

A Statistical Description of the Alginate Molecule in Terms of a Penultimate-Unit Copolymer

BJØRN LARSEN, TERENCE PAINTER,
ARNE HAUG and OLAV SMIDSRØD

Norwegian Institute of Seaweed Research, N.T.H., Trondheim, Norway

For a binary linear copolymer containing n units per chain, the 2^n different possible sequences were generated in the digital computer as numbers in the binary system. The probability of occurrence of each sequence was calculated in terms of assigned initiation and transition frequencies, the latter being determined by the identities of any combination of the three preceding units in the chain. Summation of the probabilities for all sequences of identical composition then gave the composition-distribution, and the various forms which this may take were investigated.

Comparison of the results with those observed experimentally for degraded alginate indicated that the structure of the alginate molecule closely resembles that of a copolymer formed according to the "penultimate-unit" theory of addition copolymerisation.

Alginate is a block copolymer, in which two types of long homopolymeric block, composed of residues of mannuronic and guluronic acid, respectively, are joined through long blocks in which these residues are arranged in largely alternating sequence.¹⁻³ The present paper attempts to determine the minimum number of assumptions required to describe such a structure statistically.⁴ The object is to determine the amount of information needed by the alginate synthetase system in order to give rise to such a molecule.

The treatment consists in examination of various statistical models, similar to those which have been applied to synthetic addition copolymers,^{5,6} for their capacity to explain the principal experimental observation pertaining to sequence in alginate, namely, the gradual liberation of the three types of block as discrete polymeric entities upon partial acid-hydrolysis.¹⁻³

THEORY

Nomenclature. Most papers on free-radical copolymerisation have employed a dualism in nomenclature, the statistical and the kinetic, and it is also difficult to avoid that here, without obscuring the relationship between structure and known mechanisms of heteropolysaccharide biosynthesis.

The sequence of monomer units in a copolymer can be fully defined by probabilities alone, and these take on a physical meaning only in the light of some established mechanism of copolymerisation. This may be illustrated by considering a binary linear copolymer of units A and B, n units in length, and containing a mole fraction (\bar{F}^A) of units of A.

If the distribution of A and B along the chains is random, the copolymer is fully defined by the given information alone, since the probability of any unit of any chain being an A is F^A , and is independent of the identity of any other unit in the chain. The other extreme is the situation in which the probability of a given unit being an A is dependent upon the identity of every other unit in the same chain; in this case, 2^n different probabilities, one for each complete chain, are required fully to define the copolymer.* To describe a copolymer whose molecules are all identical, for example, $2^n - 1$ of these probabilities would have to be set equal to zero; the remaining sequence, namely, that possessed by every molecule, would have a probability of occurrence of unity.

Between these two extremes, there exists a continuum of complexity, in which the probability of a given unit being an A depends upon the identities of only some of its neighbours in the same chain. The present work attempts to determine which of these neighbours are important in the case of alginate, and considers first the nearest neighbours, then the next-nearest neighbours, and then the next-next-nearest neighbours. Dependency upon them is expressed in terms of nearest, next-nearest, and next-next-nearest neighbour frequencies. However, to emphasise the essential unity with kinetic treatments, it is more convenient to refer to these quantities as persistence or transition frequencies, which are, respectively, the probabilities that, when a chain is scanned in a given direction, a given unit or group of units is followed by a unit which is similar to, or different from, that which precedes it. Thus, the persistence frequency is unity, minus the transition frequency.

Apart from the interest inherent in determining the amount of information that is required to define the structure of alginate, it is now sufficiently clear that the biosynthesis of heteropolysaccharides,⁷ almost certainly including alginate,⁸ can be treated⁴ as stepwise addition copolymerisation. It therefore seems justified tentatively to attribute a kinetic significance to the probabilities. Thus, the initiation frequencies, which are the probabilities that the first unit or units in a chain will have a particular identity, take on the significance of priming specificities, or measures of the rate at which the primer will accept one type of unit compared to another. Similarly, the persistence frequencies become measures of substrate specificity, or the rate at which

* Strictly speaking, only $2^n - 1$ different probabilities would be required, since the sum of these, subtracted from unity, would give the remaining probability.

the polymerase enzymes will perform a homogeneous addition compared to a heterogeneous one (*e.g.* $A \leftarrow A$ compared to $A \leftarrow B$).

