but different widths or force constants (see eq A8e).

We like to add that we have also tried less flexible potential functions, e.g., fourth- or sixth-order polynomials $V(q) = V_0 + V_2q^2 + V_4q^4$ $(+V_6q^6)$, but these cannot account for the "experimental" conditions (see eqs A8a–d), irrespective of the value of k in eq A8e. However, other functions with less parameters would also satisfy conditions (see eqs A8a–d), e.g., a double Morse potential

$$V = D\{\exp[-(-q + q_e)\beta] - 1\}^2 + D\{\exp[-(q + q_e)\beta] - 1\}^2 - D \quad (A9)$$

with parameters $D = 0.0098E_h$, $q_e = 0.62a_0$, $\beta = 8.4a_0^{-1}$, yielding $V^* = 25$ kJ mol^{-1} and $\hbar\omega = 900$ cm^{-1} and $V(q_e) = V(-q_e) \cong 0$. This double Morse potential would also account for the inverse isotope effect; i.e., $\hbar\omega$ would increase from 900.0 cm^{-1} to 920.4 cm^{-1} when $^{12}C^1H$ is replaced by $^{13}C^2H$.

Finally, let us consider the symmetry selection rules for the fundamental transition $|0\rangle \rightarrow |2\rangle$ depending on the dipole functions $\mu_i(q)$, $i = x, y, z$ and matrix elements $\langle 2|\mu_i|0\rangle$.

The C_s symmetry of semibullvalene along the reaction path q implies the following symmetries of the μ_i's (see Scheme I):

$$\mu_x(q) = -\mu_x(-q) \quad (A10a)$$

$$\mu_y(q) = 0 \quad (A10b)$$

$$\mu_z(q) = +\mu_z(-q) \quad (A10c)$$

For the resulting matrix elements of levels $|0\rangle$, $|2\rangle$ with gerade symmetry, we obtain

$$\langle 2|\mu_x|0\rangle = 0 \quad (A11a)$$

$$\langle 2|\mu_y|0\rangle = 0 \quad (A11b)$$

$$\langle 2|\mu_z|0\rangle \neq 0 \quad (A11c)$$

As a consequence, we have

$$\langle 2|-\mu E|0\rangle = -\langle 2|\mu_z|0\rangle E_z \quad (A11d)$$

for the dipole transition matrix element, and the golden rule for the rate of absorption

$$k(2 \leftarrow 0) = (2\pi/\hbar)\langle 2| - 0.5\mu \times E|0\rangle^2\delta(E_2 - E_0 - \hbar\omega) \quad (A12)$$

calls for z-polarized light

$$E_z = E \exp[i(k_z z - \omega t)] \quad (A13)$$

otherwise, the fundamental transition $|0\rangle \rightarrow |2\rangle$ is dipole-forbidden.

Neural Network Classification of Inductive and Resonance Effects of Substituents

Vladimír Kvasnička,* Štěpán Sklenák, and Jiří Pospíchal

Contribution from the Department of Mathematics, Faculty of Chemical Technology, Slovak Technical University, 81237 Bratislava, Czechoslovakia. *Received June 10, 1992*

Abstract: An application of feed-forward neural networks adapted by back-propagation strategy to classification of inductive and resonance effects of functional groups is described. The descriptors of functional groups, forming the input information to the neural network, are determined as numbers of times certain subgraphs (with evaluated vertices and edges) appear in the functional group under study. The hidden activities are used for cluster analysis of functional groups.

Introduction

Neural networks[1,2] are equipped by learning features that allow computers to be trained to recognize patterns in data of high dimensionality. Such patterns are manifested by effective correlations between molecular structure and property.[3–5] The neural networks do not require any formulation of rules about reactivity to make decisions. They form an internal model by extracting information directly from the properly selected examples belonging to the so-called training set. The expert systems based on neural networks can be used to roughly estimate the yields of reactions and to classify the reaction products (e.g., regioselectivity). Such an expert system is vital in any computer system for organic synthesis design.

The purpose of the present communication is to demonstrate an application of neural networks for classification and prediction of inductive and resonance effects (represented by σ constants) of functional groups. These parameters, initially introduced in physical organic chemistry,[6] describe the influence of functional groups on the reactivity of synthons[7,8] (reaction cores). Input information to the neural network, which properly represents the topology and basic physical parameters of functional groups, was chosen to consist of simple descriptors[5,9] assigned to numbers of appearance of specially labeled rooted subgraphs in the functional group. This means that the functional groups are described by graph-theoretical parameters without the necessity to calculate any additional physical or physico-chemical parameters of functional groups. The results obtained are very encouraging and

(1) (a) Simpson, P. K. *Artificial Neural Systems*; Pergamon Press: New York, 1990. (b) Wasserman, P. *Neural Computing: Theory and Practice*; Van Nostrand Reinhold: Princeton, NJ, 1989.

(2) Rumelhart, D. E.; McClelland, J. L. (and the PDP Research Group) *Parallel Distributed Processing*; MIT Press: Cambridge, MA, 1986; Vols. I and II.

(3) (a) Župan, J.; Gasteiger, J. *Anal. Chim. Acta* **1991**, *248*, 1. (b) *Tetrahedron Comput. Methodol.* **1990**, *3*, no. 1.

(4) (a) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 477. (b) Elrod, D. W.; Maggiora, G. M.; Trenary, R. G. *Tetrahedron Comput. Methodol.* **1990**, *3*, 163. (c) Luce, H. H.; Govind, R. *Tetrahedron Comput. Methodol.* **1990**, *3*, 143.

