

Autocorrelation Method Adapted To Generate New Atomic Environments: Application for the Prediction of 13-C Chemical Shifts of Alkanes

M. Nohair* and D. Zakarya*

UFR of Applied Chemistry, Department of Chemistry, Faculty of Sciences and Techniques, B.P 146,
20450 Mohammadia, Morocco

A. Berrada

Faculty of Sciences, Rabat, Morocco

Received December 7, 2000

The concept of the multifunctional autocorrelation method governing global description of molecules was changed in order to take into account the structural environment of each atom. New atomic environments are generated as possible descriptors in QSARs and can be useful for database characterization. The principles of this approach are widely explained through a case study dealing with the design of a model allowing the simulation of the carbon-13 nuclear magnetic spectra for alkanes. Carbon atoms in alkanes are described by using as structural descriptors a vector corresponding to only four components vectors of the multifunctional autocorrelation method. The statistical method used for deriving the model was a classical three-layer feedforward neural network trained by the back-propagation algorithm and multilinear regression (MLR). The predictive ability of the ANN model was tested by -10% -out(L10%O) cross-validation method, demonstrating the superior quality of the neural model. The established model allows us the prediction of the 13-C chemical shifts with success because since all types of carbons are taken into account without distinction of connectivity. The neural network possessed a 4:7:1 architecture with a sigmoid shape as a activation function. The model produced a cross-validation standard coefficient r between δ_{exp} and δ_{calc} about 0.99, while the cross-validation standard s and the mean error are equal to 0.87 and 0.60 ppm, respectively.

INTRODUCTION

Since 1980, considerable effort has been directed toward introducing the autocorrelation method^{1,2} into the field of structure–activity relationships. The goal is to represent the structure of a molecule by a set of parameters, which are unique, easy to compute, and structurally significant. In the autocorrelation method, an organic molecule is represented by means of a graph where the atoms are displayed by nodes or vertices and bonds are depicted by edges. In this graph, the distance between two nodes is defined as the smallest number of edges between them. Considering that a property of a molecule can be calculated from atomic contribution (ACi) describing this molecule, the classical autocorrelation algorithm allows for computation of all the products (ACi*ACi; ACi*ACj.....) corresponding to the different smallest internodal distances (i.e., 0 to n) in the molecular graph. The sum of these products for the same distance in the graph gives a component of the autocorrelation vector for the selected property. The general relation used to calculate the autocorrelation component is defined below (eq 1):

$$P_k = \sum_{i=0}^n (f(i)f(j))^x$$

P_k is the autocorrelation component corresponding to the topological distance of k bonds (smallest number of bonds

between i and j) to the specific property $f(i)$. Atomic contribution $f(i)$ depends on the chosen property considered under study. These properties can be based, as a example, on the van der Waals volume (V) and surface (S) to account for the size and the shape of the molecule, connectivity (number of non-hydrogen neighbors or vertex degree of the atom i) or electronegativity and charge³ to account for the electronic aspect.

n represents the number of neighborhoods of carbon atom i in the molecule. But we will shown that only four components ($k = 0$ up to 3) are enough for a correct description of atomic environment.

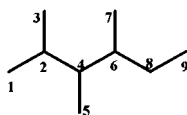
The value of x can be 1 in the classical method. But to give physical meaning to the vector component, x was taken as 0.5. For example, estimation of roughly additive molecular properties is difficult because the component zero of the autocorrelation vector is not the sum of atomic contributions. Zakarya and others⁴ have shown this in the estimation of the molar volumes of allkanes and oxygened compounds.

The aim of this paper is to investigate the ability of the autocorrelation method to generate local environment of atoms in a molecule, that is taken as descriptors in designing a QSPR model for the simulation of the 13-C chemical shifts in organic molecules. Molecules of alkanes considered in this study are nonpolar and do not contain a functional group, so a number of complexities that arise with more polar compounds are avoided. This chemical family offers widely varied carbon frameworks, and so they constitute a good example to examine whether branching affects (a source of difficulty).

* Corresponding authors e-mail: nohairm@hotmail.com (Nohair) and zakaryadriss@hotmail.com (Zakarya).

