Macromolecules **1991,** *24,* 5051-5057

Determination of Microstructure in Copolymers. Statistical Modeling and Computer Simulation of Mass Spectra

Maurizio S. Montaudo

Dipartimento di Matematica, Università di Catania, Viale Andrea Doria, 6-95125 Catania, Italy

Alberto Ballistreri and Giorgio Montaudo*

Dipartimento di Scienze Chimiche, Università di Catania, and Istituto per la Chimica e la Tecnologia dei Materiali Polimerici, Consiglio Nazionale delle Ricerche, Viale Andrea Doria, 6-95125 Catania, Italy

Received February 22, 1991; Revised Manuscript Received April 9, 1991

ABSTRACT: The recent finding that significant information on copolymer microstructure can be obtained by using mass spectrometry has stimulated our work toward developing appropriate algorithms to match the experimental mass spectra to those theoretically calculated. A novel program for the determination of microstructure in copolymers by computer simulation of mass spectra is reported. The statistical model used to describe copolymer distributions is reported in detail in the text. The computer simulation program (MACO3) is discussed, and three illustrative examples are given.

Introduction

The characterization of sequence arrangements in condensation polymers having large comonomer subunits cannot be easily achieved by current methods, including NMR, that otherwise have proven to be of general utility in the case of vinyl, olefin, and diene copolymers.[1-4]

Mass spectrometry, being able to look at the mass of individual molecules in a mixture of homologues, represents an interesting alternative, and the technique of fast atom bombardment (FABMS) has been applied to the study of the microstructure of numerous synthetic and biological copolymers.[5] The assumption on which the sequential analysis of a copolymer is based is that ions detected in FABMS originate from species already present in the polymeric sample and that fast atom bombardment does not produce fragmentation of the macromolecules.[6] FAB spectra do not yield direct structural information on large macromolecules, but this can be deduced from the structural analysis of relatively low molecular mass species. Therefore, in order to yield significant information before undergoing MS analysis, the macromolecules must be subjected to partial degradation to obtain smaller molecules.[6]

Recently FABMS has been successfully used to identify the oligomers formed in the partial degradation of several copolymers, and the results obtained have yielded detailed information on the microstructure of the materials investigated.[5,7-12] The finding that significant information on copolymer microstructure can be obtained using mass spectrometry has triggered our interest in developing suitable algorithms that would allow calculation of the copolymer sequence distribution starting from the experimental mass spectra, analogous to what is currently done in the case of NMR spectra.[1-4] This project is rendered even more attractive by the consideration that, in the case of condensation polymers, mass spectra often "see" longer sequences than NMR,[5-12] and therefore the microstructure can be defined with more accuracy. In previous work[11] we have described the best fit of the calculated abundances with the experimental mass spectra, confining ourselves to the case of poly(β-hydroxybutyrate-*co*-β-hydroxyvalerate) of bacterial origin, a two-component copolymer that we have assumed has a random distribution of units along the chain. This initial success has prompted us to extend the simulation of mass spectra to include other statistical models and also to deal with copolymers with a higher number of components in their structure.

Although the mathematical formalism is derived essentially from that used for the NMR,[1-4] a certain degree of manipulation has been necessary in order to extend it and to render it directly applicable to the simulation of mass spectra. In this respect, it constitutes a novel contribution to the literature. We report here the statistical models and describe a novel simulation program of mass spectra based on this formalism: mass analysis of copolymers (MACO3),[13] which is applicable to multicomponent copolymers. It can discriminate between Markovian and Bernoullian distributions (and therefore it can reveal random or block sequences) and is also able to distinguish a pure copolymer from a binary mixture of copolymers. To illustrate briefly the method, we have also included here three examples of mass spectra simulations. Further examples, concerning for instance the microstructure of copolymers generated by exchange reactions in polymer blends, shall be presented later.

Principle of the Method

The method that we have adopted here for the analysis of mass spectra of copolymers is very similar to that used for analyzing materials by NMR.[3,4] However, there are some important differences due to the peculiar instrumental aspects inherent in the two techniques. For example, $5!/(3!2!) = 10$ sequences corresponds to the mass number of copolymer segment A_3B_2, and in NMR all 10 contributions have, at least in principle, different chemical shifts. On the other hand, only one peak is detected in MS, and in general the intensity of each MS peak is the sum of different contributions, since there are different segment sequences that correspond to the same mass number. A peak deconvolution procedure to obtain the single contributions is therefore necessary for the MS data, and this might appear as a serious difficulty. However, the partial degradation of copolymers has recently been shown[5-12] to be able to provide mass spectra that contain mixtures of several oligomers up to high masses, and this proved to be sufficient to allow a meaningful sequential analysis of copolymers by MS.[5-12] But in the majority of

0024-9297/91/2224-5051$02.50/0

condensation polymers, NMR can hardly discern beyond the diad and triad level, and often the instrumental resolution does not match that needed in order to detect as separate signals all the existing sequence arrangements of monomeric units. Therefore, the number of NMR peaks available for sequential analysis is usually low and confined to segments (diad and triads) bearing a low discriminating power with respect to establishing the distribution of comonomer units present in the copolymer.

In practice, provided that a sufficient number of peaks corresponding to different oligomers appear in the spectra, NMR and MS spectra are both suitable for determining the microstructure of copolymers.[14]

The intensities of both NMR and MS peaks depend on the copolymer composition (i.e., the relative abundances of the components) and on the type of distribution along the chain. Therefore, assuming a theoretical distribution model and then fitting the calculated oligomer abundances with the experimental peak intensities, the copolymer composition can be determined.