(5) Kvasnička, V.; Pospíchal, J. *J. Mol. Struct. (THEOCHEM)* **1991**, *235*, 227.

(6) Hammett, L. P. *Physical Organic Chemistry*; McGraw-Hill: New York, 1970.

(7) Corey, E. J. *Pure Appl. Chem.* **1967**, *14*, 19.

(8) (a) Kvasnička, V.; Pospíchal, J. *Int. J. Quant. Chem.* **1990**, *28*, 253. (b) Koča, J.; Kratochvíl, M.; Kvasnička, V.; Matyska, L.; Pospíchal, J. *Synthon Model of Organic Chemistry and Synthesis Design.* Springer Verlag: Berlin, 1989.

(9) Zou, Y.; Johnson, M.; Tsai, C.-C. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 442.

0002-7863/93/1515-1495$04.00/0

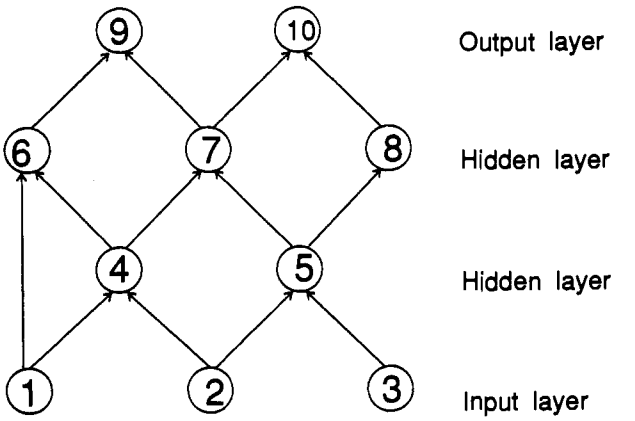

Figure 1. An illustrative example of the feed-forward neural network composed of 10 neurons and 13 connections, represented by an acyclic and connected oriented graph. Its neurons are vertically structured into one input layer, two hidden layers, and one output layer.

support the common chemical belief that the properties of molecular systems are determined mainly by structural formulas. The hidden activities of a three-layer neural network correspond to the so-called internal representation of the objects under study (in our case, the functional groups). These entities are useful for a cluster analysis[10] of functional groups which will be demonstrated in the second part of this paper. Such an analysis allows one to overcome to some extent the main drawback of the neural network paradigm, that the network cannot "explain" its predictions.

Theory of Neural Networks

A feed-forward neural network[2] applied for solving classification problems may be formally considered as an oriented and connected graph;[11] see Figure 1. The graph is composed of N vertices, *neurons*, and M edges, *connections*. The neuron set is divided into three disjoint subsets (see Figure 1):

$$V = V_I \cup V_H \cup V_O \tag{1}$$

where V_I is composed of N_I *input* neurons that are incident only with outgoing connections, V_H is composed of N_H *hidden* neurons that are incident at least with one incoming and one outgoing connection, and V_O is composed of N_O *output* neurons that are incident only with incoming connections. The neurons may be vertically organized in the so-called *layers*.

Neurons and connections of neural networks are evaluated by real numbers. We assign a *threshold coefficient* ϑ_i to each hidden or output neuron v_i and a *weight coefficient* ω_{ji} to each connection outgoing from vertex v_i and incoming to the vertex v_j. Moreover, we assign to each neuron v_i an *activity* x_i. We postulate that activities of input neurons are constant, whereas the activities of other neurons are determined by

$$x_i = f(\xi_i) \qquad (i = 1, 2, ..., N) \tag{2a}$$

$$\xi_i = \sum_j \omega_{ij} x_j + \vartheta_i \tag{2b}$$

where the summation index j runs over all neurons that are adjacent to the neuron v_i by connections that are incoming to v_i and outgoing from v_j's. We have used the following *transfer function* $f(\xi)$

$$f(\xi) = \frac{B + A \exp(-\xi)}{1 + \exp(-\xi)} \tag{3a}$$

with asymptotic boundaries specified by $A = -1$, $B = 1$. Its first derivative is determined by

$$f'(\xi) = \frac{[-A + f(\xi)][B - f(\xi)]}{B - A} \tag{3b}$$

A recurrent application of formula 2 gives the activities of all hidden and output neurons.

A *supervised adaptation process*[2] of neural network consists of looking for such threshold and weight coefficients that give, for prescribed input activities, such output activities, which are in close agreement with the required output activities. An *objective function* gives us a measure of such an agreement, based, for example, on the least-squares method. A goal of the adaptation process is finding the weight and threshold coefficients that will minimize the objective function E. This minimization may be carried out by a version of gradient method,[12] e.g., by its simplest version called the *steepest-descent method*.[12] For implementation of this method, we have to know all partial derivatives $\partial E/\partial \omega_{ij}$ and $\partial E/\partial \vartheta_i$ of the objective function with respect to the weight and threshold coefficients. It is easy to see that these partial derivatives are mutually related by

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial \vartheta_i} x_j \tag{4}$$

This means that the whole process of calculation of partial derivatives $\partial E/\partial \omega_{ij}$ and $\partial E/\partial \vartheta_i$ may be reduced to a substantially simpler calculation of $\partial E/\partial \vartheta_i$; the two-index derivatives $\partial E/\partial \omega_{ij}$ are then determined by the one-index derivatives $\partial E/\partial \vartheta_i$. The partial derivatives $\partial E/\partial \vartheta_i$ are determined by[13,14]

$$\frac{\partial E}{\partial \vartheta_i} = f'(\xi_i)\left[g_i + \sum_l \frac{\partial E}{\partial \vartheta_l} \omega_{li} \right] \text{ (for } i \in V_H \cup V_O) \tag{5}$$

where the term g_i is determined by a difference of the calculated output activity and the required activity both assigned to an output neuron v_i, and $g_i = 0$ for hidden and input neurons. The summation runs over all neurons v_l's that are forming oriented edges going from v_i to v_l.

