

# Optimisation of high performance liquid chromatography separation of neuroprotective peptides Fractional experimental designs combined with artificial neural networks

Klára Novotná, Jan Havliš\*, Josef Havel

*Masaryk University, Faculty of Science, Department of Analytical Chemistry, Kotlářská 2, 611 37 Brno, Czech Republic*

Available online 1 July 2005

## Abstract

The study of experimental design conjunction with artificial neural networks for optimisation of isocratic ion-pair reverse phase HPLC separation of neuroprotective peptides is reported. Different types of experimental designs (full-factorial, fractional) were studied as suitable input and output data sources for ANN training and examined on mixtures of humanin derivatives. The independent input variables were: composition of mobile phase, including its pH, and column temperature. In case of a simple mixture of two peptides, the retention time of the most retentive component and resolution were used as the dependent variables (outputs). In case of a complex mixture with unknown number of components, number of peaks, sum of resolutions and retention time of ultimate peak were considered as output variables. Fractional factorial experimental design has been proved to produce sufficient input data for ANN approximation and thus further allowed decreasing the number of experiments necessary for optimisation. After the optimal separation conditions were found, fractions with peptides were collected and their analysis using off-line matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS) was performed.

© 2005 Elsevier B.V. All rights reserved.

**Keywords:** Optimisation of separation; Artificial neural networks; ANN; Experimental design; Fractional experimental design; Neuroprotective peptides; HPLC; Liquid chromatography

## 1. Introduction

In the HPLC method development, the most important aspect is to achieve adequate separation of all components in reasonable time. Consequently, optimisation of different chromatographic factors, like the mobile phase composition, pH of aqueous component of the mobile phase, as well as column temperature, is critical for sufficient resolution. With the respect to the high number of factors influencing the separation, it could be difficult and time-consuming to reach optimal separation conditions, particularly using the single variable optimisation approach, when one variable is changed in time while the others are kept constant. Recently, general optimisation approach based on the combination of experimental

design (ED) and artificial neural networks (ANN) [1,2] has been proposed. The ED-ANN is based on suitable design of experiments and then the optimal separation conditions are predicted with an artificial intelligence method using input and output data of these experiments. The usefulness of ANN in the field of HPLC separation has already been reported. Petritis et al. [3] applied the ANN model to predict elution times of peptides on the basis of their amino acid sequence. The usefulness of ANN modelling in quantitative structure-retention relationships was also demonstrated by Loukas [16], Tham and Agatonovic-Kustrin [17], and Agatonovic-Kustrin et al. [18]. Another approach using the ANN is response surface modelling. In the response surface modelling, effects of different chromatographic factors are studied and optimal separation conditions are looked for [4,5,19,20].

Humanin and its derivatives were found to protect the neural cells from pathological proteins, like the amyloid

\* Corresponding author. Tel.: +420 549 494 148; fax: +420 549 492 494.  
E-mail address: [jdqh@chemi.muni.cz](mailto:jdqh@chemi.muni.cz) (J. Havliš).

precursor protein, presenilin-1 or presenilin-2 mutants, which cause the Alzheimer's disease [6]. Analytical methods for determination of these compounds have already been examined [7,8]. For the peptide separation, ion-pair (IP) reverse phase high-performance liquid chromatography (RPLC) using trifluoroacetic acid (TFA) is frequently applied [9,10] as an alternative to ion-exchange chromatography. In IP-RPLC, TFA is used as the anionic ion-pairing reagent that minimises the charge on carboxylates and enhances the hydrophobicity of the peptide.

To achieve separation of peptides with similar amino acid sequence and closed physico-chemical properties, the influence of the mobile phase composition (TFA concentration, acetonitrile content and pH of aqueous component), as well as the column temperature, was carefully investigated. Using increased temperatures for the RPLC separation of peptide mixtures has been proposed [11,12], primarily to increase the column efficiency and to make the analysis time shorter.

The aim of this work was to study and develop the effective general optimisation approach for separation of complex peptide mixtures without prior knowledge about their physico-chemical properties. Then, this approach was applied to optimise IP-RPLC separation of neuroprotective peptides such as humanin derivatives.

## 2. Theory

### 2.1. Experimental design

The experimental design means planned series of experiments with changing variables describing the experiment in the most efficient way in order to find optimal variable settings for further evaluation [13]. The primary goal is to extract the maximum amount of unbiased information regarding the factors that affect the separation process. In other words, the aim of ED is to get the best description of the response surface. The response surface is the dependence of the output response as a function of one or more factors. In case of high number of factors, factorial experimental designs (FED) are recommended [13]. Using factorial ED, not only basic effects but also interactive effects of given factors with maximum precision can be estimated. One of the factorial ED, in which further reduction of the number of necessary experiments is possible, is called fractional design (FrED).

In our case, the data from three-factor central composite designs (CCD) as well as fractional two-level three-factor designs (cube, CD, and star, SD) were taken as the relevant input data for the ANN training (Fig. 1) [13]. The experiments were performed according to the design obtained using both, FED and FrED, and then the results from these experiments (outputs), as well as separation conditions (inputs) were used to train ANN with the aim to find the approximation function between these variables and thus optimum separation conditions.

