ELSEVIER

Analytica Chimica Acta 384 (1999) 271–284

ANALYTICA CHIMICA ACTA

The neural network MolNet prediction of alkane enthalpies

Ovidiu Ivanciuc*

University 'Politehnica' of Bucharest, Department of Organic Chemistry, Faculty of Industrial Chemistry, Oficiul 12 CP 243, 78100 Bucharest, Romania

Received 7 April 1998; received in revised form 1 July 1998; accepted 7 July 1998

Abstract

MolNet, a new type of multi-layer feedforward neural network, is presented together with its application to the computation of alkane enthalpies. 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. Each atom from the molecular graph has a corresponding neuron in the input and hidden layers, respectively. 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.

Keywords: Neural network; MolNet; Structure–property model; Alkane enthalpy prediction; Molecular graph structural descriptor; Topological index

1. Introduction

There is a growing interest in the application of artificial neural networks (ANN) [1,2] in chemistry [3], in chemical engineering [4] and in biochemistry [5], mainly due to their high flexibility in modeling non-linear relationships. Various physico-chemical properties of inorganic and organic compounds were predicted in quantitative structure–property relationship (QSPR) studies involving neural networks.

The numerical representation of the chemical structure used as input by the neural networks is an important problem for the QSPR applications of neural models. Various numerical representations of organic compounds were proposed in QSPR studies using multi-layer feedforward (MLF) neural models: connection table describing the substituents [6]; modified bond-electron matrix containing as structural information the formal bond order between a pair of atoms and the atomic number Z [7]; molecular graph (topological) distance between methyl groups [8]; constitutional descriptors and topological indices [9]; numerical code [10]; counts of various molecular subgraphs (clusters) [11]; vectorial representation of the chemical structure of the substituents [12]; topo-stereochemical code describing the environment of an atom [13,14]; the three-dimensional structure encoded in the 3D MORSE (molecule representation of structures based on electron diffraction) representation [15,16]; atom type electrotopological state [17]; presence of a substituent (coded with unity) or absence (coded with zero) [18]; topological autocorrelation

*E-mail: o_ivanciuc@chim.upb.ro

PII: S0003-2670(98)00777-6

vectors [19]; and molecular similarity matrices [20,21].

Usually, the MLF networks used in QSPR studies perceive information regarding the molecular structure of the chemical compounds only through the agency of the input neurons that receive a numerical representation of the chemical structure as a vector of numerical descriptors. Therefore, the topology of the neural network (the number of neurons in the input, hidden, and output layers, together with the 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 defined by Kireev [22]; the Baskin–Palyulin–Zefirov (BPZ) neural device [23]; and the MolNet defined by Ivanciuc [24]. The three aforementioned neural models use a set of rules to build the network according to the chemical structure of each molecule examined by the ANN.

In ChemNet [22], the input and hidden neurons represent the atoms from the molecule presented to the network, while the connections between the input and hidden layers are set according to the graph distance matrix of the molecule. Atomic invariants, like the number of attached hydrogen atoms, are used as input data.

The BPZ neural device [23] contains three blocks: a sensor field, a set of eyes, and a brain. The sensor field is a matrix of neurons that corresponds to the chemical structure of the molecule presented to the neural device. In this field, the structural information that corresponds to the characteristics of atoms and bonds are encoded in signals sent to the eyes of the neural device. Each eye perceives specific information about the chemical structure of a molecule by receiving signals from selected regions of the sensor field. In an eye, the receptors receive signals from the sensor field, process the structural information, and send signals to the collectors. The signals from the collectors are transformed in the brain. The brain, that has the structure of a usual feedforward multi-layer neural network, offers at its output the computed values of the molecular properties modeled with the neural device.

In the present investigation, we use MolNet [24], a neural network related to ChemNet which is designed for the computation of molecular properties of organic compounds using atomic descriptors as input structural parameters. MolNet is applied for the computation of alkane heat enthalpies, giving good results, in calibration and prediction.

2. MolNet description

In the description and application of MolNet, we will use the molecular graph theory concepts [25,26]. 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 molecular graphs. By removing all hydrogen atoms from the chemical formula of a chemical compound containing covalent bonds, one obtains the hydrogen-depleted (or hydrogen-suppressed) molecular graph of that compound, whose vertices correspond to non-hydrogen atoms and whose edges correspond to covalent bonds. In the particular case of hydrocarbons, the vertices of the molecular graph denote carbon atoms and the edges denote carbon–carbon bonds. In this study, the expressions ‘molecular graph’ and ‘molecule’, ‘vertex’ and ‘atom’, ‘edge’ and ‘bond’ are used interchangeably.

MolNet is a multi-layer feedforward neural network that can be used to compute molecular properties on the basis of chemical structure descriptors. A specific feature of MolNet is that it changes the number of neurons and the connections between them according to the chemical structure of each molecule presented to the network. Molecules are represented by the corresponding hydrogen-suppressed molecular graph. Each non-hydrogen atom in a molecule has a corresponding unit in the input and hidden layers. The number of units in the input and hidden layers is equal to the number of vertices in the molecular graph. The output layer has only one unit, providing the calculated value of the molecular property under investigation. The network has a bias unit, connected to the hidden and output units.

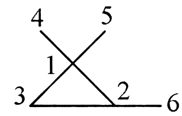
As mentioned before, with each molecule presented to the network the number and significance of the input and hidden units changes. The connections between the input and hidden layers correspond to the bonding relations between pairs of atoms. The MolNet connections between pairs of atoms exhibiting the same bonding pattern have identical weights. The bonding relationship used to generate MolNet-1 considers the type of atoms and bonds on the shortest path between a pair of atoms. An input neuron corresponding to an atom i is connected to the hidden neuron corresponding to the same atom i; these connections are classified according to the chemical nature of the atoms. Hidden input connections corresponding to the same bonding relationship between two atoms, either in the same molecule or in different molecules, have identical weights.

MolNet contains only one hidden layer. The connections between the hidden and output layers are classified according to the types of atoms as represented by the hidden neurons. The partitioning of the atoms in atom types considers the chemical nature, the hybridization state and the degree. Hidden units corresponding to atoms of the same type are connected to the output unit with identically weighted connections. The bias unit is connected to each unit in the hidden layer by 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. Also, the bias neuron is connected with the output neuron.

For a molecule with N non-hydrogen atoms, there are N^2 connections between the input and hidden layers, N connections between the hidden and output layers, 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 here. This implies that for MolNet the number of adjustable parameters is much smaller than the number of connections.

When a molecule is presented to MolNet, input unit i receives a signal representing an atomic property computed for the atom i of the respective molecular graph. Any vertex invariant computed from the structure of the molecular graph can be used as input for MolNet.

