ELSEVIER

Analytica Chimica Acta 388 (1999) 265–271

ANALYTICA
CHIMICA
ACTA

The quality of optimisation by genetic algorithms

R. Wehrens[a,*], E. Pretsch[b], L.M.C. Buydens[a]

[a]*Laboratory of Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED, Nijmegen, The Netherlands*
[b]*Laboratory of Organic Chemistry, Swiss Federal Institute of Technology (ETH), CH-8092, Zürich, Switzerland*

Received 4 September 1998; received in revised form 1 December 1998; accepted 1 December 1998

Abstract

The recently introduced quality criteria for optimisation, describing the coverage of the search and solution spaces as well as the reproducibility of both, are applied in combination with experimental design to fine-tune parameter settings and fitness function of a genetic algorithm for the structure optimisation of a heptapeptide. A series of influences of the investigated parameters are revealed by these criteria, while none of them seem significant from the fitness values of the last population alone. It is therefore suggested to apply these criteria, which are not based on the fitness value of the final population, when developing genetic algorithms. It is shown that they are easily adaptable to specific problems.

Keywords: Genetic algorithms; Quality criteria; Optimal settings; Structure optimisation

1. Introduction

Stochastic optimisation techniques like genetic algorithms (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 (see e.g., [1–3]). These methods are especially beneficial when the search space is complex with many local minima, so that conventional techniques fail to find global minima and a full search is not feasible. Although it is generally accepted that stochastic methods are the best choice in complex search spaces, there is no guarantee that they will find the global optimum. A wide variety of parameters, sometimes critically influencing the optimisation process, must be defined by the user but usually their influence is not systematically investigated. The only characteristic monitored in most cases is the development of the fitness of the best individual, which clearly provides only very little information about the search capability of a GA.

To address this problem, recently four criteria have been proposed which can be calculated using the whole population of every generation: the coverage of the whole search space, that of the solution space, defined as the relevant part of the search space, and the reproducibility of both when starting a GA several times [4]. The coverage of the search space is calculated by dividing the search space in to hypercubes of equal sizes and counting the relative number of those visited during the optimisation. The higher the coverage, the better the chance that no relevant part of the search space is missed. This criterion characterises the

---

*Corresponding author. Tel.: +31-24-365-2053; fax: +31-24-365-2653; e-mail: rwehrens@sci.kun.nl

PII: S0003-2670(99)00081-1

random component of the GA. The second criterion, the coverage of the solution space, is calculated by grouping the similar solutions into clusters and counting the number of clusters in the final populations of the pooled repeated runs. The other two criteria provide a measure of the reproducibility of the optimisation. The coverage of the total search space is characterised by principle component analysis (PCA) [5] of the entire runs. All individuals encountered during optimisation are then projected into the spaces defined by a limited number of PCs of this and other replicated runs. The ratio of within-run and between-run residuals is then calculated as a measure for the reproducibility. Ideally it should be close to unity. Finally, the reproducibility of the solution space is calculated from the cluster analysis of pooled runs by counting how many of the clusters contain individuals from different runs. The coverage of the search space and the solution space provide measures for the probability that the global optimum is found and are therefore very essential quality criteria of any GA. They are more important even in case one is not only interested in one single best solution but in all distinct good solutions, for example, in structure optimisation where several different low-energy conformations of the same species can co-exist at room temperature.

This paper shows that the quality criteria can be used as response variables in experimental design studies. Contrary to designs using only the performance of the best member of a population, the current setup yields valuable information on the optimal settings of the GA. As an additional advantage, it is shown that the form of the fitness function can be optimised as well, because the fitness values themselves are not explicitly used in the quality criteria. Finally, the paper illustrates that the criteria can be implemented with minor modifications for problems with a large number of parameters.

