Acta Chimica Hungarica – **Models in Chemistry** *130 (1), pp. 103–110 (1993)*

PREDICTION OF REGIOSELECTIVITY OF 1,3-DIPOLAR CYCLOADDITION REACTIONS BY NEURAL NETWORKS

Š. SKLENÁK, V. KVASNIČKA and J. POSPÍCHAL

Department of Mathematics, Faculty of Chemical Technology, Slovak Technical University, 81237 Bratislava, Czechoslovakia

Received September 14, 1992
In revised form November 13, 1992

Simple feed-forward three-layer neural networks adapted by back propagation algorithm are used for classification and prediction of regioselectivity of 1,3-dipolar cycloaddition reactions. The structure of products and reactants is determined by descriptors that correspond to simple fragments of structural formulae treated as molecular graphs with additional evaluation of bonds and atoms.

Introduction

Recent progress [1] of neural network paradigm, in particular, the layered feed-forward networks adapted by back-propagation strategy [2], offers new mathematical techniques equipped by learning features. This approach allows computers to be trained to recognize patterns in data of high dimensionality. Their impact on chemistry [3] is manifested not only by effective correlations between molecular structure and activity but also by simple algorithmic tools for building expert systems that, in turn, are able to classify [4,5] chemical reactions semiquantitatively or qualitatively. These expert systems can be used to address the very important question of whether a reaction will proceed, and if it will, they can be used to classify the reaction products. Such an expert system is vital in any computer system for organic synthesis design.

Neural networks, inspired by simplified models of human brain, are mathematical models used as computing systems for a broad variety of applications less tractable by other computational methods. Further description concerns mainly with one of the most common types of neural networks; the multilayered back-propagation neural network. A network does not have any previously given information about a problem in the form of formulas and rules. Instead, it is given a set of groups of inputs with a corresponding set of desired groups of outputs. The

Akadémiai Kiadó, Budapest

network then trains its parameters to match its outputs with the desired ones for given input information. Quality of matching is measured by objective functions. Such a trained network can be used for solving another problem from the same area, the solution (output information) of which is unknown.

Networks are composed of interconnected neurons. Their behavior is usually simulated by a software. Every input neuron receives an input information (a number), and accordingly sends signals to neurons in the hidden layer. Each neuron in the hidden layer sums the received signals and according to this sum sends signals to neurons of the next (hidden or output) layer. Each neuron in the output layer gives the sum of the received signals as its output information. Training procedure adjusts weight of each connection which multiply carried signal in the particular connection. Other adapted parameters are threshold coefficients, which decide for each neuron, whether the signal is substantial enough to send it over to other neurons. This adaptation is done automatically by a program during the learning process.

Choosing the right architecture of neural network for a particular problem is the major condition for its successful application. There is no precise set of instructions to be followed. However, there is a number of relevant literature on neural networks [1–5].

The purpose of the present communication is to demonstrate an interesting application of neural networks for prediction of regioselectivity of 1,3-dipolar cycloadditions [6]. The regioselectivity of these organic reactions can be explained using simple concepts of theory of molecular orbitals [7]. Neural network approach offers an alternative possibility how to solve the problem of regioselectivity of 1,3-dipolar cycloadditions. An input information (descriptors) to the neural network is the inductive F and resonance R constants [8] of functional groups, which properly characterize the neighbourhood of a fixed reaction core.

The 1,3-dipolar cycloadditions may give two products (Fig. 1) where the labels A, B, C, D, and E correspond to atoms of carbon and nitrogen (or to other atoms) and R_1 to R_5 are functional groups usually identified with alkyl, aryl or other groups. Applying neural network approach, we would like to predict which of two possible isomers is formed. If both types of isomers are experimentally observed, then the neural network is trained for the isomer with considerably greater yield. In the present communication we have used only 1,3-dipolar cycloadditions with diazomethane dipole [9].