For alkanes, the bonding relationship between two atoms depends only on the number of carbon–carbon single bonds between them. In this case, the connection types between the input and hidden layers (the IH connections) are determined from the topological distance between the carbon atoms. As an example of MolNet generation, we consider 1,1,2-trimethylcyclopropane (**1**) whose molecular graph is presented in



1

Fig. 1. The molecular graph of 1,1,2-trimethylcyclopropane.

Fig. 1. In the molecular graph of an alkane, the topological distance between two vertices i and j, d_{ij}, is equal to the number of edges (corresponding to carbon–carbon single bonds) on the shortest path between the vertices i and j [25,26]. Distances d_{ij} are elements of the distance matrix of a molecular graph G, $\mathbf{D}=\mathbf{D}(G)$. The distance matrix of the molecular graph of **1**, $\mathbf{D}(\mathbf{1})$, computed with the Floyd–Warshall algorithm [27], is presented in Table 1.

Molecule **1** contains six carbon atoms. Each carbon atom from the molecular graph **1** has a corresponding unit with the same label in the input and hidden layers of MolNet, as presented in Fig. 2(a–d). The distance matrix of 1,1,2-trimethylcyclopropane has four classes of topological distances: six distances $\mathbf{d}=0$; six for $\mathbf{d}=1$; seven for $\mathbf{d}=2$; and two for $\mathbf{d}=3$. The four types of graph distances correspond to four IH connection types or parameters that are adjusted during the learning phase. As an example of identical IH connections, consider the two following pairs of atoms: four and six; five and six. The graph distance between the atoms in the above pairs is three, as can be seen from Fig. 1 and Table 1. Therefore, for the alkane **1** there are four IH connections with identical weights between the above two pairs of atoms, as depicted in Fig. 2(d): from input neuron 4 to hidden neuron 6, from input neuron 5 to hidden neuron 6, from input neuron 6 to hidden neuron 4, and from input neuron 6 to hidden neuron 5. These four con-

Table 1
The distance matrix of the molecular graph of 1,1,2-trimethylcyclopropane (**1**)

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

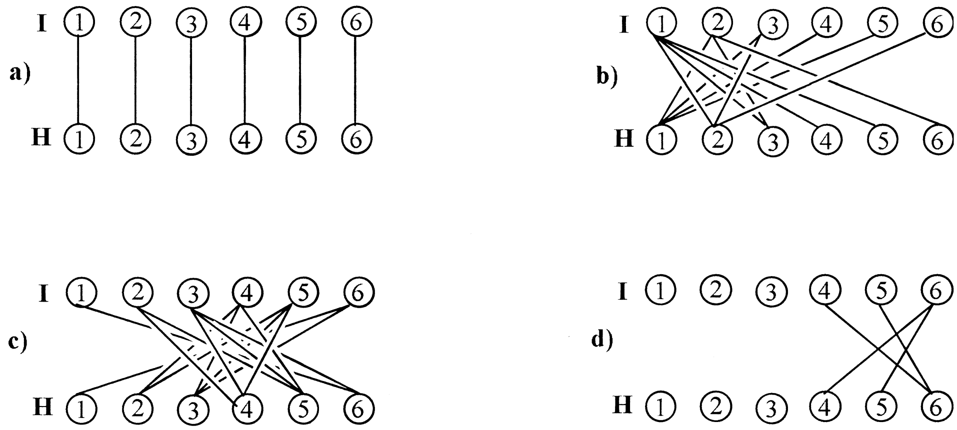


Fig. 2. The structure of the MolNet connections between the input (I) and hidden (H) layers for 1,1,2-trimethylcyclopropane; each neuron corresponds to the carbon atom with the same label from Fig. 1. The connections between atoms with the same label are presented in (a); the connections between atoms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

nections have an identical weight and correspond to the parameter for two carbon atoms situated at distance three. The four classes of identical IH connections are presented in Fig. 2(a–d): all six connections corresponding to the distance zero (Fig. 2(a)) have identical weights because all non-hydrogen atoms are carbon atoms; Fig. 2(b) presents the 12 connections between atoms situated at distance one; the 14 connections from Fig. 2(c) correspond to atoms situated at distance two; there are four connections corresponding to carbon atoms separated by three bonds, as presented in Fig. 2(d).

For alkanes, the connections between the hidden and output layers (the HO connections) are separated in classes according to the degree of the carbon atoms. Hidden units representing atoms with identical degrees, either in the same molecule or in different molecules, have HO connections with identical weights. As can be seen from Fig. 1 and Table 1, the molecular graph of 1,1,2-trimethylcyclopropane has three atoms with degree one, and one atom with degree two, three, and four, respectively. This partitioning of atoms according to their degrees gives for molecule **1** a total of four types of HO connections (adjustable weights). The connections between the bias unit and the units in the hidden layer (the BH connections) are classified according to the same rules as were used for the HO connections, giving for molecule **1** four types of BH connections. The structure of BH and HO connections is presented in Fig. 3(a–d). The bias and output connections corresponding to atoms with the degree equal to unity are presented in Fig. 3(a). The three atoms with degree one (namely atoms 4, 5, and 6) have BH connections with identical weights. As a consequence, the signal received from the bias by the units 4, 5, and 6 is identical. Also, their connections to the output unit have identical weights. The signal sent to the output unit by units 4 and 5 is identical, because the two units correspond to topologically equivalent atoms having identical connections from the input units. Although the unit 6 has an HO connection of the same type, its signal send to the output unit is different because its connections from the input layer are different. The BH and HO connections corresponding to the atoms with degree 2, 3, and 4 are depicted in Fig. 3(b–d), respectively. The bias unit has also a connection to the output unit (BO connection).

The signal flow through MolNet is briefly explained below. Consider as input atomic descriptor the number of hydrogen atoms attached to a given carbon atom. The hydrogen number vector of 1,1,2-trimethylcyclopropane is **HN**(**1**)=(0,1,2,3,3,3). After the generation of MolNet as presented here and in Figs. 2 and 3, the vector of atomic descriptors is entered into the network through the neurons in the input layer. As is usual with neural networks, the input and output values are scaled. Each input neuron receives the number of attached hydrogens for the atom with the same label from the molecular graph. In the case of

