ELSEVIER

Adaptive neuro-fuzzy inference system (ANFIS): A new approach to predictive modeling in QSAR applications: A study of neuro-fuzzy modeling of PCP-based NMDA receptor antagonists

Erdem Buyukbingol,[a,*] Arzu Sisman,[b] Murat Akyildiz,[c]
Ferda Nur Alparslan[b] and Adeboye Adejare[d]

[a]Ankara University, Faculty of Pharmacy (ECZACILIK), Department of Pharmaceutical Chemistry,
Tandogan 06100, Ankara, Turkey
[b]Middle East Technical University, Faculty of Engineering, Department of Computer Engineering, 06531, Ankara, Turkey
[c]Ankara University, Faculty of Educational Sciences, Cebeci 06590, Ankara, Turkey
[d]University of the Sciences in Philadelphia, Philadelphia College of Pharmacy, Department of Pharmaceutical Sciences,
600 South 43rd Street, Philadelphia, PA 19104, USA

Abstract—This paper proposes a new method, Adaptive Neuro-Fuzzy Inference System (ANFIS) to evaluate physicochemical descriptors of certain chemical compounds for their appropriate biological activities in terms of QSAR models with the aid of artificial neural network (ANN) approach combined with the principle of fuzzy logic. The ANFIS was utilized to predict NMDA (N-methyl-D-Aspartate) receptor binding activities of phencyclidine (PCP) derivatives. A data set of 38 drug-like compounds was coded with 1244 calculated molecular structure descriptors (clustered in 20 data sets) which were obtained from several sources, mainly from Dragon software. Prior to the progress to the ANFIS system, descriptors from the best subsets were selected using unsupervised forward selection (UFS) to eliminate redundancy and multicollinearity followed by fuzzy linear regression algorithm (FLR) which was used for variable selection. ANFIS was applied to train the final descriptors (Mor22m, E3s, R3v+, and R1e+) using a hybrid algorithm consisting of back-propagation and least-square estimation while the optimum number and shape of related functions were obtained through the subtractive clustering algorithm. Comparison of the proposed method with traditional methods, that is, multiple linear regression (MLR) and partial least-square (PLS) was also studied and the results indicated that the ANFIS model obtained from data sets achieved satisfactory accuracy.
© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Artificial Neural Networks (ANN) have appealed to the growing interest of researchers in various scientific areas due to the increasing need of adaptive intelligent systems to solve the factual prediction problems in designing of new drugs, and could be widely accepted as a QSAR methodology offering an alternative way to deal with complex and ill-defined problems in prediction of drug–receptor interactions.[1–8] The prediction of future assessments on the drug–receptor interactions or certain biological systems is often compulsory to the cost-effective studies of drug design resources. On the other hand, Fuzzy logic is a problem-solving technique that defines and generates responses based on ambiguous, incomplete and imprecise information, and fuzzy systems have fascinated the growing attention and interest in bioinformatics applications, decision making studies, pattern recognition, and data analysis.[9,10] Therefore, in this study, a new approach based on adaptive neuro-fuzzy inference system (ANFIS) is presented for prediction of NMDA receptor binding properties of PCP-related compounds.

Quantitative structure–activity/property relationship (QSAR/QSPR) methods are common and rather successful techniques in drug design. However, in cases of complex relationships, conventional QSAR/QSPR methods

often lead to insufficient or misleading information because of nonlinear relationships within the data set. Precise nonlinear functions such as the bilinear model for $\log P$ and log potency dependencies have to be used on a trial and error basis.[11] Moreover, for membrane-bound receptors (such as NMDA receptors), the biological activities often result from both membrane interaction and receptor binding, which may also lead to nonlinear dependencies. One possibility of overcoming the difficulties of such nonlinearities in QSAR/QSPR studies is the use of ANN, which has gained interest in the field of drug design.[12–22] After a proper learning procedure; ANN should be able to 'recognize' basic correlations in a given data set and to predict, for example, physicochemical properties and pharmacological activities. The creation of a wide collection of molecular descriptors based on molecular modeling techniques has turned out to be a routine part of QSAR studies. This approach has illustrated difficulty in assortment of the best descriptors to use for future predictions. ANN usage consists of a training phase in which the model descriptors are established from a set of training data. Optional but highly suggested steps include validation in which generalization capability of the model is established, followed by a judgment stage in which the biological properties of compounds are predicted using the optimized model.

The formation of numerical inputs which are generally molecular descriptors is a crucial and complicated problem. This is due to the fact that molecular descriptors are expected to characterize structural attributes interrelated to the desired biological activity as clearly as possible. Therefore, the correctness of ANN depends intensely on the level of correlation between molecular descriptors and structural characteristics. Since it is not possible to know a priori which molecular descriptors are most pertinent to the dilemma at hand, ANN is often used in concurrence with optimization techniques for feature selection, varying from simple keen attempts such as forward selection or backward elimination,[23] to more complicated methodologies such as simulated annealing,[24] genetic algorithms[25] and, most recently, artificial ant colony systems,[26] and particle swarms.[27] The problem of input variable selection is well known in the task of modeling data. In many modeling problems, for example in the context of biomedical, industrial, or environmental systems, a problem can occur when developing multivariate models and the best set of inputs to use are not known. This is particularly true when using ANNs. In this case, unrequired inputs can significantly increase learning complexity, and hence, input variable selection is required at determining which input variables are aimed for a model. The variable selection procedures based on stepwise regression[28] and partial least-squares/principal component analysis (PLS/PCA)[29] were the techniques most commonly used previously. Another technique for variable selection is genetic algorithm (GA)[30] which is also one of the most common known methods of evolutionary systems. However, in our study, a variable selection method employing a fuzzy linear regression (FLR) algorithm was used to select the set of input descriptors.[31] Before this, we would like to find the features that most accurately discriminate among the descriptor classes and will yield the highest classification accuracy. The problem of selecting a subset of relevant features in a potentially overwhelming quantity of data is classic and found in several QSAR studies. From a practical perspective, large amounts of irrelevant features affect learning algorithms in several levels and it is obvious that many of the learning problems do not scale well with the growth of irrelevant features.

This is a difficult problem which is hardly ever addressed in the QSAR studies and there is an issue of whether one is looking for the minimal set of (relevant) features. From a practical perspective, it is obvious that large amounts of irrelevant features affect learning algorithms at many levels. Therefore, we have used the unsupervised forward selection (UFS) method which is the process of choosing a subset of the original predictive descriptors by eliminating redundant and uninformative ones (multicollinearity)[32] for fuzzy linear regression studies. The UFS method uses a technique of pairwise correlation to remove redundancy and reduce multicollinearity, and is based on CORCHOP procedure[33] which rely on the independent variables are merely subjected to the procedure. Fuzzy Linear Regression technique was then applied for future variable selection of those obtained from routine UFS and this selection algorithm was performed on 20 separate descriptor subsets.

Molecular modeling studies of NMDA receptor receptor antagonism constitute an active area of research.[34] These receptors arbitrate many of the excitatory neurotransmission in the mammalian central nervous system.[35,36] Investigation of new compounds that act on NMDA receptors has been focus of research in recent years and has resulted in the innovation of numerous forms of antagonists.[37] NMDA receptor antagonists are being studied as possible compounds for alleviating diseases related to neuronal excitotoxicity including brain damage found in cerebral ischemia, epilepsy, and Parkinson and Alzheimer's Disease.[38–40] Our study involved a set of closely related structural analogues of PCP with a limited chemical space to be used for predictive value for similar NMDA ligand receptor bindings. To the extent that PCP is an important ligand in NMDA antagonism, this study is expected to give an insight into a global descriptor model of PCP-related compounds which bind to NMDA receptor. Moreover, in this study, we evaluated the use of molecular descriptors as input for fuzzy-neural system on the basis of ANFIS.[41,42]

2. Theoretical routines

2.1. Basic descriptions

The methodologies of fuzzy logic and ANNs are described in recent affirmative publications.[43,44] A Fuzzy Inference System (FIS), also known as a fuzzy-rule-based system, is an approach for building qualitative models of human knowledge and reckoning abilities without having to use exact quantitative calculations. The ANFIS[31,41] is the achievement of a Fuzzy Inference

System in the form of a multi-layer neural network. This algorithm of the fields of fuzzy logic and neural networks has been suggested to give a more powerful tool for dealing with nonlinear and ill-defined problems,[45] and to produce better generalization capabilities.

2.2. Fuzzy linear regression (FLR)

The FLR model was first introduced by Tanaka[46]. In this section, definitions related to FLR are presented, then the problem is defined. In the scope of the study, only non-fuzzy data processing is described. The conventional regression methods treat the difference between the observed data and the obtained output as the error. In FLR method, this difference is accepted as the ambiguity in the data.[44] This ambiguity is represented as the fuzzy coefficients in the fuzzy linear equations. These fuzzy coefficients are fuzzy numbers. A fuzzy regression model is used in evaluating the functional relationship between the dependent and independent variables in a fuzzy environment. Most fuzzy regression models are considered to be fuzzy outputs and parameters but non-fuzzy (crisp) inputs. In general, there are two approaches in the analysis of fuzzy regression models: linear-programming-based methods and fuzzy least-squares methods. In 1992, Sakawa and Yano[47] considered fuzzy linear regression models with fuzzy outputs, fuzzy parameters and also fuzzy inputs. They formulated multiobjective programming methods for the model estimation along with a linear-programming (LP) based approach. In this paper, linear programming based estimation method was employed to propose relevant descriptors which can be effectively utilized in NMDA binding prediction.

A fuzzy number A is represented by a center α and a width c. The membership value for a fuzzy number is given by the equation $A(x) = L((x - \alpha)/c)$. The membership function must be symmetric. In Figure 1, a triangular fuzzy number can be seen where the membership function is $A(x) = 1 - |x - \alpha|/c$.

2.2.1. Linear programming problem.
Fuzzy linear regression finds an equation for a dependent variable in which fuzzy numbers are used as coefficients of independent variables. The data handled can be either standard (with non-fuzzy output) or fuzzy data. In the scope of the study only FLR for non-fuzzy data is investigated.

Non-fuzzy data (Crisp)

The input data is given as $\vec{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})$ for non-fuzzy output y_i where i denotes the sample index. The fuzzy linear regression model produces the equation

$$Y_i = A_0 + A_1 * x_{i1} + A_2 * x_{i2} + \cdots + A_n * x_{in},$$

where Y_i is the fuzzy output and $\vec{x}_i$ is the input vector and $*$ denotes the scalar multiplication operation. The coefficients A_j (where $j = 1, \ldots, n$) are triangular fuzzy numbers with (α_j, c_j).

The coefficients in the model are computed by using the following linear programming (LP) problem. The objective function of the LP problem is;

$$\text{minimize } J(c) = \sum_{i=1}^{T} c^t |x_i|$$

and the constraints are;

$$y_i \leqslant x_i^t \alpha + |L^{-1}(h)||c^t|x_i|$$
$$y_i \geqslant x_i^t \alpha - |L^{-1}(h)||c^t|x_i|$$

where t is the size of data set, $|x_i| = (|x_1|, \ldots, |x_n|)^t$, $L(x) = 1 - |x|$, $c > 0$ and $i = 1, \ldots, n$. h denotes the degree to which the given data (y_i, x_i) are included in the output fuzzy number Y_i.[44] If a coefficient has a width c closer to zero, then the variable has less ambiguity, since the coefficient is similar to a non-fuzzy value. An example is given in Figure 2. The model in the figure is $Y_i = A_0 + A_1 x_i$ where $A_0 = (2,1)$ and $A_1 = (3, 0.5)$ and $h = 0.5$. Y_1, Y_2, and Y_3, are three fuzzy outputs for the non-fuzzy input–output pairs (y_1, x_1), (y_2, x_2), (y_3, x_3).