Table 1. Contributions of Atoms to Some Molecular Properties

property	no. of atoms in molecule (Figure 1)								
	1	2	3	4	5	6	7	8	9
connectivity	1	3	1	3	1	3	1	2	1
van der Waals volume, V (cm ³ mol ⁻¹)	13.67	6.78	13.67	6.78	13.67	6.78	13.67	10.23	13.67

**Figure 1.** A hydrogen-depleted molecular graph corresponding to skeleton of 2,3,4-trimethylhexane.

There are several ways to design QSPR models in the modeling of the NMR chemical shifts. Some of them are based on empirical models relating structural descriptors to ¹³C shifts in additive relationships or linear correlation. More recently, by introducing neural network in this field, the ¹³C chemical shifts have been predicted many times,⁵⁻¹² and the most accurate QSPR model (family of alkanes studied is that used by Lindeman and Adams¹³) is taken by J. P. Doucet and al.¹⁴ Their model was built by using descriptors derived from DARC system.¹⁵

Other models combining the electronic state of the bonded atom within the molecule with its topological nature in the context of the whole molecular skeleton were also used. Each atom is coded and represented by an index, called the electrotopological state index^{16,17}(E-state for short) which encodes the intrinsic electronic state of the atom as perturbed by the electronic influence of all other atoms in the molecule within the context of the topological character of the molecule.

METHOD

To describe a local environment of atoms in molecule, we define, like P_k , a new component P_{ik} by means of the following formula (eq 2):

$$P_{ik} = \sum_{k=0}^n (f(i)f(j))^x$$

We compute P_{ik} by fixing the atom i , P_{ik} is defined as the sum of $f(i)*f(j)$ of all chemical bonds existing between all pairs of carbon atoms i (fixed atom) and j separated by a topological distance equal to k .

Example of Computation of P_{ik} Components. 2,3,4-Trimethylhexane (Figure 1) was chosen to illustrate the calculation of components of autocorrelation vectors. The properties of different atoms present in the molecule (CH₃, CH₂, CH, and C) are given in Table 1.

We consider only the carbon atom number 4 (Figure 1) to compute its components by using descriptors based on connectivity and van der Waals volume as atomic properties

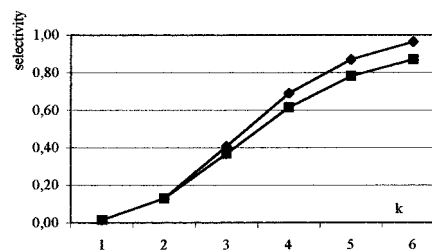
$f(i)$ = connectivity

(number of non-hydrogen neighbors of the atom)

P_{ik} is defined as C_{ik} .

*For $k=1$,

$$C_{41} = [2(3.3)^{1/2} + (3.1)^{1/2}] = 7.37$$

**Figure 2.** Selectivity vs k for both components computed by eqs 2 and 3.

This corresponds to three pairs of atoms (4, 2), (4, 6), and (4, 5) with a topological distance equal to 1.

*For $k=2$,

$$C_{42} = [3(3*1)^{1/2} + (3*2)^{1/2}] = 7.64$$

*Etc.

$f(i)$ = van der Waals volume

P_{ik} is defined as V_{ik} .

*For $k=1$,

$$V_{41} = [2(6.76*6.76)^{1/2} + (6.76*13.67)^{1/2}] = 23.18$$

*For $k=2$,

$$V_{42} = [3(6.76*13.67)^{1/2} + (6.76 \times 10.23)^{1/2}] = 37.15$$

*Etc.

Although components were used successfully as descriptors variables in many studies, they cannot describe the chemical structure with high chromatism because only atomic property of atoms i and j are considered in each interatomic distance $d(i,j)$. This drawback was avoided by the introduction of a modification based on the concept of chromatism developed in the DARC system by including the property of the intermediate atom for topological distance greater than two, as shown by the following modified component (eq 3):

$$P_{ik} = \sum_{k=0}^n [f(i)(\sum f(int_{ij}))f(j)]^{1/n}$$

int_{ij} represents the intermediate carbon atom between i and j atoms and n the number of atoms in the selected fragment.

