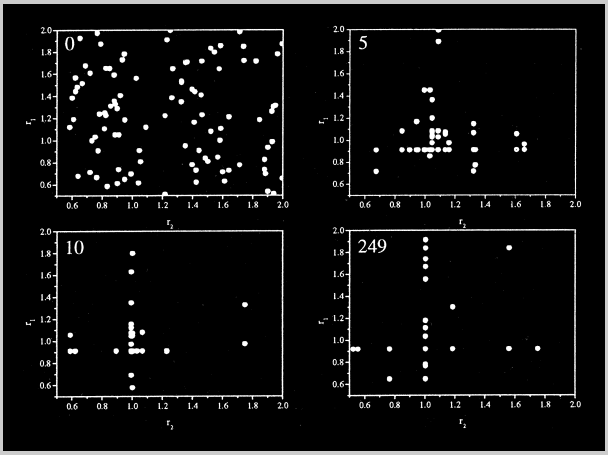
Full Paper: Here we apply evolutionary techniques to the calculation of copolymerization reactivity ratios from an inverse problem perspective. To estimate monomer reactivity ratios, we take into account the main aspects of both inverse problems and evolutionary computation techniques. Copolymers of methyl methacrylate (MMA) and α-tocopheryl methacrylate (MVE) were prepared by free radical copolymerization in dioxane solution using 2,2′-azoisobutyronitrile as the initiator. The reactivity ratios were calculated according to the general copolymerization equation using the Fineman–Röss and Kelen–Tüdos linearization methods, as well as the Tidwell–Mortimer nonlinear least-squares treatment. Reactivity ratios were compared with four different simulations of an evolutionary approach that implements a genetic algorithm. The reactivity ratios obtained with these four simulations were similar, the values being r_{MMA} = 0.92 and r_{MVE} = 1.00. Results obtained with the application of evolutionary techniques demonstrate high-quality solutions and show the convenient use by estimating monomer reactivity ratios in MMA-co-MVE (copolymer of MMA and MVE), a chain addition copolymerization system with potential biomedical applications. The numerous advantages of genetic parameters, performance, and major features of genetic algorithms, are also discussed.



Populations 0, 5, 10 and 249 from the GA2 genetic algorithm.

An Evolutionary Approach to the Estimation of Reactivity Ratios

*Dagmar Monett,**[1] *José A. Méndez,*[2] *Gustavo A. Abraham,*[2] *Alberto Gallardo,*[2] *Julio San Román*[2]

[1] Department of Computer Science, AI Lab., Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany
Fax: (+49 30) 2093 3168; E-mail: diaz@informatik.hu-berlin.de
[2] Instituto de Ciencia y Tecnología de Polímeros, CSIC. J. de la Cierva 3, 28006 Madrid, Spain

Keywords: calculations; copolymerization reactivity ratios; genetic algorithms; simulations

Introduction

The use of novel techniques has become an important means of calculating highly precise monomer reactivity ratios. Without doubt, the reliability of these studies lies obviously on the accuracy of the calculation procedures. Research in this area focuses on copolymer systems by using and comparing several methods to calculate monomer reactivity ratios. However, consensus about which method to apply (with regard to the chemical viewpoint) usually depends on numerous aspects that are related to the method being selected or to domain-dependent problems.

The classical copolymerization model, that is the terminal model, describes the instantaneous copolymer composition by means of the very well-known general copolymerization equation:

$$\frac{F_1}{F_2} = \frac{f_1}{f_2}\,\frac{(r_1 f_1 + f_2)}{(r_2 f_2 + f_1)} \tag{1}$$

or by means of the following equivalent expression:

$$F_1 = \frac{(r_1 - 1)f_1^2 + f_1}{(r_1 + r_2 - 2)f_1^2 + 2(1 - r_2)f_1 + r_2} \tag{2}$$

where F_1 and f_1 are the overall molar fractions of the monomer (1) in the copolymer and in the feed, respectively, whereas r_1 and r_2 are the reactivity ratios of monomer (1) and of monomer (2), respectively. r_1 and r_2, are determined typically by using this instantaneous copolymerization equation and analyzing series of low conversion polymerizations. Traditional methods (such as Finemann–Röss and Kelen–Tüdos procedures) are based on

  1022-1344/2002/0506–0525$17.50+.50/0

linear transformations of this equation. However, it is well established that these linear approaches are statistically invalid, although they can be used to obtain some good initial r_1 and r_2 estimates (in which case the Kelen–Tüdos method is superior to the Finemann–Röss). Nonlinear regression methods (Tidwell–Mortimer is a good example) are nowadays recommended.

So far, the main goal is to estimate kinetic values for a theoretical model and to find a method with which the unknown values can be determined. We propose to analyze this problem by considering inverse modeling and parameter estimation concepts, and by studying its solution through evolutionary approaches. With this purpose, the estimation of reactivity ratios is first posed to optimization procedures because we are interested in reactivity values such that the theoretical copolymer compositions fit measurements of copolymer compositions as well as possible. Then, we calculate the solution to the optimization by introducing a novel approach for copolymerizations based on *evolutionary techniques*.