2.3. ANFIS structure

ANFIS is a neuro-fuzzy system developed by Jang.[41,48,49] It has a feed-forward neural network structure where each layer is a neuro-fuzzy system
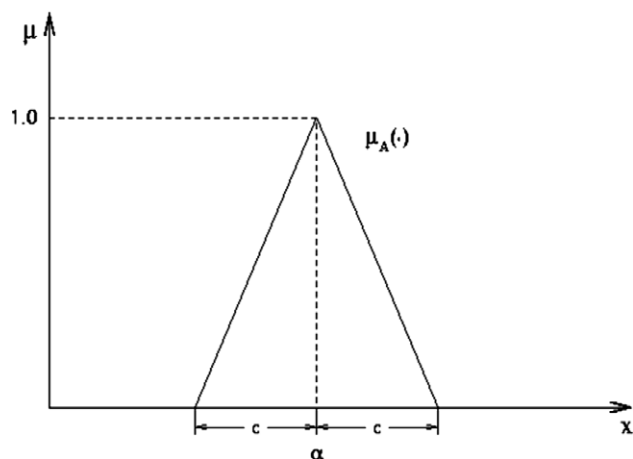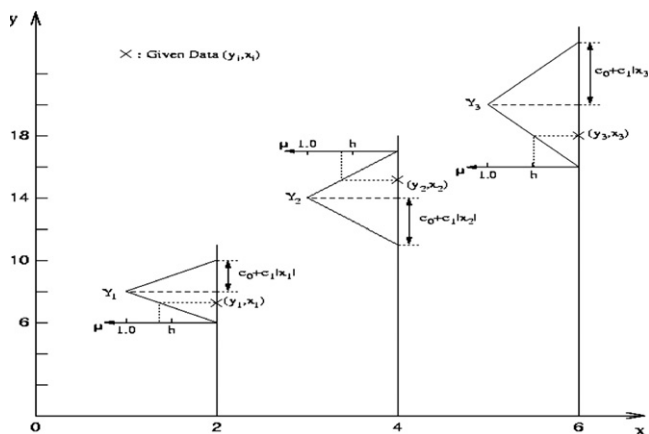

Figure 1. A triangular fuzzy number.

Figure 2. A fuzzy linear regression model.

component (Fig. 3). It simulates TSK (Takagi–Sugeno–Kang) fuzzy rule[50] of type-3 where the consequent part of the rule is a linear combination of input variables and a constant. The final output of the system is the weighted average of each rule's output. The form of the type-3 rule simulated in the system is as follows:

IF x_1 is A_1 **AND** x_2 is A_2 **AND** ... **AND** x_p is A_p

   THEN $y = c_0 + c_1x_1 + c_2x_2 + \cdots + c_px_p$

where x_1 and x_2 are the input variables, A_1 and A_2 are the membership functions, y is the output variable, and c_0, c_1, and c_2 are the consequent parameters.

The neural network structure contains six layers.

- Layer 0 is the *input* layer. It has n nodes where n is the number of inputs to the system.
- The fuzzy part of ANFIS is mathematically incorporated in the form of membership functions (MFs). A membership function $\mu_{Ai}(x)$ can be any continuous and piecewise differentiable function that transforms the input value x into a membership degree, that is to say a value between 0 and 1. The most widely applied membership function is the generalized bell (gbell MF), which is described by the three parameters, a, b, and c (Eq. 1). Therefore, Layer 1 is the *fuzzification* layer in which each node represents a membership value to a linguistic term as a Gaussian function with the mean;

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i}\right)^2\right]^{b_i}}, \qquad (1)$$

where a_i, b_i, and c_i are parameters of the function. These are adaptive parameters. Their values are adapted by means of the back-propagation algorithm during the learning stage. As the values of the parameters change, the membership function of the linguistic term, A_i changes. These parameters are called *premise parameters*. In that layer there exist $n*p$ nodes where n is the number of input variables and p is the number of membership functions. For example, if *size* is an input variable and there exist

two linguistic values for size which are SMALL and LARGE then two nodes are kept in the first layer and they denote the membership values of input variable *size* to the linguistic values SMALL and LARGE.

- Each node in Layer 2 provides the strength of the rule by means of multiplication operator. It performs AND operation.

$$w_i = \mu_{A_i}(x_0) * \mu_{B_i}(x_1)$$

Every node in this layer computes the multiplication of the input values and gives the product as the output as in the above equation. The membership values represented by $\mu_{Ai}(x_0)$ and $\mu_{Bi}(x_1)$ are multiplied in order to find the firing strength of a rule where the variable x_0 has linguistic value A_i and x_1 has linguistic value B_i in the antecedent part of Rule I.

There are p^n nodes denoting the number of rules in Layer 2. Each node represents the antecedent part of the rule. If there are two variables in the system namely x_1 and x_2 that can take two fuzzy linguistic values, SMALL and LARGE, there exist four rules in the system whose antecedent parts are as follows:

   IF x_1 is SMALL **AND** x_2 is SMALL
   IF x_1 is SMALL **AND** x_2 is LARGE
   IF x_1 is LARGE **AND** x_2 is SMALL
   IF x_1 is LARGE **AND** x_2 is LARGE

- Layer 3 is the *normalization* layer which normalizes the strength of all rules according to the equation

$$\overline{w_i} = \frac{w_i}{\sum_{j=1}^{R} w_j},$$

where w_i is the firing strength of the ith rule which is computed in Layer 2. Node i computes the ratio of the ith rule's firing strength to the sum of all rules' firing strengths. There are p^n nodes in this layer.

- Layer 4 is a layer of adaptive nodes. Every node in this layer computes a linear function where the function coefficients are adapted by using the error function of the multi-layer feed-forward neural network.

$$\overline{w_i}f_i = \overline{w_i}(p_0x_0 + p_1x_1 + p_2)$$

p^i's are the parameters where $i = n + 1$ and n is the number of inputs to the system (i.e., number of nodes in Layer 0). In this example, since there exist two variables (x_1 and x_2), there are three parameters p_0, p_1 and p_2) in Layer 4 and $\overline{w_i}$ is the output of Layer 3. The parameters are updated by a learning step. Kalman filtering based on least-squares approximation and back-propagation algorithm is used as the learning step.
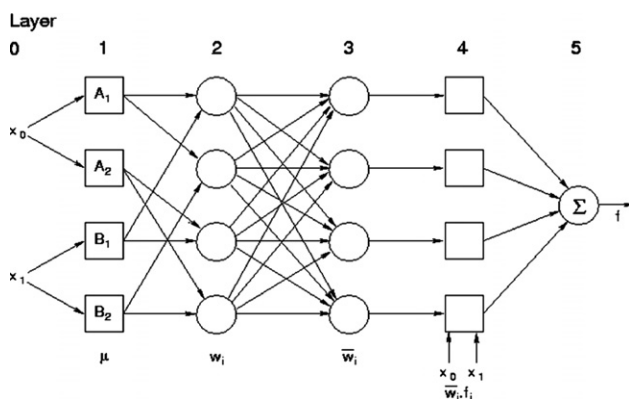

Figure 3. Basic ANFIS structure.

- Layer 5 is the *output* layer whose function is the summation of the net outputs of the nodes in Layer 4. The output is computed as:

$$\sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i},$$

where $\overline{w}_i f_i$ is the output of node *i* in Layer 4. It denotes the consequent part of rule *i*. The overall output of the neuro-fuzzy system is the summation of the rule consequences. ANFIS uses a hybrid learning algorithm in order to train the network. For the parameters in the layer 1, back-propagation algorithm is used. For training the parameters in the Layer 4, a variation of least-squares approximation or back-propagation algorithm is used.

3. Experimental

3.1. Data set and descriptor generation

This study utilized a data set, comprised of 20 subsets of PCP-like compounds, each having NMDA receptor binding affinity values (K_i).[51] The binding assays measured displacement of ^{3}H-TCP in μM (Table 1). The biological data used in this study were obtained by transforming the raw data into—log values, for example, the PCP binding activity is expressed as 7.15 by transforming the 0.07 μM value $[-\log (0.07 \times 10^{-6})]$.

The computations were carried out with Pentium-4 computers running Windows XP and Linux (Slackware, 10.2) operating systems, using programs written (C++ and Perl) by the authors. At first, the structures were submitted to a conformational search, and the geometry optimizations of all structures headed to energy minima that were accomplished by using AM1 self-consistent molecular orbital (SCF MO) method at the restricted Hartree-Fock (RHF) level of *HyperChem*[52] Program. Optimization studies of the compounds were achieved by the treatment of the Polak-Ribiere method based on conjugate gradient algorithm. Many descriptors have been formulated to study the similarity, or dissimilarity, of compounds and used to predict the relationships between physicochemical properties and biological activities in order to identify compounds with desirable or 'drug-like' properties. Although descriptors include various types of molecular fingerprints, it is quite difficult to apply all of them to predict biological activities due to difficulties in calculation procedures. However, some freely distributed programs were used to determine and calculate the descriptors. One of the excellent software written for this purpose is e-*Dragon 1.0*[53] which originated from the studies of Todeschini and Consonni.[54] *Dragon* provides molecular descriptors that are divided into 20 logical blocks. We have calculated 18 molecular descriptor categories out of 20. The excluded two data sets (the functional group and atom-centered fragment descriptors) were eliminated because the program gave zero values for the compounds. The molecular structures of compounds were first encoded using the SMILES notation, which was subsequently transformed online to the e-Dragon program to provide the 18 initial sets of molecular descriptors. Apart from Dragon-calculated descriptors, we also calculated some traditional descriptors arising from quantum mechanical and lipophilic (log *P*) and solublity calculations from different sources. Thus, we studied 20 data sets in total including those obtained from Dragon software, quantum mechanical parameters (A set of quantum-chemical descriptors, i.e., heat of formation, dipole moment, charges on nitrogen and neighboring carbon atoms, HOMO, LUMO, and HOMO–LUMO energy gap, etc., has also been obtained for the each molecule after geometry optimization procedure) obtained from *HyperChem* and *Chem3D*,[55] and lipophilicity and solubility[56] properties based on log *P* parameters acquired from several sources.[56,57]

