Genetic algorithms in chemometrics

Article *in* Journal of Chemometrics · June 2012

DOI: 10.1002/cem.2426

2 authors:

Ali Niazi
Islamic Azad University
106 PUBLICATIONS **1,367** CITATIONS
SEE PROFILE

Riccardo Leardi
Università degli Studi di Genova
134 PUBLICATIONS **4,545** CITATIONS
SEE PROFILE

JOURNAL OF CHEMOMETRICS

Received: 29 November 2011, Revised: 25 January 2012, Accepted: 1 February 2012, Published online in Wiley Online Library: 15 April 2012

(wileyonlinelibrary.com) DOI: 10.1002/cem.2426

Genetic algorithms in chemometrics

Ali Niazi[a] and Riccardo Leardi[b]*

This review covers the application of Genetic Algorithms (GAs) in Chemometrics. The first applications of GAs in chemistry date back to the 1970s, and in the last decades, they have been more and more frequently used to solve different kinds of problems, for example, when the objective functions do not possess properties such as continuity, differentiability, and so on. These algorithms maintain and manipulate a family, or population, of solutions and implement a "survival of the fittest" strategy in their search for better solutions. GAs are very useful in the optimization and variable selection in modeling and calibration because of the strong effect of the relationship between presence/absence of variables in a calibration model and the prediction ability of the model itself. This review is not a complete summary of the applications of GAs to chemometric problems; its goal is rather to show the researchers the main fields of application of GAs, together with providing a list of references on the subject. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: genetic algorithms; optimization; molecular modeling; calibration

1. INTRODUCTION

Genetic algorithms (GAs) were introduced by Holland in the early 1970s as an optimization approach, with the goal of simulating the evolutionary process of a living species [1,2]. They follow Darwin's classical rules about natural evolution and use random steps to converge to a nonrandom optimal solution [3–5].

Several papers and reviews have been published about their theory and applications in chemistry in different periods. In 1993, the journal *Science* [6] published a paper that gave a general presentation of GAs, some mathematical analysis about how GAs work and how best to use them, and some applications in modeling in several natural evolutionary systems, including immune systems. Another paper in 1995 in *Nature* [7] described a problem of molecular dynamics that had been successfully solved by GAs where conventional techniques had failed. Several reviews about GAs have been published in journals devoted to different research fields. Lucasius and Kateman [8] published a basic tutorial. Hibbert [9] wrote a review about GAs in chemistry, whereas Lucasius *et al.* [10] focused on wavelength selection. In 1998, Kemsley [11] discussed applications of GAs to multivariate classification problems, using an approach derived from the dimension-reduction method of canonical variates analysis. Several more reviews about GAs have been published; among them, we cite those by Tominaga [12], Shaffer and Small [13], Wehrens and Buydens [14], Luke [15–17], Meusinger and Himmelreich [18], Hibbert [19], Maiocchi [20], Leardi [21–23], and Hou and Xu [24].

The ultimate goal of GAs is the optimization of a given response function. These algorithms are inspired by the theory of evolution: in a living environment, the best individuals have a greater chance to survive and a greater probability to spread their genomes by reproduction. The mating of two "good" individuals causes the mixing of their genomes, which may result in a "better" offspring. The terms "good", "best", and "better" are related to the fitness of the individuals to their environment [25]. GAs have five basic steps: (i) coding of variables; (ii) initiation of population; (iii) evaluation of the response; (iv) reproduction; and (v) mutation. Steps 3–5 alternate until a termination criterion is reached; this criterion can be based on a lack of improvement in the response or simply on a maximum number of generations or on the total time allowed for the elaboration.

Coding of variables: In GAs, each variable corresponds to a gene; an experimental condition (combination of the values of the different variables) corresponds to a chromosome (sequence of genes). There are several ways to code the values of the variables; the simplest one is a direct "translation" of the numerical value to the binary code. Each experimental condition is represented by a chromosome, composed by as many genes as variables, each made by a different number of bits. As a consequence, an experimental condition will be described by the corresponding string of 0s and 1s.

Initiation of population: The original population is composed of a certain number N of chromosomes (generally in the range 20–100, according to the specific problem). After having decided the order of the genes in the chromosomes, for each gene, a sequence of 0s and 1s will be drawn. Therefore, the structure of each chromosome of the initial population is determined in a totally random way.

Evaluation of the response: For each chromosome, the response associated with the corresponding experimental conditions is evaluated. If the experimental condition lies outside the experimental domain or corresponds to an experiment impossible to perform, a null response can be given.

Reproduction: This step creates a new population of N chromosomes that can be considered as the next generation. It can be divided into two substeps: select-copy and cross-over. Select-copy

* *Correspondence to: R. Leardi, Department of Pharmaceutical and Food Chemistry and Technology, Genova University, Via Brigata Salerno (Ponte), I-16147 Genova, Italy.*
E-mail: riclea@dictfa.unige.it

a *A. Niazi*
Department of Chemistry, Islamic Azad University, Arak Branch, Arak, Iran

b *R. Leardi*
Department of Pharmaceutical and Food Chemistry and Technology, Genova University, Via Brigata Salerno (Ponte), I-16147 Genova, Italy

operator for N times randomly selects a chromosome of the population. The probability of a particular chromosome of being selected is a function of its associated response so that the best ones have a greater probability of being picked up than the worst ones. Following this step, a new population is obtained in which the best chromosomes are copied more often; this leads to a better average response. In the cross-over step, the N chromosomes forming the new population are randomly paired to form $N/2$ pairs. From each pair of "parents", two new chromosomes (the "offsprings") will be created by randomly assigning to each of them the genes of one of the two parents. As a result, the cross-over allows the exploration of new experimental conditions by mixing values of variables already tested, although in different combinations.

