Adaptive Multicomponent Analysis by Genetic Algorithms

Peter Zinn

Ruhr-Universität Bochum, Lehrstuhl für Analytische Chemie, 44780 Bochum, Germany

Received July 30, 2004

The applicability of genetic algorithms for solving multicomponent analyses is systematically examined. As a genetic algorithm (GA), the basic proposal of Goldberg is implemented in a straightforward manner to simulate multicomponent analyses in analogy to the well-established UV−vis or IR methods, especially multicomponent regression. The main focus of the study is to investigate the behavior of the genetic algorithm in order to compare it with the well-known behavior of multicomponent regression. A remarkable difference between the two methods is that the genetic algorithm method does not need any calibration procedure because of its pure searching characteristic. As important features of multicomponent systems, the degree of signal overlap (selectivity), the behavior of systems with known and unknown component numbers and qualities, and linear as well as nonlinear relationships between the analytical signal and concentration are varied within the simulations. According to multicomponent regression, recovering concentrations by a genetic algorithm is of limited applicability with the exception of systems at a low degree of signal overlap. On the other hand, the recovery of a probe spectrum in the analytical process always gives satisfactory results independent of the features of the probe system. The genetic algorithm obviously shows autoadaptive behavior in probe spectrum recovery. The quality and quantity of the resulting components may dramatically differ from the given probe, although the resulting spectrum is nearly the same. In such cases, the resulting component mixture can be interpreted as an imitation of the probe. As well probe spectra, theoretically designed spectra can also be autoadapted by genetic algorithms. The only limitation is that the desired spectrum must, of course, be incorporated into the search space defined by the involved components. Furthermore, a spectral signal is only one single property of a chemical compound or mixture. Because of the nonlinear search characteristic of genetic algorithms, any other chemical or physical property can also be treated as a desired property. Therefore, the conclusion of the study is well-founded that an old challenge of applied chemistry, namely, the development of new chemical products with desired properties, seems to be reachable under the control of genetic algorithms.

INTRODUCTION

Genetic algorithms (GAs) are typical examples of nonlinear approximations applied in optimization problems. Introduced by Holland[3] and described in detail by Goldberg,[4] GAs have also received great interest in chemistry in the past one and a half decades. Leardi[5] gives an overview of the corresponding literature up until 2001. Therein, a main part of the applications is focused on analytical chemistry. Several of these and later publications deal with different aspects of GAs applied in multicomponent analysis (MCA).

A main aspect is the application of GAs in parameter optimization, especially for wavelength selection.[6,8,10−12,15,18−22] Besides multilinear regression (MLR), partial least squares,[6,11,12,19] neural networks,[8] and principal component regression[11,12,15] are also applied as MCA solving strategies combined with GAs for wavelength selection. The second aspect is the direct application of GAs for solving MCA. Examples are described in the MCA of toxic organic compounds,[9] rare earth elements,[13] noble metals,[14] vitamins,[16] and amino acids.[17] In these examples, the nonlinear abilities of GAs and the influence of the choice of GA parameters such as mutation, crossover rate, and so forth are pointed out.

The fundamental difference between the common MCA solving methods and GAs is that GAs are a pure searching technique without the necessity for calibration of regression coefficients. The knowledge about the inner relationship of the observed signal and the analytical composition of the probe is unnecessary with GAs. Searching the composition of a probe by GA is a self-organizing and, furthermore, a self-optimizing process. The autoadaptive characteristics of GAs coinciding with their nonlinear abilities gives them great potential in MCA. Therefore, simulations in which MCA is solved by GAs allows one to find situations or applications in which the use of GAs is preferable to MLR or other multivariate methods.

When the GA simulations are interpreted, MLR is mainly chosen as the reference method because both MLR and GA deal with real physical values instead of more abstract items as applied in other multivariate techniques. MLR is the classical approach to MCA, as applied in quantitative UV−vis or IR spectroscopy. The principles of linear algebra are a well-suited instrument for solving systems of simultaneous equations in MCA and multivariate calibration. Brereton[1] as well as Martens and Naes[2] give a wide introduction to the whole variety of multivariate techniques, including MLR and the corresponding literature. The visual sample spectrum and the knowledge of the involved pure substance spectra makes the results of MLR qualitatively almost predictable. This property should help to increase the transparency of the results of the GA approach. Additionally, the chosen

* Corresponding author e-mail: Peter.Zinn@ruhr-uni-bochum.de.

10.1021/ci049763m CCC: $30.25
Published on Web 05/28/2005

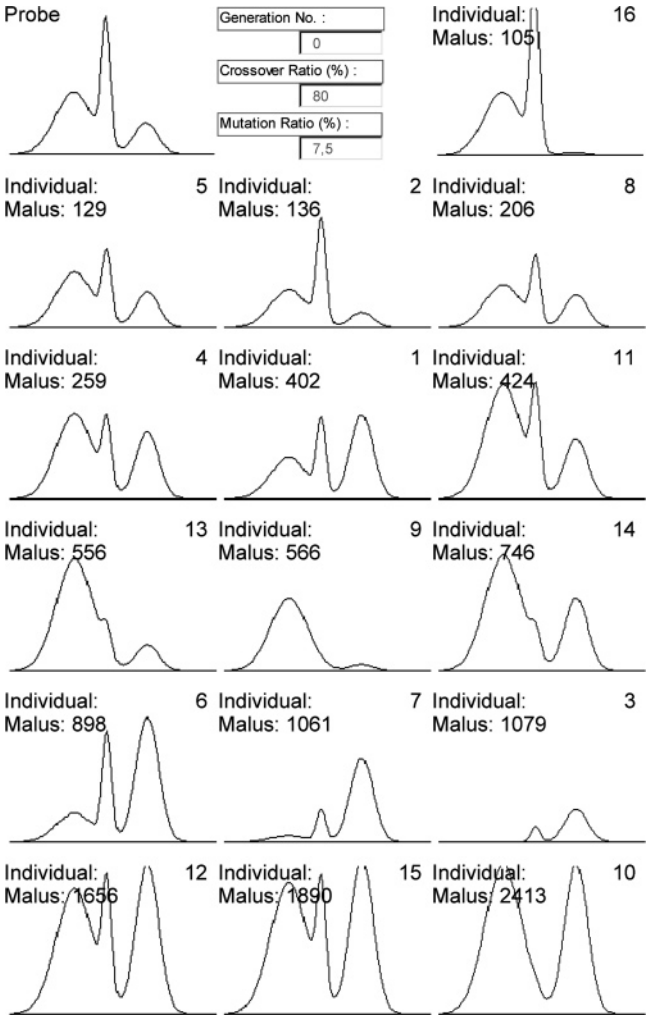

Figure 1. Probe spectrum and a typical set of 16 spectra randomly generated as a start generation. The shapes of the generated spectra are far different from the shape of the probe spectrum, with malus values ranging from 1.05 to 24.13. (In the figures, the malus values are multiplied by 100.)