3.2. Unsupervised forward selection (UFS)

Totally 1244 descriptors were utilized for the unsupervised forward selection (UFS) to remove the inter-correlating descriptors (redundancy)[32] and reduce the multicollinearity.[33]
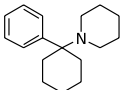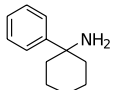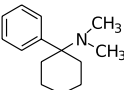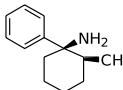
3.3. FLR tests
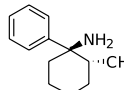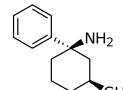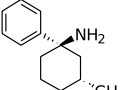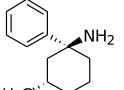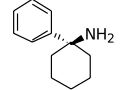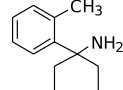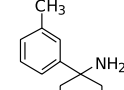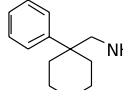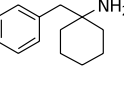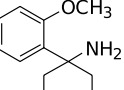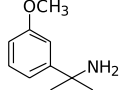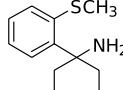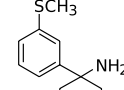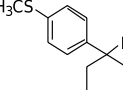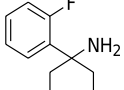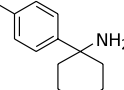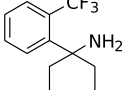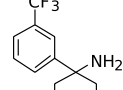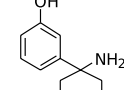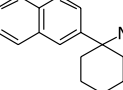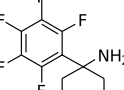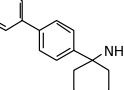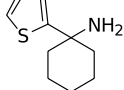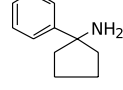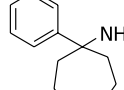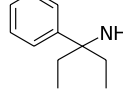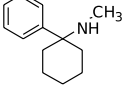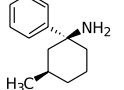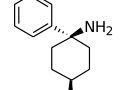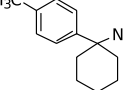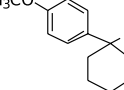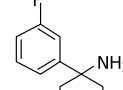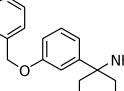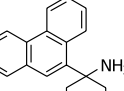
Fuzzy linear regression (FLR) was applied to data $(-\log K_i)$ and we have utilized descriptors which were classified in 20 data sets for the first FLR. For each data set, the fuzzy coefficients having largest center and smallest width values were extracted. Due to the center of the fuzzy coefficient of a value in a fuzzy regression increases, the effect of that value to the output increases accordingly. Moreover, the width of a fuzzy coefficient increases, the more the ambiguity caused by that value (The more fuzziness in the data set, the larger the coefficient width). In other words, if there are not enough data to describe the model at hand, the fuzzy coefficients obtained at the end of FLR have larger widths. The **largest** coefficient for each parameter set is highlighted as bold. The coefficients not listed in the table all have zero's center.

3.4. Performance measurement

As we have used logarithmic data to establish the predictive capability of the descriptors through the NMDA binding, it should be required to compare the accuracy of the two ANFIS accomplishments. The major concerns of the validation are the evaluation of the models, predictability and validity of the models, and consistency of the results. One of the most common methods is measurement of the Root Mean-Square Error (RMSE) which is a degree of distribution of the data. The RMSE can be calculated by the below mentioned equation where N denotes the number of data patterns, y_i^{pred} is the predicted output for pattern number i, and y_i^{obs} is the experimental output data (Eq. 2).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i^{\text{pred}} - y_i^{\text{obs}})^2} \qquad (2)$$

Table 1. The structural formulas of PCP and related compounds together with their NMDA binding values (K_i). The values in parentheses express the log ($1/K_i$)

1 (PCP) $K_i= 0.07\mu M$ (7.15)	**2** $K_i= 0.527\mu M$ (6.28)	**3** $K_i= 0.722\mu M$ (6.14)	**4** $K_i= 2.0\mu M$ (5.70)	**5** $K_i= 1.60\mu M$ (5.80)	**6** $K_i= 0.502\mu M$ (6.30)
7 $K_i= 1.20\mu M$ (5.92)	**8** $K_i= 1.64\mu M$ (5.79)	**9** $K_i= 8.40\mu M$ (5.08)	**10** $K_i= 7.80\mu M$ (5.11)	**11** $K_i= 0.595\mu M$ (6.23)	**12** $K_i= 5.10\mu M$ (5.29)
13 $K_i= 16.0\mu M$ (4.80)	**14** $K_i= 0.850\mu M$ (6.07)	**15** $K_i= 0.563\mu M$ (6.25)	**16** $K_i= 0.820\mu M$ (6.09)	**17** $K_i= 2.42\mu M$ (5.62)	**18** $K_i= 43.0\mu M$ (4.37)
19 $K_i= 10.70\mu M$ (4.97)	**20** $K_i= 5.13\mu M$ (6.03)	**21** $K_i= 12.40\mu M$ (4.91)	**22** $K_i= 8.40\mu M$ (5.08)	**23** $K_i= 0.225\mu M$ (6.65)	**24** $K_i= 34.00\mu M$ (4.47)
25 $K_i= 10.00\mu M$ (5.00)	**26** $K_i= 86.00\mu M$ (4.07)	**27** $K_i= 0.145\mu M$ (6.84)	**28** $K_i= 10.50\mu M$ (4.98)	**29** $K_i= 2.60\mu M$ (5.59)	**30** $K_i= 6.00\mu M$ (5.22)
31 $K_i= 0.347\mu M$ (6.46)	**32** $K_i= 8.00\mu M$ (5.10)	**33** $K_i= 1.10\mu M$ (5.96)	**34** $K_i= 7.30\mu M$ (5.14)	**35** $K_i= 121.00\mu M$ (3.92)	**36** $K_i= 0.675\mu M$ (6.17)
37 $K_i= 10.60\mu M$ (4.97)	**38** $K_i= 27.50\mu M$ (4.56)				

3.5. ANFIS

In this section, we attempted to investigate the possibility and effectiveness of predicting the descriptor pattern of NMDA receptor binding of PCP-related compounds with an ANFIS approach. In order to find which combination for the training, validating, and testing data sets would give better models according to their RMSE values, we aimed to categorize the data for setting up the data fragmentation into training, validation, and testing subsets with two approaches. In the first attitude, the data set contains 38 values which are sorted according to the descending output value followed by dividing the data set into eight subsets and the smallest value in each five data set was taken as validation data; the value in the medium in each subset was received as test data, and the remaining values were considered for the training data set. This is kind of heuristic way of division and the algorithms with heuristic search strategy are very simple to implement and very fast in producing results because the search space is only quadratic in terms of the number of features. Thus, the training data set contains 22 compounds, validation and test data sets contain eight compounds each (Table 2). Second, the compounds were 30 times randomly separated into train, validate, and test data sets to find out the lowest RMSE values for each. Randomization process was implemented by using *function rand* in Perl language. The data values (each data line) were extracted and written in separate training, validation, and testing data files. The heuristic and last five randomization results are shown in Table 2. For this process ANFIS was trained for 100 epochs.

The purpose of feature selection can be broadly categorized into visualization, understanding of data, data cleaning, redundancy and/or irrelevancy removal, and performance (e.g., prediction and comprehensibility) enhancement. Not every feature selection can serve all the purposes. Therefore, we implemented random-5 selection (shown bold in Table 2) due to the smaller RMSE values rather than other random data set for training, validating, and testing. The graphical representation of the selection model for ANFIS is shown in Figure 4.

Figure 4. Training, validating, and testing RMSE values obtained from random-5 data sets.

4. Results and discussion

4.1. Application of fuzzy linear regression to parameter sets

The main goal of this study was to build fuzzy regression models followed by ANFIS system, to reveal the relationship of PCP-related compounds and NMDA receptor binding, and also to evaluate risk factors of the interaction. The results were used from FLR to develop empirical guidelines for ANFIS relevance.

4.1.1. First pass FLR. The most significant descriptors (inputs) in fuzzy system were automatically identified and for the first pass of FLR, we have obtained **51** descriptors out of 1244 descriptors from the data sets. Every data set was utilized separately by the first FLR and the descriptors obtained from 20 data sets are shown in Table 3. The coefficients not listed in the following lists have all zero center and width values. The first value in parentheses indicates the center of fuzzy number (α), and the second value indicates the width of fuzzy number (c). The **larger** values of α for each of the descriptor sets were chosen for the second FLR. If the values of α are found to be close to zero, that denotes the non-fuzziness of the appropriate descriptors and the higher values indicate the significance of the related descriptors which then are chosen for the next variable selection process.

4.1.2. Second and third pass FLR. The 51 descriptors chosen by the first FLR were applied to the second process of FLR selection and the resulting twelve relevant descriptors are shown in Table 4. The highest significant twelve descriptors were then re-applied to FLR selection for the last time to reduce the descriptor to most significant parameters in order to obtain the final robust parameters due to FLR analysis.

The final selected descriptors in our system can be divided into three groups: 3D-MoRSE, WHIM, and GETAWAY molecular coding. These final descriptors selected by third pass of FLR are **Mor22m** (3D-MoRSE - signal 22/weighted by atomic masses), **E3s** (Third

Table 2. Heuristic and randomized process to obtain the subsets for the data

Type of selection	RMSE values		
	Train	Validate	Test
Heuristic	0.133148	5.546029	0.960877
Random-1	0.186286	1.963804	1.092024
Random-2	0.184986	4.377197	3.881673
Random-3	0.081462	4.479717	11.154102
Random-4	0.188901	1.252476	0.977937
Random-5	**0.132889**	**0.784274**	**0.867565**

Values indicate the RMS errors for training, validating, and testing data sets.

