

On-capillary derivatization using a hybrid artificial neural network-genetic algorithm approach

Toni Riveros,^a Grady Hanrahan,^{*b} Sarah Muliadi,^b Jennifer Arceo^b and Frank A. Gomez^{*a}

Received 8th May 2009, Accepted 20th July 2009

First published as an Advance Article on the web 3rd August 2009

DOI: 10.1039/b909143b

The first reported hybrid artificial neural network-genetic algorithm (ANN-GA) approach for the optimization of on-capillary dipeptide derivatization is presented. More specifically, genetic optimization proved valuable in the determination of effective network structure with three defined parameter inputs: (i) phthalic anhydride injection volume, (ii) time of injection, and (iii) voltage, for the maximum conversion of the dipeptide D-Ala-D-Ala by phthalic anhydride. Results obtained from the hybrid approach proved superior to an ANN model without GA optimization in terms of training data and predictive ability. The model developed will likely prove useful for the analysis of other organic-based reaction systems.

1. Introduction

The past three decades have shown great developments in capillary electrophoresis (CE) providing the analytical chemist with a vast array of applications and techniques to choose from when analyzing and/or separating mixtures of a variety of different types of materials.^{1–8} Analysis of microreactions in an open-tubular format is one area that the small sample requirements of CE has allowed for facile study of a myriad of compounds.^{9–22} A variety of both enzyme-catalyzed and organic-based reactions have been detailed that demonstrate the ease in chemically modifying a molecule with subsequent detection generally by UV-means.^{13,15} For the former, the pioneering work of Regnier and co-workers laid the ground work for the technique termed electrophoretically mediated microanalysis (EMMA).^{10–12} Here, differential electrophoretic mobilities of analyte and reagent(s) are used to merge zones of both within the capillary, thereby mimicking a ‘reaction’ but at the microscale. The technology has primarily focused on biochemical applications thereby yielding enzyme activities, Michaelis and inhibition constants.^{18,19,23}

In the latter (non-enzyme-based reactions) there has been less work which is peculiar given, one, the vast number of organic reactions that exist and, two, the potential for high-throughput syntheses (albeit at the micro-level) of vast numbers of drug and other targets.^{24,25} The separation and detection of these species is now trivial as CE can provide various separation mechanism(s) much like HPLC and can be coupled to an array of detector systems. Yet, determining the best conditions for separation can still be a tedious task complicated by the potential variables involved in methodological development. Electrophoretic-based techniques require attention be given to a host of parameters in order to ascertain the optimal possible conditions for a given

application. For example, voltage, capillary length and buffer composition are critical in the separation of molecular species. However, information processing techniques such as artificial neural networks (ANNs) that provide non-linear modeling of response surfaces and optimization of electrophoretic conditions, especially in regards to on-capillary derivatization, are limited. Few single-step organic reactions have been examined by EMMA-type methods, and of these, none has involved the use of an ANN. Incorporating ANN would be an important advancement of the microreactor concept and would broaden the types of molecules that could be studied using this analytical technique. Herein, we describe the use of an ANN in the on-capillary derivatization of a single-step organic reaction (Fig. 1). More specifically, we employed the use of a hybrid artificial neural network-genetic algorithm (ANN-GA) approach in the derivatization of the dipeptide D-Ala-D-Ala by phthalic anhydride.

2. Experimental

2.1. Data set

Thirty-two on-capillary derivatizations were used as the study data set. This data set was randomly divided into training (22 derivatizations), test (5 derivatizations) and validation (5 derivatizations) sets. During all applications, data sets were converted by normal transformation (mean 0, standard deviation 1). All data were cross-checked to ensure accuracy and validity. Both JMP (SAS, Inc., USA) and MATLAB 7.8 (The MathWorks,

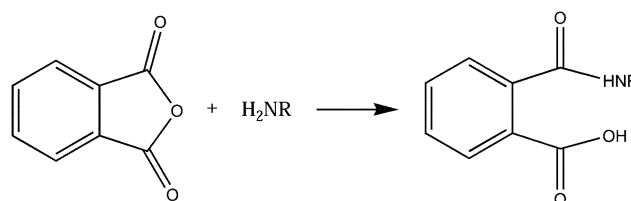


Fig. 1 Derivatization of amines by phthalic anhydride.

^aDepartment of Chemistry and Biochemistry, California State University, Los Angeles, 5151 State University Drive, Los Angeles, CA 90032, USA. E-mail: fgomez2@calstatela.edu

^bDepartment of Chemistry, California Lutheran University, 60 West Olsen Road, Thousand Oaks, CA 91360, USA. E-mail: ghanraha@clunet.edu

USA) with Neural Network and Genetic Algorithm Toolboxes were utilized.

