Revue Roumaine de Chimie, **2000**, *45* (4), 391–403

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. PART 12[1]

THE PREDICTION OF ALKANE HEAT CAPACITIES WITH THE MolNet NEURAL NETWORK

Ovidiu IVANCIUC

"Politehnica" University, Department of Organic Chemistry
Faculty of Industrial Chemistry, Oficiul 12 C.P. 12–243, 78100 Bucharest, Roumania
E-mail: o_ivanciuc@chim.upb.ro

Received December 7, 1998

MolNet, a new multi-layer feedforward neural network is presented together with its application to the computation of alkane heat capacities. The MolNet neural network changes its topology (the number of neurons in the input and hidden layers, together with the number and type of connections) according to the molecular structure of the chemical compound presented to the network. The structure of each molecule is encoded in the corresponding molecular graph that is used to set the MolNet topology. Three structural descriptors derived from the molecular graph are used as input data for the first layer of neurons, namely the degree, the distance sum, and the reciprocal distance sum.

INTRODUCTION

Conventional quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) models require the user to specify the mathematical function of the model. If these functions are highly nonlinear then considerable mathematical and numerical expertise is needed to obtain significant models. In recent years this problem was solved with artificial neural networks (ANN),[2–5] a kind of models in which the mathematical form of the relationship between the input and output data is not specified. The growing interest in their application in chemistry,[6–10] in chemical engineering,[11] and in biochemistry[12] is a result of their unique modeling features.

Many different ANN topologies are possible, but in QSPR and QSAR studies the multi-layer feedforward (MLF) networks represent the majority of applications. An important problem for the QSPR applications of neural networks is the numerical representation of the chemical structure used as input by the ANN, and various structural representations of organic compounds were used in MLF networks.[13–17] In structure-property studies a MLF network receives information related to the molecular structure of the chemical compounds only from the input neurons, and the topology of the neural network (the number of neurons in the input, hidden, and output layers, together with connections between them) is constant for all molecules presented to the network. Three new neural models, that encode into their topology the molecular structure of each compound, were recently introduced: the ChemNet proposed by Kireev;[18] the Baskin, Palyulin, Zefirov (BPZ) neural device;[19] and the MolNet defined by Ivanciuc.[1,20–22] The above three neural models use a set of rules to build the network according to the chemical structure of each molecule examined by the ANN. MolNet uses various atomic descriptors as input structural parameters, and it was successfully applied for the computation of a large number of molecular properties of organic compounds, such as boiling temperature, standard Gibbs energy of formation, density, and vaporization enthalpy. In the present investigation MolNet is applied for the computation of alkane heat capacities, giving good results both in calibration and prediction.

MolNet DESCRIPTION

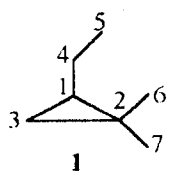
In the description and application of MolNet we will use graph theory concepts.[23-29] A graph $G = G(V,E)$ is an ordered pair consisting of two sets $V = V(G)$ and $E = E(G)$. Elements of the set $V(G)$ are called vertices and elements of the set $E(G)$, involving the binary relation between the vertices, are called edges. In this paper chemical structures are represented as hydrogen-depleted (or hydrogen-suppressed) molecular graphs obtained by removing all hydrogen atoms from the chemical formula of a chemical compound containing covalent bonds. In a molecular graph vertices correspond to non-hydrogen atoms and edges correspond to covalent bonds. In the particular case of alkanes the vertices of the molecular graph denote carbon atoms and the edges denote carbon-carbon single bonds. In this study expressions "molecular graph" and "molecule", "vertex" and "atom", "edge" and "bond" are used interchangeably.

MolNet is a new type of MLF neural network that can be used to compute molecular properties using as input data atomic descriptors.[1,20-22] In MolNet the number of units from the input and hidden layers and the connections are determined according to the structure of the molecule presented to the network. The significance of each MolNet unit and connection is derived from a set of rules that are presented below. Also, the input atomic descriptors are determined by the molecular structure.

When a molecule is presented to the network, each non-hydrogen atom in the molecule has a corresponding unit in the input and hidden layers. The output layer has one unit that offers the computed value of the molecular property under investigation. MolNet has also a bias unit connected to the hidden and output units. Using a set of rules, MolNet changes the number and significance of the units with each molecule presented to the network. A connection between an input and a hidden unit (IH connection) encodes the bonding relationship between the corresponding atoms from the molecular graph. Input-hidden connections corresponding to the same bonding relationship between two atoms, either in the same molecule or in different molecules, belong to the same connection class and they have identical weights. For the network MolNet-1 the bonding relationship considers the shortest path between a pair of atoms. Two identical paths contain the same ordered sequence of atoms and bonds; such paths belong to the same connection class. A unit that corresponds to an atom i in the input layer is connected to the unit corresponding to the same atom i in the hidden layer by a connection; these connections are classified according to the chemical nature of the atoms.

The connections from the hidden units to the output unit (HO connections) are classified according to the type of atoms represented by the hidden units. The atoms are partitioned into classes according to their atomic number Z, the hybridization state and the degree. All hidden units representing atoms of the same type, either in the same molecule or in different molecules, are connected to the output unit with connections possessing identical weights. The bias unit is connected to each hidden unit with connections (BH connections) partitioned in the same way with the connections between the hidden and output layers, i.e. according to the atom types as defined above. The bias unit is connected with the output unit (BO connection). For a molecule with N non-hydrogen atoms MolNet contains N input and N hidden units, with N^2 connections between them, N connections between the hidden and output units, N connections from the bias unit to the hidden units, and one connection from the bias unit to the output unit. Some connections may have identical weights according to the partitioning schemes described above, and the number of adjustable parameters (connection classes) is much lower than the total number of connections.

Each input unit receives an atomic invariant determined for the corresponding atom. The value of the atomic invariant is computed from the structure of the molecule presented to MolNet. Any vertex invariant of the molecular graph can be used as input for MolNet: the number of hydrogen atoms attached to each atom, the degree, the electronegativity, the atomic charge, vertex topological indices.



1

We present the MolNet structure in the case of 1-ethyl-2,2-dimethylcyclopropane **1**. In alkanes all atoms from the corresponding molecular graph represent sp^3-hybridized carbon atoms with single bonds between them. In this case the bonding relationship that determines the type of the IH connections considers only the topological distance between the carbon atoms. The topological distance between vertices i and j is denoted by d_{ij}, and is equal to the number of edges on the shortest path between the vertices i and j. Distances d_{ij} are elements of the distance matrix of the molecular graph G, **D** = **D**(G). The distance matrix of the molecular graph of **1**, **D(1)**, computed with the Floyd-Warshall algorithm,[30] is presented below.

