JOURNAL OF CHEMOMETRICS
J. Chemometrics 2001; **15**: 559–569
DOI: 10.1002/cem.651

REVIEW ARTICLE

Genetic algorithms in chemometrics and chemistry: a review

Riccardo Leardi*

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

SUMMARY

The use of genetic algorithms has been growing exponentially since Holland published the first papers about them. Thanks to the extraordinary increase in calculation power, nowadays it is possible to apply them to extremely complex problems. A considerable number of papers in which genetic algorithms have been applied have been published in several scientifical journals. This review is of course far from being a complete summary of the application of genetic algorithms to chemical problems; its goal is to show the reader the main fields of application of this technique, together with providing a list of references on the subject. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: genetic algorithms; optimization; molecular modelling; regression; curve fitting

1. INTRODUCTION

Since their first presentation by Holland in 1975, genetic algorithms (GAs) have attracted a lot of interest. The goal (trying to simulate the evolutionary process of a living species) and jargon (using typical biological terms such as 'gene', 'chromosome', 'mutation' and 'crossover' in the description of an algorithm) of GAs have helped to create something like an aura of mystery around them.

In that period the main limitation to the real development of GAs in terms of applicability was the fact that the huge amounts of computation required by them could not be handled satisfactorily by the computers then available. For almost 20 years this has been the main problem for those who would have liked to apply them to their problems but did not have the possibility of accessing a suitable computer: for 'common size' problems a mainframe would have been required, while for complex problems the computation time would have been too long even with the most powerful computers.

Since the beginning of the 1990s this major problem has been progressively removed, and nowadays every personal computer can be used to apply GAs to easy/moderate-scale problems, while the mainframes allow one to tackle very complex problems such as those typical of molecular modelling.

This is the reason why, after a first period in which the interest of the scientific community was focused mainly on the theory itself, the number of papers reporting applications of GAs to real

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

Received 3 July 2000
Accepted 9 August 2000

problems and the number of scientists and of disciplines using them have been growing exponentially.

In 1993 the journal *Science* [1] published a paper that gave a general presentation of genetic algorithms, some mathematical analysis about how GAs work and how best to use them, and some applications in modelling several natural evolutionary systems, including immune systems.

In 1995 an article in *Nature* [2] described a problem of molecular dynamics that had been successfully solved by a GA where conventional techniques had failed.

Several tutorials about GAs have been published in journals devoted to different research fields. As examples we cite those by Lucasius and Kateman [3–6], Hibbert [7], Shaffer and Small [8], Wehrens and Buydens [9] and Luke [10].

The set-up of the structure of a GA is a very critical point, and a guide leading to a good architecture is highly beneficial. Wehrens *et al*. [11,12] proposed a set of quality criteria to evaluate the performance of a GA, considering not only the best solutions suggested by the algorithm, but also the repeatability of the optimization and the coverage of the search space.

2. APPLICATION OF GAs IN MOLECULAR MODELLING

It is well recognized that the GA method is especially useful when the response hypersurface in which the optimum is searched is of high dimension and has many local optima. The high dimensionality of the problem makes the application of an exhaustive search impossible, while the presence of local optima makes direct optimization methods (e.g. steepest ascent) unreliable, since they can be caught by a local optimum and give a solution far away from the global optimum.

These characteristics make GAs particularly suitable for molecular modelling, since the energetic hypersurface is very complex, with several local minima. Therefore with the 'standard' methods the optimized structure can depend heavily on the chosen starting point.

In the case of a medium-size protein (100 residues), if there are no constraints, the number of conformations can be approximated to (5 torsion angles per residue $\times$ 5 likely values per torsion angle)$^{100} = 25^{100}$. In such a large search space, clearly beyond the capacity of any supercomputer, the GA is able to find suboptimal solutions. This means that, although it cannot be proven that the GA has really found the optimal solution, some of the results obtained by the GA practically surpass any previously known solution. This can be of much help in non-polynomial complete problems where no analytical solution of the problem is available, as shown by Schulze-Kremer [13].

In 1992, Blommers *et al*. [14] applied a GA to derive the solution structure of the photodimer *cis. syn*-dUp[]dT.

In 1993, McGarrah *et al*. [15] studied the behaviour of the GA method as a global search technique for finding low-energy conformations. They evaluated different hybrid methods that combine coarse GA global search with local gradient minimization, and the effects of the GA parameters on the final results. In the same journal issue, Judson *et al*. [16] used a GA search procedure for finding low-energy conformations of small/medium organic molecules (1–12 rotatable bonds) and compared the performances against the CSEARCH algorithm in Sybyl. They demonstrated that for molecules with more than eight rotatable bonds the GA method is more efficient computationally, and as the number of rotatable bonds increases, the relative efficiency of the GA method grows.