For feed-forward neural networks, the relation 5 offers immediately the well-known recurrent relations used in the so-called *back-propagation adaptation*.[2,15] The partial derivatives $\partial E/\partial \vartheta_i$, for $i \in V_O$, are simply determined by $\partial E/\partial \vartheta_i = f'(\xi_i) g_i$. These derivatives are then used for the evaluation of derivatives, $\partial E/\partial \vartheta_i$, where the index i corresponds to neurons adjacent to output neurons. The same procedure is recurrently repeated until all derivatives $\partial E/\partial \vartheta_i$ assigned to hidden neurons are calculated. Finally, knowing all partial derivatives $\partial E/\partial \vartheta_i$, then the derivatives $\partial E/\partial \omega_{ij}$ are simply determined by (4).

The above-outlined method of calculation of partial derivatives may be simply generalized for more than one pair of sets of input–output activities.

The steepest-descent minimization method accelerated by the so-called momentum method[2] is based on the following updated formula,

$$\omega_{ij}^{(k+1)} = \omega_{ij}^{(k)} - \lambda \frac{\partial E}{\partial \omega_{ij}} + \mu \Delta \omega_{ij}^{(k)} \tag{6}$$

where the positive parameter $\lambda > 0$ should be sufficiently small to ensure the convergence of adaptation process and simultaneously sufficiently large to achieve the fast convergence. The momentum parameter μ is taken from the semi-open interval (0, 1], usually $\mu = 0.7$–0.9. Finally, the term $\Delta \omega_{ij}^{(k)}$ corresponds to the previous change of weight. We have had very good numerical experiences[5,16] if the adaptation process of neural networks was carried out by more sophisticated versions of the gradient method,[12] e.g., by the method of conjugate gradients or by the method of variable metric.

(10) Anderberg, M. R. *Cluster Analysis for Applications*; Academic Press: New York, 1973.
(11) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
(12) Polak, E. *Computational Methods in Optimization*; Academic Press: New York, 1971.
(13) Pineda, F. J. *Phys. Rev. Lett.* **1987**, *59*, 2229.
(14) Kvasnička, V.; Sklenák, S.; Pospíchal, J. *J. Mol. Struct. (THEOCHEM)*, in press.
(15) Minsky, M.; Papert, S. *Perceptron. An Introduction to Computational Geometry*; MIT Press: Cambridge, MA, 1969.
(16) (a) Kvasnička, V. *Chem. Pap.* **1990**, *44*, 19. (b) Kvasnička, V. *J. Math. Chem.* **1991**, *6*, 63.

Table I. Descriptors of Illustrative Examples in Figure 4

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
NO_2	0	1	0	0	4	2	0	0	2	0	0	0	0	0
CH_2Ph	0	1	2	0	0	0	0	1	0	0	0	0	0	0
SO_2Ph	0	2	0	0	4	2	0	1	2	0	0	0	0	0
$NHCOCH_3$	1	1	1	0	0	1	0	0	0	2	2	3	0	1

Since the objective function E is highly nonlinear, its minimization is a nonstandard numerical task. Usually, a local minimum is achieved where the value of objective function is much greater than a required one. This means that it is necessary to carry out at least a few adaptation processes with randomly generated initial values of threshold and weight coefficients. Then we select those coefficients that give the lowest (positive) value of minimized objective functions for the forthcoming *active* process of neural network. The input activities in the active process are determined by descriptors of objects taken from the so-called *testing set* (a set composed of ordered pairs x_I/x_O that are different from those included in the training set).

The extrapolation of the adapted neural network outside the training set is another critical point of applications of neural networks as a classifier of objects from the testing set. The active process often gives results of classification with much lower precision than one requires. Then the adaptation process of neural network should be repeated with new randomly generated threshold and weight coefficients. New coefficients obtained as a result of this adaptation process are again used for the forthcoming active process. If the resulting classification of objects does not satisfy the required precision repeatedly, then attention must be directed to the topology of the neural network or to descriptors (input activities) of objects from training and testing sets. Likely, the topology of the neural network (number of hidden layers and numbers of hidden neurons situated at these layers) is inadequate to the problem under study or the descriptors do not reflect properly the internal structure of objects.

Descriptors and Functional Groups

Functional groups may be formally considered as rooted molecular graphs[17] in which the vertices (atoms) and edges (bonds) are evaluated by the additional symbols that specify their physical nature. The root of these molecular graphs corresponds to the atom immediately attached to the substituted molecule. The approach presented is restricted to functional groups that are attached to molecules by a single bond and do not hold a positive or negative charge. We used a modification of the approach suggested by Johnson et al.[9] This modified description of input parameters for the neural network helped to predict the meta products of nitration[5] and ^{13}C NMR chemical shifts[14] in a series of monosubstituted benzenes. The atom and bond labels are specified in Figure 2. In the next step, we introduce 14 molecular subgraphs, each labeled with only one atom or bond; see Figure 3. These labeled subgraphs are used for simple construction of descriptors that characterize the univalent functional groups as inputs of neural networks. The entry of the descriptor is equal to the number of times the particular subgraph appears in the molecular graph assigned to the functional group as shown in Figure 4 and Table I.