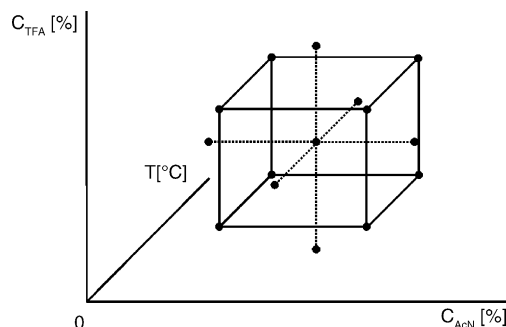


Fig. 1. The scheme of central composite design. Solid lines—cube design, dotted lines—star design.

### 2.2. Artificial neural network

Artificial neural network is a computational modelling tool that consists of groups of highly interconnected processing elements called neurons or nodes. ANNs are inspired by the architecture of cerebral networks [14]. The neurons are arranged in series of layers: one input layer with nodes representing independent variables, one output layer with nodes representing dependent variables, and several hidden layers (usually 1–3) which associate the inputs with outputs. Each neuron from one layer is connected with each neuron in the next layer (Fig. 2). As relevant inputs, as well as outputs for ANN training, data coming from ED are used.

The training is carried out by adjusting strength of connections between neurons with the aim to adapt the outputs of whole network to be closer to the desired outputs, or to minimise the sum of the squared error (SSE) of the training data. SSE is an error function composed by squaring the difference between sets of target and actual values, and by adding these together. SSE computed at the output side is propagated backwards from this layer to the hidden layer and at the end to the input layer. During the training phase, each neuron in the hidden layer sums its input signals  $x_i$  after multiplying them by the strengths of the respective connections called weights ( $w_{ij}$ ), and computes its output  $y_j$  as a function of the sum, Eq. (1):

$$y_j = f \left( \sum w_{ij} x_i \right) \quad (1)$$

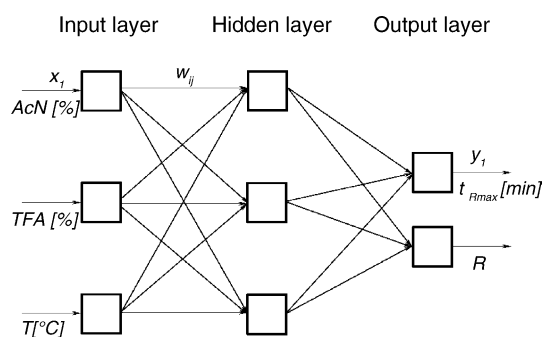


Fig. 2. Scheme of three-layer network used in Case I.

where  $f$  is the activation function (Eq. (2)) that is necessary to transform the weighted sum of all signals connecting with a neuron.

$$f = \frac{1}{1 + e^{(1-x)/z}} \quad (2)$$

where  $z$  is the gain determining the slope of the sigmoid function.

The training phase is finished when SSE is minimised across all training cases. Once ANN has been trained, it has good predictive capability and ability to accurately describe the response surface even without any knowledge about the physical and chemical background of the modelled system. Since ANNs are learning to associate the inputs with the outputs, the network may get over-trained, i.e. does not approximate the well studied system. Therefore, to keep independent check of the ANN performance, relevant verification and test data sets, different from training data set, should be created.

### 3. Experimental

#### 3.1. Chemicals

The neuroprotective peptide samples—[G<sup>14</sup>]-humanin (HNG): MAPRGFSCLLLLTGEIDLVPKRR, [W<sup>14</sup>]-humanin (HNW): MAPRGFSCLLLLTWEIDLVPKRR, and the peptide mixture of (HNG) derivatives (number of individual peptides estimated by MALDI-TOF-MS: nine dominant components, 25% of content unidentified; described elsewhere [8]), used in this study were purchased from Clonestar Biotech (Brno, Czech Republic). Bradykinin, renin substrate tetradecapeptide, angiotensin I, adrenocorticotrophic hormone (fragments 18–39), pancreatic bovine insulin,  $\alpha$ -cyano-4-hydroxycinnamic acid and trifluoroacetic acid (TFA) were from Sigma (Steinheim, Germany). HPLC grade acetonitrile was from Merck (Darmstadt, Germany). All reagents, except for peptides, were of the analytical grade purity. Deionised water used to prepare all solutions was produced in a commercial apparatus from Premier MFG'D Systems (Phoenix, AZ, USA).

#### 3.2. Instrumentation

The Shimadzu HPLC 10A VP system (Kyoto, Japan) consisting of a system controller (SCL-10A VP), a degasser (GT-154), a pump (LC-10AD VP), a column oven (CTO-10AS VP), a diode array detector (SPD M10A VP) and a fraction collector (FRC-10A) was used. The system was controlled by Class-VP 6.12 SP5 software. Commercial pre-packed analytical HPLC column Luna C8(2) (3  $\mu$ m particle size, 150 mm  $\times$  4.6 mm I.D.) as well as the guard column (4  $\times$  2 I.D.) containing the same packing material were from Phenomenex (Torrance, CA, USA).