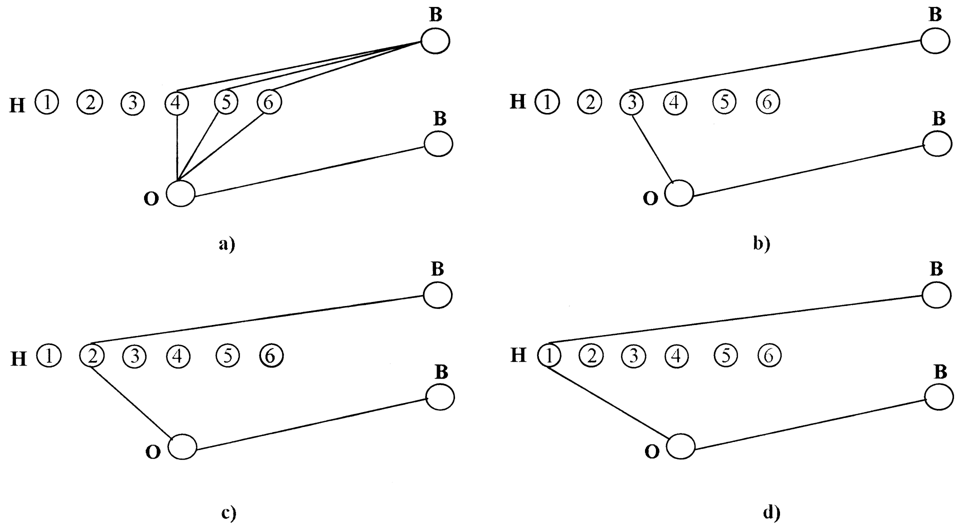


Fig. 3. The structure of the MolNet connections between the hidden (H) and output (O) layers for 1,1,2-trimethylcyclopropane; the bias neuron is labeled with B. The connections to/from atoms with the degree 1,2,3 and 4 are presented in (a), (b), (c) and (d), respectively.

1,1,2-trimethylcyclopropane, input Neuron 1 receives the value zero, neuron 2 the value one, neuron 3 the value two, and neurons 4, 5, and 6 all receive the same value – three. In this way, the **HN** vector is entered into MolNet and is then propagated through the network following the rules of MLF **ANN**. The signal that is offered by the output neuron represents the computed value of 1,1,2-trimethylcyclopropane enthalpy. During the calibration phase this value is compared with the experimental one and the adjustable parameters are modified according to the backpropagation algorithm or any other suitable non-linear optimization method.

Considering all connection types present in the case of 1,1,2-trimethylcyclopropane, the total number of adjustable weights is: 4 (IH connections) + 4 (BH connections) + 4 (HO connections) + 1 (BO connection) = 13. MolNet can contain more parameters, corresponding to bonding relationships that are not present in this example, but are found in one or more molecules from the calibration set. A certain connection type has the same weight in all molecules that contain it.

MolNet is an MLF neural network and its use involves two phases: a calibration (learning) and a prediction phase. In the calibration phase, the weights of all connection types are optimized (adjusted) in order to estimate with high precision the investigated molecular property. The optimization of the weights can be made with a non-linear optimization algorithm selected from the large set used in neural network calibration. One can use global optimization algorithms (random search, simulated annealing or genetic algorithms), the simplex algorithm, direction-set methods (Powell's method), methods that require the computation of the first derivatives like 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 [1].

In MolNet, the weights are adjusted after the presentation of each molecule. In a molecule, all connections from the same class are adjusted with the same value obtained by a summation of individual gradients and application of the usual backpropagation with momentum equation. If a connection type is absent from a certain molecule its value does not change after the presentation of that molecule to the network. The calibration phase stops when a certain convergence criterion is satisfied.

In the prediction phase, MolNet computes the molecular properties with the weights determined in the calibration phase. If the set of molecules used in the prediction phase contains bonding relationships (connection types) that are absent in the molecules used in the calibration phase, these bonding relationships are neglected in predicting the molecular property.

3. MolNet operation

3.1. Data set

MolNet is tested in a QSPR investigation for the estimation of alkane enthalpies. As it is important to determine the MolNet prediction power, the patterns are separated into a calibration (learning) set and a prediction (test) set. In this way, it is possible to determine the MolNet precision in predicting the enthalpy for alkanes that are not used in the calibration of the neural model. The calibration set contains 109 alkanes, and the prediction set 25 alkanes between C_6 and C_{10}. The structure and experimental enthalpies of the alkanes used in the present investigation are taken from the literature [8] and are reported in Tables 2 and 3. The separation of the alkanes in the calibration and the prediction sets is identical with that used in Ref. [8].

Table 2
Alkanes used in MolNet calibration, experimental enthalpies, and calibration residuals for MolNet network using **DEG**, **DS**, and **RDS** input atomic descriptors

No.	Hydrocarbon	Enthalpy H_f at 300 K (kJ/mol)			
		exp.	**DEG** res.	**DS** res.	**RDS** res.
1	2	3	4	5	6
1	3-methylpentane	26.32	−0.94	−0.77	0.41
2	2,2-dimethylbutane	25.40	−1.11	−0.56	−0.54
3	2,3-dimethylbutane	24.77	−0.70	−1.63	−0.13
4	3-methylhexane	30.71	−0.15	0.28	0.32
5	3-ethylpentane	31.71	0.68	1.22	1.77
6	2,2-dimethylpentane	29.50	−0.70	−0.31	0.10
7	2,3-dimethylpentane	28.62	−0.79	−1.30	−0.12
8	2,4-dimethylpentane	29.58	−0.41	−0.79	0.72
9	3,3-dimethylpentane	29.33	−0.55	−0.35	0.05
10	2,2,3-trimethylbutane	28.28	0.10	−0.63	0.15
11	*n*-octane	38.12	1.66	1.58	−0.26
12	2-methylheptane	35.82	1.07	0.77	−0.29
13	3-methylheptane	35.31	0.68	0.52	−0.64
14	2,4-dimethylhexane	33.76	0.26	−0.25	−0.07
15	2,5-dimethylhexane	33.39	0.67	−0.66	−0.74
16	3,3-dimethylhexane	33.43	−0.32	−0.36	−0.57
17	3,4-dimethylhexane	32.47	−1.10	−1.33	−1.34
18	3-ethyl-2-methylpentane	34.31	0.65	0.50	0.87
19	3-ethyl-3-methylpentane	33.26	−0.60	−0.53	−0.49
20	2,2,3-trimethylpentane	32.13	−0.26	−1.03	−0.47
21	2,2,4-trimethylpentane	32.55	−0.18	−1.04	0.16
22	2,3,3-trimethylpentane	32.17	−0.31	−0.98	−0.41
23	2,3,4-trimethylpentane	32.55	0.14	−0.84	0.35
24	2,2,3,3-tetramethylbutane	31.84	0.60	−0.61	0.03
25	2-methyloctane	40.42	0.98	0.43	−0.58
26	3-methyloctane	39.92	0.72	−0.11	−0.93
27	3-ethylheptane	40.71	1.93	1.02	0.11
28	4-ethylheptane	40.50	1.66	0.70	−0.11
29	2,2-dimethylheptane	38.83	1.32	0.70	0.20
30	2,3-dimethylheptane	37.82	0.16	−0.44	−1.09
31	2,4-dimethylheptane	38.16	0.37	−0.02	−0.49
32	2,5-dimethylheptane	37.53	0.15	−0.78	−1.36