The statistical analysis of copolymers makes use of Bernoullian and Markovian statistics in order to calculate and to characterize the microstructure of copolymer samples.[1-4,11,14] Bernoullian models describe a random distribution of monomer units, where the probability of monomer addition is independent of the outcome of any previous addition. The first-order Markovian model is more general, since it describes distributions that may result when the probability of monomer addition is dependent upon a single preceding event, and can be applied to discriminate among block, alternating, and random copolymers. In some cases it is also necessary to distinguish if a copolymer sample consists of a single copolymer or is actually a mixture of two copolymers (this question may often rise when evaluating commercial samples that may have been mixed to achieve certain properties or when dealing with copolymers of bacterial origin where the bacteria have been fed with two or more different carbon sources).[5,10,11,14] In order to cope with this problem, one has to use statistical models that are able to discriminate a single Bernoullian (or Markovian) copolymer from a mixture of two Bernoullian (or two Markovian) copolymers. To discriminate among the possible distributions in a copolymer, one must therefore generate the theoretical oligomer distribution arising from each one of the three statistical models discussed above and compare them with the experimental MS spectra.

In the following section are discussed the mathematical models that we have used in our work.

Statistical Models

Copolymers are formed from two components in many cases but may be made up of several components, as in biological macromolecules. Since the first-order Markovian model represents the most general case (the Bernoullian model can be called a zero-order Markov), it appears appropriate to start by defining this model for a multicomponent copolymer and to derive the others from it by imposing successive simplifications. A Markov distribution is completely defined when the transition probability matrix (**P** matrix) is given. The **P** matrix is a square matrix of dimension N, where N is the number of different components that occur in the chain. Each matrix element P_{ij} describes the probability of finding the component "i" after the component "j". For example, a Markov distribution for a four component copolymer (A–D) is defined when the 16 **P** matrix elements P_{AA}, P_{AB}, P_{AC}, P_{AD}, P_{BA}, P_{BB}, P_{BC}, P_{BD}, P_{CA}, P_{CB}, P_{CC}, P_{CD}, P_{DA}, P_{DB}, P_{DC}, and P_{DD} are given. The intensity of the MS peak associated with the hypothetical sequences AABBACD, AABBCAD, ABABDAC is

$$\begin{aligned} I_{AABBACD} &= hS_AP_{AA}P_{AB}P_{BB}P_{BA}P_{AC}P_{CD} \\ I_{AABBCAD} &= hS_AP_{AA}P_{AB}P_{BB}P_{BC}P_{CA}P_{AD} \\ I_{ABABDAC} &= hS_AP_{AB}P_{BA}P_{AB}P_{BD}P_{DA}P_{AC} \end{aligned} \tag{1}$$

where the factor h depends on the instrumental response and where S_A, the probability that the sequence starts with A, is given by

$$\begin{aligned} S_A = {} & (-P_{BA}P_{CB}P_{DC} + P_{BA}P_{DB}P_{CC} - P_{BA}P_{DB} + \\ & P_{CA}P_{BB}P_{DC} - P_{CA}P_{DB}P_{BC} - P_{CA}P_{DC} - P_{DA}P_{BB}P_{CC} + \\ & P_{DA}P_{BB} + P_{DA}P_{CB}P_{BC} + P_{DA}P_{CC} - P_{DA})/(P_{AA}P_{BB}P_{CC} - \\ & P_{AA}P_{BB}P_{DC} - P_{AA}P_{BB} - P_{AA}P_{CB}P_{BC} + P_{AA}P_{CB}P_{DC} + \\ & P_{AA}P_{DB}P_{BC} - P_{AA}P_{DB}P_{CC} + P_{AA}P_{DB} - P_{AA}P_{CC} + \\ & P_{AA}P_{DC} + P_{AA} - P_{BA}P_{AB}P_{CC} + P_{BA}P_{AB}P_{DC} + P_{BA}P_{AB} + \\ & P_{BA}P_{CB}P_{AC} - P_{BA}P_{CB}P_{DC} - P_{BA}P_{DB}P_{AC} + P_{BA}P_{DB}P_{CC} - \\ & P_{BA}P_{DB} + P_{CA}P_{AB}P_{BC} - P_{CA}P_{AB}P_{DC} - P_{CA}P_{BB}P_{AC} + \\ & P_{CA}P_{BB}P_{DC} + P_{CA}P_{DB}P_{AC} - P_{CA}P_{DB}P_{BC} + P_{CA}P_{AC} - \\ & P_{CA}P_{DC} - P_{DA}P_{AB}P_{BC} + P_{DA}P_{AB}P_{CC} - P_{DA}P_{AB} + \\ & P_{DA}P_{BB}P_{AC} - P_{DA}P_{BB}P_{CC} + P_{DA}P_{BB} - P_{DA}P_{CB}P_{AC} + \\ & P_{DA}P_{CB}P_{BC} - P_{DA}P_{AC} + P_{DA}P_{CC} - P_{DA} - P_{BB}P_{CC} + \\ & P_{BB}P_{DC} + P_{BB} + P_{CB}P_{BC} - P_{CB}P_{DC} - P_{DB}P_{BC} + \\ & P_{DB}P_{CC} - P_{DB} + P_{CC} - P_{DC} - 1) \end{aligned} \tag{1'}$$

Therefore, assuming given transition probabilities, P_{ij}, eq 1 can be used to calculate the mass number and the corresponding intensity of all possible copolymer segments. If two or more sequences happen to have the same mass number, they will contribute to the intensity of the same peak, and the corresponding theoretical intensities must be summed (a computer can automatically perform these operations by iterating over different sequences).