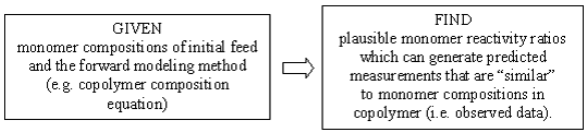
Evolutionary algorithms are *probabilistic algorithms based on principles of evolution and heredity*. Genetic algorithms (GAs) are one type of probabilistic algorithm.[1] GAs have already been successfully applied to optimization problems. Thus, GAs are considered to estimate copolymerization reactivity ratios. Their major features (e.g., genetic parameters, fitness functions and performance) are described in this paper. Finally, we present and discuss an application for the methyl methacrylate/α-tocopheryl methacrylate (MMA-*co*-MVE) copolymer system in order to show how GAs can be used for copolymerizations.

Theory

Inverse Modeling and Parameter Estimation Concepts

The estimation of copolymerization reactivity ratios is an example of a *coefficient inverse problem* from mathematics also known as *classical parameter estimation problem*: constant multipliers (i.e., reactivity ratios) in a governing model should be found. The general nature of an inverse problem is to determine unknown causes based on observations of their effects. This means that we might try to infer values for the reactivity ratios by observing the copolymer compositions based on the initial feed compositions.

The inversion problem for copolymerization systems can be described as seen in Scheme 1. A solution to this problem can be formulated as finding a least squares fit of a set of predicted measurements to the monomer compositions in copolymer over time, or a minimization with respect to the monomer reactivity ratios. Accordingly to this, let us define the function ψ: $\Re^m \rightarrow \Re$, with m being the number of experiences over time, to be the l_2-norm of



Scheme 1. The inversion problem for copolymerization systems.

the difference between predicted and observed monomer compositions in copolymer as follows:

$$\psi^2 = \| F^{\text{pred}} - F^{\text{obs}} \|_2 = \sum_{i=1}^{m} (F_i^{\text{pred}} - F_i^{\text{obs}})^2 \tag{3}$$

where F^{pred} are the predicted data or theoretical composition values calculated by forward modeling (i.e., by considering the copolymer composition equation), and F^{obs} are the observed data or experimental measurements of monomer compositions in copolymer. The Euclidean distance is the square root of the l_2-norm (i.e., ψ). It is the measure of model fitting we use.

Let P be the parameter vector containing the parameters to be estimated. The dimension of P is the number of parameters to be estimated (e.g., 2 for a free radical copolymerization of two monomers). The system could be described as a vector function s, i.e., $s(P, F^{\text{pred}}) = 0$. Following our assumptions above, the vector P contains the monomer reactivity ratios, and F^{pred}, the molar fractions of the monomers calculated with the values of P.

In practice, observed measurements F^{obs} are such that $F^{\text{obs}} = F^{\text{pred}} + \varepsilon$. By substituting this expression in Equation (3) we obtain:

$$\psi = \sqrt{\sum_{i=1}^{m} \varepsilon_i^2} \tag{4}$$

The values of ε_i might indicate that measurement errors are present, such as errors when reading scales which are almost unavoidable. Consequently, an absolute accuracy of the measurements is not always fulfilled.

Our purpose is to estimate P using F^{obs}. The function $s(P, F^{\text{obs}}) = 0$ is not valid any more, since $\varepsilon \neq 0$. In this case, we should instead use a function such as:

$$\sigma(P, F_1^{\text{obs}}, F_2^{\text{obs}}, ..., F_m^{\text{obs}}) \tag{5}$$

that has to be optimized using the Equation (4).

Optimization Techniques

The inversion problem can be formulated in terms of an optimization, namely

$$\text{minimize } \sigma, \text{ subject to } \psi < \delta \tag{6}$$

Table 1. Some disadvantages when using optimization methods.

Source	Problems
Optimization methods	Traditional optimization methods: an initial approach for the global optimum finding is needed. In some cases, the optimum is not a global one and strongly depends on the initial approach. Such methods that use derivatives, gradients and similar expressions, cannot always identify the global optimum.
Real systems	Sometimes, analytical information about the problem being solved (like derivatives) is not available. Noisy and time-varying objective function values are present (regarding measurements, for example). Also, the set of feasible solutions is only a subset of the domain of the variables (e.g., restrictions concerning practical situations like avoiding reactivity ratios with negative values).

where $\delta \in \Re$ is an accuracy level for the acceptance of the solutions. The idea here is to find a combination of parameters on P such that the values of ε_i are close to zero, which implies that ψ is also close to zero. However, optimization problems are very difficult to deal with: for most optimization methods any of the situations from Table 1 is present. Furthermore, one of the ways to approach these problems is by using exact methods but their computational complexity makes them really intractable.

