ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
28 (2002) 581–590

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Application of the artificial neural network in quantitative structure–gradient elution retention relationship of phenylthiocarbamyl amino acids derivatives

S.Y. Tham [a], S. Agatonovic-Kustrin [b,*]

[a] *School of Pharmaceutical Science, USM, Penang 11800, Malaysia*
[b] *School of Pharmc. Molecular and Biomedical Sciences, University of South Australia, North Terrace, Adelaide SA 5000, Australia*

Received 10 August 2001; received in revised form 5 October 2001; accepted 7 October 2001

Abstract

Quantitative structure–retention relationship(QSRR) method was used to model reversed-phase high-performance liquid chromatography (RP-HPLC) separation of 18 selected amino acids. Retention data for phenylthiocarbamyl (PTC) amino acids derivatives were obtained using gradient elution on ODS column with mobile phase of varying acetonitrile, acetate buffer and containing 0.5 ml/l of triethylamine (TEA). Molecular structure of each amino acid was encoded with 36 calculated molecular descriptors. The correlation between the molecular descriptors and the retention time of the compounds in the calibration set was established using the genetic neural network method. A genetic algorithm (GA) was used to select important molecular descriptors and supervised artificial neural network (ANN) was used to correlate mobile phase composition and selected descriptors with the experimentally derived retention times. Retention time values were used as the network's output and calculated molecular descriptors and mobile phase composition as the inputs. The best model with five input descriptors was chosen, and the significance of the selected descriptors for amino acid separation was examined. Results confirmed the dominant role of the organic modifier in such chromatographic systems in addition to lipophilicity (log P) and molecular size and shape (topological indices) of investigated solutes. © 2002 Published by Elsevier Science B.V.

Keywords: Amino acids; Phenylthiocarbamyl derivatives; Artificial neural networks; Quantitative structure–retention relationship

1. Introduction

The basic concept of chemistry is that there is a relationship between bulk properties of compounds and their molecular structure. This provides a connection between the macroscopic and the microscopic properties of matter. The prediction of physicochemical and biological properties

* Corresponding author. Tel.: +61-8-8302-2934; fax: +61-8-8302-2389.
E-mail address: nena.kustrin@unisa.edu.au (S. Agatonovic-Kustrin).

0731-7085/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0731-7085(01)00690-2

of organic molecules is the main objectives of the quantitative structure–property/activity relationships (QSPRs/QSARs) [1]. QSPRs/QSARs are mathematical models that attempt to correlate the molecular structure of compounds and their biological, chemical, and physical properties. Among the most extensively studied properties are the chromatographic ones. It is considered that the same basic intermolecular interactions determine the behavior of chemical compounds in both biological and chromatographic environments [2]. Predicting chromatographic behavior from molecular structure of solutes resulted in the quantitative structure–retention relationships (QSRR) methodology. QSRR [3] are statistically derived relationships between the chromatographic parameters determined for a representative series of analytes in given separation systems and the molecular descriptors accounting for the structural differences among the investigated analytes. Such relationships may provide insight into the molecular mechanism of separation in a given chromatographic system, generate knowledge about the various interactions taking place between the solute and the stationary phase, evaluate physicochemical properties of analytes and identify the most informative structural descriptors. QSRR can be used to predict relative biological activities within a set of drugs, and to predict the retention of the solutes in a mixture and thus to optimize chromatographic separation [4]. Another area that can benefit from the QSRR approach is structure testing, e.g. the predicted retention of a hypothetical structure could be compared with the actual retention of the unknown component.

Three main types of QSRR have been employed. The oldest type correlates logarithms of retention factors ($\log k$) with the logarithms of *n*-octanol–water partition coefficients ($\log P$) [5]. The second type of QSRR is based on the solvatochromic comparison method and the so-called linear solvation energy relationships (LSERs) [6–8]. The third type of QSRR equation describes the chromatographic retention values in terms of quantum chemical indices and or other structural descriptors from calculation chemistry [9,10]. High-performance liquid chromatography (HPLC) retention data have been used as a pseudo-molecular descriptor to estimate the water solubility of aromatic hydrocarbons and organic non-electrolytes [11], hydrophobicity or octanol–water partition coefficient ($\log P$) [12,13] and for accurate estimations of pK_a [14,15]. Computer simulation methods were used to predict separation as a function of simultaneous change in pH and solvent strength for reversed-phase high-performance liquid chromatography (RP-HPLC) [16,17].

The major aim in this study was to develop QSRR, via artificial neural network (ANN) as data modeling tool, capable to predict and describe the retention capability of 18 amino acids in a given RP- HPLC system. The usefulness of ANNs for modeling retention times in HPLC optimization to correlate the chromatographic behavior of solutes (capacity factors) with mobile phase composition and pH has been previously investigated [18].