More recently, Jin *et al*. [17,18] efficiently elucidated backbone conformational features observed in the global minimum energy structure of a pentapeptide and searched the conformation space of [Met]-enkephalin. GAs had already been applied to the same molecule, by Lee *et al*. [19].

Other applications of GAs to conformational searching are reported in the papers by Fontain [20], Mestres and Scuseria [21], Hermann and Suhai [22], Meza *et al*. [23], Pullan [24], Niesse and Magne

[25], Wang *et al*. [26], Hartke [27] and Kariuki *et al*. [28], while a general description of the application of GAs to molecular modelling is given by Brodmeier and Pretsch [29].

Another field in which GAs are being applied is computer-aided molecular design (CAMD), with the goal of designing molecules with desired properties and activities, for which the traditional approach often requires a trial-and-error procedure involving a combinatorially large number of potential candidate molecules. A presentation of a GA in CAMD is given by Venkatasubramanian and Sundaram [30], while Venkatasubramanian *et al*. [31,32] apply a GA to polymer design. Other applications are reviewed by Devillers [33].

Sundaram and Venkatasubramanian [34] show very effectively that, since the performance of a GA in CAMD is highly dependent on the parameter setting, an optimal or desirable setting can be defined only on a run-to-run basis.

In CAMD the property prediction based on the structural characteristics of the molecular subunits is defined as the forward problem, while the inverse problem is the construction of a molecular structure given a set of desired macroscopic properties. De Weijer *et al*. [35] solve the forward problem by a neural network-based approach, while the application of a GA was found to be very effective in the inverse problem, i.e. in finding hypothetical physical structures of polyethylene terephthalate yarns corresponding to a certain combination of mechanical and shrinkage properties.

The same hybrid system of a backpropagation neural network and a GA was applied by Venkatasubramanian *et al*. [36] for locating optimal targets for the design of fairly complex polymer molecules, by Devillers and Putavy [37] for designing organic molecules presenting a specific biodegradability, by Burden *et al*. [38] for finding more active dihydrofolate reductase inhibitors, and by Sundaram *et al*. [39] for designing fuel additives. Hopfinger and Patel [40] describe the application of GAs to establish reliable quantitative structure–activity relationships (QSARs) and to a molecular diversity experiment, and Meusinger and Moros [41] apply GAs for the determination of quantitative structure–octane rating relationships of hydrocarbons.

Putavy *et al*. [42] use a classical GA for the selection of a set of aromatic substituents highly representative of a physicochemical parameter space.

Another complex problem is the obtention of atomic-level models of receptor sites. Walters and Muhammad [43] used a GA to alter and optimize the atom types of the receptor site in such a way as to maximize the correlation between calculated drug–receptor binding and measured drug activity.

Jones *et al*. [44] report three applications of GAs in chemical structure handling and molecular recognition. In the first one a GA was found to be highly efficient in searching the conformational space of small three-dimensional molecules for pharmacophoric patterns, by searching substructures in databases of three-dimensional compounds. In the second one a GA for docking flexible ligands into partially flexible protein sites has been developed, while in the third one a GA has been used to superimpose flexible molecules automatically.

A program for three-dimensional intermolecular interactions of molecular clusters or docking problems has been developed by Xiao and Williams [45], and further applications of GAs to docking problems are reported in the papers by Clark and Ajay [46], Judson *et al*. [47], Vieth *et al*. [48], Morris *et al*. [49] and Wang *et al*. [50].

A good tutorial about the use of GAs in the problem of protein folding, by Schulze-Kremer, can be found on the net [51], while among the papers reporting applications in this field, those by Ebeling and Nadler [52] and Krasnogor *et al*. [53,54] can be cited.

Tufféry *et al*. [55] compared the results obtained by different search algorithms (simulated annealing, simple and modified GAs and a heuristic combinatorial approach) in optimizing the conformations of protein side-chains.

According to Van Kampen and Buydens [56], in the structure elucidation of a heptapeptide in torsion angle space the recombination is not always effective, since the crossover is unable to

recombine the building blocks that should produce improved trial solutions. Therefore in such a problem the GA is basically based on selection and point mutation and is outperformed by the more sophisticated mutation-selection scheme imposed by simulated annealing, the latter converging about three times faster.

A combination of GA and neural network (NN) has been used by Hunger and Huttner [57] in the optimization and analysis of force field parameters applied to *tripod* metal templates. Significant time savings have been obtained by partly substituting the evaluation of the fitness function computed from structures calculated by force field methods with results obtained by a trained neural network.

Reijmers *et al*. [58,59] used GAs for the construction of phylogenetic trees of G-protein coupled receptor sequences.

A comprehensive and updated list of references on 'Evolutionary algorithms in computer-aided molecular design' has been compiled by Clark [60].

3. APPLICATION OF GAs IN REGRESSION

GAs have found widespread application in several fields involving regression problems.