Table 2 (*Continued*)

No.	Hydrocarbon	Enthalpy H_f at 300 K (kJ/mol)			
		exp.	**DEG** res.	**DS** res.	**RDS** res.
1	2	3	4	5	6
33	2,6-dimethylheptane	37.99	0.16	−0.39	−0.83
34	3,3-dimethylheptane	38.20	0.68	0.20	−0.46
35	3,4-dimethylheptane	37.02	−0.70	−1.12	−1.75
36	3,5-dimethylheptane	38.07	0.40	−0.17	−0.71
37	3-ethyl-3-methylhexane	37.36	−0.32	−0.73	−1.11
38	4-ethyl-2-methylhexane	39.25	1.74	1.01	0.57
39	2,2,4-trimethylhexane	36.61	0.32	−0.32	−0.39
40	2,2,5-trimethylhexane	36.36	0.69	−0.49	−0.67
41	2,3,3-trimethylhexane	36.28	−0.17	−0.77	−0.86
42	2,3,4-trimethylhexane	36.86	0.21	−0.34	−0.24
43	2,3,5-trimethylhexane	36.02	−0.18	−1.10	−1.05
44	2,4,4-trimethylhexane	36.44	0.02	−0.67	−0.52
45	3,3,4-trimethyhexane	35.98	−0.65	−1.16	−1.16
46	3,3-diethylpentane	38.37	0.68	0.48	0.37
47	3-ethyl-2,2-dimethylpentane	36.15	−0.14	−0.79	0.75
48	3-ethyl-2,3-dimethylpentane	36.61	−0.07	−0.54	−0.25
49	2,2,3,3-tetramethylpentane	35.86	0.38	−0.64	−0.09
50	2,2,3,4-tetramethylpentane	35.06	−0.34	−1.44	−0.65
51	2,3,3,4-tetramethylpentane	36.23	0.55	−0.37	0.50
52	3-ethyloctane	45.31	2.10	0.44	1.04
53	4-ethyloctane	45.10	2.14	−0.13	0.57
54	2,2-dimethyloctane	43.43	1.36	0.57	1.58
55	2,4-dimethyloctane	42.76	0.55	0.66	0.27
56	2,5-dimethyloctane	41.92	−0.16	−0.34	−0.59
57	3,4-dimethyloctane	41.80	−0.29	−0.82	−0.68
58	3,5-dimethyloctane	42.47	0.43	−0.01	0.02
59	3,6-dimethyloctane	41.63	−0.37	−1.07	−0.81
60	4,4-dimethyloctane	42.30	0.77	−0.12	0.15
61	4,5-dimethyloctane	41.51	−0.47	−1.00	−1.06
62	4-*n*-propylheptane	44.85	2.09	−0.69	0.23
63	4-isopropylheptane	43.10	1.48	0.04	0.31
64	2-methyl-3-ethylheptane	43.30	1.59	0.58	0.82
65	2-methyl-4-ethylheptane	43.64	1.92	0.56	1.00
66	3-methyl-4-ethylheptane	42.47	0.72	−0.54	−0.14
67	3-methyl-5-ethylheptane	43.35	1.62	0.59	0.86
68	2,2,3-trimethylheptane	41.30	0.74	0.21	0.91
69	2,2,4-trimethylheptane	41.05	0.73	0.17	0.69
70	2,2,5-trimethylheptane	40.50	0.31	−0.75	0.14
71	2,2,6-trimethylheptane	40.96	0.32	−0.41	0.92
72	2,3,3-trimethylheptane	41.00	0.37	0.15	0.55
73	2,3,4-trimethylheptane	40.96	0.07	0.07	0.16
74	2,3,5-trimethylheptane	40.12	−0.62	−0.75	−0.68
75	2,3,6-trimethylheptane	39.75	−1.08	−1.07	−0.84
76	2,4,4-trimethylheptane	40.54	0.31	0.11	0.05
77	2,4,5-trimethylheptane	39.98	−0.71	−0.77	−0.75
78	2,4,6-trimethylheptane	41.13	0.25	0.71	0.65
79	2,5,5-trimethylheptane	40.54	0.26	−0.32	0.23
80	3,3,5-trimethylheptane	40.46	0.00	−0.34	−0.12
81	3,4,4-trimethylheptane	40.08	−0.36	−0.71	−0.60
82	3,4,5-trimethylheptane	41.14	0.21	0.09	0.21

Table 2 (*Continued*)

No.	Hydrocarbon	Enthalpy H_f at 300 K (kJ/mol)			
		exp.	**DEG** res.	**DS** res.	**RDS** res.
1	2	3	4	5	6
83	2-methyl-3-isopropylhexane	40.46	−0.09	−0.50	−0.53
84	3,3-diethylhexane	42.43	1.04	−0.07	0.49
85	3,4-diethylhexane	43.68	1.98	0.72	1.23
86	2,2-dimethyl-3-ethylhexane	40.50	0.33	−0.50	−0.26
87	2,2-dimethyl-4-ethylhexane	41.92	1.69	0.74	1.25
88	2,3-dimethyl-3-ethylhexane	40.71	0.19	−0.28	0.05
89	2,3-dimethyl-4-ethylhexane	42.43	1.68	1.25	1.51
90	2,4-dimethyl-4-ethylhexane	40.29	−0.09	−0.57	−0.28
91	3,3-dimethyl-4-ethylhexane	39.92	−0.45	−1.04	−0.79
92	3,4-dimethyl-4-ethylhexane	40.42	−0.21	−0.65	−0.25
93	2,2,3,3-tetramethylhexane	40.00	0.66	0.42	0.82
94	2,2,3,4-tetramethylhexane	39.25	−0.28	−0.51	−0.14
95	2,2,3,5-tetramethylhexane	39.54	0.31	0.05	0.52
96	2,2,4,5-tetramethylhexane	38.83	−0.22	−0.69	−0.28
97	2,2,5,5-tetramethylhexane	39.37	0.80	−0.45	0.63
98	2,3,3,4-tetramethylhexane	40.04	0.28	0.17	0.62
99	2,3,3,5-tetramethylhexane	39.16	−0.13	−0.22	0.01
100	2,3,4,4-tetramethylhexane	38.87	−0.74	−0.90	−0.53
101	2,3,4,5-tetramethylhexane	40.71	0.91	0.99	1.28
102	3,3,4,4-tetramethylhexane	39.87	0.35	0.03	0.49
103	2,4-dimethyl-3-isopropylpentane	39.25	−0.25	−0.16	0.00
104	2-methyl-3,3-diethylpentane	39.25	−1.11	−1.46	−1.09
105	2,2,3-trimethyl-3-ethylpentane	38.41	−0.83	−1.28	−0.86
106	2,2,4-trimethyl-3-ethylpentane	38.87	−0.16	−0.56	−0.42
107	2,3,4-trimethyl-3-ethylpentane	38.58	−1.08	−1.23	−0.65
108	2,2,3,3,4-pentamethylpentane	38.62	0.08	−0.43	0.44
109	2,2,3,4,4-pentamethylpentane	38.81	0.87	0.12	0.55

3.2. Number of adjustable parameters

MolNet has a variable topology that changes with each molecule presented to the network. Therefore, the number of adjustable parameters (connections) depends on the structure of the molecules from the learning set. In view of the learning set of 109 alkanes having the maximum graph distance of seven between two carbon atoms, there are eight IH connection classes. The degree of the carbon atoms is between one and four, giving four HO connection types and four BH adjustable weights. The total number of adjustable connections 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 6.4, a value that is high enough to eliminate the danger of overfitting.