Table 3. Descriptors obtained from the first FLR

Data set	Descriptors		Ref.
1. Constitutional descriptors (21)	Mv = (2.11792, 0)	Mv = Mean atomic van der Waals volume (scaled on Carbon atom)	a
	Mp = (6.41735, 0)	Mp = Mean atomic polarizability (scaled on Carbon atom)	
2. Topological descriptors (74)	J=(0.923415, 0)	J = Balaban distance connectivity index	b,c
	BLI=(1.95652, 0)	BLI = Kier benzene-likeliness index	
3. Walk and Path Counts descriptors (43)	MWC10 = (0.446926, 0.067383)	MWC10 = Molecular walk count of order 10	d
	PCR = (0.328483, 0)	PCR = Ratio of multiple path count over path count	
4. Connectivity Indices (33)	X5A = (27.52, 0)	X5A = Average connectivity index chi-5	c
	X5Av = (49.1179, 0)	X5Av = Average valence connectivity index chi-5	
5. Information Indices (47)	Xindex=(4.07989, 0)	Xindex = Balaban X index	e,f
	BIC3=(2.90046, 0)	BIC3 = Bond information content (neighborhood symmetry of 3-order)	
6. 2D-Autocorrelations (96)	MATS5v = (2.63566, 0.756997)	MATS5v = Moran autocorrelation-lag 5/weighted by atomic van der Waals volumes	g,h
	MATS1e = (2.34841, 0.802463)	MATS1e = Moran autocorrelation-lag 1/weighted by atomic Sanderson electronegativities	
	MATS8p = (1.35378, 0.188721)	MATS8p = Moran autocorrelation-lag 8/weighted by atomic polarizabilities	
	GATS1v = (2.16918, 0)	GATS1v = Geary autocorrelation-lag 1/weighted by atomic van der Waals volumes	
	GATS3v = (1.14392, 0)	GATS3v = Geary autocorrelation-lag 3/weighted by atomic van der Waals volumes	
7. Edge adjacency indices (106)	EEig01x =(1.28984, 0)	EEig01x = Eigenvalue 01 from edge adj. matrix weighted by edge degrees	i
8. BCUT descriptors (64)	BEHm1 = (1.19505, 0)	BEHm1 = Highest eigenvalue n. 1 of Burden matrix/weighted by atomic masses	j
	BELm7 = (0.49801, 0)	BELm7 = Lowest eigenvalue n. 7 of Burden matrix/weighted by atomic masses	
9. Topological Charge Indices (17)	JGI2 = (44.8189, 1.23355)	JGI2 = Mean topological charge index of order2	k
	JGI3 = (32.757, 0)	JGI3 = Mean topological charge index of order3	
10. Eigenvalue-based indices (44)	LP1=(1.9810, 0)	LP1 = Lovasz-Pelikan index (leading eigenvalue)	l,m
	VEA2=(4.88121, 0)	VEA2 = Average eigenvector coefficient sum from adjacency matrix	
11. Randic molecular profiles (41)	DP01 = (1.16507, 0.0382321)	DP01 = Molecular profile no. 01	n
	SHP2 = (3.69109, 0.263984)	SHP2 = Average shape profile index of order 2	
12. Geometrical descriptors (38)	FDI = (2.08395, 0)	FDIfolding degree index	o
	RCI = (0.889407, 0)	RCIJug RC index	
	DISPe = (0.817223, 0.333446)	DISPe = dCOMMA2 value/weighted by atomic Sanderson electronegativities	
13. RDF Descriptors (109)	RDF010m = (5.35146, 0.11741)	RDF010m = Radial Distribution Function-1.0/weighted by atomic masses	p
14. 3 D-MorSE descriptors (160)	Mor09u = (1.10528, 0)	Mor09u = 3D-MoRSE—signal 09/unweighted	q
	Mor14u = (2.34189, 0)	Mor14u = 3D-MoRSE—signal 14/unweighted	
	Mor30u = (1.8916, 0)	Mor30u = 3D-MoRSE—signal 30/unweighted	
	Mor22m = (2.40836, 0)	Mor22m = 3D-MoRSE—signal 22/weighted by atomic masses	
	Mor26m = (2.16008, 0)	Mor26m = 3D-MoRSE—signal 26/weighted by atomic masses	
	Mor23e = (0.998615, 0)	Mor23e = 3D-MoRSE—signal 23/weighted by atomic Sanderson electronegativities	
	Mor10p = (2.68542, 0)	Mor10p = 3D-MoRSE—signal 10/weighted by atomic polarizabilities	
	Mor29p = (1.26781, 0)	Mor29p = 3D-MoRSE—signal 29/weighted by atomic polarizabilities	
15. WHIM descriptors (99)	G1e = (5.65215, 0)	G1e = 1st component symmetry directional WHIM index/weighted by atomic Sanderson electronegativities	r
	L3p = (2.42598, 0)	L3p = 3rd component size directional WHIM index/weighted by atomic polarizabilities	
	G2s = (2.88017, 0)	G2s = 2st component symmetry directional WHIM index/weighted by atomic electrotopological states	

Table 3 (continued)

Data set	Descriptors		Ref.
	E3s = (4.393, 0)	E3s = 3rd component accessibility directional WHIM index/weighted by atomic electrotopological states	
	Dp = (3.92512, 0)	Dp = D total accessibility index/weighted by atomic polarizabilities	
16. GETAWAY descriptors (196)	R5u+ = (9.80995, 0)	R5u+ = R maximal autocorrelation of lag 5/Unweighted	s
	R3v+ = (71.7628, 0)	R3v+ = R maximal autocorrelation of lag 3/weighted by atomic van der Waals volumes	
	R1e+ = (7.09672, 0)	R1e+ = R maximal autocorrelation of lag 1/weighted by atomic Sanderson electronegativities	
	R1p+ = (17.8103, 0)	R1p+ = R maximal autocorrelation of lag 1/weighted by atomic polarizabilities	
17. Charge descriptors (14)	Qmean = (10.7343, 0)	Qmean = Mean absolute charge (charge polarization)	t,u
	RNCG = (7.83754, 0.367559)	RNCG = Relative negative charge	
18. Molecular properties (8)	Ui = (0.965911, 0)	Ui = Unsaturation index	a,v
19. Lipophilicity (13)	$\log P_{Chem3D}$ = (3.26752, 0.304488)	$\log P_{Chem3D}$ = log P value obtained from Chem3D	55,57
	A log Ps = (1.1447, 0)	A log Ps = Lop value from A log Ps 2.1 software	
20. Quantum mechanical descriptors (17)	QC = (8.00264, 2.40012)	QC = Charge on the carbon atom attached to nitrogen atom, calculated by HyperChem ver 5.1	52

The number in parentheses indicates number of descriptors of those data set.
[a] Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, Germany, 2000.
[b] Balaban A.T. *Chem. Phys. Lett.* **1982,** *89,* 399.
[c] Kier, L.B.; Hall, L.H. Molecular Connectivity in Structure–activity Analysis. RSP-Wiley, Chichester, UK, 1986.
[d] Rucker, G.; Rucker, C. *J. Chem. Inf. Comput. Sci.* **2000,** *40,* 99.
[e] Balaban, A.T.; Balaban, T.-S., *J. Math. Chem.* **1991,** *8,* 383.
[f] Magnuson, V. R.; Harriss, D. K.; Basah, S.C. Studies in Physical and Theoretical Chemistry, Kinq, R.B., ed., Elsevier, Amsterdam, The Netherlands, 1983, pp. 178–191.
[g] Moran, P.A.P. *Biometrika,* **1950,** *37,* 17.
[h] Geary, R.C. *lncorp. Statist.* **1954,** *5,* 115.
[i] Estrada, E. *J. Chem. Inf. Comput. Sci.* **1997,** *37,* 320.
[j] Burden. F.R. *J. Chem. Inf. Comput. Sci.* **1989,** *29,* 225.
[k] Rios-Santamarina, I.; Garcia-Domenech, R.; Galvez, J.; Cortijo, J.; Santamaria, P.; Marcillo, E. *Bioorg. Med. Chem. Lett.* **1998,** *8,* 477.
[l] Lovasz, L.; Pelikan, J. *Period. Math. Hung.* **1973,** *3,* 175.
[m] Balaban, A.T; Ciubotariu, D.; Medeleanu, M. *J. Chem. Inf. Comput. Sci.* **1991,** *31,* 517.
[n] Randic, M. *J. Chem. Inf. Comp. Sci.* **1995,** *35,* 372.
[o] Robinson, D. D.; Barlow, T. W.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1997,** *37,* 939.
[p] Hemmer, M. C.; Steinhauer, V.; Gasteiger, J. *Vibrat. Spect.* **1999,** *19,* 151.
[q] Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. *J. Chem Inf. Comput. Sci.* **1996,** *36,* 1030.
[r] Todeschini, R.; Gramatica, P. *Quant. Struct. Act. Relat.* **1997,** *16,* 120.
[s] Consonni, V.; Todeschini, R.; Pavan, M. *J. Chem. Inf. Comp. Sci.* **2002,** *42,* 682.
[t] Karelson, M.; Lobanov, V. S.; Katritzky, A. R. *Chem. Rev.* **1996,** *96,* 1027.
[u] Stanton. D. T.; Jurs. P. C. *Anal. Chem.* **1990,** *62,* 2323.
[v] Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. *J. Chem. Inf. Comput. Sci.* **1989,** *29,* 163.

component accessibility directional WHIM index/weighted by atomic electrotopological states), **R3v+** (R maximal autocorrelation of lag 3/weighted by atomic van der Waals volumes), and **R1e+** (R maximal autocorrelation of lag 1/weighted by atomic Sanderson electronegativities) which are shown in bold in Table 4. The 3D-Morse descriptors (Molecule Representation of Structures based on Electron diffraction) were obtained on the basis of the molecular transformation equation used in electron diffraction.[58]

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i=1} A_i A_j \frac{\sin(sr_{ij})}{sr_{ij}}$$

$I_{(s)}$ is the scattered electron intensity, A_i, A_j is the atomic properties (e.g., total charge) of atoms i and j, resp., s is the diffraction angle, r_{ij} is the distance between atoms i and j, N is the number of atoms in a molecule.

This method allows the transformation of the 3D structure into a fixed-length code. The code is independent of translation and rotation of the molecule and can be applied to QSAR studies. The 3D-MoRSE code as a molecular transforming, illustrates large capability features for the sign of molecular structures regarding to their independent size of the molecule which allocate the structural diversity as well as certain allowed alterations can be represented in different atomic properties such as atomic number, mass, partial charge, and polarizability, which reflects the flexibility in explanation of the molecules for their biological activities. The descriptor, **Mor22m** represented the 22nd signal of the 3D-MoRSE weighted with atomic masses. It is obvious that as a composite descriptor, **Mor22m**, can encode several physicochemical and structural characteristics within a molecule and it is quite complicated to take out the descriptor information related to the fact as to which molecular aspects are

Table 4. Second and third pass FLR results of data set

Data set no.	Data		
	Descriptor name	Fuzzy number of the second pass objective function	Definition of descriptors
4	X5Av	(0, 0.893817)	Connectivity indices
			Average valence connectivity index chi-5
6	MatS5v	(0, 0)	2D-Autocorrelations
			Moran autocorrelation—lag 5/weighted by atomic van der Waals volumes
6	MatS1e	(1.7936, 0)	2D-Autocorrelations
			Moran autocorrelation—lag 1/weighted by atomic
			Sanderson electronegativities
8	Belm7	(0, 0)	BCUT descriptor
			Lowest eigenvalue n. 7 of Burden matrix/weighted by atomic masses
14	**Mor22m**	**(5.25924, 0.86743)**	**3D-MoRSE descriptor**
			3D-MoRSE—signal 22/weighted by atomic masses
14	Mor23e	(1.37054, 0)	3D-MoRSE descriptor
			3D-MoRSE—signal 23/weighted by atomic Sanderson electronegativities
14	Mor29p	(1.53668, 0)	3D-MoRSE descriptor
			3D-MoRSE—signal 29/weighted by atomic polarizabilities
15	G2s	(0, 0.608203)	WHIM descriptor
			Second component symmetry directional
			WHIM index/weighted by atomic electrotopological states
15	**E3s**	**(16.8944, 0)**	**WHIM descriptor**
			Third component accessibility directional
			WHIM index / weighted by atomic electrotopological states
16	**R3v+**	**(161.954, 0)**	**GETAWAY descriptor**
			R maximal autocorrelation of lag 3/weighted by atomic van der Waals volumes
16	**R1e+**	**(9.98291, 0)**	**GETAWAY descriptor**
			R maximal autocorrelation of lag 1/weighted by atomic
			Sanderson electronegativities
20	QC	(0, 0.25568)	Quantum mechanical descriptor
			Charge on the carbon atom attached to nitrogen atom,
			calculated by HyperChem ver 5.1

The bold indicates the descriptor chosen by the third FLR variable selection.

directly important for the drug–receptor interaction.[59] In our study, the appearance of the 3D-MoRSE signal over the 22nd stage weighted by atomic masses might indicate the above-average value distribution on the spatial arrangements of substituents within the molecular environments. Herein, the importance of the atomic masses, which appeared in the 22rd stage out of 32 dimensional space, might be taken into account for the ability of the atomic masses information should be related to the alterations in the spatial arrangements of the substitution patterns within the electron diffraction properties. It is well known that diffraction is a phenomenon which characterized by all types of waves and electrons has wave characteristics and researches on the electron diffraction showed that the angles of the diffracted electrons and the intensity of the diffracted electron beams can be used to determine the structure of the molecules through its environments. The electron distribution regarding the atomic masses of substituents might then be a factor in influencing the binding of the PCP-related compounds on the NMDA receptor site. Consequently, one of the contributions of the electron diffracting properties of the molecules might be an indication of the molecule to move into the transition state from the normal state which is made by the lowest energy pattern, and therefore, incomplete resonance to transitional states would occur sufficient to decrease the interatomic distances appreciably. We know that structures of molecules in different environments are not necessarily the same, and sometimes they are strikingly different. Therefore, the 3D-MoRSE descriptors based on electron diffraction can be valuable information for the process of molecular transform within the interface position to binding the receptor site.