One of the most important steps in a calibration is the selection of the relevant variables. The size of the search domain (with v variables, $2^v - 1$ combinations are possible) and the presence of many local optima make the GA one of the suggested methods.

It is interesting to notice that several authors have published papers about feature selection by GAs, each of them using a different GA structure, sometimes rather far from the 'standard' algorithm. This demonstrates the need to modify the algorithm according to the peculiarities of the problem to be solved. In the case of feature selection, for instance, a chromosome is made by a very high number of genes (as many as the variables), each of them being just 1 bit long (0 = variable absent, 1 = variable present). Leardi *et al*. [61] use a simulated data set to show that a GA can always find the global maximum of a simple problem, in a time much shorter than the time required by a full search. Lucasius *et al*. [62] showed that a GA generally performs better than simulated annealing and stepwise regression; on the other hand, Hörchner and Kalivas [63] demonstrated that simulated annealing can give the same results.

Wise *et al*. [64,65] also developed GAs for feature selection.

Wallet *et al*. [66] solve the problem of selecting a minimal model which correctly predicts the response by applying a GA using a two-criteria population management scheme.

Broadhurst *et al*. [67] applied GAs to pyrolysis mass spectrometry, with the goal of determining the optimal subset of variables to give the best possible prediction or determining the optimal subset of variables to produce a model with a predictive ability higher than or equal to a given value (both in MLR and in PLS models).

The method proposed by Bangalore *et al*. [68] leads to the selection of wavelengths and to the definition of the PLS model size. To do that, a 'model size' gene taking on the integer value corresponding to the number of latent variables to be used in building the calibration model is added to the genes coding the presence or absence of each variable in the model.

Jouan-Rimbaud *et al*. [69] successfully applied a GA to the problem of wavelength selection for MLR calibration, while Arcos *et al*. [70] obtained a set of wavelengths able to perform a PLS calibration of mixtures of indomethacin and acemethacin, in spite of the fact that the two compounds have almost identical spectra.

A more complex optimization was performed by Shaffer and Small [71]. They apply a GA to the NIR analysis of glucose in biological matrices, optimizing at the same time five important variables: the position and width of the bandpass filter, the starting and ending points of the spectral range

submitted to the PLS regression, and the number of latent variables employed in the calibration model.

In spectroscopic infrared imaging applied to discriminate between different materials, the selection of a limited number of spectroscopic wavelengths guaranteeing the optimal discrimination makes the acquisition and processing time much faster. This goal has been achieved using a GA by van den Broek *et al*. [72].

Depczynski *et al*. [73] devised a method for multicomponent analysis by near-infrared spectrometry by combining wavelet coefficient regression with a GA.

Although spectral data sets are the most common field of application of GAs for feature selection, owing to the very large search domain, also in the case of non-spectral variables some good results can be obtained, as reported by Aishima *et al*. [74].

Overfitting is the greatest risk when applying GAs to feature selection. This aspect has been taken into account especially by Leardi [75,76], Leardi and Lupiáñez González [77] and Jouan-Rimbaud *et al*. [78].

A different use of GAs in regression is shown by Rogers [79]. In its genetic function approximation the GA selects a whole model whose 'building blocks', called basis functions, are the original features or any possible transformation as sines, square terms or rectangular terms. The same approach is followed by Ozdemir *et al*. [80] for the calibration of multiple instruments, showing that genetic regression is able to produce acceptable models using fewer spectra from the secondary instrument than PLS. The same group also used genetic regression to correct for wavelength drift in single- and multi-instrument calibration [81] and to obtain a multi-instrument calibration in UV-vis spectroscopy [82].

Vankeerberghen *et al*. [83] applied a GA to the problem of robust regression and outlier detection.

Several applications can be found in the study of Quantitative Structure–Activity Relationships (QSAR). Dunn and Rogers [84] combine PLS with the model-generating ability of a GA to create genetic partial least squares, while a hybrid method combining GAs and neural networks was used by Van Helden *et al*. [85]. Li *et al*. [86] used non-linear PLS combined with a GA in a study of the fungicidal activity of a series of *O*-etil-*N*-isopropylphosphoro(thioureido)thioates, obtaining results better than those of the reference, while Hou *et al*. [87] developed a QSAR program combining a GA with multiple linear regression and cross-validation.

4. MISCELLANEOUS

GAs can be successfully applied in curve fitting, as shown by De Weijer *et al*. [88]. They have a good search accuracy (they approach the globally optimal solution irrespective of the diverse starting conditions) and a poor search precision (they often fail in finding the very optimum solution), while the traditional local optimization techniques have a good search precision and a poor search accuracy. Therefore better overall performances can be attained with hybrid algorithms in which, for instance, the starting point for a local search method is given by the best GA solution. In the hybrid algorithm used by Hibbert [89], the GA is iterated with a steepest descent procedure, each providing a starting point for the other.