The composition of individual peptide fractions was checked by means of matrix assisted laser

desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS). All measurements of mass spectra were performed on AXIMA CFR (Kratos Analytical, Manchester, UK) mass spectrometer, equipped with the nitrogen laser (Franklin, MA, USA) wavelength of 337 nm.

#### 3.3. Procedures

##### 3.3.1. Calculation procedures

The experimental designs were devised using Statistica 6.1 software (StatSoft, Tulsa, OK, USA). For neurocomputing, TRAJAN Neural Network Simulator version 3.0D (Trajan Software Ltd., Durham, UK) was employed. Additional experiments, so called verification and test data sets, were conducted to confirm reliability of ANN performance. Verification data set consisted of five sets of input and output data, whereas test data set of four. Inputs were randomly chosen inside the space given by input variable limits and related outputs were experimentally obtained. Then, these were submitted to once trained ANN and SSE for these data sets were evaluated.

##### 3.3.2. Analytical procedures

Individual peptides of model peptide mixture (*Case I*) as well as the complex peptide mixture (*Case II*) were for the RPLC analysis dissolved in deionised water (1 mg/ml). Composition of the mobile phase, as well as its pH, was changed in accordance with proposed ED. The temperature of the column was also included in this study. All experiments were performed using isocratic elution at the flow rate 1 ml/min in triplicate.

The mass spectrometric analysis was done in this way: 1  $\mu$ l of matrix was added to 1  $\mu$ l of the sample solution and mixed by pipetting up and down directly on the sample target. Then, the sample target was allowed to dry in the air stream and introduced to the mass spectrometer. The energy of the laser was changed in relative units, where 180 units correspond to the maximum power of 6 mW and 0 units to 0 mW. External calibration with bradykinin, renin substrate tetradecapeptide, angiotensin I, adrenocorticotrophic hormone (fragments 18–39) and bovine insulin was used. As a matrix,  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA) dissolved in acetonitrile and water (1:1) was utilised.

### 4. Results and discussion

In this study, we modelled two situations relevant for separation of peptide mixture of unknown composition by means of IP-RPLC. First, we dealt with separation of simple mixture of G<sup>14</sup>-humanin (HNG) and W<sup>14</sup>-humanin (HNW) assigned as *Case I*. The difference in the structure of these two peptides is just one amino acid of 24 total, where we do not expect changes in protonation properties, but only in structure. Thus, these two peptides represent a challenge for separation, as well as this case is similar to many other real cases of peptide

Table 1  
Experimental conditions (inputs) and results (outputs) used for optimisation of separation in the *Case I*

| Experiment | AcN (% v/v) | TFA (% v/v)  | <i>T</i> (°C) | <i>t</i> <sub>Rmax</sub> (min) | <i>R</i>     |
|------------|-------------|--------------|---------------|--------------------------------|--------------|
| 1          | 28.6        | 0.070        | 35            | 62.20                          | 15.90        |
| 2          | <b>30.0</b> | <b>0.050</b> | <b>23</b>     | <b>19.11</b>                   | <b>16.83</b> |
| 3          | <b>30.0</b> | <b>0.050</b> | <b>47</b>     | <b>6.05</b>                    | <b>17.55</b> |
| 4          | <b>30.0</b> | <b>0.090</b> | <b>23</b>     | <b>46.43</b>                   | <b>15.29</b> |
| 5          | <b>30.0</b> | <b>0.090</b> | <b>47</b>     | <b>41.75</b>                   | <b>15.36</b> |
| 6          | 32.0        | 0.036        | 35            | 5.07                           | 10.65        |
| 7          | 32.0        | 0.070        | 15            | 11.68                          | 14.48        |
| 8          | <b>32.0</b> | <b>0.070</b> | <b>35</b>     | <b>11.07</b>                   | <b>15.48</b> |
| 9          | <b>32.0</b> | <b>0.070</b> | <b>35</b>     | <b>10.83</b>                   | <b>15.44</b> |
| 10         | 32.0        | 0.070        | 55            | 10.33                          | 16.44        |
| 11         | 32.0        | 0.100        | 35            | 27.60                          | 20.96        |
| 12         | <b>34.0</b> | <b>0.050</b> | <b>23</b>     | <b>2.98</b>                    | <b>6.48</b>  |
| 13         | <b>34.0</b> | <b>0.050</b> | <b>47</b>     | <b>3.03</b>                    | <b>7.17</b>  |
| 14         | <b>34.0</b> | <b>0.090</b> | <b>23</b>     | <b>14.36</b>                   | <b>16.40</b> |
| 15         | <b>34.0</b> | <b>0.090</b> | <b>47</b>     | <b>14.22</b>                   | <b>18.13</b> |
| 16         | 35.4        | 0.070        | 35            | 4.70                           | 9.55         |

Central composite design: 1–16; **cube design**: in bold; *star design*: in italics.

separation. In *Case II*, we studied rather complicated mixture of HNG derivatives (with different peptide chain length) where the exact number of peptides/analytes in the separated mixture was relatively high.