Prediction and Classification of Inductive and Resonance σ Constants by Neural Networks

The inductive and resonance σ constants σ_I and σ_R in physical organic chemistry[6] are very important parameters that characterize the electronic properties of functional groups. They offer a quantitative description of qualitative conceptions used in organic chemistry[18] when the influence of functional groups on the molecular skeleton (e.g., benzene ring) is considered. Quantum-chemical approaches[19] explaining the physical nature of these constants are not very successful. This lack of success may be due to an inadequate consideration of environmental effects. Effective procedures for quantification of inductive and resonance effects were developed by Gasteiger et al.[20] in the early 80's.

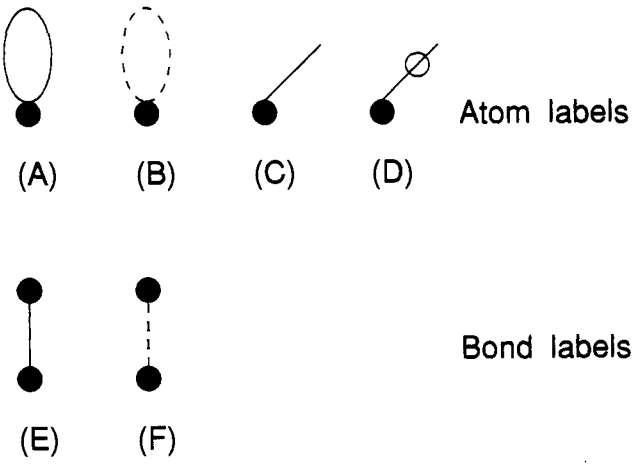

Figure 2. Atom and bond labels used for the construction of functional group descriptors. The labels correspond to: (A) the lone electron pair, (B) the main quantum number decreased by one, (C) the hydrogen atom attached to an atom, (D) the phenyl group attached to an atom, (E) the σ bond, and (F) the π bond.

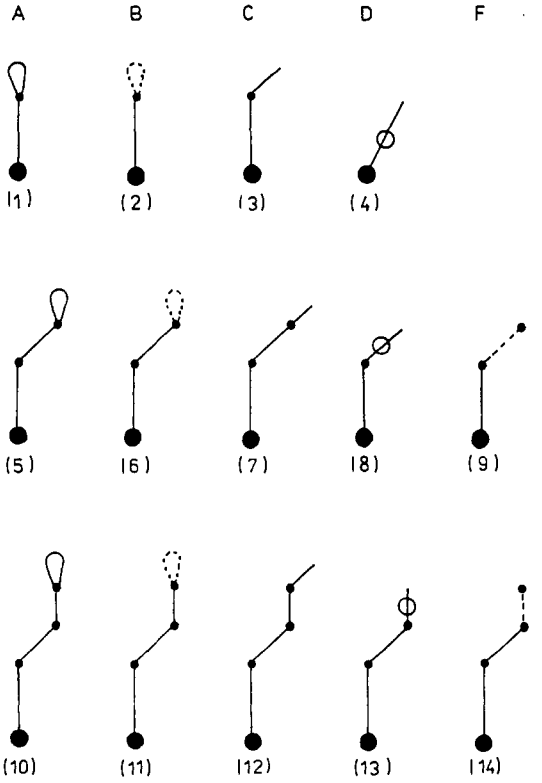

Figure 3. Fourteen molecular subgraphs that classify rooted molecular graphs assigned to univalent functional groups. The bold dots (roots) correspond to atoms that are immediately attached to an atom of the substituted molecule.

(17) Balaban, A. T., Ed. *Chemical Applications of Graph Theory*; Academic Press: London, 1976.
(18) Ingold, C. K. *Structure and Mechanism in Organic Chemistry*; Cornell University Press: Ithaca, NY, 1969.
(19) Hehre, W. J.; Taft, R. W.; Topson, R. D. *Prog. Phys. Org. Chem.* **1976**, *12*, 159.