examples are of a basic character in order to recognize the new qualities that are expected of the application of GA in MCA.

METHODOLOGY OF ADAPTIVE MULTICOMPONENT ANALYSIS

The implementation of the GA is straightforward in analogy to Goldberg's suppositions.[4] An additive mixture of Gaussian curves serves as a simulated spectrum of a multicomponent probe. The spectrum is divided into 128 equidistant values for the wavelength and the absorbance scale. The absorbance of each component is the product of the absorption coefficient at the corresponding wavelength and the concentration factor ranging from 0 to 127. A typical probe spectrum can be seen in the upper left corner of Figure 1.

In the first step of the GA, a start generation of, for example, 16 individual mixture spectra with randomly distributed concentration factors are generated. Figure 1 gives an example of a start generation of spectra. As expected, each spectrum is far different from the given probe spectrum. In the second step, each of the individual spectra is compared with the probe spectrum. Therefore, the corresponding sum of squares of errors is assigned to each individual as a malus value. Dependent on the malus values, the best fitting individuals have the highest probability to be selected in a spinning-wheel procedure to build the next generation of spectra. The winners of the spinning wheel become the subjects of the genetic operations. Crossover and mutation procedures are applied to the concentration values in order to build the new generation of mixtures. The GA stops if an individual spectrum of the new generation fits the probe spectrum within given tolerances.

Table 1. Parameters for Generating the Spectrum of the Standard Probe[a]

	component 2	component 3	component 4
maximum absorption coefficient	1024	1024	1024
band position	40	60	85
bandwidth	16	4	10
probe concentration	50	100	25
scaling factor	0.00001		

[a] The values have arbitrary units. Band positions and probe concentrations both range from 0 to 127. The value of the scaling factor is responsible for obtaining realistic absorbance values. The numeration of the components is in accordance with the other tables.

SIMULATIONS OF SOLVING MULTICOMPONENT ANALYSIS BY A GENETIC ALGORITHM

The GA is implemented under Visual Basic and includes a Microsoft Excel user interface. Figure 1 shows a typical Excel sheet showing a generation of spectra. Other sheets are available for the evolution statistics, history, control, and so forth. After some system validation, the number of individuals per generation was set to 16, the crossover rate to 80%, and the mutation rate to 7.5%. These adjustments were retained through the MCA simulations. Because the genetic search depends on probabilities, exact results of the concentration values are not to be expected. Therefore, all simulation runs were performed 12 times to get an idea about the precision of the method. All simulation runs were stopped after 25 generations.

SIMULATION OF WELL-DEFINED MULTICOMPONENT SYSTEMS

Probe Recovery with Known Components at a Low Degree of Overlap. MLR gives fairly good analytical precision when applied to well-defined multicomponent systems. Within such systems, nearly 100% of the measured signals are associated to the compound concentrations by calibration. In analogy, a first GA simulation should give an answer about the precision of the recovery of the concentration values. To generate a standard spectrum of the target probe, the parameters of Table 1 are used. Figure 1 shows the appropriate probe spectrum containing three single bands with a small degree of overlap. Each band belongs to one component. It is obvious that such a completely defined multicomponent system gives exact results for the recalculation of the concentrations by MLR, as reproduced in Table 1, namely, 50, 100, and 25 arbitrary concentration units.

When GA is applied, exactly recovered concentrations cannot be expected because of GA's probabilistic nature. Therefore, statistical values are used to describe the results

Table 2. Statistics for a Three-Component System Recovery by GA over 12 Simulation Runs at a Low Degree of Overlap[a]

	concn 2	concn 3	concn 4	minimum malus	average malus
probe	50	100	25		
simulation					
mean value	50.00	97.83	26.58	0.12	1.69
standard deviation	3.84	9.45	5.92	0.09	0.79
minimum value	44	78	18	0.02	0.33
maximum value	56	112	34	0.31	2.64

[a] Components 2, 3, and 4 have band positions at 40, 60, and 85 arbitrary wavelength units, respectively. The mean values of the recovered concentrations are significant and near to the expected probe concentrations.

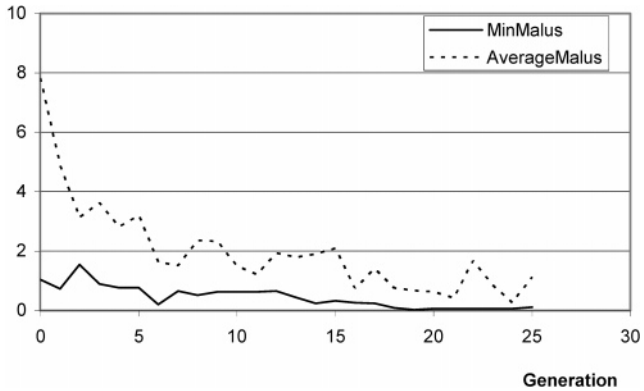

Figure 2. Development of the minimum and average malus in simulation 9 of 12 runs of the three-component system corresponding to Table 2. The smallest minimum malus has a value of 0.03 in generation 19. The malus is defined as the sum of squares of errors between the probe spectrum and the spectrum of the individual mixture.

in Table 2. Each statistical value refers to 12 single values belonging to the best generation of each simulation run. The best generation is defined by having the smallest malus of the run. Because of the simplicity of the probe spectrum, the mean concentrations found are not far different from the exact values. The values of the corresponding standard deviations and ranges between maximum and minimum concentrations are in agreement with the given spectral parameters. As expected with MLR, the band with the highest sensitivity (component 2) shows the smallest standard deviation of the corresponding concentration. Table 2 also gives the statistics of the minimum and average malus values over the 12 simulation runs. These values are important for the comparison with the following multicomponent systems.