All **P** matrix elements P_{ij} must be positive; furthermore, they must satisfy the following property:

$$\sum_{i=1}^{N} P_{ij} = 1 \qquad j = 1, 2, 3, ..., N \tag{2}$$

In a similar way, the sum of the probabilities that the sequence starts with A, S_A, starts with B, S_B, starts with C, S_C, and starts with D, S_D, must obey the condition

$$S_A + S_B + S_C + S_D = 1 \tag{3}$$

The latter equation, combined with the relationships

$$\begin{aligned} S_AP_{AA} + S_BP_{BA} + S_CP_{CA} + S_DP_{DA} &= S_A \\ S_AP_{AB} + S_BP_{BB} + S_CP_{CB} + S_DP_{DB} &= S_B \\ S_AP_{AC} + S_BP_{BC} + S_CP_{CC} + S_DP_{DC} &= S_C \end{aligned}$$

constitutes a system of four linear equations in four unknowns from which the above eq 1′ for S_A can be derived.

Let us designate by $\rho(A)$, $\rho(B)$, $\rho(C)$, and $\rho(D)$ the relative abundances of components A–D in the copolymer. The probability that the sequence starts with a certain component is directly proportional to the relative abundance of that component. If concentrations are expressed as molar fractions (so that $\rho(A) + \rho(B) + \rho(C) + \rho(D) = 1$), one obtains

$$S_A = \rho(A) \qquad S_B = \rho(B) \qquad S_C = \rho(C) \qquad S_D = \rho(D) \tag{4}$$

It is of great interest to calculate, by means of eq 1, the theoretical intensities of the MS peak associated with long sequences of like monomers and to compute the number-average lengths of like monomers,[1] which are a measure the degree of "blockiness" of a certain copolymer with respect to one of its components. In our four-component copolymer example the number-average lengths of component A, $\langle n_A \rangle$, of component B, $\langle n_B \rangle$, of component C, $\langle n_C \rangle$, and of component D, $\langle n_D \rangle$ are defined by

$$\langle n_A \rangle = (\sum_i iI_{(A)_i})/(\sum_i I_{(A)_i}) \quad \langle n_B \rangle = (\sum_i iI_{(B)_i})/(\sum_i I_{(B)_i})$$
$$\langle n_C \rangle = (\sum_i iI_{(C)_i})/(\sum_i I_{(C)_i}) \quad \langle n_D \rangle = (\sum_i iI_{(D)_i})/(\sum_i I_{(D)_i}) \tag{5}$$

In the case of a two-component (AB) copolymer, a Markov distribution needs only four **P** matrix elements, P_{AA}, P_{AB}, P_{BA}, and P_{BB}, to be defined. The intensity of the MS peaks associated with the sequences BAABAABAA and ABBBBBBBB is

$$I_{BAABAABAA} = hS_B P_{BA} P_{AA} P_{AB} P_{BA} P_{AA} P_{AB} P_{BA} P_{AA}$$
$$I_{ABBBBBBBB} = hS_A P_{AB} P_{BB} P_{BB} P_{BB} P_{BB} P_{BB} P_{BB} P_{BB} \tag{6}$$

$$S_A = P_{BA}/(P_{AB} + P_{BA}) \qquad S_B = P_{AB}/(P_{AB} + P_{BA}) \tag{7}$$

Equation 6, therefore, corresponds to eq 1 for the two-component case, and a set of equations similar to eqs 2–5 can be written, allowing the computation of theoretical peak intensities. In particular, expressions for $\langle n_A \rangle$ and $\langle n_B \rangle$ can be written, after a simple calculation,[1] in the form

$$\langle n_A \rangle = 1/P_{AB} \qquad \langle n_B \rangle = 1/P_{BA} \tag{8}$$

Let us now proceed from the Markovian to a simpler distribution. Bernoullian statistics can be considered as a particular case of Markovian statistics when the probability of monomer addition is proportional to monomer relative abundance; that is, when

$$P_{AA} = k\rho(A), \quad P_{AB} = k\rho(B), \quad P_{AC} = k\rho(C), \quad P_{AD} = k\rho(D)$$
$$P_{BA} = k\rho(A), \quad P_{BB} = k\rho(B), \quad P_{BC} = k\rho(C), \quad P_{BD} = k\rho(D)$$
$$P_{CA} = k\rho(A), \quad P_{CB} = k\rho(B), \quad P_{CC} = k\rho(C), \quad P_{CD} = k\rho(D)$$
$$P_{DA} = k\rho(A), \quad P_{DB} = k\rho(B), \quad P_{DC} = k\rho(C), \quad P_{DC} = k\rho(D) \tag{9}$$

(here k is a proportionality factor). Inserting eq 9 in eq 1, we obtain the intensity of MS peaks associated with sequences AABBACD, AABBCAD, and ABABDAC in the Bernoullian distribution chain model:

$$I_{AABBACD} = h[\rho(A)]^3[\rho(B)]^2[\rho(C)][\rho(D)]$$
$$I_{AABBCAD} = h[\rho(A)]^3[\rho(B)]^2[\rho(C)][\rho(D)]$$
$$I_{ABABDAC} = h[\rho(A)]^3[\rho(B)]^2[\rho(C)][\rho(D)] \tag{10}$$

Equation 10 requires explicitly that all the sequences associated to one mass number produce spectral signals having the same intensity, and, therefore, the total peak intensity must arise from the sum of equal contributions. This peculiar property allows computation of theoretical mass spectra using a procedure that iterates over different mass numbers, instead of iterating over different sequences as in the Markovian case (see above). In order to implement such a procedure, one must compute the number (Γ) of distinct sequences associated with the same mass number. For an oligomer $A_mB_nC_pD_q$, a calculation due to Leibniz shows that

$$\Gamma = \frac{(m + n + p + q)}{m!n!p!q!} \tag{11}$$

Therefore, for four components and $h = 1$ we obtain

$$I_{A_mB_nC_pD_q} = \frac{(m + n + p + q)!}{m!n!p!q!}[\rho(A)]^m[\rho(B)]^n[\rho(C)]^p \times [\rho(D)]^q \tag{12}$$

Each oligomer peak in the MS spectrum is identified by the indexes (m, n, p, q), and the intensity of such a peak is computed by using eq 12 in the hypothesis of a four-component copolymer that has a Bernoullian distribution of comonomer units.

