was found in 3 of 28 inactive molecules, n_{IA}. Thus, the probability P_A of its realization in this class is about 0.84. As seen from the
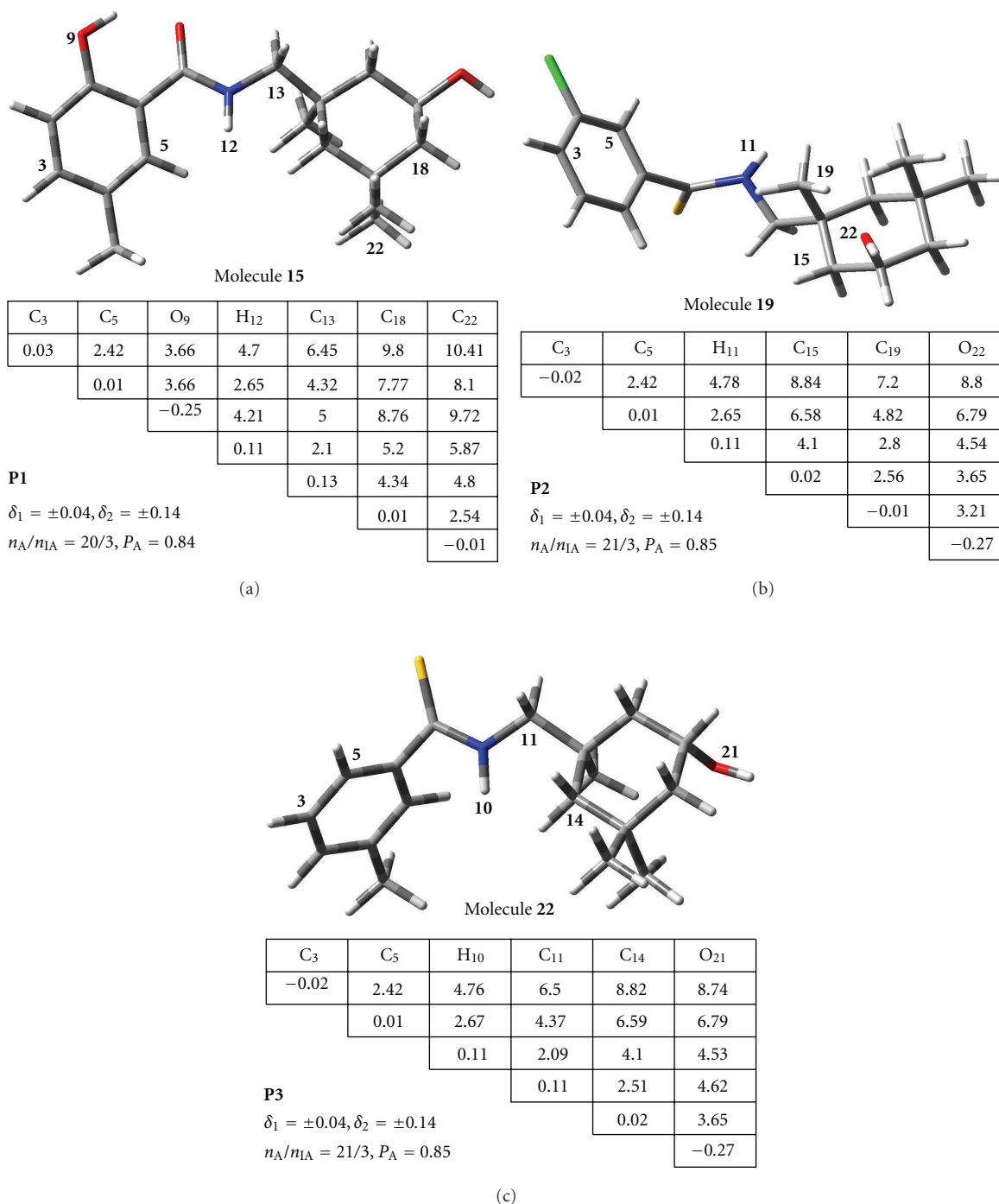
The quantum chemical parameters were calculated by AOMix program [24, 25] by using the data obtained and calculated at the level of B3LYP/6-31G(d,p) method (See Tables 1–3 in supplementary Material available online at doi:10.1155/2010/693031 which shows some quantum chemical descriptors, mulliken charges and Fukui functions of the studied molecules calculated using the B3LYP/6-31G (d,p) basis set) .

Molecule **15**

C$_3$	C$_5$	O$_9$	H$_{12}$	C$_{13}$	C$_{18}$	C$_{22}$
0.03	2.42	3.66	4.7	6.45	9.8	10.41
	0.01	3.66	2.65	4.32	7.77	8.1
		−0.25	4.21	5	8.76	9.72
			0.11	2.1	5.2	5.87
			0.13	4.34	4.8	
			0.01	2.54		
			−0.01			

P1

$\delta_1 = \pm0.04, \delta_2 = \pm0.14$

$n_A/n_{IA} = 20/3, P_A = 0.84$

(a)

Molecule **19**

C$_3$	C$_5$	H$_{11}$	C$_{15}$	C$_{19}$	O$_{22}$
−0.02	2.42	4.78	8.84	7.2	8.8
	0.01	2.65	6.58	4.82	6.79
		0.11	4.1	2.8	4.54
		0.02	2.56	3.65	
		−0.01	3.21		
		−0.27			