Another typical evolution behavior shows the development of both malus values during a simulation, as reproduced in Figure 2. A relatively strong descent of the malus values in the early evolution period is followed by some horizontal fluctuation later on. After a few generations, not only are acceptable candidates for a solution found, but the similarity to the probe has also increased for the whole generation. According to the last statement, Figure 3 shows the spectra of generation 19 in simulation run 12. The comparison with the start generation in Figure 1 explicitly shows a tremendous improvement in fitting the probe spectrum of several individual spectra of that generation accompanied by small individual malus values.

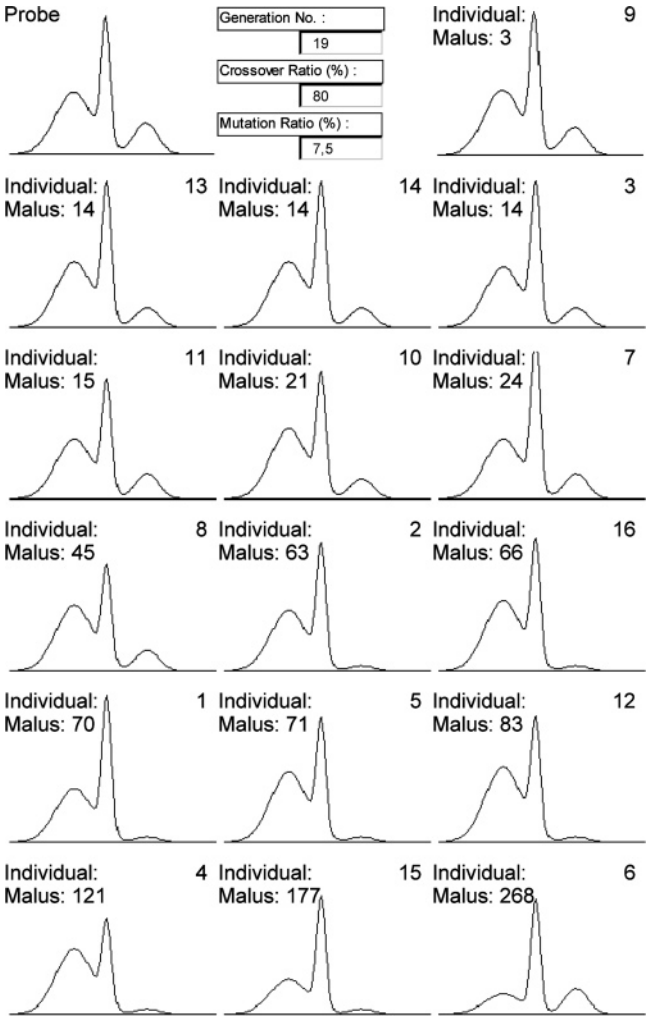

Figure 3. Probe spectrum and the 16 spectra of generation 18 of a simulation run. After genetic modifications from generation to generation, the shapes of the spectra are fairly good adaptations of the shape of the probe spectrum, with malus values now ranging from 0.03 to 2.68.

The résumé of this simulation is that MLR is superior to GA in solving MCA with a small degree of signal overlap. MLR derives exact results of concentration values if a well-defined multicomponent system is examined. GA obtains concentrations that are not exactly correct but not far off from the expected values. By analyzing real probes, the additional influence of the experimental error has to be considered regardless of the applied solving technique. Thereby, MLR will also show statistical deviations from the expected results and, additionally, a calibration always has to be performed.

Probe Recovery with Known Components at a High Degree of Overlap. The precision of solving MCA by MLR decreases with the degree of overlap.[23] To investigate the behavior of GA in such a situation, the position of probe component 3 is set to 80 instead of 60 with a wavelength difference of 5 in comparison to component 4. As can be seen in Figure 4, the bands of components 3 and 4 are widely overlapped and cannot be localized as individual bands. Figure 4 also reproduces the spectra of the 18th generation of run 6 with several acceptable malus values for the recovery of the probe spectrum. The statistical results over all 12 runs are gathered in Table 3. When the results of Table 3 are