There are several possibilities how to approach the optimisation of the separation with regard to the goal (minimal run time, maximal resolution, etc.). Usually, optimum separation means that all components of the sample are separated in a reasonable time. The time of analysis is limited by the retention time of the most retentive component. Another important aspect of separation is resolution, which should be in a case of two components at least 1.5. Hence, as the output data, we chose two criteria in *Case I*: retention time of a more retentive component (*t*<sub>Rmax</sub>) that should be as low as possible, and resolution that should possess the greatest value while reasonable maximum peak retention time is maintained. In *Case II*, the situation is more complicated as it was mentioned before, especially the unknown number of separated components. As in *Case I*, we followed the retention time of the most retentive component (*t*<sub>Rmax</sub>). Since evaluation of the resolution between each pair of peaks is quite complex, sum of resolutions ( $\sum R$ ) was monitored. However, these parameters did not describe the system completely. To appraise sum of resolution, it is necessary to know the number of peaks. Therefore, the number of (significant) peaks (*n*) observed in chromatogram was monitored as the most important factor. Determination of the number of peaks is quite unusual, but these variables with respect to the others enable better description of the system. We also preferably used the isocratic elution not to compromise the selectivity of separation and thus possible maximal number of peaks.

#### 4.1. Case I

In the first attempt to optimise IP-RPLC separation of the mixture of HNG and HNW using FED, preliminary four-factor central composite design was devised to map out the

space of the experimental conditions setting and to estimate the limits of factors. The preliminary CCD included the mobile phase composition: acetonitrile content in the range 0–80% (v/v), TFA concentration in the range 0.02–0.1% (v/v), pH in the range 2–4 and column temperature in the range 10–55 °C. According to the respective results, the limits of factors were adjusted and new CCD was proposed. The pH of aqueous component of the mobile phase was excluded from this CCD because it was observed that pH influences mainly the time of analysis. The time of analysis decreased with the decreasing value of pH while the resolution changed slightly ( $\Delta\text{pH} = 0.1$  resulted in  $\Delta R \sim 0.5\%$ ). Thence, the limit pH value for the given column equal to 2 was chosen. Therefore, three factors were considered in the new CCD. The content of acetonitrile was changed in the range from 28% to 36%, the content of TFA was changed in the range from 0.04% to 0.1%. The range of the column temperature was changed from 10 to 55 °C. The input, the mobile phase composition and the column temperature, as well as the output data (*t*<sub>Rmax</sub>, *R*) that were used for ANN training, are listed in Table 1. The architecture of ANN was optimised by monitoring of the dependence of SSE on the number of nodes in hidden layer. The increase of number of hidden nodes was stopped when SSE did not decrease more. Furthermore, SSE of verification as well as test data sets were also monitored (Table 2). For both *Cases*, the optimal architecture of ANN was found comprising of three neurons in the input layer,

Table 2  
The values of SSE of the training, verification and data sets

| Data set    | Training | Verification | Test   |
|-------------|----------|--------------|--------|
| Case I—CCD  | 1.7280   | 4.7660       | 2.6250 |
| Case I—CD   | 1.4760   | 3.4600       | 4.8300 |
| Case I—SD   | 2.3500   | 3.9060       | 3.5240 |
| Case II—CCD | 1.3110   | 2.8150       | 0.9847 |
| Case II—CD  | 0.7375   | 3.2610       | 1.0930 |
| Case II—SD  | 1.3740   | 3.3470       | 1.9630 |

