Revue Roumaine de Chimie, **1999**, *44* (3), 299-309

ARTIFICIAL NEURAL NETWORKS APPLICATIONS. PART 10.[1]

MolNet PREDICTION OF ALKANE GIBBS ENERGIES

Ovidiu IVANCIUC

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

Received October 7, 1997

The alkane Gibbs energies are computed with the MolNet neural network. MolNet is a new type of neural network that 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. Three molecular graph invariants are used as input data for the first layer of neurons, namely the degree, the distance sum, and the reciprocal distance sum.

INTRODUCTION

Usual quantitative structure-property relationships (QSPR) and quantitative structure-activity relationships (QSAR) studies 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 class of nonlinear 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,[3] in chemical engineering,[4] and in biochemistry[5] is a result of their unique modeling features.

An important problem for the chemical applications of neural networks remains the numerical representation of the chemical structure. Various structural representations of organic compounds were used in recent 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] topological distance;[8] constitutional descriptors and topological indices;[9] numerical code;[10] 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] 3D MORSE (Molecule Representation of Structures based on Electron diffraction);[15,16] atom type electrotopological state;[17] presence of a substituent (coded with 1) or absence (coded with 0);[18] topological autocorrelation vectors.[19]

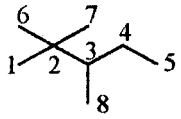
Kireev[20] proposed a new neural network, ChemNet, in which the input and hidden units represent the atoms from the molecule presented to the network, while the connections between the input and hidden layers are set according to the graph distance matrix of the molecule. In the present investigation we use a related kind of neural network, MolNet,[1] which is destined to the computation of molecular properties of organic compounds using atomic descriptors as input structural parameters. MolNet is applied for the computation of alkane Gibbs energies, giving good results both in calibration and prediction.

MolNet DESCRIPTION

MolNet is a new type of neural network[1] that changes the topology according to the structure of each molecule presented to the network. Each non-hydrogen atom in the molecule has a corresponding unit in the input and hidden layers, while the output layer has only one unit, representing the molecular property under investigation. The network has a bias unit, connected to the hidden and output units. With each molecule presented to the network the number and significance of the input and hidden units change. The connections between the input and hidden layers correspond to the bonding relationships of the atoms, with identically weighted connections for pairs of atoms exhibiting the same bonding pattern. For the network MolNet-1 the bonding relationship considers the type of atoms and bonds on the shortest path between a pair of atoms. Also, 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. Input-hidden connections corresponding to the same bonding relationship between two atoms either in the same molecule or in different molecules have identical weights.

The connections between the hidden and output layers are classified according to the partitioning of the atoms by their atomic number Z, the hybridization state and the degree. This partitioning scheme of the connections defines an atom-type contribution to the molecular property that is investigated. We have to point here that even for atoms in the same class their contribution to the molecular property depends also on the signal received from the input layer, signal that can be different for atoms in the same class. The bias neuron is connected to each neuron 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, and $N + 1$ connections from the bias neuron. Some connections may have identical weights according to the partitioning schemes described above.

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 of the molecular graph can be used as input for MolNet.



1

Fig. 1 – The molecular graph of 2,2,3-trimethylpentane **1**.

The connection types between the input and hidden layers (the IH connections) for alkanes are determined from the topological distance between the carbon atoms. We present the structure of MolNet for 2,2,3-trimethylpentane **1** whose molecular graph is presented in Figure 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. Distances d_{ij} are elements of the distance matrix of a molecular graph G, $\mathbf{D}(G)$. The distance matrix of the molecular graph of **1**, $\mathbf{D}(\mathbf{1})$, computed with the Floyd-Warshall algorithm,[21] is presented in Table 1.

Table 1

The distance matrix of the molecular graph of 2,2,3-trimethylpentane **1**

	1	2	3	4	5	6	7	8
1	0	1	2	3	4	2	2	3
2	1	0	1	2	3	1	1	2
3	2	1	0	1	2	2	2	1
4	3	2	1	0	1	3	3	2
5	4	3	2	1	0	4	4	3
6	2	1	2	3	4	0	2	3
7	2	1	2	3	4	2	0	3
8	3	2	1	2	3	3	3	0

Each carbon atom from the molecular graph **1** corresponds to a unit with the same label in the input and hidden layers of MolNet, as presented in Figure 2a-e. The distance matrix of 2,2,3-trimethylpentane has five classes of topological distances, from 0 to 4, corresponding to five IH connection types or parameters that are adjusted during the learning phase. For example, the topological distance between the pairs of atoms (1 and 5), (5 and 6), (5 and 7) is equal to 4. Therefore, for the alkane **1** there are six IH connections with identical weights between the above three pairs of atoms. These six connections have an identical weight and correspond to the parameter for two carbon atoms situated at distance 4. The classes of identical IH connections are presented in Figure 2a-e: all 8 connections corresponding to the distance 0 (Figure 2a) have identical weights because all non-hydrogen atoms are carbon atoms; Figure 2b presents the 14 connections between atoms situated at distance 1; the 20 connections from Figure 2c correspond to atoms situated at distance 2; there are 16 connections corresponding to carbon atoms separated by three bonds, as presented in Figure 2d; Figure 2e depicts the 6 connections between the three pairs of carbon atoms situated at distance 4.