Mutation: Although the cross-over operator is active at the gene level (whole genes are involved), the mutation takes place at bit level. To do this, for each bit of each chromosome, a random number is drawn to decide whether it has to be affected by a mutation. If so, the bit will be flipped (it will become 0 if it was 1 and vice versa). This operator allows the "jump" to new regions of the experimental domain and avoids the risk of being stuck in some specific conditions (if a gene is the same in all the chromosomes of the population, without mutations the value of the corresponding variable will stay the same forever).

After the reproductions and the mutations, the new generation replaces the previous one and the algorithm continues from the evaluation of the response. Figure 1 shows a flowchart of a GA.

In this paper, the authors will review the applications of GAs in three different areas (optimization; quantitative structure-activity relationship (QSAR) and molecular modeling; multivariate calibration); a list of miscellaneous examples in chemometrics will also be given.

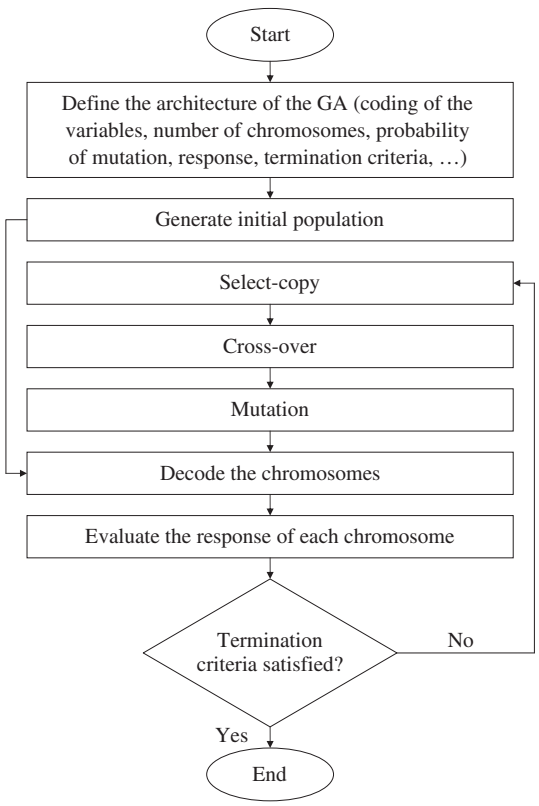

Figure 1. Flow chart of a genetic algorithm (GA).

2. APPLICATION OF GENETIC ALGORITHMS IN OPTIMIZATION

Stochastic optimization techniques such as GAs are gaining increasing popularity in various fields of chemistry, and the number of papers describing successful applications continues to grow at a quick rate [26–29]. These methods are especially beneficial when the search space is complex with many local minima (or maxima) so that conventional techniques fail to find the global minimum (or maximum) and a full search is not feasible. Although it is generally accepted that stochastic methods are the best choice in complex search space, there is no guarantee that they will find the global optimum [29].

Hibbert [30] used GAs to optimize the rate coefficients for the hydrolysis of adenosine 5′-triphosphate by fitting a kinetic model to concentration versus time data. The fastest convergence to a good optimum is achieved by a hybrid GA in which a steepest descent, pseudo-Newton procedure is iterated with an incest-preventing GA, each providing a starting point for the other. In a study by Hartke [31], a GA is used to find the global minimum energy structure for Si_4 on an empirical potential energy surface. Given a suitable encoding of the cluster geometry, and an exponential scaling of the potential energy values to obtain a fitness function, the GA can successfully optimize all degrees of freedom. With the number of potential energy function evaluations as a measure, the GA is more economical than either a set of traditional local minimizations or a molecular dynamics-simulated annealing approach.

Other applications of GAs to optimization are reported in the papers by Weber *et al.* [32], Jiang *et al.* [33], Van Kampen *et al.* [34], Niesse and Mayne [35], Shaffer and Small [13], Lavine *et al.* [36], Hanger and Huttner [37], Smith and Gemperline [38], Kabrede and Hentschke [39], and Chen *et al.* [40].

In 2005, Babic *et al.* [41] reported a method for optimization of a thin layer chromatography separation on the basis the use of GA, and in 2006, Yu *et al.* [42] reported an application of GA to optimize the buffer system of micellar electrokinetic capillary chromatography for separating the active components contained in Chinese medicine. Chedly *et al.* in 2009 [43] used a GA for multiobjective optimization of molded foams characteristics. The effects of injection process parameters on the properties of molded foams are investigated. The input optimization parameters considered are injection temperature, mold temperature, injection speed, plasticization back pressure, and screw rotation speed during the plasticization phase. The output optimization parameters considered are density, shock absorption, and acoustic absorption. Finally, models are used to carry out multiobjective optimization of injected foam characteristics in the presence of a few constraints on decision variables. This optimization is carried out using a very robust technique, Nondominated Sorting Genetic Algorithm II. Several two-objective functions involving sometimes the maximization and other times the minimization of foam characteristics have been studied to illustrate the procedures and explain and interpret the results obtained.

Recently, several papers described applications of GAs in optimization such as Madaeni *et al.* [44], Cano-Odena *et al.* [45], Shi and Xue [46], and Vadood *et al.* [47]. Bhatti *et al.* in 2011 [48] described response surface methodology and artificial neural network (ANN) approach for electrocoagulation of copper from simulated wastewater. Multiobjective optimization for maximizing the copper removal efficiency and minimizing the energy

consumption was carried out using GAs over the ANN model. The optimization procedure resulted in the creation of nondominated optimal points that gave an insight regarding the optimal operating conditions of the process.

Milani and Milani [49] presented a simple closed form equation for the prediction of cross-linking of ethylene propylene diene monomer rubber during accelerated sulfur vulcanization. To estimate numerically the degree of cross-linking, kinetic model constants are evaluated through a simple data fitting, performed on experimental rheometer curves. The fitting procedure is a new one and is achieved using an ad-hoc GA, provided that a few points, strictly required to estimate model unknown constants with sufficient accuracy, are selected from the whole experimental curve. To assess the results obtained with the model proposed, a number of different compounds are analyzed, for which experimental or numerical data are available from the literature. The important cases of moderate and strong reversions are also considered, experiencing a convincing convergence of the analytical model proposed.