Dane *et al*. [90] applied GAs to the curve fitting required by glancing incidence X-ray reflectometry, obtaining results better than those obtained with a currently used method, reducing the amount of human effort and expertise required for analysing reflectivity measurements, and reducing the probability of overlooking feasible solutions. The same group [91] implemented a two-step fundamental parameter method for model-free analysis of thin-layered materials by X-ray fluorescence spectrometry, in which a gradient technique is used to refine the results of a GA used

to obtain the number of layers and, for each layer, an estimate of the elementary concentrations and thickness.

Wise *et al*. [92] applied a GA to dynamic model identification. Comparing the results with those obtained by artificial neural networks and non-linear biased regression, they found that the GA gave the best compromise between performance and model complexity.

In the field of experimental design a GA has been successfully applied by Broudiscou *et al*. [93] to the problem of finding a good-quality experimental matrix when dealing with several qualitative variables, each having several levels. In a problem with six variables having seven, six, six, five, three and three levels respectively (11 340 possible points), the Fedorov exchange algorithm is hardly applicable. The GA allowed the obtention of a very good experimental matrix containing just 28 experiments, while with the classical methods the lowest number of experiments to be performed would have been 49.

The simultaneous optimization of several parameters is always a quite challenging task, becoming very difficult when the response surfaces are characterized by several local maxima. Wienke *et al*. combined the target vector criterion with a GA to obtain the simultaneous optimization of the intensities of six atomic emission lines of trace elements in alumina powder as a function of spectroscopic excitation conditions [94] and to improve simultaneously six properties of a biochemical strip for human blood glucose determination as a function of 12 chemical and technological parameters [95].

Instead of a feature selection, sometimes an object selection can be required, with the goal of obtaining a subsample fully representative of the whole sample. Pizarro Millán *et al*. [96] and Tominaga [97] obtained interesting results when applying a GA.

In clustering, GAs have been applied by Lucasius *et al*. [98] and Jiang *et al*. [99].

In classification, van Kampen *et al*. [100] developed a classifier system based on GA methodology for the automatic extraction of classification rules from a database of about 6000 ion chromatography method examples.

Tominaga [101] compared the results of seven classification methods on a very large QSAR data set (three categories, 156 variables, 12 242 objects in the training set, 960 objects in the validation set); the best results in prediction were obtained by the *k*-nearest-neighbour method, followed by a combination of GA and *k*-nearest-neighbour.

Shaffer and Small [102] applied a GA to the optimization of piecewise linear discriminants for the automated classification of Fourier-transformed infrared remote-sensing data.

Chen *et al*. [103] developed a GA for the optimization of non-linear mapping; their algorithm introduced oriented mutation and immigration in order to enhance the efficiency of the search process and reduce the risk of premature convergence.

De Gracia *et al*. [104] used a neural network and a GA jointly to model and optimize a sequential injection flow system of analysis for colorimetric iron(III) determination in water samples. The system was efficiently optimized in terms of sensitivity, linearity and sampling rate.

GAs have also been applied in combinatorial chemistry. Weber *et al*. [105] found highly biologically active compounds from a library of 160 000 possible combinations (Ugi-type reactions of an isocyanide, an aldehyde, an amine and a carboxylic acid to be chosen among 10, 40, 10 and 40 respectively) by performing just 400 reactions.

Hartnett *et al*. [106] developed a GA–simplex hybrid approach for the determination of stability constants using calorimetric and polarographic data obtained from literature sources. The GA determined both the most suitable equilibrium model for the system studied and the values of the stability constants and the heat of formation for the calorimetric studies.

Parczewski *et al*. [107] applied a GA to the determination of physicochemical parameters and concentrations of analytes from titration data. They concluded that GAs are very useful, especially in

finding the starting values of parameters for further refinement by an optimization method of more local scope, such as the simplex method.

5. CONCLUSIONS

The ever-increasing number of papers in which GAs are applied to very different fields of chemistry and chemometrics shows the effectiveness and validity of this technique. The advantage of GAs over the 'classical' techniques becomes greater the greater the complexity of the problem, especially owing to the good balance between exploration and exploitation. The results obtained by a GA can be greatly improved after 'hybridization' with a standard technique, typically having a very poor exploration ability and a very high exploitation potential. This allows one to define with great precision the global optimum, whose location has been effectively found by the GA. It is anyway to be highlighted that it is not possible to define a 'best' GA architecture to be used in all applications, since the optimal structure is extremely problem-dependent.