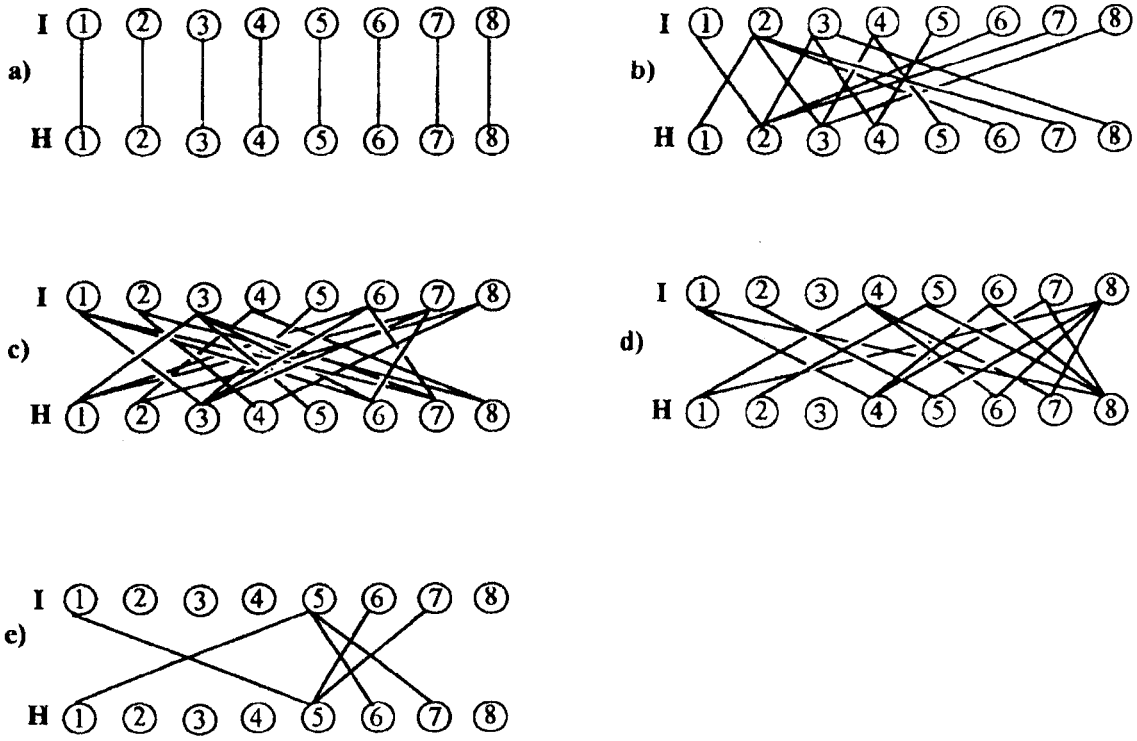


Fig. 2 – The structure of the MolNet connections between the input (**I**) and hidden (**H**) layers for 2,2,3-trimethylpentane; each neuron corresponds to the carbon atom with the same label from Figure 1. The connections between atoms with the same label are presented in a); the connections between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

The connections between the hidden and output layers (the HO connections) are separated in sets according to the degree of the carbon atoms: atoms with identical degree have connections with identical weights to the output unit. The molecular graph of 2,2,3-trimethylpentane contains 5 atoms with degree 1, and one atom with degree 2, 3, and 4, respectively; therefore, there are 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 used for the HO connections, giving in the case of molecule **1** four adjustable weights. The structure of BH and HO connections is presented in Figure 3a-d: the bias and output connections to atoms with the degree equal to 1 are presented in Figure 3a; those connecting the atoms with degree 2, 3, and 4 are depicted in Figure 3b-d, respectively. From Figure 3a one can see that the four atoms with degree 1 (namely atoms 1, 5, 6, and

8) have HO connections with identical weights; also, their connections to the bias unit have identical weights. The bias unit has also a connection to the output unit (BO connection). The total number of adjustable weights for 2,2,3-trimethylpentane is: 5 (IH connections) + 4 (BH connections) + 4 (HO connections) + 1 (BO connection) = 14.