Table 1
Levels of GA parameter settings

Parameter	Code	Level 1 (−)	Level 2 (+)
Selection pressure	sep	low	high
Population size	pop	50	150
Maximum generations	gen	200	800
Crossover type	xot	two-point	uniform
Crossover probability	xop	0.5	0.85
Mutation probability	mup	0.04	0.08
Violations weight (A)	vio	1	2

2. Experimental

Dermorphine was recently used in a study comparing the behaviour of GA and simulated annealing optimisation [6]. The structure was optimised by minimising distance constraint violations obtained from NMR measurements [7], and atom overlaps, both expressed in energy units. The error function is a linear combination of both:

$$E = AE_{\mathrm{viol}} + E_{\mathrm{overlap}}. \tag{1}$$

The weighting factor A was set to 2.0 in [6]. The influence of GA parameter settings was investigated by experimental designs using the number of evaluations needed to reach the optimal fitness and fitness measures as response variables. Fitnesses of the best string in a population have been used in combination with factorial designs [8], but usually GAs are fine-tuned using a one-at-a-time approach, or standard search settings are taken. This can lead to sub-optimally configured GAs with bad performance.

In the present paper, six GA parameters and the weighting factor A are fine-tuned for the structure optimisation of dermorphine. These, and the two levels at which they are set, are gathered in Table 1 (see [9] for more information on GA search parameters). In all experiments, sigmoid fitness scaling, tournament selection, and elitism of 10% of the population are applied. Torsion angles are represented by 10-bit Gray code binary numbers. Low selection pressure is achieved by choosing a tournament size of 2, and high selection pressure by a tournament size of 4.

A Plackett-Burman experimental design [10] for seven factors is used (Table 2), where the quality criteria are used as response variables. For each parameter setting of the GA, five replicate runs were performed to calculate the quality criteria where only the random seed (and therefore the initial population as well) is changed. To be able to calculate the standard errors from the experimental design, each set of five runs was repeated with other random seeds. In total, 80 experiments ($8\times5\times2$) were performed in the fine-tuning of the settings.

Table 2
Plackett-Burman design for seven factors (plus and minus signs indicate high and low levels for the variables, respectively)

Experiment	sep	pop	gen	xot	xop	mup	vio
1	+	+	−	+	−	−	−
2	−	+	+	−	+	−	−
3	−	−	+	+	−	+	−
4	−	−	−	+	+	−	+
5	+	−	−	−	+	+	−
6	−	+	−	−	−	+	+
7	+	−	+	−	−	−	+
8	−	−	−	−	−	−	−

All calculations were performed on SUN Sparc 1 workstations using the software package GATES [11,12]. The GA strings and results of each evaluation were written to a file and processed afterwards with the statistical package R [13].

3. Results and discussion

3.1. Fitness measures and experimental design

To contrast experimental design results using energies with those using quality criteria as response variables, the effects of the Plackett-Burman design in Table 2 on several fitness measures are depicted in Fig. 1. The energy of the best string in the last population, and the mean and median energies of strings in the last population are used, respectively. To be able to compare the different energies, for the calculations in this figure the van der Waals error and violation error are given equal weightage (*A* from Eq. (1) equals 1); during the GA optimisation itself, of course, *A* has the value indicated by the experimental design in Table 2.

No significant effect on the energy of the best solution can be detected. The only parameter that has a significant effect is the mutation rate (median energy). A higher mutation rate causes a more diverse population, so this is exactly what one would expect. No evidence is found, however, that a higher mutation rate also leads to better solutions (no effect on the best solution). These results seem to indicate that the GA parameters do not influence the search very much.

3.2. Adaption of the quality criteria to dermorphine

The implementation of the quality criteria as defined in [4] can be adapted to other systems such as dermorphine quite easily, as shown below. Although the implementation differs in some cases because of the much larger size of the problem, the main ideas are still used.