REFERENCES

1. Forrest S. Genetic algorithms: principles of natural selection applied to computation. *Science* 1993; **261**: 872–878.
2. Maddox J. Genetics helping molecular dynamics. *Nature* 1995; **376**: 209.
3. Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 1. Concepts, properties and context. *Chemometrics Intell. Lab. Syst.* 1993; **19**: 1–33.
4. Lucasius CB, Kateman G. Understanding and using genetic algorithms. Part 2. Representation, configuration and hybridization. *Chemometrics Intell. Lab. Syst.* 1994; **25**: 99–145.
5. Lucasius CB, Kateman G. Gates towards evolutionary large-scale optimization: a software-oriented approach to genetic algorithms—I. General perspective. *Comput. Chem.* 1994; **18**: 127–136.
6. Lucasius CB, Kateman G. Gates towards evolutionary large-scale optimization: a software-oriented approach to genetic algorithms—II. Toolbox description. *Comput. Chem.* 1994; **18**: 137–156.
7. Hibbert DB. Genetic algorithms in chemistry. *Chemometrics Intell. Lab. Syst.* 1993; **19**: 277–293.
8. Shaffer RE, Small GW. Learning optimization from nature: simulated annealing and genetic algorithms. *Anal. Chem.* 1997; **69**: 236A–242A.
9. Wehrens R, Buydens LMC. Evolutionary optimization: a tutorial. *Trends Anal. Chem.* 1998; **17**: 193–203.
10. Luke BT. An overview of genetic methods. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 35–66.
11. Wehrens R, Pretsch E, Buydens LMC. Quality criteria of genetic algorithms for structure optimization. *J. Chem. Info. Comput. Sci.* 1998; **38**: 151–157.
12. Wehrens R, Pretsch E, Buydens LMC. The quality of optimization by genetic algorithm. *Anal. Chim. Acta* 1999; **388**: 265–271.
13. Schulze-Kremer S. Genetic algorithms for protein tertiary structure prediction. In *Parallel Problem Solving from Nature II*, Maenner R, Manderick B (eds). North-Holland: Amsterdam, 1992; 391–400.
14. Blommers MJJ, Lucasius CB, Kateman G, Kaptein R. Conformational analysis of a dinucleotide photodimer with the aid of the genetic algorithm. *Biopolymers* 1992; **32**: 45–52.
15. McGarrah DB, Judson RS. An analysis of the genetic algorithm method of molecular conformation determination. *J. Comput. Chem.* 1993; **14**: 1385–1395.
16. Judson RS, Jaeger EP, Treasurywala AM, Peterson ML. Conformational searching methods for small molecules. II. Genetic algorithm approach. *J. Comput. Chem.* 1993; **14**: 1407–1414.
17. Jin AY, Leung FY, Weaver DF. Development of a novel genetic algorithm search method (GAP 1·0) for exploring peptide conformational space. *J. Comput. Chem.* 1997; **18**: 1971–1984.
18. Jin AY, Leung FY, Weaver DF. Three variations of genetic algorithm for searching biomolecular conformation space: comparison of GAP 1·0, 2·0 and 3·0. *J. Comput. Chem.* 1999; **20**: 1329–1342.
19. Lee J, Scheraga H, Rackovsky S. New optimization method for conformational energy calculation. Polypeptides: conformational space annealing. *J. Comput. Chem.* 1997; **18**: 1222–1232.
20. Fontain E. The problem of atom-to-atom mapping. An application of genetic algorithms. *Anal. Chim. Acta* 1992; **265**: 227–232.