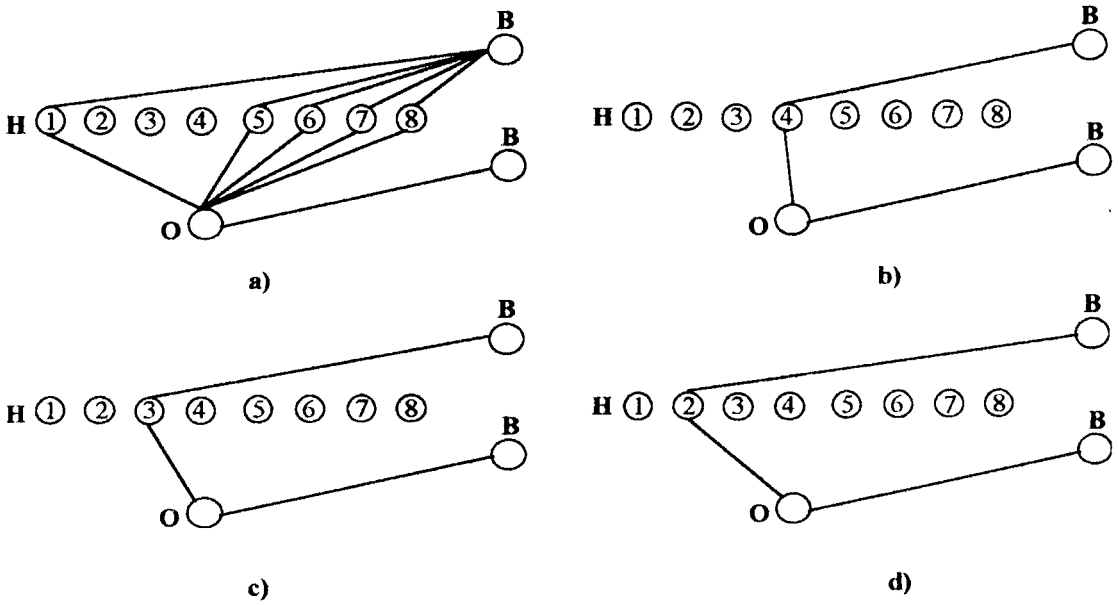


Fig. 3 – The structure of the MolNet connections between the hidden (**H**) and output (**O**) layers for 2,2,3-trimethylpentane; the bias neuron is labeled with **B.** The connections to/from atoms with the degree 1, 2, 3 and 4 are presented in a), b), c) and d), respectively.

Because MolNet is a MLF neural network, its use involves two phases: a learning and a prediction phase, respectively. In the learning phase the weights are adjusted with the backpropagation algorithm after the presentation of each molecule. If a connection type is absent from a certain molecule its value is not changed after the presentation of that molecule to the network. 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. In the prediction phase the molecular properties are computed with the weights determined in the learning phase. If the set of molecules used in the prediction phase contains bonding relationships that are absent in the molecules used in the learning phase these bonding relationships are neglected in predicting the molecular property.

MolNet OPERATION

Data Set. MolNet is tested in a QSPR investigation of a structural determination of alkane Gibbs energies. Because it is important to estimate 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 Gibbs energy for alkanes that are not used in the calibration of the model. The learning set contains 109 alkanes, while the test set contains 25 alkanes between C_6 and C_{10}. The structure and experimental Gibbs energies of the alkanes used in the present investigation are taken from the literature[8] and are reported in Tables 2 and 3.

Number of Adjustable Parameters. MolNet has a variable topology, and the number of adjustable parameters (connections) depends on the structure of the entire learning set of molecules. Because for the learning set of 109 alkanes the maximum graph distance between two carbon atoms is 7, there are 8 IH connection types. The degree of the carbon atoms is from 1 up to 4, and this gives 4

Table 2

Alkanes used in MolNet calibration, experimental Gibbs energies, and calibration residuals for MolNet network using **RDS** input atomic descriptor