2.2. Chemical reagents

Stock solutions of phthalic anhydride (5 mM) were prepared by dissolving the solid in acetonitrile. Stock solutions of D-Ala-D-Ala (5 mM) were prepared by dissolving in acetonitrile–water 4 : 1 v/v. Sample solutions of phthalic anhydride (1 mM) and D-Ala-D-Ala (1 mM) were prepared by dissolving stock solution in 20 mM phosphate buffer and adjusted to pH 9.4 by the addition of NaOH.

2.3. Instrumental procedure

In this work, all experimental analyses were carried out using a P/ACE MDQ CE instrument (Beckman Instruments, Fullerton, CA, USA). The capillary tubing was of uncoated fused silica with an internal diameter of 50 μm , a length from inlet to detector of 40 cm and a length from detector to outlet of 10.2 cm. Sample solutions (6 nL) containing D-Ala-D-Ala were introduced by pressure injection (5 seconds; 0.5 psi) into the capillary equilibrated with buffer. Next, samples of phthalic anhydride, (3.6 nL, 8.4 nL) were introduced by pressure injection (3, 7 seconds; 0.5 psi) and electrophoresis was run at a given contact voltage

(5 and 15 kV) for the desired mixing time (0.2 and 1.0 min). The voltage was adjusted to 25 kV to complete elution of all species.

2.4. ANN-GA strategy

Fig. 2 shows the ANN-GA strategy employing Levenberg–Marquardt back-propagation with a sigmoid transfer function used to search for optimal neural network architecture, and to aid in optimizing experimental conditions for maximum conversion to product. As ANN input, three parameters (phthalic anhydride injection volume, time of injection and voltage) were selected on the basis of previous univariate studies. In addition, we carefully considered our knowledge of the derivatization on known experimental protocols for providing a foundation for the chosen parameters. Genetic optimization has proven valuable in the determination of efficient neural network structure through a population of individuals, which evolves toward optimum solutions through genetic operators (selection, crossover and mutation).^{26–29} Shown in Fig. 2 is the schematic flow of the ANN-GA approach with user-defined inputs, genetic algorithm optimization and training components incorporating the operators defined above. ANN provided a non-linear model paradigm and the GA allowed optimization of the input space of the ANN through the combined efforts of topology selection, network training, evolution of weights, and node determination. The GA operated according to a general