In the case of a two-component copolymer, the intensity of the MS peak due to A_mB_n oligomers is given by the well-known Newton formula:

$$I_{A_mB_n} = \frac{(m + n)!}{m!n!}[\rho(A)]^m[\rho(B)]^n \tag{13}$$

We now turn to the mathematical model valid for the case of a mixture of two copolymers, say, C1 and C2. Let us suppose that the mixture is made of a fraction X of copolymer C1 and a fraction $1 - X$ of copolymer C2 and that one or both chains obey first-order Markov statistics. We call N the largest number between the number of components of C1 and C2, $P_{ij}^{(1)}$ the **P** matrix elements that define the Markov distribution in C1, and $P_{ij}^{(2)}$ the **P** matrix elements in C2. Since the overall probability must give 1, we have

$$\sum_{i=1}^{N} P_{ij}^{(1)} = 1, \quad \sum_{i=1}^{N} P_{ij}^{(2)} = 1 \quad j = 1, 2, 3, ..., N$$

Let us focus on four-component (ABCD) copolymers. The intensity of the MS peak associated with the sequence AABBACD is

$$I_{AABBACD} = h[XS_A^{(1)}P_{AA}^{(1)}P_{AB}^{(1)}P_{BB}^{(1)}P_{BA}^{(1)}P_{AC}^{(1)}P_{CD}^{(1)} + (1 - X)S_A^{(2)}P_{AA}^{(2)}P_{AB}^{(2)}P_{BB}^{(2)}P_{BA}^{(2)}P_{AC}^{(2)}P_{CD}^{(2)}] \tag{14}$$

where $S_A^{(1)}$ is the probability that a generic sequence starts with A in the distribution defined by the matrix $P^{(1)}$ while $S_A^{(2)}$ is the corresponding probability in the distribution defined by the matrix $P^{(2)}$. If $\rho_1(A),\rho_1(B)\rho_1(C)\rho_1(D)$ is the copolymer C1 composition and $\rho_2(A),\rho_2(B)\rho_2(C)\rho_2(D)$ is the composition of copolymer C2, then the following relations hold:

$$\rho_1(A) = S_A^{(1)}, \quad \rho_1(B) = S_B^{(1)}, \quad \rho_1(C) = S_C^{(1)}, \quad \rho_1(D) = S_D^{(1)}$$
$$\rho_2(A) = S_A^{(2)}, \quad \rho_2(B) = S_B^{(2)}, \quad \rho_2(C) = S_C^{(2)}, \quad \rho_2(D) = S_D^{(2)}$$

Formula 14 can be used to compute mass number and corresponding intensities of all possible sequences, thus allowing the generation of a theoretical mass spectrum.

Let us consider the case when the chain of copolymer C1 and the chain of copolymer C2 can both be described by two Bernoullian distributions. In this case the intensity of the mass spectrum peak associated with the oligomer

Table I
Experimental[a] and Calculated[b] Relative Amounts of the Methanolysis Products from Sample 1,[10] Contained in Some HPLC Peaks

oligomer	m/z	obsd[a] from FABMS	calcd[b] for HV/HH/HN/HU mole ratio 2/26/70/2	2/31/65/2	2/33/63/2	0/40/60/0
			Peak 15			
C_{36} tetramers	679	27	45	35	29	21
C_{41} pentamers	779	51	43	48	50	55
C_{46} hexamers	879	19	11	15	18	22
C_{51} heptamers	979	3	1	2	3	2
AF			0.345	0.146	0.077	0.132
			Peak 16			
C_{43} pentamers	807	50	63	54	50	42
C_{48} hexamers	907	39	31	37	39	45
C_{53} heptamers	1007	11	6	9	11	13
AF			0.255	0.083	0.009	0.155
			Peak 17			
C_{45} pentamers	835	20	34	25	22	12
C_{50} hexamers	935	50	47	48	48	50
C_{55} heptamers	1035	24	16	22	24	31
C_{60} octamers	1135	6	3	5	6	7
AF			0.292	0.101	0.047	0.174
			Peak 18			
C_{52} hexamers	963	42	54	44	43	30
C_{57} heptamers	1063	42	35	40	41	47
C_{62} octamers	1163	14	9	14	14	20
C_{67} nonamers	1263	2	2	2	2	3
AF			0.234	0.084	0.052	0.233
			Peak 19			
C_{54} hexamers	991	18	27	18	16	7
C_{59} heptamers	1091	42	47	46	44	42
C_{64} octamers	1191	26	21	27	29	37
C_{69} nonamers	1291	12	4	8	9	12
C_{74} decamers	1391	2	1	1	1	2
AF			0.244	0.106	0.097	0.284
AF global			0.279	0.101	0.062	0.198

[a] Relative intensities of MNa^+ ions in the FABMS spectrum.[10] [b] Relative intensities of the methanolysis products calculated by MACO3, applying subroutine BERNGEN (eq 12 in the text).