D(1)

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	0	1	1	1	2	2	2
2	1	0	1	2	3	1	1
3	1	1	0	2	3	2	2
4	1	2	2	0	1	3	3
5	2	3	3	1	0	4	4
6	2	1	2	3	4	0	2
7	2	1	2	3	4	2	0

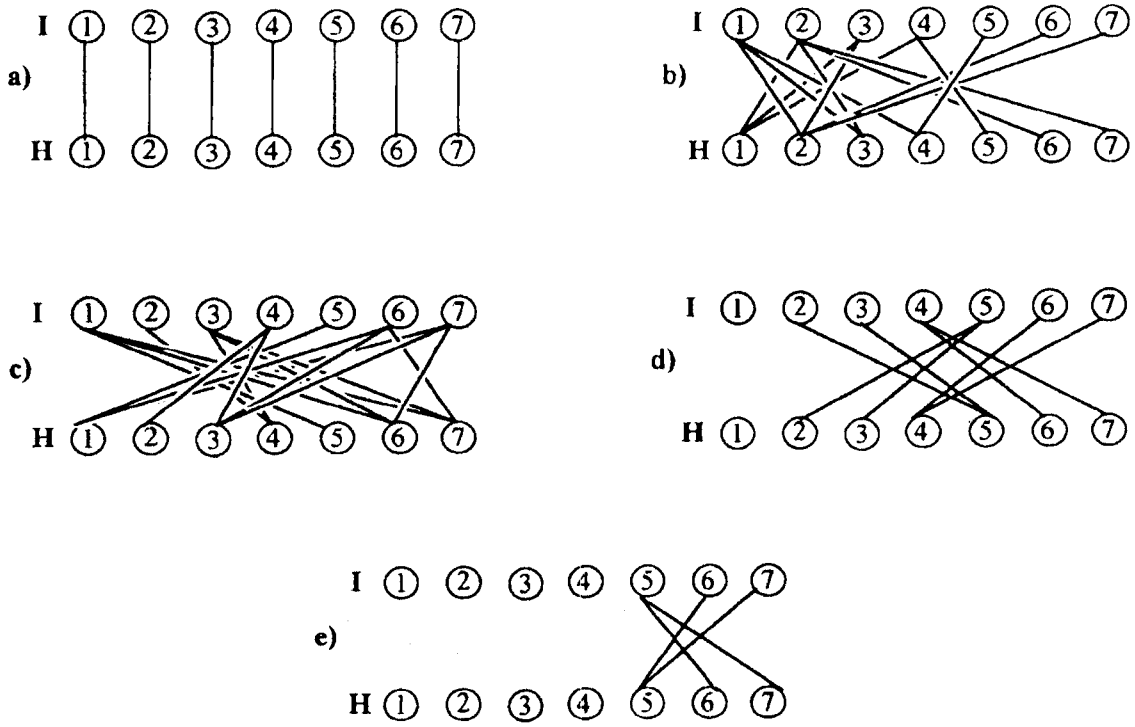


Fig. 1 – The structure of the MolNet connections between the input (**I**) and hidden (**H**) layers for 1-ethyl-2,2-dimethylcyclopropane **1**; each neuron corresponds to the carbon atom with the same label from the molecular graph of **1**. The connections between atoms with the same label are presented in a); the connections between atoms situated at distances 1, 2, 3 and 4 are presented in b), c), d) and e), respectively.

Because **1** has 7 carbon atoms, the corresponding MolNet has 7 input and 7 hidden units. Each unit from the input and hidden layers of MolNet corresponds to the atom with the same label in the molecular graph **1**, as presented in Fig. 1a-e. The distance matrix of **1** has 5 classes of topological distances, from 0 to 4. The 5 distance classes correspond to 5 IH connection types (i.e. five parameters that are adjusted in the MolNet calibration phase). A bonding relationship between two atoms i and j corresponds to two identical IH connections: the first from input unit i to hidden unit j, and the second

from input unit j to hidden unit i. In **1** there are two pairs of atoms situated at distance 4, namely (5 and 6) and (5 and 7). The 4 IH connections with identical weights between the two pairs are depicted in Fig. 1e. In Fig. 1a-e we present the structure of IH connections according to the classes of identical weights: there are 7 connections corresponding to the distance 0 (Fig. 1a), which in our case have identical weights because all non-hydrogen atoms are carbon atoms; Fig. 1b presents the 14 connections between the 7 pairs of atoms situated at distance 1; the 16 connections corresponding to the 8 pairs of atoms situated at distance 2 are presented in Fig. 1c; the 8 connections in Fig. 1d are bonding the 4 pairs of atoms situated at distance 3; Fig. 1e depicts the 4 connections representing relations between the 2 pairs of atoms situated at distance 4.

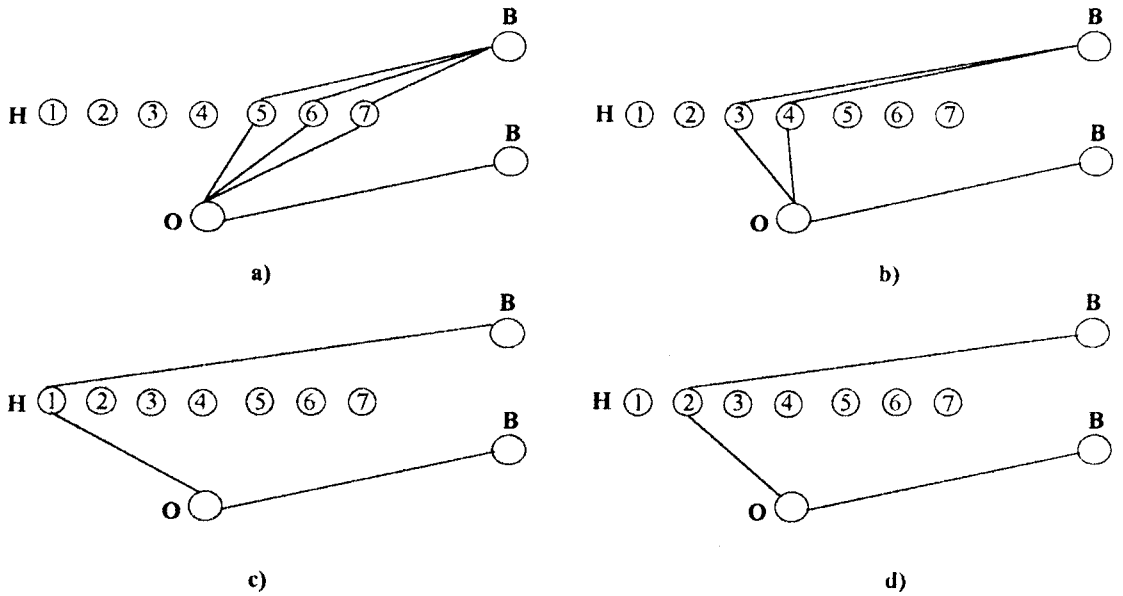


Fig. 2 – The structure of the MolNet connections between the hidden (**H**) and output (**O**) layers for 1-ethyl-2,2-dimethylcyclopropane **1**; the bias neuron is labeled with **B**. The connections to/from atoms with the degree 1, 2, 3 and 4 are presented in a), b), c) and d), respectively.