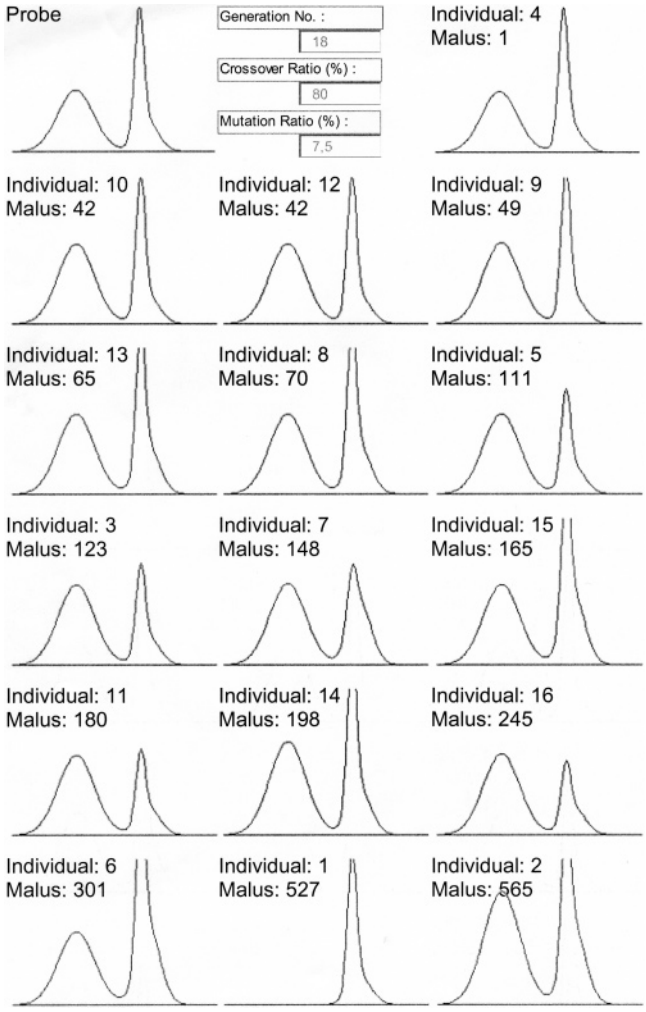

Figure 4. Probe spectrum of the three-component probe with a large degree of overlap and the 16 spectra of generation 18 of a corresponding simulation run. After genetic modifications from generation to generation, the shapes of the spectra are again fairly good adaptations of the shape of the probe spectrum, with malus values ranging from 0.01 to 5.65. Also, in the case of strong band overlap, the GAs give results similar to those in the case of low band overlap.

Table 3. Statistics for a Three-Component System Recovery by GA over 12 Simulation Runs at a High Degree of Overlap[a]

	concn 2	concn 3	concn 4	minimum malus	average malus
probe	50	100	25		
simulation					
mean value	50.08	95.83	27.25	0.06	1.91
standard deviation	1.88	11.22	5.24	0.07	1.05
minimum value	48	74	19	0.00	0.43
maximum value	55	107	36	0.21	3.62

[a] Components 3 and 4 have a large degree of overlap with band positions at 80 and 85 arbitrary wavelength units, respectively. The mean values of the recovered concentrations of components 3 and 4 are less near to the expected probe concentrations. Nonetheless, the minimum malus is smaller than that in the case of higher band selectivity, as reproduced in Table 2. The probability of finding a similar spectrum obviously increases with an increasing degree of overlap.

compared with those of Table 2, it must be recognized that the concentration-describing parameters as well as the malus values are very similar, solving the more-or-less overlapped system by GA. The greatest difference shown is the evolution

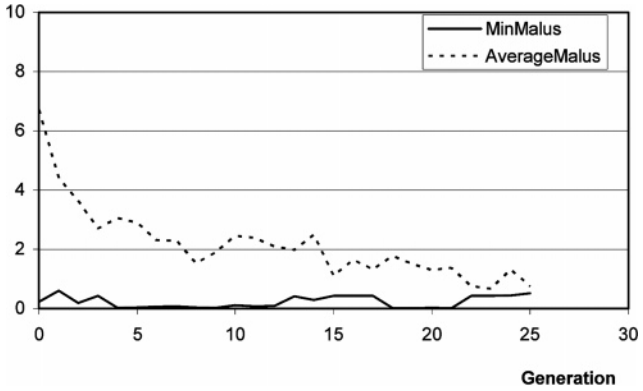

Figure 5. Development of the minimum and average malus in a typical simulation run of the strongly overlapping three-component system, corresponding to Figure 4. The smallest minimum malus has a value of 0.01 in generation 18. The descent of the minimum malus is less distinct than that in slightly overlapping systems because the probability of finding combinations of single-component bands fitting the encasing band is already growing in early generations.

of the minimum malus value. The descent of the minimum malus in the previous simulation is no longer observed, as reproduced in Figure 5. Horizontal fluctuations of the minimum malus now prevail. The reason for this behavior is that the probability of finding combinations of bands fitting the encasing band grows with an increasing degree of overlap. In such cases, different combinations of concentrations will reproduce the same spectrum. Therefore, in early generations, the probability of finding a compatible combination of concentrations is already high.

The consequence for strongly overlapping spectra is that GA is able to recover a given spectrum but not the corresponding real concentrations. The found mixture concentrations imitate the original probe spectrum. An increasing degree of overlap increases the ability of GA to find imitators. This is an important property of the GA technique that should be kept in mind for the summary discussion.

SIMULATION OF UNDEFINED MULTICOMPONENT SYSTEMS

Probe Recovery with Additional Components at a Low Degree of Overlap. As undefined multicomponent systems, we will utilize probes that are characterized by an unknown number of probe components with unknown qualities in addition to the already treated probe component quantities. To analyze such a probe by MLR, first the number and quality of the components have to be ascertained, followed by the calibration procedure. When the GA analysis of such a probe is simulated, the standard parameters of Table 1 again serve to generate the target probe. The simulation is now carried out by offering two additional components to the GA with band centers at 10 and 110 arbitrary wavelengths units. This system is well-conditioned because the degree of overlap is small and the three probe components are incorporated within all five offered components. The results of the 12 simulation runs are reproduced in Table 4. The mean values of the probe concentrations are found to be near to the expected ones, whereas the mean concentrations of the two disturbing components are minimized, but with a greater distance to the expected value of zero. The high standard deviations of the two disturbing concentrations in

Table 4. Statistics for a Three-Component Probe Recovery by GA over 12 Simulation Runs at a Low Degree of Overlap and Five Offered Components[a]

	concn 1	concn 2	concn 3	concn 4	concn 5	minimum malus	average malus
probe	0	50	100	25	0		
simulation							
mean value	10.83	49.25	101.50	23.42	9.17	0.49	3.88
standard deviation	8.42	4.67	9.68	7.34	7.27	0.28	0.96
minimum value	1	40	80	12	0	0.23	1.79
maximum value	28	56	112	34	22	1.12	5.06

[a] All five components 1, 2, 3, 4, and 5 have a small degree of overlap with band positions at 10, 40, 60, 85, and 110 arbitrary wavelength units, respectively. The mean values of the recovered probe concentrations 2, 3, and 4 are significant and near to the expected values. The concentrations of the additional components 1 and 5 are minimized as expected, but they are not negligible. The remaining concentrations are responsible for the increasing malus values. The concentrations 1 and 5 have low significance, indicated by the high relative standard deviations. Nevertheless, within the simulation runs, several candidates could be found with additional concentrations near 0 and with acceptable malus values.

relation to the associated mean values indicate their low significance to the reconstruction of the probe spectrum. Figure 6 gives a visual impression of the remaining quantities of the two disturbing components. The remaining disturbing concentrations are also responsible for enlarging both malus values. This is not surprising because the higher variability of the five-component system causes a much larger search space for possible combinations of concentrations. (The search space increases from $128^3 \approx 2$ million to $128^5 \approx 34$ billion combinations.) Although the search space has largely increased, the number of generations (25) and individuals per generation (16) are kept unchanged.