The structure and biological properties of a protein are determined by the primary sequence of amino acids. The principal source of variation between proteins is the length (and hence molecular weight) and the sequence of amino acids. In order to obtain the primary sequence of a protein it is usually necessary to determine the amino acid content of the protein after hydrolysis. The amino acids liberated by hydrolysis can be identified and quantified using chromatographic methods. Traditional methods of analysis involve ion-exchange chromatography and ninhydrin detection. Modern chromatographic approaches are based upon HPLC using hydrophobic 'reverse phase' columns. Several derivatisation methods are in common use: such as dansyl derivatives [19], dabsyl [20], *o*-pthalaldehyde (OPA) [21] derivatives, phenylisothiocyanate (PITC) derivatives [22–25] or 9-fluorenylmethyl chloroformate (Fmoc) [26] derivatives. Not all amino acids can be detected with the same sensitivity, different derivatisation chemistries giving differential sensitivity. For example, OPA–lys derivatives are unstable and OPA cannot detect proline or hydroxyproline unless they are previously oxidized with sodium hypochlorite. PITC derivatisation is not very good at detecting cysteines as cysteic acid and

other forms of cysteine resulting from hydrolysis give poor separation on HPLC. Despite this, PITC derivatisation is extensively used precolumn derivatization reagent for amino acids followed by RP-HPLC. It provides greater efficiency, sensitivity down to picomole level, ease of use and higher speed of analysis as well as specificity for primary and secondary amines.

HPLC separation of amino acid derivatives can generally be accomplished by using acetonitrile as the organic modifier and gradient elution on reversed phase octadecylsilane column. Acetonitrile is the preferred organic solvent because of its low viscosity, relatively high volatility, UV transparency and high degree of selectivity. The development of the elution profile can be often very tedious and time-consuming. Besides, the gradient profile can vary from column to column. By combining HPLC with ANN, a gradient profile for a complex mixture can be obtained, thus making the prediction easier to achieve.

1.1. Artificial neural networks

An ANN is a biologically inspired computational model that simulates the way in which human brain processes information. The network learns through experience with appropriate learning exemplars by detecting the patterns and relationships in data, not from pre-programming. An ANN is constituted from hundreds of single units, artificial neurons, organized in layers and connected with adjustable coefficients or weights. There is always one input, one output layer and there should be at least one hidden layer between them. Each PE has weighted inputs, transfer function and one output. The strength of connections, called weights act like a synaptic strength. The number of hidden units is optimized. The number of hidden units influences the number of connections. The more weights, the more powerful the network, and the more training data are needed to adequately train the network. The smallest model that is not significantly improved with additional hidden units or hidden layers will provide the system with the best generalization and the fastest training times. Careful feature selection and scaling of the inputs affects the complexity of the problem and selection of the best neural network model.

The behavior of a neural network is determined by the transfer functions of its neurones, by the learning rule, and by the architecture. We have used a supervised network with back-propagation learning rule and multilayer perceptron (MLP) architecture and genetic algorithm (GA) input selection. MLPs are general-purpose, flexible, nonlinear models. In this model, the inputs are fully connected to the hidden layer and hidden layer neurons are fully connected to the outputs. MLP models compute the output as a sum of non-linear transformations of linear combinations of the inputs. Detailed descriptions of this type of the ANN model have been published [27–29].

2. Experimental

2.1. Materials

The amino acid standards and anhydrous sodium acetate were purchased from Sigma Chemical Co. (St Louis, MO, USA) and Ajax Chemicals (Auburn, NSW, Australia) respectively. Methanol, AR and ethanol were purchased from JT Baker (Phillipsburg, NJ, USA) whilst triethylamine, (TEA, synthesis grade) and PITC were obtained from E. Merck (Darmstadt, Germany). TEA and PITC were both redistilled before use. Redistilled PITC was stored at -20 °C until use. Unless stated otherwise, distilled water was used throughout the experiment.

2.2. Instrumentation

The HPLC solvent delivery system consisted of a Gilson Model 305 main pump (Villiers le Bel, France) and a Gilson Model 302 secondary pump, a Gilson Model 802C manometric module, a Gilson Model 811B dynamic mixer. The chromatographic system consisted of a Rheodyne 7125 injector (Cotati, CA, USA), a homepacked reversed phase Whatman Partisil C18 column (25 cm $\times$ 4.6 mm ID, 10 μm) and guard column thermostatted at 30 °C. Detection was performed on a Shimadzu Spd-6A ultraviolet detector and

recorded with a Hitachi D2500 integrator-plotter (Tokyo, Japan).