Experiments with copolymers are in general resource consuming (e.g., chemical substances and labor materials) and usually require expensive, complex equipment. Moreover, it is difficult to have the same conditions to repeat similar reactions (i.e., environment, people, products). Also, if a procedure of calculating monomer reactivity ratios is needed, the following situations are present most of the time: (i) an appropriate method to be used should be selected; (ii) the user should know its characteristics or theory; (iii) he/she should know how to use the method (e.g., simulations by means of computers); (iv) he/she should have a particular implementation at hand if needed.

Unfortunately, some other disadvantages might also arise:

- Real problems (i.e., copolymerizations) are complex and cannot always be symbolically represented (i.e., as complete as possible through an exact theory).
- Popular techniques that are still used have no statistical validity and should not be applied in further quantitative uses.[2]
- Other methods transform the original problem mathematically, deriving a not entirely satisfactory prediction of their properties.
- Conventional techniques based on compositional analysis of initial feed and copolymer formed might consider other problems caused by conversion requirements.[3]
- Different experimental designs might lead to different results for the same system being analyzed.
- Experimental errors with regard to equipment precision or human interaction might be presented.

Thus, the application of other approaches where the above problems are not present becomes a necessity. Evolutionary algorithms have emerged as effective and efficient methods to improve solutions by optimization. For these reasons, they have been successfully applied to such domains. Particularly, GAs, a special kind of evolutionary approach, have been well adapted to optimization problems and their solutions outperform other numerical approaches most of the times.

Genetic Algorithms (GAs)

The essential idea of GAs is to perform a search on a space of potential solutions represented by a *population* of individuals. Individuals undergo probabilistic *genetic operators* such as *recombination* and *mutation* to simulate analogous processes in nature. A fitness function evaluates the "quality" of individuals and the best ones are *selected* to *survive*. When a stop criterion is satisfied, then the more adapted individual represents the best solution for the problem being solved.

In terms of GAs suited to optimization procedures, minimizing the function σ from Equation (6) is equivalent to obtain *individuals* with high *fitness* values. Then, the aim of the genetic search is to find a parameter combination so that the forward modeling fits the observed data as closely as possible. With this purpose, a transformation such as the following could be used,

$$\xi = 1/(\psi + K) \tag{7}$$

where $\xi : \Re \rightarrow \Re$ is a monotonically decreasing function corresponding to the fitness of an individual, and K is a convenient parameter to avoid indeterminations. In this way, the values of ψ and ξ are inversely proportional, whereas a value of K close to zero maintains the denominator values close to ψ, thus considering the simplest transformation $\xi = 1/\psi$. Then, the following statement yields:

$$\text{minimize } \sigma \Longleftrightarrow \text{maximize } \xi \tag{8}$$

which means that individuals with high fitness values will have the smallest error values. Thus, they will be the more adapted from the population.

The pseudo-code in Figure 1 presents these ideas by considering two non-overlapping populations P and P' (P stands for the current population and P' for the surviving one). The basic genetic cycle works as follows: first the initial population P is created at random (there are also other ways to initialize populations), the fitness of indivi-

```
begin GA;
 t = 0;                          // Initializing time
 // Initialize an usually random population of individuals:
 random P( t );
 // Evaluate the fitness of all individuals:
 evaluate P( t );
 // Calculate statistics:
 statistics P( t );
 while not done do   // Testing for termination criterion
   t = t + 1;                    // Increasing time
   // Select a sub-population for offsprings production:
   P' = select P( t );
   // Stochastically perturb genes of selected parents
   // and recombine them:
   recombine P'( t );
   // Evaluate the new fitnesses:
   evaluate P'( t );
   // Select the survivors for next generations:
   P = survive P( t ), P'( t )
   // Calculate statistics:
   statistics P( t );
 endwhile;
end GA;
```

Figure 1. Pseudo-code of a genetic algorithm.

duals is evaluated, and general statistics of the process are calculated (e.g., best individual up to now). Then, a new population P' is created by selecting individuals from the old one. There are a number of ways to implement selection mechanisms. For example, following a *Roulette Wheel* procedure, individuals are chosen in proportion to their fitness values: the higher the fitness, the higher the possibility to be selected. Selected individuals undergo mutations (new genetic information is introduced) and crossover operators (genetic information is exchanged among parents). The fitness of new individuals and statistics are once again calculated. The process ends when a termination criterion is satisfied (e.g., while at least one individual satisfying Equation (6) is found).

General Features of GAs

Encoding Mechanism

The encoding of individuals refers to how the genetic information is stored. For instance, two of the most applied mechanisms make use of bit strings (bits of 0's and 1's) or real values to represent the *genes* on a *chromosome*. We assume the second approach which make use of real genes as in previous studies:[4] a chromosome is a vector of g real genes, each gene varying in a real interval and representing possible values for the monomer reactivity ratios.

