Quantitative structure–conditions–property relationship studies. Neural network modelling of the acid hydrolysis of esters

Nataliya M. Halberstam, Igor I. Baskin, Vladimir A. Palyulin and Nikolai S. Zefirov*

Department of Chemistry, M. V. Lomonosov Moscow State University, 119992 Moscow, Russian Federation.
E-mail: vap@org.chem.msu.su, zefirov@org.chem.msu.su

10.1070/MC2002v012n05ABEH001620

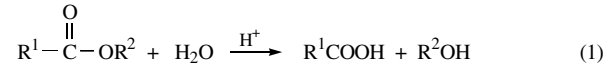
A new approach to the quantitative description of organic compound reactivity based on the neural network modelling of structure–reaction conditions–rate constant relationships was demonstrated for the acid hydrolysis of esters.

Nowadays, when considerable amounts of experimental data have been accumulated in virtually all areas of chemistry, there is the acute need of processing huge and disembodied data sets for the purpose of predicting various properties of chemical compounds. The use of artificial neural networks is one of the most promising approaches to this processing.[1] A main advantage of neural networks is their ability of building non-linear multiparametric models for structure–property relationships with an unknown generic type.

Since typically a property of a chemical compound depends on medium conditions non-linearly, we presume that it is possible to expand the neural network approach to structure–property relationships by adding characteristics of a medium to input parameters of neural networks.[2,3] Thus, properties can be predicted under different conditions. Such characteristics of a medium could be, for example, temperature, pressure, concentration of a component, solvent composition, *etc*.

Previously,[4] we have suggested a concept of building structure–conditions–property relationship models and demonstrated its validity for predicting physico-chemical properties of hydrocarbons under different conditions (temperature, pressure, *etc.*). In this work, this concept is applied to predicting the reactivity of chemical compounds, as exemplified by predicting the kinetics of acid hydrolysis of esters at various temperatures and solvent compositions.

The kinetics and the mechanisms of the acid hydrolysis of carboxylic acid esters are of considerable attention to chemists. The general equation of this process is as follows:



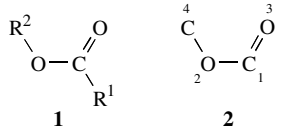
$$R^1-\overset{\overset{O}{\|}}{C}-OR^2 + H_2O \xrightarrow{H^+} R^1COOH + R^2OH \qquad (1)$$

On the basis of this reaction series, Hammett has suggested an equation for the quantitative description of electronic effects of substituents for *para*- and *meta*-substituted phenols.[5] Later, Taft has utilised the reactions of the acid and base hydrolysis of substituted acetic acid esters for describing steric effects of substituents.[6,7] Nowadays, a huge amount of data on reaction rate constants for different temperatures and characteristics of media has been accumulated for a great variety of esters.[8]

However, in spite of a huge amount of experimental data, all attempts of building linear models for predicting rate constants of the acid hydrolysis of esters has been undertaken only for very small series of compounds under constant reaction conditions.[9–11]

In this work, we have explored the possibility of generalising information on the hydrolysis kinetics of carboxylic acid esters. We have utilised data[8] on reaction rate constants measured at different temperatures (from 0 to 154 °C) in binary water–solvent mixtures (non-water component concentration varied from 0 to 98%) for developing a neural network model. The logarithms of the hydrolysis rate constant for esters with different substituents lie in a range from –7.53 to –0.17.

Since our primary task was the development of models that could be applied to any carboxylic acid ester, we used theoretically derived descriptors for describing the electronic structure of esters rather than experimental ones known only for a small number of compounds. For this purpose, we selected quantum-chemical descriptors, which were computed by the PM3 method with geometry optimization. We computed the energies of frontier orbitals as molecular descriptors; we also calculated minimum, maximum and average values (for all atoms in the molecule, as well as for the separate atom types, such as carbon, oxygen, nitrogen and sulfur) for electronic densities and superdelocalizabilities of atoms and bonds, atomic charges and free valency indexes. For calculating local descriptors for all esters **1**, maximum common fragment **2** was chosen, and a set of the following descriptors was computed for each of its atoms: the Mulliken atomic charge, electrophilic and nucleophilic electron densities and superdelocalizabilities, the free valency index.[12]



1 **2**

For describing reaction conditions, we used the temperature and concentration of organic solvent components. To describe the properties of organic solvents, the parameters proposed by Palm,[13] namely, general acidity, general basicity, polarity, and

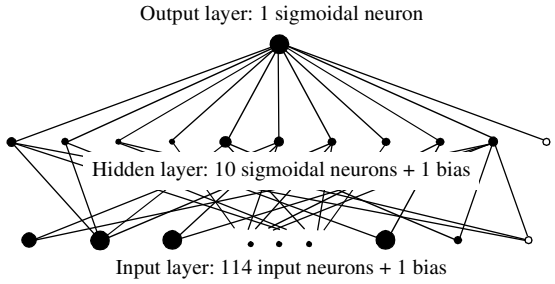


Figure 1 Architecture of the neural net used for modelling.