3. APPLICATION OF GENETIC ALGORITHMS IN QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP/MOLECULAR MODELING

Quantitative structure-activity relationship and quantitative structure–property relationship (QSPR) studies are essentially applied to chemometrics, pharmacodynamics, pharmacokinetics, toxicity, and so on. A major step in constructing QSAR/QSPR models is finding one or more molecular descriptors. A wide variety of descriptors have been reported to be used in QSAR analysis. Whether by traditional methods or multivariate-based techniques, the success of a modeling study depends also on the selection of variables (molecular descriptors) and on the representation of information. Variables should represent the maximum information in activity variations, and collinearity among them must be kept to a minimum. Among different variable selection strategies, GAs are an interesting, flexible, and widely used alternative [50,51].

In 1998, Hou *et al.* applied a GA to the QSAR research of pyrrolobenzothiazepinones and pyrrolobenzoxazepinones inhibitory activities with non-nucleoside HIV-1 reverse transcriptase [52]. In 1999, Meusinger and Moros [53] determined the influence of the molecular structure of organic compounds on their knocking behavior by using a nonbinary GA. Results obtained by GA were significantly better than those obtained by multiple linear regression (MLR). The molecular structures of 240 potential gasoline components were described by 16 different structural groups. Partial octane numbers were calculated for the structural groups related to the substance classes paraffins, naphthenes, olefins, aromatics, and oxygenates. The sum of the calculated partial octane numbers supplies the octane number of the compound. An MLR, a neural network, and a GA were used for the computations of the connections between the structural groups and the knock ratings. Results obtained by GA were significantly better than those obtained by MLR.

In 1999, Hou *et al.* applied GAs to the structure-activity correlation study of a group of non-nucleoside HIV-1 inhibitors and some cinnamamides [54,55]. In these studies, it has been demonstrated that GAs are very useful in data analysis and that they can be applied as a very powerful technique in QSAR. The authors developed a QSAR program combining a GA with MLR and cross-validation.

Some studies of GAs applied to QSAR/QSPR are reported in the papers by Hoffman *et al.* [56], Ros *et al.* [57], Hemmateennejad *et al.* [58–60], Niculescu [61], Fatemi *et al.* [62], Kompani-Zareh [63], Guo *et al.* [64], Niazi *et al.* [65], and Wang *et al.* [66].

Ghasemi and Ahmadi [67] applied GAs for variable selection in a QSAR study of a series of pure nonionic surfactants containing linear alkyl, cyclic alkyl, and alkeyphenyl ethoxylates. Modeling of cloud point of these compounds as a function of the theoretically derived descriptors was established by MLR and partial least squares (PLS) regression. The results indicate that GA is a very effective variable selection approach for QSPR analysis. The comparison of the two regression methods used showed that PLS has better prediction ability than MLR.

Jalali-Heravi and Kyani [68] applied GA-KPLS (kernel PLS) as a novel nonlinear feature selection method in QSAR study. This technique combines GA as a powerful optimization method with KPLS as a robust nonlinear statistical method for variable selection. This feature selection method is combined with ANN to develop a nonlinear QSAR model for predicting activities of a series of substituted aromatic sulfonamides as carbonic anhydrase II inhibitors. Superiority of this method (GA-KPLS-ANN) over MLR and GA-PLS-ANN (in which a linear feature selection method has been used) indicates that the GA-KPLS approach is a powerful method for the variable selection in nonlinear systems. Gharagheizi [69] reported using GA-based MLR for solubility parameter studies. Recently, several papers have been published by Riahi *et al.* [70], Ghavami *et al.* [71], Goodarzi *et al.* [72], Afiuni-Zadeh and Azimi [73], and Hao *et al.* [74].

4. APPLICATIONS OF GENETIC ALGORITHMS IN MULTIVARIATE CALIBRATION

Multivariate calibration is used to develop a quantitative relationship between the predictor variables in X and the response variable(s) in Y. Recently, multivariate calibration underwent several enhancement/extensions [75,76] that have found widespread use in analytical science. Nowadays, spectral data are perhaps the most common type of data to which chemometric techniques are applied. Owing to the development of new instrumentation, data sets in which each object is described by several hundreds of variables can be easily obtained. Calibration methods, being based on latent variables, allow taking into account the whole spectrum without having to perform a previous feature selection. In the last decades, it has anyway been recognized that an efficient feature selection can be highly beneficial both to improve the predictive ability to the model and to greatly reduce its complexity.

One of the greatest problems in multivariate analysis is to select the combination of variables that produces the best result. This goal is attained through the elimination of those variables that produce noise or that, although giving good information, are strictly correlated with other already selected variables. Feature selection is very important both in studies of correlation and in studies of classification and modeling.

Genetic algorithms have found widespread application in several fields involving multivariate calibration because one of the most important steps in a calibration is the selection of the relevant variables. Leardi *et al.* [77] published one of the very first papers about the application of GAs to variable selection. Lucasius

and Kateman [78] showed that a GA generally performs better that simulated annealing and stepwise regression; on the other hand, Horchner and Kalivas [79] demonstrated that simulated annealing can give the same results. Wise *et al.* [80] also developed a GA for feature selection.

Broudiscou *et al.* [81] described a new technique based on GAs for constructing experimental designs; also, in 1996, Jouan-Rimbaud *et al.* [82] studied the random correlation in variable selection using GA in multivariate calibration. Several papers about the application of GAs in multivariate calibration were published before 2000 [83–90].

In 2001, Liu and Wang [91,92] used a GA for the quantitative analysis of overlapped spectra in Fourier transform infrared spectroscopy (FTIR) data, and Yoshida *et al.* [93] used a GA for feature selection in mass data. Leardi *et al.* [94] used a GA for variable selection for multivariate calibration for predicting concentrations in polymer films in FTIR data, and several researchers [95–115] published papers in which they used GAs for variable selection in different fields such as spectroscopy, electrochemistry, and chromatography.