The selectivity¹⁸ (number of different values/number of total compounds) is tested for the whole set. We found that it increases with the increase of k and it is sufficiently high (>80%) for $k=5$. Since we were interested in comparing the selectivity power, results given in Figure 2 show that it could be improved by using as structural descriptors the vector of components using the eq 2.

We have mentioned that the modification introduced in the autocorrelation method is based on the DARC system. In this work we try to give a new method for describing an

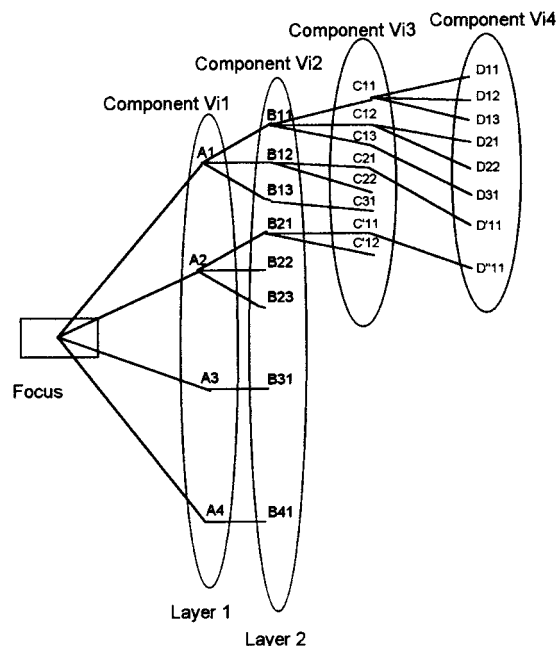


Figure 3. Trace of the studied population (208 differing environments in 65 compounds).

atomic environment by combining the autocorrelation method and the DARC system. So it was of interest to illustrate how to obtain a component of the modified autocorrelation method from the concept of the DARC system.

In the DARC system, the environment of the atom is concentrically described in terms of discrete and ordered atom sites which are given individual contributions to the considered property (Here are the chemical shifts.). The environment of the resonating carbon is described in concentric layers A, B, C, and D, then we constitute the hyperstructure of the population studied by superposing all topological description of carbon atoms. Figure 3 represents the trace or the hyperstructure of a population of alkanes containing 208 differing environments in 65 compounds. The sites in the environment are labeled by indices which specify the order induced in this concentric organization. For example, on the second layer B_{ij}, the first index refers to the position of the preceding layer (A_i) to which they are bonded. The second index j corresponds to the order for sites linked to the same seed site (B₁₂ is the second B site linked to the first A site). According to the hyperstructure, schematized in the graph of the Figure 3, it appears that each atomic environments considered can be represented by a string of 27 (ordered) binary values (1 if the corresponding position is occupied, 0 otherwise) characterizing univocally each individual environment. An application example of these topological descriptors is given in Figure 4.

Starting from this description, the procedure of computing the components of modified autocorrelation method is illustrated as follows: $V_{i1} = \sum A_{ij}$, $V_{i2} = \sum B_{ij}$, $V_{i3} = \sum C_{ij}$, and $V_{i4} = \sum D_{ij}$... etc., and we add the component V_{i0} encoding the intrinsic environment of the considered atom. Each vector is computed by using eq 3. It is important to indicate that the number of descriptors constituting the inputs values both of linear regression and neural network was reduced to six. We will show later in this paper that four components (V_{i0} – V_{i3}) are sufficient to obtain good results.