In the following discussion, kinetic terminology will be employed alongside the statistical, with the reservation that the significance of the hypothetical rate-constants is subject to modification. Accordingly, a copolymer whose structure is fully defined by the nearest-neighbour frequencies is treated according to the classical "ultimate-unit" theory of copolymerisation,⁹ which assumes that the rate of addition of a given monomer unit to the end of a growing chain is influenced only by the identity of the terminal unit of that chain. Copolymers whose structures must be defined in terms of both the nearest and the next-nearest neighbour frequencies are treated according to the "penultimate-unit" theory,^{5,6,10,11} which assumes that only the last two units in a growing chain are able to influence the rate of addition of the next unit. When the nearest, next-nearest, and next-next-nearest neighbours are all considered, the copolymer is treated as a "penpenultimate-unit" model,^{6,10,11} which assumes an influence of the last three units in a chain upon the nature of the next addition.

The essential identity of the kinetic treatment with the statistical theory of sequence may be illustrated by reference to the most complex situation mentioned earlier, in which every complete chain must be assigned a separate probability of occurrence. In terms of the kinetic model, this is equivalent to the statement that the rate of addition of a given unit is determined by every preceding unit in the chain. Thus, the probability that the first unit added to the primer is an A is a single figure. For the second unit that is added, two probabilities are needed, one for the circumstance in which the first was an A, the other for when the first was a B. The third unit requires four different probabilities, and so on. The total number of different probabilities is then the sum of the n terms of a geometrical progression, with unity as the first term, and 2 as the common ratio, and is accordingly $2^n - 1$, the same result as before.

Definitions and assumptions. Biosynthesis is considered to occur by stepwise addition of monomer units to one end of a growing chain, and to be initiated by a primer, P. The two monomers are A and B, and the chains are linear and of infinite length. Composition-distribution is considered for fragments, formed by random depolymerisation of these long chains. Only the distribution for fragments of a given degree of polymerisation, n , is considered, since summation over all values of n does not change the form of the distribution.¹² The donor substrates are assumed to be present in excess, giving rise to zero-order kinetics.

For the ultimate-unit model, which is also treated elsewhere,^{4,12} there are two priming specificities, P_A and P_B , these being the probabilities that the first unit added to the primer is an A or a B, respectively. Since termination is ignored, $P_A + P_B = 1$. The growth of the chains is governed by four zero-order rate-constants, $K_{A \leftarrow A}$, $K_{B \leftarrow B}$, $K_{B \leftarrow A}$, and $K_{A \leftarrow B}$, where $K_{B \leftarrow A}$, for example, is the rate of addition of a unit of A to a chain whose terminal unit is a B. These constants are expressed as substrate specificities a and b , where $a = K_{A \leftarrow A} / K_{A \leftarrow B}$ and $b = K_{B \leftarrow B} / K_{B \leftarrow A}$. The two nearest-neighbour or

persistence frequencies, $p(\text{AA})$ and $p(\text{BB})$ are accordingly $a/(1+a)$ and $b/(1+b)$, respectively.

For the penultimate-unit model, four priming specificities, P_{AA} , P_{BB} , P_{AB} , and P_{BA} are recognised, where P_{AA} , for example, is the probability that the first two units of any chain are both A's. In this case, there are eight zero-order rate-constants, in terms of which four substrate specificities are defined, respectively, as $a_1 = K_{\text{AA} \leftarrow \text{A}} / K_{\text{AA} \leftarrow \text{B}}$, $a_2 = K_{\text{BA} \leftarrow \text{A}} / K_{\text{BA} \leftarrow \text{B}}$, $b_1 = K_{\text{BB} \leftarrow \text{B}} / K_{\text{BB} \leftarrow \text{A}}$, and $b_2 = K_{\text{AB} \leftarrow \text{B}} / K_{\text{AB} \leftarrow \text{A}}$. In the notation used throughout, it is assumed that the chain is initiated at the left, and grows towards the right. The four persistence frequencies are then, $p(\text{AAA}) = a_1/(1+a_1)$, $p(\text{BAA}) = a_2/(1+a_2)$, $p(\text{BBB}) = b_1/(1+b_1)$, and $p(\text{ABB}) = b_2/(1+b_2)$.