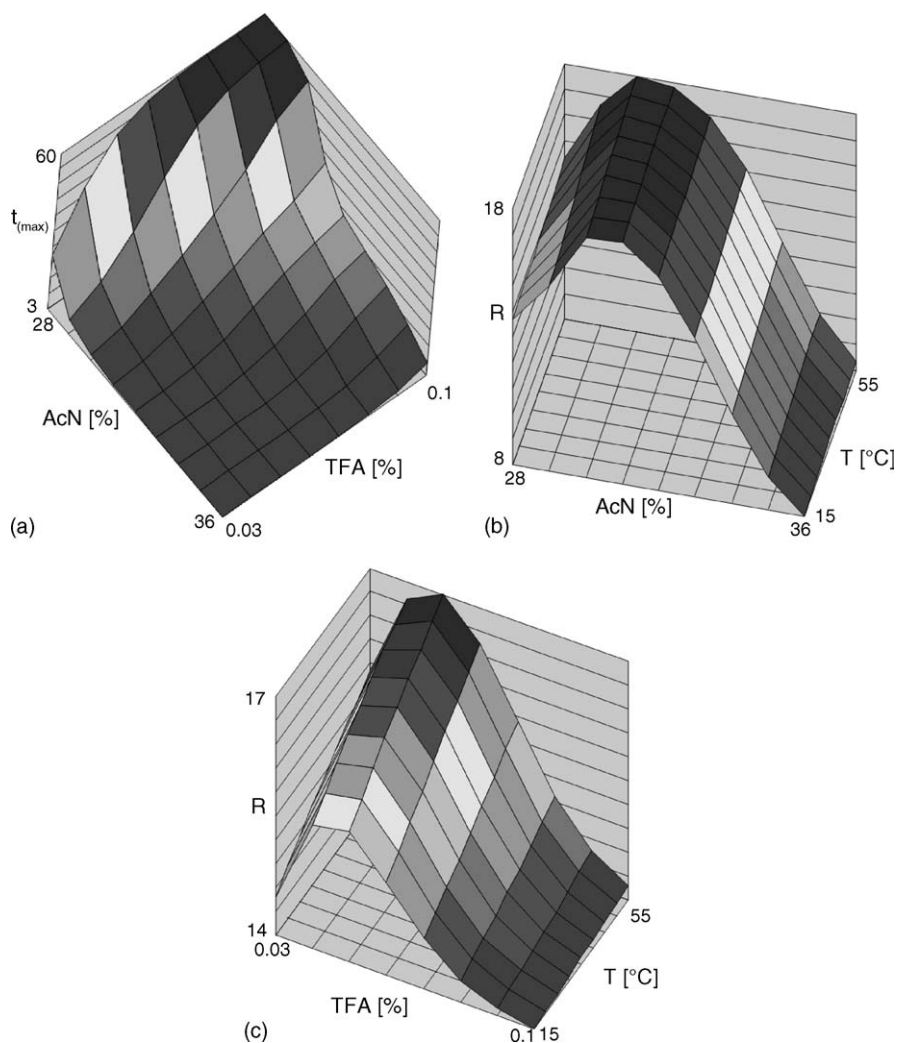


Fig. 3. Response surface diagrams for *Case II*: (a)  $t_{Rmax}$  as a function of acetonitrile and TFA content; (b) resolution as a function of column temperature and content of acetonitrile and (c) resolution as a function of column temperature and content of TFA.

and three neurons in the hidden layer. In *Case I* the architecture comprises two neurons in output layer (two output variables). ANN was trained and used to predict the peak with maximal retention time and the resolution across the range of eluent conditions within the search area. Then, six response surfaces were created as a three dimensional graphs of dependence of output variables ( $t_{Rmax}$ ,  $R$ ) on the input variables (AcN content, TFA content, and temperature). The analysis of the response surfaces indicated that the optimal value for resolution was found with 30% acetonitrile, 0.04% TFA and 55 °C while  $t_{Rmax}$  was below 15 min (Fig. 3). The experiment was conducted under the optimal separation conditions (see Fig. 4) and the determined maximum peak retention time (retention time of HNW) and the resolution were  $11.5 \pm 0.4$  min (predicted value  $8.5 \pm 1.7$  min) and  $15.7 \pm 1.8$  (predicted value  $14.9 \pm 3.0$ ), respectively (Table 3).

In the following step, the prediction capability of ANN using the data from the FrED was examined. The central composite design was divided into the central two-level three-factor cube design (CD) and the star design (SD). The data

of CD and SD used for ANN training are listed in Table 1 (in bold and italics, respectively). The proper ANN were trained and the results implicit from the analysis of the response surface diagrams were similar to the previous proposal. In case of CD, the predicted  $t_{Rmax}$  and the resolution were  $4.0 \pm 1.5$  and  $11.4 \pm 4.4$ , respectively. In the case of SD, the predicted values for  $t_{Rmax}$  and the resolution were  $13.6 \pm 3.1$  and  $14.5 \pm 3.3$ , respectively. As it is obvious in Table 3, the RSD of the predicted and measured values of the maximum retention time as well as the resolution is the highest in case of the cube design. With regard to the low RSD values coming from the SD data, this design was used as a suitable starting fractional design in optimisation of HNG derivatives mixture (*Case II*).

#### 4.2. Case II

In the first step, the star design of experiments was proposed and applied. The input as well as the output data for ANN training are listed in Table 4 in italics. In *Case II*, there