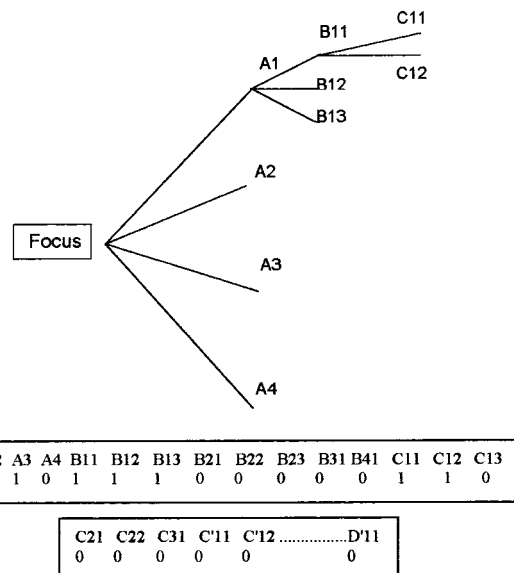


Figure 4. Ordered topological descriptor.

EXPERIMENTAL SECTION

The data set (molecules of alkanes with two up to nine carbon atoms) contains 243 values of chemical shifts ranging from 6.5 ppm in ethane up to 53 ppm in 2,2-dimethyl-3-ethylpentane. We consider all molecules studied by Lindeman and Adams,¹¹ and we have added four molecules with up to four atoms. All values are checked by consulting the database ACD labs,¹⁹ and we removed all molecules having double values.

The size of the data set is enough of a correct investigation of the ¹³C NMR shift of the family of alkanes. All types of carbons are taken into account without distinction of connectivity. The chemical shifts do not change significantly from the maximal values of the data set.

In previous work, the same data set was investigated by J. P. Doucet and al.¹² The study included a topological description of the environment of the resonating carbon according to the concept of the DARC system¹³ and a computational neural network.

To simplify the computation of the components, molecules were coded by means of the Smiles system²⁰ and storage as input files. The computer program used to compute the components represents an algorithm for the construction of the connectivity matrix of any molecule from its Smiles code. The neural network was performed from the QUIKNET package.

Linear Model. A linear modeling method was first employed to investigate the behavior of each component. Two different LRA calculations have been carried out. While the first calculation was based on the components from eq 1, the second calculation included components generated from eq 2. To further validate stability of a model, the whole data set was decomposed to a training set containing 203 different environments for the adaptation process of the model and a testing set used for the verification of the adapted model. The obtained results of linear regression are summarized in Table 3.

According to Gore's method, V_{i0} contributes negatively (50%). Both V_{i1} and V_{i2} contribute with 20% for each one, V_{i3} has a minor contribution, as expected for the change in

Table 2. Statistical Analysis Related to the Studied Family of Alkanes

number of molecules	65
number de \neq environments	243
primary carbons	93 δ (6.50 à 26.37 ppm)
secondary carbons	98 δ (16.30 à 56.50 ppm)
tertiary carbons	29 δ (24.60 à 52.10 ppm)
quaternary carbons	23 δ (28.00 à 37.30 ppm)
δ minimal	6.5 ppm
δ maximal	53 ppm
δ average	26.37 ppm

Table 3. Results of Linear Regression

type of eq	training set		testing set	
	r^2	s	r^2	s
eq 2	0.97	1.62	0.98	1.51
eq 3	0.98	1.30	0.98	1.26

chemical shift due to substitution of an hydrogen by a methyl group at the α to ϵ carbon atoms.²¹ Substitution of a hydrogen by a carbon causes a deshielding in the α and β positions (positive contribution for V_{1i} and V_{12}), while a shielding increases with increasing of the electron charge density of the considered carbon (negative contribution for V_{10}). The chemical shift appears largely determined by V_{10} perturbed by the influence of all other atoms. Valence-bond calculations of 13-C chemical shift parameters of alkanes denote that the main contribution for the 13-chemical shift is represented by the paramagnetic screening,²¹ and then the obtained model is considered as a function between the structural environment of this property for each carbon and the chemical shift 13-C. Usually, any interaction must decrease with the increase of the distance between the interacting particles, so we define a new model in which both components V_{4i} and V_{15} will not be considered, and we noticed that the quality of regression was not significantly changed.

The error associated with the prediction set for the second model represents a significant improvement but leads to 25% shifts farther than 1 from the experimental values. In addition to that, three carbon atoms present bad prediction ($s > 4$ ppm).