21. Mestres J, Scuseria GE. Genetic algorithms: a robust scheme for geometry optimizations and global minimum structure problems. *J. Comput. Chem.* 1995; **16**: 729–742.
22. Hermann F, Suhai S. Energy minimization of peptide analogues using genetic algorithms. *J. Comput. Chem.* 1995; **16**: 1434–1444.
23. Meza JC, Judson RS, Faulkner TR, Treasurywala AM. A comparison of direct search method and a genetic algorithm for conformational searching. *J. Comput. Chem.* 1996; **17**: 1142–1151.
24. Pullan WJ. Energy minimization of mixed argon–xenon microclusters using a genetic algorithm. *J. Comput. Chem.* 1997; **18**: 1096–1111.
25. 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.
26. Wang J, Hou T, Chen L, Xu X. Conformational analysis of peptides using Monte Carlo simulations combined with the genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 347–351.
27. Hartke B. Global cluster geometry optimization by a phenotype algorithm with niches: location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.* 1999; **20**: 1752–1759.
28. Kariuki BM, Calcagno P, Harris KDM, Philp D, Johnston R. Evolving opportunities in structure solution from powder diffraction data—crystal structure determination of a molecular system with twelve variable torsion angles. *Angew. Chem. Int. Ed.* 1999; **38**: 831–834.
29. Brodmeier T, Pretsch E. Application of genetic algorithms in molecular modeling. *J. Comput. Chem.* 1994; **15**: 588–595.
30. Venkatasubramanian V, Sundaram A. Genetic algorithms: introduction and applications. In *Encyclopedia of Computational Chemistry*. Wiley: New York, 1998; 1115–1127.
31. Venkatasubramanian V, Chan K, Caruthers JM. Computer-aided molecular design using genetic algorithms. *Comput. Chem. Engng* 1994; **18**: 833–844.
32. Venkatasubramanian V, Chan K, Caruthers JM. On the performance of genetic search for large-scale molecular design. *Proc. PSE'94*, 1994; 1001–1006.
33. Devillers J. Genetic algorithms in computer-aided molecular design. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 1–34.
34. Sundaram A, Venkatasubramanian V. Parametric sensitivity and search-space characterization studies of genetic algorithms for computer-aided polymer design. *J. Chem. Info. Comput. Sci.* 1998; **38**: 1177–1191.
35. De Weijer AP, Lucasius CB, Buydens LMC, Kateman G. Using genetic algorithms for an artificial neural network model inversion. *Chemometrics Intell. Lab. Syst.* 1993; **20**: 45–55.
36. Venkatasubramanian V, Sundaram A, Chan K, Caruthers JM. Computer-aided molecular design using neural networks and genetic algorithms. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 271–302.
37. Devillers J, Putavy C. Designing biodegradable molecules from the combined use of a backpropagation neural network and a genetic algorithm. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 303–314.
38. Burden FR, Rosewarne BS, Winkler DA. Predicting maximum bioactivity by effective inversion of neural networks using genetic algorithms. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 127–137.
39. Sundaram A, Ghosh P, Venkatasubramanian V, Caruthers J, Daly DT. Design of fuel-additives using hybrid neural networks and evolutionary algorithms. *Proc. Int. Conf. on Foundations of Computer-Aided Process Design, FOCAPD'99*, Snowmass, CO, July 1999; 478–481.
40. Hopfinger AJ, Patel HC. Application of genetic algorithms to the general QSAR problem and to guiding molecular diversity experiments. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 131–157.
41. Meusinger R, Moros R. Determination of quantitative structure–octane rating relationships of hydrocarbons by genetic algorithms. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 67–78.
42. Putavy C, Devillers J, Domine D. Genetic selection of aromatic substituents for designing test series. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 243–269.
43. Walters DE, Muhammad TD. Genetically evolved receptor models (GERM): a procedure for construction of atomic-level receptor site models in the absence of a receptor crystal structure. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 193–210.
44. Jones G, Willett P, Glen RC. Genetic algorithms for chemical structure handling and molecular recognition. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 211–242.
45. Xiao YL, Williams DE. GAME: genetic algorithm for minimization of energy, an interactive program for three-dimensional intermolecular interactions. *Comput. Chem.* 1994; **18**: 199–201.