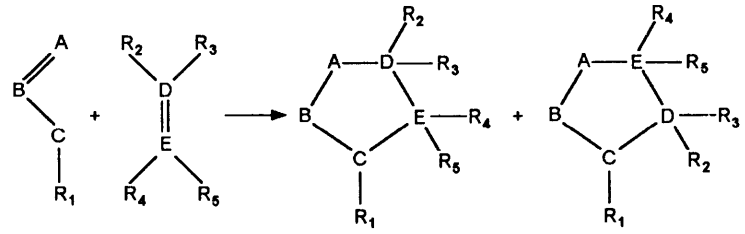


Fig. 1. General scheme of 1,3-dipolar cycloadditions

Design of neural network

The feed-forward neural network used in our study [2] contained three layers composed of four input neurons, five hidden neurons, and one output neuron, respectively. The descriptors specifying functional groups are inductive (F) and resonance (R) constants often used in physical organic chemistry for quantification of influence of functional groups on benzene ring. The chemical reactivity of diazomethane and its derivatives, arbitrarily chosen for our study, is similar. Therefore, we have considered all used diazomethanes as the same chemical entity. This considerable simplification is supported by the observation that HOMO and LUMO coefficients on strategic atoms and the corresponding energies are very similar for all derivatives of diazomethane.

It is clear that the regioselectivity is driven by HOMO and LUMO orbitals, which are influenced by substituents near the reaction core. The inductive and resonance constants are the most obvious descriptors of electronic influence of the substituent, so it is probable that they will influence also HOMO and LUMO and therefore regioselectivity, too. However, this dependence is not straightforward and cannot be described by a simple formula. Therefore, we have used a neural network to find this relation.

As an input information (descriptors) was used only inductive F and resonance R constants of functional groups of strategic double bond C=C of dipolarophile (Fig. 2). A general scheme of a product of 1,3-dipolar cycloaddition is displayed in Fig. 3, where R' and R" are functional groups placed on diazoalkane R'–(R")C=N=N. For 1,3-cycloaddition, two isomers may exist (Fig. 3, schemes **A** and **B**). We have to stress that isomer **A** is transformed to isomer **B** by changing the functional groups $R_1 \leftrightarrow R_4$ and $R_2 \leftrightarrow R_3$.

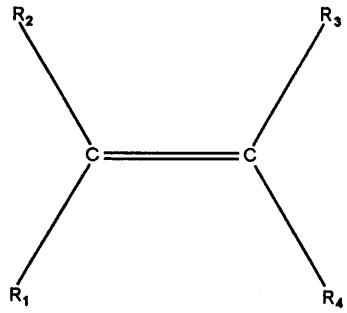
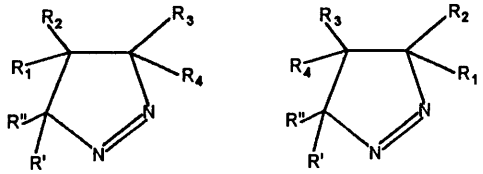


Fig. 2. Labeling of functional groups in dipolarophile



A **B**

Fig. 3. Labeling of functional groups in product isomers

The functional groups R_1, R_2, R_3, and R_4 may be for both templates assigned to actual substituents of the dipolarophile (see Fig. 3). In our forthcoming considerations, to keep the uniqueness of this assignment, the functional groups R_1, R_2, R_3, and R_4 will be related only to the template in Fig. 3, scheme **A**. Figure 4 displays both formed ($R_1=R_2=H$ $R_3=CH_3$, $R_4=COOCH_3$) and forbidden ($R_1=COOCH_3$, $R_2=CH_3$, $R_3=R_4=H$) products.

Four descriptors (input activities of the used neural network) are determined as follows:

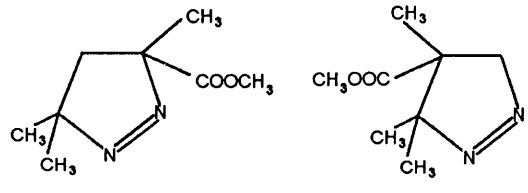
d_1 = sum of inductive constants of R_1 and R_2,

d_2 = sum of resonance constants of R_1 and R_2,

d_3 = sum of inductive constants of R_3 and R_4,

d_4 = sum of resonance constants of R_3 and R_4.

Actual numerical values of F and R constants are taken from monograph [8].