$A_mB_nC_pD_q$ will be the weighted sum of the two contributions, say, $I^{(1)}$ and $I^{(2)}$, arising, respectively, from copolymer C1 and C2:

$$I_{A_mB_nC_pD_q} = XI^{(1)} + (1 - X)I^{(2)} \quad (15)$$

Expressions for $I^{(1)}$ and $I^{(2)}$ can be easily obtained from eq 12, and, substituting these expressions in eq 15, we get

$$I_{A_mB_nC_pD_q} = \frac{(m+n+p+q)!}{m!n!p!q!}\{X[\rho_1(A)]^m[\rho_1(B)]^n[\rho_1(C)]^p \times [\rho_1(D)]^q + (1-x)[\rho_2(A)]^m[\rho_2(B)]^n[\rho_2(C)]^p[\rho_2(D)]^q\} \quad (16)$$

This equation is valid for a Bernoullian sample made of a binary mixture of copolymers, each copolymer having four components.

Equation 16 reduces to the simple form

$$I_{A_mB_n} = \frac{(m+n)!}{m!n!}\{X[\rho_1(A)]^m[\rho_1(B)]^n + (1-X)[\rho_2(A)]^m \times [\rho_2(B)]^n\} \quad (17)$$

in the case of a binary mixture of two-component copolymers that follow the Bernoullian model.

MACO3 Computer Program[13]

Formulas 1, 5, 6, 12–14, 16, and 17 have been implemented in a computer program, MACO3,[13] written in Standard ANSI Fortran that runs on a MICROVAX computer. The program is structured in a main program and various subroutines. In particular, (a) subroutine MARKGEN is based on eq 1 and deals with multicomponent copolymers that follow Markov statistics, (b) subroutine MARKTWO is based on eq 6 and deals with two-component copolymers that follow Markov statistics, (c) subroutine BERNGEN is based on eq 12 and deals with multicomponent copolymers that follow Bernoullian statistics, (d) subroutine BERNTWO is based on eq 13 and deals with two-component copolymers that follow Bernoullian statistics, (e) subroutine BINMIXGEN is based on eq 16 and deals with binary mixtures of copolymers that follow multicomponent Bernoullian statistics, (f) subroutine BINMIXTWO is based on eq 17 and deals with binary mixtures of copolymers that follow two-component Bernoullian statistics, (g) subroutine BINMARK is based on eq 14 and deals with binary mixtures of copolymers that follow multicomponent Markov statistics, (h) subroutine BEST-FIT calls another subroutine, MINPACK1-LMDIF1, belonging to the Argonne Library (MINPACK1-LMDIF1 is a minimization subroutine), and (i) subroutine NAV computes number-average lengths of like monomers, according to formula 5. MACO3 accepts as input the number of components, N, the chosen mathematical model, and the kind of computation (i.e., THEORY or MINIMUM) that must be performed. When user selects the option THEORY, MACO3 asks for the parameters that define the chosen distribution. Once these are given, MACO3 calls one of the above subroutines that generate mass spectra theoretical intensities and displays the result. When the user selects the option MINIMUM, the program asks for the experimental MS spectrum and stores given data. Then it calls subroutine BEST-FIT, which starts generating many different theoretical distributions by varying the parameters that define the distribution and by calling the subroutine that corresponds to the chosen model. For each set of parameters the agreement factor, AF, between experimental and theoretical intensity

Table II
Experimental[a] and Calculated[b] Relative Amounts of the Methanolysis Products from Sample 2,[10] Contained in Some HPLC Peaks

oligomer	*m/z*	obsd[a] from FABMS	calcd[b] for HV/HC/HH/HO/HN//HD/HU mole ratio		
			2/6/35/40/2/1	2/6/16/42/28/4/2	1/5/16/25/50/2/1
			Peak 22		
C_{33} tetramers	637	75	83	74	86
C_{37} pentamers	723	23	16	24	13
C_{41} hexamers	809	2	1	2	5
AF			0.144	0.021	0.190
			Peak 23		
C_{34} tetramers	651	34	47	36	55
C_{38} pentamers	737	23	19	24	16
C_{39} pentamers	751	35	30	33	26
C_{42} hexamers	823	2	1	2	1
C_{43} hexamers	837	6	3	5	2
AF			0.285	0.064	0.435
			Peak 24		
C_{35} tetramers	665	22	40	30	51
C_{40} pentamers	765	63	53	57	44
C_{44} hexamers	851	15	7	13	5
AF			0.316	0.150	0.520
			Peak 25		
C_{36} tetramers	679	7	14	11	26
C_{41} pentamers	779	49	53	45	49
C_{45} hexamers	865	16	11	16	9
C_{46} hexamers	879	24	20	25	15
C_{50} heptamers	965	4	2	3	1
AF			0.181	0.110	0.388
			Peak 26		
C_{42} pentamers	793	45	59	48	66
C_{47} hexamers	893	47	37	44	31
C_{51} heptamers	979	8	4	8	3
AF			0.278	0.073	0.413
			Peak 27		
C_{43} pentamers	807	25	38	27	49
C_{48} hexamers	907	51	42	42	37
C_{52} heptamers	993	9	7	11	5
C_{53} heptamers	1007	13	12	18	9
C_{57} octamers	1093	2	1	2	0
AF			0.272	0.177	0.481
			Peak 28		
C_{44} pentamers	821	10	23	17	35
C_{49} hexamers	921	56	53	49	48
C_{54} heptamers	1021	29	22	29	16
C_{58} octamers	1107	5	2	5	1
AF			0.250	0.159	0.460
			Peak 29		
C_{45} pentamers	835	6	9	8	17
C_{50} hexamers	935	45	50	39	51
C_{55} heptamers	1035	37	30	34	24
C_{59} octamers	1121	5	4	7	3
C_{60} octamers	1135	7	7	12	5
AF			0.150	0.146	0.317
			Peak 30		
C_{51} hexamers	949	34	43	32	53
C_{56} heptamers	1049	49	43	45	36
C_{61} octamers	1149	13	13	20	10
C_{65} nonamers	1235	4	1	3	1
AF			0.191	0.144	0.378
AF global			0.235	0.122	0.403

