Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1076 (2005) 163–169
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Application of artificial neural networks for prediction of retention factors of triazine herbicides in reversed-phase liquid chromatography

Fabrizio Ruggieri [a,*], Angelo Antonio D'Archivio [a], Giuseppe Carlucci [b], Pietro Mazzeo [a]

[a] *Università degli Studi di L'Aquila, Dipartimento di Chimica, Ingegneria Chimica e Materiali, Via Vetoio, 67010 Coppito (L'Aquila), Italy*
[b] *Università degli Studi "G. D'Annunzio" Chieti, Dipartimento di Scienze del Farmaco, Facoltà di Farmacia, Via dei Vestini, Chieti 66100, Italy*

Received 12 November 2004; received in revised form 11 April 2005; accepted 20 April 2005
Available online 4 May 2005

Abstract

In this paper a quantitative structure-retention relationship (QSRR) method is used to model reversed-phase high-performance liquid chromatography (HPLC) behaviour of a series of triazine herbicides and their metabolites. Accurate description of the retention factors in terms of four descriptors related to the analytes and to the mobile phase is achieved by means of an artificial neural network (ANN). For comparison, a QSRR model is derived by multilinear regression (MLR). Validation of the two models shows a better ability in prediction of the ANN as compared with the MLR method. A solid-phase extraction (SPE) procedure allowing the simultaneous determination of the five triazinic compounds in groundwater analysis is also presented. The observed recoveries from water samples range between 85 and 100% for ng/ml concentration levels of all analytes.

Keywords: Quantitative structure-retention relationships; Artificial neural networks; HPLC optimisation; Triazine herbicides

1. Introduction

Triazines, owing to their extensive use as herbicides in modern agriculture, can be dispersed in surface and spring water at trace levels [1–3]. As a consequence of proven carcinogenic and endocrine disrupting action of these and other potentially hazardous compounds resulting from human activity, monitoring of groundwater has become an important aspect of environmental and health safeguard. Triazines are subjected to various abiotic and biotic degradation processes [4], and consequently, quantification of the metabolic products provides an additional analytical index to check water contamination.

High-performance liquid chromatography (HPLC) based on reversed stationary phase, coupled with a suitable preliminary sample preparation step able to concentrate the analytes and remove possible interferences, is one of the most powerful techniques for detection and quantification of triazine herbicides and their metabolites in water environment [2,5,6].

In the framework of the progress of chromatography, much effort has been concentrated in the last years to develop expert systems able to predict with good accuracy the retention behaviour of the analytes, providing an automatic means for the optimisation of chromatographic performance. In this perspective, quantitative structure-retention relationships (QSRR) methods [7,8] have been proposed, with the major aim of finding a mathematical model relating to the retention of a given analyte to physicochemical and structural parameters (descriptors). Besides practical application in optimisation strategies, QSRR studies can significantly contribute to get some insight into the molecular mechanism of separation [9–11].

Statistical treatment of QSRR multivariate data, consisting of a set of observed retention values and descriptors for a number of test molecules, is generally based on multilinear regression (MLR) [7,10–14]. In recent years, artificial neural networks (ANN) [15,16] have become a very popular and powerful chemometric tool to solve chemical problems,

* Corresponding author. Tel.: +390862433772; fax: +390862433753.
E-mail address: darchivi@univaq.it (F. Ruggieri).

0021-9673/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2005.04.038

including optimisation of chromatographic analysis [17–23]. As compared with multivariate regression, ANN does not require knowledge of a mathematical model before fitting of the data. Thus, it is particularly useful in the case of hidden nonlinearity inside the data variables.

In the present paper, ANN was used to develop a QSRR model for the prediction of the retention factor *k* of triazinic herbicides. In addition to the effect of the molecular structure of the analytes on the retention behaviour, as expressed by suitable descriptors, our attention was focused on the influence of pH and composition of the mobile phase, that are some of the operative parameters optimised in HPLC in order to achieve adequate separation and analysis time. The ability in prediction of the best ANN model was compared with that given by MLR.

A solid-phase extraction (SPE) procedure allowing simultaneous preconcentration of the five analytes in groundwater samples was also proposed. As alternative to common sorbents, i.e. porous silica particles surface-bonded with C_{18} or other hydrophobic groups, we used a macroporous copolymer formed by [poly(divinylbenzene-co-*N*-vinylpyrrolidone)], exhibiting both hydrophilic and lipophilic retention characteristics.

2. Method

2.1. Involved parameters

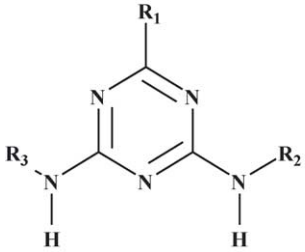
The triazinic herbicides used as test analytes in the present study are summarised in Fig. 1. The QSRR model was built by using descriptors related to the analyte and descriptors related to the eluent as inputs. The analyte descriptors were: the logarithm of the *n*-octanol–water partition coefficient ($\log K_{ow}$, taken from literature [2,24]), which is the standard hydrophobicity index widely used in QSRR research, and the total dipole moment (μ), related to the charge distribution within the molecule, obtained by ab-initio calculations. The descriptors related to the eluent were the eluent composition expressed by the percentage of methanol (%MeOH) and pH. In addition to $\log K_{ow}$ and μ, some other physico-chemical properties calculated from the molecular structure (molecular weight, refractive index, molar volume and polarisability), were considered in a preliminary step. However, information carried by these descriptors appeared to largely overlap with that provided by $\log K_{ow}$ and μ, as indicated by absolute values of coefficients of mutual correlation of these parameters with $\log K_{ow}$ (between 0.86 and 0.97) and μ (between 0.75 and 0.85). The influence of the above molecular properties on the retention behaviour was also evaluated by applying MLR based on a stepwise procedure, in which the number of descriptors to be selected and the order of entry are based on statistical criteria (see below). The regression model with the best statistics was that including only $\log K_{ow}$ and μ as analyte descriptors. These were finally chosen as the optimal parameters to describe the molecular properties.

2.2. Artificial neural networks analysis

Details on principles, functioning and applications of artificial neural networks can be found in references [15] and [16].

ANNs are computational models designed to simulate the way in which the human brain processes information. They consist of simple processing units (or neurons) linked with weighted modifiable interconnections. The neurons are generally organised into a layered structure, formed by one input layer, one output layer and at least one hidden layer. In a feed-forward network the signals are propagated from the input layer through the hidden layer(s) to the output layer. The feed-forward ANN architecture adopted in the present study consists of four inputs (the descriptors defined above) and one output (*k* values) connected to each other by one hidden layer with six neurons.

In addition to the network topology, an important component of most neural networks is a learning rule. A learning rule allows the network to adjust its connection weights in order to associate given inputs with corresponding outputs. The training of the network has been carried out by using a back-propagation algorithm, in which the network reads inputs and outputs from a proper data set (training set) and iteratively computes weights and biases in order to minimise the sum of squared differences between predicted and target values. The training is stopped when the error in prediction reaches a desired level of accuracy. However, if the network is left to train too long, it will overtrain and lose the ability



analyte	R_1	R_2	R_3
desisopropylatrazine	Cl	H	CH_2CH_3
desethylatrazine	Cl	H	$CH(CH_3)_2$
simazine	Cl	CH_2CH_3	CH_2CH_3
atrazine	Cl	CH_2CH_3	$CH(CH_3)_2$
prometon	OCH_3	$CH(CH_3)_2$	$CH(CH_3)_2$

Fig. 1. Structure of the triazine herbicides used in this work.