Let N_{Iter} be the total number of GA iterations (i.e., how many times the "while" cycle from Figure 1 is executed). The population, $P(t) = v_1^t, \ldots, v_n^t$ $(1 \leq t \leq N_{\text{Iter}}, t \in \aleph)$, comprises n individuals v_i^t $(1 \leq t \leq n, i \in \aleph)$ that represent possible solutions of the problem for each GA iteration t. The number of genes on a free radical copolymerization is equal to 2 ($g = 2$) and we can assume that

$$v_i^t = r_i^t = (r_{i_1}^t, r_{i_2}^t) \quad r_{i_j}^t \in \Re, \quad j = 1, 2 \tag{9}$$

Consequently, intervals that parameters belong to are defined as follows:

$$r_{i_j}^t \in I_j = [I_{i_j}, u_{i_j}] \quad l_{i_j}, u_{i_j} \in \Re \tag{10}$$

where l and u denote appropriate lower and upper limits for each interval I, respectively.

For example, a population might have 50 individuals or 50 pairs of monomer reactivity ratios ($n = 50$), each one being a probable solution to the problem. In this case, the chromosomes belong to a product set, denoted by $I_1 \times I_2$, and consist of the search space or the widest region where the monomer reactivity ratios are defined. The major question is which pair results in a better approximation to the measurement values. It can be assumed that reasonable estimates of monomer reactivity ratios are available to serve as starting points for the construction of the intervals. Although the genetic search does not necessarily need initial estimates, appropriate values from other methods can be also used.

Fitness Function

For copolymerizations, the fitness function could be related to Equation (7). The function ψ corresponds to the difference ("distance") between predicted and observed monomer compositions in copolymer as mentioned above. Then, individuals (or pairs of monomer reactivity ratios) that generate minimal solutions for the distance (i.e., ψ) are the more adapted in the population and have a greater fitness (i.e., ξ).

Genetic Operators

As we already discussed, genetic operators are applied to the entire population. The most commonly used ones are crossover (i.e., genetic information between parents is exchanged to produce new offsprings) and mutation (i.e., genetic information is stochastically perturbed, thus new genetic material is introduced). For example, Figure 2 shows the simplest crossover operator and how genetic material is exchanged between two individuals. In contrast, mutation operators perturb genes only on a chromosome, e.g., by considering new values on their intervals.

Genetic operators are applied depending on probability values. Both crossover and mutation probabilities are expected to be numbers between 0 and 1. This means, if a randomly generated number in $[0,1]$ is less than the crossover probability, then the parents can recombine their chromosomes. Otherwise, the parents remain the same in the next generation. Mutation probability behaves similar. Examples, implementations, and theory about genetic operators have been extensively described in the literature.[1,5]

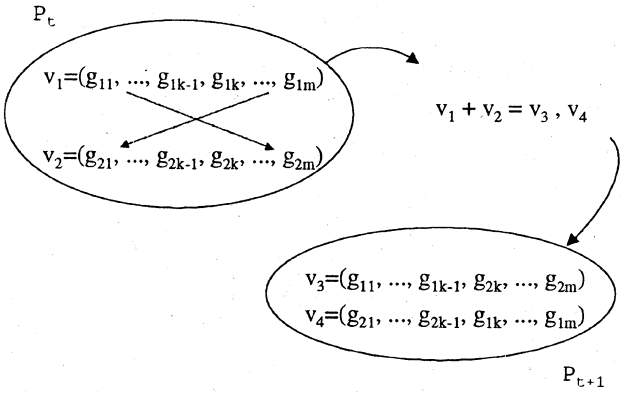


Figure 2. Simple one-point crossover on GAs. When crossing their genes, parents v_1 and v_2 produce offsprings v_3 and v_4 ("+" stands for the crossover operator).

Other GA Features

Sometimes it is desirable to maintain the best already obtained individuals to survive in further populations. Elitist techniques are conceived with this purpose. Besides, different kinds of GAs have also been studied, as have replacement strategies to replace old individuals by the fittest ones, hybrid GAs (e.g., GAs combined with other local optimization procedures), parallel GAs (to exploit the advantages of implicit parallelism on GAs as well as to model more complex populations), etc. Many systems, programs, as well as implementations codes for GAs are available.[1, 6]

Experimental Part

Materials

α-Tocopherol (Vitamin E) was supplied by Merck. Methacryloyl chloride (Acros) was distilled before use, triethylamine (Fluka) was refluxed over potassium hydroxide, and purified by distillation. The purified product was stored over molecular sieve 4 Å exhaustively dried at vacuum. MMA was supplied by Merck and was distilled after filtration though a column of alumina. 2,2′-Azobis(isobutyronitrile) was purified by fractional crystallization from methanol, m.p. = 104 °C. 1,4-Dioxane was dried over sodium hydroxide for 24 h and then was distilled at 104 °C. Other pure-grade reagents were used as purchased.

Copolymerization

MVE was synthesized as described in a previous paper.[7] Briefly, *α*-tocopherol was reacted with methacryloyl chloride in presence of triethylamine as catalyst. The product was purified by dissolving it in toluene and was passed through a chromatographic column containing silica gel (Merck 0.040–0.063 mm thickness), using a chloroform/hexane mixture 9:1 as mobile phase. The purified fraction was analyzed by thin layer chromatography, using the same eluent and pre-coated silica gel sheets supported on aluminium (Merck) as the stationary phase.