The second descriptor which appeared from the third variable selection process by FLR, was **E3s** (Third component accessibility directional WHIM index/ weighted by atomic electrotopological states), a WHIM (Weighted Holistic Invariant Molecular Descriptors) descriptor which covers the relevant molecular 3D information relevant to molecular size, shape, symmetry, and more likely, atom distribution with respect to invariant reference frames.[60,61]

$$S_{jk} = \frac{w_i(q_{ij} - \overline{q}_j)(q_{ik} - \overline{q}_k)}{\sum_{i=1}^{A} w_i}$$

S_{jk} is the weighted covariance between the jth and kth atomic coordinates, A is the number of atoms, w_i is the weight of the ith atom, q_{ij} and q_{ik} represent the jth and kth coordinates ($j,k = x,y,z$) of the ith atom respectively, and $\overline{q}$ is the corresponding average value.

The corresponding average value.

Depending on the weighting system, different covariance matrices and hence different principal axes are obtained.

Generally, the WHIM descriptors provide a variety of principal axes with respect to defined atomic properties. For each weighting system, a set of statistical indices are calculated on the atoms projected onto the principal axes. Six different weighting systems can be measured; (1) the unweighted case, u ($w_i = 1$; i = 1, n, where A is the number of atoms for each compound), (2) atomic mass, m, (3) the van der Waals volume, v, (4) the Sanderson atomic electronegativity, e, (5) the atomic polarizability, p, and (6) the electrotopological state indices.

The WHIM descriptor, **E3s**, appeared in our fuzzy logic model indicating the relevant feature of electrotopological state factors which might be considered in PCP-related compounds with NMDA receptor bindings. It is quite clear that descriptors at the sight might not be considered for the 3D-molecular structure directly as some of them could be attributed to more realistic approaches, like lipophilicity descriptors, that is, $\log P$. However, most of the topological, geometrical 3D descriptors might act as molecular sub-fragments which define the importance of fragment-patterns to interacting with receptors.[62] Although three-dimensional descriptors derived from the chemical graph theory have been widely used for the modeling of physicochemical and biological properties of organic compounds, at present, little attention should be paid on their possible applications to defining the ligand–receptor interactions. However, in some cases electrotopological information implemented to give topological and/or geometrical graph features might be beneficial to account for the noncovalent interactions which could be determinant for the interaction of ligands with the receptor site. In our case, the appearance of **E3s** descriptor might indicate the hydrophobic interaction in certain degrees between ligand and receptor site would be one of the possibilities. In our model, the latter two descriptors resulted from the GETAWAY pool. The **R3v+** indicates R maximal autocorrelation of lag 3/weighted by atomic van der Waals volumes, and **R1e+** indicates R maximal autocorrelation of lag 1/weighted by atomic Sanderson electronegativities. The ranking of the descriptors in the ANFIS developed model, which is determined by the fuzzy number of center, α (the first number in the parentheses in Table 4), is in the following decreasing order: **R3v+** ($\alpha = 161.954$), **Es3** ($\alpha = 16.8944$, **R1e+** ($\alpha = 9.98291$), and **Mor22m** ($\alpha = 5.25924$). As can be seen, one of the GETAWAY descriptors (**R3v+**) is found to be a main contributor to the NMDA binding pattern of PCP-related compounds. However, like other topological and/or geometrical descriptors which encode information about size, shape, symmetry, atom distribution in several features (van der Waals, Sanderson, polarizability, and electrotopological states), it is certainly not feasible to conclude on the statement that identifies the manner of the relationship between a precise nature of a molecule and the receptor binding. On the other hand, it is quite obvious to indicate the importance of the GETAWAY descriptor in defining the PCP-compounds on the NMDA binding activity. The second GETAWAY parameter, **R1e+**, appeared in this ANFIS model to a lesser extent, but this might also indicate the role of such descriptors in this complex interaction.

The GETAWAY (GEometry, Topology, and Atom Weights AssemblY) descriptors mainly aim to attain three-dimensional molecular geometry provided by the molecular influence (H-type) and/or influence/distance (R-type) matrixes and atom relatedness by topology with chemical information by using different atomic weighting schemes (unit weights, mass, polarizability, and electronegativity) [lit G]. In our study, the appeared descriptors are derived from R-type of GETAWAYs and can be expressed as follows;

$$R_k(w) = \sum_{i=1}^{A-1} \sum \frac{\sqrt{(h_{ii})(h_{jj})}}{r_{ij}}(w_i)(w_j)\delta(k, d_{ij})$$
$$k = 1, 2, \ldots, d,$$

where $R_k(w)$ is the w-weighted kth order autocorrelation index, d_{ij} is the topological distance between atoms i and j, d is the topological diameter, δ, is the delta Dirac function.[63]

The R-GETAWAY descriptors encode the geometrical and topological information specified by both molecular influence and distance matrixes powered by special atomic properties. The selected descriptors are the result of taking into account the atomic positions and their spatial arrangements which include additional specific weights, that is, in our case van der Waals volume for **R3v+** and Sanderson electronegativities for **R1e+,** the topological distances of three (lag 3) and one (lag 1), respectively, occurred in the 3D-molecular space. It is well known that the van der Waals volume is a widely used descriptor in modeling physicochemical properties which are closely related to several biological interactions such as intestinal absorption[64] and blood–brain barrier (BBB) penetration.[65] This property is the sum of the union of the van der Waals spheres of the atoms within a molecule.

The van der Waals molecular volume (vdWV) is one of the essential characteristics for the understanding of molecular behavior since short range dispersion forces play a major role in the binding of drug molecules to receptors. The emergence of the combination of geometrical and topological properties with van der Waals volume might lead to a dependence on the volume and hydrophobicity is not completely unprecedented. The cumulative energetic contributions from van der Waals property might be large relative to those from hydrogen bondings and the interaction of the compounds with the NMDA receptor site would be the shape complementary of geometrical/topological features with vdWV. Indeed, the activity interaction of PCP-related compounds on the NMDA receptor site is most likely related to the ability of the compounds to cross over the BBB where in the uptake of a drug into the brain is a complex process. In our model, no lipophilicity-related descriptors such as $\log P$ and RM emerged from the variable selection to indicate the initial step (crossing BBB) of ligand–receptor interaction. It is known that the moderately lipophilic compounds are able to pass the BBB by passive diffusion and the hydrogen bonding properties of drugs significantly influence their particular central nervous

system uptake profiles.[66] However, the obtained GET-AWAY descriptors have limited ability to fully explain this approximation. Yet it might be considered that the relative effectiveness of R-GETAWAY descriptor weighted by vdWV at the third lag would a beneficial but weak indicator for explaining such complex systems like NMDA binding pattern of PCP-related compounds. In spite of the relative weakness of such explanation, the inclusion of van der Waals volume into the topological influence/distance indices should lead to a connection between the molecular size and the molecular shape, and to a certain extent help to elucidate the interaction.

The second GETAWAY descriptor, **R1e+**, was in the third rank of the final data set. This descriptor was also weighed by Sanderson electronegativity which indicates charge distribution within a molecule.[67] It is interesting to note that the **R1e+** should have a limited influence on the NMDA binding due to little extent (lag 1) of charge distribution inside the molecule attributed to NMDA binding of PCP compounds. The negative charge mostly

focused on the nitrogen atom of the molecule set and the importance of this property might be considered merely within the neighboring atoms of the nitrogen. Therefore, the charge distribution around the nitrogen atom, more likely depends on spatial arrangement of neighboring atoms around nitrogen.

4.2. ANFIS results

ANFIS has proven to be excellent function approximation tool. In this study, we applied to NMDA receptor binding of PCP-related compounds ANFIS that implements a first order Sugeno-style fuzzy system. The results showed that using of ANFIS network that constructed possible become good results at less outlay of the designed study. First of all, a total of 22 compounds of data were selected from the total of 38 compounds obtained in the NMDA receptor binding study for purpose of training in ANFIS (Table 5). The other validation and test data sets include eight data values each. The four descriptors obtained from third pass FLR were applied to ANFIS which yielded following

Table 5. Experimental ANFIS results for training, validating, and testing data sets for data according to random-5 selection