Hydrocarbon	Gibbs energy ΔG at 300 K (kJ/mol)	
	Exp.	Residual
3-M-C_5	– 2.12	–3.08
2,2-M_2-C_4	– 7.42	– 3.48
2,3-M_2-C_4	– 1.77	1.17
3-M-C_6	6.60	– 0.05
3-E-C_5	12.70	1.28
2,2-M_2-C_5	2.10	– 1.92
2,3-M_2-C_5	7.60	1.47
2,4-M_2-C_5	4.90	– 1.34
3,3-M_2-C_5	4.80	0.66
2,2,3-M_3-C_4	6.30	5.31
n-C_8	17.67	– 0.89
2-M-C_7	13.37	1.77
3-M-C_7	13.79	0.61
2,4-M_2-C_6	13.07	– 0.03
2,5-M_2-C_6	11.40	5.97
3,3-M_2-C_6	15.13	– 1.18
3,4-M_2-C_6	18.43	0.46
3-E-2-M-C_5	20.68	– 2.64
3-E-3-M-C_5	24.36	1.78
2,2,3-M_3-C_5	19.45	1.23
2,2,4-M_3-C_5	15.70	– 0.16
2,3,3-M_3-C_5	20.04	0.62
2,3,4-M_3-C_5	20.76	1.70
2,2,3,3-M_4-C_4	24.04	5.93
2-M-C_8	21.60	– 0.86
3-M-C_8	22.00	– 1.37
3-E-C_7	26.40	0.26
4-E-C_7	26.80	0.30
2,2-M_2-C_7	19.50	3.29
2,3-M_2-C_7	23.50	2.37
2,4-M_2-C_7	20.80	– 3.30
2,5-M_2-C_7	18.20	1.08
2,6-M_2-C_7	19.80	3.43
3,3-M_2-C_7	22.00	0.21
3,4-M_2-C_7	24.90	– 1.91
3,5-M_2-C_7	22.00	– 0.41
3-E-3-M-C_6	30.50	– 1.30
4-E-2-M-C_6	24.50	0.28
2,2,4-M_3-C_6	23.60	0.24
2,2,5-M_3-C_6	15.30	3.07
2,3,3-M_3-C_6	29.40	– 0.61
2,3,4-M_3-C_6	28.60	– 1.57
2,3,5-M_3-C_6	22.20	0.74
2,4,4-M_3-C_6	26.60	– 1.19
3,3,4-M_3-C_6	31.40	– 3.04
3,3-E_2-C_5	43.30	0.17
3-E-2,2-M_2-C_5	37.50	– 0.88
3-E-2,3-M_2-C_5	36.80	– 4.21
2,2,3,3-M_4-C_5	39.00	– 1.95
2,2,3,4-M_4-C_5	36.70	– 0.72
2,3,3,4-M_4-C_5	39.70	– 0.64
3-E-C_8	34.90	– 0.09
4-E-C_8	33.40	– 0.49
2,2-M_2-C_8	27.70	3.25
2,4-M_2-C_8	28.80	-2.88
2,5-M_2-C_8	26.90	– 2.12
3,4-M_2-C_8	33.00	– 0.35
3,5-M_2-C_8	29.10	– 4.09
3,6-M_2-C_8	28.90	0.42
4,4-M_2-C_8	31.90	– 0.99
4,5-M_2-C_8	35.30	– 0.02
4-nP-C_7	38.20	4.73
4-iP-C_7	37.90	3.35
2-M-3-E-C_7	35.70	0.07
2-M-4-E-C_7	31.60	– 0.40
3-M-4-E-C_7	36.90	– 1.29
3-M-5-E-C_7	33.10	– 0.83
2,2,3-M_3-C_7	34.80	3.11
2,2,4-M_3-C_7	31.90	– 1.67
2,2,5-M_3-C_7	23.80	– 0.12
2,2,6-M_3-C_7	24.20	1.37
2,3,3-M_3-C_7	37.30	2.25
2,3,4-M_3-C_7	37.20	– 1.62
2,3,5-M_3-C_7	30.30	– 1.49
2,3,6-M_3-C_7	28.50	1.32
2,4,4-M_3-C_7	35.90	2.76
2,4,5-M_3-C_7	36.10	1.10
2,4,6-M_3-C_7	28.40	– 4.36
2,5,5-M_3-C_7	25.80	– 2.08
3,3,5-M_3-C_7	34.10	– 2.19
3,4,4-M_3-C_7	40.30	– 2.89
3,4,5-M_3-C_7	39.70	1.83
2-M-3-iP-C_6	46.80	2.91
3,3-E_2-C_6	51.00	3.28
3,4-E_2-C_6	45.00	1.67
2,2-M_2-3-E-C_6	44.40	1.58
2,2-M_2-4-E-C_6	36.10	1.63
2,3-M_2-3-E-C_6	45.00	– 2.40
2,3-M_2-4-E-C_6	42.10	– 1.41
2,4-M_2-4-E-C_6	42.30	0.64
3,3-M_2-4-E-C_6	50.00	0.09
3,4-M_2-4-E-C_6	47.60	3.71
2,2,3,3-M_4-C_6	48.70	– 1.15
2,2,3,4-M_4-C_6	46.10	– 2.51
2,2,3,5-M_4-C_6	32.10	– 4.43
2,2,4,5-M_4-C_6	32.80	– 2.49
2,2,5,5-M_4-C_6	21.10	– 2.46
2,3,3,4-M_4-C_6	49.10	– 4.20
2,3,3,5-M_4-C_6	41.20	– 1.67
2,3,4,4-M_4-C_6	49.20	– 2.95
2,3,4,5-M_4-C_6	42.70	– 1.75
3,3,4,4-M_4-C_6	58.90	0.31
2,4-M_2-3-iP-C_5	64.30	4.12
2-M-3,3-E_2-C_5	60.30	– 1.33
2,2,3-M_3-3-E-C_5	66.60	3.52
2,2,4-M_3-3-E-C_5	59.00	– 0.38
2,3,4-M_3-3-E-C_5	61.00	– 1.86
2,2,3,3,4-M_5-C_5	68.80	1.17
2,2,3,4,4-M_5-C_5	63.80	– 0.18