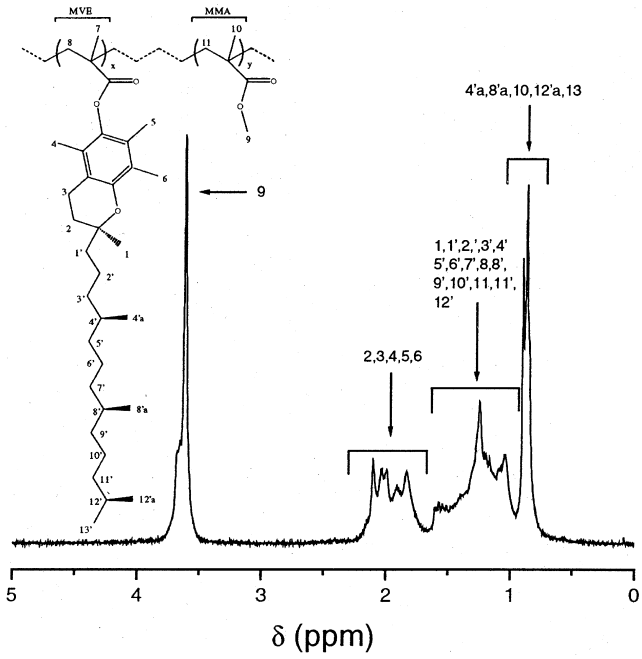


Figure 3. Chemical structure of MMA-co-MVE and ^{1}H NMR spectrum (F_1 = 0.872).

The copolymerization reaction was performed in dioxane solution at 50 °C in Pyrex glass flasks under purified nitrogen atmosphere. Comonomer and initiator concentrations were 1 mol · L^{-1} and 1.5 wt.-%, respectively. After selected reaction times, the flask contents were poured into a large excess of methanol. The precipitated copolymer was filtered and dried under vacuum until constant weight was attained. All the copolymer systems isolated have molecular weight $\overline{M}_n$ higher than 20 kg · mol^{-1} and polydispersity index of 2.5-2.8, determined by size exclusion chromatography. The molar fraction of monomers units incorporated in the copolymers was determined by ^{1}H NMR spectroscopy (Gemini 200). The spectra were recorded at room temperature in 10% w/v $CDCl_3$ solutions (an example is shown in Figure 3 together with the chemical structure). The analysis was performed by the integration of the characteristic resonance signals assigned to the methoxy protons CH_3O— (3.40–3.85 ppm) of MMA structural unit, denoted as A_{OCH_3}, and all the signals associated to MVE and the rest of signals of MMA (0.50–2.40 ppm), denoted as A_t. The molar fraction of MMA (F_1) was calculated by using the following equation derived from the chemical structure of the copolymer:

$$F_1 = \frac{\frac{A_{OCH_3}}{3}}{\frac{A_{OCH_3}}{3} + \left(\frac{A_t - \left(\frac{A_{OCH_3}}{3} \cdot 5\right)}{54}\right)} \qquad (11)$$

Genetic Algorithm Settings

Four different computational simulations with GAs (GA1, GA2, GA3, and GA4) were run. Their basic genetic features

remained the same except when considering their termination criteria and the intervals from Equation (10), where the individuals vary. In GA1, the first simulation, the threshold level for the acceptance of the solutions (i.e., δ from Equation (6)) was set to a value of the Euclidean distance less than the obtained with the linearization methods. In this case, each reactivity ratio varied in [0.0, 10.0] representing a first wide region of definition of the genes. In GA2, the threshold remained the same as in GA1. However, intervals for the parameters were reduced to [0.5, 2.0] by considering the zones where individuals from GA1 were concentrated. The threshold in GA3 was set to the fitness value of the best individual from GA2. The intervals, in this case, were reduced to [0.9, 0.95] and [0.9, 1.1] expecting a more approximated solution. Similarly, the threshold in GA4 was set to the fitness value of the best individual from GA3. In the fourth simulation the intervals from GA3 were preserved without change.

Populations overall comprised 100 uniformly distributed individuals, which were initialized at the first generation using the Mersenne-Twister pseudo random number generator.[8] The selection of individuals used a fitness-based tournament. The crossover operator was the one-point simple arithmetic crossover with probability $p_c = 0.8$, while the mutation operator was the uniform mutation with probability $p_m = 0.1$.

In our simulations, the performance of the GAs was evaluated on a Sun Ultra 5/10 computer with an UltraSPARC-IIi 400 MHz processor, 131072 KB of RAM, and SunOS Release 5.6 Version Generic-105181-19 operating system.

Results and Discussion

Different derivatives of vitamin E are known for their antioxidant properties, which have been related to the preservation of cells and can be used as anti-aging agents.[9] Recently, several reports were devoted to the synthesis and characterization of novel polymeric systems bearing vitamin E[7, 10–12] due to their important biomedical application. The copolymers of MVE with MMA have been developed to incorporate this antioxidant additive to acrylic bone cement formulations in order to avoid the cytotoxicity associated to the residual monomers. In this sense, the understanding of the copolymerization and the determinations of the reactivity ratios are key issues in the design of this material.