In the penpenultimate-unit model, there are eight priming specificities, sixteen rate-constants, and eight substrate specificities which are defined as

$$\begin{aligned} a_1 &= K_{\text{AAA} \leftarrow \text{A}} / K_{\text{AAA} \leftarrow \text{B}}, & a_2 &= K_{\text{BAA} \leftarrow \text{A}} / K_{\text{BAA} \leftarrow \text{B}}, & a_3 &= K_{\text{ABA} \leftarrow \text{A}} / K_{\text{ABA} \leftarrow \text{B}}, \\ a_4 &= K_{\text{BBA} \leftarrow \text{A}} / K_{\text{BBA} \leftarrow \text{B}}, & b_1 &= K_{\text{BBB} \leftarrow \text{B}} / K_{\text{BBB} \leftarrow \text{A}}, & b_2 &= K_{\text{ABB} \leftarrow \text{B}} / K_{\text{ABB} \leftarrow \text{A}}, \\ b_3 &= K_{\text{BAB} \leftarrow \text{B}} / K_{\text{BAB} \leftarrow \text{A}}, & \text{and } b_4 &= K_{\text{AAB} \leftarrow \text{B}} / K_{\text{AAB} \leftarrow \text{A}}. \end{aligned}$$

The computer program. It has previously been recognised by Price¹⁰ that any sequence of two monomers can be represented by some number in the binary system. For a binary linear copolymer, n units in length, there are 2^n different possible sequences, which can be written as 2^n binary numbers of n digits each, beginning with zero, and ending with the equivalent of $2^n - 1$. The programming technique can be illustrated by reference to the penultimate-unit model, from which it will be clear how it was readily extended to treat the penpenultimate-unit case, and how it could be further extended to still more complex situations.

In the binary numbers, an 0 was regarded as a unit of A, and a 1 as a unit of B. The sequence $(\text{A})_n$ was then a row of n 0's, that is, simply zero in the binary system, while the sequence $(\text{B})_n$ was a row of n 1's, which is the equivalent, in binary numbers, of $2^n - 1$. As the numbers were generated in the computer, they were sorted into $(n+1)$ groups, according to the number of 1's that they contained. Thus, each group contained sequences of identical composition. At the same time, the groups were subdivided according to whether the first two digits were 00, 01, 10, or 11. For every number, a product, obtained by multiplication of $(n-1)$ factors, was calculated: if a number began with 00, P_{AA} was chosen as the first factor; if the next digit was then another 0, P_{AA} was multiplied by $a_1/(1+a_1)$; if the next digit was a 1, the product of the first two factors was multiplied by $1/(1+a_1)$; if the next digit was a 1, the factor $b_2/(1+b_2)$ was then applied; and so on. Separate summation of the products for all chains of the same composition then gave the composition-distribution. The sum of all the factors for all 2^n chains was, of course, unity. A typical print-out for a simple case is shown in Table 1.

Mean composition. The mole fraction (F^{A}) of A in the copolymer was calculated from the composition-distribution by applying the factor $(n-x)/n$ to the fraction of all chains containing x units of B and $(n-x)$ units of A, and summing the products over all values of x (Table 1). It must be noted that the result obtained is not necessarily identical with that given by the well-known copolymer composition equations,^{5,6,10,11} which are valid for

Table 1. Penultimate-unit model. Computer print-out for $n=6$, $a_1=b_1=6$, $a_2=b_2=0.1667$, and $P_{AA}=P_{AB}=P_{BA}=P_{BB}=0.25$. The computer cannot print subscripts or lower-case letters. The symbol -03 after a number means that the decimal point must be moved three places to the left. The first column gives the different possible compositions, x/n , where x is the number of units of B in the chain. The second to fifth columns inclusive give the fractions of the total molecules whose first two units are AA, AB, BA, or BB, respectively, and also the sums of these, which must be identical with P_{AA} , etc. The sixth column is obtained by horizontal summation of the fractions of all chains having a given composition, and represents the composition-distribution; the sum of these must be unity. The seventh column gives the figures in the sixth column, multiplied by $(n-x)/n$; the sum of these is the mean composition of the polymer.

X/N	PAA	PAB	PBA	PBB	SUM	SUM * (N-X)/X
0.0000, +00	1.3494, -01	0.0000, +00	0.0000, +00	0.0000, +00	1.3494, -01	1.3494, -01
1.6667, -01	5.2479, -02	2.2494, -02	2.2494, -02	0.0000, +00	9.7467, -02	8.1222, -02
3.3333, -01	5.0081, -02	3.0615, -02	3.0616, -02	3.7491, -03	1.1506, -01	7.6708, -02
5.0000, -01	8.7476, -03	1.4378, -01	1.4378, -01	8.7476, -03	3.0506, -01	1.5253, -01
6.6667, -01	3.7491, -03	3.0616, -02	3.0615, -02	5.0081, -02	1.1506, -01	3.8354, -02
8.3333, -01	0.0000, +00	2.2494, -02	2.2494, -02	5.2479, -02	9.7467, -02	1.6244, -02
1.0000, +00	0.0000, +00	0.0000, +00	0.0000, +00	1.3494, -01	1.3494, -01	0.0000, +00
SUM	2.5000, -01	2.5000, -01	2.5000, -01	2.5000, -01	1.0000, +00	5.0000, -01