Goicoechea and Olivieri [95] presented a new method for wavelength interval selection with a GA to improve the predictive ability of PLS calibration. It involves separately labeling each of the selected sensor ranges with an appropriate inclusion ranking. The new approach intends to alleviate overfitting without the need of preparing an independent monitoring sample set. A theoretical example is worked out to compare the performance of the new approach with previous implementations of GAs. Two experimental data sets are also studied: target parameters are the concentration of glucuronic acid in complex mixtures studied by Fourier transform mid-infrared spectroscopy and the octane number in gasolines monitored by near-infrared spectroscopy. Ghasemi *et al.* [98] proposed GAs for selecting wavelengths for PLS calibration using spectrophotometric method. The method is based on the development of the reaction between the analytes and Zincon reagent. A series of synthetic solutions containing different concentrations of copper and zinc were used to check the prediction ability of the GA-PLS models.

Majidi *et al.* [104] used GAs for potential selection in differential pulse voltammetry method in simultaneous determination of cysteine, tyrosine, and tryptophan on the unmodified glassy carbon electrode. The main difficulty in the analysis of these analytes in the same samples is the high degree of overlapping of the voltammograms. The relationships between the currents and the concentrations are complex and highly nonlinear. The predictive ability of principal component regression (PCR), PLS, GA-PLS, and principal component-artificial neural network (PC-ANN) were examined for simultaneous determination of three amino acids. For a regression model, everything that does not help in constructing the model may be considered as noise. PC-ANN and GA-PLS use significant data and show superiority over other applied multivariate methods.

5. MISCELLANEOUS APPLICATIONS

Genetic algorithms were employed in curve fitting [116]. In 1995, Benedetti and Morosetti [117] reported the application of a GA to search for optimal and suboptimal RNA secondary structures. In 1996, Dods *et al.* [118] used a GA approach for fitting polyatomic spectra. Kariuki *et al.* [119] described the development of GAs for solving crystal structures directly from powder diffraction data.

Hervas *et al.* [120] coupled GAs and pruning computational neural networks for the selection of the number of inputs required to correct temperature variations in kinetic-based determinations. Giro *et al.* [121] developed a new methodology to design conducting polymers on the basis of the use of GAs coupled to negative factor counting techniques. The authors showed the results for a case study of polyanilines, one of the most important families of conducting polymers. The methodology proved to be able of generating automatic solutions for the problem of determining the optimum relative concentration for binary and ternary disordered polyaniline alloys exhibiting metallic properties.

Maeder *et al.* [122] reported the application of GAs to the task of determining initial parameter estimates that lie near the global optimum. In iterative nonlinear least squares fitting, the reliable estimation of initial parameters that lead to convergence to the global optimum can be difficult. Irrespective of the algorithm used, poor parameter estimates can lead to abortive divergence or in rare cases convergence to a local optimum. For the determination of the parameters of complex reaction mechanisms, where often little is known about what value these parameters should take, the task of determining good initial estimates can be time consuming and unreliable. In this contribution, the methodology of applying a GA to the task of determining initial parameter estimates that lie near the global optimum is explained. A generalized GA was implemented according to the methodology, and the results of its application are also given. The parameter estimates obtained were then used as the starting parameters for a gradient search method, which quickly converged to the global optimum. The GA was successfully applied to both simulated kinetic measurements where the reaction mechanism contained one equilibrium constant and two rate constants to be fitted and to kinetic measurements of the complexation.

Fatemi *et al.* [123] used GAs in kinetic modeling and reaction mechanism studies. This study is focused on the development of a systematic computational approach that implements GA to find the optimal rigorous kinetic models. This model consists of eight continuous parameters (e.g., Arrhenius and Van't Hoff parameters) and six discrete parameters representing the order of the reaction with respect to each concentration. The optimal values of these parameters have been obtained on the basis of GA. Furthermore, the best type of Genetic operators and their corresponding parameters for this type of problems have been obtained on the basis of a comprehensive study of the effect of these parameters on the efficiency of the GA.

Gianoli *et al.* [124] reported the application of GAs in kinetic modeling, and also, Sadi and Dabir [125] applied GAs for the determination of kinetic parameters of free radical polymerization of vinyl acetate by multiobjective optimization technique. Harris [126] studied applications of GAs for obtaining structure solution from powder X-ray diffraction data, and Guruprasad and Behera [127] applied GAs to textile.

Acknowledgement

Financial support from the Italian Ministry of University and Research (PRIN 2008, CUP:D31J0000020001) is gratefully acknowledged.