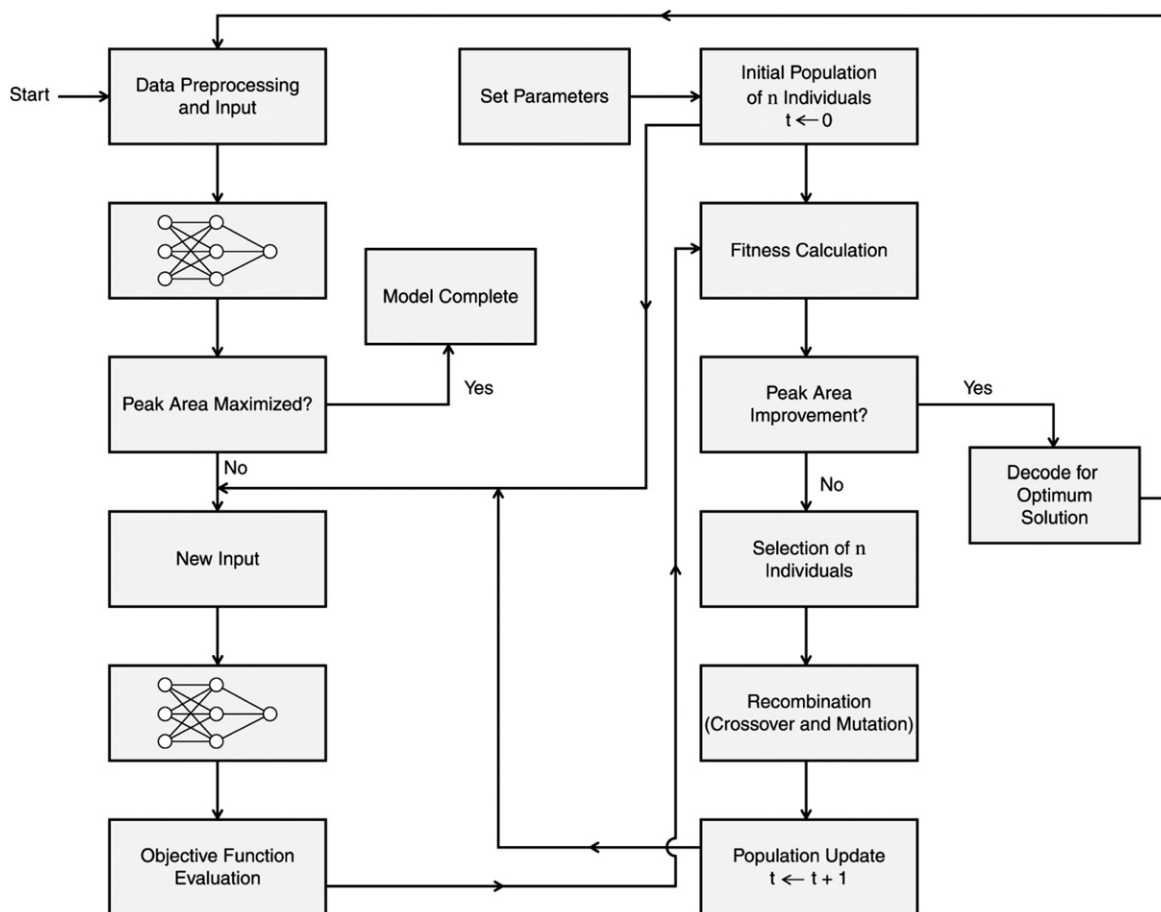


Fig. 2 Schematic of the hybrid artificial neural network-genetic algorithm (ANN-GA) method.

two-step process: (i) initialization of the population and evaluation of each member of the initial population, and (ii) reproduction until a stopping condition was met. More detailed description is provided below.

An initial population of random binary strings, each of which represented a specific neural network topology and set of training parameters, were defined. The fitness of each trained network was calculated by:

$$\text{Fitness} = W_1 f_e(\hat{e}) + W_2 f_t(\hat{t}) \quad (1)$$

where $f_e(\hat{e})$ = the error between real output and ANN output, $f_t(\hat{t})$ = the training time of ANN, and W_1 and W_2 = suitable values of weight. According to the fitness observed, the GA selected a new group of strings, which ultimately represented the parents of the next generation of the GA with an assigned probability of reproduction.

The strings were then subjected to the evolutionary operators defined above. Crossover is one of the most important GA operators that combines two chromosomes (parents) to produce a new chromosome (offspring). The crossover process occurs during evolution according to a user-definable crossover probability. For this study, a two-point crossover approach, one that calls for two points to be selected on the parent organism string, was utilized. The mutation process ensures that the probability of searching regions in the problem space is never zero and prevents complete loss of genetic material through reproduction and crossover. The above process was repeated until the maximum number of generations was reached. Once reached, the best string that gave the maximum fitness or minimum mean square error (*MSE*) was considered appropriate. The *MSE* was calculated as:

$$MSE = SSE/n \quad (2)$$

where *SSE* is the sum of squares error given by:

$$SSE = \sum_{i=1}^n \{w_i (y_i - f_i)^2\} \quad (3)$$

where y_i is the observed data value and f_i is the predicted value from the fit. The parameter w_i is the weighting applied to each data point, usually $w_i = 1$. For objective appreciation, synaptic weights and bias were initialized to 0.01 and 0.1, respectively. GA configuration values included a population size of 30, maximum generation of 100, a crossover probability of 0.5, and a mutation probability of 0.01.

3. Results and discussion

The optimal ANN architecture, realized in 20 generations, included three hidden neurons and an *MSE* of 0.193. Fig. 3 shows a plot of *MSE* versus generations for this optimized genetic algorithm run. Fig. 4 presents a plot of experimental peak areas versus peak areas predicted by ANN-GA for training and verification of the optimum architecture. As justification for our model, we ran a back-propagated ANN without GA optimization and compared this to our ANN-GA in terms of training data. ANN training was an iterative process in which the network was given the desired inputs along with correct outputs for those inputs. Weights were evaluated and adjusted to produce

the correct output within a reasonable error margin. Training continued until the total error across the training set was below a specified maximum.

The ANN model proved inferior to the ANN-GA approach in terms of fitting ($r^2 = 0.87$ ($n = 10$) vs. $r^2 = 0.96$ ($n = 10$)) as well as predicting abilities ($r^2 = 0.90$ ($n = 10$) vs. $r^2 = 0.94$ ($n = 10$)). The ANN model was also inferior in terms of the *MSE* (mean *MSE* = 0.227 ($n = 10$) vs. mean *MSE* = 0.197 ($n = 10$)). Sensitivity analyses of ANN-GA were performed to specify appropriate GA parameters (random number, crossover probability, mutation probability) to ensure efficiency of the solution process and global optimality of the obtained solution.