The copolymerization of MMA (monomer 1) with MVE (monomer 2) in dioxane solutions was studied in a wide range of compositions with feed molar fractions (f_1) ranging from 0.3 to 0.9 (feed molar fractions of 0.2 or lower do not polymerize in the conditions used in this work, probably due to the radical-capture effect of the vitamin E derivative). The reaction time was initially regulated to reach conversions lower than 10 wt.-%, in order to satisfy the differential copolymerization equation.[13] The data of molar composition of the initial comonomer mixture used and the obtained copolymers are summarized in Table 2 (determined as described in experimental). From this Table, approximated values of the reactivity ratios r_1 and r_2 were obtained by means of the Fineman–Röss[14] and Kelen–Tüdos[15] linearization methods, as well as Tidwell–Mortimer[16] nonlinear least-squares treatment. Approximations for the reactivity ratios were also selected among the more adapted individuals from the GA simulations. The scores presented in Table 3 correspond to all the reactivity ratios mentioned above. The most reliable reactivity ratios were found to be close to 1 for both comonomers, that means the monomer distribution in the copolymer is expected to be random and the reaction is not far from the ideal copolymerization. These values have been drawn in Figure 4 together with the 95% confidence limits. As expected, the Finemann–Röss linearization gives the poorest estimation, followed by the Kelen–Tüdos method.

Table 2. Composition data of the free radical copolymerization of MMA and MVE (where f_1 and F_1 are the molar fractions of MMA in the monomer feed and in the copolymer, respectively).

Feed[a] f_1	Copolymer[a] F_1	Conversion %
0.9000	0.872	7
0.9004	0.867	6
0.7998	0.800	8
0.8004	0.793	9
0.7009	0.678	9
0.7007	0.707	7
0.5998	0.606	7
0.5998	0.589	9
0.4998	0.508	8
0.5017	0.469	10
0.3995	0.386	9
0.4003	0.376	9
0.3036	0.304	8
0.3032	0.297	8

[a] The copolymer data error is associated to the NMR measurement while the feed data error is related to the more accurate weighting procedure. This is the reason for the decimal discrepancy between both columns.

Table 3. Calculated reactivity ratios for the composition data of the MMA-*co*-MVE copolymerization system.

Method	Reactivity ratios		
	r_1	r_2	$r_1 \cdot r_2$
Fineman–Röss	0.70 ± 0.03	0.61 ± 0.14	0.43
Kelen–Tüdos	0.80 ± 0.12	0.86 ± 0.09	0.69
Tidwell–Mortimer	0.92	1.00	0.92
GA1 (1576)[a]	0.92	1.00	0.92
GA2 (249)	0.92	1.00	0.92
GA3 (100)	0.92	1.00	0.92
GA4 (64)	0.92	1.00	0.92

[a] The number of GA iterations where the best values were achieved are given in parentheses.

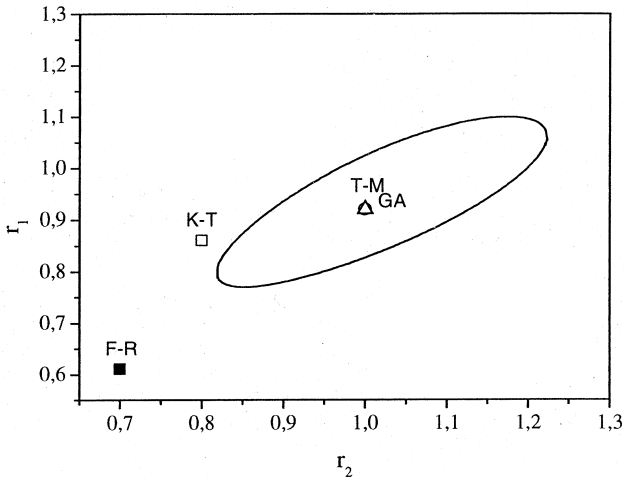


Figure 4. Reactivity ratios r_1 and r_2 obtained by linear, nonlinear least-squares and GA4 methods.

Table 4. Error estimations by different numerical criteria.

Method	Criterion		
	l_2-norm[a)]	Euclidean distance[b)]	Quadratic mean deviation[c)]
Fineman–Röss	1.5600×10^{-2}	1.24901×10^{-1}	1.11430×10^{-3}
Kelen–Tüdos	4.6950×10^{-3}	6.85202×10^{-2}	3.35358×10^{-4}
Tidwell–Mortimer	3.2407×10^{-3}	5.69274×10^{-2}	2.31481×10^{-4}
GA1	3.2408×10^{-3}	5.69285×10^{-2}	2.31490×10^{-4}
GA2	3.2408×10^{-3}	5.69276×10^{-2}	2.31482×10^{-4}
GA3	3.2408×10^{-3}	5.69278×10^{-2}	2.31484×10^{-4}
GA4	3.2407×10^{-3}	5.69274×10^{-2}	2.31481×10^{-4}

a) $l_2\text{-norm} = \psi^2$.
b) $l_2\text{-norm}^{1/2} = \psi$.
c) $l_2\text{-norm/m} = \psi^2/m$, where m is the number of experimental data.