Table 3

Alkanes used in MolNet prediction, experimental Gibbs energies, and prediction residuals for MolNet network using **RDS** input atomic descriptor

Hydrocarbon	Gibbs energy ΔG at 300 K (kJ/mol)	
	Exp.	Residual
2-M-C_5	– 4.05	– 6.11
n-C_7	9.50	1.47
2-M-C_6	4.90	2.03
4-M-C_7	17.40	0.23
3-E-C_6	18.53	– 0.12
2,2-M_2-C_6	12.15	4.67
2,3-M_2-C_6	17.20	3.27
4-M-C_8	21.00	– 4.41
4,4-M_2-C_7	25.80	– 2.48
3-E-2-M-C_6	26.40	– 2.57
3-E-4-M-C_6	29.90	– 2.81
2,2,3-M_3-C_6	27.20	1.51
3-E-2,4-M_2-C_5	37.80	– 1.83
2,2,4,4-M_4-C_5	35.60	2.73
2,3-M_2-C_8	32.00	2.43
2,6-M_2-C_8	26.90	– 0.95
2,7-M_2-C_8	28.20	1.39
3,3-M_2-C_8	30.50	1.63
2-M-5-E-C_7	31.20	0.53
3-M-3-E-C_7	38.20	1.68
4-M-3-E-C_7	38.60	– 1.99
4-M-4-E-C_7	40.30	1.25
3,3,4-M_3-C_7	38.70	– 2.43
2,5-M_2-3-E-C_6	33.60	– 2.32
2,2,4,4-M_4-C_6	44.10	– 1.32

HO connection classes 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 large ratio between the number of alkanes in the learning set and the number of adjustable weights indicates that there is no danger of overfitting.

Input Data. The input data for the neurons in the input layer are atomic topological descriptors namely the degree **DEG**, the distance sum **DS**,[22,23] and the reciprocal distance sum **RDS**.[24,25] The degree of the atom i is computed with the equation:

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

where **A** is the adjacency matrix. The degree vector of 2,2,3-trimethylpentane is **DEG(1)** = {1, 4, 3, 2, 1, 1, 1, 1}.

The distance sum of the atom i is the sum of the elements in the row i (or column i) of the distance matrix **D**:

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

The distance sum vector of 2,2,3-trimethylpentane is **DS(1)** = {17, 11, 11, 15, 21, 17, 17, 17}.

The reciprocal distance sum of the atom i is defined by the equation:

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

where **RD** is the reciprocal distance matrix. The reciprocal distance sum vector of 1 is **RDS(1)** = = {3.417, 5.333, 5.000, 4.000, 2.917, 3.417, 3.417, 3.333}.

Learning Method. The training of the ANNs is performed with the standard backpropagation method,[2] until the convergence is obtained, i.e., the correlation coefficient between experimental and calculated alkane Gibbs energy values improves by less than 10^{-5} in 100 epochs. One epoch corresponds to one complete presentation of the 109 molecules in the learning set. The connections are updated 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 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. Both learning rate and momentum

values are maintained constant during the training phase. In all cases the learning phase stops after a few hundred epochs and the results are slightly influenced by the initial random set of weights.

Activation Functions. The most commonly used activation function in chemical applications of neural networks is the sigmoid that takes values between 0 and 1. For large negative arguments its value is close to 0, and practice demonstrated that learning with the backpropagation algorithm is difficult in such conditions. To overcome this deficiency of the sigmoid function, the hyperbolic tangent (tanh) which takes values between –1 and 1 is used in the present study. Both the tanh and the sigmoid activation functions are very flat when the absolute value of the argument is greater than 10 when the derivative has an extremely small value, leading to a poor sensitivity of the two activation functions to large positive or negative arguments. This behavior is an important cause 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 function; therefore, for the output layer we use also a linear activation function. A new type of activation function is the symmetric logarithmoid,[26,27] defined by the formula: $\mathrm{Act}(z) = \mathrm{sign}\,(z)\ln(1 + |z|)$. 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. The symlog activation function is used for the unit from the output layer.