REFERENCES

1. Holland JH. Adaptation in Natural and Artificial Systems. The University of Michigan Press, Michigan, 1975.
2. Mitchell M. An introduction to genetic algorithms. The MIT Press, Massachusetts, 1999.
3. Otto M. Chemometrics. Wiely-VCH Verlag GmbH and Co.: Weinheim, 2007.
4. Massart DL, Vandeginste BGM, Buydens LMC, De Long S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics, Part A. Elsevier Science: Amsterdam, 1997.
5. Vandeginste BGM, Massart DL, Buydens LMC, De Long S, Lewi PJ, Smeyers-Verbeke J. Handbook of Chemometrics and Qualimetrics, Part B. Elsevier Science: Amsterdam, 1998.
6. Forrest S. Genetic algorithms: principles of natural selection applied to computation. *Science* 1993; **261**: 872–878.
7. Maddox J. Genetics helping molecular dynamics. *Nature* 1995; **376**: 209.
8. Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 1: concepts, properties and context. *Chemometr. Intell. Lab* 1993; **19**: 1–33.
9. Hibbert DB. Genetic algorithms in chemistry. *Chemometr. Intell. Lab* 1993; **19**: 277–293.
10. Lucasius CB, Beckers MLM, Kateman G. Genetic algorithms in wavelength selection: a comparative study. *Anal. Chim. Acta* 1994; **286**: 135–153.
11. Kemsley EK. A genetic algorithm (GA) approach to the calculation of canonical variates. *Trends Anal. Chem.* 1998; **17**: 24–34.
12. Tominaga Y. Representative subset selection using genetic algorithms. *Chemometr. Intell. Lab* 1998; **43**: 157–163.
13. Shaffer RE, Small GW. Learning optimization from nature: simulated annealing and genetic algorithms. *Anal. Chem.* 1997; **69**: 236A-242A.
14. Wehrens R, Buydens LMC. Evolutionary optimization: a tutorial. *Trends Anal. Chem.* 1997; **17**: 193–203.
15. Luke BT. An overview of genetic methods. In Genetic Algorithms in Molecular Modeling, Devillers J (ed.). Academic Press: New York, 1996; 35–66.
16. Luke BT. Genetic algorithms and beyond. *Data Handl. Sci. Techn.* 2003; **23**: 3–54.
17. Luke BT. Applying genetic algorithms and neural networks to chemometric problems. *Data Handl. Sci. Techn.* 2003; **23**: 343–375.
18. Meusinger R, Himmelreich U. Neural networks and genetic algorithms applications in nuclear magnetic resonance spectroscopy. *Data Handl. Sci. Techn.* 2003; **23**: 281–321.
19. Hibbert DB. Hybrid genetic algorithms. *Data Handl. Sci. Techn.* 2003; **23**: 55–68.
20. Maiocchi A. Genetic algorithms in molecular modeling: a review. *Data Handl. Sci. Techn.* 2003; **23**: 109–139.
21. Leardi R. Genetic algorithms in chemometrics and chemistry: a review. *J. Chemometr.* 2001; **15**: 559–569.
22. Leardi R. Genetic algorithm-PLS as a tool for wavelength selection in spectral data sets. *Data Handl. Sci. Techn.* 2003; **23**: 169–196.
23. Leardi R. Genetic algorithms in chemistry. *J. Chromatogr. A* 2007; **1158**: 226–233.
24. Hou T, Xu X. Applications of genetic algorithms to computer-aided drug design. *Prog. Chem.* 2004; **16**: 35–41.
25. Jouan-Rimbaud D, Massart DL, Leardi R, De Noord OE. Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Anal. Chem.* 1995; **67**: 4295–4301.
26. Clark DE, Westhead DR. Evolutionary algorithms in computer-aided molecular design. *J. Comput. Aid. Mol. Des.* 1996; **10**: 337–358.
27. Devillers J. Genetic Algorithms in Molecular Modeling. Principles of QSAR and Drug Design. Academic Press: New York, 1996.
28. Judson RS. Genetic algorithms and their use in chemistry. In Review in Computational Chemistry, Lipkowitz KB, Boyd DB (eds). VCH Publishers: New York, 1997.
29. Wehrens R, Prestsch E, Buydens LMC. The quality of optimization by genetic algorithms. *Anal. Chim. Acta* 1999; **388**: 265–271.
30. Hibbert DB. Ahybrid genetic algorithm for the estimation of kinetic parameters. *Chemometr. Intell. Lab* 1993; **19**: 319–329.
31. Hartke B. Global geometry optimization of clusters using genetic algorithms. *J. Phys. Chem.* 1993; **97**: 9973–9976.
32. Weber L, Wallbaum S, Broger C, Gubernator K. Optimization of the biological activity of combinatorial compound libraries by a genetic algorithm. *Angew. Chem.* 1995; **34**: 2280–2282.
33. Jian JH, Wang JH, Song XH, Yu RQ. Network training and architecture optimization by a recursive approach and modified genetic algorithm. *J. Chemometr.* 1996; **10**: 253–267.