Compound	Data sets	Mor22m	E3s	R3v+	R1e+	Observed output	Predicted output
3	TRAINING	0.051000	0.113000	0.022000	0.105000	6.140000	6.139697
7		0.214000	0.077000	0.027000	0.112000	5.920000	5.603503
8		0.214000	0.077000	0.027000	0.112000	5.790000	5.603503
11		0.165000	0.068000	0.025000	0.118000	6.230000	6.224845
12		0.100000	0.077000	0.022000	0.155000	5.290000	5.288793
13		0.202000	0.098000	0.026000	0.106000	4.800000	4.799737
14		0.072000	0.046000	0.028000	0.107000	6.070000	6.069716
15		0.113000	0.059000	0.027000	0.122000	6.250000	6.250716
18		0.372000	0.061000	0.021000	0.123000	4.370000	4.369695
19		0.078000	0.043000	0.025000	0.120000	4.970000	4.971082
21		-0.034000	0.159000	0.021000	0.113000	4.910000	4.910119
22		-0.102000	0.139000	0.025000	0.132000	5.080000	5.079987
23		0.181000	0.058000	0.022000	0.201000	6.650000	6.649004
24		0.252000	0.051000	0.025000	0.113000	4.470000	4.470214
25		0.245000	0.017000	0.024000	0.152000	5.000000	5.000208
28		0.269000	0.055000	0.024000	0.162000	4.980000	4.980001
30		0.113000	0.119000	0.024000	0.139000	5.220000	5.220469
32		0.214000	0.077000	0.027000	0.112000	5.100000	5.603503
34		0.182000	0.067000	0.022000	0.127000	5.140000	5.143012
36		0.162000	0.055000	0.025000	0.128000	6.170000	6.171011
37		0.077000	0.074000	0.026000	0.105000	4.970000	4.970892
38		0.162000	0.127000	0.024000	0.106000	4.560000	4.560245
1	VALIDATING	0.351000	0.092000	0.019000	0.102000	7.150000	7.073497
4		0.192000	0.130000	0.025000	0.116000	5.700000	4.679674
5		0.192000	0.130000	0.025000	0.116000	5.800000	4.679674
6		0.214000	0.077000	0.027000	0.112000	6.300000	5.603503
9		0.216000	0.067000	0.025000	0.125000	5.080000	5.665701
16		0.199000	0.048000	0.025000	0.120000	6.090000	5.825874
31		0.151000	0.101000	0.025000	0.117000	6.460000	5.331493
35		0.061000	0.059000	0.023000	0.125000	3.920000	4.588726
2	TESTING	0.164000	0.071000	0.025000	0.129000	6.280000	6.359962
10		0.190000	0.054000	0.022000	0.107000	5.110000	6.597614
17		0.262000	0.062000	0.024000	0.122000	5.620000	4.546155
20		0.125000	0.054000	0.025000	0.129000	6.030000	5.916266
26		0.188000	0.044000	0.023000	0.116000	4.070000	5.649781
27		0.059000	0.077000	0.026000	0.159000	6.840000	6.631722
29		0.140000	0.087000	0.024000	0.123000	5.590000	5.690946
33		0.216000	0.067000	0.025000	0.125000	5.960000	5.665701
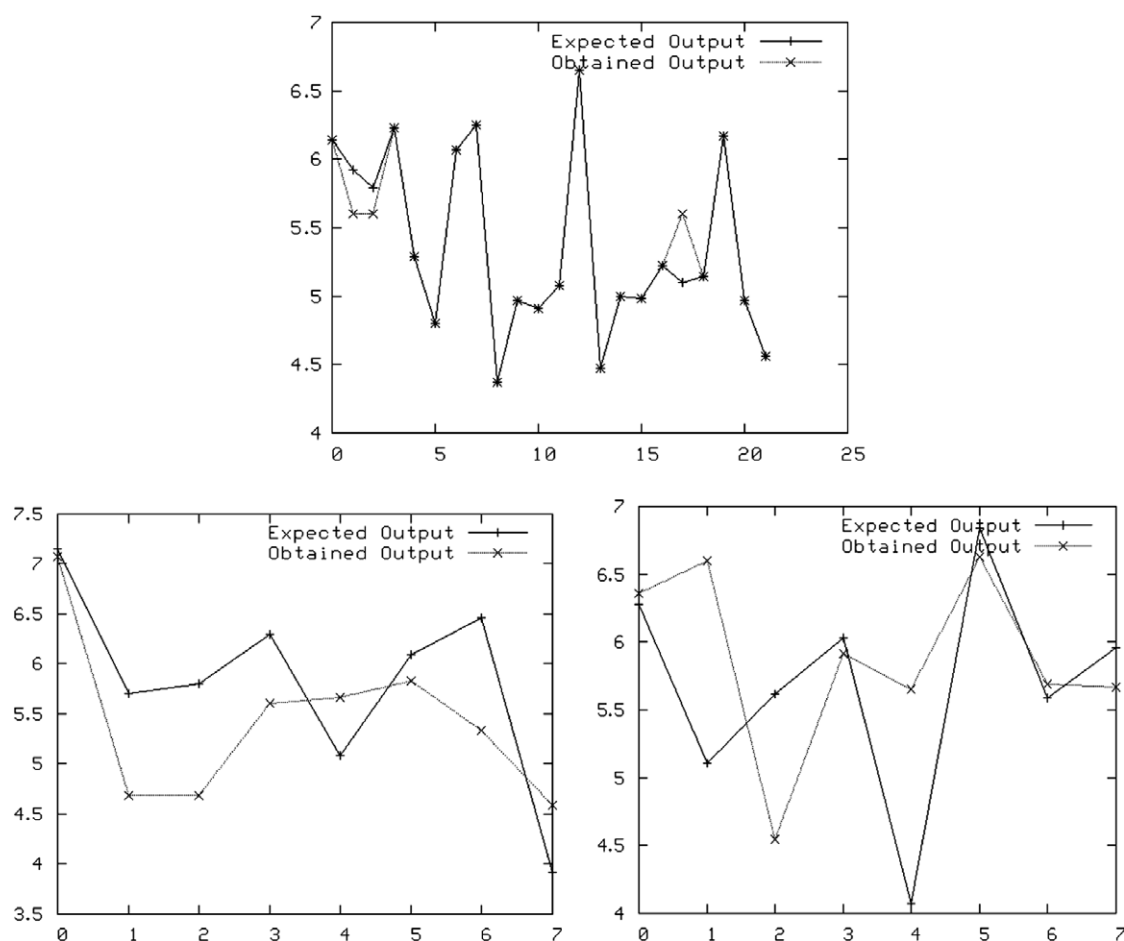
Figure 5. ANFIS results, graphics for data indicating the RMSE values obtained from training, validating, and testing data.

errors for the data sets. We applied the random-5 selection for the training, validation, and test data sets as it might be more feasible among the selection basis according to the obtained RMSE values (Table 2). ANFIS code was written in C language with a small modification in order to have predictions and graphical outputs printed, when testing the network.

Figure 5 indicates the RMS errors obtained from the ANFIS data sets of training, validating, and testing models. For the data set, ANFIS formed successful outcome in training as shown in the first row and the left illustration. This indicates that both observed (obtained) and predicted (expected) values are found in a perfect match for the training data set. Although a perfect fit has occurred in the training data set, the validating and testing data sets gave rather acceptable fits depending on the data (below left and right graphical representation in Fig. 5). Some divergences can be seen both in the validating and testing data. However, ANFIS, in general, appeared to recognize the overall system and is a good fit for training, validating, and testing models.

Table 6. Final gaussian membership function (MF) of the descriptors obtained from the data

DATA MFs			
Mor22m	a	b	c
SMALL	0.2204006281	2.0041232843	−0.0885630587
LARGE	0.1379102915	2.0055224290	0.4183146113
E3s	a	b	c
SMALL	0.0533873943	2.0007062167	−0.0057486985
LARGE	0.0384929956	2.0026339749	0.1509415046
R3v+	a	b	c
SMALL	−0.0261152019	2.0016172424	−0.0311087319
LARGE	−0.0155390714	2.0030468935	0.0513890754
R1e+	a	b	c
SMALL	0.0084724110	2.0013102459	0.0954503796
LARGE	−0.0380588099	2.0018350415	0.1927073175

1/(1+(((x+0.102000)/0.237000)**2)**2.00
1/(1+(((x-0.372000)/0.237000)**2)**2.0)

1/(1+(((x+0.0885630587)/0.2204006281)**2)**2.0041232843)
1/(1+(((x-0.4183146113)/0.1379102915)**2)**2.0055224290)

1/(1+(((x-0.017000)/0.071000)**2)**2.0)
1/(1+(((x-0.159000)/0.071000)**2)**2.0)

1/(1+(((x+0.0057486985)/0.0533873943)**2)** 2.0007062167)
1/(1+(((x-0.1509415046)/0.0384929956)**2)**2.0026339749)

1/(1+(((x-0.021000)/0.003500)**2)**2.0)
1/(1+(((x-0.028000)/0.003500)**2)**2.0)

1/(1+(((x+0.0311087319)/-0.0261152019)**2)**2.0016172424)
1/(1+(((x-0.0513890754)/-0.0155390714)**2)**2.0030468935)

1/(1+(((x-0.105000)/0.048000)**2)**2.0)
1/(1+(((x-0.201000)/0.048000)**2)**2.0))

1/(1+(((x-0.0954503796)/0.0084724110)**2)**2.0013102459)
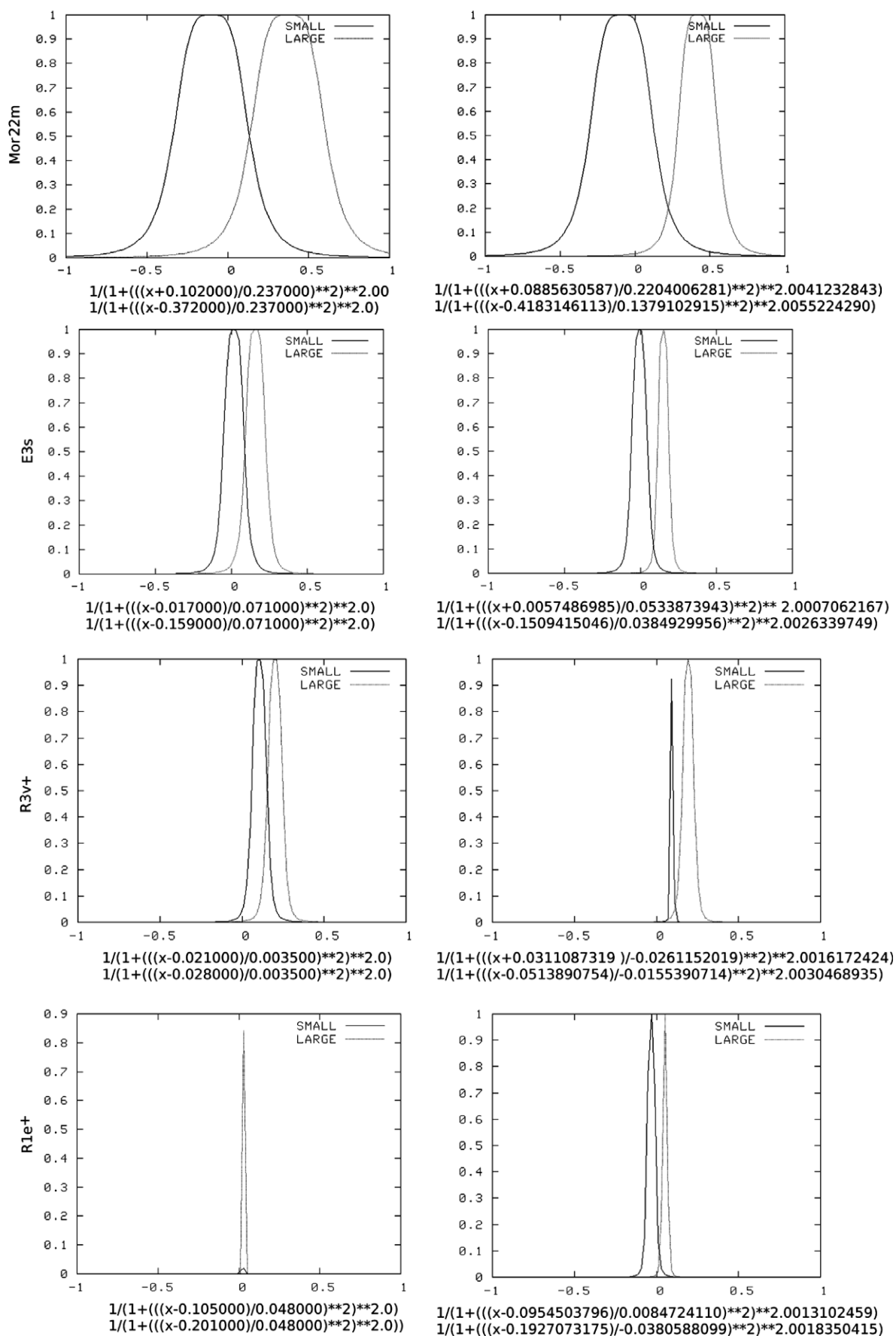1/(1+(((x-0.1927073175)/-0.0380588099)**2)**2.0018350415)

Figure 6. Graphical representation of Gaussian MF of the descriptors of the data. The bold line indicates SMALL (upper equation) and the light line is for LARGE (below equation) membership functions.