2.3. Packing of Whatman Partisil C18 column

Approximately 3.8 g of the Whatman Partisil C18 (10 μm) packing material was suspended in 35 ml of methanol and ultrasonicated for 2 min. The slurry, shaken for 10 s to ensure homogeneity, was poured into the reservoir of the packer (Shandon Column Packer, UK). The pressure of the packer was set at 580 bar. After 10 ml of the solvent had flowed through the column, it was turned 180° and a further 40 ml of the solvent was allowed to flow through before the column was finally packed.

2.4. Preparation of standard phenylthiocarbamyl (PTC) amino acid derivatives

Stock solutions, each containing 250 μmol/ml of amino acid, were prepared in distilled water. About 10 μl aliquots ($\equiv 2.5$ μmol) of the stock solution was added to freshly prepared 20 μl of ethanol–water–TEA (2:2:1) mixture, shaken thoroughly and the mixture was dried under vacuum by using an Edward Model E2M8 Vacuum pump (Sussex, England). About 20 μl of the freshly prepared derivatisaton mixture of ethanol–TEA–water–PITC (7:1:1:1) was added to each dried sample, shaken thoroughly and allowed to stand at room temperature for 20 min. Excess reagent and by-products were removed under vacuum. The PTC–amino acids were then stored at -20 °C until use.

2.5. Isocratic separation

Dried PTC–amino acid derivatives were restored in 1 ml of filtered methanol, AR. The solutions were ten times diluted with the mobile phase and centrifuged. About 20 ul aliquots of the PTC amino acid derivatives were injected into the HPLC system. The mobile phase consisted of acetonitrile: acetate buffer (pH 6.35, 0.14 M containing 0.5 ml/l TEA) in varying ratios. The flow rate was fixed at 1 ml/min. The set of 18 amino acids were divided into two clusters. The first cluster contained aspartic acid (asp), glutamic acid (glu), serine (ser), glycine (gly), threonine (thr), alanine (ala), histidine (his), proline (pro) and arginine (arg), whilst the second cluster consisted of tyrosine (tyr), valine (val), methionine (met), cystine (cyt), isoleucine (ile), leucine (leu), phenylalanine (phe), tryptophan (trp) and lysine (lys). The mobile phase composition varied from 5:95 to 15:85 acetonitrile:buffer for the first cluster and from 15:85 to 30:70 for the second cluster.

2.6. Gradient separation

After the elution order of the PTC amino acids was determined using the isocratic separation, a gradient profile was developed to optimize a complete separation of the PTC amino acid derivatives. The solvent system consisted of two eluents, solvent A (acetate buffer, pH 6.35, 0.14 M containing 0.5 ml/l TEA) and solvent B (60% acetonitrile in water). The total flow rate used was at 1 ml/min. The gradient profile used was first linear gradient from 0 to 5.5% B in 5 min and then eluent was maintained at 5.5% B from 5 to 17 min and finally a second gradient from 5.5 to 100% B from 17 to 20 min. The washing step at 100% B was programmed to run for 30 min to wash away any residual contaminants that could be present in the sample. The system was returned to 100% A at 65 min for the next injection.

2.7. ANN structure optimization

A standard feed-forward network, with back-propagation rule and MLP model architecture [30] was chosen. An initial neural network consisting of 56 input–output sets and two hidden layers was constructed. Although it is possible to approximate any function with just one layer of hidden units, a huge number of hidden units may be required. An additional layer was used to reduce the number of hidden units and, consequently, the number of required weights. A total of 36 molecular descriptors including constitutional, topological, chemical, geometrical and quantum chemical descriptors were calculated for each amino acid. The 56 experimentally derived retention times were used as the output of the

ANN and calculated molecular descriptors and mobile phase composition were used as the inputs. The number of inputs, number of hidden layers and hidden neurons was optimized. Connections or units were eliminated during training based on sensitivity reports, highest coefficient of multiple determination and minimal error for the external prediction data set. Using a GA and sensitivity analysis the number of inputs was reduced from 37 to 5. Input selection has reduced the size and complexity of the network and focused the training on the most important data. This also reduced the training time and improved the network performance.

Output sensitivity was used to evaluate the influence of selected descriptors on chromatographic separation. Sensitivity reports show the sensitivity of the output variables, as a percentage, to the changes in the corresponding input variables. If the direction of the change in the output variable is always the same as the change in investigated descriptor then the average sensitivity is positive. The set of percentages also reveals the effect that a change in a particular input has on output.

Before each training run, data sets were split randomly into three separate groups, the training (40 data sets), testing (eight data sets) and validation set (eight data sets), and both weights and biases were initialized with random values. The results of the five runs were averaged. During training, the performance of the ANN was evaluated with testing data. The training set was used to train the network and the testing set was used to monitor overtraining the network. Training was stopped when the training root mean squared error (RMS) fails to improve over a given number of training cycles and when the testing RMS error started to increase. Validation set was used to evaluate the trained model.