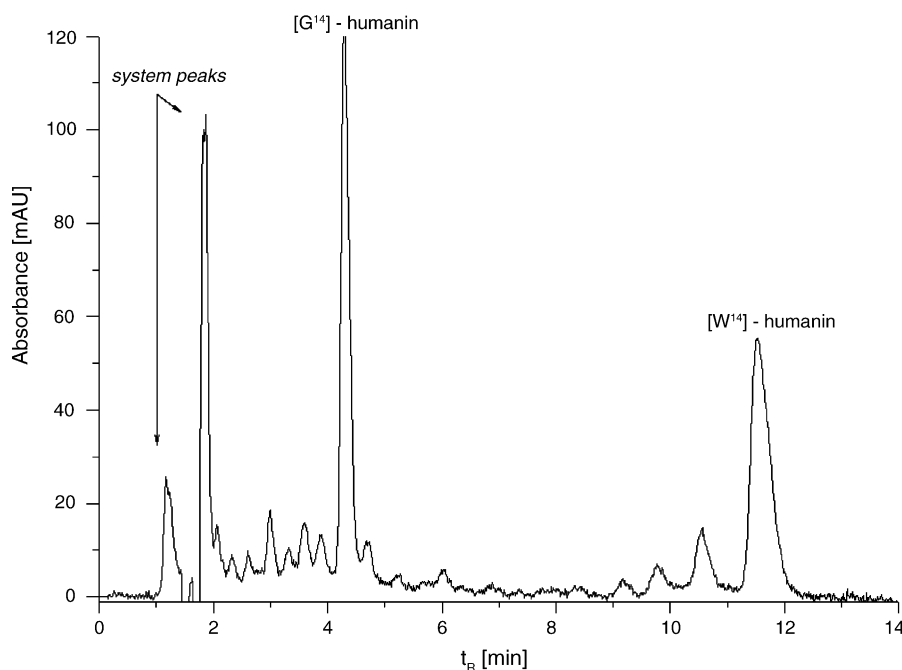


Fig. 4. Chromatogram of the *Case I*. Separation conditions: Luna C8(2) (3  $\mu$ m, 150 mm  $\times$  4.6 mm I.D.), 0.04% TFA, 30% acetonitrile, 55  $^{\circ}$ C, 1 ml/min,  $\lambda$  = 200 nm.

Table 3  
Measured and predicted data for *Case I*

| Data                     | $t_{Rmax}$ (min) | RSD (%) | $R$  | RSD (%) |
|--------------------------|------------------|---------|------|---------|
| Measured                 | 11.5             | –       | 15.7 | –       |
| Central composite design | 8.5              | 13      | 14.9 | 2       |
| Cube design              | 4.0              | 33      | 11.4 | 14      |
| Star design              | 13.6             | 9       | 14.5 | 4       |

RSD—relative standard deviation.

were three neurons in output layer (three output variables), thus the optimal ANN structure was 3:3:3. The trained ANN was then used to predict the number of peaks, the maximum peak retention time, as well as the sum of resolution across

the range of eluent conditions within the search area. Note that the most important criterion in the optimisation of HNG derivatives mixture separation was the number of peaks. Optimal separation conditions were estimated from the analysis of nine response surface diagrams and they were: 0.04% TFA, 29% AcN and 55  $^{\circ}$ C. The analysis under these conditions was run (Fig. 5). Since the prediction capability was not fully satisfactory (Table 5), this data point was added to the training set and the ANN was trained again. Using the newly trained ANN, unfortunately, the results were not improved.

Therefore, in the next step, it was decided to return to the more complex, but better surface describing CCD. The ANN with the optimal architecture using the data from CCD was

Table 4  
Experimental conditions (inputs) and results (outputs) used for optimisation of separation for in *Case II*

| Experiment | AcN (% , v/v) | TFA (% , v/v) | $T$ ( $^{\circ}$ C) | $n$       | $t_{Rmax}$ (min) | $\sum R$     |
|------------|---------------|---------------|---------------------|-----------|------------------|--------------|
| 1          | 28.6          | 0.070         | 35                  | 8         | 29.05            | 27.18        |
| 2          | <b>30.0</b>   | <b>0.050</b>  | <b>23</b>           | <b>10</b> | <b>10.19</b>     | <b>16.46</b> |
| 3          | <b>30.0</b>   | <b>0.050</b>  | <b>47</b>           | <b>13</b> | <b>9.97</b>      | <b>20.94</b> |
| 4          | <b>30.0</b>   | <b>0.090</b>  | <b>23</b>           | <b>8</b>  | <b>21.10</b>     | <b>26.14</b> |
| 5          | <b>30.0</b>   | <b>0.090</b>  | <b>47</b>           | <b>8</b>  | <b>20.18</b>     | <b>23.17</b> |
| 6          | 32.0          | 0.036         | 35                  | 5         | 2.81             | 4.57         |
| 7          | 32.0          | 0.071         | 15                  | 7         | 4.38             | 8.93         |
| 8          | 32.0          | <b>0.071</b>  | 35                  | 7         | <b>4.51</b>      | <b>9.25</b>  |
| 9          | 32.0          | <b>0.071</b>  | 35                  | 7         | <b>4.49</b>      | <b>9.98</b>  |
| 10         | 32.0          | 0.071         | 55                  | 7         | 4.45             | 9.60         |
| 11         | 32.0          | 0.100         | 35                  | 7         | 8.06             | 16.37        |
| 12         | <b>34.0</b>   | <b>0.050</b>  | <b>23</b>           | <b>3</b>  | <b>4.70</b>      | <b>8.97</b>  |
| 13         | <b>34.0</b>   | <b>0.050</b>  | <b>47</b>           | <b>5</b>  | <b>4.67</b>      | <b>12.13</b> |
| 14         | <b>34.0</b>   | <b>0.090</b>  | <b>23</b>           | 7         | <b>5.63</b>      | <b>10.17</b> |
| 15         | <b>34.0</b>   | <b>0.090</b>  | <b>47</b>           | 7         | <b>5.80</b>      | <b>10.53</b> |
| 16         | 35.4          | 0.068         | 35                  | 3         | 2.23             | 2.91         |

Central composite design: 1–16; **cube design**: in bold; *star design*: in italics.  $n$ : number of peaks and  $\sum R$ : sum of resolutions.