The type of the HO connections are determined by the structure of the molecule presented to MolNet. In alkanes the HO connections are classified according to the degree of the carbon atoms: hidden units representing atoms with identical degrees are linked to the output unit with connections having identical weights. Because the molecular graph of 1-ethyl-2,2-dimethylcyclopropane **1** contains 3 atoms with degree 1, 2 atoms with degree 2, one atom with degree 3, and one atom with degree 4, the 7 HO connections belong to four classes (*i.e.* adjustable parameters). The BH connections are classified according to the same rules utilized for the HO connections, giving in the case of molecule **1** four adjustable weights. For example, there are 3 atoms with degree 1 in 1-ethyl-2,2-dimethylcyclopropane, namely atoms 5, 6, and 7. Fig. 2a presents their connections to the output unit; obviously, the 3 HO connections have identical weights. The same Fig. 2a depicts the 3 identical BH connections to the 3 hidden units representing atoms with degrees 1. The structure of BH and HO connections is presented in Fig. 2a-d: the bias and output connections to atoms with the degree equal to 1 are presented in Fig. 2a; those connecting the atoms with degree 2, 3, and 4 are depicted in Fig. 2b-d, respectively. The bias unit has also a connection to the output unit, presented in Fig. 2. The total number of adjustable weights for 1-ethyl-2,2-dimethylcyclopropane is: 5 (IH connections) + 4 (BH connections) + 4 (HO connections) + 1 (BO connection) = 14.

MolNet is a MLF network and its use involves two phases: a calibration (learning) and a prediction phase, respectively. In the calibration phase the weights are optimized (adjusted) with a nonlinear optimization algorithm. The selection of the algorithm depends on the number of input units, on the number of patterns in the calibration set and on the number of parameters. One can use global opti-

mization algorithms (such as simulated annealing or genetic algorithms), the simplex algorithm, direction-set methods (Powell's method), methods that require the computation of the first derivatives such as conjugate gradient methods (Fletcher-Reeves or Polak-Ribiere) or quasi-Newton (variable metric) methods (Davidon-Fletcher-Powell or Broyden-Fletcher-Goldfarb-Shanno). For the present investigation we have selected the most widely used method in the optimization of neural networks, the backpropagation with momentum algorithm. The weights are modified after the presentation of each molecule. If a connection type is absent from a certain molecule its value is not modified after the presentation of that molecule to the network. In a molecule all connections belonging to the same class are adjusted with the same quantity obtained by a summation of individual gradients and application of the usual backpropagation with momentum equation.

In the prediction phase MolNet uses the weights determined in calibration to compute the investigated molecular property for molecules that were not present in the calibration set. If the set of molecules used in the prediction phase contains bonding relationships that are absent in the molecules used in the calibration phase these bonding relationships are neglected in predicting the molecular property.

The main differences between ChemNet and MolNet are presented below:

(a) The bonding relationship that determines the topology of the IH connections is different in MolNet and ChemNet. For example, ChemNet uses identical connections for C–C–C, C=C–C, and C=C=C, while in this case MolNet considers three different IH connections.

(b) MolNet uses connections between the units with the same label in the input and hidden layers, while in ChemNet such connections are missing.

(c) The structure of the HO connections in ChemNet is described only for the networks that compute atomic properties. MolNet is used to compute molecular properties and has a topology of the HO connections related to the molecular structure.

(d) In ChemNet the BH connections have identical values. In MolNet the connections from the bias to a hidden unit are classified with the same rules used for the HO connections, namely according to the atomic number Z, the hybridization state and the degree of the atom represented by the hidden unit.

(e) When computing molecular properties, ChemNet can have more than one hidden layer, while MolNet has always only one hidden layer.

MolNet OPERATION

Data Set. The MolNet capability to offer useful QSPR models is tested in an investigation of the structural determination of alkane heat capacities. The structure and experimental heat capacities of the alkanes used in the present investigation are taken from the literature[13] and are reported in Tables 1 and 2. In order to estimate the MolNet prediction power the alkanes are separated into a calibration (learning) set containing 109 alkanes, and a prediction (test) set containing 25 alkanes. We have retained the separation in the calibration and prediction sets used in the literature.[13] The use of the prediction set makes possible the determination of the MolNet ability of predicting the heat capacity for alkanes that are not used in the calibration of the model.

Number of Adjustable Parameters. The number of adjustable parameters (connections) of MolNet depends on the structure of the molecules from the learning set. Because the molecules in the learning set contain only carbon atoms, there is only one IH connection class representing connections between unit i from the input layer with unit i from the hidden layer. For the learning set of 109 alkanes the maximum graph distance between two carbon atoms is 7, giving a total of 8 IH connection classes. The degree of the carbon atoms is from 1 up to 4, and this gives 4 HO connection types and 4 BH adjustable connections. The total number of adjustable weights for the alkane learning set is: 8 (IH connections) + 4 (BH connections) + 4 (HO connections) + 1 (BO connection) = 17. The ratio between the number of alkanes in the learning set and the number of adjustable weights is 6.41 indicating that there is no danger of overfitting.

Table 1

Alkanes used in MolNet calibration, experimental heat capacities at 300 K, J/mol grad, (Cp_{exp}), and calibration residuals ($Cp_{exp} - Cp_{cal}$) for MolNet network using **DEG** (DEG_{res}), **DS** (DS_{res}), and **RDS** (RDS_{res}) input atomic descriptors