To summarize this simulation example, it has to be noticed that the GA solving procedure works without any modifications, independent of the number and quality of the probe components. When the number of substances offered as potential probe components increases, the malus values will increase too. Nevertheless, in the investigated multicomponent system characterized by a small degree of band overlap, several individuals with concentrations near to the probe composition could be observed.

Probe Recovery with Additional Components at a High Degree of Overlap. When the degree of overlap is enlarged, it is to be expected that the recovery of the probe composition by GA might only succeed imperfectly. To investigate such a multicomponent system, the positions of the disturbing bands of components 1 and 5 were set to 35 and 45, respectively, strongly overlapping probe component 2 with 40 arbitrary wavelength units. An example of an adapted generation of spectra is shown in Figure 7, and the results of the corresponding 12 simulation runs are summarized in Table 5. The concentration means of probe components 3 and 4 are still significant because of their low overlap degree. The less significant three overlapping compounds reinforce the expectation of an incomplete recovery of the probe concentrations. On the other hand, the comparison of Tables 4 and 5 indicates better minimum malus values of the overlapping system. This coincides with the already discussed three-component system of Table 3. The recovery of the probe spectrum is more successful than the recovery of the probe concentrations. Again, the tendency to imitate the

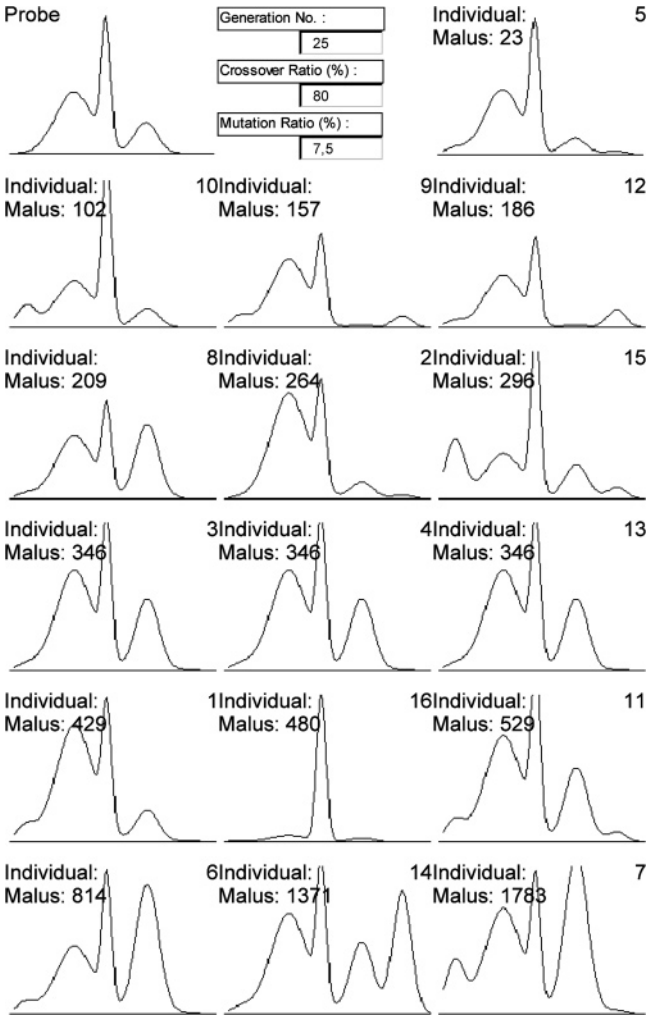

Figure 6. Probe spectrum of the three-component probe with a small degree of overlap and the 16 spectra of generation 25 of a corresponding simulation run including two additional slightly overlapping bands. The additional bands surround the probe bands at the low and high ends of the spectrum. After 25 generations, the additional bands are not completely minimized to 0 and small remaining absorptions can be recognized, resulting in increased malus values ranging from 0.23 to 17.83. Nevertheless, acceptably recovered probe spectra can be found.

probe with a different composition of concentrations but the same spectrum is clearly seen. Table 6 gives three typical patterns of recovered probe concentrations: Example I is similar to the probe composition. In example III, component 2 is completely substituted by 1 and 5. In example II, component 2 is partly substituted. More or less, substitution of probe component 2 by 1 and 5 increases the probability of finding a mixture with a spectrum similar to the probe spectrum.

Just as in the simulation worked out for defined multicomponent systems, the simulation of solving undefined systems with a high degree of overlap by GA shows its ability to find solutions imitating a given probe.

SIMULATION OF A NONLINEAR MULTICOMPONENT SYSTEM

Nonlinear Probe Recovery with a Missing Probe Component. In the simulations so far, the relationship between the absorbance as the measured signal and the