P2

$\delta_1 = \pm0.04, \delta_2 = \pm0.14$

$n_A/n_{IA} = 21/3, P_A = 0.85$

(b)

Molecule **22**

C$_3$	C$_5$	H$_{10}$	C$_{11}$	C$_{14}$	O$_{21}$
−0.02	2.42	4.76	6.5	8.82	8.74
	0.01	2.67	4.37	6.59	6.79
		0.11	2.09	4.1	4.53
		0.11	2.51	4.62	
		0.02	3.65		
		−0.27			

P3

$\delta_1 = \pm0.04, \delta_2 = \pm0.14$

$n_A/n_{IA} = 21/3, P_A = 0.85$

(c)

FIGURE 3: ETSC of the pharmacophores: **P1**, **P2**, and **P3** found relative to active molecule **15**, **19**, and **22**.

pharmacophore structure, the active feature **2**, **P2**, consists of the 6 atoms (C$_3$, C$_5$, H$_{11}$, C$_{15}$, C$_{19}$, and O$_{22}$).

Molecules **15** and **22** were chosen as templates and compared with the rest of the molecules in the series. Pharmacophore **1** (**P1**) was found in 20 of 30 active molecules having the probability P_A = 0.84 and pharmacophore **3** (**P3**) were found in 21 of 30 active molecules having the probability P_A = 0.85 as seen in Figure 3 and statistical estimates for the pharmacophores are given in Table 2.

Superimposition of three template compounds that correspond to the calculated pharmacophores **P1–P3** is shown in Figure 4. In 3D space, three separate regions can be indicated for each active molecule, where atoms of three pharmacophores can be found.

These regions are shown in Figure 4 by dotted lines. One of them is formed by the atoms belonging to the phenyl ring, while the other two are formed by atoms that represent the heterocyclic rings. It is quite possible that the atoms from

TABLE 2: Statistical estimates for the pharmacophores.

Template Molecules	Phamacoph	n_A	n_{IA}	P_A	P_{IA}
15	P1	20	3	0.84	0.16
19	P2	21	3	0.85	0.15
22	P3	21	3	0.85	0.15
39	AP1	2	15	0.16	0.84
48	AP2	1	12	0.19	0.81
65	AP3	1	15	0.11	0.89
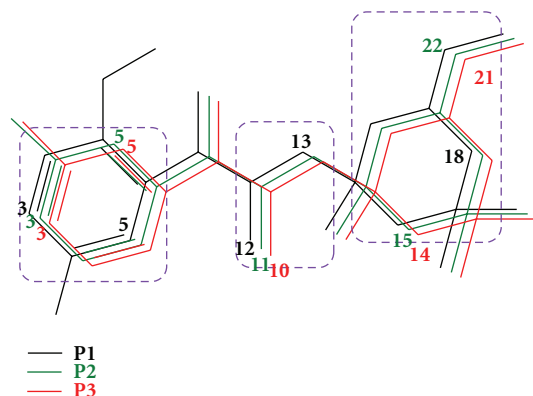
FIGURE 4: Template compounds **15**, **19** and **22** alignment by superposing the P1, P2, and P3 features.

these regions play an important role in the ligand-receptor interaction.

3.2. Antipharmacophores Calculation. Antipharmacophores (alone with pharmacophores) are also of interest in this study, as those parts of molecules that are responsible for the considerable decrease or complete loss of the activity To find antipharmacophores, inactive molecules (**39, 48,** and **65**) were selected as template molecules (their structures are given in Figure 5). Again, both protonated and unprotonated forms of molecules (**39, 48** and **65**) were studied. As an example, **AP1, AP2,** and **AP3** antipharmacophores are shown in the figure by their numbers, while corresponding submatrices are given nearby.

The antipharmacophores consists of **AP1** 5 atoms, **AP2** and **AP3** 4 atoms (see Figure 5). **AP1** enters the structures of 15 inactive molecules and are found 2 active molecules (AP1, AP2, and AP3, see Table 2). The probability level of them are realized between 81% and 89%, because **AP1, AP2,** and **AP3** are found in, respectively 15, 12, 15 from 28 inactive molecules in total. The distribution of the atoms are as shown in Figure 5.

The molecular orbitals (HOMO, LUMO) for template molecules taken as P1, P2, P3, AP1, AP2, and AP3 are shown in Figure 6, but we could not find any relationship between pharmacophore group and the molecular orbitals (HOMO and LUMO).

When comparing the structures of the pharmacophores and antipharmacophores, one can pay attention to the differences in their spatial and electron characteristics. Thus, pharmacophores and antipharmacophores can play their role in the activity prediction only if both types of fragments participate in the process of prognosis. Thus, the set of activity/inactivity fragments found as the result of this study forms a basis for a system of the analgesic activity prediction.

As seen from Figure 7, the pharmacophores **P1–P3** and antipharmacophores **AP1–AP3,** found as a result of the ETM application, were used as a basis for a system formation that is capable of the prediction of the activity of the thiobenzamide and quinolizidine derivatives.

4. Combined ETM-ETM-ANNs Approach (ETM-ANNs)

The algorithm for the data resulting from the ETM calculations analysis (ETM-data) extends the ideas of volume-learning algorithm [26]. This method is also implemented as an iterative application of the Kohonen SOMs [27] and associative neural networks (ASNNs) [28]. The algorithm evaluates the "weights" of fragments represented by the ETMCs that have been obtained from the ETM calculations. To do this, their projections on the Kohonen's maps corresponding to each ETMC are calculated and a weight of each fragment's presence in a molecule is determined [19, 29]. The calculated weights are used for the ANN training. The original implementation of two last algorithms was used.