3.2.1. Coverage of the search space: “Coverage”

The number of different solutions that is actually considered is a measure for how well the search space is covered. This, in turn, is related to the chance of missing important solutions. For structure optimisation problems, where the parameters to be optimised are torsion angles in the range between -180 and 180 degrees, this is calculated as follows. The range for each parameter is divided into a number of equally sized parts (e.g., with two parts this would mean a *high*

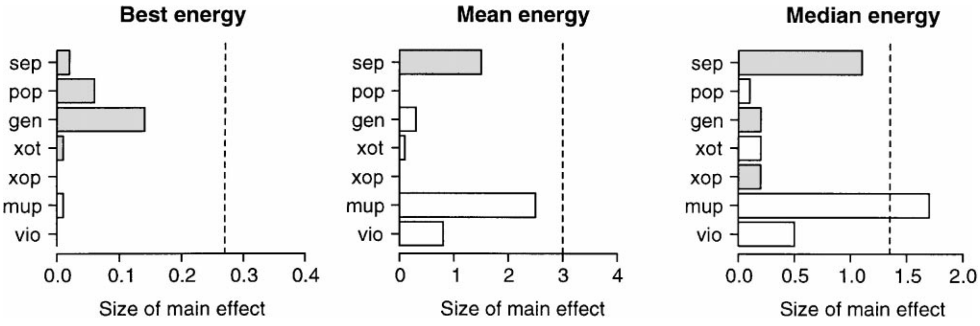

Fig. 1. Main effects of parameters on energy values. Effects with a negative sign are indicated in gray. The dashed line indicates three times the estimated standard error.

and a *low* value) and all combinations (each defining a hypercube in high-dimensional space) are counted. In [4] this was done for five replicate runs at once. For the larger dermorphine, this criterion is calculated for each replicate run separately, allowing the estimation of the variability of this criterion. Moreover, it is expressed here as the number of parts of the search space that have been visited divided by the total number of evaluations. In this case, each torsion angle range is divided in parts of 60°.

3.2.2. Coverage of the solution space: "Nr. of clusters"

This criterion counts the number of distinct good solutions that are found, since in general the total number of good solutions is unknown. For this, the final solutions (all elements in the last populations of the replicate runs) are clustered using Euclidean distances in the space spanned by the parameters that are optimised (here the 31 torsion angles). A greedy cluster method was used [4] that only takes into account solutions below an energy threshold. The best unclustered solution forms the cluster centre and all other unclustered solutions that are within a certain predefined distance of the centre are also included in the cluster. This process continues until no more unclustered solutions are available satisfying the energy threshold. In this case, separate thresholds for both distance violations and atom overlap were used, and were set to values of 1 and 3 energy units, respectively.

3.2.3. Reproducibility of search space coverage: "Reproducibility"

This criterion is used in exactly the same way as in [4]. Since PCA on all populations would be prohibitive in terms of computer memory and time, 20 evenly distributed generations are taken here. This was shown not to alter the results significantly.

3.2.4. Reproducibility of the coverage of the solution space: "Inter-run distance"

The clusters defined with the second criterion consist of members of replicate runs. Poor reproducibility means that each run finds different solutions, whereas a good reproducibility indicates that the same good solutions are found consistently. Therefore, in [4] this criterion is formed simply by calculating the number of clusters that is "touched" in more than one replicate run. In cases with a larger number of parameters that are optimised and a large number of local optima, this may lead to seemingly irreproducible search behaviour, as depicted in Fig. 2. Clearly, the GA is not able to find the same solutions repeatedly. Therefore, in this case we use a related measure, the mean smallest distance between clusters from different runs. If one cluster contains elements from several runs, these runs have a distance of 0 to each other. Both implementations are depicted in Fig. 3.

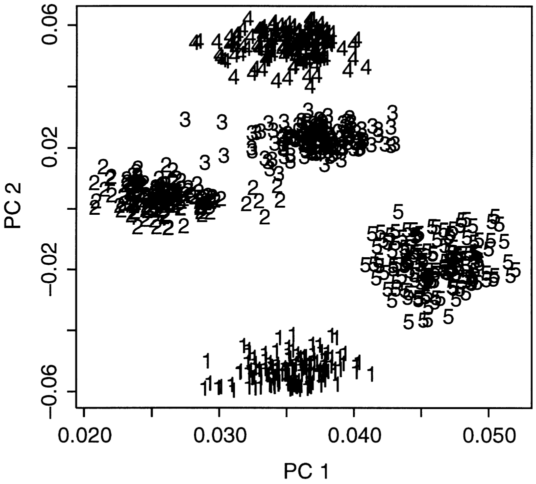

Fig. 2. Principal component representation of the solutions in the last populations of five replicate runs (indicated by the numbers). Experimental settings: second line in Table 2.