From the data patterned by the ANN-GA, a response surface was generated for the two interactive factors (phthalic anhydride injection volume and voltage), with interaction significance determined by an initial 2^3 full factorial screening design (data not shown). The response surface profiling facility was used for determining optimum values resulting in a predicted peak area maximum (180 092). Validation of experimental results was accomplished by performing electrophoresis of D-Ala-D-Ala and phthalic anhydride at experimental conditions dictated by the ANN-GA model. A series of three replicate experiments using the ANN-GA model predicted that optimal conditions were performed. Here, sound statistical evidence was obtained by

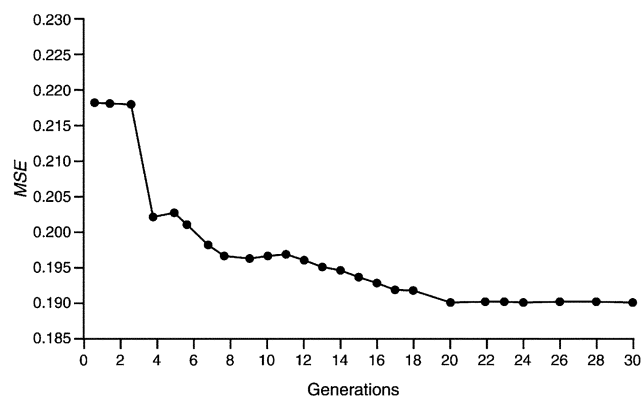


Fig. 3 Plot of best mean square error (*MSE*) versus number of generations.

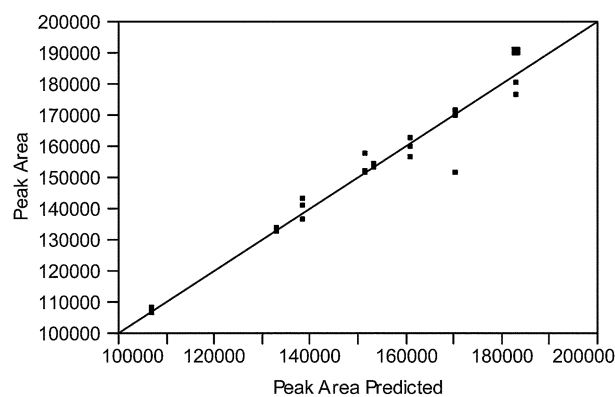


Fig. 4 Plot of experimental peak areas versus peak areas predicted by ANN-GA for training and verification.

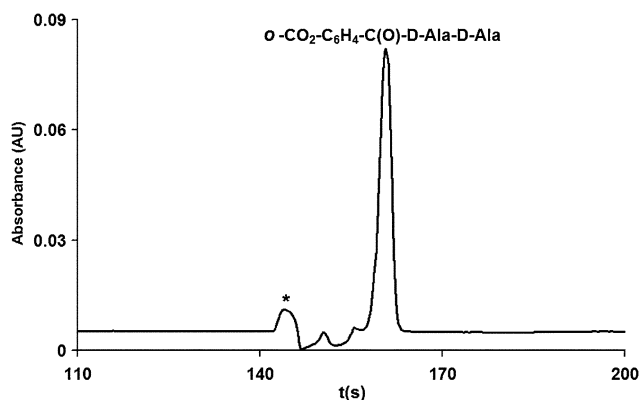


Fig. 5 Representative electropherogram of the reaction product *o*-CO₂-C₆H₄-D-Ala-D-Ala derivatized by phthalic anhydride. Reaction was carried out in 20 mM phosphate buffer (pH 9.4). The total analysis time was 6.2 min at 25 kV (145.4 μA) using a 40.0 cm (inlet to detector) coated capillary. The peak marked * is unreacted peptide.

good RSD values (typically <5%) between replicates and overall agreement (replicate 1 = 4.7%, replicate 2 = 2.5% and replicate 3 = 6.9%) between model predicted and experimental derivatization. A representative electropherogram from replicate 3 is shown in Fig. 5. Here, the derivatization of D-Ala-D-Ala by phthalic anhydride is nearly complete as little peptide is observed in the electropherogram.