4.1. Selforganizing Map of Kohonen (SOM). A SOM is neural network trained using unsupervised learning to create a nonlinear projection of high-dimensional input samples to a lower two-dimensional output space that is represented by a two-dimensional array of neurons, called a map [27]. The map seeks to preserve the topological properties of the input space. Like most artificial neural networks, SOMs operate in two modes: training and mapping. Training builds the map using input examples. It is a competitive process, also called vector quantization. During the SOM training, input vectors with similar properties in the high dimensional space are mapped to the same (or to the nearby) neurons on the two-dimensional space. Thus, by considering all input vectors projected to the same output neuron it is possible to determine clusters of vectors having similar properties in the high-dimensional space. Mapping automatically classifies a new input vector.

4.2. Associative Neural Network (ASNN). The ASNN represents a combination of an ensemble of feed-forward neural networks and the k-nearest neighbor method (kNN). Usually, the neural network ensemble consists of 100–200 networks when one hidden layer is used. This method uses the correlation between ensemble responses (each molecule is represented in space of neural network models as a vector of model predictions) as a measure of distance amid the analyzed cases for the nearest neighbor technique. Thus, ASNN performs kNN in space of ensemble residuals. This provides an improved prediction by the bias correction of the neural network ensemble [28].
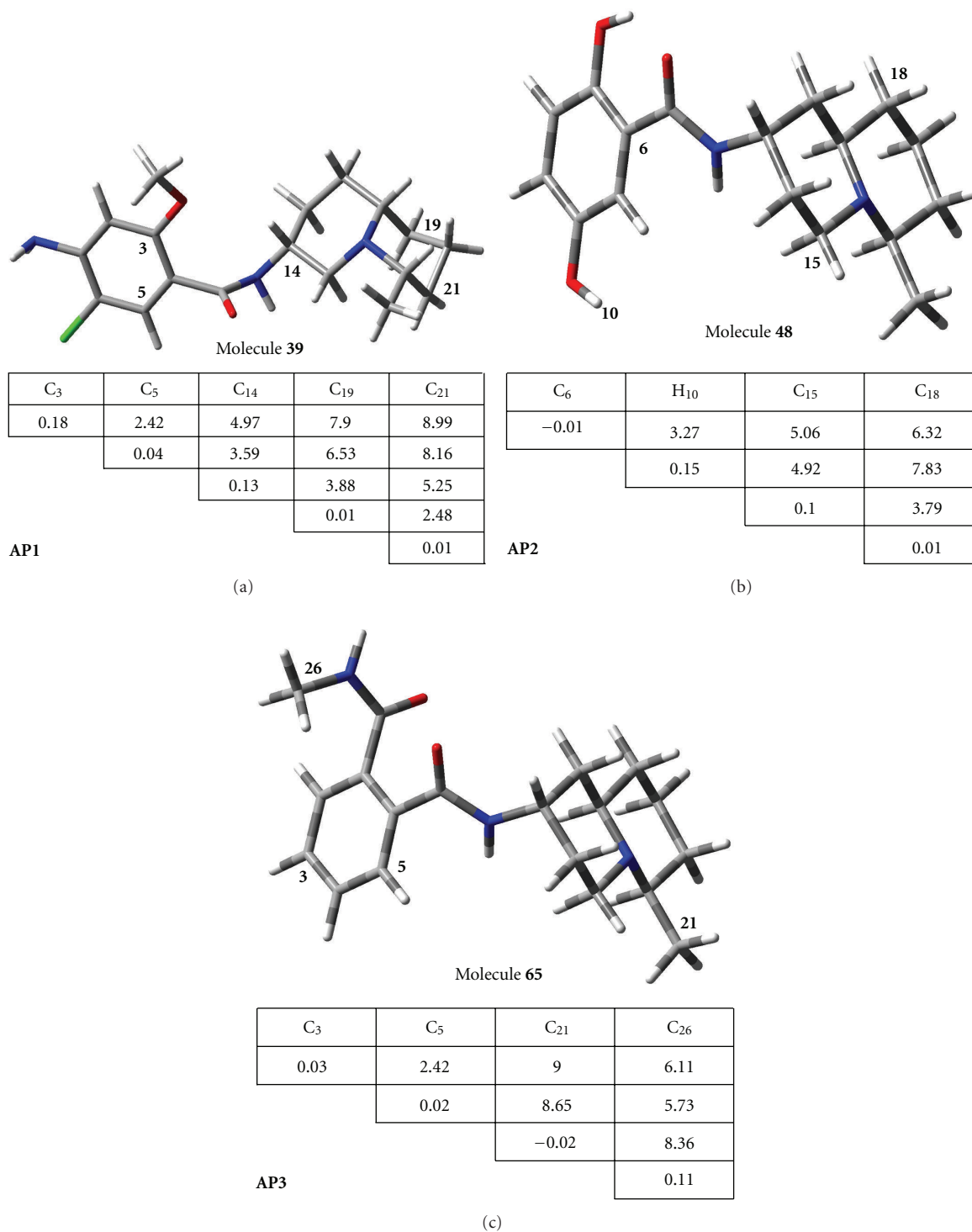
Molecule **39**

C_3	C_5	C_{14}	C_{19}	C_{21}
0.18	2.42	4.97	7.9	8.99
	0.04	3.59	6.53	8.16
		0.13	3.88	5.25
			0.01	2.48
AP1				0.01

(a)

Molecule **48**

C_6	H_{10}	C_{15}	C_{18}
-0.01	3.27	5.06	6.32
	0.15	4.92	7.83
		0.1	3.79
AP2			0.01

(b)

Molecule **65**

C_3	C_5	C_{21}	C_{26}
0.03	2.42	9	6.11
	0.02	8.65	5.73
		-0.02	8.36
AP3			0.11

(c)

FIGURE 5: ETSC of the antipharmacophores: **AP1**, **AP2,** and **AP3** found relative to active molecule **39**, **48,** and **65**.