Table II. Experimental and Estimated σ_I and σ_R Using Neural Network

		exp		est				exp		est	
no.	X	σ_I	σ_R	σ_I	σ_R	no.	X	σ_I	σ_R	σ_I	σ_R
					Training Set						
1	Br	0.44	−0.17	0.42	−0.17	34	CH_2CH_3	−0.05	−0.10	−0.10	−0.11
2	F	0.43	−0.34	0.41	−0.36	35	OCH_2CH_3	0.22	−0.44	0.22	−0.47
3	SO_2F	0.75	0.22	0.67	0.20	36	$NHCH_2CH_3$	−0.11	−0.51	−0.07	−0.50
4	SF_5	0.57	0.15	0.60	0.16	37	$SO_2CH_2CH_3$	0.54	0.22	0.49	0.21
5	IO_2	0.63	0.20	0.71	0.22	38	$P(CH_3)_2$	−0.08	0.39	−0.08	0.39
6	NO_2	0.67	0.16	0.60	0.17	39	$PO(CH_3)_2$	0.37	0.19	0.41	0.18
7	N_3	0.30	−0.13	0.39	−0.12	40	$C(OH)(CF_3)_2$	0.28	0.05	0.29	0.06
8	H	0.00	0.00	−0.02	0.00	41	$OCH(CH_3)_2$	0.30	−0.72	0.27	−0.67
9	OH	0.29	−0.64	0.32	−0.60	42	$CH_2Si(CH_3)_3$	−0.15	−0.07	−0.15	−0.07
10	SH	0.28	−0.11	0.27	−0.16	43	Ph	0.08	−0.08	0.09	−0.08
11	NH_2	0.02	−0.68	0.04	−0.63	44	NHPh	−0.02	−0.38	−0.03	−0.39
12	NHOH	0.06	−0.40	0.16	−0.36	45	$NHSO_2Ph$	0.21	−0.18	0.22	−0.15
13	SO_2NH_2	0.41	0.19	0.59	0.20	46	COPh	0.30	0.16	0.32	0.17
14	$N{=}CCl_2$	0.23	−0.08	0.27	−0.08	47	CO_2Ph	0.33	0.13	0.28	0.10
15	CF_3	0.38	0.19	0.36	0.20	48	N=CHPh	0.09	−0.63	0.14	−0.56
16	OCF_3	0.38	0.00	0.36	0.00	49	CH=NPh	0.31	0.13	0.14	0.11
17	SO_2CF_3	0.73	0.26	0.68	0.25	50	NHCOPh	0.09	−0.27	0.05	−0.33
18	CN	0.51	0.19	0.47	0.16	51	C≡CPh	0.12	0.05	0.19	0.09
19	NCS	0.51	−0.09	0.36	−0.12	52	CH=CHPh	0.06	−0.12	0.06	−0.15
20	CHO	0.31	0.13	0.25	0.14	53	$B(OH)_2$	−0.07	0.18	0.02	0.21
21	CH_2Br	0.10	0.05	0.09	0.04	54	CH_2OH	0.00	0.00	0.00	0.00
22	$CONH_2$	0.24	0.14	0.27	0.11	55	SCH_3	0.20	−0.18	0.18	−0.15
23	CH=NOH	0.25	−0.13	0.34	−0.11	56	$CH{=}CH_2$	0.07	−0.08	0.05	−0.07
24	CH_3	−0.04	−0.13	−0.08	−0.14	57	$NHCOCH_3$	0.28	−0.26	0.15	−0.28
25	$NHCONH_2$	0.04	−0.28	0.14	−0.28	58	$N(CH_3)_2$	0.10	−0.92	0.04	−0.87
26	$SOCH_3$	0.52	0.01	0.45	0.00	59	CH_2OPh	0.02	0.02	0.07	0.04
27	SO_2CH_3	0.54	0.22	0.53	0.25	60	$NHNH_2$	0.17	−0.71	0.07	−0.86
28	OSO_2CH_3	0.39	0.00	0.43	0.01	61	$OCOCH_3$	0.41	−0.07	0.38	−0.07
29	$NHCH_3$	−0.11	−0.74	0.02	−0.78	62	N=NPh	0.28	0.13	0.32	0.11
30	$NHSO_2CH_3$	0.25	−0.20	0.32	−0.17	63	CH_2OCH_3	0.01	0.02	0.00	0.01
31	C≡CH	0.19	0.05	0.27	0.05	64	$CH_2CH_2CH_3$	−0.06	−0.08	−0.03	−0.06
32	CH_2CN	0.21	−0.18	0.18	−0.18	65	$C(CH_3)_3$	−0.07	−0.13	−0.06	−0.12
33	$COCH_3$	0.32	0.20	0.21	0.17	66	$Si(CH_3)_3$	−0.04	−0.04	−0.00	−0.03
					Testing Set						
1	Cl	0.41	−0.15	0.39	−0.21	12	SCH_2CH_3	0.23	−0.18	0.22	−0.06
2	I	0.40	−0.19	0.44	−0.15	13	$CH(CH_3)_2$	−0.05	−0.10	−0.09	−0.11
3	SCF_3	0.35	0.18	0.51	0.14	14	OPh	0.34	−0.35	0.31	−0.39
4	SCN	0.36	0.19	0.26	0.32	15	SO_2Ph	0.56	0.18	0.60	0.19
5	CO_2H	0.33	0.15	0.35	0.15	16	$N(Ph)_2$	0.07	−0.29	−0.05	−0.14
6	CH_2Cl	0.10	0.03	0.05	0.03	17	NHCN	0.26	−0.18	0.26	0.00
7	CH_2I	0.09	0.03	0.11	0.07	18	NO	0.50	0.45	0.55	0.27
8	NHCHO	0.25	−0.23	0.10	−0.24	19	OCH_3	0.26	−0.51	0.28	−0.67
9	CF_2CF_3	0.44	0.11	0.27	0.33	20	OCOPh	0.23	−0.08	0.31	−0.03
10	$SCOCH_3$	0.36	0.11	0.35	0.26	21	CH_2Ph	−0.08	−0.01	−0.10	0.11
11	CO_2CH_3	0.33	0.15	0.38	0.12						

Substituents and their experimental inductive and resonance σ constants were selected from Hansch and Leo.[21] This set of functional groups was divided into a disjoint training set (66 functional groups) and testing set (21 functional groups); see Table II. Substituents with complicated cyclic substructures have been excluded from the sets. The choice for the testing set was directed to obtain a sufficiently diverse variety of types of substituents. The used feed-forward neural network is composed of three layers. The juxtaposed input, hidden, and output layers are fully connected by oriented connections. The input layer is composed of 14 neurons that correspond to descriptors (input activities). The output layer contains two neurons; their activities correspond to inductive and resonance σ constants. We have varied the number of neurons in the hidden layer from 6 to 10, and best results have been obtained for $N_H = 8$. Since the σ constants range from -1 to 1, the constants A and B from the transfer function 3 should be fixed as $A = -1$ and $B = 1$. The adaptation process based on formula 6 is carried out for $\lambda = 0.05$ and $\mu = 0.8$. After 5000–8000 iterations, values of $E = 0.1$ and $|\text{grad } E| = 10^{-4}$ are achieved. The σ constants for substituents from the training set as well as the testing set predicted by adapted neural network are given in Table II.