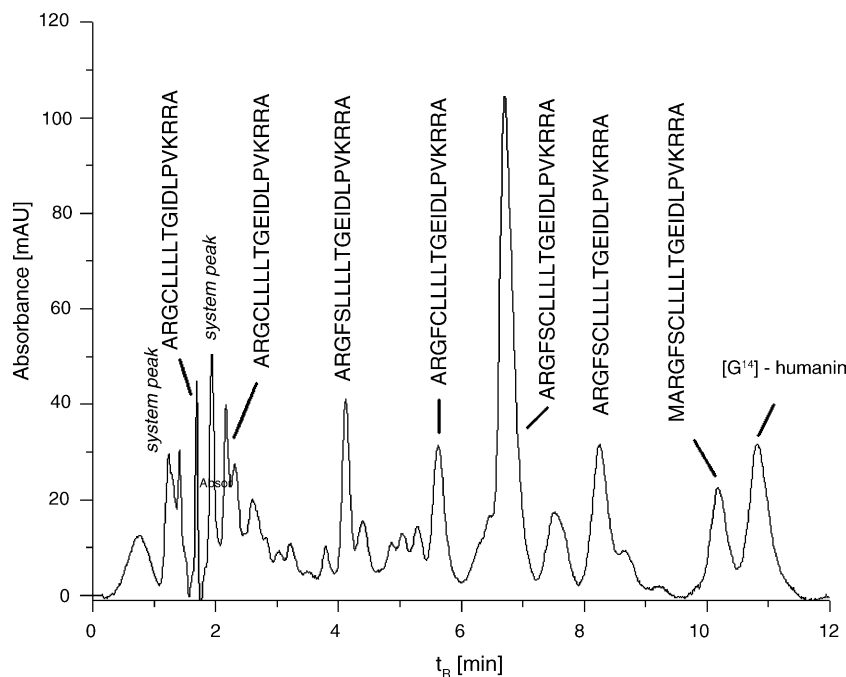


Fig. 5. Chromatogram of the *Case II*. Separation conditions: Luna C8(2) (3  $\mu$ m, 150 mm  $\times$  4.6 mm I.D.), 0.04% TFA, 29% acetonitrile, 55  $^{\circ}$ C, 1 ml/min,  $\lambda$  = 200 nm.

trained and then the response surface analysis and prediction was done. The analysis of diagrams confirmed that the optimal separation conditions found in the case of SD were correct. The predictive capability was very good, considering the number of peaks and the resolution, but unsatisfactory in case of the maximum retention time where RSD was over 30%.

Since the data of the cube design as a part of the central composite design have already been measured, new ANN using these training data was trained. The results concerning the optimal separation conditions were again in agreement with the previous ones (data not shown).

#### 4.3. MALDI-TOF mass spectrometry of G14-humanin derivatives

When the optimal separation conditions were found, individual fractions with peptides were collected and analysed. The composition of the fractions was analysed off-line by means of the matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF-MS). Using

MALDI-TOF-MS, eight peptides in the mixture were identified in accordance to mass spectrometric analysis of unseparated mixture [8]. The identification of the other peptides was probably disabled by insufficient separation of peptides (Fig. 5) or their non-specific co-elutions, where both effects could cause mutual suppression of ionisation [15].

## 5. Conclusions

The combination of ED, the FED as well as the FrED, with ANN has been found to be general and effective tool for optimisation of the RPLC separation of peptides mixtures. Unlike to approaches using hard or semi-hard models for prediction of retention times [21,22], combination of ED-ANN allows to optimise separation conditions regardless of what we know about the structure and other physico-chemical properties of peptidic analytes. Another advantage of the use of ED is that it does not require excessive number of experiments in comparison with the single variable approach where at least 14 experiments are needed.

In *Case I*, using the star design only eight experiments were sufficient to reach the optimal separation conditions. The ANN trained with the data set coming from the star design has good predictive capability and ability to model the response surface of outputs. In *Case II*, the results were quite complex. This mixture can, in some way, model complicated peptide matrix, e.g. protein digest (10–30 individual tryptic peptides) or biological fluids, where the exact number of peptides is not known and the primary goal of the analysis is to separate as many peptides as possible, e.g. before mass spectrometric analysis.

Table 5  
Measured and predicted data for *Case II*

| Data                     | <i>n</i> | RSD (%) | <i>t</i> <sub>Rmax</sub> (min) | RSD (%) | <i>R</i> | RSD (%) |
|--------------------------|----------|---------|--------------------------------|---------|----------|---------|
| Measured                 | 12       | –       | 10.8                           | –       | 35.8     | –       |
| Central composite design | 12       | 0.7     | 18.3                           | 35      | 23.8     | 17      |
| Cube design              | 13       | 3       | 16.4                           | 26      | 25.1     | 15      |
| Star design              | 8        | 19      | 15.9                           | 23      | 18.7     | 24      |

RSD—relative standard deviation.