The avoidance of overfitting/overtraining has been shown to be an important factor for improvement of generalisation ability and correct selection of variables in ASNN. the current study used the early stopping over ensemble (ESE) technique to accomplish this. A detailed description of ESE can be found elsewhere [30]. In brief, each analysed artificial neural network ensemble (ANNE) was composed of $M = 100$ networks. The values calculated for analysed cases were averaged over all M neural networks, and their means were used for computing statistical coefficients with targets. We used a subdivision of the initial training set into two equal learning/validation subsets. The first set was used to train the neural network while the second one was used to control the training process. Two stopping points
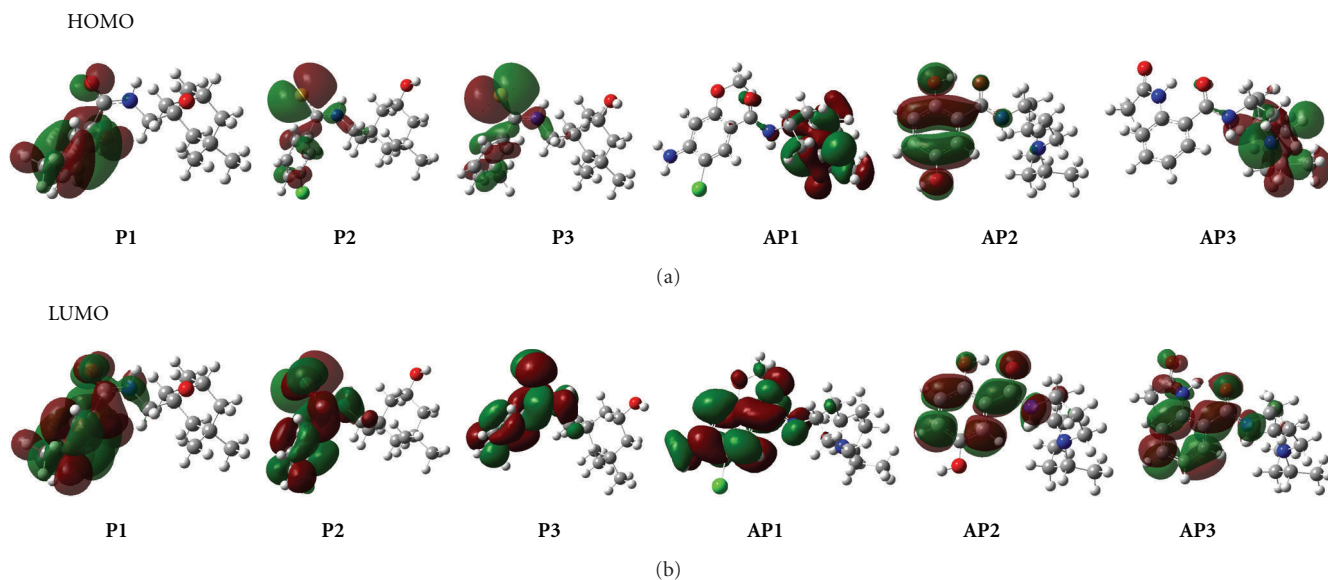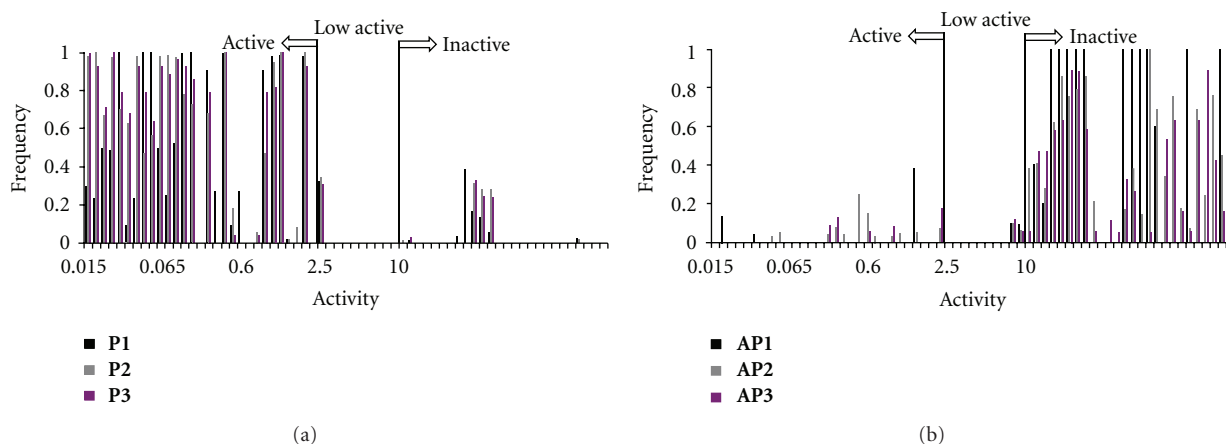
HOMO

P1   P2   P3   AP1   AP2   AP3

(a)

LUMO

P1   P2   P3   AP1   AP2   AP3

(b)

FIGURE 6: HOMO, LUMO for pharmacophore groups.