Hydrocarbon	Cp_{exp}	DEG_{res}	DS_{res}	RDS_{res}	Hydrocarbon	Cp_{exp}	DEG_{res}	DS_{res}	RDS_{res}
3-M-C_5	140.88	-11.77	-5.55	-0.42	2,4-M_2-C_8	239.40	8.41	3.93	4.25
2,2-M_2-C_4	142.26	-11.29	-2.45	-5.25	2,5-M_2-C_8	231.80	1.44	-3.76	-2.67
2,3-M_2-C_4	140.21	-10.06	-4.89	-5.14	3,4-M_2-C_8	229.30	-2.72	-6.11	-4.72
3-M-C_6	164.50	-2.87	-1.56	0.26	3,5-M_2-C_8	238.30	5.97	2.21	1.54
3-E-C_5	166.80	-1.63	-1.22	5.49	3,6-M_2-C_8	229.60	-1.61	-7.06	-6.30
2,2-M_2-C_5	167.70	-1.04	2.21	0.81	4,4-M_2-C_8	239.30	7.72	5.00	6.22
2,3-M_2-C_5	161.80	-4.27	-5.41	-3.15	4,5-M_2-C_8	230.10	-2.41	-4.18	-4.49
2,4-M_2-C_5	171.70	5.02	5.02	0.61	4-nP-C_7	237.70	3.93	-4.70	-4.04
3,3-M_2-C_5	166.70	-1.69	-0.17	4.33	4-iP-C_7	239.20	5.42	0.16	1.06
2,2,3-M_3-C_4	164.20	-1.77	-0.95	-0.90	2-M-3-E-C_7	238.50	5.05	-0.69	5.56
n-C_8	188.70	3.14	-1.87	-6.22	2-M-4-E-C_7	243.40	9.77	3.97	3.59
2-M-C_7	188.20	2.04	-1.89	-5.60	3-M-4-E-C_7	236.20	2.09	-0.06	-1.17
3-M-C_7	186.82	0.12	-2.86	-6.26	3-M-5-E-C_7	240.90	7.54	3.56	4.82
2,4-M_2-C_6	193.35	6.54	2.65	-0.08	2,2,3-M_3-C_7	232.50	0.89	-2.28	4.15
2,5-M_2-C_6	186.52	1.73	-3.46	-5.48	2,2,4-M_3-C_7	234.70	2.25	0.38	3.91
3,3-M_2-C_6	191.96	2.99	1.26	1.81	2,2,5-M_3-C_7	230.50	0.16	-3.50	1.31
3,4-M_2-C_6	182.72	-4.58	-8.57	-9.03	2,2,6-M_3-C_7	234.80	4.16	1.21	6.97
3-E-2-M-C_5	193.05	4.58	0.56	5.43	2,3,3-M_3-C_7	235.10	2.89	1.01	4.30
3-E-3-M-C_5	189.07	-0.74	-3.63	1.06	2,3,4-M_3-C_7	237.60	3.94	4.29	4.66
2,2,3-M_3-C_5	186.77	-1.06	-5.03	-1.88	2,3,5-M_3-C_7	233.90	1.18	-0.20	-0.58
2,2,4-M_3-C_5	189.45	0.86	-2.21	-1.03	2,3,6-M_3-C_7	228.50	-3.51	-6.33	-4.18
2,3,3-M_3-C_5	188.20	0.40	-4.36	-2.56	2,4,4-M_3-C_7	238.90	5.00	5.72	6.06
2,3,4-M_3-C_5	192.72	6.18	-0.17	1.12	2,4,5-M_3-C_7	234.10	0.76	0.72	0.59
2,2,3,3-M_4-C_4	188.28	0.94	-3.07	-3.04	2,4,6-M_3-C_7	246.30	12.73	11.43	11.60
2-M-C_8	210.90	4.64	-5.81	-7.13	2,5,5-M_3-C_7	234.20	3.27	0.89	2.44
3-M-C_8	209.70	1.54	-6.97	-9.67	3,3,5-M_3-C_7	234.10	1.46	1.57	-1.29
3-E-C_7	213.00	2.43	-3.46	-4.49	3,4,4-M_3-C_7	235.60	1.70	3.69	3.97
4-E-C_7	214.30	3.07	-2.05	-6.93	3,4,5-M_3-C_7	235.10	1.22	3.22	1.15
2,2-M_2-C_7	212.40	3.56	-3.84	0.82	2-M-3-iP-C_6	231.80	-2.75	-4.04	-3.25
2,3-M_2-C_7	207.70	-1.52	-7.81	-8.79	3,3-E_2-C_6	242.50	7.40	8.80	7.58
2,4-M_2-C_7	217.10	6.61	1.97	-0.22	3,4-E_2-C_6	246.90	12.03	11.92	10.10
2,5-M_2-C_7	208.20	-0.47	-7.00	-9.34	2,2-M_2-3-E-C_6	227.70	-5.61	-6.86	-3.86
2,6-M_2-C_7	210.40	1.89	-5.75	-6.72	2,2-M_2-4-E-C_6	236.10	4.00	1.55	1.79
3,3-M_2-C_7	214.00	4.05	-0.29	-1.46	2,3-M_2-3-E-C_6	238.20	3.71	4.75	4.09
3,4-M_2-C_7	206.80	-3.69	-7.22	-9.89	2,3-M_2-4-E-C_6	243.00	8.70	9.74	8.59
3,5-M_2-C_7	214.60	4.65	0.28	-4.83	2,4-M_2-4-E-C_6	235.00	0.89	1.99	-1.83
3-E-3-M-C_6	214.10	1.70	-0.63	-2.57	3,3-M_2-4-E-C_6	228.20	-6.46	-4.16	-4.50
4-E-2-M-C_6	219.70	9.09	4.41	0.64	3,4-M_2-4-E-C_6	235.50	0.65	3.37	1.26
2,2,4-M_3-C_6	210.70	0.53	-3.77	-4.04	2,2,3,3-M_4-C_6	238.20	5.17	4.99	6.78
2,2,5-M_3-C_6	209.10	1.61	-5.15	-3.68	2,2,3,4-M_4-C_6	229.40	-3.73	-2.32	-3.47
2,3,3-M_3-C_6	213.30	2.27	-1.80	-2.11	2,2,3,5-M_4-C_6	235.80	4.24	3.86	4.76
2,3,4-M_3-C_6	214.00	3.54	-0.77	-3.08	2,2,4,5-M_4-C_6	229.20	-2.27	-2.46	-2.95
2,3,5-M_3-C6	212.50	3.22	-2.09	-4.76	2,2,5,5-M_4-C_6	229.80	1.98	-1.02	3.38
2,4,4-M_3-C_6	213.50	2.11	-1.18	-2.90	2,3,3,4-M_4-C_6	241.50	7.25	9.23	7.72
3,3,4-M_3-C_6	210.50	-1.16	-4.70	-5.20	2,3,3,5-M_4-C_6	234.00	0.55	1.86	0.23
3,3-E_2-C_5	217.86	3.97	1.79	8.15	2,3,4,4-M_4-C_6	231.80	-2.36	0.10	-2.66
3-E-2,2-M_2-C_5	205.00	-7.42	-10.96	-3.61	2,3,4,5-M_4-C_6	243.10	9.36	11.55	9.00
3-E-2,3-M_2-C_5	213.40	0.82	-3.47	0.59	3,3,4,4-M_4-C_6	238.00	3.55	4.91	4.01
2,2,3,3-M_4-C_5	213.34	1.48	-4.60	-1.18	2,4-M_2-3-iP-C_5	234.50	-1.55	0.60	2.51
2,2,3,4-M_4-C_5	208.50	-2.72	-8.57	-4.58	2-M-3,3-E_2-C_5	224.80	-11.31	-8.52	-3.78
2,3,3,4-M_4-C_5	219.50	8.22	1.45	3.57	2,2,3-M_3-3-E-C_5	223.70	-11.29	-10.72	-6.46
3-E-C_8	235.80	4.86	-4.01	-0.77	2,2,4-M_3-3-E-C_5	227.30	-7.79	-6.23	-1.93
4-E-C_8	236.50	4.20	-4.53	-3.51	2,3,4-M_3-3-E-C_5	229.00	-6.56	-5.03	-2.70
2,2-M_2-C_8	235.10	7.78	2.56	7.94	2,2,3,3,4-M_5-C_5	234.30	-0.15	-1.27	0.90
					2,2,3,4,4-M_5-C_5	234.20	0.52	0.65	2.87