EXECUTION TIME 0.859 SECONDS

infinitely long chains. It becomes identical only when the priming specificity is identical with the probability of occurrence of the same unit or group of units during the stationary state of chain growth.¹²

Since the present work is concerned with composition-distributions among fragments produced by random depolymerisation of very long chains, the priming specificities must be set equal to the stationary-state probabilities. For the ultimate-unit case, P_A is simply set equal to F^A (Ref. 12).

For the penultimate-unit model, it is necessary to know the probability of finding the pairs of units AA, AB, BA, and BB at any point along an infinite chain. Now the condition of a stationary state implies that the probability of any single unit being an A is F^A . The probability of finding two A's together is, therefore, F^A multiplied by the persistence frequency, $p(AA)$. In the penultimate unit model, this quantity is a function in both a_1 and a_2 , since it is the probability that an A will be followed by an A, regardless of the identity of the preceding unit. The exact relationship has previously been derived from first principles by Ham,⁶ and states that $p(AA)=p(BAA)/[p(AAB)+p(BAA)]$. This and the corresponding relationships given by Ham for $p(BB)$, $p(AB)$, and $p(BA)$ lead to the following values, which must be given to P_{AA} , P_{AB} , P_{BA} , and P_{BB} in order to fulfil the required condition:

$$P_{AA} = \frac{F^A(1 + a_1)a_2}{1 + 2a_2 + a_1a_2} \quad (1); \quad P_{AB} = \frac{F^A(1 + a_2)}{1 + 2a_2 + a_1a_2} \quad (2)$$

$$P_{BA} = \frac{F^B(1 + b_2)}{1 + 2b_2 + b_1b_2} \quad (3); \quad P_{BB} = \frac{F^B(1 + b_1)b_2}{1 + 2b_2 + b_1b_2} \quad (4)$$

It may be noted that $P_{AA}+P_{AB}=F^A$; and that $P_{BB}+P_{BA}=F^B$. Moreover, since^{5,6}

$$F^A/F^B = (1 + 2a_2 + a_1a_2)(1 + b_2)/(1 + 2b_2 + b_1b_2)(1 + a_2)$$

which is the copolymer-composition equation for the penultimate-unit case, it is readily seen that $P_{AB} = P_{BA}$.

For the penultimate-unit model, the frequencies of occurrence of the eight different possible triplets are required, and Ito and Yamashita¹¹ have derived relationships which permit calculation of these quantities in terms of the substrate specificities used here. The expressions are, however, unwieldy, and will not be required because only symmetrical distributions, with each priming specificity set at 0.125, are needed for the purposes of the present discussion.

Average sequence-lengths. The expressions for these quantities, valid for long chains, are needed. Those for homopolymeric blocks can be written down directly. For example, in the ultimate-unit model, the statement that $a=20$ implies that an average of 20 homogeneous additions of an A on to an A will be made before a heterogeneous addition of a B on to an A is made, thus terminating the sequence. Since the sequence must begin with an A that has terminated a sequence of B's, the average length (\bar{N}_1^A) of the blocks of A's is $(1+a)$.

Similarly, in the penultimate-unit case, the average length of all A-blocks containing 2 or more units of A is $(2+a_1)$; and in the penultimate-unit model, the average length of all A-blocks containing 3 or more units of A is $(3+a_1)$. For reasons which will become clear in the discussion, however, it is useful to regard the last unit of the block, in a penultimate-unit copolymer, as belonging to a separate, alternating sequence, and the last two units of the block, in a penultimate-unit model, as belonging to a separate sequence beginning with (AAB). The average lengths of the A-blocks, (\bar{N}_2^A) and (\bar{N}_3^A), in the two models are accordingly both formulated as $(1+a_1)$.