4.2.1. Membership values of data. The fuzzy part of AN-FIS is mathematically incorporated in the form of membership functions (MFs) and this study also presents a Gaussian membership function (MF) values that are related to the results obtained from ANFIS model as explained in Figure 3. The initial value for the adaptation of parametric step was applied as 0.1. The Gaussian MF values of every input parameter within the architecture were divided into two areas, that is, SMALL and LARGE (Table 6), and Figure 6 illustrates the initial and final Gaussian membership functions of four descriptors for the ANFIS model.

ANFIS dynamically constructs the initial (input) and final (output) membership functions based on the nature of data. The degree an object belongs to a fuzzy set is denoted by a membership value between 0 and 1. The left side of Figure 6 shows the initial membership functions identified by competitive learning where Gaussian functions are exploited to represent the knowledge about its landmarks. On the other hand, it is evident that interpretability of the right side graphical representation of final MFs is rather lost after the optimization procedure. The graphical representation of Gaussian MFs of the descriptors was drawn according to Eq. 1 as defined in the ANFIS Structure of Theoretical Routines Section. As can be seen from Figure 6, the initial and final MFs of the descriptors, **Mor22m**, **E3s**, and **R3v+** have, alterations in describing the SMALL and LARGE values, whereas descriptor **R1e+** was found to be unchanged in the two areas. These changes might indicate that the first three descriptors rationalize much better than the **R1e+** descriptor. Among the four descriptors selected by the third FLR, the highest coefficient belongs to **R3v+** (161.954, 0), which indicates the most relevant descriptor for the model. The interesting point appeared from Table 6 and reveals that the c-parameter of the function of the first pertinent descriptors has negative value for the SMALL MFs.

Since our goal is to obtain the simplest logical description of the data obtained from Gaussian membership

function values, rather than description of the network mapping, we are using the parameter patterns to define the descriptor behavior within the ANFIS model. The number $\mu_A(x)$ represents the grade of membership of element x in set A, with larger values denoting higher degrees of data set membership. The fuzzy variable **Mor22m** denoted by 3D-MoRSE—signal 22/weighted by atomic masses can be seen in Figure 6 to extract the very similar grade for SMALL and LARGE expression at the initial state of the Gaussian MF, whereas the final state indicates a drop in the LARGE expression which might be related to the loss of potency of this descriptor during the training process. Indeed, the lowering influence (SMALL) pattern of this descriptor remaining after final Gaussian MF was obtained. The second descriptor, **E3s**, initially started with rather high values both for SMALL (0.35) and LARGE (0.7) terms. After training, ANFIS model for the system obtained the third component accessibility directional WHIM index/weighted by atomic electrotopological states. They showed a drastic drop both in SMALL (0.08) and LARGE (0.25) eventually for the final stage. The two GETAWAY descriptors extracted from the third fuzzy selection appeared as the relevant descriptors from the FLR. However, ANFIS recognized them separately for affecting the overall system definition. It is obvious that all stages of the fuzzy inference system are affected by the choice of certain features of the descriptors. The peak (MAX) values for SMALL and LARGE MFs are 3.5×10^{-6} (S), 5.0×10^{-6} (L) for initial and 0.017 (S), 0.02 (L) for the final period, respectively. A very intense raise both in SMALL and LARGE MF values seemed to indicate the fact that GETAWAY explanation of maximal autocorrelations of lag 3, which is weighted by van der Waals volume, was adaptively adjusted by monitoring and processing period of ANFIS according to the learning environment of the model. For the last descriptor, **R1e+**, the loss of significance occurred as the initial MFs were drastically dropped from 1.0 (S), and 0.9 (L) values to 0.01 (S), and 0.02 (L) at the final phase. After the results are obtained from the initial and final Gaussian membership functions, the order of

Table 7. Consequent parameters and rules for data

Rule No.	p_0	p_1	p_2	p_3	p_4
1	11.8690313634	0.4554232859	0.0985055495	1.3399733990	3.9410266940
2	−3.9695244312	11.6109092375	0.1284307177	26.1575951939	−1.8075129291
3	27.9936530694	−1.9185486053	0.1062813277	3.2988308905	4.8787087854
4	28.9968712583	27.4890007261	−0.3474818752	30.0908822833	−0.9629215238
5	2.7788091501	1.0344223769	0.4778477722	2.8082157087	26.1504277539
6	−1.2044262528	−8.0774534387	−0.0791655116	2.9176229021	−6.3333776672
7	3.1210497236	−2.5947454497	−0.5253669260	−0.3836794254	−8.9605435128
8	5.9576149929	−12.6549714449	0.3881395559	4.2597243717	12.0949556319
9	1.7280963488	0.8810846693	0.1790299240	1.1843616502	8.9776869891
10	14.4257044062	1.8874543342	−0.0523830662	4.9295386004	−1.8235003783
11	−3.8072135331	0.4152135031	−0.1791985287	0.2177042997	−2.0377467487
12	10.0545155877	5.4414120110	0.0411947327	5.3851577124	0.5247787108
13	1.0584564975	0.0946692801	0.1194111784	0.5177997675	4.5696626182
14	0.7075355766	−0.9227225410	−0.0908287930	−0.2369460987	−5.5300983016
15	2.5339209888	0.7219252311	0.3053289877	1.3447060527	12.2426478846
16	1.7538555109	−0.7899596275	0.1287898777	0.7906551017	3.1132511511

IF *Mor22m* is A_1 **AND** *E3s* is A_2 **AND** *R3v+* is A_3 **AND** *R1e+* is A_4 **THEN** $y = p_0 + p_1$ *Mor22m* $+ p_2$ *E3s* $+ p_3$ *R3v+* $+ p_4$ *R1e+* where A_1, A_2, A_3, and A_4, can be SMALL (S) or LARGE (L).

the descriptors related to the relevance of the model obtained occurred as follows; R3v+ > Mor22m > E3s > R1e+.

4.2.2. Fuzzy rules for the data. The consequent parameters (Table 7) obtained from ANFIS indicate the relevance of existing model to adopt the explanation of the biological activity related to the chosen descriptors. These parameters (p_0–p_4) were obtained due to the first order Sugeno fuzzy model rules as the four descriptors resulted in 16 (2^4) rules (neuronal parameters at layer 4) for the data set. We are aware of the fact that such fuzzy rules have been widely used for knowledge representation purposes in neural network applications.

4.2.3. Comparison with multiple linear regression (MLR) and partial least-squares (PLS). In this study, we set as an aim to measure the predictive ability of the ANFIS model by comparison with multiple linear regression (MLR) and partial least-squares (PLS) methods. On the basis of this test and all the other information presented here, it appears that the ANFIS model described here is very superior for predicting NMDA binding of PCP-related compounds. The differences in the results of predictions done using the various models are a function of the modeling approach employed, the descriptors used, and the data set of compounds. The ANFIS model presented here obviously gives the best statistical results. A summary of the comparisons of ANFIS with multiple linear regression (MLR) and partial least-squares (PLS) is given in Table 8. The consistency of the ANFIS model as compared with MLR and PLS methods was revealed by cross validation[68] test quantified with predictive Q^2. The Q^2 values measure the goodness of the predictions of the held out cases exactly in the same way as R^2 does with the cases included in the modeling phase. But R^2 is always lower and may be even negative if the predictions are worse than just using the average value of the response. The Q^2 value should be at least 0.3–0.4 in order to assess that the model has statistically significant

Table 8. Comparison of ANFIS with multiple regression (MLR) and partial least-square (PLS) analyses

Data sets	Compound	Observed output	Predicted output		
			ANFIS	MLR	PLS
Training	3	6.140000	6.139697	5.13571	5.13623
	7	5.920000	5.603503	5.73249	5.73062
	8	5.790000	5.603503	5.73249	5.73062
	11	6.230000	6.224845	5.48547	5.48489
	12	5.290000	5.288793	5.50394	5.50668
	13	4.800000	4.799737	5.66525	5.66361
	14	6.070000	6.069716	5.43342	5.43119
	15	6.250000	6.250716	5.60042	5.59928
	18	4.370000	4.369695	5.39921	5.40004
	19	4.970000	4.971082	5.26212	5.26182
	21	4.910000	4.910119	5.22376	5.22545
	22	5.080000	5.079987	5.59845	5.59945
	23	6.650000	6.649004	5.96141	5.96612
	24	4.470000	4.470214	5.47452	5.47341
	25	5.000000	5.000208	5.57408	5.57526
	28	4.980000	4.980001	5.88867	5.89035
	30	5.220000	5.220469	5.76753	5.76866
	32	5.100000	5.603503	5.73249	5.73062
	34	5.140000	5.143012	5.30098	5.30211
	36	6.170000	6.171011	5.51306	5.51294
	37	4.970000	4.970892	5.36577	5.36437
	38	4.560000	4.560245	5.55951	5.55892
Validating	1	7.150000	7.073497	5.12915	5.12996
	4	5.700000	4.679674	5.80727	5.80664
	5	5.800000	4.679674	5.80727	5.80664
	6	6.300000	5.603503	5.73249	5.73062
	9	5.080000	5.665701	5.61769	5.61730
	16	6.090000	5.825874	5.45350	5.45287
	31	6.460000	5.331493	5.61823	5.61771
	35	3.920000	4.588726	5.17212	5.17304
Testing	2	6.280000	6.359962	5.60371	5.60367
	10	5.110000	6.597614	5.05757	5.05767
	17	5.620000	4.546155	5.53174	5.53150
	20	6.030000	5.916266	5.46666	5.46670
	26	4.070000	5.649781	5.18799	5.18807
	27	6.840000	6.631722	5.87073	5.87202
	29	5.590000	5.690946	5.49543	5.49565
	33	5.960000	5.665701	5.61769	5.61730
Q^2			0.4876	0.08760587	0.0876013
RMSE			0.546051704	0.728645781	0.728647623

prediction ability. In this study, the Q^2 values of the models are calculated by the below equation:

$$Q^2 = 1 - \frac{\sum_{i=1}^{N}\left(Y_{\text{obs}} - Y_{\text{pred}}\right)^2}{\sum_{i=1}^{N}\left(Y_{\text{obs}} - Y_{\text{mean}}\right)^2},$$

where Y_{obs} and A_{pred} are the predicted and the observed biological activities, respectively, Y_{mean} is average observed biological activity. Another validation analysis of the comparison of ANFIS with other conventional methods is RMSE (Root-Mean-Square Error) which was explained previously, as an indicator of reliability or accuracy of the models. RMSE is computed (Eq. 2) on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the stochastic nature of the model. Among the three approaches, ANFIS outperformed the other two methods significantly with Q^2 values of 0.4876 for ANFIS, 0.0876 both for MLR and PLS, and with RMSE values of 0.5461 for ANFIS, 0.7286 for MLR, and 0.7286 for PLS. The MLR and PLS calculations were done by using the Tanagra program which is free data mining program developed at University of Lyon.[69]

5. Conclusion

In this study, we investigated the possibility and effectiveness of predicting the decriptor pattern of NMDA receptor binding of PCP-related compounds with adaptive neuro-fuzzy inference system (ANFIS). In straightforwardness, there are two primary advantages for use of a fuzzy logic-based modeling algorithm when compared to conventional methods. First is that the fuzzy logic reduces the hitches of modeling and examination of complex system and second is that fuzzy logic is appropriate for integrating the chemical archetype to biological activity prediction. Therefore, neuro-fuzzy approach provides a means of training a variety of bias functions to emulate knotty multi-dimensional mapping functions and is used to forecast the pattern recognition.