Table 2

Alkanes used in MolNet prediction, experimental heat capacities at 300 K, J/mol grad, (Cp_{exp}), and prediction residuals ($Cp_{exp} - Cp_{pr}$) for MolNet network using **DEG** (DEG_{res}), **DS** (DS_{res}), and **RDS** (RDS_{res}) input atomic descriptors

Hydrocarbon	Cp_{exp}	DEG_{res}	DS_{res}	RDS_{res}	3-E-2,4-M_2-C_5	209.00	-3.19	-7.28	-2.39
2-M-C_5	143.01	-9.85	-2.48	-4.72	2,2,4,4-M_4-C_5	215.77	3.73	-0.47	5.76
n-C_7	166.00	-1.70	2.49	0.80	2,3-M_2-C_8	230.50	0.35	-5.47	-1.41
2-M-C_6	165.40	-0.96	0.76	0.84	2,6-M_2-C_8	231.90	1.94	-3.32	-3.15
4-M-C_7	188.03	0.16	-1.79	-5.08	2,7-M_2-C_8	233.20	5.93	0.87	1.98
3-E-C_6	190.58	2.18	-0.35	-0.76	3,3-M_2-C_8	237.10	6.48	1.28	5.71
2,2-M_2-C_6	189.33	2.54	-0.04	1.16	2-M-5-E-C_7	234.30	1.63	-4.38	0.96
2,3-M_2-C_6	185.18	-1.26	-5.50	-6.57	3-M-3-E-C_7	236.20	3.32	1.10	3.35
4-M-C_8	210.40	1.55	-5.25	-9.02	4-M-3-E-C_7	238.60	4.35	3.23	5.93
4,4-M_2-C_7	217.20	5.22	3.36	3.09	4-M-4-E-C_7	239.20	5.00	4.58	2.66
3-E-2-M-C_6	216.10	4.96	0.64	-0.10	3,3,4-M_3-C_7	233.60	0.55	1.53	1.89
3-E-4-M-C_6	215.20	3.48	0.58	-0.88	2,5-M_2-3-E-C_6	240.80	7.15	4.84	5.08
2,2,3-M_3-C_6	209.90	0.07	-4.79	-3.54	2,2,4,4-M_4-C_6	239.20	6.55	7.81	5.44

Input Data. Three atomic topological descriptors are used as input data for the input layer neurons, namely the degree **DEG**,[24,26] the distance sum **DS**,[31,32] and the reciprocal distance sum **RDS**.[33-36] The degree of the atom i is computed with the equation:[24,26]

$$\mathbf{DEG}_i = \sum_{j=1}^{N} \mathbf{A}_{ij}$$

where **A** is the adjacency matrix. The degree vector of 1-ethyl-2,2-dimethylcyclopropane is **DEG(1)** = = {3, 4, 2, 2, 1, 1, 1}. This graph invariant gives information on the local structure of the molecular graph, and has a fairly high degeneracy, i.e. non-equivalent atoms have identical degrees. For example, in **1** the non-equivalent atoms 3 and 4 have the degree 2.

In a molecular graph G the distance sum of the atom i, $\mathbf{DS}_i = \mathbf{DS}_i(G)$, is the sum of the elements in the row i (or column i) of the distance matrix $\mathbf{D} = \mathbf{D}(G)$:[31,32]

$$\mathbf{DS}_i = \sum_{j=1}^{N} \mathbf{D}_{ij}$$

The distance sum vector of 1-ethyl-2,2-dimethylcyclopropane is **DS(1)** = {9, 9, 11, 12, 17, 14, 14}.

The reciprocal distance sum of the atom i in a molecular graph G is defined by the equation:[33-36]

$$\mathbf{RDS}_i = \sum_{j=1}^{N} \mathbf{RD}_{ij}$$

where **RD** = **RD**(G) is the reciprocal distance matrix of G. The reciprocal distance sum vector of 1-ethyl-2,2-dimethylcyclopropane is **RDS(1)** = {4.50000, 4.83333, 3.83333, 3.66667, 2.66667, 3.08333, 3.08333}.

Learning Method. The MolNet training is made with the standard backpropagation with momentum method,[3,4] until the convergence is obtained. The calibration converges when the correlation coefficient between experimental and calculated alkane heat capacities improves by less than 10^{-5} in 100 epochs; this convergence condition was applied with good results for the calibration of usual MLF networks.[37-39] One epoch corresponds to one complete presentation of the 109 molecules in the learning set. The connection weights are adjusted after the presentation of each molecule. Random values between –0.1 and 0.1 are used as initial weights. The learning process is very sensitive to the learning rate and momentum values, and a small, non-zero, momentum value is necessary for the convergence of the calibration phase.[1,20-22] The learning rates are equal to 0.05 for both the hidden and output layers, and the momentum is set between 0.30 and 0.05 for all activation functions used in this

study. The learning rate and momentum values are maintained constant during the training phase. In all cases the convergence of the calibration phase is obtained in a few hundreds epochs and the results are not greatly influenced by the initial random set of weights.

Activation Functions. The sigmoid, which is the most commonly used activation function, takes values between 0 and 1. For large negative arguments its value is close to 0, and learning with the backpropagation algorithm is difficult in such conditions. The present study uses a better function, the hyperbolic tangent (tanh), that takes values between –1 and 1. When the absolute value of the argument is greater than 10 the tanh and the sigmoid activation functions are very flat. In such conditions the derivatives of the two functions have an extremely small value, leading to a poor sensitivity of the two activation functions to large positive or negative arguments. This property is one of the causes of the very slow rates of convergence during the calibration of neural networks with algorithms that use the derivative of the activation function, such as the backpropagation algorithm. A linear output activation function, with a constant sensitivity, overcomes the problems of the tanh function, and gives better prediction results in MLF networks.[37,39,40] For the output layer we use also a linear activation function. A new type of activation function is the symmetric logarithmoid,[41,42] defined by the formula: $Act(z) = \text{sign}(z) \ln(1 + |z|)$. The symmetric logarithmoid (symlog) is a monotonically increasing function with the maximum sensitivity near zero and with a monotonically decreasing sensitivity away from zero. Because its output is not restricted to a finite range of values this function is sensitive to large positive or negative arguments. The symlog function is used for the output layer of MolNet.