Computational Neural Network. The main advantage of using neural networks in QSPR models is their capacity to offer a nonlinear mapping of the structural descriptors to physicochemical property. It is interesting to see how much is gained with ANNs in the prediction power of chemical shift in alkanes. The computational neural network used in this study was three-layer (input-hidden-output), fully connected, feed-forward networks. A sigmoid transformation of the descriptor values is performed in the first layer, the input layer. The output layer consisted of one neurone and represents the 13-C chemical shift. Although there are neither theoretical nor empirical rules to determine the number of hidden layers nor the number of neurones in this layer, one layer seems to be sufficient in most chemical applications of ANNs. Each neurone of a given layer (except the input one) takes a value Y calculated by using the following transfer function (eq 4)

$$O_i = 1/(1 + \exp(-(\sum W_{ij}O_j + \theta\theta_j)))$$

where O_i and O_j are the outputs of neurones i and j , respectively, W_{ij} is the weight connecting neurones i and j ,

and θ_j is the bias of neurone j . The weights of connections between the neurones were initially assigned with random values uniformly. Training was initiated and followed by examining the RMS error (RMS stands for the root mean error, that is the square root of the overage residual) for the total set and for both the training and the test set as well. Training was stopped when there was not further improvement in the test set RMS error. We also computed the correlation coefficient between the observed and predicted values.

To establish the optimal size of the hidden layer (the number of the neurones), a first application uses the whole of components V_{ik} ($k = 0-5$) as input values by applying a cross-validation technique. This was carried out iteratively. We started from one neurone in the hidden layer, the statistical indices of the correlation between experimental and predicted chemical shift for the whole data set (243 chemical shift) improves with the increase of their number, the optimal number of neurones of the hidden layer was found seven with a number of iterations of 2000. Any improvement was observed when this number increases over seven. The modeling power of the ANN model can be improved by using networks with a higher number of hidden layer, but in QSPR studies it is important to take into consideration that MLF network are universal approximators, which are capable of arbitrarily accurate approximations to arbitrary mappings, provided a sufficiently number of hidden units is available. In a recent work,^{21,22} based on empirical observations, it has been proposed to use for QSPR studies only network with ρ parameter greater than 2 (ρ is the ration of the number of patterns in the training to the number of connections), to ensure that network can give reliable prediction. The number with seven neurones in the hidden layer has 57 connections, and a ρ value greater than 4 ($\rho = 4.26$).

To compare the performance of the ANN model with the statistical results of the MLR equation, we have used the correlation coefficient r , the standard deviation s , and the mean residual $mres$ of the linear correlation between experimental and predicted chemical shifts

$$s^2 = \frac{\sum (\delta_{obs} - \delta_{calc})^2}{n}$$

$$R^2 = 1 - \frac{\sum (\delta_{obs} - \delta_{calc})^2}{\sum (\delta_{obs} - \delta_{MEAN})^2}$$

$$mres = \sum \frac{|\delta_{exp} - \delta_{calc}|_i}{n}$$

δ_{MEAN} is computed from the calculated values, i.e., $\delta_{13-Cexp} = a*\delta_{13-Cpred} + b$.

For prediction, the network is trained on a learning set comprising only a part of the total set previously used in linear model (203 shifts for the training and 40 for the test set). The network is able to calculate the chemical shift for the data test with 0.66 ppm. The correlation coefficient between observed and calculated values is 0.99. Only six carbon atoms present a deviation more than 1 but does not exceed 1.4 ppm.

Table 4. Comparison between Results of LRA Model and ANN Model

analysis	LRA		ANN (5-7-1)	
	training set	testing set	training set	testing set
r^2	97	98	99	99
s	1.62	1.51	0.57	0.71
no. of outlier, ppm				
$3 \leq \delta_{\text{calc}}$		3		0
$2 \leq \delta_{\text{calc}} < 3$		2		0
$1 \leq \delta_{\text{calc}} < 2$		11		6

Table 5. Percentage of Prediction in the Interval [0–0.5 ppm] and [0–1 ppm] by Removing Respectively Components V_{i4} , V_{i5} and V_{i4} and V_{i5}

descriptor removed	SD (r^2 , rms)	percentage of deviation	
		[0–0.5 ppm]	[0–1 ppm]
none	(0.99; 0.69)	71%	92%
V_{i5}	(0.99; 0.69)	60%	85%
V_{i4} and V_{i5}	(0.99; 0.72)	56%	85%

Comparing results of neural network and linear model, we see that the neural network calculation gives better results than linear regression. The standard error is slightly smaller than it is for linear model. The obtained results are summarized in Table 4.