3.3. Input data

In the present study, three types of atomic topological indices were tested as input structural descriptors. Each structural descriptor was investigated in a separate test, independent of the remaining two descriptors. The three atomic topological descriptors entering in the input neurons are the degree **DEG** [25,26], the distance sum **DS** [28,29], and the reciprocal distance sum **RDS** [30,31]. The degree of a vertex i from a molecular graph G is the sum over row i (or column i) of its adjacency matrix $\mathbf{A}=\mathbf{A}(G)$ [25,26]:

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

The degree vector of 1,1,2-trimethylcyclopropane is $\mathbf{DEG}(\mathbf{1})$=(4,3,2,1,1,1). As presented in the MolNet

Table 3
Alkanes used in MolNet prediction, experimental enthalpies, and prediction residuals for MolNet network using **DEG**, **DS**, and **RDS** input atomic descriptors

No.	Hydrocarbon	Enthalpy H_f at 300 (kJ/mol)			
		Exp.	**DEG** res.	**DS** res.	**RDS** res.
1	2-methylpentane	26.61	−1.03	−0.58	0.47
2	*n*-heptane	33.56	1.32	2.32	0.83
3	2-methylhexane	31.21	0.68	0.76	0.53
4	4-methylheptane	35.06	0.11	0.31	−0.66
5	3-ethylhexane	36.07	1.28	1.47	0.61
6	2,2-dimethylhexane	34.23	0.94	0.50	0.02
7	2,3-dimethylhexane	33.05	−0.27	−0.73	−0.90
8	4-methyloctane	39.71	0.62	−0.11	−1.12
9	4,4-dimethylheptane	37.53	−0.02	−0.44	−1.04
10	3-ethyl-2-methylhexane	38.70	1.15	0.56	0.04
11	3-ethyl-4-methylhexane	38.07	0.33	−0.06	−0.46
12	2,2,3-trimethylhexane	36.61	0.28	−0.27	−0.53
13	3-ethyl-2,4-dimethylpentane	36.07	−0.52	−1.00	−0.61
14	2,2,4,4-tetramethylpentane	36.44	1.43	0.19	0.80
15	2,3-dimethyloctane	42.43	0.06	0.03	0.03
16	2,6-dimethyloctane	42.09	−0.35	−0.41	−0.25
17	2,7-dimethyloctane	42.58	−0.19	0.71	0.20
18	3,3-dimethyloctane	42.80	1.00	−0.01	0.98
19	2-methyl-5-ethylheptane	42.93	1.27	0.15	0.55
20	3-methyl-3-ethylheptane	42.13	0.58	−0.41	0.19
21	4-methyl-3-ethylheptane	42.51	0.83	−0.10	0.12
22	4-methyl-4-ethylheptane	41.46	0.07	−1.23	−0.73
23	3,3,4-trimethylheptane	40.46	−0.17	−0.41	−0.11
24	2,5-dimethyl-3-ethylhexane	41.34	0.91	0.37	0.43
25	2,2,4,4-tetramethylhexane	40.29	1.51	0.96	1.07

description section, when the **DEG** vector of 1,1,2-trimethylcyclopropane is used as input descriptor, the network has the topology presented in Figs. 2 and 3. The element *i* of the **DEG** vector enters in the input neuron *i* that corresponds to atom *i* from the molecular graph, i.e. the value four goes to input neuron 1, the value three enters in the input neuron 2, the value two goes to input neuron 3, while neurons 4, 5, and 6 all receive the same input value one.

The distance sum of the atom *i* from a molecular graph *G*, $\mathbf{DS}_i=\mathbf{DS}_i(G)$, is equal to the sum of the elements in the *i*th row (or *i*th column) of the distance matrix of *G*, $\mathbf{D}=\mathbf{D}(G)$ [28,29]:

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

The distance sum vector of 1,1,2-trimethylcyclopropane is **DS**(1)=(6,7,8,10,10,11).

The reciprocal distance sum of the atom *i* from a molecular graph *G*, $\mathbf{RDS}_i=\mathbf{RDS}_i(G)$, is defined as follows [30,31]:

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

where $\mathbf{RD}=\mathbf{RD}(G)$ is the reciprocal distance matrix of *G*. The reciprocal distance sum vector of **1** is **RDS**(**1**)=(4.50000, 4.00000, 3.50000, 2.83333, 2.83333, 2.66667).

3.4. The learning method

The training (calibration) of MolNet is accomplished with the help of standard backpropagation with momentum method [1], until convergence is obtained, i.e. the correlation coefficient between experimental and calculated alkane enthalpy values improves by $<10^{-5}$ in 100 epochs. A complete presentation of the 109 alkanes in the learning set corresponds to one epoch. Random values between -0.1 and 0.1 are used as initial weights. The connection weights are updated after the presentation of each molecule. Both, the learning rate and momentum values are maintained constant during the training phase. The learning process is sensitive to the learning rate and momentum values, and small learning rates are used, equal to 0.05, for both the hidden and output layers. The momentum is set between 0.30 and 0.05 for all activation functions used in this study. In all the cases, the learning phase stops after a few hundreds epochs and the results are not greatly influenced by the initial random set of weights.

3.5. Activation functions

The most commonly used activation function in neural network studies is the sigmoid that takes values between zero and one. The main drawback of the sigmoid is that for large negative arguments its value is close to zero, and practice demonstrated that learning with the backpropagation algorithm is difficult in such conditions. To overcome this deficiency of the sigmoid a related function, the hyperbolic tangent (tanh) which takes values between -1 and $+1$, is used in the present investigation. The two activation functions, the sigmoid and tanh, are very flat when the absolute value of the argument is >10. In such situations, the derivative of the activation function has an extremely small value, leading to a poor sensitivity of the sigmoid and the tanh functions to large positive or negative arguments. This is one of the causes of the very slow rates of convergence during the training of neural networks with algorithms that use the derivative of the activation function (e.g. the backpropagation algorithm). A linear output activation function overcomes the problems of the sigmoidal and tanh functions, and we use it for the output layer. A new type of activation function is the symmetric logarithmoid [32,33], defined by the formula: $Act(z)=\mathrm{sign}(z)\ln(1+|z|)$. 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. We use the symlog function for the output layer of units.