Preprocessing of the Data. The input vectors (**DEG**, **DS**, or **RDS**) and the output values (representing the target heat capacity) are linearly scaled between –0.9 and 0.9. For the tanh output activation function the scaling is required by the range of values of the function between –1 and 1, while for the unbounded functions (linear and symlog) the experience showed that a linear scaling improves the calibration process.

Performance Indicators. The MolNet results are evaluated both for the network calibration and prediction. The quality of MolNet calibration is estimated by comparing the calculated alkane heat capacities obtained at the end of the calibration phase ($C_{p,cal}$) with the target values ($C_{p,exp}$), while the predictive quality is estimated with a set of alkanes that were not used in the calibration phase by comparing the predicted ($C_{p,pr}$) and experimental values. In order to compare the results of MolNet networks calibrated with different learning rate and momentum values we use the correlation coefficient r and the standard deviation s of the linear correlation between experimental and calculated (in calibration or prediction) heat capacities: $C_{p,exp} = A + B \cdot C_{p,cal/pr}$.

MolNet COMPUTATION of ALKANE HEAT CAPACITIES

In a QSPR study with a usual MLF neural network the number of hidden units is determined during the calibration phase. MolNet has a topology that depends on the structure of the molecule presented to the network. The number of adjustable parameters in MolNet is determined by the bonding relationships between atoms and by the types of atoms from the calibration set of molecules. In MolNet the number of hidden units or the number of connections are no longer variables that can be optimized during the calibration phase. MolNet accepts as input any atomic property, and the calibration and prediction results depend on the atomic invariant used to feed the network. This study investigates the alkane heat capacities estimation with MolNet networks that use three input atomic descriptors, namely **DEG**, **DS**, and **RDS**. These descriptors can be readily computed for any molecule from the structure of the corresponding molecular graph.

The first set of MolNet networks is obtained using as input data the **DEG** atomic descriptor. The corresponding calibration and prediction results for alkane heat capacities are presented in Table 3. The calibration correlation coefficient, r_{cal}, takes values between 0.980 to 0.986, and the calibration

Table 3

MolNet calibration and prediction results for the computation of alkane heat capacities using DEG input atomic descriptor. The table reports the number of training epochs, the hidden and output momentum (ah and ao), the output activation function, the calibration and prediction standard deviation (scal and spr) and correlation coefficient (rcal and rpr). All networks were provided with the tanh hidden activation function

Epoch	Hidden Momentum (α_h)	Output Activation Function	Output Momentum (α_o)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
800	0.30	linear	0.30	4.96	0.980	2.90	0.995
900	0.30	linear	0.15	4.77	0.982	2.79	0.995
900	0.30	linear	0.10	4.71	0.982	2.77	0.995
900	0.30	linear	0.05	4.66	0.983	2.74	0.995
800	0.15	linear	0.05	4.46	0.984	3.70	0.991
900	0.10	linear	0.05	4.63	0.983	2.71	0.995
700	0.05	linear	0.05	4.47	0.984	3.64	0.991
900	0.15	linear	0.15	4.75	0.982	2.77	0.995
900	0.15	linear	0.10	4.70	0.982	2.75	0.995
900	0.10	linear	0.10	4.68	0.982	2.73	0.995
400	0.30	symlog	0.30	4.84	0.981	4.20	0.989
900	0.30	symlog	0.15	4.78	0.982	4.29	0.988
1000	0.30	symlog	0.10	4.76	0.982	4.31	0.988
1100	0.30	symlog	0.05	4.74	0.982	4.32	0.988
1000	0.15	symlog	0.05	4.70	0.982	4.21	0.988
1100	0.10	symlog	0.05	4.69	0.982	4.19	0.989
1200	0.05	symlog	0.05	4.68	0.983	4.17	0.989
800	0.15	symlog	0.15	4.73	0.982	4.19	0.989
900	0.15	symlog	0.10	4.71	0.982	4.19	0.989
1000	0.10	symlog	0.10	4.70	0.982	4.17	0.989
500	0.30	tanh	0.30	4.94	0.980	4.14	0.989
300	0.30	tanh	0.15	4.85	0.981	4.07	0.989
400	0.30	tanh	0.10	4.83	0.981	4.08	0.989
900	0.30	tanh	0.05	4.85	0.981	4.32	0.988
1000	0.15	tanh	0.05	4.80	0.982	4.21	0.988
1100	0.10	tanh	0.05	4.79	0.982	4.18	0.989
800	0.05	tanh	0.05	4.78	0.982	4.14	0.989
400	0.15	tanh	0.15	4.82	0.981	4.02	0.989
800	0.15	tanh	0.10	4.83	0.981	4.20	0.988
600	0.10	tanh	0.10	4.12	0.986	3.83	0.990

standard deviation, s_{cal}, is in the range 4.12 and 4.96. The prediction correlation coefficient, r_{pr}, is larger than r_{cal} with values between 0.988 and 0.995, while the prediction standard deviation, s_{pr}, is in the range 2.73 to 4.32. As a general trend, the prediction results are of higher quality than the calibration statistical indices. Overall, the best calibration and prediction results are obtained with linear output function, followed by the tanh and symlog output functions. The alkane heat capacity residuals computed with **DEG** input data, a linear output function, a hidden and an output momentum $\alpha_h = \alpha_o = 0.10$ are presented in column 4 of Table 1 for calibration, and in Table 2 for prediction. This example, with $r_{cal} = 0.982$, $s_{cal} = 4.68$, $r_{pr} = 0.995$, and $s_{pr} = 2.73$, is selected because it offers good calibration and prediction results From the whole set of 109 molecules used in calibration, one finds 7 alkanes with residuals greater than 10 J K^{-1} mol^{-1}, presented here together with their residuals: 3-methylpentane with −11.77, 2,2-dimethylbutane with −11.29, 2,3-dimethylbutane with −10.06, 2,4,6-trimethylheptane with 12.73, 3,4-diethylhexane with 12.03, 2-methyl-3,3-diethylpentane with −11.31, and 2,2,3-trimethyl-3-ethylpentane with −11.29. If we consider that an outlier pattern in a QSPR model has an absolute resi-

dual 3 times greater than the standard deviation, it follows that in this case, when $3s_{cal} = 14.04$, there is no outlier. The threshold of 10 J K^{-1} mol^{-1} is used only to compare the results obtained with the three input structural descriptors, namely **DEG**, **DS**, and **RDS**. The greatest prediction residual is obtained for 2-methylpentane with −9.85 J K^{-1} mol^{-1}.