Up to now, the architecture of neural network includes six parameters as inputs, but we have shown in LRA that only four descriptors V_{i0} ($i = 1-3$) give the same precision in the prediction of the testing data. Let us evaluate the effect of this new description in ANNs. After omission of one descriptor, we calculate the error for each of the 40 chemical shifts and classify them into a discrete category (here in [0–0.5 ppm], [0–1 ppm]). The standard error and the percentage of the prediction in the considered precision interval are presented in Table 5. In the above results of Table 4, P_{ik} components up to $k = 3$ represent factors of prime importance in the modeling of the 13-C chemical shift in alkanes.

The goal of a QSPR study is to develop a model of the investigated phenomena that can give reliable predictions for new patterns that were not used in the calibration of the mathematical model, so it is necessary to estimate the prediction capabilities of the ANNs with a leave-10%-out cross-validation method. In the leave-10%-out technique 10% of the patterns are selected and extracted from the entire set. Then the ANN model is calibrated with a learning set consisting of the remaining 90% of the data, and the neural model obtained is used to predict patterns in the prediction set. This procedure is repeated 10 times until all patterns are selected in a prediction set once and only once. A linear regression between experimental and predicted values leads to the following results:

- The cross-validation standard coefficient r between δ_{exp} and δ_{calc} is 0.99, while the cross-validation standard s and the mean error are equal to 0.87 and 0.60 ppm, respectively.
- The ANN model gives excellent prediction results, with only two cases in which the neural model offers a computed value greater than 3 ppm (Table 6).

CONCLUSION

As a conclusion it seems to us from this study that the modified autocorrelation method and neural networks are

Table 6. Statistical Results, Outliers, and Their Residuals for ANNs with 4–7–1 Architecture for the Prediction by the Leave-10%-Out Cross-Validation Method

analysis	
no. of patterns	243
range	6.50–56.50
mean	26.40
SD, s	0.87
mean error	0.60
no. of outliers, ppm	3
$3 \leq \delta_{\text{calc}} < 4$	5
$2 \leq \delta_{\text{calc}} < 3$	37
$1 \leq \delta_{\text{calc}} < 2$	56
$0.5 \leq \delta_{\text{calc}} < 1$	142
$\delta_{\text{calc}} < 0.5$	$\delta_{\text{calc}} < 0.5$

efficient to estimate a 13-C chemical shift with small topological parameters (four components of autocorrelation method). Efficiency of the topological descriptors generated from the multifunctional autocorrelation method is not well demonstrated here in comparison with other models. However, our aim is the only illustration as to how we have adapted it to generate new atomic environments and useful descriptors for description of environments of specific carbon atoms in alkanes in QSARs or QSPRs. Linear regression analysis and neural network calculations have been useful in developing mathematical models that predict 13-C NMR chemical shifts in alkanes from components generated by using autocorrelation method. The neural model provides better results than the MLR model, demonstrating that there exists a nonlinear relationship between the structural descriptors and the investigated 13-C NMR chemical shift.

The success of this method depends on the correct description brought by each component when it is modified by taking into account all carbon atom existing in a considered fragment with a topological greater than two.

ACKNOWLEDGMENT

M. Nohair and D. Zakarya thank the ITODYS laboratory (Paris-Jussieu) for the using of the database ACD and the QUIKNET package.