Preprocessing of the Data. Each component of the input (**DEG**, **DS**, or **RDS** vector) and output (representing the target Gibbs energy value) patterns is linearly scaled between -0.9 and 0.9. We have to point out that for the tanh output 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.

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 Gibbs energies at the end of the calibration phase (G_{cal}) with the target values (G_{exp}), while the predictive quality is estimated with a set of alkanes that were not used in the calibration phase by comparing the predicted (G_{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) Gibbs energies: $\Delta G_{exp} = A + B \cdot \Delta G_{cal/pr}$.

MolNet PREDICTION OF ALKANE GIBBS ENERGIES

Table 4 presents the calibration and prediction results obtained when MolNet was trained with the **DEG** atomic descriptor as input data. The calibration correlation coefficient, r_{cal}, is in the range 0.948 to 0.977, and the calibration standard deviation, s_{cal}, takes values between 3.23 and 4.89. The prediction correlation coefficient, r_{pr}, takes values between 0.940 and 0.979, while the prediction standard deviation, s_{pr}, is in the range 2.49 to 4.17. Overall, the best calibration and prediction results are obtained with linear output function, followed by the tanh output function.

The calibration and prediction results obtained when MolNet is trained with the **DS** atomic descriptor as input data are presented in Table 5. In the calibration phase r_{cal} is between 0.941 and 0.963, and s_{cal} takes values between 4.11 and 5.16. In the prediction phase r_{pr} takes values between 0.944 and 0.973, and s_{pr} is in the range 2.82 to 4.03. For all MolNet networks trained with **DS** input data the calibration and prediction results are of lower statistical quality than those obtained with **DEG** as input data.

Table 6 presents the calibration and prediction results obtained when MolNet is trained with **RDS** atomic descriptor as input data. In the calibration phase r_{cal} is between 0.958 and 0.989, and s_{cal} takes values between 2.23 and 4.41. In the prediction phase r_{pr} takes values between 0.920 and 0.978, and s_{pr} is in the range 2.56 to 4.80. The Gibbs energy residuals computed with linear output function, a

Table 4

MolNet calibration and prediction results for the computation of alkane Gibbs energies using **DEG** input atomic descriptor. The table reports the number of training epochs, the hidden and output momentum (α_h and α_o), the output activation function, the calibration and prediction standard deviation (s_{cal} and s_{pr}) and correlation coefficient (r_{cal} and r_{pr}). 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}
1400	0.30	linear	0.30	3.39	0.975	3.76	0.952
1200	0.30	linear	0.15	3.32	0.976	3.64	0.955
1200	0.30	linear	0.10	3.30	0.976	3.61	0.955
1400	0.30	linear	0.05	3.28	0.977	3.58	0.956
1600	0.15	linear	0.05	3.25	0.977	3.50	0.958
1700	0.10	linear	0.05	3.24	0.977	3.49	0.958
1900	0.05	linear	0.05	3.23	0.977	3.48	0.959
200	0.15	linear	0.15	3.93	0.966	2.49	0.979
1600	0.15	linear	0.10	3.26	0.977	3.53	0.957
1800	0.10	linear	0.10	3.26	0.977	3.53	0.957
2000	0.30	symlog	0.30	4.25	0.961	4.17	0.940
1900	0.30	symlog	0.15	4.24	0.961	4.10	0.942
1800	0.30	symlog	0.10	4.23	0.961	4.09	0.942
1200	0.30	symlog	0.05	3.83	0.968	2.91	0.971
1900	0.15	symlog	0.05	4.18	0.962	3.96	0.946
2000	0.10	symlog	0.05	4.16	0.962	3.92	0.947
1800	0.05	symlog	0.05	4.16	0.962	3.92	0.947
1700	0.15	symlog	0.15	4.19	0.962	3.97	0.946
2000	0.15	symlog	0.10	4.19	0.962	3.97	0.946
1900	0.10	symlog	0.10	4.17	0.962	3.93	0.947
1900	0.30	tanh	0.30	4.10	0.963	3.91	0.947
1700	0.30	tanh	0.15	4.89	0.948	2.83	0.973
1400	0.30	tanh	0.10	3.71	0.970	2.81	0.973
1700	0.30	tanh	0.05	4.20	0.962	4.00	0.945
1600	0.15	tanh	0.05	3.71	0.970	2.83	0.973
1900	0.10	tanh	0.05	4.14	0.963	3.87	0.948
2000	0.05	tanh	0.05	4.13	0.963	3.85	0.949
1800	0.15	tanh	0.15	4.13	0.963	3.90	0.948
1500	0.15	tanh	0.10	3.70	0.970	2.82	0.973
1800	0.10	tanh	0.10	4.13	0.963	3.86	0.949