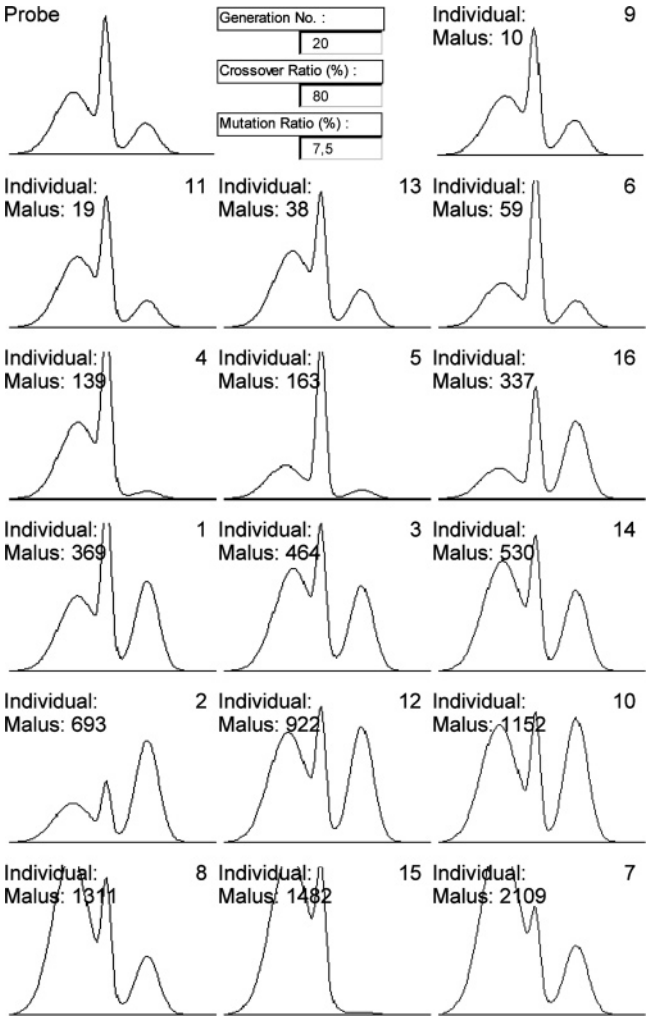

Figure 7. Probe spectrum of the three-component probe with a small degree of overlap and the 16 spectra of generation 20 of a corresponding simulation run including two additional strongly overlapping bands. The additional bands closely surround the leftmost probe component. Several minimum malus values are smaller than those in the slightly overlapping system of Figure 5 because the probability of finding combinations of bands fitting the encasing band is higher in comparison to the system of Figure 4.

Table 5. Statistics for a Three-Component Probe Recovery by GA over 12 Simulation Runs at a High Degree of Overlap and Five Offered Components[a]

	concn 1	concn 2	concn 3	concn 4	concn 5	minimum malus	average malus
probe simulation	0	50	100	25	0		
mean value	15.08	19.00	99.67	27.67	19.92	0.17	5.66
standard deviation	7.45	13.92	9.38	7.18	10.93	0.12	2.91
minimum value	2	0	88	15	1	0.00	2.65
maximum value	30	48	118	40	34	0.44	14.20

[a] The additional components 1 and 5 strongly overlap with probe component 2. The band positions from components 1−5 are 35, 40, 60, 85, and 45 arbitrary wavelength units, respectively. Again, the mean values of the undisturbed probe concentrations 3 and 4 are significant and near to the expected values. Now, the concentrations of the overlapping components 1, 2, and 5 are less significant. On the other hand, the minimum malus has decreased because of the higher probability of finding good combinations for the imitation of the pure concentration 2.

analytical concentration was treated as linear. In a final simulation, the behavior of the GA regarding nonlinear signals could be proven. Therefore, the transmittance T was examined as a measured signal instead of the absorbance $A = -\log T = \epsilon cd$. As a consequence, changing the measured signal effects the concentration c in a nonlinear way.

Table 6. Typical Results Emphasizing the Tendency to Imitate the Probe when Applying GAs in MCA[a]

	concn 1	concn 2	concn 3	concn 4	concn 5	minimum malus
probe simulation	0	50	100	25	0	
example I	3	38	98	27	2	0.13
example II	10	30	101	23	19	0.12
example III	18	0	88	28	34	0.10

[a] The results stem from the simulation system described in Table 5. The spectrum recovery in each example is generally correct, as indicated by low malus values, whereas the recovery of the concentrations show three typical patterns: The result of example I is very similar to the probe composition. In example III, component 2 is completely substituted by 1 and 5. In example II, component 2 is partly substituted by 1 and 5.

Again, the standard probe was used to generate the target spectrum. Recovering the target spectrum of probe component 2 at 40 arbitrary wavelength units was not offered to the GA. The other four components are left unchanged with band positions at 35, 60, 85, 45 arbitrary wavelength units, respectively. Under these conditions, an imitation of the band of component 2 by components 1 and 5 is to be expected. Figure 8 gives a typical example of a spectra generation under these conditions. Individual 4 is nearly perfectly adapted to the probe spectrum even if the GA is searching in a nonlinear system. The complete results of the 12 simulation runs are reproduced in Table 7. All of the four components involved in the probe rebuilding process, probe components 3 and 4 as well as the imitating components 1 and 5, are of significance. The relative mean standard deviations of their concentrations are about 20% and smaller. The mean concentrations of probe components 3 and 4 are near the expected values. Because of the symmetric arrangement of the imitating components 1 and 5 around the missing band of probe component 2, no significant differences between 1 and 5 can be observed. The minimum and average malus values are small, whereas the influence of the nonlinear transformation between absorbance and transmittance was not examined in detail.

It can be concluded that GA is successfully applied in solving nonlinear MCA. Even if the number and quality of the components are unknown, GA tries to find solutions that imitate the given probe spectrum.

DISCUSSION

The starting point of the presented investigation is the idea to apply GAs in solving an analytical problem, namely MCA. Several results of interest could be worked out by simulating systematically varied multicomponent systems. How are these results to be interpreted? The overview of the elaborated results in Table 8 may clarify the situation. The order of the lines in Table 8 corresponds to the discussed simulations. The values of the simulation influence and of the resulting parameters are characterized by "+" and "−". For example, the first line should be interpreted in this way: If the measurement linearity is high (+), the number and