3.6. Preprocessing of the data

Each component of the input (**DEG**, **DS**, or **RDS** vector) and output (representing the target enthalpy value) patterns is linearly scaled between -0.9 and 0.9. For the tanh activation function, the scaling is required by the range of values of the function, while for the unbounded functions (linear and symlog) the experience showed that a linear scaling improves the learning process.

3.7. Performance indicators

The performances of MolNet are evaluated both for the network calibration and prediction. The quality of MolNet calibration is estimated by comparing the calculated alkane enthalpies at the end of the calibration phase ($H_{\mathrm{f\ cal}}$) with the target (experimental) values ($H_{\mathrm{f\ exp}}$), while the predictive quality is estimated with a set of alkanes that were not used in the calibration phase by comparing the predicted ($H_{\mathrm{f\ pr}}$) and experimental values. In order to compare the performance of different MolNet networks, we use the correlation coefficient r and the standard deviation s of the linear correlation between experimental and calculated (in calibration or prediction) enthalpies: $H_{\mathrm{f\ exp}}=A+B\times H_{\mathrm{f\ cal/pr}}$.

4. MolNet computation of alkane enthalpies

For the usual MLF neural network, the number of hidden layers and the number of units in each hidden layer is determined during the calibration phase. Since MolNet has a topology that depends on the structure of the molecule presented to the network, the optimization of the number of hidden units is no longer a problem. MolNet accepts as input any atomic property, and the calibration and prediction results depend on the atomic invariant used to feed the network. As presented in the previous section, this study investigates 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.

Table 4 contains the calibration and prediction results obtained when MolNet parameters are optimized using as input data the **DEG** atomic descriptor.

Table 4
MolNet calibration and prediction results for the computation of alkane enthalpies using **DEG** input atomic descriptor. The table reports the number of training epochs, the hidden and output momentum (α_h and α_0), the output activation function, the calibration and prediction standard deviation (s_{cal} and s_{pr}) and correlation coefficient (r_{cal} and r_{pr}). All the networks were provided with the tanh hidden activation function

Epoch	Hidden momentum (α_h)	Output activation function	Output momentum (α_0)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1900	0.30	linear	0.30	0.82	0.982	0.72	0.986
1700	0.30	linear	0.15	0.77	0.984	0.72	0.986
1900	0.30	linear	0.10	0.76	0.985	0.72	0.986
2000	0.30	linear	0.05	0.75	0.985	0.72	0.986
1700	0.15	linear	0.05	0.75	0.985	0.71	0.987
1900	0.10	linear	0.05	0.75	0.985	0.70	0.987
1800	0.05	linear	0.05	0.75	0.985	0.70	0.987
1700	0.15	linear	0.15	0.77	0.984	0.71	0.987
2000	0.15	linear	0.10	0.76	0.985	0.71	0.987
1800	0.10	linear	0.10	0.76	0.985	0.71	0.987
1100	0.30	symlog	0.30	1.01	0.973	1.06	0.970
800	0.30	symlog	0.15	0.91	0.978	0.87	0.980
700	0.30	symlog	0.10	0.91	0.978	0.87	0.980
700	0.30	symlog	0.05	0.90	0.978	0.87	0.980
900	0.15	symlog	0.05	0.89	0.979	0.85	0.981
900	0.10	symlog	0.05	0.89	0.979	0.85	0.981
900	0.05	symlog	0.05	0.89	0.979	0.85	0.981
800	0.15	symlog	0.15	0.90	0.978	0.85	0.981
900	0.15	symlog	0.10	0.89	0.978	0.85	0.981
900	0.10	symlog	0.10	0.89	0.979	0.85	0.981
600	0.30	tanh	0.30	0.96	0.976	0.78	0.984
1000	0.30	tanh	0.15	1.05	0.971	1.02	0.972
1500	0.30	tanh	0.10	0.93	0.977	0.78	0.984
800	0.30	tanh	0.05	0.91	0.978	0.78	0.984
1200	0.15	tanh	0.05	0.90	0.978	0.77	0.984
1300	0.10	tanh	0.05	0.90	0.978	0.77	0.984
1300	0.05	tanh	0.05	0.90	0.978	0.76	0.984
1100	0.15	tanh	0.15	0.92	0.977	0.77	0.984
1300	0.15	tanh	0.10	0.92	0.978	0.77	0.984
1300	0.10	tanh	0.10	0.91	0.978	0.77	0.984

The calibration correlation coefficient, r_{cal}, is in the 0.971–0.985 range, and the calibration standard deviation, s_{cal}, takes values between 0.75 and 1.05. The prediction correlation coefficient, r_{pr}, takes values between 0.970 and 0.987, while the prediction standard deviation, s_{pr}, is in the 0.70–1.06 range. Overall, the best calibration and prediction results are obtained with linear output function, followed by the tanh output function. The alkane enthalpy residuals computed with **DEG** input data, a linear output function, a hidden momentum, α_h=0.10, and an output momentum, α_0=0.05, are presented in column 4 of Table 2 for calibration, and in Table 3 for prediction. This example, selected because it offers good calibration and prediction results, has the following statistical indices: r_{cal}=0.985, s_{cal}=0.75, r_{pr}=0.987, and s_{pr}= 0.70. From the whole set of 109 alkanes used in calibration, one finds 13 cases with residuals >1.5 kJ/mol, presented here together with their residuals: *n*-octane with 1.66, 3-ethylheptane with 1.93, 4-ethylheptane with 1.66, 4-ethyl-2-methylhexane with 1.74, 3-ethyloctane with 2.10, 4-ethyloctane with 2.14, 4-*n*-propylheptane with 2.09, 2-methyl-3-ethylheptane with 1.59, 2-methyl-4-ethylheptane with 1.92, 3-methyl-5-ethylheptane with 1.62, 3,4-diethylhexane with 1.98, 2,2-dimethyl-4-ethylhexane with 1.69, and 2,3-dimethyl-4-ethylhexane with 1.68. Usually, in a QSPR model an outlier is defined as the pattern

Table 5
MolNet calibration and prediction results for the computation of alkane enthalpies using **DS** input atomic descriptor. The notations are the same as in Table 4