A **B**

Fig. 4. Examples of allowed (**A**) and forbidden (**B**) products

Example

F and R constants for $-H$, $-CH_3$, and $-COOCH_3$ are [8]

$F_H=0$, $F_{CH_3}=-0.04$, $F_{COOCH_3}=0.33$

$R_H=0$, $R_{CH_3}=-0.13$, $R_{COOCH_3}=0.15$.

This means that molecular structures displayed in Fig. 4 are determined by descriptors,

$d_1=0.0$, $d_2=0.0$, $d_3=0.29$, $d_4=0.02$ (allowed)

$d_1=0.29$, $d_2=0.02$, $d_3=0.0$, $d_4=0.0$ (forbidden).

The allowed isomer was trained for output equal to 1, the forbidden isomer was trained for output equal to 0. When there was only one possible isomer ($R_1=R_4$ and $R_2=R_3$), the neural network was twice trained for output equal to 1.

Selection of 1,3-dipolar cycloadditions

An arbitrarily chosen set of 42 1,3-dipolar cycloadditions [9] was divided into two disjoint subsets. The training subset was composed of 35 reactions while the remaining reactions formed the test set (Table I and Fig. 3, scheme **A**). In some special cases the presented products correspond to unstable intermediates transformed by the forthcoming reaction steps into more stable compounds. Moreover, in our study we did not consider the possible different tautomeric forms of the products.

Table I

Allowed products of 1,3-cycloaddition

No.	R', R"	R_1, R_2	R_3, R_4	Obs.[1]	Pred.[2]
		Training set			
1	CH_3, CH_3	H, H	CH_3, $COOCH_3$	A	A
2	H, H	H, H	$(CH_2)_{1-4}$-$COOCH_3$, $COOCH_3$	A	A
3	H, H	H, H	SCH_3, $COOCH_3$	A[3]	A
4	H, H	H, H	H, CH_2-CN	A[4]	A
5	H, H	H, H	H, $CH(OEt)_2$	A	A
6	CH=C(CN)COOEt, H	H, H	H, $COOCH_3$	A[4]	A
7	H, H	H, H	H, H	A	A
8	H, H	H, H	n-C_4H_9, H	A	A
9	H, H	H, H	CH_2-Ph, H	A	A
10	H, H	H, H	H, Ph,	A	A
11	H, H	Ph, H	H, $COOCH_3$	A[4]	A
12	H, H	H, H	H, Ph-R	A	A
13	H, H	Ar, H	H, Ar	A[4]	A
14	$(CH_3)_3Si$, H	H, H	H, CN	A[4]	A
15	Ph, $PO(MeO)_2$	H, H	H, $COCH_3$	A[4]	A
16	H, Ph	H, H	Ph, COOEt	A[3]	A
17	COOEt, H	H, H	Cl, COH	A[3]	A
18	H, H	Ph, H	H, $COOCH_3$	A[4]	A
19	CH_3, CH_3	CH_3, H	CH_3, $COOCH_3$	A	A
20	H, H	CH_3, CH_3	CN, CN	A	A
21	H, H	H, Ph	CN, $CONH_2$	A	A
22	H, H	H, Ph	COOEt, COOEt	A	A
23	H, H	H, i-Pr	CN, $COOCH_3$	A	A
24	H, H	H, t-Bu	CN, CN	A	A
25	H, H	H, p-Cl-Ph	$COCH_3$, $COOCH_3$	A	A
26	H, H	H, Ph	CN, $PO(OEt)_2$	A	A
27	H, H	H, Ph	COPh, $COCH_3$	A	A
28	H, H	H, $COOCH_3$	$COOCH_3$, $COOCH_3$	A	A
29	H, H	F, Ph	CN, CN	A[3]	A
30	H, H	F, Ph	CN, COOEt	A[3]	A
31	H, H	Cl, Ph	CN, COOEt	A[3]	A
32	H, H	F, Ph	COOEt, COOEt	A[3]	A
33	H, H	Ph, H	NO_2, CH_3	A	A
34	H, H	Ph, H	NO_2, Br	A[3]	A
35	H, H	H, H	H, $PO(OEt)_2$	A	A
		Testing set			
36	CF_3, H	H, H	CH_3, H	A	A
37	H, H	H, H	H, OEt	A	A
38	COOEt, H	H, H	H, COH	A[4]	A
39	COOEt, H	H, H	CH_3, COH	A[3]	A
40	H, H	Ph, H	H, COPh	A[4]	A
41	CH_3, H	H, p-NO_2-Ph	$COOCH_3$, $COOCH_3$	A	A
42	H, H	Cl, Ph	CN, CN	A[3]	A