Table 5

MolNet calibration and prediction results for the computation of alkane Gibbs energies using **DS** input atomic descriptor. The notations are explained in Table 4

Epoch	Hidden momentum (α_h)	Output activation function	Output momentum (α_o)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1800	0.30	linear	0.30	4.60	0.954	3.58	0.956
2000	0.30	linear	0.15	4.40	0.958	3.43	0.960
1800	0.30	linear	0.10	4.34	0.959	3.38	0.961
1900	0.30	linear	0.05	4.28	0.960	3.34	0.962
1700	0.15	linear	0.05	4.17	0.962	3.22	0.965
1900	0.10	linear	0.05	4.14	0.963	3.18	0.966
2000	0.05	linear	0.05	4.11	0.963	3.15	0.966
1600	0.15	linear	0.15	4.28	0.960	3.30	0.963
1700	0.15	linear	0.10	4.23	0.961	3.26	0.964
1600	0.10	linear	0.10	4.22	0.961	4.01	0.945
1500	0.30	symlog	0.30	4.60	0.954	3.13	0.967
1900	0.30	symlog	0.15	5.16	0.941	4.03	0.944
1800	0.30	symlog	0.10	4.74	0.951	3.06	0.968
1500	0.30	symlog	0.05	4.44	0.957	3.00	0.969
1700	0.15	symlog	0.05	4.67	0.952	3.02	0.969
1900	0.10	symlog	0.05	4.66	0.952	3.02	0.969
1100	0.05	symlog	0.05	4.36	0.959	2.82	0.973
2000	0.15	symlog	0.15	4.72	0.951	3.04	0.967
1500	0.15	symlog	0.10	4.41	0.957	2.93	0.971
1800	0.10	symlog	0.10	5.02	0.945	4.03	0.944
1700	0.30	tanh	0.30	4.78	0.950	3.28	0.963
1200	0.30	tanh	0.15	4.64	0.953	3.17	0.966
1300	0.30	tanh	0.10	4.61	0.954	3.15	0.966
1300	0.30	tanh	0.05	4.58	0.954	3.13	0.967
1100	0.15	tanh	0.05	4.52	0.955	3.01	0.969
1200	0.10	tanh	0.05	4.50	0.956	2.99	0.969
1000	0.05	tanh	0.05	4.49	0.956	2.96	0.970
1000	0.15	tanh	0.15	4.58	0.954	3.04	0.969
1300	0.15	tanh	0.10	4.54	0.955	3.05	0.968
1200	0.10	tanh	0.10	4.53	0.955	3.02	0.969

Table 6

MolNet calibration and prediction results for the computation of alkane Gibbs energies using **RDS** input atomic descriptor. The notations are explained in Table 4

Epoch	Hidden momentum (α_h)	Output activation function	Output momentum (α_o)	s_{cal}	r_{cal}	s_{pr}	r_{pr}
1700	0.30	linear	0.30	3.80	0.969	4.58	0.927
1800	0.30	linear	0.15	3.13	0.979	3.39	0.961
1700	0.30	linear	0.10	3.67	0.971	4.34	0.935
900	0.30	linear	0.05	3.62	0.972	4.24	0.938
2000	0.15	linear	0.05	2.23	0.989	2.60	0.977
1700	0.10	linear	0.05	3.29	0.977	3.60	0.956
500	0.05	linear	0.05	3.70	0.970	3.84	0.949
1100	0.15	linear	0.15	3.63	0.971	4.27	0.937
800	0.15	linear	0.10	3.62	0.972	4.24	0.938
900	0.10	linear	0.10	3.45	0.974	3.90	0.948
2000	0.30	symlog	0.30	4.03	0.965	4.53	0.929
1800	0.30	symlog	0.15	2.47	0.987	2.65	0.976
300	0.30	symlog	0.10	4.41	0.958	5.14	0.907
200	0.30	symlog	0.05	4.19	0.962	4.49	0.930
1900	0.15	symlog	0.05	3.86	0.968	4.16	0.940
1800	0.10	symlog	0.05	3.86	0.968	4.18	0.940
1900	0.05	symlog	0.05	2.40	0.988	2.56	0.978
1700	0.15	symlog	0.15	3.89	0.967	4.24	0.938
1800	0.15	symlog	0.10	2.43	0.987	2.60	0.977
2000	0.10	symlog	0.10	3.88	0.967	4.23	0.938
1000	0.30	tanh	0.30	4.29	0.960	5.19	0.905
1600	0.30	tanh	0.15	4.14	0.963	4.80	0.920
2000	0.30	tanh	0.10	4.04	0.965	4.57	0.928
1800	0.30	tanh	0.05	2.46	0.987	2.61	0.977
1700	0.15	tanh	0.05	2.43	0.987	2.56	0.978
1900	0.10	tanh	0.05	4.08	0.964	4.83	0.919
2000	0.05	tanh	0.05	3.94	0.966	4.40	0.933
1800	0.15	tanh	0.15	4.08	0.964	4.69	0.924
1500	0.15	tanh	0.10	4.05	0.964	4.72	0.922
1900	0.10	tanh	0.10	2.45	0.987	2.57	0.978