46. Clark KP, Ajay. Flexible ligand docking without parameter adjustment across four ligand–receptors complexes. *J. Comput. Chem.* 1995; **16**: 1210–1226.
47. Judson RS, Tan YT, Mori E, Melius C, Jaeger EP, Traesurywala AM, Mathiowetz A. Docking flexible molecules: a case study of three proteins. *J. Comput. Chem.* 1995; **16**: 1405–1419.
48. Vieth M, Hirst JD, Dominy BN, Daigler H, Brooks III CL. Assessing search strategies for flexible docking. *J. Comput. Chem.* 1998; **19**: 1623–1631.
49. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 1998; **19**: 1639–1662.
50. Wang J, Hou T, Chen L, Xu X. Automated docking of peptides and proteins by genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 281–286.
51. Schulze-Kremer S. Genetic algorithms and protein folding. Available: http://www.techfak.uni-bielefeld.de/bcd/Curric/ProtEn/ [28 June 2000].
52. Ebeling M, Nadler W. Protein folding: optimized sequences obtained by simulated breeding in a minimalist model. *Biopolymers* 1997; **41**: 165–180.
53. Krasnogor N, Marcos DH, Pelta D, Risi WA. Protein structure prediction as a complex adaptive system. *Int. Workshop on Frontiers in Evolutionary Algorithms (FEA'98), Fourth Joint Conf. on Information Sciences (JCIS'98) Proc*., vol. 2, 1998; 441.
54. Krasnogor N, Hart WE, Smith J, Pelta DA. Protein structure prediction with evolutionary algorithms. In *GECCO99: Proceedings of the Genetic and Evolutionary Computation Conference*, Banzhaf W, Daida J, Eiben AE, Garzon MH, Honovar V, Jakaiela M, Smith RE (eds). Morgan Kaufman: 1999; 1596–1601.
55. Tufféry P, Etchebest C, Hazout S, Lavery R. A critical comparison of search algorithms applied to the optimization of protein side-chain conformations. *J. Comput. Chem.* 1993; **14**: 790–798.
56. Van Kampen AHC, Buydens LMC. The ineffectiveness of recombination in a genetic algorithm for the structure elucidation of a heptapeptide in torsion angle space. A comparison to simulated annealing. *Chemometrics Intell. Lab. Syst.* 1997; **36**: 141–152.
57. Hunger J, Huttner G. Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks. *J. Comput. Chem.* 1999; **20**: 455–471.
58. Reijmers TH, Wehrens R, Daeyaert FD, Lewi PJ, Buydens LMC. Using genetic algorithms for the construction of phylogenetic trees: application to G-protein coupled receptor sequences. *BioSystems* 1999; **49**: 31–43.
59. Reijmers TH, Wehrens R, Buydens LMC. Quality criteria of genetic algorithms for construction of phylogenetic trees. *J. Comput. Chem.* 1999; **20**: 867–876.
60. Clark DA. Evolutionary algorithms in computer-aided molecular design. Available: http://panizzi.shef.ac.uk/cisrg/links/ea_bib.html [28 June 2000].
61. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J. Chemometrics* 1992; **6**: 267–281.
62. Lucasius CB, Beckers MLM, Kateman G. Genetic algorithms in wavelengths selection: a comparative study. *Anal. Chim. Acta* 1994; **286**: 135–153.
63. Hörchner U, Kalivas JH. Further investigation on a comparative study of simulated annealing and genetic algorithm for wavelengths selection. *Anal. Chim. Acta* 1995; **311**: 1–13.
64. 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.
65. Wise MB, Gallagher NB, Eschbach PA. Application of a genetic algorithm to variable selection for PLS models. *EUCHEM Conf*., Gothenburg, June 1996.
66. Wallet BC, Marchette DJ, Solka JL, Wegman EJ. A genetic algorithm for best subset selection in linear regression. *Computer Science and Statistics*, 1997; **28**: 545–550.
67. Broadhurst 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, with applications to pyrolysis mass spectrometry. *Anal. Chim. Acta* 1997; **348**: 71–86.
68. Bangalore AS, Shaffer RE, Small GW. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. *Anal. Chem.* 1996; **68**: 4200–4212.
69. Jouan-Rimbaud D, Massart DL, Leardi R, De Noord O. Genetic algorithms as a tool for wavelength selection in multivariate calibration. *Anal. Chem.* 1995; **67**: 4295–4301.
70. Arcos MJ, Ortiz MC, Villahoz B, Sarabia LA. Genetic-algorithm-based wavelengths selection in

multicomponent spectrometric determinations by PLS: application on indomethacin and acemethacin mixture. *Anal. Chim. Acta* 1997; **339**: 63–77.
71. Shaffer RE, Small GW, Arnold MA. Genetic algorithm-based protocol for coupling digital filtering and partial least-squares regression: application to the near-infrared analysis of glucose in biological matrices. *Anal. Chem.* 1996; **68**: 2663–2675.
72. Van den Broeck WHAM, Wienke D, Melssen WJ, Buydens LMC. Optimal wavelength range selection by a genetic algorithm for discrimination purposes in spectroscopic infrared imaging. *Appl. Spectrosc.* 1997; **51**: 1210–1217.
73. Depczynski U, Jetter K, Molt K, Niemöller A. Quantitative analysis of near infrared spectra by wavelet coefficient regression using a genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1999; **47**: 179–187.
74. Aishima T, Togari N, Leardi R. Selection of aroma components to predict sensory quality of Kenyan black teas using genetic algorithms for multiple regression models. *Food Sci. Technol. Int.* 1996; **2**: 124–126.
75. Leardi R. Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection. *J. Chemometrics* 1994; **8**: 65–79.
76. Leardi R. Genetic algorithms in feature selection. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 67–86.
77. Leardi R, Lupiáñez González A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometrics Intell. Lab. Syst.* 1998; **41**: 195–207.
78. Jouan-Rimbaud D, Massart DL, de Noord OE. Random correlation in variable selection for multivariate calibration with a genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1996; **35**: 213–220.
79. Rogers D. Some theory and examples of genetic function approximation with comparison to evolutionary techniques. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 87–107.
80. Ozdemir D, Mosley M, Williams R. Hybrid calibration models: an alternative to calibration transfer. *Appl. Spectrosc.* 1998; **52**: 599–603.
81. Ozdemir D, Mosley M, Williams R. Effect of wavelength drift on single- and multi-instrument calibration using genetic regression. *Appl. Spectrosc.* 1998; **52**: 1203–1209.
82. Ozdemir D, Williams R. Multi-instrument calibration with genetic regression in UV-visible spectroscopy. *Appl. Spectrosc.* 1999; **53**: 210–217.
83. Vankeerberghen P, Smeyers-Verbeke J, Leardi R, Carr CL, Massart DL. Robust regression and outlier detection for non-linear models using genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1995; **28**: 73–87.
84. Dunn WJ, Rogers D. Genetic partial least squares in QSAR. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 109–130.
85. Van Helden SP, Hamersma H, Van Geerestein VJ. Prediction of the progesterone receptor binding of steroids using a combination of genetic algorithms and neural networks. In *Genetic Algorithms in Molecular Modeling*, Devillers J (ed.). Academic Press: New York, 1996; 159–192.
86. Li T, Mei H, Cong P. Combining nonlinear PLS with the numeric genetic algorithm for QSAR. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 177–184.
87. Hou TJ, Wang JM, Xu XJ. Application of a genetic algorithm on the structure–activity correlation study of a group of non-nucleoside HIV-1 inhibitors. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 303–310.
88. De Weijer AP, Lucasius CB, Buydens LMC, Kateman G, Heuvel HM, Mannee H. Curve fitting using natural computation. *Anal. Chem.* 1994; **66**: 23–31.
89. Hibbert DB. A hybrid genetic algorithm for the estimation of kinetic parameters. *Chemometrics Intell. Lab. Syst.* 1993; **19**: 319–329.
90. Dane AD, Veldhuis A, de Boer 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.
91. Dane AD, van Sprang HA, Buydens LMC. A two-step approach toward model-free X-ray fluorescence analysis of layered materials. *Anal. Chem.* 1999; **71**: 4580–4586.
92. Wise BM, Holt BR, Gallagher NB, Lee S. A comparison of neural networks, non-linear biased regression and a genetic algorithm for dynamic model identification. *Chemometrics Intell. Lab. Syst.* 1995; **30**: 81–89.
93. Broudiscou A, Leardi R, Phan-Tan-Luu R. Genetic algorithm as a tool for selection of D-optimal design. *Chemometrics Intell. Lab. Syst.* 1996; **35**: 105–116.
94. Wienke D, Lucasius CB, Kateman G. Multicriteria target vector optimization of analytical procedures using a genetic algorithm. Part I. Theory, numerical simulation and application to atomic emission spectroscopy. *Anal. Chim. Acta* 1992; **265**: 211–225.
95. Wienke D, Lucasius CB, Kateman G. Multicriteria target vector optimization of analytical procedures using

