



Combination of artificial neural network technique and linear free energy relationship parameters in the prediction of gradient retention times in liquid chromatography

M.H. Fatemi^{a,*}, M.H. Abraham^b, C.F. Poole^c

^a Department of Chemistry, Mazandaran University, Babolsar, Iran

^b Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom

^c Department of Chemistry, Wayne State University, Detroit, MI 48202, USA

ARTICLE INFO

Article history:

Received 23 January 2008

Received in revised form 29 February 2008

Accepted 6 March 2008

Available online 13 March 2008

Keywords:

Artificial neural network

LFER parameters

Linear free energy relationship

Quantitative structure–retention relationship

Gradient retention

Multiple linear regression

ABSTRACT

In this work multiple linear regression (MLR) and artificial neural network (ANN) were used to predict the gradient retention times of diverse sets of organic compounds in four separate data sets. Descriptors which were used as inputs of these models are five linear free energy relationship (LFER) solute parameters including E , S , A , B and V . In the first step eight separate multiple linear regression and artificial neural network models were used to predict the gradient retention time for each gradient condition separately. Results obtained in this step reveal that there are significant relations between LFER parameters and gradient retention times of solutes in liquid chromatography. Then MLR and ANN were applied to develop more general models in which several different gradient elution conditions were used. The performances of these models are compared in terms of their standard errors and also correlation analysis. The results obtained reveal that although there are no significant differences between ANN and MLR in separate modeling of the gradient retention times, ANN has a significant superiority over MLR models in developing the general models for various gradient elution conditions. The results of sensitivity analysis on ANN models indicate that the order of importance for input terms in separate ANN models is $V_x > B > S > E > A$ and in the case of combined ANN model is $V_x > B > t_g > S > E > A$, which are in agreement with the order of percentage of significance terms that obtained from the MLR models.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The most common mode of separation in high-performance liquid chromatography (HPLC) is isocratic elution. In this method, the mobile phase composition is unchanged during the separation. Although isocratic elution is the simplest technique and is highly reproducible, yet in specific cases it may appear inappropriate. The disadvantages of the isocratic mode are; poor resolution of early-eluting bands, broadening of late-eluting bands to the point of difficult detection, tailing peaks and unnecessarily long separation times. This is often overcome by gradient elution, in other words, by changing the strength of the eluent over the course of the separation. Gradient elution offers several advantages: total analysis time can be significantly reduced, overall resolution of a mixture is increased, peak shape is improved (less tailing) and effective sensitivity is increased because there is little variation in peak shape. More importantly, it provides the maximum resolution

per unit of time for mixtures of a wide polarity. In order to find appropriate gradient conditions, trial-and-error optimizations are frequently used, although they are particularly slow and inefficient. These limitations may explain the effort that has been invested in computer-assisted strategies. Some software packages that include gradient optimization facilities, such as Dry-Lab (Rheodyne, USA) [1,2], Preopt-W (University of Santiago de Compostela, Spain) [3], Osiris (Datalys, France) [4] and ChromSword (Agilent, USA) [5] are currently available. However, although gradient optimization has reached a routine level, some topics still remain problematic. The main factor to be enhanced in order to obtain realistic optimizations is the ability to predict retention times as accurately as possible.

Unfortunately our understanding of retention in gradient elution is still rather limited. In fact, there are few studies on the relationships between retention and the physicochemical properties of solutes [6–8]. The existence and determination of such relationships are obviously very important because it offers the ability to estimate or predict the retention and selectivity of solutes during method development. In particular, if a quantitative relationship can be established between retention and the physicochemical properties of solutes, it will allow the prediction of retention of solutes with known properties in gradient elution, or

* Corresponding author at: Department of Chemistry, Mazandaran University, P.O. Box 453, Babolsar, Iran. Tel.: +98 112 5242931; fax: +98 112 4242002.

E-mail address: mhfatemi@umz.ac.ir (M.H. Fatemi).

enable the search for solutes with desired retention. For example, if it is required to insert an internal standard in a desired position for a gradient separation, the quantitative relationship between retention and the physicochemical properties of solutes can be established by using selected reference solutes. Once such a relationship has been established, the retention of solutes in a database can be calculated, and the solutes that match the desired retention can be selected as the internal standard. Although there are certain errors in the prediction, the benefits of the approach are obviously extremely significant.

Prediction and quantitative evaluation of properties is the main task of chemometrics [9,10]. To test the predictive potency of chemometric models, chromatography appears a unique and most suitable physicochemical system [11]. That is because in chromatographic systems all the measurement conditions can be kept constant for a large, statistically representative series of structurally diverse analytes, and in chromatographic retention parameters, which can easily be collected in a precise and reproducible manner, can be correlated to the chemical structure of individual solutes. Evaluation of retention in terms of chemical structure of analytes and of physicochemical properties of both the mobile and the stationary phase is known under the acronym QSRRs: quantitative structure–retention relationships. QSRRs are statistically derived relationships between chromatographic parameters and the quantities (descriptors) characterizing molecular structure of the analytes. QSRRs have found application to: (i) gain insight into the molecular basis of the separation mechanism for a given chromatographic system; (ii) identify the most informative structural descriptors of analytes; (iii) evaluate complex physicochemical properties of analytes, e.g., lipophilicity; (iv) evaluate properties of stationary phases; (v) predict relative differences in biological activities within a set of congeneric drugs or other xenobiotics; (vi) predict retention for a new analyte [12].