3. Results and discussion

The elution order for the first cluster of PTC-derivatives of amino acids was established by using isocratic RP-HPLC system and varying the content of the organic modifier, acetonitrile. As the acetonitrile content gradually increased from 2:98 to 5:95 (acetonitrile:water), retention times of the PTC-amino acid derivatives were determined. Elution order was asp, glu, ser, gly, thr, ala, his, pro and arg. The elution order for the second cluster of PTC-amino acids was found to be tyr,

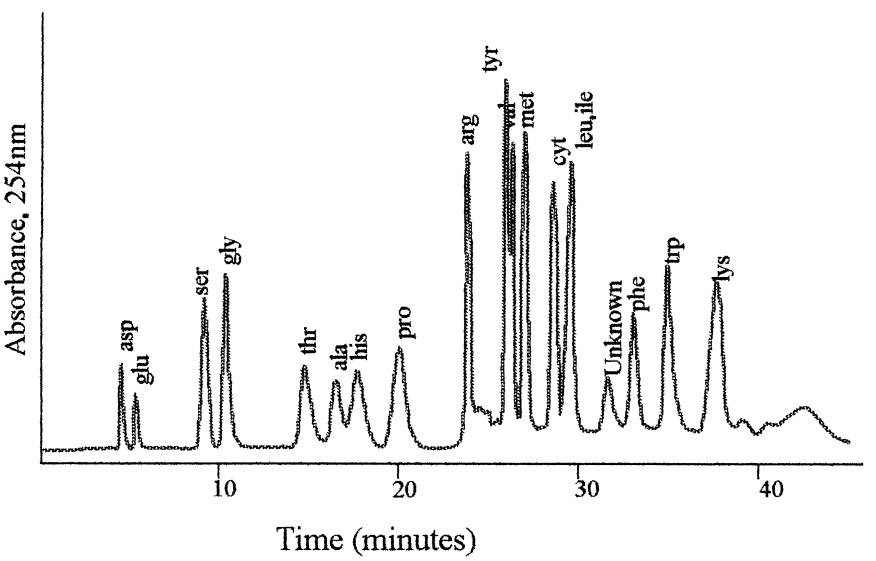

Fig. 1. HPLC chromatogram showing the separation of PTC derivatives of amino acids on a Whatman Partisil ODS column using gradient elution.

Table 1
Averaged retention time sensitivity

Descriptor	Sensitivity (%)
Mobile phase composition	0.42
χ_1	0.19
Log P	0.22
χ_{v2}	0.09
χ_0	0.07

val, met, cyt, ile (leu), phe, trp and lys. The gradient profile for the separation of the 18 PTC amino acid derivatives was then developed (Fig. 1). Based on the column and mobile phase conditions used, no satisfactory separation was achieved for PTC–leu and PTC–ile.

The first step in developing QSRR was to calculate numerical descriptors. A total of 36 calculated structural features including constitutional, topological, geometrical, quantum chemical and physicochemical descriptors were generated for each of the 18 amino acids. The next step was to select descriptors important in chromatographic separations. A subset of descriptors that best encodes the retention times was selected using a GA [31–33] and correlated to the experimentally measured retention times using ANNs. The ANN model with five input descriptors was found to have the best predictive performance. The model had one hidden layer with two neurons, thus producing a 5–2–1 architecture. Other architectures were examined, but they produced poorer quality neural network models.

Selection of the important molecular descriptors and examination of the variable contribution to the model through output sensitivity is an important aspect of QSRR study, not only for ranking the relative importance of each variable and calculating its statistical significance, but also as a means of refining the model by variable selection. The sign and size of the input sensitivity report from the ANN model denoted in what way and how much each variable contributes to the final predicted value.

The QSPR that was developed indicates that the mobile phase composition, lipophilicity (log P) and molecular shape and size (topological shape indices, molecular connectivity indices) are the most important in chromatographic separation (Table 1).

It has been shown [34–36] that, in binary aqueous/organic mobile phases on a RPLC column, the retention of a single solute can be modeled as a function of the mobile phase composition over a limited range of mobile phase compositions. As expected the mobile phase composition (percent of organic modifier in mobile phase) was the main factor effecting the separation of amino acids with the output sensitivity of over 40%.

Partitioning is one of its most important and fundamental properties for a chemical compound. Octanol/water partition coefficient (log P) and water solubility are critical in understanding the tendency of drugs to cross biological membranes (absorption from the gastrointestinal tract, blood brain barrier). Therefore, log P is frequently used in QSAR studies as a measure of the lipophilic character of the molecules and their biological activity [37]. Retention in the reversed-phase liquid chromatography system is characterized by bulkiness and the polar properties of the solute and may be related to the solubility of solutes [38]. Thus, a measure of lipophilicity or hydrophilicity (log P, the octanol/water partition coefficient) should correlate fairly well with retention time in such a system. Bulk properties include molecular weight, volume, surface area, density, and molecular length, width and depth.