[a] Relative intensities of MNa^+ ions in the FABMS spectrum.[10] [b] Relative intensities of the methanolysis products calculated by MACO3, applying subroutine BERNGEN (eq 12 in the text).

is computed according to the formula

$$\mathrm{AF} = \left(\frac{\sum_i (Y_i^{\mathrm{exp}} - Y_i^{\mathrm{calc}})^2}{\sum_i [Y_i^{\mathrm{exp}}]^2} \right)^{1/2} \tag{18}$$

The process continues until the set of parameters that gives the minimum AF is found. The best fit composition is computed and displayed, together with the corresponding AF. The capabilities of MACO3 have been tested on several sets of experimental data. Some illustrative examples are discussed in the following section.

Microstructure of Copolymers by Simulation of Mass Spectra: Some Examples

As a first example of copolymer microstructure determination, we have chosen a four-component copolyester

Table III
Experimental[a] and Calculated[b] Relative Amounts of the Methanolysis Products from Sample 3,[8] Contained in Some HPLC Peaks

oligomer	m/z	obsd[a] from FABMS	calcd[b] for $P_{HB,HB} = 0.9152$ $P_{HV,HB} = 0.5000$ $P_{HB,HV} = 0.0848$ $P_{HV,HV} = 0.5000$	calcd[b] for $P_{HB,HB} = 0.855$ $P_{HV,HB} = 0.855$ $P_{HB,HV} = 0.145$ $P_{HV,HV} = 0.145$
Peak 15				
$(HB)_8$	743	35	53	33
$(HB)_6(HV)$	671	42	26	45
$(HB)_4(HV)_2$	599	20	15	19
$(HB)_2(HV)_3$	527	3	6	3
AF			0.415	0.088
Peak 16				
$(HB)_9$	829	29	50	28
$(HB)_7(HV)$	757	41	26	44
$(HB)_5(HV)_2$	685	24	16	23
$(HB)_3(HV)_3$	613	6	8	5
AF			0.469	0.069
Peak 17				
$(HB)_{10}$	915	20	46	24
$(HB)_8(HV)$	843	42	27	43
$(HB)_6(HV)_2$	771	27	18	26
$(HB)_4(HV)_3$	699	11	9	7
AF			0.571	0.114
Peak 18				
$(HB)_{11}$	1001	22	41	21
$(HB)_9(HV)$	929	36	27	40
$(HB)_7(HV)_2$	857	27	18	29
$(HB)_5(HV)_3$	785	12	10	9
$(HB)_3(HV)_4$	713	3	4	1
AF			0.442	0.119
Peak 19				
$(HB)_{12}$	1087	21	38	18
$(HB)_{10}(HV)$	1015	37	27	38
$(HB)_8(HV)_2$	943	27	19	31
$(HB)_6(HV)_3$	871	12	11	11
$(HB)_4(HV)_4$	799	3	5	2
AF			0.411	0.109
Peak 20				
$(HB)_{13}$	1173	5.1	4.5	5.0
$(HB)_{11}(HV)$	1101	9.1	8.7	11.9
$(HB)_9(HV)_2$	1029	8.5	6.6	10.8
$(HB)_7(HV)_3$	957	5.9	4.1	4.7
$(HB)_5(HV)_4$	885	2.6	2.1	10.0
$(HB)_3(HV)_5$	813	1.0	0.8	0.1
$(HB)_{14}$	1259	4.5	9.7	4.3
$(HB)_{12}(HV)$	1187	9.1	8.5	11.0
$(HB)_{10}(HV)_2$	1115	9.1	6.7	11.1
$(HB)_8(HV)_3$	1043	7.0	4.4	5.5
$(HB)_6(HV)_4$	971	3.5	2.4	1.4
$(HB)_4(HV)_5$	899	1.5	1.1	0.1
$(HB)_2(HV)_6$	827	1.2	0.4	0
$(HB)_{15}$	1345	4.8	9.6	3.7
$(HB)_{13}(HV)$	1273	8.6	8.3	10.1
$(HB)_{11}(HV)_2$	1201	9.1	6.9	11.2
$(HB)_9(HV)_3$	1129	6.5	4.6	6.3
$(HB)_7(HV)_4$	1057	2.6	2.6	1.9
AF			0.431	0.250
Peak 21				
$(HB)_{16}$	1431	13	28	10
$(HB)_{14}(HV)$	1359	30	26	28
$(HB)_{12}(HV)_2$	1287	32	22	34
$(HB)_{10}(HV)_3$	1215	21	15	21
$(HB)_8(HV)_4$	1143	4	9	7
AF			0.395	0.098
Peak 22				
$(HB)_{17}$	1517	13	29	9
$(HB)_{15}(HV)$	1445	31	28	29
$(HB)_{13}(HV)_2$	1373	36	25	37
$(HB)_{11}(HV)_3$	1301	20	18	25
AF			0.358	0.141
AF global			0.443	0.114

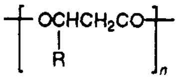
[a] Relative intensities of MNa^+ ions in the FABMS spectrum.[10] [b] Relative intensities of the methanolysis products calculated by MACO3, applying subroutine MARKTWO (eq 6 in the text).