Table 4

MolNet calibration and prediction results for the computation of alkane heat capacities using **DS** input atomic descriptor. The notations are explained in Table 3

Epoch	Hidden Momentum (α_h)	Output Activation Function	Output Momentum (α_o)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1500	0.30	linear	0.30	5.25	0.978	5.33	0.981
1400	0.30	linear	0.15	5.16	0.979	5.29	0.982
1500	0.30	linear	0.10	5.13	0.979	5.28	0.982
1600	0.30	linear	0.05	5.11	0.979	5.28	0.982
500	0.15	linear	0.05	4.65	0.983	3.86	0.990
600	0.10	linear	0.05	4.63	0.983	3.98	0.990
400	0.05	linear	0.05	4.61	0.983	3.81	0.991
1700	0.15	linear	0.15	5.02	0.980	5.06	0.983
300	0.15	linear	0.10	4.66	0.983	3.82	0.991
400	0.10	linear	0.10	4.65	0.983	3.85	0.990
600	0.30	symlog	0.30	5.27	0.978	5.25	0.982
600	0.30	symlog	0.15	5.25	0.978	5.28	0.982
1900	0.30	symlog	0.10	5.25	0.978	5.34	0.981
2000	0.30	symlog	0.05	5.25	0.978	5.33	0.981
1600	0.15	symlog	0.05	5.18	0.979	5.19	0.982
1700	0.10	symlog	0.05	5.16	0.979	5.17	0.983
1500	0.05	symlog	0.05	5.14	0.979	5.13	0.983
800	0.15	symlog	0.15	5.18	0.979	5.13	0.983
1700	0.15	symlog	0.10	5.18	0.978	5.21	0.982
2000	0.10	symlog	0.10	5.16	0.979	5.16	0.983
400	0.30	tanh	0.30	5.43	0.976	5.51	0.980
600	0.30	tanh	0.15	5.36	0.977	5.48	0.980
1500	0.30	tanh	0.10	5.35	0.977	5.45	0.981
2000	0.30	tanh	0.05	5.50	0.976	5.61	0.979
600	0.15	tanh	0.05	5.30	0.978	5.33	0.981
200	0.10	tanh	0.05	5.19	0.978	4.99	0.984
900	0.05	tanh	0.05	5.24	0.978	5.22	0.982
600	0.15	tanh	0.15	5.31	0.977	5.19	0.982
2000	0.15	tanh	0.10	4.34	0.985	4.50	0.987
700	0.10	tanh	0.10	5.29	0.978	5.26	0.982

The MolNet calibration and prediction results obtained by using as input data the **DS** atomic descriptor are presented in Table 4. In the calibration phase r_{cal} is in the range 0.976 and 0.985, while s_{cal} takes values between 4.34 and 5.50. In the prediction phase r_{pr} takes values between 0.979 and 0.991, and s_{pr} is in the range 3.81 to 5.61. The results obtained with **DS** input data are not as good as those obtained with **DEG** input descriptors. The difference between the three output activation functions is small, with slightly better results for the linear function. The alkane heat capacity residuals computed with **DS** input data, a linear output function, a hidden and an output momentum $\alpha_h = \alpha_o = 0.05$ are presented in column 5 of Table 1 for calibration, and in Table 2 for prediction. This example has the following statistical indices: $r_{cal} = 0.983$, $s_{cal} = 4.61$, $r_{pr} = 0.991$, and $s_{pr} = 3.81$. The calibration results are better than those obtained with the **DEG** input data, with only 5 alkanes having an absolute residual greater than 10 J K^{-1} mol^{-1}. The set of 5 alkanes is presented here with their corresponding residuals: 3-ethyl-2,2-dimethylpentane with −10.96, 2,4,6-trimethylheptane with 11.43, 3,4-diethylhexane with

Table 5

MolNet calibration and prediction results for the computation of alkane heat capacities using **RDS** input atomic descriptor. The notations are explained in Table 3

Epoch	Hidden Momentum (α_h)	Output Activation Function	Output Momentum (α_o)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
900	0.30	linear	0.30	4.96	0.980	4.18	0.989
1000	0.30	linear	0.15	4.87	0.981	4.06	0.989
1100	0.30	linear	0.10	4.85	0.981	4.02	0.989
1100	0.30	linear	0.05	4.84	0.981	3.99	0.990
2000	0.15	linear	0.05	4.91	0.981	4.05	0.989
800	0.10	linear	0.05	4.72	0.982	3.84	0.990
900	0.05	linear	0.05	4.69	0.982	3.83	0.990
2000	0.15	linear	0.15	4.77	0.982	3.99	0.990
1900	0.15	linear	0.10	4.78	0.982	3.98	0.990
1800	0.10	linear	0.10	4.80	0.982	3.98	0.990
1700	0.30	symlog	0.30	5.41	0.977	4.83	0.985
1700	0.30	symlog	0.15	5.82	0.973	5.78	0.978
600	0.30	symlog	0.10	5.29	0.978	4.71	0.986
1100	0.30	symlog	0.05	4.84	0.981	3.98	0.990
200	0.15	symlog	0.05	5.23	0.978	4.54	0.987
2000	0.10	symlog	0.05	5.19	0.978	4.64	0.986
1800	0.05	symlog	0.05	4.82	0.981	3.97	0.990
1100	0.15	symlog	0.15	4.88	0.981	4.01	0.990
1900	0.15	symlog	0.10	4.84	0.981	3.97	0.990
2000	0.10	symlog	0.10	4.85	0.981	3.99	0.990
1700	0.30	tanh	0.30	5.48	0.976	4.89	0.984
400	0.30	tanh	0.15	4.93	0.981	4.16	0.989
1300	0.30	tanh	0.10	4.91	0.981	4.13	0.989
700	0.30	tanh	0.05	4.86	0.981	4.08	0.989
700	0.15	tanh	0.05	5.28	0.978	4.63	0.986
400	0.10	tanh	0.05	4.84	0.981	4.11	0.989
1600	0.05	tanh	0.05	4.82	0.981	4.03	0.989
1300	0.15	tanh	0.15	4.88	0.981	4.09	0.989
600	0.15	tanh	0.10	4.86	0.981	4.08	0.989
1400	0.10	tanh	0.10	4.85	0.981	4.07	0.989

11.92, 2,3,4,5-tetramethylhexane with 11.55, and 2,2,3-trimethyl-3-ethylpentane with −10.72. The largest prediction error is obtained for 2,2,4,4-tetramethylhexane with 7.81 J K^{-1} mol^{-1}.