hidden momentum $\alpha_h = 0.15$ and an output momentum $\alpha_o = 0.05$ are presented in Tables 2 and 3, column 4. This example is selected because it has good calibration and prediction results. Overall, the results obtained with **RDS** input data are better than those obtained with **DEG** and **DS** atomic descriptors. The results obtained with the three output activation functions are close and from the data reported in Table 6 it is not possible to prefer a certain output function.

CONCLUSIONS

The MolNet neural network represents a new type of multi-layer feedforward ANN that gives good results in predicting alkane Gibbs energies. MolNet changes its topology (the number of units in the input and hidden layer, and the number and type of connections) according to the molecular structure of the chemical compound presented to the network.[1,28] Each non-hydrogen atom in the molecule has a corresponding unit in the input and hidden layers, while the output layer contains only one unit that provides the computed value of the molecular property.

Three atomic descriptors, namely **DEG, DS** and **RDS**, were used as input data, with best results obtained with the **RDS** index. MolNet networks were trained with the tanh hidden activation function and three output activation functions: linear, symlog and tanh. The results obtained with the three output activation functions are close and therefore it is not possible to prefer a certain output function.

ACKNOWLEDGEMENT. We acknowledge the partial financial support of this research by the Ministry of Research and Technology under Grant 381 TA10 and by the Ministry of Education under Grant 5001 TB10.

REFERENCES

1. Part 9: O. Ivanciuc, *Rev. Roum. Chim.*, **1998**, *43*, 885–894.
2. D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, **1986**, *323*, 533–536.
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 and R. G. Trenary, *J. Chem. Inf. Comput. Sci.*, **1990**, *30*, 477–484.
7. D. W. Elrod, G. M. Maggiora and R. G. Trenary, *Tetrahedron Comput. Methodol.*, **1990**, *3*, 163–174.
8. A. A. Gakh, E. G. Gakh, B. G. Sumpter and D. W. Noid, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 832–839.
9. A. T. Balaban, S. C. Basak, T. Colburn and G. D. Grunwald, *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, 1118–1121.
10. D. Cherqaoui and D. Villemin, *J. Chem. Soc. Faraday Trans.*, **1994**, *90*, 97–102.
11. D. Cherqaoui, D. Villemin, A. Mesbah, J.-M. Cense and V. Kvasnička, *J. Chem. Soc. Faraday Trans.*, **1994**, *90*, 2015–2019.
12. F. R. Burden, *Quant. Struct.-Act. Relat.*, **1996**, *15*, 7–11.
13. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye and J. P. Doucet, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 644–653.
14. O. Ivanciuc, J.-P. Rabine, D. Cabrol-Bass, A. Panaye and J. P. Doucet, *J. Chem. Inf. Comput. Sci.*, **1997**, *37*, 587–598.
15. J. H. Schuur, P. Selzer and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 334–344.
16. J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer and V. Steinhauer, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 1030–1037.
17. L. H. Hall and C. T. Story, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 1004–1014.
18. S. Hatrik and P. Zahradnik, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 992–995.
19. H. Bauknecht, A. Zell, H. Bayer, P. Levi, M. Wagener, J. Sadowski and J. Gasteiger, *J. Chem. Inf. Comput. Sci.*, **1996**, *36*, 1205–1213.
20. D. B. Kireev, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 175–180.
21. B. Mohar and T. Pisanski, *J. Math. Chem.*, **1988**, *2*, 267–277.
22. A. T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399–404.
23. A. T. Balaban, *Pure Appl. Chem.*, **1983**, *55*, 199–206.
24. O. Ivanciuc, *Rev. Roum. Chim.*, **1989**, *34*, 1361–1368.
25. O. Ivanciuc, T.-S. Balaban and A. T. Balaban, *J. Math. Chem.*, **1993**, *12*, 309–318.
26. A. B. Bulsari and H. Saxén, *Neurocomputing*, **1991**, *3*, 125–133.
27. A. B. Bulsari and H. Saxén, *Neural Network World*, **1991**, *4*, 221–224.
28. O. Ivanciuc, *Anal. Chim. Acta*, **1999**, *384*, 271–284.