REFERENCES AND NOTES

- (1) Moreau, G.; Broto, P. the autocorrelation of a topological structure: A new molecular descriptor. *Nouv. J. Chim.* **1980**, *4*, 359–360.
- (2) Moreau, G.; Broto, P. Autocorrelation of molecular structures, application to SAR studies. *Nouv. J. Chim.* **1980**, *4*, 757–764.
- (3) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (4) Zakarya, D.; Tiyal, F.; Chastrette, M. Use of the multifunctional autocorrelation method to estimate molar volumes of alkanes and oxygenated compounds. Comparison between components of autocorrelation vectors and topological indices. *J. Phys. Org. Chem.* **1993**, *6*, 574–582.
- (5) Ivanciuc, O.; Rabine, J. P.; Cabrol-Bass, D. 13-C NMR Chemical shift sum prediction for alkanes using neural networks. *Comput. Chem.* **1997**, *21*(6), 437–443.
- (6) Panaye, A.; Doucet, J. P.; Fan, B. T.; Feuillebois, E.; Rahali El Azzouzi, S. Artificial neural network simulation of 13-C NMR shifts for methyl substituted cyclohexanes. *Chem. Intell. Labor. Syst.* **1994**, *24*, 129–135.
- (7) Ivanciuc, O.; Rabine, J. P.; Cabrol-Bass, D.; Panaye, A.; Doucet, J. P. 13-C NMR Chemical Shift Prediction of sp^2 Carbon Atoms in Acyclic Alkenes Using Neural Networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 644–653.
- (8) Ivanciuc, O.; Rabine, J. P.; Cabrol-Bass, D.; Panaye, A.; Doucet, J. P. 13-C NMR Chemical Shift Prediction of the sp^3 Carbon Atoms in the a Position relative to the Double Bond in Acyclic Alkenes. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 587–598.

- (9) Svozil, D.; Pospichal, J.; Kvasnicka, V. Neural Network Prediction of Carbon-13 NMR chemical shifts of Alkanes. *J. Chem. Inf. Comput. Sci.* **1995**, 35, 924–928.
- (10) Kvasnicka, V. An application of Neural Networks in Chemistry. Prediction of 13-C NMR Chemical shifts. *J. Math. Chem.* **1991**, 6, 63–76.
- (11) Meiler, J.; Meusinger, R.; Will, M. Fast determination of C-13 NMR chemical shifts using artificial neural networks. *J. Chem. Inf. Comput. Sci.* **2000**, 40(5), 1169–1176.
- (12) Le Bret, C. A general C-13 NMR spectrum predictor using mining data techniques. *SAR QSAR Environ. Res.* **2000**, 211, 3–4.
- (13) Lindeman, L. P.; Adams, J. Q. Carbon-13 Nuclear Magnetic Resonance. II. Chemical shifts for the Paraffins through C₉. *Anal. Chem.* **1971**, 43, 1245–1252.
- (14) Doucet, J. P.; Panaye, A.; Feuilleaubois, E.; Ladd, P. Neural Networks and 13-C Shift Prediction. *J. Chem. Inf. Comput. Sci.* **1993**, 33, 3, 320–324.
- (15) Dubois, J. E.; Mercier, C.; Panaye, A. Darc System and computer aided design. *Acta Pharm. Jugosl.* **1986**, 36, 135–169.
- (16) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Comput. Sci.* **1995**, 35, 1039–1045.
- (17) Hall, L. H.; Kier, L. B. An Electrotopological State Index for Atoms in Molecules. *J. Math. Chem.* **1991**, 7, 229–241.
- (18) Razinger, M.; Chrétien, J. R.; Dubois, J. E. *J. Chem. Inf. Comput. Sci.* **1985**, 25, 23.
- (19) Weininger, D. SMILES, a chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, 28, 31–36.
- (20) www.acdlabs.com.
- (21) Breitmaier, E.; Voelter, W. *Carbon-13 NMR Spectroscopy*; VCH: New York, 1990.
- (22) Burns, J. A.; Whitesides, G. M. Feed-Forwards Neural Networks in chemistry: Mathematical Systems For classification and Pattern Recognition. *Chem. Rev.* **1993**, 93, 8, 2591–2601.
- (23) Livingstone, D. J.; Manallack, D. T. Statistics Using Neural Network: Chance Effects. *J. Med. Chem.* **1993**, 36, 1295–1297.

CI000159V