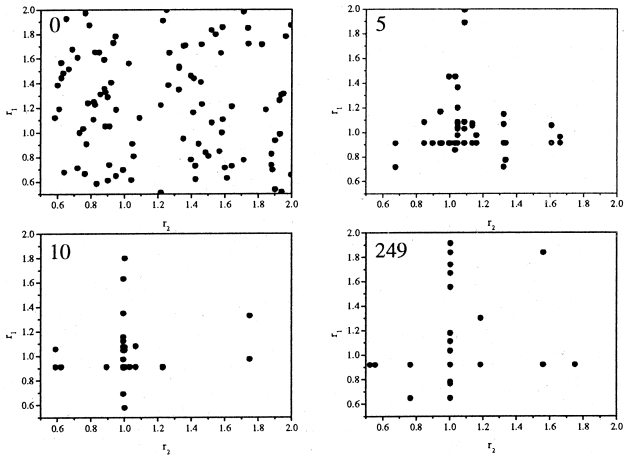


Figure 5. Populations 0, 5, 10 and 249 from the GA2 genetic algorithm.

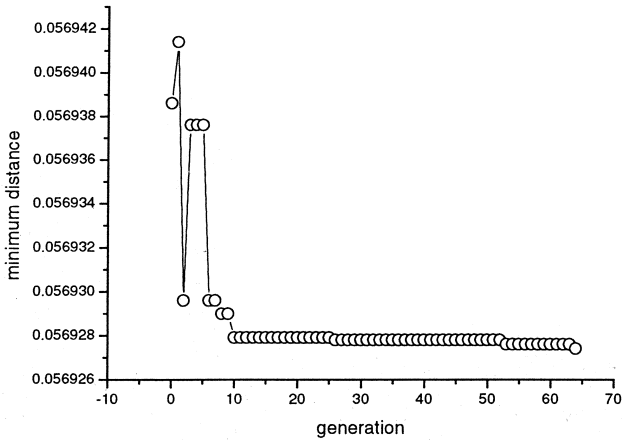


Figure 6. Minimum distance ψ versus generation t for the GA4 genetic algorithm.

On the other hand, the evolutionary approach proposed in this work is able to select reliable pairs of data, as good as those obtained by the nonlinear regression. In this sense, Table 4 shows the error estimations for each method by considering the most frequently used numerical techniques to assess them: the l_2-norm, the Euclidean distance, and the quadratic mean deviation. As can be seen in Table 4, the GA simulations outperform the distances between theoretical and experimental data. For example, a quadratic mean deviation of 2.31481×10^{-4} for the GA4 shows a great precision. Looking at the error estimations, we can infer that general copolymer properties could be predicted in a precise way under an evolutionary performance.

Figure 5 shows populations at generations 0, 5, 10 and 249 for the GA2 simulation. In the interest of brevity, intermediate generations have not been displayed. At the start of the process (population 0) there are many different chromosomes. Crossover plays an important role here, when the population is diverse: individuals exchange their characteristics producing new offsprings with similar genetic information. However, the population converges to a set of very similar chromosomes: overlapping genes of individuals appears (note the superposition of points in populations 5, 10 and 249). This means that crossover has little effect in advanced populations. However, as chromosomes start to converge, mutation helps to preserve a reasonable level of diversity. This is why, in later generations, the mutation operator is more effective producing a population that contains individuals with new genetic material, which occasionally contributes to the best solution for the problem. In particular, the tendency is to create two well-defined lines in advanced populations: a horizontal line that contains the best estimations for r_1, and a vertical one with the best estimations for r_2. Note that the point where both lines intercept at population 249 represents the best chromosome (i.e., the best solution).

According to our expectations, the minimum distances per generation in all GA simulations decreased. As an example, Figure 6 shows how the minimum distance var-

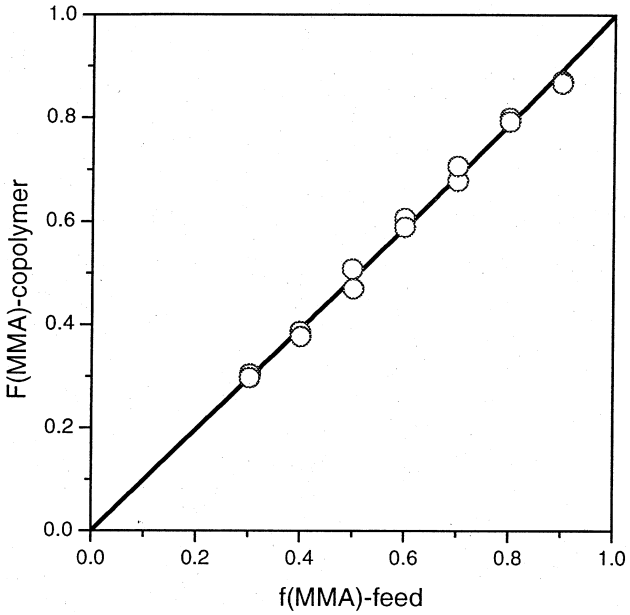


Figure 7. Theoretical composition diagram calculated using r_1 = 0.92 and r_2 = 1.00 in the copolymerization equation, together with the experimental data.