3.3. Quality criteria and experimental design

From the experimental designs using fitness measures as response variables it could be concluded that most parameters have no influence on the optimisation. However, this conclusion is not justified, as can be seen in Fig. 4, where the main effects on the four quality criteria are displayed. Clearly, several are significant.

Factors having a significant influence on the coverage are the selection pressure, the crossover probability and the mutation probability. The latter two have significant positive effects: both are related to mechanisms that produce new genetic material in the pool, the mutation by a completely random process, the crossover a random process guided by the fitness of the parent strings. A high selection pressure has a negative

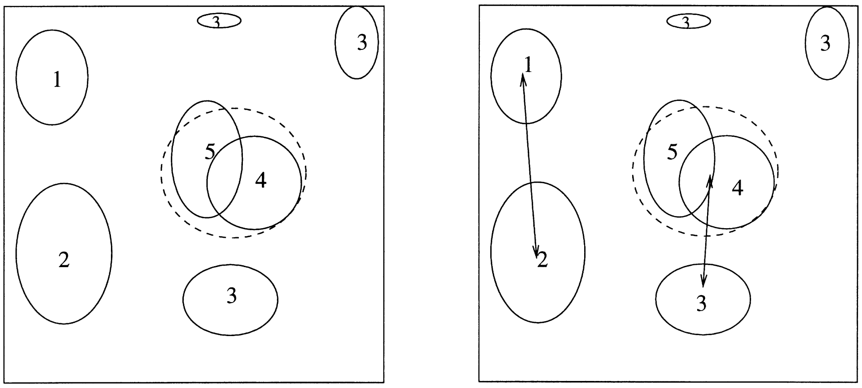

Fig. 3. Two ways to define reproducibility of the coverage of the relevant search space. In the left picture (corresponding with the criterion from [4]), one cluster is indicated consisting of members from two replicate runs (runs indicated with numbers). In the right picture distances to the closest clusters, containing elements from other runs, are indicated. The mean of these distances is used as the current criterion.

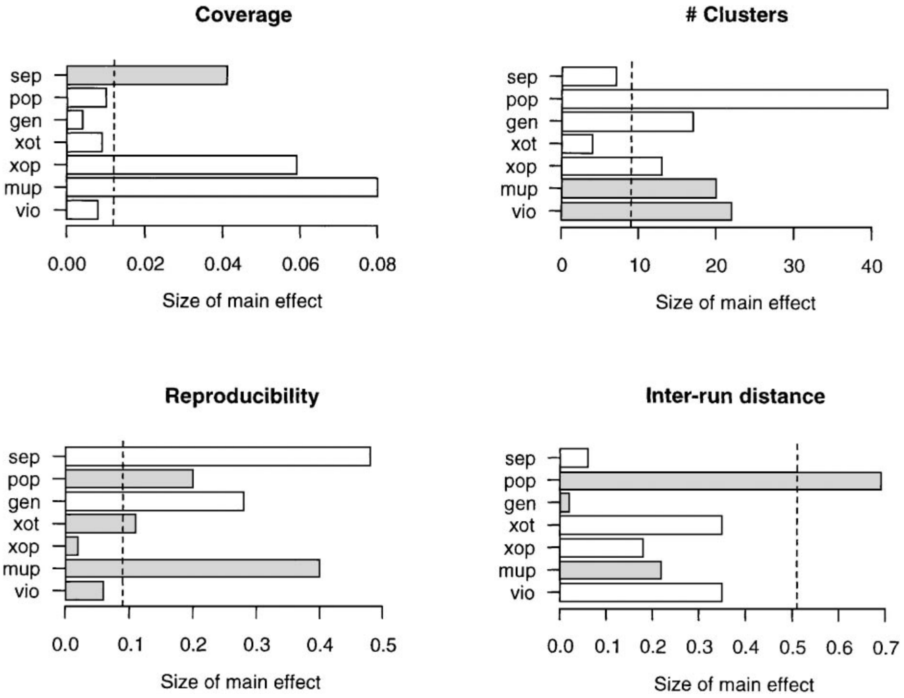

Fig. 4. Main effects of parameters on the quality criteria. Effects with a negative sign are indicated in gray. The dashed line indicates three times the estimated standard error.

effect on the coverage because solutions with high energies do not survive in the population, thus decreasing the diversity.