It is known that the retention value of a solute is related to the partition and absorption process [39]. The great advantage of reversed-phase high-performance liquid chromatography RP-HPLC is that it can yield a relatively precise and reproducible data so that retention values (retention time, capacity factor) can be used as hydrophobic descriptors. On the other hand hydrophobicity of a substance can be determined with classical reversed-phase liquid chromatography (RP-LC) systems. In recent years there has been an increasing tendency to apply RP-HPLC retention factors (capacity, retention time) in QSPRs/QSARs [40] as a measure of lipophilicity. As expected, developed QSRR suggests that the increase in lipophilicity (log P) increases retention time of amino acids.

Over the last 10 years variety of topological and shape descriptors emerged as alternative descriptors in quantitative structure–activity studies [41,42]. The most popular topological indices are molecular connectivity indices [43,44]. A large number of studies have demonstrated that many physicochemical and biological properties correlate with the connectivity index [45]. The main advantage of the graph theoretical approach to the prediction of properties is that it permits the interpretation of results in terms of structurally related concepts. In spite of that, the most important criticism of the so-called topological indices is concerned with their physical meaning [46].

Molecular connectivity is a method of molecular structure quantification based only on bonding and branching patterns rather than physical or chemical characteristics. Weighted counts of substructure fragments are incorporated into numerical indices and structural features (size, branching, unsaturation, heteroatom content and cyclicity) are encoded. These indices are related to the number of atoms and how they are connected in a molecule. Only the carbon or heavy atoms are taken into consideration and the connectivity indices are derived from the hydrogen-suppressed graph of the molecule. Each atom is represented by a vertex in the graph, while the bonds becomes edges. Valence connectivity index [47] uses the same invariant but modifies vertex degrees to account for heteroatoms by using the number of valence electrons in the corresponding atom.

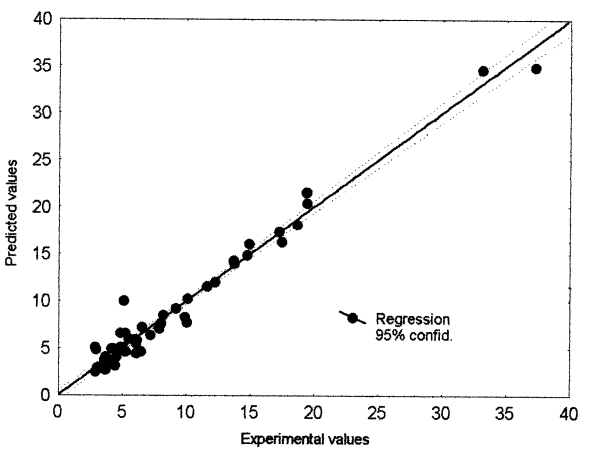

Fig. 2. Linear relationship between the experimentally derived and retention time values predicted by the QSRR model.

Connectivity indices are descriptor of molecular structure, a descriptor of size and shape based on a count of groupings of skeletal atoms, weighted by degree of skeletal branching. Zeroth-order (atomic) χ_0 represents the sum over all vertexes and conveys information about the number of atoms in a molecule. It is shown that the increase in χ_0 decreases retention time due to the increase in molecular size. First order, χ_1 (molecular, one bond paths) index is the sum of all bonds and second-order (path) molecular connectivity χ_2 is the sum over subgraphs. Increase in χ_1, the first order (bond) connectivity index increase chromatographic retention time. Molecular connectivity index of the first order, χ_1, encodes single bond properties. It is a weighted count of bonds, related to the types and position of branching in the molecule. χ_2, the second order (path, two bond fragments) connectivity index is derived from fragments of two-bond length. It also provides information about types and position of branching and may be indication of the amount of structural flexibility and bulkiness. In the analysis of hydrocarbons the molecular connectivity indices correlate well with the retention indices in the stationary phase of low polarity [48]. The chromatographic process of separation results from the forces that operate between solute molecules and the molecule of the stationary phase. The interaction of the carbon atoms with the stationary phase is determined by its electrical properties and by the steric hindrance of other carbon atoms attached to it. The number of carbon atoms in linear alkanes is correlated to their chromatographic retention and increase in χ_1. However, the branched alkanes do not present this linear relationship. The steric effects of their neighboring groups decrease retention of the tertiary and quaternary carbon atoms [49]. Increase in branching (χ_{V2}) increases surface area and molecular volume [50] and decrease retention time. On the other hand low values of χ_{V2} are found for more elongated molecules or these with only one branching atom. Increase in the length of the carbon chain, non-polar portion of the molecule, results in the increase in lipid-solubility (log P).