To this end, such a study provides guidance for the development and application of ANN-GA tools that can increase the robustness and transparency, and therefore confidence and acceptance of ANN modeling techniques as methods for optimizing on-capillary derivatization of single-step organic reactions. Current work is focused on expanding the present methodology for analysis of other organic-based reaction systems including multi-step ones, and in examining other parameters associated with said reactions. Finally, we acknowledge that as a general-purpose optimization tool, genetic algorithms will likely be applicable to, and aid in, other types of neural networks for which evaluation functions can be derived.

Acknowledgements

The authors gratefully acknowledge financial support for this research by grants from the National Science Foundation (CHE-

0515363, DMR-0351848 and CBET-0723271). The authors also acknowledge support from the John Stauffer Charitable Trust and the Swenson Summer Research Internship Program.

References

- 1 L. Clohs and K. M. McErlane, *J. Pharm. Biomed. Anal.*, 2001, **24**, 545–554.
- 2 N. A. Guzman, *Anal. Bioanal. Chem.*, 2004, **378**, 37–39.
- 3 C. L. Flurer, *Electrophoresis*, 2001, **22**, 4249–4261.
- 4 W. Thormann, R. Theurillat, M. Wind and R. Kudvee, *J. Chromatogr.*, 2001, **924**, 429–437.
- 5 L. K. Amundsen and H. Siren, *Electrophoresis*, 2007, **28**, 99–113.
- 6 M. Hong, A. Cassely, A. Mechref and M. V. Novotny, *J. Chromatogr., B: Biomed. Sci. Appl.*, 2001, **752**, 207–216.
- 7 B. M. Busby and G. Vigh, *Electrophoresis*, 2005, **26**, 3849–3860.
- 8 J. Simal-Gándara, *Crit. Rev. Anal. Chem.*, 2004, **34**, 85–94.
- 9 Z. Glatz, *J. Chromatogr., B*, 2006, **841**, 23–37.
- 10 B. J. Harmon, D. H. Patterson and F. E. Regnier, *Anal. Chem.*, 1993, **65**, 2655–2662.
- 11 D. H. Patterson, B. J. Harmon and F. E. Regnier, *J. Chromatogr.*, 1994, **662**, 389–394.
- 12 D. H. Patterson, B. J. Harmon and F. E. Regnier, *J. Chromatogr.*, 1996, **732**, 119–132.
- 13 D. S. Zhao and F. A. Gomez, *Electrophoresis*, 1998, **19**, 420–426.
- 14 D. S. Zhao and F. A. Gomez, *Chromatographia*, 1997, **44**, 514–520.
- 15 E.-S. Kwak, S. Esquivel and F. A. Gomez, *Anal. Chim. Acta*, 1999, **397**, 183–190.
- 16 Y. Zhang, R. El-Maghrabi and F. A. Gomez, *Analyst*, 2000, **125**, 685–689.
- 17 L. Z. Avila and G. M. Whitesides, *J. Org. Chem.*, 1993, **58**, 5508–5512.
- 18 S. Van Dyck, A. Van Schepdael and J. Hoogmartens, *Electrophoresis*, 2002, **23**, 2854–2859.
- 19 A. R. Whisnant, S. E. Johnston and S. D. Gilman, *Electrophoresis*, 2000, **21**, 1341–1348.
- 20 Q. Xue and E. Yeung, *Nature*, 1995, **373**, 681–683.
- 21 L. J. Cole and R. T. Kennedy, *Electrophoresis*, 1995, **16**, 549–556.
- 22 H.-T. Chang and E. S. Yeung, *Anal. Chem.*, 1993, **65**, 2947–2951.
- 23 K. J. Miller, I. Leesong, J. Bao, F. E. Regnier and F. E. Lytle, *Anal. Chem.*, 1993, **65**, 3267–3270.
- 24 Y. Zhang and F. A. Gomez, *Electrophoresis*, 2000, **21**, 3305–3310.
- 25 M. Azad, C. Silverio, Y. Zhang, V. Villareal and F.A., *J. Chromatogr.*, 2004, **1027**, 193–204.
- 26 M. Mitchell, *An Introduction to Genetic Algorithms*, 1999, MIT Press, Cambridge, MA, p. 205.
- 27 E. Richards, C. Bessant and S. Saini, *Analyst*, 2004, **129**, 355–358.
- 28 T. Gao, L.-L. Shi, H.-B. Li, S.-S. Zhao, H. Li, S.-L. Sun, Z.-M. Su and Y.-H. Lu, *Phys. Chem. Chem. Phys.*, 2009, **11**, 5124–5129.
- 29 M. Jalali-Heravi and A. Kyani, *Eur. J. Med. Chem.*, 2007, **42**, 649–659.