Table IV
Experimental[a] and Calculated[b] Relative Amounts of the Partial Pyrolysis (TG–FAB) Products from Sample 3[11]

oligomer	m/z	obsd[a] from FABMS	calcd[b] for $P_{HB,HB} = 0.9152$ $P_{HV,HB} = 0.5000$ $P_{HB,HV} = 0.0848$ $P_{HV,HV} = 0.5000$	calcd[b] for $P_{HB,HB} = 0.855$ $P_{HV,HB} = 0.855$ $P_{HB,HV} = 0.145$ $P_{HV,HV} = 0.145$
$(HB)_2$	143	69	78	73
$(HB)(HV)$	171	27	15	25
$(HV)_2$	199	4	7	2
AF			0.214	0.068
$(HB)_3$	215	64	72	63
$(HB)_2(HV)$	243	30	17	32
$(HB)(HV)_2$	271	6	8	5
$(HV)_3$	299	0	3	0
AF			0.221	0.034
$(HB)_4$	287	55	66	54
$(HB)_3(HV)$	315	32	19	36
$(HB)_2(HV)_3$	343	10	10	9
$(HB)(HV)_3$	371	3	4	1
$(HV)_4$	399	0	1	0
AF			0.264	0.077
$(HB)_5$	359	45	60	46
$(HB)_4(HV)$	387	35	20	39
$(HB)_3(HV)_2$	415	16	11	13
$(HB)_2(HV)_3$	443	4	5	2
$(HB)(HV)_4$	471	0	2	0
AF			0.367	0.086
$(HB)_6$	431	40	55	39
$(HB)_5(HV)$	459	38	21	40
$(HB)_4(HV)_2$	487	17	13	17
$(HB)_3(HV)_3$	515	5	6	4
$(HB)_2(HV)_4$	543	0	5	0
AF			0.399	0.040
AF global			0.249	0.055

[a] Relative intensities of $(M - H)^-$ ions in the FABMS spectrum.[11] [b] Relative intensities of partial pyrolysis products calculated by MACO3, applying subrouting MARKTWO (eq 6 in the text).

sample of bacterial origin studied by means of FABMS[10] (sample 1). This copolyester has the structure



$$\left[\mathrm{OCHCH_2CO} \right]_n \text{ with } \mathrm{R} \text{ on CH}$$

containing units with higher *n*-alkyl pendant groups (R), which are identified by the following abbreviations: HV = β-hydroxyvalerate, HH = β-hydroxyheptanoate, HN = β-hydroxynonanoate, and HU = β-hydroxyundecanoate.

The experimental technique used[10] to study the sample was the following: the sample was subjected to partial degradation by methanolysis, the oligomers were separated by HPLC, and each chromatographic fraction was analyzed by FABMS.

The second and third columns of Table I show the experimental data reported[10] for each chromatographic fraction. These data were analyzed by means of the program MACO3. The statistical model selected as a first hypothesis was a pure Bernoullian copolymer made of four components, and therefore the Leibniz formulas (eqs 11 and 12) had to be used. The program performed a minimization, and the result was a best fit composition of $\rho(HV) = 2$, $\rho(HH) = 33$, $\rho(HN) = 63$, and $\rho(HU) = 2$ with a global agreement factor AF = 0.062. This overall composition is consistent with the experimental results.[10] The theoretical spectrum that corresponds to the above composition is reported in Table I (column six). Some theoretical spectra have been generated from compositions different from the best fit. The results (reported in Table I, columns four, five, and seven) are self-explanatory.

Thereafter, we simulated a binary mixture of two Bernoullian copolymers, each having four components (eq 16). The set of parameters that gave the minimum AF was found to be

$$\rho_1(\mathrm{HV}) = 2,\ \rho_1(\mathrm{HH}) = 33,\ \rho_1(\mathrm{HN}) = 62,\ \rho_1(\mathrm{HU}) = 2;\quad X = 0.30$$

$$\rho_2(\mathrm{HV}) = 3,\ \rho_2(\mathrm{HH}) = 32,\ \rho_2(\mathrm{HN}) = 63,\ \rho_2(\mathrm{HU}) = 2;\quad 1 - X = 0.70$$

with an agreement factor AF = 0.061. These figures identify two essentially identical copolymer compositions for the mixture hypothesized. Furthermore, this composition is the same as that found for a Bernoullian copolymer (Table I, column one), and we concluded that the more complex model is not necessary and that a description of sample 1 in terms of a pure Bernoullian copolymer made of four components is completely satisfactory.

As a second example, we report the statistical analysis of another copolyester of bacterial origin[10] (sample 2), which is a seven-component copolymer having the structure given above. Following the nomenclature used above we shall refer to the components as HV–HC–HH–HO–HN–HD–HU (HC = β-hydroxycaproate, HO = β-hydroxyoctanoate, HD = β-hydroxydecanoate).

The experimental technique used[10] to study the sample was the following: the sample was subjected to partial degradation by methanolysis, the oligomers were separated by HPLC, and each chromatographic fraction was analyzed by FABMS.

The first three columns of Table II show the experimental data.[10] We chose a pure Bernoullian distribution having seven components; therefore, the Leibniz formulas (eqs 11 and 12) had to be used. A minimization was carried over, and the result was a best fit composition of $\rho(\mathrm{HV}) = 2$, $\rho(\mathrm{HC}) = 6$, $\rho(\mathrm{HH}) = 16$, $\rho(\mathrm{HO}) = 42$, $\rho(\mathrm{HN}) = 28$, $\rho(\mathrm{HD}) = 4$, and $\rho(\mathrm{HU}) = 2$, with an agreement factor AF = 0.122. This overall composition is consistent with the experimental results.[10] In Table II (column five) is reported the theoretical spectrum corresponding to the best fit composition. The latter is considered very satisfactory, as compared to the numbers in columns four and six in Table II, which correspond to slightly different compositions. Due to the complexity of the copolymer structure (seven components), no attempt was made to simulate other statistical models.