Table 2
Retention times for 16 amino acids by using isocratic RP-HPLC system and varying the content of the organic modifier predicted using ANN and measured

Amino acid	Mobile phase composition	Retention time (min)	
		Experimental	Predicted
ala	70:30	2.57	2.90
ala	80:20	3.73	3.34
ala	85:15	4.83	4.65
ala	90:10	7.59	7.07
ala	95:5	13.64	14.29
ala	85:15	4.88	4.65
ala	80:20	3.72	3.34
arg	85:15	6.08	4.62
arg	90:10	9.84	7.66
arg	95:5	19.56	20.32
asp	85:15	2.44	4.85
glu	85:15	2.43	5.04
ser	85:15	3.75	4.93
ser	90:10	5.13	6.01
ser	85:15	3.82	4.93
gly	70:30	2.42	2.49
gly	80:20	3.19	2.70
gly	70:30	2.42	2.493
gly	80:20	3.18	2.707
gly	85:15	4.11	4.09
gly	90:10	5.64	5.858
thr	85:15	4.60	5.06
thr	90:10	6.90	6.30
thr	95:5	12.18	11.95
hys	85:15	4.71	9.98
hys	90:10	7.63	7.331
hys	95:5	13.71	14.00
pro	85:15	4.48	6.56
pro	90:10	7.75	7.57
pro	95:5	14.78	14.9
pro	85:15	4.83	6.56
arg	95:5	19.60	21.49
arg	85:15	5.70	5.43
arg	90:10	9.69	8.25
arg	95:5	19.55	21.49
arg	98:2	38.19	34.87
tyr	85:5	8.96	9.17
val	70:30	3.11	3.68
val	80:20	5.75	5.75
val	85:15	9.917	10.19
val	70:30	3.114	3.686
val	80:20	5.75	5.75
met	85:15	11.52	11.53
cys	85:15	17.33	17.38
cys	70:30	3.23	4.04
cys	80:20	6.24	7.19
leu	85:15	18.80	18.08
ile	70:30	4.00	3.09
ile	80:20	7.92	8.48
phe	85:15	33.88	34.55
trp	80:20	14.89	16.05
lys	70:30	5.68	4.43
lys	80:20	17.59	16.26

As expected, the model shows a strong correlation (up to $R = 0.97$) between predicted and experimentally measured flux values (Fig. 2). Since the slope ($b = 0.965$; $t_b = 1.67$, $t_{0.05} = 2.67$) was not significantly different from unity, the method did not show proportional error. In other words, the sensitivity was the same for measured and predicted values (Table 2). A proportional error would lead to a change in b so that the difference between b and unity gives an estimate of the proportional error. The intercept was not significantly different from zero ($a = 0.28$; $t_a = 1.13$) indicating the absence of systematic error and method bias.

4. Conclusion

A five-descriptor nonlinear computational neural network model has been developed for the estimation of chromatographic retention time values for a data set of 18 amino acids. The training set RMS error was 1.773 and the testing set RMS error was 0.8377. Based on the RMS errors of the training and testing sets and high correlation of predicted versus experimentally derived M/P values ($R > 0.97$, it is clear that a link exists between structure and chromatographic separation. The strength of this link was measured by the quality of the model prediction. With an RMS error of 1.898, the validation data set ensures the quality of the model.

References

[1] J.W. McFarland, C.M. Berger, S.A. Froshauer, S.F. Hayashi, S.J. Hecker, B.H. Jaynes, M.R. Jefson, B.J. Kamicker, C.A. Lipinski, K.M. Lundy, C.P. Reese, C.B. Vu, Quantitative structure-activity relationships among macrolide antibacterial agents: in vitro and in vivo potency against *Pasteurella multocida*, J. Med. Chem. 40 (1997) 1340–1346.

[2] R. Kaliszan, Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987.

[3] R. Kaliszan, M.A. van Straten, M. Markuszewski, C.A. Cramers, H.A. Claessens, Molecular mechanism of retention in reversed-phase high-performance liquid chromatography and classification of modern stationary phases by using quantitative structure–retention relationships, J. Chromatogr. A 855 (1999) 455–486.