A larger population size has a beneficial effect on the number of different solutions that is found, as have a high number of generations and a high crossover rate. Higher mutation rates or a higher weight factor for the violation error term in Eq. (1) lead to fewer distinct solutions. From Fig. 1 it can be concluded that this is caused by the less optimal members of the population; the best members are comparable to those of runs with low mutation rates.

The reproducibility of the search space coverage is decreased significantly by a higher selection pressure and more generations.[1] Since no measures are taken to prevent premature convergence, a high selection pressure may cause the GA to give too much attention to strings that seem to be very good. These then dominate the whole population. Especially in later stages of the optimisation, replicate runs may have committed themselves to different parts of the search space. Better reproducibility is achieved with high mutation rates, a two-point crossover method, and small populations.

Larger populations have a diminishing effect on the distances between clusters found in different runs. This is related to the number of clusters that is found; the more clusters are found, the more probable it is that some of them are close together. The other effects on the inter-run cluster distance are not significant.

By varying the settings for the parameters the adverse effect of one parameter may be compensated for by another; no single parameter shows significant effects for all criteria. The criteria described above provide a way to guide this process of finding optimal parameter settings. In this case the following conclusions can be drawn about the GA configuration. It is easily seen that a low selection pressure (i.e. binary tournament selection), a large population size, uniform crossover, a high crossover probability and a low violations weight only have significant beneficial effects for the four quality criteria. A larger number of generations leads to more distinct solutions at the expense of reproducibility; higher mutation rates have exactly the opposite effect. To minimise run time, therefore, one would perform runs with a small number of generations and low mutation rates. This would also decrease coverage somewhat. As noted in [4], it is very important to keep diversity in GA populations. Several specialised operators can be used for this. The most used ones are sharing, crowding [14], and forced mutations when two strings are too much alike [15]. To counter the effect of the low mutation rates, coverage could easily be increased by using one of these operators.

3.4. Comparison of optimal GA settings

For completeness, we here compare the results of the optimisation using fitness values and the number of evaluations needed as described in [6] and the approach using the quality criteria. Since different parameters and parameter levels have been included in the experimental designs, the results are only presented here as an illustration. The optimal settings are gathered in Table 3. In both cases 800 generations were used, and five replicate runs were performed. The same five random seeds were used so that the search started from the same initial populations in both sets of experiments. Whereas in this work tournament selection was used with tournament sizes of 2 and 4, in [6] rank-based threshold selection was used. The optimal threshold was found to be 25% of the population (intermediate level) [6]. Here, no attempt was made to optimise the number of bits for the representation of torsion angles. Ten bits were used in all experiments. In [6] seven bits were used (intermediate level).

The results of the runs expressed in the four quality criteria, as well as on some fitness measures and the number of evaluations needed to find the best solutions are gathered in Table 4. Clearly, the current approach yields better results, not only on the four quality criteria but also on the energy of the best solutions. Since the diversity in the population is maintained throughout the search (see the large mean and median energies in the last population), premature convergence is avoided. Trying to minimise the mean or median error in a population or the number of evaluations needed to find the best solution does the exact opposite.

Table 3
Optimal GA parameter settings from [6] (first column) and this work (second column)

Parameter	From [6]	This work
Selection pressure	Intermediate	High
Population size	50	150
Crossover type	Uniform	two-point
Crossover probability	0.7	0.85
Mutation probability	0.05	0.08
Violations weight (A)	2	1

---

[1] The criterion is defined as a dissimilarity measure; therefore, a positive effect in Fig. 4 is decreasing reproducibility.