(a)

(b)

FIGURE 7: Frequency of the fragments' occurrencies in the compounds studied: (a) for Pharmacophores **P1**, **P2,** and **P3**; (b) for antipharmacophores **AP1**, **AP2,** and **AP3**.

were used to test network performance measured by root mean square error (RMSE). The first point (early stopping) determined a best fit of a network to the validation set while the second point corresponded to the error minimum for the learning set, and as a rule coincides with the end of the network training. The quality of each final model was assessed by the leave-one-out cross-validation method (LOO). By the method, each molecule was removed from the training set, and the remaining set was used to separate molecules into classes of activity, thereby predicting the activity of this molecule and evaluating the quality of the decision rule. The detailed description of the used methods can be found in [30, 31].

Sensitivity analysis methods estimate the rate of change in the output of a model caused by the changes of the model inputs. It is mainly used to determine which input descriptor is more important or sensible to achieve accurate output values. It is also used to understand the behavior of the modeled system that to evaluate of the model applicability and to determine the stability of a model. To evaluate the importance of the original ETMC fragments, we have used sensitivity analysis methods named pruning algorithms [32, 33]. The pruning algorithms introduce some measures of importance of weights matrix of ASNN by so called "sensitivities" (S). Since details of the sensitivity methods can be found in literature, we only give here the short description of the used method. The sensitivity of input neuron i was calculated as,

$$S_i = \sum_{j=1}^{n_j} \left(\frac{w_{ji}}{\max_a |w_{ja}|} \right)^2 \cdot S_j, \tag{1}$$

where, S_j was a sensitivity of the jth neuron in the upper layer and $\max_a$ was taken over all weights ending at neuron j. This was a recurrent formula calculating the sensitivity of a unit on layer s via the neuron sensitivities on layer $s+1$. The sensitivities of output layer neurons were set to 1 [32].

The sensitivities of all neurons were calculated and the ETM descriptors associated with input neurons having smallest sensitivities was deleted.

4.3. Results of the ETM-ANN Approach Application.

A dataset containing 65 molecules was used. Fifty-two of the compounds were used for the model development, and 13 randomly selected compounds (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, and 65) were used for the model validation. For the data, 210 fragments were selected. The importance of the detected fragments for the observed activity was evaluated by using pruning methods. The most part of the ETMC fragments (as descriptors) were detected as nonsignificant and removed by the pruning algorithms. As the result, only 13 ETMC-fragments were chosen from 210 fragments in total as the most important ones. By this, ASNNs classified correctly 92%, or 48 compounds from 52 in training set and 69%, or 9 compounds from external test set.

4.4. Model Interpretation.

The ETM-ASNNs models can be predominantly applied as virtual screening tools and also be useful for elucidating the essential structural and physicochemical requirements for activity. Usually, the main QSAR and pharmacophore elucidation approaches use three-dimensional steric and electrostatic potentials or chemical feature-based pharmacophoric ngerprints whereas the model reported in this paper uses 2D ngerprints (Mulliken atomic charges, Wiberg indices and bond length), in addition to several global physiochemical descriptors in the ASNN model development. In view of these differences, the current ETM-ASNN model reveals inherent features in terms of abstract substructural features (ETMS) that discriminate active and inactive molecules. This list of substructures preferentially present in actives and inactives (activity is specifically dependent on its conformational disposition on the cyclohexanol ring. In the trans position, it is dependent on the cyclohexanol group and the phenyl group and the substituents on the aromatic ring (especially the groups on C_3 and C_5 prevent the activity) can serve as a look-up guide to chemists during a hit-to-lead or lead optimization campaign. In general, substructures that have an incremental effect on activity (Figure 3) include the following: substituted (C_3, C_5, H11, C15, C_{19}, and O_{22}) and substructures that have a decremental effect on activity include the following: substituted (C_3, C_5, H_{14}, C_{19}, and O_{21}).

The ASSN models based on descriptors set can be useful for understanding of the descriptor H1 influenza inhibitory activity of analysed molecules. The descriptors such as, chemical potential (μ), hardness (η), softness (s), electrophilicity index (ω), and condensed fukui functions, and quantum chemical descriptors such as highest occupied molecular orbital (HOMO) energy and lowest unoccupied molecular orbital (LUMO) energy, dipole moment, Mulliken charges are chemical reactivity descriptor [32–36]. For example, HOMO plays the role of the binding ability of the molecules. HOMO energy of molecules having high activity is lower than nonactive ones.

4.5. QSAR Study.

There are two main alternative approaches to random testing to find compounds with superior properties for correlation and quantitative prediction of chemical and physical properties from structure. One consists of theoretical calculations using quantum and statistical mechanics, and the other is QSAR/QSPR. The major goal of QSAR or QSPR studies is to find a mathematical relationship between activity or any other property under investigation (eg LD_{50}, pKa etc.) and one or more descriptors related to the structure of the molecule. While some of the descriptors can themselves be experimental properties of the molecule, it is generally more useful to use descriptors derived mathematically from either 2D or 3D molecular structure, since this allows any relationship derived to be extended to the prediction of the property or activity for unavailable compounds. If an acceptable model of this type can be found, it can guide the synthetic chemist in the choice between alternative hypothetical structures. More fundamentally, such studies can illuminate, or even elucidate the mechanism by which the property or activity in question is related to the chemical structure.