[4] E.R. Collantes, W. Tong, W.J. Welsh, W.L. Zielinski, Use of moment of inertia in comparative molecular field analysis to model chromatographic retention of nonpolar solutes, Anal. Chem. 68 (1996) 2038–2043.
[5] M. Hanna, V. de Biasi, Vern, B. Bond, Estimation of the partitioning characteristics of drugs: a comparison of a large and diverse drug series utilizing chromatographic and electrophoretic methodology, Anal. Chem. 70 (1998) 2092.
[6] L.C. Tan, P.W. Carr, M.H. Abraham, Study of retention in reversed-phase liquid chromatography using linear solvation energy relationships. I: the stationary phase, J. Chromatogr. A 752 (1996) 1–18.
[7] K. Valko, N. Plass, C. Bevan, D. Reynolds, M.H. Abraham, Relationships between the chromatographic hydrophobicity indices and solute descriptors obtained using reversed-phase diol, nitrile, cyclodextrin and immobilized artificial membrane-bonded high-performance liquid chromatography columns, J. Chromatogr. A 797 (1998) 41–55.
[8] Á. Sándi, M. Nagy, L. Szepesy, Characterization of reversed-phase columns using the linear free energy relationship; III. Effect of the organic modifier and the mobile phase composition, J. Chromatogr. A 893 (2000) 215–234.
[9] J. Dai, L. Jin, S. Yao, L. Wang, Prediction of partition coefficient and toxicity for benzaldehyde compounds by their capacity factors and various molecular descriptors, Chemosphere 42 (2001) 890–899.
[10] T. Suzuki, S. Timofei, B.E. Iuoras, G. Uray, P. Verdino, W.M.F. Fabian, Quantitative structure–enantioselective retention relationships for chromatographic separation of arylalkylcarbinols on Pirkle type chiral stationary phases, J. Chromatogr. A 922 (2001) 13–23.
[11] P. de Voogt, J.W. Wegener, U.A. Brinkman, H. Gover, Retention of neutral and basic heteroaromatic hydrocarbons in RPLC systems and its use in predictive studies. I. Concentration of the organic modifier, Sci. Total Environ. 109-110 (1991) 69–87.
[12] L. Escuder-Gilabert, J.M. Sanchis-Mallols, S. Sagrado, M.J. Medina-Hernández, R.M. Villanueva-Camañas, Chromatographic quantitation of the hydrophobicity of ionic compounds by the use of micellar mobile phases, J. Chromatogr. A 823 (1998) 549–559.
[13] L.P. Sarna, P.E. Hodge, G.R.B. Webster, Octanol–water partition coefficients of chlorinated dioxins and dibenzofurans by reversed-phase HPLC using several C_{18} columns, Chemosphere 13 (1984) 975–983.
[14] H.Y. Ando, T. Heimbach, pK_a determinations by using a HPLC equipped with DAD as a flow injection apparatus, J. Pharm. Biom. Anal. 16 (1997) 31–37.
[15] D.S. Ashton, C.R. Beddell, G.S. Cockerill, K. Gohil K, C. Gowrie, J.E. Robinson, M.J. Slater, K. Valko, Binding measurements of indolocarbazole derivatives to immobilised human serum albumin by high-performance liquid chromatography, J. Chromatogr. B Biomed. Appl. 677 (1996) 194–198.
[16] J.A. Lewis, J.W. Dolan, L.R. Snyder, I. Molnar, Computer simulation for the prediction of separation as a function of pH for reversed-phase high-performance liquid chromatography. II. Resolution as a function of simultaneous change in pH and solvent strength, J. Chromatogr. 592 (1992) 197–208.
[17] J.A. Lewis, D.C. Lommen, W.D. Raddatz, J.W. Dolan, L.R. Snyder, I. Molnar, Computer simulation for the prediction of separation as a function of pH for reversed-phase high-performance liquid chromatography. I. Accuracy of a theory-based model, J. Chromatogr. 592 (1992) 183–195.
[18] S. Agatonovic-Kustrin, M. Zecevic, L.j. Zivanovic, Use of ANN modelling in structure–retention relationships of diuretics in RP-HPLC, J. Pharm. Biom. Anal. 21 (1999) 95–103.
[19] C. De Jong, G.H. Hughes, E.V. Wieringen, K.J. Wilson, Amino acid analyses by high-performance liquid chromatography. An evaluation of the usefulness of precolumn Dns derivatization, J. Chromatogr. 241 (1982) 345–359.
[20] R. Kneck, J.Y. Chang, Liquid chromatographic determination of amino acids after gas-phase hydrolysis and derivatization with (dimethylamino) azobenzenesulfonyl chloride, Anal. Chem. 58 (1986) 2375–2379.
[21] P. Lindroth, K. Mopper, HPLC determination of subpicomole amounts of amino acids by precolumn fluorescence derivatization with *o*-phthaldialdehyde, Anal. Chem. 51 (1979) 1667–1674.
[22] S. Tsuasawa, J. Kondo, F. Sakiyama, Isocratic separation of PTH–amino acids at picomole level by reverse-phase HPLC in the presence of sodium dodecylsulfate, J. Biochem. 97 (1985) 701–704.
[23] H. Scholze, Determination of phenylthiocarbamyl amino acids by reversed-phase high-performance liquid chromatography, J. Chromatogr. 350 (1985) 453–460.
[24] K. Hayakawa, J. Oizumi, Isocratic separation of phenylthiohydantoin–amino acids by reversed-phase high-performance liquid chromatograohy, J. Chromatogr. 487 (1989) 161–166.
[25] M. Abrahamsson, K. Gröningsson, Separation of PTH–amino acids by isocratic high-performance liquid chromatography, J. Chromatogr. 154 (1978) 313–317.
[26] R.A. Bank, E.J. Jansen, B. Beekman, J.M. te Koppele, Amino acid analysis by reverse-phase high-performance liquid chromatography: improved derivatization and detection conditions with 9-fluorenylmethyl chloroformate, Anal. Biochem. 240 (1996) 167–176.
[27] M. Egmont-Petersen, J.L. Talmo, A. Hasman, Assessing the importance of features for multi-layer perceptrons, Neural Networks 11 (1998) 623–635.
[28] G. Bologna, C. Pellegrini, Three medical examples in neural network rule extraction, Phys. Med. 13 (1997) 183–187.
[29] T.F. Rathbun, S.K. Rogers, M.P. DeSimio, MLP iterative construction algorithm, Neurocomputing 17 (1997) 195–216.