[1]observed isomer; [2]predicted isomer; [3]unstable intermediate product that undergoes further reaction; [4]product is converted to the other tautomer by a shift of the proton and the double bond between two nitrogens

Results and discussion

The 3-layer neural network has been adapted by standard back-propagation method [2] for all 35 reactions (patterns) which form the so-called training set. For an initialization of the neural network we have used random numbers. The dipole was always diazomethane or its simple derivatives for all these reactions. The hidden layer was composed of 4 to 6 hidden neurons, the best classification of reactions from the training set was achieved with 5 hidden neurons, all 35 (training set) reactions having been correctly classified. The adaptation was finished with the value of minimized objective function (error) $E = 1.3 \times 10^{-7}$.

The output of the neural network is given by the value of the single output neuron. This neuron was trained to give a binary answer yes/no. When the result should be yes, i.e., 1, the worst observed difference for the reported results of neural network was on the fourth place (0.9997). For the "no" answer, when the other isomer was observed, the difference was similar. As there is no predefined way how to describe two possible isomers, in Table I the observed isomer is always described in such a way, that its description corresponds to pattern **A**. Of course, the neural network was trained to predict both 1's and 0's for the observed and unobserved isomer, respectively.

This simplification is admissible, because arbitrarily chosen experimental results report only one of two possible isomers. The predicted products in the whole Table I correspond to observed products (marked **A**). The other isomer **B** was neither observed, nor predicted in the cases studied. It should be mentioned, that for differently substituted diazomethanes the other isomer could also be produced in some cases. However, this is a rare case and the yield of the concurrent isomer is generally not significant. As we did not use descriptors for substituents of diazomethane, we would not be able to distinguish these cases.

From the results it can be concluded that inductive and resonance constants are powerful enough to describe the basic influence of substituents on regioselectivity. Other basic descriptors like steric effect were not necessary. More descriptors may describe substituents more adequately, but our goal was to prove that even with substantial simplifications, neural networks can produce reasonable answers.

The adapted neural network with 5 hidden neurons was used for the test set. All 7 reactions have been correctly classified. The value of objective function (error)

was $E = 1.7 \times 10^{-10}$, which is accidentally turned to be better value than for the training set.

The neural network simulation software was written by the authors in TurboPascal, it is described in [10] and available by request.

References

[1] ECKMILLER, R., MALSBURG, C. v. D. (Eds), *Neural Computers*, Springer-Verlag, Berlin, 1988
[2] RUMELHART, D. E., MCCLELLAND, J. L.: *Parallel Distributed Processes*, Vols 1–2, MIT Press, Cambridge, MA, 1986
[3] ZUPAN, J., GASTEIGER, J.: *Anal. Chim. Acta*, **248**, 1 (1991)
[4] ELROD, D. W., MAGGIORA, G. M., TRENARY, R. G.: *J. Chem. Inf. Comput. Sci.*, **30**, 477 (1990)
[5] ELROD, D. W., MAGGIORA, G. M., TRENARY, R. G.: *Tetrahedron Comput. Methodol.*, **3**, 163 (1990)
[6] PADWA, E. (Ed.): *1,3-Dipolar Cycloaddition Chemistry*, Vols 1–2, John Wiley & Sons, New York, 1984
[7] FLEMING, I.: *Frontier Orbitals and Organic Chemical Reactions*, John Wiley & Sons, Chichester, Sussex, 1976
[8] HANSCH, C., LEO, A.: *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, New York, 1979
[9] REGITZ, M., HEYDT, H.: In *1,3-Dipolar Cycloaddition Chemistry*, Vol. 1, Chap. 4., John Wiley & Sons, New York, 1984
[10] KVASNIČKA, V.: *J. Math. Chem.*, **6**, 63 (1991)