The mathematical foundation of the QSAR is based on the principle of polylinearity. Multiple linear regression is a common method used in QSAR studies. The QSAR equations can be obtained by forward stepwise multiple regression shown below.

Activity $= f$(physicochemical properties and/or structural properties)

or

$$f(\text{biological activity}) = f(\text{electronic}) + f(\text{steric})$$
$$+ f(\text{hydrophobic}) + f(\text{structural})$$
$$+ f(\text{theoretical}), \tag{2}$$

or

$$\text{Biological activity} = a_0 + a_1 D_1 + a_2 D_2 + a_3 D_3 + \cdots a_n D_n, \tag{3}$$

where D_1, D_2, D_3 and D_n are descriptors, n is the number of descriptors. The intercept (a_0) and the regression coefficient of the descriptors are determined using the least squares method.

In this study, by using the multiple regression, several QSAR models were performed with the help of some of the calculated descriptors and the compounds activity data from the different skeletons. Below are the regression equations.

Skeleton I

$$A_{\text{Theor}} = 10.2752\Delta E^1 + 328.707\omega + 47.11191\mu$$
$$- 1.71759 * \text{Polarizability} + 0.25237 E^2$$
$$+ 509.4651, \tag{4}$$
$$R^2 = 0.9999, \quad \text{standard error} = 0.$$

TABLE 3: Comparison of experimental and theoretical the activity by using the multiple regression analysis for compounds in Skeleton I.

No	Theoretical activity	Experimental activity	No	Theoretical activity	Experimental activity
1	0.060605	0.06	9	−47.0767	NA
2	−205.41	NA	10	−34.05735	NA
3	−150.779	NA	11	−120.032	NA
4	−34.0955	NA	12	42.74012	NA
5	−142.783	NA	13	0.050604	0.05
6	−221.083	NA	14	4.50064	4.5
7	−288.973	NA	15	0.030675	0.03
8	−146.076	NA	16	0.400708	0.4
9	−47.0767	NA	17	0.900642	0.9
10	−34.05735	NA			

TABLE 4: Comparison of experimental and theoretical the activity by using the multiple regression analysis for compounds in Skeleton II.

No	Theoretical acitivity	Experimental activity	No	Theoretical acitivity	Experimental activity
18	2.404503	0.2	28	−3.3007	0.03 ± 0.07
19	−2.23585	0.02	29	4.28845	1.0
20	2.566251	0.05	30	21.714	25
21	−2.42417	0.065 ± 0.035	31	−0.14402	2.0
22	3.207149	0.015 ± 0.007	32	2.274756	0.15 ± 0.071
23	1.730689	0.038	33	1.167479	0.5
24	−1.0243	0.09	34	−1.9877	1.8
25	−0.26089	NA	35	3.457774	2.5
26	−1.73024	0.15	36	52.1551	NA
27	0.338793	0.023 ± 0.011	37	−6.66909	NA

Skeleton II

$$A_{\text{Theor}} = 178.0305\Delta E + 93.2522\mu - 171.2043\eta$$
$$+ 716.9193S + 44.3081\omega - 6.0477 * \text{dipole}$$
$$- 0.256 * \text{polariz} + 0.0015E^2,$$
$$R^2 = 0.8475, \quad \text{standard error} = 3.4522.$$

(5)

Skelelton III

$$A_{\text{Theor}} = 196.5666E_{\text{HOMO}} + 300.7505\Delta E$$
$$+ 53.5951\eta + 1547.8892S$$
$$+ 135.0100 * \omega - 15.4192 * \text{dipole}$$
$$+ 0.1275 * \text{polariz} + 0.00174E^2,$$
$$R^2 = 0.9538, \quad \text{standard error} = 1.5980.$$

(6)

From the three equations above, it shows that statistically the most significant one is the correlated parameters in Skeleton I with $R^2 = 0.9999$. There are some deviations

for skeleton II and III. It also implies that the activity of the compounds in the skeletons does not only depend on the quantum chemical descriptors but may be influenced by other parameters.

Below are tables of theoretically calculated activities and experimental activities (Tables 3, 4, and 5). Skeleton I shows better correlation between the theoretically calculated activities and experimental activities while skeletons II and III do not show good agreement between them.

For each compound, 19 QSAR descriptors such as molecular weight, E_{HOMO}, E_{LUMO}, ΔE, chemical potential, softness, electrophilicity index, dipole moment, and others were calculated. Preliminary study by ASNNs for all descriptors resulted in model with good statistical qualities (classified correctly 87%, or 45 compounds from 52). Only 3 compounds in the external test set were predicted incorrectly (classified correctly 77%, or 10 compounds from 13).

Thus, the comparison of both models allowed us to conclude that both approaches provide practically similar results using the different set of descriptors. It should be noted that three identical molecules (5, 10, and 65) were classified incorrectly by both models. As it is seen, the approach presented in this paper has shown quite satisfactory results. This fact tells in favor of workability of both found models. These models can be applied to the design of new potent drugs.

TABLE 5: Comparison of experimental and theoretical the activity by using the multiple regression analysis for compounds in Skeleton III.