[30] H. White, Artificial Neural networks: Approximation and Learning Theory, Blackwell, Oxford, UK, 1992.
[31] P. Willet, Genetic algorithms in molecular recognition and design, Trends Biotechnol. 13 (1995) 516–521.
[32] S.S. So, M. Karplus, Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural networks, J. Med. Chem. 39 (1996) 1521–1530.
[33] S.S. So, M. Karplus, Three-dimensional quantitative structure–activity relationship from molecular similarity matrices and genetic neural networks, J. Med. Chem. 40 (1997) 4347–4359.
[34] M. Shibukawa, Theoretical interpretation of the retention of system peaks in partition chromatography with a mobile phase containing electrolytes, J. Chromatogr. A 696 (1995) 165–172.
[35] L.R. Synder, C.S. Horvath, High-performance Liquid Chromatography, vol. 1, Academic Press, New York, 1980, p. 207.
[36] L. Zhen, Z. Hanfa, Y. Mingliang, N. Jianyi, Z. Yukui, Effects of organic modifiers on retention mechanism and selectivity in micellar electrokinetic capillary chromatography studied by linear solvation energy relationships, J. Chromatogr. A 863 (1999) 69–79.
[37] R.N. Smith, C. Hanch, M.M. Ames, Selection of a reference partitioning system for drug design work, J. Pharm. Sci. 64 (1975) 599.
[38] R. Gami-Yilinkou, R. Kaliszan, High performance liquid chromatographic (HPLC) measures of hydrophobicity as determined by means of new HPLC columns, Pol. J. Pharmacol. Pharm. 44 (1992) 515–525.
[39] C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
[40] J. Dai, L. Jin, S. Yao, W. Liansheng, Prediction of partition coefficient and toxicity for benzaldehyde compounds by their capacity factors and various molecular descriptors, Chemosphere 42 (2001) 899–907.
[41] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floc'h, R. Buelow, Computer-assisted rational design of immunosuppressive compounds, Nat. Biotechnol. 16 (1998) 748–752.
[42] D. Gorse, A. Rees, M. Kaczorek, R. Lahana R, Molecular diversity and its analysis, Drug Discovery Today 4 (1999) 257–264.
[43] M. Randic, S.C. Basak, Optimal molecular descriptors based on weighted path numbers, J. Chem. Inf. Comput. Sci. 39 (1998) 261–266.
[44] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[45] H. Hong, D. Zhou, S.K. Han, L.S. Wang, Retention of aromatic sulfur-containing compounds on RP-HPLC: correlation with partition coefficients and molecular connectivity indices, J. Liq. Chromatogr. Rel. Technol. 19 (1996) 1617–1629.
[46] Z. Mihalic, S. Nikolic, N. Trinajstic, Comparative study of molecular descriptors derived from the distance matrix, J. Chem. Inf. Comput. Sci. 32 (1992) 28–37.
[47] E. Estrada, Generalization of topological indices, Chem. Phy. Lett. 336 (2001) 248–252.
[48] R. Kalisman, Quantitative Structure–Chromatographic Retention Relationships, Wiley, New York, 1987.
[49] E.F.V. Heinzen, M.F. Soares, R.A. Yunes, Semi-empirical topological method for the prediction of the chromatographic retention of *cis*- and *trans*-alkene isomers and alkanes, J. Chromatogr. A 849 (1999) 495–506.
[50] A. Verloop, W. Hoogenstraaten, J. Tysker, in: E.J. Ariens (Ed.), Drug Design, vol. 7, Academic Press, New York, 1976.