a genetic algorithm. Part II. Polyoptimization of the photometric calibration graph of dry glucose sensors for quantitative clinical analysis. *Anal. Chim. Acta* 1993; **271**: 253–268.
96. Pizarro Millán C, Forina M, Casolino C, Leardi R. Extraction of representative subsets by potential function method and genetic algorithms. *Chemometrics Intell. Lab. Syst.* 1998; **40**: 33–51.
97. Tominaga Y. Representative subset selection using genetic algorithms. *Chemometrics Intell. Lab. Syst.* 1998; **43**: 157–163.
98. Lucasius CB, Dane AD, Kateman G. On k-medoid clustering of large data sets with the aid of a genetic algorithm: background, feasibility and comparison. *Anal. Chim. Acta* 1993; **282**: 647–669.
99. Jiang JH, Wang JH, Chu X, Yu RQ. Clustering data using a modified integer genetic algorithm (IGA). *Anal. Chim. Acta* 1997; **354**: 263–274.
100. Van Kampen AHC, Ramadan Z, Mulholland M, Hibbert DB, Buydens LMC. Learning classification rules from an ion chromatography database using a genetic based classifier system. *Anal. Chim. Acta* 1997; **344**: 1–15.
101. Tominaga Y. Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, ANNs, and KNN. *Chemometrics Intell. Lab. Syst.* 1999; **49**: 105–115.
102. Shaffer RE, Small GW. Genetic algorithms for the optimization of piecewise linear discriminants. *Chemometrics Intell. Lab. Syst.* 1996; **35**: 87–104.
103. Chen ZP, Jiang JH, Li Y, Yu RQ. Nonlinear mapping using real-valued genetic algorithm. *Chemometrics Intell. Lab. Syst.* 1999; **45**: 409–418.
104. De Gracia J, Saravia MLMFS, Araújo AN, Lima JLFC, Del Valle M, Poch M. Evaluation of natural computation techniques in the modelling and optimization of a sequential injection flow system for colorimetric iron(III) determination. *Anal. Chim. Acta* 1997; **348**: 143–150.
105. Weber L, Wallbaum S, Broger C, Gubernator K. Optimization of the biological activity of combinatorial compound libraries by a genetic algorithm. *Angew. Chem. Int. Ed. Engl.* 1995; **34**: 2280–2282.
106. Hartnett MK, Bos M, van der Linden WE, Diamond D. Determination of stability constants using genetic algorithms. *Anal. Chim. Acta* 1995; **316**: 347–362.
107. Parczewski A, Lucasius CB, Kateman G. Evolutionary determination of physico-chemical parameters and concentrations of analytes from titration data. *Fresenius J. Anal. Chem.* 1994; **348**: 626–632.