The analysis of the response surfaces diagrams, obtained from the ANN trained with the data sets coming from SD, CD, as well as from CCD, displayed the similar optimal separation conditions settings for IP-RPLC. Though, from the point of view of the prediction precision, the relative standard deviations of predicted outputs were varying in a range from 0.7% to 35%. Using CCD, RSD of predicted values of the number of peaks, the main criterion, and the resolution was reasonable, 0.7% and 17%, respectively. Concerning the RSD of the predicted  $t_{Rmax}$ , the value was over 30%. In the case of FrED, the relative standard deviations of predictions of all outputs for SD were relatively high (>20%) while using CD, only RSD of predicted  $t_{Rmax}$  was over 20%.

To conclude, FED produced sufficient input data sets to optimise IP-RPLC separation of peptides for even complicated systems. Furthermore, we have shown that FrEDs were adequate to reach satisfactory optimisation while halving the number of experiments needed for FED. Thus, FrED can be utilised with benefit instead of FED for optimisation of RPLC separation of peptides by ED-ANN. As it can be seen in the results obtained by optimisation of the simple mixture separation, the FrED represented by central two-level star design was found to be suitable. For the more complex systems such as *Case II*, cube design standing was found more appropriate.

## Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, MSM 143100011 and by the Grant Agency of the Czech Republic, grant no. 305/03/1100 and 203/02/1103.

## References

- [1] J. Havel, E.M. Peña, A. Rojas-Hernández, J.P. Doucet, A. Panaye, J. Chromatogr. A 793 (1998) 317.
- [2] J. Havliš, E.J. Madden, A.L. Revilla, J. Havel, J. Chromatogr. B 755 (2001) 185.
- [3] K. Petritis, L.J. Kangas, P.L. Ferguson, G.A. Anderson, L. Pasatolic, M.S. Lipton, K.J. Auberry, E.F. Strittmatter, Y.F. Shen, R. Zhao, R.D. Smith, Anal. Chem. 75 (2003) 1039.
- [4] S. Agatonovic-Kustrin, M. Zecevic, L.J. Zivanovic, I.G. Tucker, J. Pharm. Biomed. Anal. 17 (1998) 69.
- [5] H.J. Metting, P.M.J. Coenegracht, J. Chromatogr. A 728 (1996) 47.
- [6] Y. Hashimoto, Y. Ito, T. Niikura, Z. Shao, M. Hata, F. Oyama, I. Nishimoto, Biochem. Biophys. Res. Commun. 283 (2001) 460.
- [7] O. Šedo, K. Novotná, J. Havel, Rapid Commun. Mass. Spectrom. 18 (2004) 339.
- [8] K. Novotná, O. Šedo, J. Havel, J. Appl. Biomed. 2 (2004) 113.
- [9] W.S. Hancock, R.C. Chloupek, J.J. Kirkland, L.R. Snyder, J. Chromatogr. A 686 (1994) 31.
- [10] D.C. Guo, C.T. Mant, R.S. Hodges, J. Chromatogr. 386 (1987) 205.
- [11] N.M. McKern, H.K. Edskes, D.D. Shukla, Biomed. Chromatogr. 7 (1993) 15.
- [12] Y. Chen, C.T. Mant, R.S. Hodges, J. Chromatogr. A 1010 (2003) 45.
- [13] E. Morgan, Chemometrics: Experimental Design, Wiley, New York, 1997.
- [14] D.E. Rumelhart, G.E. Hinton, R.J. Williams, in: D.E. Rumelhart, J.L. McClelland (Eds.), Parallel Distributed Processing: Exploration in the Microstructure of Cognition, vol. 1, MIT Press, Cambridge, MA, 1986, p. 318 (Chapter 8).
- [15] J. Preisler, P. Hu, T. Rejtar, B.L. Karger, Anal. Chem. 72 (2000) 4785.
- [16] Y.L. Loukas, J. Chromatogr. A 904 (2000) 119.
- [17] S.Y. Tham, S. Agatonovic-Kustrin, J. Pharm. Biomed. Anal. 28 (2002) 581.
- [18] S. Agatonovic-Kustrin, M. Zecevic, L.J. Zivanovic, J. Pharm. Biomed. Anal. 21 (1999) 95.
- [19] E. Marengo, V. Gianotti, S. Angioi, M.C. Gennaro, J. Chromatogr. A 1029 (2004) 57.
- [20] T. Vasiljević, A. Onjia, D. Čokeša, M. Laušević, Talanta 64 (2004) 785.
- [21] R. Kalisz, T. Bączek, A. Buciniński, B. Buszewski, M. Sztupecka, J. Sep. Sci. 26 (2003) 271.
- [22] O.V. Krokhin, R. Craig, V. Spicer, W. Ens, K.G. Standing, R.C. Beavis, J.A. Wilkins, Mol. Cell. Proteomics 3 (2004) 908.