Quality assortment is necessary since the set of candidate attributes is very large for a relatively small set of molecules. 'Data strip mining' then affords a method for taking out important features from descriptive data sets based on successive sensitivity analyses with a random gauge variable. Data strip mining attributes selection can be carried out with extrapolative models based on neural networks. Therefore, in our study, we applied the fuzzy linear regression (FLR) technique for selecting descriptors from each data subset followed by ANFIS approach to the chosen descriptors. The FLR gave the most promising molecular descriptors which are derived either from two- or three-dimensional molecular structure indices based on 3D-MoRSE, WHIM, and GETAWAY descriptors. These classes of descriptors seem to be naturally related to NMDA binding activity as they associate with the atomic masses, van der Waals characteristics, electrotopological states, and Sanderson electronegativities of the PCP compounds. Using the most relevant descriptors obtained from FLR, we compared the ANFIS results with multiple linear regression (MLR) and partial least-square (PLS) analyses to find out the relevance of the fuzzy logic approach in such a very complex system like NMDA receptor binding pattern of PCP-related compounds.

ANFIS routine performs the descriptors better than the multiple linear regression (MLR) and partial least-squares (PLS) analyses. The main advantage of the fuzzy approach however, is that the qualitative model concepts are explicitly available in the form of the inference rules and membership functions. As pointed out by Tessem and Davidsen[70] fuzzy set theory may prove to be a useful alternative for conventional methods to incorporate qualitative concepts in simulation models, such as the table functions now provided by simulation tools. Conditions in favor of applying fuzzy logic are: qualitative expert knowledge, dependency of input variables, nonlinearity of relationships, and the absence of mathematical models.

These results suggest that a comprehensive QSAR model of PCP compounds on NMDA receptor binding may be built by using only a few appropriate parameters, although the relevance of the identified descriptors to the continuous-scale is yet to be shown. Further work will focus on populating a database with continuous-scale of compounds which are clustered to both structural types and relative activities, and expansion of the database. New predictive FLR-ANFIS modeling can be expected to be useful in screening larger sets of compounds for their potential binding on NMDA receptors, and thus may suggest a useful order of priorities in drug design and experimental testing.

References and notes

1. Aoyama, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 905.
2. Andrea, T. A.; Kalayeh, H. *J. Med. Chem.* **1991**, *34*, 2824.
3. So, S.-S.; Richards, W. G. *J. Med. Chem.* **1992**, *35*, 3201.
4. Ajay, A. *J. Med. Chem.* **1993**, *36*, 3565.
5. Wikel, J. H.; Dow, E. R. *Bioorg. Med. Chem. Lett* **1993**, *3*, 645.
6. Gasteiger, J.; Zupan, J. *Angew. Chem., Int. Ed. Engl.* **1993**, *105*, 503.
7. Manallak, D. T.; Ellis, D. D.; Livingston, D. J. *J. Med. Chem.* **1994**, *37*, 3758.
8. So, S.-S.; Karplus, M. *J. Med. Chem.* **1996**, *39*, 1521.
9. Exner, T. E.; Brickmann, J. *J. Mol. Model* **1997**, *3*, 321.
10. Exner, T. E.; Keil, M.; Brickmann, J. Fuzzy Set Theory and Fuzzy Logic and Its Application to Molecular Recognition. In *Chemoinformatics—From Data to Knowledge.*; Gasteiger, J., Engel, T., Eds.; Wiley-VCH: Weinheim, 2003; pp 1216–1238.
11. Kubinyi, H.; Kehrhahn, O. H. *Arzneim.-Forsch* **1978**, *28*, 598.

4282

E. Buyukbingol et al. / Bioorg. Med. Chem. 15 (2007) 4265–4282

12. Fernández, M.; Caballero J. Bioorg. Med. Chem. 2006, 14, 280.
13. Valkova, I.; Vracko, M.; Basak, S. C. Anal. Chim. Acta 2004, 509, 179.
14. Weekes, D.; Fogel, G. B. Biosystems 2003, 72, 149.
15. López-Rodríguez, M. L.; JoséMorcillo, M.; Fernández, E.; Rosado, M. L.; Orensanz, L.; Beneytez, M. E.; Manzanares, J.; Fuentes, J. A.; Schaper, K.-J. Bio. Med. Chem. Lett. 1999, 9, 1679.
16. Schneider, G.; Wrede, P. Prog. Biophys. Mol. Biol. 1998, 70, 175.
17. Ghoshal, N.; Achari, B.; Ghoshal, T. K. Bioorg. Med. Chem. Lett. 1997, 7, 877.
18. Tetko, I. V.; Luik, A. I.; Poda, G. I. J. Med. Chem. 1993, 36, 811.
19. Manallack, D. T.; Ellis, D. D.; Livingstone, D. J. J. Med. Chem. 1994, 37, 3758.
20. Gini, G.; Lorenzini, M.; Benfenati, E.; Grasso, P.; Bruschi, M. J. Chem. Inf. Comput. Sci. 1999, 39, 1076.
21. Jaen-Oltra, J.; Salabert-Salvador, M. T.; Garcia-March, F. J.; Perez-Gimenez, F.; Tomas-Vert, F. J. Med. Chem. 2000, 43, 1143.
22. So, S.-S.; Karplus, M. J. Med. Chem. 1996, 39, 5246.
23. Schneider, G.; Coassolo, P.; Lave, T. J. Med. Chem. 1999, 42, 5072.
24. Sutter, J. M.; Dixon, S. L.; Jurs, P. C. J. Chem. Inf. Comput. Sci. 1995, 35, 77.
25. Pullan, W. J. J. Chem. Inf. Comput. Sci. 1997, 37, 1189.
26. Izrailev, S.; Agrafiotis, D. K. SAR QSAR Environ. Res. 2002, 13, 417.
27. Agrafiotis, D. K.; Cedeno, W. J. Med. Chem. 2002, 45, 1098.
28. Steyerberg, E. W.; Eijkemans, M. J. C.; Habbema, J. D. F. J. Clin. Epidemiol. 1999, 52, 935.
29. Hoskuldsson, Agnar. Chemom. Intell. Lab. Syst. 2001, 55, 23.
30. Hasegawa, K.; Kimura, T.; Funatsu, K. J. Chem. Inf. Comput. Sci. 1999, 39, 112.
31. Jang, J.-S. R. IEEE Trans. on Neural Networks 1992, 3, 714.
32. Whitley, D. C.; Ford, M. G.; Livingstone, D. J. J. Chem. Inf. Comp. Sci. 2000, 40, 1160.
33. Livingstone, D. J.; Rahr, E. Quant. Struct-Act. Relat. 1989, 8, 103.
34. Elhallaoui, M.; Laguerre, M.; Carpy, A.; Ouazzani, F. C. J. Mol. Model 2002, 8, 65.
35. Monaghan, D. T.; Bridges, R. J.; Cotman, C. W. Annu. Rev. Pharmacol. Toxicol. 1989, 29, 365.
36. Johnson, R. L.; Koerner, J. F. J. Med. Chem. 1988, 31, 2057.
37. Bräuner-Osborne, H.; Egebjerg, J.; Nielsen, B.; Madsen, U.; Krogsgaard-Larsen, P. J. Med. Chem. 2000, 43, 2609.
38. Shinozaki, H. Prog. Neurobiol. 1988, 30, 399.
39. Hargreaves, R. J.; Hill, R. G.; Iversen, L. L. Acta Neurochir Suppl. 1994, 60, 15.
40. Parsons, C. G.; Danysz, W.; Quack, G. Drug News Perspect 1998, 11, 523.
41. Jang, J.-S. R. IEEE Trans. on Systems, Man and Cybernetics 1993, 23, 665.
42. Loukas, Y. L. J. Med. Chem. 2001, 44, 2772.
43. Sarimveis, H.; Alexandridis, A.; Tsekouras, G.; Bafas, G. A. Ind. Eng. Chem. Res. 2002, 41, 751–759.
44. Terano, T.; Asai, K.; Sugeno, M. Fuzzy Systems Theory and Its Applications; Academic Press: San Diego, 1992.
45. Frattale-Mascioli, F. M.; Martinelli, G. A. Signal Process. 1998, 64, 347–358.
46. Tanaka, H.; Uejima, S.; Asai, K. IEEE Trans. Systems Man and Cybernetics 1982, 12, 903.
47. Sakawa, M.; Yano, H. Fuzzy Sets and Systems 1992, 47, 173.
48. Jang, J.-S. R. Neuro-Fuzzy and Soft Computing; Prentice-Hall: New Jersey, 1997.
49. Jang, J.-S. R.; Sun, C.-T. Proc. IEEE 1995, 83, 378.
50. Sugeno, M.; Kang, G. T. Fuzzy Sets and Systems 1988, 28, 15.
51. Thurkauf, A.; De Costa, B.; Yamaguchi; Mattson, M. V.; Jacobson, A. E.; Rice, K. C.; Rogawski, M. A. J. Med. Chem. 1990, 33, 1452.
52. HyperChem Version 5.1, www.hyper.com.
53. Virtual Computational Chemistry Laboratory, http://146.107.217.178/lab/edragon/index.html.
54. Todeschini, R.; Consonni, V. Dragon, Rel. 2.1. for Windows; Milano, Italy.
55. ChemOffice Pro ver.5. CambridgeSoft, 2001.
56. Tetko, I. V.; Tanchuk, V. Yu.; Kasheva, T. N.; Villa, A. E. P. J. Chem. Inf. Comput. Sci. 2001, 41, 246.
57. Tetko, I. V.; Poda, G. J. Med. Chem. 2004, 47, 5601.
58. Schuur, J. H.; Selzer, P.; Gasteiger, J. J. Chem. Inf. Comput. Sci. 1996, 36, 334.
59. Yap, C. W.; Chen, Y. Z. J. Chem. Inf. Model. 2005, 45, 982–992.
60. Todeschini, R.; Consonni, V. Handbook of Molecular Descriptors; Wiley-VCH: Weinheim, 2000.
61. Table 3—literature.
62. Klopman, G.; Buyukbingol, E. Mol. Pharm. 1988, 34, 852.
63. Consonni, V.; Todeschini, R.; Pavan, M. J. Chem. Inf. Comput. Sci. 2002, 42, 682.
64. Abraham, M. H.; Chadha, H. S.; Martins, F.; Mitchell, R. C.; Bradbury, M. W.; Gratton, J. A. Pest. Sci. 1999, 55, 78.
65. Zhao, Y. H.; Abraham, M. H.; Le, J.; Hersey, A.; Luscombe, C. N.; Beck, G.; Sherborne, B.; Cooper, I. Pharm. Res. 2002, 19, 1444.
66. Begley, D. J. J. Pharm. Pharmacol. 1996, 48, 136.
67. Fedorowicz, A.; Zheng, L.; Singh, H.; Demchuk, E. Int. J. Mol. Sci. 2004, 5, 56.
68. Fernandez, M.; Caballero, J. Bioorg. Med. Chem. 2006, 14, 280.
69. http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html.
70. Tessem, B.; Davidsen, P. I. System Dyn. Rev. 1994, 10, 49.