No	Theoretical acitivity	Experimental activity	No	Theoretical acitivity	Experimental activity
38	38.2127	1–3	52	−0.52359	NA
39	64.19797	NA	53	0.78188	NA
40	57.64826	NA	54	−11.3917	NA
41	1.941236	1.6–3.1	55	5.240327	4
42	20.35258	NA	56	−7.94074	NA
43	0.257742	0.4–0.8	57	11.32183	NA
44	1.530218	0.8	58	12.43476	NA
45	0.690143	0.8	59	4.718806	NA
46	0.643486	0.2-0.3	60	6.527825	8
47	0.757012	0.6	61	−20.8307	10–20
48	3.174496	NA	62	−7.68755	NA
49	4.757657	NA	63	10.20552	10
50	2.120876	2	64	43.48481	5–10
51	−0.07988	0.5	65	−23.5418	NA

5. Conclusion

A series of thiobenzamide and quinolizidine derivatives activity-binding affinity is studied by means of the ETM and ANN, which takes into account both structural and electronic characteristics of molecules. Based on pharmacophores and antipharmacophores calculated by the ETM-software (as submatrices containing important spatial and quantum chemistry characteristics), a system for the activity prognostication is developed. The system was tested on a few molecules with molecular skeletons other than those that were characteristic of the training sets. It allows for identifying the presence/absence of the analgesic activity (with probabilities 84%-85%) in molecules with diverse structures and predicting the level of the activity.

QSAR models were performed with the help of some of the calculated descriptors and the compounds activity data from the different skeletons by using the multiple regression. Skeleton 1 shows better correlation between the theoretically calculated activities and experimental activities while skeletons 2 and 3 do not show good agreement between them.

QSAR descriptors such as molecular weight, as E_{HOMO}, E_{LUMO}, ΔE, chemical potential, softness, electrophilicity index, dipole moment, and others were calculated to compare classical and ETM model. Thus, the comparison of both models allowed us to conclude that both approaches provide practically similar results using the different set of descriptors. It should be noted that three identical molecules (5, 10, 65) were classified incorrectly by both models. As it is seen, the approach presented in this paper has shown quite satisfactory results. This fact tells in favor of workability of both found models. These models can be applied to the design of new potent drugs.

Acknowledgments

References

[1] K. L. Yu, A. F. Torri, G. Luo et al., "Structure-activity relationships for a series of thiobenzamide influenza fusion inhibitors derived from 1,3,3-trimethyl-5-hydroxy-cyclohexylmethylamine," *Bioorganic and Medicinal Chemistry Letters*, vol. 12, no. 23, pp. 3379–3382, 2002.

[2] K. L. Yu, E. Ruediger, G. Luo et al., "Novel quinolizidine salicylamide influenza fusion inhibitors," *Bioorganic and Medicinal Chemistry Letters*, vol. 9, no. 15, pp. 2177–2180, 1999.

[3] G. Stamatiou, G. B. Foscolos, G. Fytas et al., "Heterocyclic rimantadine analogues with antiviral activity," *Bioorganic and Medicinal Chemistry*, vol. 11, no. 24, pp. 5485–5492, 2003.

[4] K. D. Combrink, H. B. Gulgeze, K.-L. Yu et al., "Salicylamide inhibitors of influenza virus fusion," *Bioorganic and Medicinal Chemistry Letters*, vol. 10, no. 15, pp. 1649–1652, 2000.

[5] W. J. Brouillette, S. N. Bajpai, S. M. Ali et al., "Pyrrolidinobenzoic acid inhibitors of influenza virus neuraminidase: modifications of essential pyrrolidinone ring substituents," *Bioorganic and Medicinal Chemistry*, vol. 11, no. 13, pp. 2739–2749, 2003.

[6] A. S. Dimoglo, "Compositional approach to electronic structure description of chemical compounds, oriented computer analysis of structure-activity relation," *Khimiko Farmatsevtlcheskii Zhurnal*, vol. 4, pp. 438–444, 1985.

[7] N. M. Shvets, "Applied program system for the prognosis of biological activity of chemical compounds: development and use," *Computer Science Journal of Moldova*, vol. 1, pp. 101–110, 1993.

[8] N. M. Shvets, "The study of data and control flows and the user interface organization in an applied system used in chemistry and medicine for the biological activities prediction," *Computer Science Journal of Moldova*, vol. 3, pp. 301–311, 1997.

[9] A. S. Dimoglo, P. F. Vlad, N. M. Shvets, and M. N. Coltsa, "Structure-ambergris odour relationship investigation in a mixed series of decalin and non-decalin compounds: the electronic-topological approach," *New Journal of Chemistry*, vol. 25, no. 2, pp. 283–288, 2001.

[10] A. S. Dimoglo, N. M. Shvets, I. V. Tetko, and D. J. Livingstone, "Electronic-topological investigation of the structure—acetylcholinesterase inhibitor activity relationship in the series of *n*-benzylpiperidine derivatives," *Quantitative Structure-Activity Relationships*, vol. 20, no. 1, pp. 31–45, 2001.

[11] A. S. Dimoglo, M. Y. Gorbachov, T. I. Lesnik, M. Saraçoglu, Y. Güzel, and I. Yildirim, "Investigation of the relationship between chemical structure and Anti-HIV-1 activity in a class of nucleoside analogues: electron-topological approach," *Current Medicinal Chemistry*, vol. 4, no. 1, pp. 23–34, 1997.

[12] A. S. Dimoglo, P. F. Vlad, and N. M. Shvets, "Electronic-topolocigal investigations of the relationship between chemical structure and ambergris odor," *New Journal of Chemistry*, vol. 19, pp. 1217–1226, 1995.