Table 1 Neural network modelling of acid hydrolysis rate constants (compounds were randomly selected from the validation set).

Record no.	R^1	R^2	*t*/°C	Solvent	C^a_{org}	lg k_{exp}	lg k_{calc}
319	CH_2Cl	CH_2Cl	5	water	—	–5.10	–5.09
648	$(CH_2)_7OMe$	Me	25	water–methanol	50.0	–4.35	–4.23
712	*p*-$MeOC_6H_4$	Et	99.4	water–ethanol	31.6	–4.25	–4.15
1139	$(CH_2)_2CO_2Et$	Me	50	water–acetone	7.2	–3.49	–3.75
1463	Me	$CHClCH_2Cl$	25	water–acetone	23.7	–5.02	–4.78
1740	CH_2CH_2Br	Et	50	water–1,4-dioxane	23.5	–4.08	–4.13
1927	(*trans*-2-COOH)C_6H_{11}	Me	25	water–acetone	36.2	–3.37	–3.35
2060	Me	CH_2Ph	40	water–glycerol	5.6	–3.45	–3.54

[a] C_{org} is the concentration (in molar %) of an organic component in the solvent.

Table 2 Sensitivity values of main descriptors.

Sensitivity	Descriptor
868.6	Temperature
747.2	Charge on the α-carbon atom in substituent R[1]
663.2	LUMO electron density at the carbonyl oxygen atom
–542.7	Nucleophilic delocalisability at the carbonyl oxygen atom
341.3	Polarizability (*P*)
124.8	General acidity (*E*)
–111.2	General basicity (*B*)
65.2	Polarity (*Y*)
–64.9	Concentration of the solvent organic component

polarizability, were used. For neural network modelling, we utilised experimental data on hydrolysis rate constants measured in the following solvents and solvent mixtures:[8] water, water–methanol, water–ethanol, water–ethylene glycol, water–acetone, water–1,4-dioxane, water–dimethylsulfoxide and water–glycerol. Since water was one of the components in all cases, only the concentration of an organic component was used for describing the solvent mixture composition.

In this work, a data set containing 2092 records was used. Each record contained the logarithm of the experimental hydrolysis rate constant and 114 descriptors characterising chemical structures and reaction conditions. The whole set of records was split randomly into a training set (1883 records) and a validation set (209 records).

A feed-forward three-layered network with the 114-10-1 architecture was used in this study (Figure 1). One bias neuron was added to the input layer, and one, to the hidden layer.[1] The network was trained using the generalised delta-rule algorithm with a learning factor of 0.25 and a momentum of 0.9. The predictive performance of the network was assessed by computing the root-mean-square errors of prediction for records from the validation set. The training was stopped after the beginning of overtraining, and a model providing the lowest prediction errors on the validation set was chosen. It was characterised by the following parameters (see also Table 1 and Figure 2): $R = 0.9669$, $S_t = 0.2710$, $S_v = 0.3417$, where R is the correlation coefficient, S_t and S_v are root-mean-square errors on the training and validation sets, respectively (in log units of reaction rate constant).

For each input neuron of the trained neural network, the mean value of the first derivative of the logarithm of the hydrolysis rate constant with respect to the corresponding descriptor was computed.[14] The values for the most informative descriptors are listed in Table 2. The analysis of hydrolysis condition descriptors indicates that the most important influence on the rate constant is exerted by the temperature. The rise in temperature leads to acceleration of the reaction, while an increase in the concentration of an organic component in a binary mixture with water leads to the opposite effect. An analogous effect of the concentration of an organic component on the hydrolysis reaction rate was observed previously.[15] This effect can be explained by the slowing down of the first reaction step (ester protonation) as a result of a decrease in the concentration of protonating ions in the reaction solution. From the four parameters characterising the influence of solvent, the main contribution belongs to polarizability *P*, which is followed by general acidity *E*, general basicity *B* and polarity *Y*. The influence of the polarizability, acidity and polarity on the reaction rate is positive, while the influence of the basicity is negative. We suggest the following putative explanation of this fact: a solvent with high polarizability and acidity facilitates the second irreversible step of the hydrolysis reaction (the C–O bond cleavage) and stabilises the resulting carbonyl ion.