Table 4
Quality measures for the optimised GAs from [6] (first column) and this work (second column)

Quality measure	From [6]	This work
Energy of best string	3.172	2.948
Mean energy of last population	4.775	13.02
Median energy of last population	3.251	8.008
First generation in which best found	156	402
Coverage	0.628	0.834
Number of distinct solutions	66	117
Reproducibility	3.843	2.839
Mean between-run distance	6.234	5.949

4. Conclusion

The quality criteria presented here address an important issue in the practical use of optimisation methods such as GAs. Questions on the best evaluation or fitness function, and optimal settings of search parameters are not easily answered by merely looking at the quality of the best solution. The quality criteria aim to provide guidance with this process.

It is shown that experimental designs using conventional measures such as the energy of the best solution or the number of evaluations needed to reach the best solution offer only little information and may even yield misleading results. The proposed criteria yield much more detailed information about the search process. Moreover, since they effectively use only the values of the parameters that are being optimised, they can be used to fine-tune fitness functions.

The criteria are defined in general terms, making it possible to change their implementation according to the application at hand. Thus, they can be applied to a range of problems with different characteristics. In the present form, they are focused to problems where more than one solution is important, a situation that is often encountered in chemistry [4]. Similar criteria could also be applied to other stochastic optimisation methods such as simulated annealing, and evolutionary strategies. Research in this field is in progress.

In summary, the quality criteria provide a measure to compare in an objective way the performance of optimisation methods, independent of the implementation of the evaluation criterion. This in turn provides insight in the behaviour of the optimisation method and gives the possibility of fine-tuning in combination with experimental design methods.

References

[1] D.E. Clark, D.R. Westhead, Evolutionary algorithms in computer-aided molecular design, J. Comput. Aided Mol. Design 10 (1996) 337–358.
[2] J. Devillers (Ed.), Genetic algorithms in molecular modeling. Principles of QSAR and drug design, Academic Press, New York, 1996.
[3] R.S. Judson, Genetic algorithms and their use in chemistry, in: K.B. Lipkowitz, Donald B. Boyd, (Eds.), Reviews in computational chemistry, VCH Publishers, New York, 1997, pp. 1–73.
[4] R. Wehrens, E. Pretsch, L.M.C. Buydens, Quality criteria of genetic algorithms for sturcture optimisation, J. Chem. Inf. Comput. Sci. 38(2) (1998) 151–157.
[5] J.E. Jackson, A user's guide to principal components, Wiley, New York, 1991.
[6] A.H.C. van Kampen, L.M.C. Buydens, The ineffectiveness of recombination in a genetic algorithm for the structures elucidation of a heptapeptide in torsion angle space. A comparison to simulated annealing, Chemom. Intell. Lab. Systems 36 (1997) 141–152.
[7] K. Wuethrich, NMR of proteins and nucleic acids, Wiley, New York, 1986.
[8] R.E. Shaffer, G.W. Small, M.A. Arnold, Genetic algorithm-based protocol for coupling digital filtering and PLS regression: application to the NIR analysis of glucose in biological matrices, Anal. Chem. 68 (1996) 2663–2675.
[9] C.B. Lucasius, G. Kateman, Understanding and using genetic algorithms. Part 2: Representation, configuration and hybridization, Chemom. Intell. Lab. Systems 25 (1994) 99–145.
[10] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke (Eds.), Handbook of Chemometrics and Qualimetrics: Part A. Data Handling in Science and Technology, vol. 20A, Elsevier, Amsterdam, 1998.
[11] C.B. Lucasius, G. Kateman, GATES: Towards evolutionary large-scale optimization: A software-oriented approach to genetic algorithms. Part 1: General perspective, Comput. Chem. 18(2) (1994) 127–136.
[12] C.B. Lucasius, G. Kateman, GATES: Towards evolutionary large-scale optimization: A software-oriented approach to genetic algorithms. Part 2: Toolbox description, Comput. Chem. 18(2) (1994) 137–156.
[13] R. Ihaka, R. Gentleman, A language for data analysis and graphics, J. Comput. Graphic. Statist. 5(3) (1996) 299–314.
[14] D.E. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, Reading, MA, 1989.
[15] T. Brodmeier, E. Pretsch, Application of genetic algorithms in molecular modelling, J. Comput. Chem. 15(6) (1994) 588–595.