Epoch	Hidden momentum (α_h)	Output activation function	Output momentum (α_0)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1600	0.30	linear	0.30	0.78	0.984	1.04	0.971
1700	0.30	linear	0.15	0.75	0.985	0.99	0.973
1200	0.30	linear	0.10	0.74	0.985	0.98	0.974
1700	0.30	linear	0.05	0.74	0.986	0.97	0.975
2000	0.15	linear	0.05	0.65	0.989	0.75	0.985
500	0.10	linear	0.05	0.73	0.986	0.92	0.977
1900	0.05	linear	0.05	0.72	0.986	0.92	0.977
1800	0.15	linear	0.15	0.74	0.985	0.96	0.975
1700	0.15	linear	0.10	0.66	0.989	0.75	0.985
400	0.10	linear	0.10	0.74	0.985	0.92	0.977
1100	0.30	symlog	0.30	0.87	0.980	1.13	0.965
1400	0.30	symlog	0.15	0.85	0.981	1.11	0.967
1600	0.30	symlog	0.10	0.85	0.981	1.11	0.967
1500	0.30	symlog	0.05	0.85	0.981	1.10	0.967
1600	0.15	symlog	0.05	0.84	0.981	1.08	0.969
1700	0.10	symlog	0.05	0.83	0.981	1.07	0.969
2000	0.05	symlog	0.05	0.83	0.982	1.07	0.969
1700	0.15	symlog	0.15	0.84	0.981	1.09	0.968
1400	0.15	symlog	0.10	0.84	0.981	1.08	0.968
1500	0.10	symlog	0.10	0.84	0.981	1.08	0.969
1000	0.30	tanh	0.30	0.89	0.979	1.16	0.964
1900	0.30	tanh	0.15	0.77	0.984	0.96	0.975
700	0.30	tanh	0.10	0.87	0.979	1.13	0.966
1000	0.30	tanh	0.05	0.87	0.980	1.12	0.966
1400	0.15	tanh	0.05	0.85	0.981	1.10	0.968
1700	0.10	tanh	0.05	0.85	0.981	1.09	0.968
600	0.05	tanh	0.05	0.86	0.980	1.09	0.968
1300	0.15	tanh	0.15	0.86	0.980	1.11	0.967
1400	0.15	tanh	0.10	0.86	0.980	1.10	0.967
1900	0.10	tanh	0.10	0.85	0.981	1.10	0.968

with an absolute residual three times greater than the standard deviation. In this case, when $3s_{cal}$=2.25, there is no outlier. The threshold of 1.5 kJ/mol is used only to compare the results obtained with the three input structural descriptors, namely **DEG**, **DS**, and **RDS**. The prediction residuals are small, and the greatest one is obtained for 2,2,4,4-tetramethylhexane with 1.51 kJ/mol.

The second test is made using the **DS** atomic descriptor as input data. The MolNet calibration and prediction results are presented in Table 5. In the calibration phase r_{cal} is between 0.979 and 0.989, and s_{cal} takes values between 0.66 and 0.89. In the prediction phase r_{pr} takes values between 0.964 and 0.985, and s_{pr} is in the 0.75–1.16 range. The linear output function offers the best calibration and prediction results. The alkane enthalpy residuals computed with **DS** input data, a linear output function, a hidden momentum, α_h=0.15, and an output momentum, α_o=0.10, are presented in column 5 of Table 2 for calibration, and in Table 3 for prediction. This example has the following statistical indices: r_{cal}=0.989, s_{cal}=0.66, r_{pr}=0.985, and s_{pr}=0.75. The calibration results are better than those obtained with the **DEG** input data, with only two alkanes with an absolute residual >1.5 kJ/mol, namely 2,3-dimethylbutane with −1.63, and *n*-octane with 1.58. The largest prediction error is obtained for *n*-heptane with a residual of 2.32 kJ/mol. The residuals for the other alkanes in the prediction set are <1.5 kJ/mol.

Table 6 presents the calibration and prediction results obtained when the MolNet parameters are

Table 6
MolNet calibration and prediction results for the computation of alkane enthalpies using **RDS** input atomic descriptor. The notations are the same as in Table 4

Epoch	Hidden momentum (α_h)	Output activation function	Output momentum (α_0)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1600	0.30	linear	0.30	0.74	0.985	0.69	0.987
1700	0.30	linear	0.15	0.72	0.986	0.67	0.988
1800	0.30	linear	0.10	0.71	0.987	0.65	0.989
1700	0.30	linear	0.05	0.71	0.986	0.66	0.988
1200	0.15	linear	0.05	0.70	0.987	0.65	0.989
1600	0.10	linear	0.05	0.70	0.987	0.65	0.989
2000	0.05	linear	0.05	0.72	0.986	0.68	0.988
1900	0.15	linear	0.15	0.74	0.985	0.70	0.987
1600	0.15	linear	0.10	0.71	0.987	0.66	0.988
600	0.10	linear	0.10	0.75	0.985	0.80	0.983
700	0.30	symlog	0.30	0.88	0.979	0.89	0.979
1900	0.30	symlog	0.15	0.85	0.981	0.84	0.981
500	0.30	symlog	0.10	0.86	0.980	0.87	0.980
1800	0.30	symlog	0.05	0.84	0.981	0.85	0.981
2000	0.15	symlog	0.05	0.83	0.982	0.84	0.981
1800	0.10	symlog	0.05	0.89	0.979	0.89	0.979
500	0.05	symlog	0.05	0.84	0.981	0.85	0.981
1900	0.15	symlog	0.15	0.84	0.981	0.85	0.980
2000	0.15	symlog	0.10	0.84	0.981	0.83	0.982
1800	0.10	symlog	0.10	0.84	0.981	0.83	0.982
1400	0.30	tanh	0.30	0.89	0.979	0.87	0.980
1900	0.30	tanh	0.15	0.85	0.981	0.81	0.983
900	0.30	tanh	0.10	0.86	0.980	0.84	0.981
1800	0.30	tanh	0.05	0.84	0.981	0.80	0.983
1100	0.15	tanh	0.05	0.85	0.981	0.83	0.982
900	0.10	tanh	0.05	0.85	0.981	0.83	0.982
1900	0.05	tanh	0.05	0.85	0.981	0.81	0.982
1100	0.15	tanh	0.15	0.86	0.981	0.84	0.981
1000	0.15	tanh	0.10	0.85	0.981	0.83	0.982
1700	0.10	tanh	0.10	0.84	0.981	0.93	0.977

optimized using the **RDS** atomic descriptor as input data. In the calibration phase, r_{cal} is between 0.979 and 0.987, and s_{cal} takes values between 0.70 and 0.89. In the prediction phase, r_{pr} takes values between 0.977 and 0.989, and s_{pr} lies in the 0.65–0.93 range. The statistical indices obtained with the three output activation functions and reported in Table 6 show again that the linear output function gives the best results. The enthalpy residuals computed with a linear output function, a hidden momentum, α_h=0.10, and an output momentum, α_o=0.05, are presented in column 6 of Tables 2 and 3. This example has the following statistical indices: r_{cal}=0.987, s_{cal}=0.70, r_{pr}=0.989, and s_{pr}=0.65. In the calibration set of 109 alkanes, there are four cases with absolute residuals >1.5 kJ/mol, presented here with their corresponding residuals: 3-ethylpentane with 1.77, 3,4-dimethylheptane with −1.75, 2,2-dimethyloctane with 1.58, and 2,3-dimethyl-4-ethylhexane with 1.51. The largest prediction residual is obtained for 4-methyloctane with −1.12 kJ/mol.