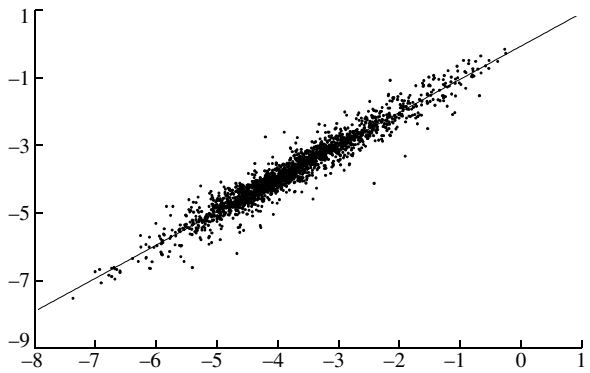


Figure 2 Results of the neural network modelling of hydrolysis rate constant (abscissa, experimental values of the logarithms of the rate constant; ordinate, calculated values).

Analysis of the most informative quantum-chemical descriptors revealed a significant positive influence of the charge on the α-carbon atom in substituent R[1], as consistent with generally accepted ideas of the reaction mechanism.[16] In addition, a considerable positive influence of the LUMO electron density at the carbonyl oxygen atom and a considerable negative influence of the nucleophilic delocalisability at the same atom on the reaction rate also take place.

Thus, the use of artificial neural networks makes it possible to predict hydrolysis rate constants for arbitrary esters at arbitrary temperature and solvent composition with a high precision, as well as to analyse the revealed relationships. The results of this study demonstrate that this approach can be used for developing reliable quantitative models for the reactivity of organic compounds.

This work was supported by INTAS (grant no. 00-0363).

References

1 J. Zupan and J. Gasteiger, *Neural Networks in Chemistry*, Wiley-VCH, Weinheim, 1999.
2 N. M. Halberstam, I. I. Baskin, V. A. Palyulin and N. S. Zefirov, *Trudy VII Vserossiiskoi konferentsii 'Neirokomp'yutery i ikh primenenie'* (*Proceedings of the VII All-Russian Conference 'Neurocomputers and Their Applications'*), Moscow, 2001, p. 423 (in Russian).
3 N. M. Halberstam, I. I. Baskin, V. A. Palyulin and N. S. Zefirov, *Tezisy II Vserossiiskoi konferentsii 'Molekulyarnoe modelirovanie'* (*Abstracts of the II All-Russian Conference 'Molecular Modelling'*), Moscow, 2001, p. 62 (in Russian).
4 N. M. Halberstam, I. I. Baskin, V. A. Palyulin and N. S. Zefirov, *Dokl. Akad. Nauk*, 2002, **384**, 202 [*Dokl. Chem.* (*Engl. Transl.*), 2002, **384**, 140].
5 L. P. Hammett, *Physical Organic Chemistry Reaction Rates, Equilibria, and Mechanisms*, McGraw-Hill, New York–London, 1940, p. 184.
6 R. W. Taft, *J. Am. Chem. Soc*., 1952, **74**, 3120.
7 R. W. Taft, *J. Am. Chem. Soc*., 1953, **75**, 4231.
8 *Tablitsy konstant skorosti i ravnovesiya geteroliticheskikh organicheskikh reaktsii* (*Tables of Rate and Equilibrium Constants of Heterolytic Organic Reactions*), ed. V. A. Palm, VINITI, Moscow, 1975, vol. 1, p. 7 (in Russian).
9 M. S. Silver, *J. Am. Chem. Soc*., 1961, **83**, 404.
10 O. A. Seoud and M. F. Martins, *J. Phys. Org. Chem*., 1995, **8**, 637.
11 H. Neuvonen and K. Neuvonen, *J. Chem. Soc., Perkin Trans. 2*, 1999, 1497.
12 M. Karelson, V. Lobanov and A. Katritzky, *Chem. Rev*., 1996, **96**, 1027.
13 V. A. Palm, *Osnovy kolichestvennoi teorii organicheskikh reaktsii* (*Quantitative Theory of Organic Reactions*), Khimia, Leningrad, 1977, p. 106 (in Russian).
14 I. I. Baskin, A. O. Ait, N. M. Halberstam, V. A. Palyulin and N. S. Zefirov, *SAR and QSAR in Env. Res*., 2002, **13**, 35.
15 V. R. Stimson, *J. Chem. Soc*., 1955, 2010.
16 C. K. Ingold, *Structure and Mechanism in Organic Chemistry*, Cornell University Press, Ithaca, 1969.

Received: 21st June 2002; Com. 02/1946