Table III. Statistical Interpretation of Results from Table II

	std error	mean error	mean error dev	correl coeff
		Training Set		
σ_I	0.064	0.049	0.042	0.957
σ_R	0.031	0.021	0.023	0.994
		Testing Set		
σ_I	0.077	0.057	0.052	0.924
σ_R	0.108	0.083	0.070	0.905

The correlation coefficient for both sets is above 0.9 (see Table III). A small anomaly exists in these coefficients: the greater the correlation between the required and computed resonance σ constants for the training set, the lower the correlation between the required and computed inductive constants for the testing set. The same observation holds for errors. Naturally, the results are better for training than for testing sets.

Cluster Analysis of Activities of Hidden Neurons for Different Substituents

An application of three-layer feed-forward neural networks for classification and prediction purposes offers as a byproduct the

(20) (a) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219. (b) Hutchings, M. G.; Gasteiger, J. *Tetrahedron Lett.* **1983**, *24*, 2541. (c) Gasteiger, J.; Saller, H. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687.

(21) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley: New York, 1979.

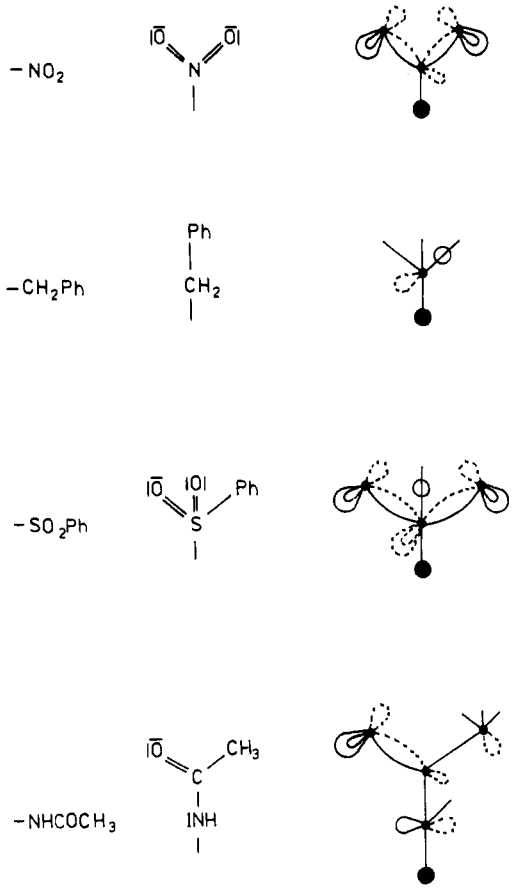

Figure 4. Four illustrative examples of labeled rooted molecular graphs assigned to univalent functional groups.

very interesting possibility of how to use hidden activities for a cluster analysis[10] of functional groups. These hidden activities may be formally considered[2] as an internal representation of classified functional groups produced by the adapted neural network. Let us assume that a functional group (indexed by $1 \le k \le N$) is represented by a vector composed of hidden activities,

$$\mathbf{x}_k = (x_1^{(k)}, x_2^{(k)}, ..., x_{N_H}^{(k)}) \tag{7}$$

The set of vectors $\{\mathbf{x}_k; k = 1, 2, ..., N\}$ of hidden activities assigned to all substituents is normalized by standard statistical procedure:

$$\tilde{x}_i^{(k)} = \frac{x_i^{(k)} - \bar{x}_i}{\sigma_i} \text{ (for } k = 1, 2, ..., N \text{ and } i = 1, 2, ..., N_H) \tag{8}$$

where $\bar{x}_i$ is the arithmetic mean and σ_i is the standard deviation, both assigned to the ith component of all vectors,

$$\bar{x}_i = \frac{\sum_k x_i^{(k)}}{N} \qquad \sigma_i^2 = \frac{\sum_k (x_i^{(k)} - \bar{x}_i)^2}{N - 1} \tag{9}$$

The distance between two normalized vectors $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{x}}_l$ is determined by

$$d(\tilde{\mathbf{x}}_k, \tilde{\mathbf{x}}_l) = [\sum_i (\tilde{x}_i^{(k)} - \tilde{x}_i^{(l)})^2]^{1/2} \tag{10}$$

The used cluster analysis is based on the following recurrent procedure. First (initialization), we create from each normalized vector $\tilde{\mathbf{x}}_k$ a cluster; i.e., in this step we have N clusters composed of one functional group represented by a normalized vector. Two clusters with minimum distance (calculated by (10)) are amalgamated in a new cluster, and both original clusters are removed from the forthcoming considerations. A vector assigned to this new cluster is calculated as an arithmetic mean of vectors of all functional groups that belong to the cluster. This procedure is recurrently repeated until we have achieved a prescribed number $N' < N$ of clusters (see Figure 5). The resulting clustering of