[13] E. Sim, A. Dimoglo, N. Shvets, and V. Ahsen, "Electronic-topological study of the structure-activity relationship in a series of piperidine morphinomimetics," *Current Medicinal Chemistry*, vol. 9, no. 16, pp. 1537–1545, 2002.

[14] A. S. Dimoglo, Y. M. Chumakov, B. N. Dobrova, and M. Saracoglu, "Electron-topological investigation of the structure-antitumor activity relationship of thiosemicarbazone derivatives," *Drug Research*, vol. 47, no. 4, pp. 415–419, 1997.

[15] E. Oruç, S. Rollas, F. Kandemirli, N. M. Shvets, and A. S. Dimoglo, "1,3,4 Thiadazole derivatives synthesis, structure elucidation and structure-antituberculosis activity relationships investigations," *Journal of Medicinal Chemistry*, vol. 47, no. 27, pp. 6760–6767, 2004.

[16] A. Altun, M. Kumru, and A. Dimoglo, "Study of electronic and structural features of thiosemicarbazone and thiosemicarbazide derivatives demonstrating anti-HSV-1 activity," *Journal of Molecular Structure*, vol. 535, pp. 235–246, 2001.

[17] A. Altun, K. Golcuk, M. Kumru, and A. F. Jalbout, "Electron-conformational study for the structure-hallucinogenic activity relationships of phenylalkylamines," *Bioorganic and Medicinal Chemistry*, vol. 11, no. 18, pp. 3861–3868, 2003.

[18] F. Kandemirli, N. Tokay, N. M. Shvets, and A. S. Dimoglo, "Investigation of structure-activity relationship on 17-spirolactone derivatives: the electronic-topological approach," *Farmaco*, vol. 57, no. 7, pp. 601–607, 2002.

[19] F. Kandemirli, M. Saraçoglu, and V. Kovalishyn, "Human acetylcholinesterase inhibitors: electronic-topological and neural network approaches to the structure-activity relationships study," *Mini-Reviews in Medicinal Chemistry*, vol. 5, no. 5, pp. 479–487, 2005.

[20] M. Saracoglu and F. Kandemirli, "Investigation of structure-activity relationships of tacrine analogues: electronic-topological method," *The Open Medicinal Chemistry Journal*, vol. 2, pp. 75–80, 2008.

[21] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785–789, 1988.

[22] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[23] M. J. Frisch, G. W. Trucks, H. B. Schlegel et al., *Gaussian 09, Revision A.1*, Gaussian, Inc., Wallingford, Conn, USA, 2009.

[24] S. I. Gorelsky, *AOMix: Program for Molecular Orbital Analysis*, University of Ottawa, Canada, 2009, http://www.sg-chem.net/.

[25] S. I. Gorelsky and A. B. P. Lever, "Electronic structure and spectra of ruthenium diimine complexes by density functional theory and INDO/S. Comparison of the two methods," *Journal of Organometallic Chemistry*, vol. 635, no. 1-2, pp. 187–196, 2001.

[26] I. V. Tetko, V. V. Kovalishyn, and D. J. Livingstone, "Volume learning algorithm artificial neural networks for 3D QSAR studies," *Journal of Medicinal Chemistry*, vol. 44, no. 15, pp. 2411–2420, 2001.

[27] T. Kohonen, *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*, Springer, New York, NY, USA, 2001.

[28] I. V. Tetko, "Neural network studies. 4. Introduction to associative neural networks," *Journal of Chemical Information and Computer Sciences*, vol. 42, no. 3, pp. 717–728, 2002.

[29] A. Dimoglo, V. Kovalishyn, N. Shvets, and V. Ahsen, "The structure—inhibitory activity relationships study in a series of cyclooxygenase-2 inhibitors. A combined electronic-topological and neural networks approach," *Mini-Reviews in Medicinal Chemistry*, vol. 5, no. 10, pp. 879–892, 2005.

[30] I. V. Tetko and A. E. P. Villa, "Efficient partition of learning data sets for neural network training," *Neural Networks*, vol. 10, no. 8, pp. 1361–1374, 1997.

[31] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network Studies. 1. Comparison of overfitting and overtraining," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 5, pp. 826–833, 1995.

[32] I. V. Tetko, A. E. P. Villa, and D. J. Livingstone, "Neural network studies. 2. Variable selection," *Journal of Chemical Information and Computer Sciences*, vol. 36, no. 4, pp. 794–803, 1996.

[33] V. V. Kovalishyn, I. V. Tetko, A. I. Luik, V. V. Kholodovych, A. E. P. Villa, and D. J. Livingstone, "Neural network studies. 3. Variable selection in the cascade-correlation learning architecture," *Journal of Chemical Information and Computer Sciences*, vol. 38, no. 4, pp. 651–659, 1998.

[34] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, NY, USA, 1989.

[35] R. Parthasarathi, J. Padmanabhan, U. Sarkar et al., "Toxicity analysis of benzidine through chemical reactivity and selectivity profiles: a DFT approach," *Internet Electronic Journal of Molecular Design*, vol. 2, no. 12, pp. 798–813, 2003.

[36] R. Parthasarathi, J. Padmanabhan, V. Subramanian, B. Maiti, and P. K. Chattaraj, "Chemical reactivity profiles of two selected polychlorinated biphenyls," *Journal of Physical Chemistry A*, vol. 107, no. 48, pp. 10346–10352, 2003.