34. Van Kampen AHC, Buydens LMC, Lucasius, CB, Blommers MJJ. Optimization of metric matrix embedding by genetic algorithms. *J. Biomol.* 1996; **7**: 214–224.
35. Niesse JA, Mayne HR. global optimization of atomic and molecular clusters using the space-fixed modified genetic algorithm method. *J. Comput. Chem.* 1997; **18**: 1233–1244.
36. Lavin BK, Moores A, Helfend LK. Genetic algorithm for pattern recognition analysis of pyrolysis gas chromatographic data. *J. Anal. Appl. Pyrol.* 1999; **50**: 47–62.
37. Hanger J, Huttner G. Optimization and analysis of force field parameters by combination genetic algorithms and neural networks. *J. Comput. Chem.* 1999; **20**: 455–471.
38. Smith BM, Gemperline PJ. Wavelength selection and optimization of pattern recognition methods using the genetic algorithm. *Anal. Chim. Acta* 2000; **423**: 167–177.
39. Kabrede H, Hentschke R. An improved genetic algorithm for global optimization and its application to sodium chloride clusters. *J. Phys. Chem. B* 2002; **106**: 10089–10095.
40. Chen XG, Li X, Kong L, Ni JY, Zhao RH, Zou HF. Application of uniform design and genetic algorithm in optimization of reversed-phase chromatographic separation. *Chemometr. Intell. Lab* 2003; **67**: 157–166.
41. Babic S, Horvat AJM, Kastelan-Macan M. Use of a genetic algorithm to optimize TLC separation. *J. Planar Chromat.* 2005; **18**: 112–117.
42. Yu K, Lin Z, Cheng Y. optimization of the buffer system of micellar electrokinetic capillary chromatography for the separation of the active components in Chinese medicine 'SHUANGDAN' granule by genetic algorithm. *Anal. Chim. Acta* 2006; **562**: 66–72.
43. Chedly S, Chettah A, Ichchou MN. Multiobjective optimization of molded LDPE foams characteristics using genetic algorithm. *J. Appl. Ploym. Sci.* 2009; **114**: 358–368.
44. Madaeni SS, Hasankiadeh NT, Kurdian AR, Rahipour A. Modeling and optimization of membrane fabrication using artificial neural network and genetic algirthm. *Sep. Purif. Technol.* 2010; **76**: 33–43.
45. Cano-Odena A, Spilliers M, Dedroog T, De Grave K, Raman J, Vankelecom IFJ. Optimization of cellulose acetate nanaofilteration membrane for micropollutant removal via genetic algorithms and high throughout experimentation. *J. Membrane Sci.* 2011; **366**: 25–32.
46. Shi J, Xue X. Optimization design of electrodes for anode-supported solid oxide fuel cells via genetic algorithm. *J. Electrochem. Soc.* 2011; **158**: B143-B151.
47. Vadood M, Semnani D, Morshed M. Optimization of acrylic dry spinning production line by using artificial neural network and genetic algorithm. *J. Appl. Polym. Sci.* 2011; **120**: 735–744.
48. Bhatti MS, Kapoor D, Kalia RK, Reddy AS, Thukral AK. RSM and ANN modeling for electrocoagulation of copper from simulated wastewater; multi objective optimization using genetic algorithm approach. *Desalination* 2011; **274**: 74–80.
49. Milani G, Milani F. EPDM accelerated sulfur vulcanization: A kinetic model based on a genetic algorithm. *J. Math. Chem.* 2011; **49**: 1357–1383.
50. Zupan J, Novic M. General type of a uniform and reversible representation of chemical structures. *Anal. Chim. Acta* 1997; **348**: 409–418.
51. Kompani-Zareh M, Mirzaei M. Genetic algorithm-based method for selection conditions in multivariate determination of povidone-iodine using hand scanner. *Anal. Chim. Acta* 2004; **521**: 231–236.
52. Hou TJ, Wang JM, Li YY, Xu XY. Application of genetic algorithm to the QSAR research of pyrrolobenzothiazepinones and pyrrolobenzoxazepinone-novel and specific non-nucleoside HIV-1 reverse transcription inhibitors. *Chin. Chem. Lett.* 1998; **9**: 651–654.
53. Meusinger R, Moros R. Determination of quantitative structure-octane rating relationships of hydrocarbons by genetic algorithms. *Chemometr. Intell. Lab* 1999; **46**: 67–78.
54. Hou TJ, Wang JM, Xu XJ. Applications of genetic algorithms on the structure-activity correlation study of a group of nin-nucleoside HIV-1 inhibitors. *Chemometr. Intell. Lab* 1999; **45**: 303–310.
55. Hou TJ, Wang JM, Liao N, Xu XJ. Applications of genetic algorithms on the structure-activity relationship analysis of some cinnamamides. *J. Chem. Inf. Comp. Sci.* 1999; **39**: 775–781.
56. Hoffman BT, Kopajtic T, Katz JL, Newman AH. 2D QSAR modeling and preliminary database searching for dopamine transporter