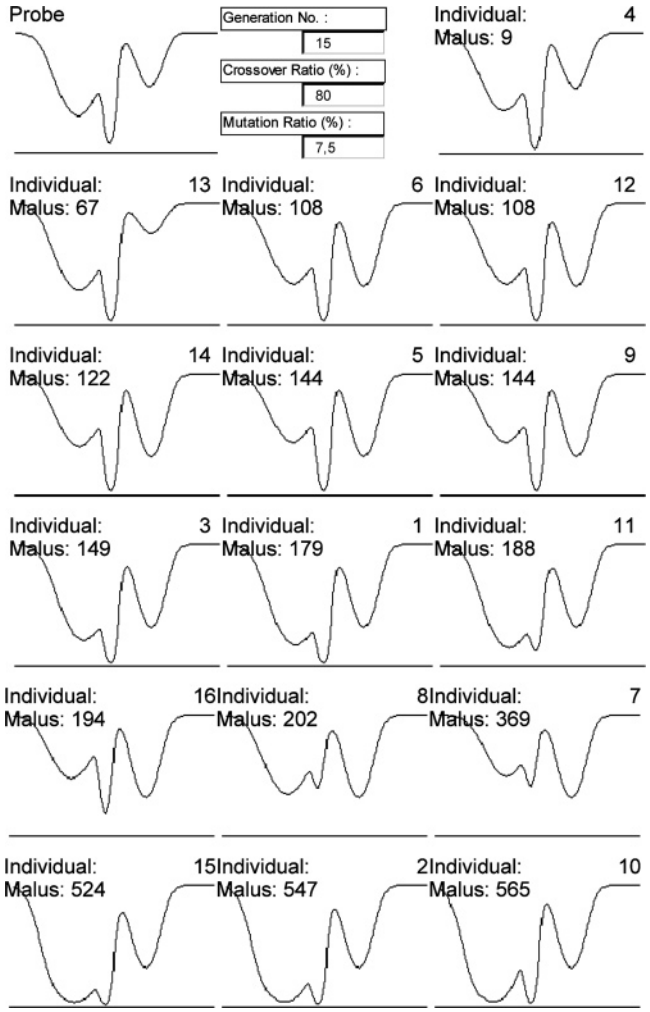

Figure 8. Example of a simulation run replacing absorbance by transmittance as the measured signal. During the simulation, the left-most probe component is eliminated and substituted by the two additional components of Figure 7. Also, in this nonlinear situation, the GA finds imitations for the probe composition, recovering the spectrum shape of the probe.

Table 7. Statistics for a Three-Component Probe Recovery over 12 Simulation Runs for a Nonlinear Genetic Search[a]

	concn 1	concn 2	concn 3	concn 4	concn 5	minimum malus	average malus
probe	0	50	100	25	0		
simulation							
mean value	26.50		98.67	25.50	27.83	0.11	1.41
standard deviation	5.39		18.17	3.61	5.81	0.07	0.58
minimum value	13		69	18	22	0.02	0.55
maximum value	33		127	32	44	0.21	2.26

[a] The three probe components 2, 3, and 4 are unchanged, whereas in the recovery, component 2 is not offered but substituted by the overlapping components 1 and 5. During the recovery, the band positions from component 1−5 are 35, (no value), 60, 85, and 45 arbitrary wavelength units, respectively. Also, the absorbance is substituted by the transmittance as the analytical signal to examine a nonlinear relationship between the signal and the concentration. The low malus values indicate that the genetic search is unaffected by nonlinear signal transformations. The replacement of component 2 by a combination of 1 and 5 is successful because the substituted part of the probe spectrum is covered by the combination's search space.

qualities of the components are known (+), and the selectivities of the components are high (+), then the resulting quantities and the spectral congruence are good approximations (+) of the probe concentrations and spectrum. As a further example, in the last line, the influencing parameters are small and unknown (−) with the consequence of a wrong concentration pattern (−) but, still, a good reproduction of the probe spectrum (+).

Table 8. Overview of the Simulation Parameters and Results for Solving Multicomponent Analyses Applying GAs[a]

simulation parameters				results	
measurement linearity + linear/ − nonlinear	component number + known/ − unknown	quality of components + known/ − unknown	selectivity of components + high/ − low	quantity of components + accurate/ − false	spectral congruence + high
+	+	+	+	+	+
+	+	+	−	−	+
+	−	−	+	+	+
+	−	−	−	−	+
−	−	−	−	−	+

[a] The method recovers accurate quantities of the components only for high selective probes (low degree of overlap). On the other hand, the recoveries of the probe spectra are successful in all different situations independent from the values of the simulation parameters. The method is obviously able to find imitations of a probe composition that adapt the probe spectrum. (As a prerequisite, the spectrum must, of course, be situated within the search space defined by the involved components.)

From MLR, it is well-known that the reconstruction of a spectrum from strongly overlapping component bands can result in several solutions. Different mixtures can have the same or similar spectra. That is an advantage for the spectrum recovery but a disadvantage for the concentration analysis. According to MLR, Table 8 makes it obvious that GA may also fail to reproduce an analytical concentration pattern. The reproduction of the qualitative and quantitative probe compositions is restricted to systems with high selectivities between the probe components. On the other hand, the analytical abilities of GAs are, by far, surpassed in reproducing spectral similarities. GA finds, in all of the different simulation situations, good congruences of the approximated spectra to the given probe spectra. Additionally, spectrum recovery by GA is a completely autoadaptive process without calibration or any other external intervention. This behavior is not limited to spectral similarities only but can also be extended to nearly all other physical or chemical properties of multicomponent systems. Leaving the analytical query behind, the presented examination has demonstrated the ability of GA to find chemical imitators with similar or even the same properties as the original. Furthermore, it is not necessary to imitate a given probe. If a special property, for example, a color spectrum, is just designed, GA is able to find multicomponent systems adapting this spectrum.

As a necessary condition of the genetic search, it should be mentioned that the designed property must, of course, be incorporated into the search space defined by all offered components. One can only win in a genetic lottery if the winning numbers are within the lottery's range. It is, by far, not trivial to estimate if a solution lies within the search space of a spectrum recovery or not, especially when the spectra are influenced by additional components, such as pH, redox behavior, ligand coordination, and so forth. Even within the described basic examples, the number of possible band combinations are up to the billion range. On the other hand, even if the designed property lies in the search space, the

conditions under which the property could be found are, a priori, not known.

Although objections may be raised against the incompleteness of Table 8, it seems to be reliable that GAs are a universal tool for approximating properties of mixtures, for example, spectra of multicomponent systems. The conclusion is well-founded that the application of GAs offers promising abilities to develop new chemical products with desired properties.