Table 5 presents the calibration and prediction results obtained when the MolNet parameters are optimized using the **RDS** atomic descriptor as input data. In the calibration phase r_{cal} is between 0.973 and 0.982, and s_{cal} takes values between 4.69 and 5.82. In the prediction phase r_{pr} takes values between 0.978 and 0.990, and s_{pr} is in the range 3.83 to 5.78. The statistical indices reported in Table 5 show that there is a small difference between the results obtained with the three output activation functions. The alkane heat capacity residuals computed with **RDS** input data, a linear output function, a hidden and an output momentum $\alpha_h = \alpha_o = 0.05$ are presented in Tables 1 and 2, column 6. The statistical indices of this example are $r_{cal} = 0.982$, $s_{cal} = 4.69$, $r_{pr} = 0.990$, and $s_{pr} = 3.83$. In the calibration set of 109 alkanes there are 2 cases with absolute residuals greater than 10 J K^{-1} mol^{-1}, presented here with their residuals: 2,4,6-trimethylheptane with 11.60, and 3,4-diethylhexane with 10.10. The above two alkanes have also large residuals for the **DEG** and **DS** input data. The largest prediction residual is obtained for 4-methyloctane with −9.02 J K^{-1} mol^{-1}.

Using three input structural descriptors, namely **DEG**, **DS**, and **RDS**, the MolNet network provides good calibration and prediction results for the alkane heat capacities. Although some alkanes exhibit a residual larger than 10 J K^{-1} mol^{-1}, there is no outlier identified in the QSPR models reported

here. All three structural descriptors give comparable results for the calibration, while for the prediction the best results are obtained with the **DEG** descriptor. MolNet networks provided with a linear output function give better calibration and prediction results.

In general, there is no systematic deviation for certain types of alkanes, small or large, linear or highly branched. An inspection of the residuals shows that for certain alkanes a large residual is computed with one of three descriptors while the residuals computed with the other two descriptors are small. For example, the **DEG** calibration residuals of 3-methylpentane, 2,2-dimethylbutane, 2,3-dimethylbutane are fairly large, while for the same three alkanes, the **DS** and **RDS** calibration residuals are small. This shows that there is no special trend in these cases, the only difference being in the way the molecular structure is reflected in the 3 atomic descriptors. In a small number of cases the residuals computed with all three descriptors are fairly large. Two such examples are 2,4,6-trimethylheptane and 3,4-diethylhexane, whose calibration residuals are large for all three input descriptors. Because the MolNet results depend heavily on the input data, other atomic descriptors computed from the molecular graph, representing in new ways the chemical structure, must be tested in order to identify the best one.

ACKNOWLEDGEMENT. Financial support was obtained from the Ministry of Research and Technology under Grant 310 TA10 and from the Ministry of the National Education under Grant 7001 T34.

REFERENCES

1. Part 11: O. Ivanciuc, *Rev. Roum. Chim.*, **1999**, *44*, 000–000.
2. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.*, **1982**, *79*, 2554–2558.
3. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Nature*, **1986**, *323*, 533–536.
4. D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge, MA, USA, 1986.
5. P. D. Wasserman, *Neural Computing*, Van Nostrand Reinhold, New York, **1989**.
6. J. Zupan and J. Gasteiger, *Anal. Chim. Acta*, **1991**, *248*, 1–30.
7. J. Gasteiger and J. Zupan, *Angew. Chem. Int. Ed. Engl.*, **1993**, *32*, 503–527.
8. J. A. Burns and G. M. Whitesides, *Chem. Rev.*, **1993**, *93*, 2583–2601.
9. J. Zupan and J. Gasteiger, *Neural Networks for Chemists*, VCH, Weinheim, **1993**.
10. O. Ivanciuc, *Molecular Graph Descriptors Used in Neural Network Models.* In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 697–777.
11. A. B. Bulsari (Ed.), *Neural Networks for Chemical Engineers*, Elsevier, Amsterdam, **1995**.
12. J. Devillers (Ed.), *Neural Networks in QSAR and Drug Design*, Academic Press, London, **1996**.
13. A. A. Gakh, E. G. Gakh, B. G. Sumpter, and D. W. Noid, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 832–839.
14. A. T. Balaban, S. C. Basak, T. Colburn, and G. D. Grunwald, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 1118–1121.
15. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye, and J. P. Doucet, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 644–653.
16. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye, and J. P. Doucet, *J. Chem. Inf. Comput Sci.*, **1997**, *37*, 587–598.
17. O. Ivanciuc, J.-P. Rabine, and D. Cabrol-Bass, *Comput. Chem.*, **1997**, *21*, 437–443.
18. D. B. Kireev, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 175–180.
19. I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 715–721.
20. O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 885–894.
21. O. Ivanciuc, *Rev. Roum. Chim.*, **1999**, *44*, 000–000.
22. O. Ivanciuc, *Anal. Chim. Acta*, **1999**, *384*, 271–284.
23. M. V. Diudea and O. Ivanciuc, *Molecular Topology*, Comprex, Cluj, Romania, **1995**.
24. O. Ivanciuc and A. T. Balaban, *Graph Theory in Chemistry.* In: *The Encyclopedia of Computational Chemistry*, Eds.: P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner. John Wiley & Sons, Chichester, **1998**, pp. 1169–1190.
25. A. T. Balaban and O. Ivanciuc, *Historical Development of Topological Indices.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, **1999**, pp. 21–57.
26. O. Ivanciuc and A. T. Balaban, *The Graph Description of Chemical Structures.* In: *Topological Indices and Related Descriptors in QSAR and QSPR,* Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, **1999**, pp. 59–167.

27. O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, *Vertex- and Edge-Weighted Molecular Graphs and Derived Structural Descriptors*. In: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, **1999**, pp. 169–220.
28. O. Ivanciuc and T. Ivanciuc, *Matrices and Structural Descriptors Computed from Molecular Graph Distances. In: Topological Indices and Related Descriptors in QSAR and QSPR*, Eds.: J. Devillers and A. T. Balaban. Gordon and Breach Science Publishers, The Netherlands, **1999**, pp. 221–277.
29. O. Ivanciuc, T. Ivanciuc, and M. V. Diudea, *SAR QSAR Environ. Res.*, **1997**, *7*, 63–87.
30. B. Mohar and T. Pisanski, *J. Math. Chem.*, **1988**, *2*, 267–277.
31. A. T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399–404.
32. A. T. Balaban, *Pure Appl. Chem.*, **1983**, *55*, 199–206.
33. O. Ivanciuc, *Rev. Roum. Chim.*, **1989**, *34*, 1361–1368.
34. T. S. Balaban, P. A. Filip, and O. Ivanciuc, *J. Math. Chem.*, **1992**, *11*, 79–105.
35. O. Ivanciuc, T. S. Balaban, and A. T. Balaban, *J. Math. Chem.*, **1993**, *12*, 309–318.
36. M.V. Diudea, O. Ivanciuc, S. Nikolić, and N. Trinajstić, *MATCH (Commun. Math. Comput. Chem.)*, **1997**, *35*, 41–64.
37. O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 347–354.
38. O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 545–551.
39. O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 775–783.
40. O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 255–260.
41. A. B. Bulsari and H. Saxén, *Neurocomputing*, **1991**, *3*, 125–133.
42. A. B. Bulsari and H. Saxén, *Neural Network World*, **1991**, *4*, 221–224.