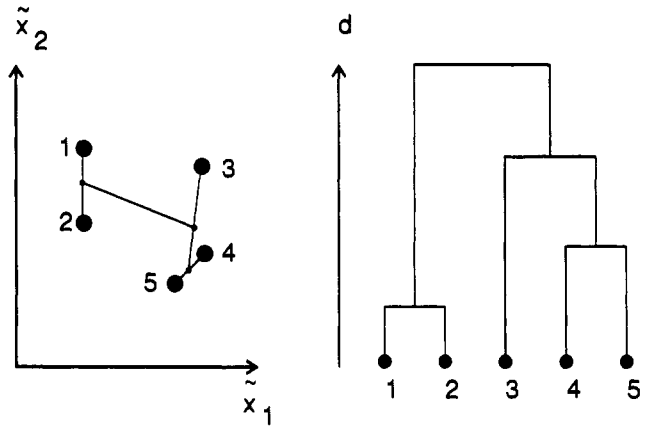

Figure 5. Demonstration of cluster analysis of five objects represented by dots in a plane $\tilde{x}_1$–$\tilde{x}_2$. In the first step, objects 4 and 5 are joined into a new cluster represented by a dot placed at middle position of 4 and 5. In the second step, objects 1 and 2 are joined into a new cluster. In the third step, cluster {4,5} and object 3 are joined together. Finally, in the fourth step, clusters {1,2} and {3,4,5} are joined into a cluster composed of all five objects.

Table IV. Clustering of Functional Groups into 10 Clusters

cluster	functional groups
1	{Br, Cl, F, I, OH, SH, OCH_3, SCH_3, OCH_2CH_3, OPh, SCH_2CH_3}
2	{SO_2F, IO_2, NO, NO_2, $B(OH)_2$, NHOH, SO_2NH_2, CN, CHO, CO_2H, $CONH_2$, $SOCH_3$, SO_2CH_3, $COCH_3$, NNPh, $SO_2CH_2CH_3$, SO_2Ph, COPh}
3	{SF_5, CF_3, CH_2Br, CH_2Cl, CH_2I, CO_2CH_3, $PO(OCH_3)_2$, CO_2Ph}
4	{N_3, NCS, SCN, NHCN, NHCHO, $NHCONH_2$, $SCOCH_3$, OSO_2CH_3, $NHSO_2CH_3$, $OCOCH_3$, $NHCOCH_3$, OCOPh, $NHSO_2Ph$, NHCOPh}
5	{H, NH_2, $NHNH_2$, CH=NOH, CH_2OH, $NHCH_3$, Ph, $N(Ph)_2$, C≡CH, CH_2OCH_3, $NHCH_2CH_3$, $N(CH_3)_2$, CH_2OPh, C≡CPh, NHPh, N=CHPh, CH=NPh}
6	{$N{=}CCl_2$, OCF_3, SO_2CF_3, SCF_3, CF_2CF_3, $C(OH)(CF_3)_2$}
7	{CH_3, CH=CH_2, CH_2CH_3, $P(CH_3)_2$, $CH(CH_3)_2$, $CH_2CH_2CH_3$, $Si(CH_3)_3$, $C(CH_3)_3$, CH_2Ph, CH=CHPh}
8	{CH_2CN}
9	{$OCH(CH_3)_2$}
10	{$CH_2Si(CH_3)_3$}

Table V. Clustering of Functional Groups into Five Clusters

cluster	functional groups
1	{Br, Cl, F, I, OH, SH, OCH_3, SCH_3, $OCH(CH_3)_2$, OCH_2CH_3, SCH_2CH_3, OPh}
2	{SO_2F, SF_5, IO_2, NO, NO_2, H, $B(OH)_2$, NH_2, NHOH, SO_2NH_2, $NHNH_2$, N=CCl_2, CF_2, OCF_3, SO_2CF_3, SCF_3, CN, CHO, CO_2H, CH_2Br, CH_2Cl, CH_2I, $CONH_2$, CH=NOH, CH_2OH, $SOCH_3$, SO_2CH_3, $NHCH_3$, CF_2CF_3, C≡CH, CH_2CN, $COCH_3$, CO_2CH_3, CH_2OCH_3, $NHCH_2CH_3$, $SO_2CH_2CH_3$, $N(CH_3)_2$, $PO(OCH_2)_2$, $C(OH)(CF_3)_2$, Ph, NNPh, SO_2Ph, NHPh, COPh, CO_2Ph, N=CHPh, CH=NPh, CH_2OPh, C≡CPh, $N(Ph)_2$}
3	{N_3, NCS, SCN, NHCN, NHCHO, $NHCONH_2$, OSO_2CH_3, $NHSO_2CH_3$, $SCOCH_3$, $OCOCH_3$, $NHCOCH_3$, $NHSO_2Ph$, OCOPh, NHCOPh}
4	{CH_3, CH=CH_2, CH_2CH_3, $P(CH_3)_2$, $CH(CH_3)_2$, CH_2Ph, $CH_2CH_2CH_3$, $Si(CH_3)_3$, $C(CH_3)_3$, CH=CHPh}
5	{$CH_2Si(CH_3)_3$}

N functional groups into a smaller number N' of clusters should reflect a similarity between them. All 87 functional groups for 10 and 5 clusters are displayed in Tables IV and V.

Results and Discussion

One of the main problems of applications of neural networks in chemical reactivity studies is a proper representation of functional groups via their descriptors. The initial method suggested by Johnson et al.[9] was expanded in our work. It is also based on

simple atomic and bond labels that reflect their physical nature and in some (additive) manner their topology. Our recent results[5,14] and the results presented in this paper support the assumption that these descriptors are useful for a proper representation of functional groups for purposes of neural network applications.