Because of the discussed high potential of GAs in product development, we have extended the pure simulation to a real chemical application. Therein, the simulation kernel of the GA stays completely unchanged. The virtual spectra and concentrations of the simulation have been substituted by real substances and spectrometers to establish a practical chemical application. As a result, we have established a system for the adaptation of dye mixtures to the color of a probe.[24] The system consists of a probe sampler, a UV−vis spectrometer, and a computer with an implementation of the GA. The sampler produces the individual dye mixtures of each generation and performs the transport to the spectrometer. The color of the probe, represented by its spectrum, serves as the desired property, and the spectrum of each individual of a generation determines the individual fitness of the corresponding mixture. All together, it is controlled by the computer's feedback loop, depending on the GA. This in vitro evolution process is still self-organizing and self-optimizing. Because the system is fully automated, external intervention can completely be avoided. The corresponding paper is in preparation.

REFERENCES AND NOTES

(1) *Chemometrics. Data Analysis for the Laboratory and Chemical Plant*; Brereton, R. G., Ed.; John Wiley & Sons: Chichester, U. K., 2003.

(2) *Multivariate Calibration*; Martens, H., Naes, T., Eds.; John Wiley & Sons: Chichester, U. K., 1989.

(3) *Adaption in Natural and Artificial Systems*; Holland, J. H., Ed.; University of Michigan Press: Ann Arbor, MI, 1975.

(4) *Genetic Algorithms in Search, Optimisation and Machine Learning*; Goldberg, D. E., Eds.; Addison-Wesley: Reading, MA, 1989.

(5) Leardi, R. Genetic Algorithms in Chemometrics and Chemistry: A review. *J. Chemom.* **2001**, *15*, 559−569.

(6) 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.

(7) Shao, X.; Sun, L. Resolution of multicomponent overlapping NMR signals using an immune algorithm and genetic algorithm. *Anal. Lett.* **2002**, *35*, 2375−2387.

(8) Dieterle, F.; Kieser, B.; Gauglitz, G. Genetic algorithms and neural networks for the quantitative analysis of ternary mixtures using surface plasmon resonance. *Chemom. Intell. Lab. Syst.* **2003**, *65*, 67−81.

(9) Liu, F.; Wang, J. Using genetic algorithm for quantitative analysis of overlapped spectra in FTIR. *Guangpuxue Yu Guangpu Fenxi* **2001**, *21*, 607−610.

(10) Xu, L.; Zhang, W.-J. Comparison of different methods for variable selection. *Anal. Chim. Acta* **2001**, *446*, 477−483.

(11) Kawakami Harrop Galvao, R.; Fernanda Pimentel, M.; Cesar Ugulino Araujo, M.; Yoneyama, T.; Visani, V. Aspects of the successive projections algorithm for variable selection in multivariate calibration applied to plasma emission spectrometry. *Anal. Chim. Acta* **2001**, *443*, 107−115.

(12) Saldanha, T. C. B.; de Araujo, M. C. U.; de Barros Neto, B.; Chame, H. C. Simultaneous analysis of Co_2^+, Cu_2^+, Mn_2^+, Ni_2^+ and Zn_2^+ in the ultraviolet region using 4-(pyridyl-2-azo)resorcinol and multivariate calibration. *Anal. Lett.* **2000**, *33*, 1187−1202.

(13) Wang, H.; 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 Huaxue* **1999**, *27*, 953−956.

(14) Yang, B.; Wang, H.; Chen, S.; Sha, Z. Simultaneous multicomponent determination by genetic algorithm. *Changchun Keji Daxue Xuebao* **1999**, *29*, 193−196.

(15) Depczynski, U.; Jetter, K.; Molt, K.; Niemoller, A. Quantitative analysis of near-infrared spectra by wavelet coefficient regression using a genetic algorithm. *Chem. Intell. Lab. Syst.* **1999**, *47*, 179−187.

(16) Li, Z.; Zeng, G.; Xia, Z.; Li, M.; Muramatsu, Y.; Wu, X.; Matsumoto, S. Application of genetic algorithm to UV spectroscopy for simultaneous multicomponent analysis of vitamins. *Fenxi Yiqi* **1998**, 46−49.

(17) Xia, Z.; Hu, F.; Qiu, X.; Shi, L.; Li, Z. Genetic algorithms and ultraviolet spectroscopy as applied to multicomponent analysis of amino acids. *Chongqing Daxue Xuebao, Ziran Kexueban* **1998**, *21*, 107−112.

(18) Bohren, A.; Sigrist, M. W. SILC−an algorithm for calibration and analysis of multicomponent absorption spectra with considerable abscissa errors. *Spectrochim. Acta, Part A* **1998**, *54A*, 1049−1058.

(19) Ozdemir, D.; Mosley, M.; Williams, R. Hybrid calibration models: an alternative to calibration transfer. *Appl. Spectrosc.* **1998**, *52*, 599−603.

(20) Arcos, M. J.; Ortiz, M. C.; Villahoz, B.; Sarabia, L. A. Genetic-algorithm-based wavelength selection in multicomponent spectrometric determinations by PLS: application on indomethacin and acemethacin mixture. *Anal. Chim. Acta* **1997**, *339*, 63−77.

(21) Lucasius, C. B.; Beckers, M. L. M.; Kateman, G. Genetic algorithms in wavelength selection: a comparative study. *Anal. Chim. Acta* **1994**, *286*, 135−53.

(22) Lucasius, C. B.; Kateman, G. Genetic algorithms for large-scale optimization in chemometrics: an application. *Trends Anal. Chem.* **1991**, *10*, 254−261.

(23) Bergmann, G.; von Oepen, B.; Zinn, P. Improvement in the Definitions of Sensitivity and Selectivity. *Anal. Chem.* **1987**, *59*, 2522−5227.

(24) Zinn, P. Verfahren und Anlage zur Lösung von Aufgaben der Adaptiven Chemie. Europäische Patentanmeldung PCT/EP 03/07895.

CI049763M