ies with the number of generations for the GA4, where the best individual was found. Note that the value of the minimum distance decreases drastically at first generations (until generation 10); however, as generations continue, the minimum distance converges to the best solution at generation 64. These behavior could also be explained by considering the effect of the genetic operators mentioned before.

The average composition diagram shown in Figure 7 has been drawn with the data obtained from the GA4 simulation using the Lewis-Mayo classical copolymerization equation (i.e., Equation (2)). In addition, the experimental composition data obtained at low conversion (Table 2) which fit the theoretical diagram adequately are also depicted.

Conclusions

Genetic algorithms (GAs) are excellent evolutionary methods for the accurate determination of reactivity ratios in addition polymerization reactions. The application of GAs to the determination of r_1 and r_2 for the free radical copolymerization of MMA and MVE gives the best optimized approach when considering a GA involving just one-point simple arithmetic crossover and uniform mutation. This approach has proved to be powerful enough and the termination criteria is always satisfied for the selected class of problems. As a useful and effective method for any coefficient inverse problem, we are currently applying this methodology to high conversion copolymerization data.

Acknowledgement: We wish to acknowledge the scholarship support A/99/03570 from the *German Academic Exchange Service* (DAAD) for a research period to *D. Monett* at the *Institute of Informatics*, *Humboldt University of Berlin*. *J. A. Méndez* thanks to CICYT for the fellowship awarded to him. *G. A. Abraham* is grateful to MEC of Spain which provided a grant for his stay in Spain. Financial support from the *Comisión Interministerial de Ciencia y Tecnología*, CICYT (MAT 2001-1618) is also gratefully acknowledged.

Received: January 1, 2002
Revised: April 14, 2002
Accepted: April 19, 2002

[1] D. E. Goldberg, "*Genetic Algorithms in Search, Optimization, and Machine Learning*", Addison-Wesley, Reading, MA 1989.
[2] K. F. O'Driscoll, P. M. Reilly, *Makromol. Chem., Macromol. Symp.* **1987**, *10/11*, 355.
[3] R. van der Meer, H. N. Linssen, A. L. German, *J. Polym. Sci., Polym. Chem. Ed.* **1978**, *16*, 2915.
[4] D. Monett, in: *Proceedings of the GECCO'01 Graduate Student Workshop*, Genetic and Evolutionary Computation Conference (GECCO 2001), San Francisco, CA, USA 2001, p. 429.
[5] Z. Michalewiçz, "*Genetic Algorithms + Data Structures = Evolution Programs*", 3rd edition, Springer-Verlag, Berlin 1999.
[6] E. Schöneburg, F. Heinzmann, S. Feddersen, "*Genetische Algorithmen und Evolutionsstrategien*", Addison-Wesley, Bonn 1994.
[7] C. Ortiz, B. Vázquez, J. San Román, *Polymer* **1998**, *39*, 4107.
[8] M. Matsumoto, T. Nishimura, in: *ACM Transactions on Modeling and Computer Simulations*: *Special Issue on Uniform Random Number Generation*, January 1998, vol. 8, no. 1, p. 3.
[9] K. K. Chan, N. Cohen, J. P. De Noble, A. C. Specian, Y. G. Jr. Saucy, *J. Org. Chem.* **1976**, *41*, 3497.
[10] A. Bulai, C. Ortiz, A. Gallardo, A. Lozano, J. San Román, *Polymer* **1998**, *24*, 6163.
[11] B. Vázquez, C. Ortiz, J. San Román, M. A. Plasencia, A. López Bravo, *J. Biomater. Appl.* **2000**, *14*, 367.
[12] J. A. Méndez, M. R. Aguilar, G. A. Abraham, B. Vázquez, M. Dalby, L. Di Silvio, J. San Román, *J. Biomed. Mater. Res.* **2002**, *in press*.
[13] A. E. Hamielec, J. F. Macgregor, A. Penlidis, in: *Comprehensive Polymer Science*, G. Allen, J. C. Bevington, C. Booth, C. Price, Eds., Pergamon Press, Oxford 1989, vol. 3, p. 17.
[14] M. Fineman, S. D. Ross, *J. Polym. Sci.* **1950**, *5*, 259.
[15] T. Kelen, R. Tüdos, *J. Macromol. Sci.* **1975**, *A9*, 1.
[16] P. W. Tidwell, G. A. Mortimer, *J. Polym. Sci.* **1965**, *A3*, 369.