FABMS data on bacterial copolyesters have been obtained in the literature,[8,10,11] by two methods. In the first method,[8,10] the partially degraded copolymer sample was fractionated by HPLC and then subjected to FABMS analysis. Although very reliable, this is a time-consuming procedure and an alternative (to avoid HPLC fractionation) has been tried,[11] and the mixture of oligomers generated in the partial pyrolysis was directly subjected to FABMS analysis.[11] Two independent sets of data, collected by HPLC–FABMS[8] and TG–FABMS,[11] respectively, are available in the case of PHB-*co*-PHV copolymer (HB = β-hydroxybutyrate, HV = β-hydroxyvalerate; sample 3), and we therefore applied our simulation method to the two groups of experimental MS data.

Although it is known that sample 3 follows a Bernoullian distribution,[11] we selected a two-component first-order Markov distribution. Since the experimental composition[8,11] of sample 3 is $\rho(\mathrm{HV}) = 85.5$ and $\rho(\mathrm{HB}) = 14.5$, we considered two different **P** matrices, P_α and P_β, respectively

$$P_\alpha \rightarrow P_{\mathrm{HB,HB}} = 0.9152,\ P_{\mathrm{HB,HV}} = 0.0848,\quad P_{\mathrm{HV,HB}} = 0.5000,\ P_{\mathrm{HV,HV}} = 0.5000$$

$$P_\beta \rightarrow P_{\mathrm{HB,HB}} = 0.8550,\ P_{\mathrm{HB,HV}} = 0.1450,\quad P_{\mathrm{HV,HB}} = 0.8550,\ P_{\mathrm{HV,HV}} = 0.1450$$

which simulate two copolymers having the same composition (as one can verify by substituting in eq 7 values for P_α and for P_β) but different distribution. In fact, inserting these figures in eq 8, one finds that number-average lengths for P_α ($\langle n_\mathrm{A}\rangle = 11.7$, $\langle n_\mathrm{B}\rangle = 2.0$) are almost double the number-average lengths for P_β ($\langle n_\mathrm{A}\rangle = 6.89$, $\langle n_\mathrm{B}\rangle = 1.17$). Furthermore, the matrix P_α corresponds to a nondegenerate Markovian distribution, bu P_β does not. In fact, P_β satisfies eq 10, which identifies a Bernoullian distribution.

In Table III we report a comparison between experimental HPLC–FABMS data[8] and computed MS data (obtained using eq 6) for the two distributions. The computed intensities for P_α show a poor agreement with experimental data, as underlined by the high value of the agreement factor AF = 0.250. Instead, the computed MS peak intensities for P_β show a much better AF. We conclude that the simulation method is reliable and that it is able to select between different statistical models.

In Table IV we report experimental TG–FABMS intensities[11] concerning sample 3 together with the theoretical intensities (generated using eq 6 for P_α and for P_β). From a comparison between figures in Tables III and IV, it can be seen that in both cases the agreement for **P** matrix P_β (generating a Bernoullian distribution) turns out to be at least a factor of 4 better than the AF value for P_α, thus demonstrating the consistency of the two groups of data and the power of our computer simulation method.

Acknowledgment. Partial financial support from the Italian Ministry for University and for Scientific and Technological Research (MURST) and from the National Council of Research (CNR) (Rome), Finalized Project of fine and secondary chemistry, is gratefully acknowledged.

References and Notes

(1) Randall, J. C. *Polymer Sequence Determination*; Academic Press: New York, 1977; and references therein.
(2) Bovey, F. A. *Acc. Chem. Res.* **1968**, *1*, 175.
(3) Price, F. P. *J. Chem. Phys.* **1962**, *36*, 209.
(4) Tonelli, A. E. *NMR Spectroscopy and Polymer Microstructure*; VCH Publishers: New York, 1989.
(5) Montaudo, G. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 95.
(6) Ballistreri, A.; Garozzo, D.; Giuffrida, M.; Montaudo, G. *Anal. Chem.* **1987**, *59*, 2024.
(7) Montaudo, G.; Scamporrino, E.; Vitalini, D. *Macromolecules* **1989**, *22*, 623; **1989**, *22*, 627.
(8) Ballistreri, A.; Garozzo, D.; Giuffrida, M.; Impallomeni, G.; Montaudo, G. *Macromolecules* **1989**, *22*, 2107.
(9) Montaudo, G.; Puglisi, C.; Samperi, F. *Polym. Bull.* **1989**, *21*, 483.
(10) Ballistreri, A.; Impallomeni, G.; Montaudo, G.; Lenz, R.; Kim, B.; Fuller, R. C. *Macromolecules* **1990**, *23*, 5059.
(11) Ballistreri, A.; Garozzo, D.; Giuffrida, M.; Montaudo, G.; Montaudo, M. S. *Macromolecules* **1991**, *24*, 1231.
(12) Montaudo, G.; Scamporrino, E.; Vitalini, D. *Macromolecules* **1991**, *24*, 376.
(13) Source listing of program MACO3 is available on request.
(14) Kamiya, N.; Yamamoto, Y.; Inoue, Y.; Chujo, R.; Doi, Y. *Macromolecules* **1989**, *22*, 1676.

Registry No. (HV)(HH)(HN)(HU) (copolymer), 134905-18-3; (HV)(HH)(HN)(HU)(HO)(HC)(HO) (copolymer), 134905-19-4; (HB)(HV) (copolymer), 80181-31-3.