inhibitors using genetic algorithm variable selection of Molconn Z descriptors. *J. Med. Chem.* 2000; **43**: 4151–4159.
57. Ros F, Pintore M, Chretien JR. Molecular descriptor selection combining genetic algorithms and fuzzy logic: application to database mining procedure. *Chemometr. Intell. Lab* 2002; **63**: 15–26.
58. Hemmateenejad B, Miri R, Akhond M, Shamsipur M. QSAR study of the calcium channel antagonist activity of some recently synthesized dihydropyridine derivatives: an application of genetic algorithm for variable selection in MLR and PLS methods. *Chemometr. Intell. Lab* 2002; **64**: 91–99.
59. Hemmateenejad B, Akhond M, Miri R, Shamsipur M. Genetic algorithm applied to the selection of factors in principal component-artificial neural networks: application to QSAR study of calcium channel antagonist activity of 1,4-dihydropyridines. *J. Chem. Inf. Comp. Sci.* 2003; **43**: 1328–1334.
60. Hemmateenejad B. Optimal QSAR analysis of the carcinogenic activity of drugs by correlation ranking and gentic algorithm-based PCR. *J. Chemometr.* 2004; **18**: 475–485.
61. Niculescu SP. Artificial neural networks and genetic algorithms in QSAR. *J. Mol. Struct. (THEOCHEM)* 2003; **622**: 71–83.
62. Fatemi MH, Jalali-Heravi M, Konuze E. Prediction of bioconcentration factor using genetic algorithm and artificial neural network. *Anal. Chim. Acta* 2003; **486**: 101–108.
63. Kompani-Zareh M. A QSPR study of boiling point of saturated alcohols using genetic algorithm. *Acta Chim. Slov.* 2003; **50**: 259–273.
64. Guo W, Cai W, Shao X, Pan Z. Application of genetic stochastic resonance algorithm to quantitative structure-activity relationship study. *Chemometr. Intell. Lab* 2005; **75**: 181–188.
65. Niazi A, Jameh-Bozorghi S, Nori-Shargh D. Prediction of acidity constants of thiazolidine-4-carbozylic acid derivatives using Ab initio and genetic algorithm-partial least squares. *Turk. J. Chem.* 2006; **30**: 619–628.
66. Wang J, Krudy G, Xie XQ, Wu C, Holland G. Genetic algorithm-optimized QSPR model for bioavailability, protein binding, and urinary excretion. *J. Chem. Inf. Model.* 2006; **46**: 2674–2683.
67. Ghasemi J, Ahmadi S. Combination of genetic algorithm and partial least squares for cloud point prediction of nonionic surfactants from molecular structures. *Ann. Chim.* 2007; **97**: 69–83.
68. Jalali-Heravi M, Kyani A. Application of genetic algorithm-kernel partial least squares as a novel nonlinear feature selection method: activity of carbonic anhydrase II inhibitors. *Eur. J. Med. Chem.* 2007; **45**: 649–659.
69. Gharagheizi F. QSPR studies for solubility parameter by means of genetic algorithm-based multivariate linear regression and generalized regression neural network. *QSAR Comb. Sci.* 2008; **27**: 165–170.
70. Riahi S, Ganjali MR, Pourbasheer E, Norouzi P. QSPR study of GC retention indices of essential oil compounds by multiple linear regression with a genetic algorithm. *Chromatographia* 2008; **67**: 917–922.
71. Ghavami R, Najafi A, Sajadi M, Djannaty F. Genetic algorithm as variable selection procedure for the simulation of 13 C nuclear magnetic resonance spectra of flavonoid derivatives using multiple linear regression. *J. Mol. Graph. Model.* 2008; **27**: 105–115.
72. Goodarzi M, Freitas MP, Wu CH, Duchowicz PR. pKa modeling and prediction of series of pH indicators through genetic algorithm-least square support vector regression. *Chemometr. Intell. Lab* 2010; **101**: 102–109.
73. Afiuni-Zadeh S, Azimi G. A QSAR for modeling of 8-azaadenine analogues proposed as Al adenosine receptor antagonists using genetic algorithm coupling adaptive neuro-fuzzy inference system. *Anal. Sci.* 2010; **26**: 897–902.
74. Hao, M, Li Y, Wang Y, Zhang S. Prediction of P2Y12 antagonists using a novel genetic algorithm-support vector machine coupled approach. *Anal. Chim. Acta* 2011; **690**: 56–63.
75. Gabrielsson J, Trygg J. Recent developments in multivariate calibration. *Crit. Rev. Anal. Chem.* **36**; 2006: 243–255.
76. Wold S, Trygg J, Berglund A, Antii H. Some recent developments in PLS mg. *Chemometr. Intell. Lab* 2001; **58**: 131–151.
77. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J. Chemometr.* 1992; **6**: 267–281.
78. Lucasius CB, Kateman G. Genetic algorithms for large-scale optimization in chemometrics: an application. *Trends Anal. Chem.* 1991; **10**: 254–261.
79. Horchner U, Kalivas JH. Further investigation on a comparative study on simulated annealing and genetic algorithm for wavelengths selection. *Anal. Chim. Acta* 1995; **311**: 1–13.
80. Wise BM, Gallagher NB, Eschbach PA, Sharpe SW, Griffin JW. Optimization of prediction error using genetic algorithms and continuum regression: determination of the reactivity of automobile emissions from FTIR spectra. *Fourth Scand. Symp. on Chemometrics (SSC4)*, Lund, June 1995.
81. Broudiscou A, Leardi R, Phan-Tan-Luu R. Genetic algorithm as a tool for selection of D-optimal design. *Chemometr. Intell. Lab* 1996; **35**: 105–116.
82. Jouan-Rimbaud D, Massart DL, De Noord OE. Random correlation in variable selection for multivariate calibration with a genetic algorithm. *Chemometr. Intell. Lab* 1996; **35**: 213–220.
83. Brodhurst D, Goodacre R, Jones A, Rowland JJ, Kell DB. Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression. *Anal. Chim. Acta* 1997; **348**: 71–86.
84. Acros MJ, Alonso C, Ortiz MC. Genetic-algorithm-based potential selection in multivariate voltammetric determination of idomethacin and acemethacin by partial least squares. *Electrochim. Acta* 1998; **43**: 479–485.
85. Leardi R, Lupianez Gonzalez A. Genetic algorithm applied to feature selection in PLS regression: How and when to use them. *Chemometr. Intell. Lab* 1998; **41**: 195–207.
86. Ding Q, Small GW, Arnold MA. Genetic algorithm-based wavelength selection for the near-infrared determination of glucose in biological matrixes: initialization strategies and effects of spectral resolution. *Anal. Chem.* 1998; **70**: 4472–4479.
87. Frost VJ, Molt K. Use of genetic algorithm for factor selection in principal component regression. *J. Near Infrared Spec.* 1998; **6**: A185-A190.
88. Wang J, Xian R, Yang B, Wang D, Wang Y, Chen S. Application of genetic algorithm-spectrophotometric method for the multicomponent simultaneous determination of rare earth elements in geological samples. *Fenxi Huazue* 1999; **27**: 955–956.
89. Roger JM, Bellon-Maurel V. Using genetic algorithms to select wavelengths in near-infrared spectra: application to sugar content prediction in cherries. *Appl. Spectrosc.* 2000; **59**: 1313–1320.
90. Leardi R. Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. Chemometr.* 2000; **14**: 643–655.
91. Liu F, Wang JD. Using genetic algorithm for quantitative analysis of overlapped spectra in FTIR spectra. *Spectroscopy Spectral Anal.* 2001; **21**: 609–610.
92. Liu F, Wang JD. Application of a genetic algorithm to quantitative analysis of overlapped FTIR spectra. *Spectrosc. Lett.* 2001; **34**: 13–24.
93. Yoshida H, Leardi R, Funatsu K, Varmuza K. Feature selection by genetic algorithms for mass spectral classifiers. *Anal. Chim. Acta* 2001; **446**: 485–494.
94. Leardi R, Seasholtz MB, Pell RJ. Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data. *Anal. Chim. Acta* 2002; **461**: 189–200.
95. Goicoechea HC, Olivieri AC. Wavelength selection for multivariate calibration using a genetic algorithm: a novel initialization strategy. *J. Chem. Inf. Comp. Sci.* 2002; **45**: 1146–1153.
96. Dieterle F, Kieser B, Gauglitz G. Genetic algorithms and neural networks for quantitative analysis of ternary mixtures using surface plasmon resonance. *Chemometr. Intell. Lab* 2003; **65**: 67–81.
97. Chen K, Li T, Lu P. Application of genetic algorithms in resolution of chromatogram. *Fenxi Huaxue* 2003; **31**: 158–162.
98. Ghasemi J, Niazi A, Leardi R. Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture. *Talanta* 2003; **59**: 311–317.
99. Goicoechea HC, Olivieri AC. A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy. *J. Chemometr.* 2003; **17**: 338–345.
100. Lestander TA, Leardi R, Geladi P. Selection of near infrared wavelengths using genetic algorithms for the determination of seed moisture content. *J. Near Infrared Spec.* 2003; **11**: 433–446.
101. Abdollahi H, Bagheri L. Simultaneous spectrophotometric determination of vitamin K3 and 1,4-naphthoquinone after cloud point extraction by using genetic algorithm based wavelength selection-partial least squares regression. *Anal. Chim. Acta* 2004; **514**: 211–218.
102. Abdollahi H, Bagheri L. Simultaneous spectrophotometric of p-benzoquinone and chloranil after microcrystalline naphthalene