The MolNet calibration and prediction results obtained with the three structural descriptors are good, with no case of poorly computed alkane enthalpy. In calibration, the **DEG** descriptor gives the largest number of alkanes with a residual >1.5 kJ/mol. The **DS** descriptor gives the best calibration results, while the **RDS** descriptor offers the best prediction results. We have to mention here that, as is also apparent from the definition, **DEG** is a very simple atomic descriptor which considers only the number of neighbors of an atom in the molecular graph. The **DS** and **RDS** descriptors are more complex, reflecting the influence of atoms situated at greater topological distance.

MolNet has a much lower number of adjustable parameters, compared with a usual MLF NN that has the same structure, i.e. the same number of neurons in each layer and the same number of connections between neurons. In the example presented in this paper, 1,1,2-trimethylcyclopropane, MolNet has 13 adjustable parameters. An MLF network with the same structure, i.e. with six input neurons, six hidden neurons, and one output neuron, has 49 connections (adjustable parameters). The lower number of adjustable parameters for MolNet is an important practical advantage, because the learning is fast and usually takes a few hundreds iterations, as can be seen from Tables 4–6.

There is no systematic deviation for certain types of alkanes, small or large, linear or highly branched. An inspection of the residuals shows that, generally, for a certain alkane with a large residual computed with one of three descriptors the residuals computed with the other two descriptors are small. For example, the **DEG** calibration residuals of 3-ethyloctane, 4-ethyloctane, and 4-*n*-propylheptane are fairly large, as presented here. For the same three alkanes, the **DS** and **RDS** calibration residuals are small, showing that there is no special trend in these cases. The only difference is in the way the molecular structure is reflected in the three atomic descriptors. Other atomic descriptors computed from the molecular graph, representing the chemical structure in new ways, must be tested in order to identify the best MolNet input data.

The results presented here and in previous communications [24] show that MolNet is a neural model suited for the prediction of molecular properties. The MolNet topology reflects the chemical structure of each molecule presented to the neural network, giving a high flexibility to QSPR studies. Any atomic property can be used as input data, from topological descriptors to quantum indices.

Acknowledgements

Financial support for this work 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] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Nature 323 (1986) 533.
[2] P.D. Wasserman, Neural Computing, Van Nostrand Reinhold, New York, 1989.
[3] J. Zupan, and J. Gasteiger, Neural Networks for Chemists, VCH, Weinheim, 1993.
[4] A.B. Bulsari (Ed.), Neural Networks for Chemical Engineers, Elsevier, Amsterdam, 1995.
[5] J. Devillers (Ed.), Neural Networks in QSAR and Drug Design, Academic Press, London, 1996.
[6] D.W. Elrod, G.M. Maggiora, R.G. Trenary, J. Chem. Inf. Comput. Sci. 30 (1990) 477.
[7] D.W. Elrod, G.M. Maggiora, R.G. Trenary, Tetrahedron Comput. Methodol. 3 (1990) 163.
[8] A.A. Gakh, E.G. Gakh, B.G. Sumpter, D.W. Noid, J. Chem. Inf. Comput. Sci. 34 (1994) 832.
[9] A.T. Balaban, S.C. Basak, T. Colburn, G.D. Grunwald, J. Chem. Inf. Comput. Sci. 34 (1994) 1118.
[10] D. Cherqaoui, D. Villemin, J. Chem. Soc. Faraday Trans. 90 (1994) 97.
[11] D. Cherqaoui, D. Villemin, A. Mesbah, J.-M. Cense, V. Kvasnička, J. Chem. Soc. Faraday Trans. 90 (1994) 2015.
[12] F.R. Burden, Quant. Struct.-Act. Relat. 15 (1996) 7.
[13] O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye, J.P. Doucet, J. Chem. Inf. Comput. Sci. 36 (1996) 644.
[14] O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye, J.P. Doucet, J. Chem. Inf. Comput. Sci. 37 (1997) 587.
[15] J.H. Schuur, P. Selzer, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 334.
[16] J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, J. Chem. Inf. Comput. Sci. 36 (1996) 1030.
[17] L.H. Hall, C.T. Story, J. Chem. Inf. Comput. Sci. 36 (1996) 1004.
[18] S. Hatrik, P. Zahradnik, J. Chem. Inf. Comput. Sci. 36 (1996) 992.
[19] H. Bauknecht, A. Zell, H. Bayer, P. Levi, M. Wagener, J. Sadowski, J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996) 1205.
[20] S.-S. So, M. Karplus, J. Med. Chem. 40 (1997) 4347.
[21] S.-S. So, M. Karplus, J. Med. Chem. 40 (1997) 4360.
[22] D.B. Kireev, J. Chem. Inf. Comput. Sci. 35 (1995) 175.
[23] I.I. Baskin, V.A. Palyulin, N.S. Zefirov, J. Chem. Inf. Comput. Sci. 37 (1997) 715.
[24] O. Ivanciuc, MolNet neural network application in structure–property studies, The 23rd Chemistry Conference, 8–10 October 1997, Călimăneşti, Valcea, Romania.
[25] M.V. Diudea, O. Ivanciuc, Molecular Topology, Comprex, Cluj, Romania, 1995.
[26] O. Ivanciuc, A.T. Balaban, Graph theory in chemistry, in P.v.R. Schleyer (Ed.), Encyclopedia of Computational Chemistry, Wiley, 1998.
[27] B. Mohar, T. Pisanski, J. Math. Chem. 2 (1988) 267.
[28] A.T. Balaban, Chem. Phys. Lett. 89 (1982) 399.
[29] A.T. Balaban, Pure Appl. Chem. 55 (1983) 199.
[30] O. Ivanciuc, Rev. Roum. Chim. 34 (1989) 1361.
[31] O. Ivanciuc, T.-S. Balaban, A.T. Balaban, J. Math. Chem. 12 (1993) 309.
[32] A.B. Bulsari, H. Saxén, Neurocomputing 3 (1991) 125.
[33] A.B. Bulsari, H. Saxén, Neural Network World 4 (1991) 221.