One of the main tasks of chemometrics is the quantitative evaluation and the prediction of properties [9,10]. As regards physicochemical properties, chromatographic systems seem ideal [11], because it is relatively easy to maintain constant measurement conditions, a large and diverse set of analytes can be used, and retention parameters can be collected in a precise and reproducible manner. Retention is a property of both the analyte and the mobile and stationary phase. For the case of a fixed mobile/stationary phase system, and evaluation of the various physicochemical properties that govern retention is known as quantitative structure–retention relationships, QSRRs. These are statistically derived relationships between chromatographic parameters and quantities (descriptors) characterizing the molecular structure of the analytes. QSRRs can be used to obtain information as to the factors that govern separation, especially structural features of the analytes, and hence can be used to predict retentions for new analytes [12]. Successful prediction of isocratic retention data by means of QSRR equations has been reported by several authors, Carr and co-workers [13,10], Forgacs and Cserhati [14], Valko et al. [15], Park et al. [16] and others [17,18]. For example Kaliszan et al. [19–21] demonstrated some QSRR models for prediction of liquid chromatographic retention times employing the following analyte descriptors: (i) total dipole moment, μ ; (ii) electron excess charge of the most negatively charged atom, δ_{Min} ; (iii) water-accessible molecular surface area, A_{was} . They assumed the following physical meaning of individual descriptors: μ accounts for the dipole–dipole and dipole-induced dipole attractive interactions of the analyte with the components of the competing mobile and stationary phase; δ_{Min} reflects the ability of analytes to participate in polar interactions of the charge-transfer and hydrogen-bonding type with the mobile and stationary phases; A_{was} describes the strength of dispersive interactions of analyte with the mobile and stationary phases. For example Kaliszan et

al. [19–21] demonstrated QSRR models for prediction of liquid chromatographic retention times employing using total molecular dipole moment, electron excess charge of the most negatively charged atom, and water-accessible molecular surface area as molecular descriptors. Also they developed another theoretical model using the above descriptors for prediction of gradient retention times [22,23]. The main drawback of their approach is that the specifications of their models were varied by changing the gradient elution conditions. Therefore it was necessary to derive separate models for each condition.

Other QSRR model for calculation of isocratic retention was later proposed by Galushko et al. [24]. In that model, the molecular bulkiness-dependent interactions of analytes with the components of the chromatographic system are accounted for by the partial molar volume descriptor, V . That descriptor appears to be a fairly reliable measure of structurally nonspecific inputs to retention. Unfortunately, less reliable and rather obscure is another structural descriptor proposed by Galushko et al. [24], ΔG , which is thought to reflect differences in “electrostatic” intermolecular interactions involving the analytes. Another structural descriptor proposed by Galushko et al. [24], is ΔG which is thought to reflect differences in “electrostatic” intermolecular interactions involving the analytes. Baczek and Kaliszan [22], however, have termed this descriptor ‘less reliable and rather obscure’.

In the 1980s much interest was focused on the so-called solvatochromic parameters of analytes as a means of retention prediction. In 1976 Kamlet and Taft [25,26] introduced the solvatochromic method to evaluate the relative polarity of solvents. The original theory was adapted to chromatography by Abraham and co-workers in the form of linear free energy relationships (LFER) [27,28]. The principle of a generalized LFER is based on a simple and conceptually explicit model of solute–solvent interactions, in which the solute solvation processes are identified and dissected into four types of solute–solvent interactions [29–32]. They include the cavity formation–dispersive interactions, dipolarity–polarizability interactions, and hydrogen bonding interactions. These interactions occur in both mobile and stationary phases in chromatography, and retention is the result of the difference in the solute’s interactions with both mobile and stationary phases. Because each solute possesses a unique set of physicochemical properties, it shows a different retention for a chromatographic system. The measurement of retention for solutes of known physicochemical properties, allows the correlation of retention with their LFER properties. The outcome of a LFER analysis is a set of regression coefficients that relate retention with the physicochemical properties of solutes. Accordingly, knowing the properties of different solutes, their retention can be predicted. One of the more widely accepted symbolic representation of the LFER model was proposed by Abraham in the form of following equation:

$$\log k = c + eE + sS + aA + bB + vV \quad (1)$$

in which k is the solute retention factor ($(t_R - t_0)/t_0$). The capital letters, E , S , A , B , and V are the solute descriptors independent of the mobile and/or stationary phase used. E is the solute excess molar refraction modeling the solute polarizability due to n - and/or π -electrons in excess of that of a comparable sized n -alkane, S is the solute descriptor for the dipolar character and also polarizability of the molecule, A and B , are respectively, the H-bond solute acidity (H-donor) and basicity (H-acceptor) descriptors and V is the McGowan’s characteristic molecular volume calculated using the solute structure [33,34]. The lowercase letters, c , e , s , a , b and v are the system parameters or constants reflecting the difference in solute interactions between the mobile and the stationary phase. The c constant is the intercept obtained in the regression calculation; it depends on the experimental system used (nature of the organic modifier, phase ratio) and not on the solute. The e

Download English Version:

<https://daneshyari.com/en/article/1205690>

Download Persian Version:

<https://daneshyari.com/article/1205690>

[Daneshyari.com](https://daneshyari.com)