extraction using genetic algorithm-based wavelength selection-partial least squares regression. *Anal. Sci.* 2004; **20**: 1701–1706.
103. Kompani-Zareh M, Farrokhi-Kurd S. Genetic algorithm applied to the selection of conditions for the simultaneous quantification of three-food colorants using a hand scanner. *Microchim. Acta* 2005; **150**: 77–85.
104. Majidi MR, Jouyban A, Asadpour-Zeynali K. Genetic algorithm based potential selection in simultaneous voltammetric determination of isoniazid and hydrazine by using partial least squares and artificial neural networks. *Electroanalysis* 2005; **17**: 915–918.
105. Zinn P, Adaptive multicomponent analysis by genetic algorithms. *J. Chem. Inf. Model.* 2005; **45**: 880–887.
106. Reybes C, De Souza S, Sabatier R, Figueres G, Vidal B. Selection of discriminant wavelength intervals in NIR spectrometry with genetic algorithms. *J. Chemometr.* 2006; **20**: 136–145.
107. Niazi A, Soufi A, Mobarakabadi M. Genetic algorithm applied to selection of wavelength in partial least squares for simultaneous spectrophotometric determination of nitrophenol isomers. *Anal. Lett.* 2006; **39**: 2359–2372.
108. Ghasemi J, Ebrahimi DM, Hejazi L, Leardi R, Niazi A. Simultaneous kinetic-spectrophotometric determination of sulfide and sulfite by partial least squares and genetic algorithms variable selection. *J. Anal. Chem.* 2007; **62**: 348–354.
109. Carneiro RL, Braga JWB, Bottoli CBG, Poppi RJ. Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides and metabolites in wine. *Anal. Chim. Acta* 2007; **595**: 51–58.
110. Tewari JC, Dixit V, Cho BK, Malik KA. Detemination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy. *Spectrochim. Acta A* 2008; **71**: 1119–1127.
111. Fei Q, Li M, Wang B, Huan Y, Feng G, Ren Y. Analysis of cefalexin with NIR spectrometry coupled to artificial neural networks with modified genetic algorithm for wavelength selection. *Chemometr. Intell. Lab* 2009; **97**: 127–131.
112. Zou X, Zhao J, Mao H, Shi J, Yin X, Li Y. Genetic algorithm interval partial least squares regression combined successive projection algorithm for variable selection in near-infrared quantitative analysis of pigment in cucumber leaves. *Appl. Spectrosc.* 2010; **64**: 786–794.
113. Csefalvayova L, Pelikan M, Kralj Cigic I, Kolar J, Strli M. Use of genetic algorithms with multivariate regression for determination of gelatine in historic papers based on FT-IR and NIR spectral data. *Talanta* 2010; **82**: 1784–1790.
114. Arakawa M, Yamashita Y, Funatsu K. Genetic algorithm-based wavelength selection method for spectral calibration. *J. Chemometr.* 2011; **25**: 10–19.
115. De Weijer AP, Lucasius CB, Buydens LMC, Kateman G, Heuvel HM, Mannee H. Curve fitting using natural computation. *Anal. Chem.* 1994; **66**: 23–31.
116. Dane AD, Veldusi A, de Beer DKG, Leenaers AJG, Buydens LMC. Application of genetic algorithms for characterization of thin layer materials by glancing incidence X-ray refractometry. *Physica B* 1998; **253**: 254–268.
117. Benedetti G, Morosetti S. A genetic algorithm to search for optimal and suboptimal RNA secondary structures. *Biophys. Chem.* 1995; **55**: 253–259.
118. Dods J, Gruner D, Brumer P. A genetic algorithm approach to fitting polyatomic spectra via geometry shifts. *Chem. Phys. Lett.* 1996; **261**: 612–619.
119. Kariuki BM, Johnston RL, Harris KDM, Psallidas K, Ahn S, Serrano-Gonzalez H. Application of a genetic algorithm in structure determination from powder diffraction data. *Match* 1998; **38**: 123–135.
120. Hervas C. Algar JA, Silva M. Correction of temperature variations in kinetic-based determinations by use of pruning computational neural networks in conjucation with gentic algorithms. *J. Chem. Inf. Comp. Sci.* 2000; **40**: 724–731.
121. Giro R, Cyrillo M, Galvao DS. Designing conducting polymers using genetic algorithms. *Chem. Phys. Lett.* 2002; **366**: 170–175.
122. Maeder M, Neuhold YM, Puxty G. Applications of a genetic algorithm: near optimal estimation of the rate and equilibrium constants of complex reaction mechanism. *Chemometr. Intell. Lab* 2004; **70**: 193–203.
123. Fatemi S, Masoori M, Bozorgmehry Boozarjomehry R. Application of genetic algorithm in kinetic modeling and reaction mechanism studies. *Iran. J. Chem. Chem. Eng.* 2005; **24**: 37–46.
124. Gianoli SI, Puxty G, Fisher U, Maeder M, Hungerbuchler K. Empirical kinetic modeling of on line simultaneous infrared and calorimetric measurement using a Pareto optimal approach and multi-objective genetic algorithm. *Chemometr. Intell. Lab* 2007; **85**: 47–62.
125. Sadi M, Dabir B. Application of genetic algorithm to determine kinetic parameters of free radical polymerization of vinyl acetate by multi-objective optimization technique. *Iran. J. Chem. Chem. Eng.* 2007; **26**: 29–37.
126. Harris KDM. Fundamentals and applications of genetic algorithms for structure solution from powder X-ray diffraction data. *Comp. Mat. Sci.* 2009; **45**: 16–20.
127. Guruprasad R, Behera BK. Genetic algorithms and its application to textile. *Textile Asia* 2009; **40**: 35–38.