The descriptors are constructed directly from the structural formulas of functional groups in a way immediately reflecting their structural features. We believe that this is the main superiority of the used descriptors with respect to other potentially applicable descriptors for chemical-reactivity studies (e.g., Ugi's[22] BE matrix used by Elrod[4a,b] or ad hoc chosen physico-chemical parameters used by Luce[4c]). Moreover, for those special cases dealing with specific chemical-reactivity effects (e.g., regioselectivity or site selectivity), the structural descriptors used may be directly enlarged by additional parameters. Those parameters could correspond to the properties of reactants or functional groups determined by the whole structure (e.g., inductive and resonance σ constants, steric hindrance parameters, charges of σ or π skeleton,[20] a kind of topological index, etc.).

Inductive and resonance σ constants represent a very important approach to describe properties of functional groups, in particular their influence to synthons[7,8] undertaking chemical reactions. We have demonstrated that these parameters are useful for classification by feed-forward neural networks. The descriptors are also adequate in the classification of the influence of functional groups to reaction cores. The adapted neural network was tested for its ability to predict inductive and resonance σ constants for 21 functional groups, which form the testing set (see second part of Table II). In most cases, the sign of the predicted σ constants is retained, and for the functional groups $-N(Ph)_2$, -NHCN, and $-CH_2Ph$, one constant has the opposite sign. Two of these groups contain phenyl, and therefore, their precise description should likely involve additional descriptors that reflect more adequately their anomalous physical properties.

Activities of hidden neurons correspond[2] to an internal representation of functional groups. These activities were used for simple cluster analysis of functional groups. That is, the whole training set is divided in disjoint subsets, where these clusters are composed of functional groups with similar hidden-activity vectors. For these purposes, the training set was composed of all 87 functional groups (see Table II), and the neural network was again subjected to the adaptation process. After 5000 iterations, the value of the objective function yielded $E = 0.2$ and $|\text{grad } E| = 10^{-4}$. The hidden activities of all 87 functional groups were used as an input for cluster analysis; the obtained results for 5 and 10 clusters are presented in Tables IV and V. The obtained clustering of functional groups is chemically plausible as they have similar structural features and similar inductive and resonance properties. Moreover, the present results are closely related to the cluster analysis of functional groups carried out by Hansch and Leo[21] based on different properties that characterize the functional groups (lipophilicity, molecular refractivity, inductive and resonance σ constants, and binary H-bonding).

In conclusion, it seems that the neural network approach gives both a useful and simple mathematical model for the classification and prediction of molecular properties manifested by organic chemical reactivity. This approach allows the construction of formal methods that solve in some binary manner the difficult problem of chemical reactivity for purposes of computer-aided organic synthesis design.

(22) Ugi, I.; Bauer, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, V. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 111.

The Equilibrium Conformation of Ethyl Isocyanate Revisited

Miklós Fehér,* Tibor Pasinszki,† and Tamás Veszprémi†

Contribution from the Institut für Physikalische Chemie, Universität Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland, and Inorganic Chemistry Department, The Technical University of Budapest, H-1521 Budapest, Hungary. *Received June 12, 1992*

Abstract: The equilibrium conformation of ethyl isocyanate was reinvestigated by molecular orbital calculations using second-order Møller-Plessett perturbation theory and a 6-31G** basis set. It was shown that the inclusion of electron correlation is of crucial importance in the prediction of the equilibrium structure for this molecule. According to the calculations, there are two stable conformers, the gauche and the trans, of which the gauche has lower energy. The cis form does not correspond to a local energy minimum. A new interpretation for the experimental microwave spectra is suggested. The known infrared spectrum was also assigned using frequency calculations.

Introduction

The geometric structure of pseudohalides is still subject to uncertainty in those cases where experimental information on the geometry is not available. This is caused by the fact that ab initio calculations seem to have difficulties in predicting the correct geometry of these molecules.[1-3] The most probable reason for this effect is that the potential energy curve of pseudohalides is rather shallow around the minimum.[1,2] In order to circumvent this problem, it is often necessary to use higher than usual accuracy in the gradient calculation.

For all pseudohalides investigated so far, it was necessary to include polarization functions and correlation energy also had to be incorporated into the calculations to obtain geometries similar to the experimental ones. On using Gaussian basis sets of quality 3-21G or better in an SCF calculation but without polarization functions and including the effects of correlation, always a linear C-N-C and N-C-O angle was obtained for methyl and tertiary-butyl isocyanates.[1,2] In a publication by Sullivan, Durig, Durig, and Cradock,[4] however, ab initio calculations predicted these angles to be less than 180° in ethyl isocyanate using a 3-21G basis

* Author to whom correspondence should be addressed at Universität Basel.
† The Technical University of Budapest.

(1) Veszprémi, T.; Pasinszki, T.; Fehér, M. *J. Chem. Soc., Faraday Trans. 1* **1991**, *87*, 3805.
(2) Pasinszki, T.; Veszprémi, T.; Fehér, M. *Chem. Phys. Lett.* **1992**, *189*, 245.
(3) Mack, H. G.; Oberhammer, H. *Chem. Phys. Lett.* **1989**, *157*, 436.
(4) Sullivan, J. F.; Durig, D. T.; Durig, J. R.; Cradock, S. *J. Phys. Chem.* **1987**, *91*, 1770.

0002-7863/93/1515-1500$04.00/0