To calculate the average lengths of other kinds of sequence, it is convenient to invoke a method of reasoning due to Price.¹⁰ This consists in the recognition that the average length of any given sequence of repeating units is simply the probability of occurrence of the repeating unit, divided by the probability that it will be terminated. Still more generally, it is now proposed that, subject to the semantic qualifications just mentioned, the average length of any kind of sequence is simply $1/(1-\omega)$, where ω is the probability that the sequence will be propagated (that is, the probability that one repeating unit of the sequence will be followed by another of the same kind). It is immediately seen that this leads directly to the modified formulae for (\bar{N}_2^A) and (\bar{N}_3^A) given above.

For alternating sequences of the type $(AB)_n$, ω becomes $[1/(1+a)][1/(1+b)]$ in the ultimate-unit case, $[1/(1+a_2)][1/(1+b_2)]$ in the penultimate-unit case, and $[1/(1+a_3)][1/(1+b_3)]$ in the penultimate-unit case. The three resultant expressions are referred to as \bar{N}_1^{AB} , \bar{N}_2^{AB} and \bar{N}_3^{AB} respectively. Expressions for other kinds of average sequence-length follow in a straightforward manner. In that for \bar{N}_3^{AAB} , for example, ω becomes $[1/(1+b_4)][a_3/(1+a_3)][1/(1+a_2)]$. It should be noted that the values obtained must be multiplied by the number of monomer units in the repeating unit considered, if the average sequence-length is required in terms of monomer units.

Composition-distribution in long chains: a method of approximation

Since the number of different possible chains, and hence the computer running-time, varied as 2^n , these became astronomically large for long chains. Even with $n=10$, the running-time was 13–15 seconds. For chains longer than 15 units, it was necessary to seek some method of economy. This quest was guided by the consideration that, for large n , the absolute composition-distribution, in which every possible composition is plotted separately, is of little practical interest in view of the limited resolving power of physical methods of fractionation. It is more relevant to know the fraction of the total number of molecules having a composition falling within a given range.

In an earlier paper,¹² which included an analytical treatment of composition-distribution in ultimate-unit copolymers, the procedure was adopted of summing the fractions of all the chains having compositions falling within a specified range (5 % or 10 %), and of plotting them as a single point. When this was done, it was noted that, if a and b were large, and were increased in direct proportion to n , the composition-distribution, expressed in this way, remained virtually unchanged. This therefore provided a means of obtaining, with a high degree of accuracy, such distributions for very long chains.

This method of approximation is now examined further, in an attempt to determine how it may be applied to the penultimate-unit and more complex models. A suggestion as to its physical basis is due to Stockmayer,¹³ who, assuming long chains and ultimate-unit effects only, showed that the variance about the mean is proportional to κ/λ , where λ is the degree of polymerisation, and κ , in the present notation, is $F^A b + F^B a$. This quantity, κ , was described as "a certain kind of average sequence-length". Since $F^A = (1+a)/(2+a+b)$ and $F^B = (1+b)/(2+a+b)$, these mole fractions will suffer little change when a and b are large, and are both increased by the same factor. A proportionate increase in a , b , and n should therefore bring about very little change in the variance, as was observed.¹²

An exact expression for the variance, valid for chains of any length, was later derived by Frensdorff and Pariser,¹⁴ and this, together with the expressions just given for F^A and F^B , was used to write a computer program, allowing computation of sets of values of a , b , and n , all giving the same variance and mean composition. The results show that the approximation still holds good for short chains, provided, as before, that a and b are large (Table 2). Appreciable inaccuracy in the assumption of strict proportionality arises

Table 2. Ultimate-unit model. Sets of values of a , b , and n , all giving a relative variance (V) of 0.7955 ± 0.0005 .

n	$a=b$	n	$a=b$
4	4.0000	20	20.367
8	8.1056	24	24.419
12	12.198	28	28.550
16	16.271	32	32.584

Table 3. Ultimate-unit model. Sets of values of a , b , and n , all giving a relative variance (V) of 0.091733 ± 0.00005 .

a	b	n
2.0000	1.0000	10
3.7980	2.1987	20
5.5856	3.3904	30
7.3695	4.5797	40
9.1541	5.7694	50

when a and b are relatively small, and especially when they are different (Table 3).

Pursuing Stockmayer's suggestion, another kind of average sequence length is now considered. If, in the general formula, $1/(1-\omega)$, mentioned in the previous section, ω is set equal to $[a/(1+a)]^2$, the result must be the number-average length of all blocks containing an integral number of non-overlapping (AA) doublets (that is to say, half the number-average length, expressed as singlets, of all sequences containing an even number of monomer units). It follows that, given an ultimate-unit copolymer X , for which $a=a'$, and a similar copolymer Y , for which $a=a''$, the consequence of setting $[a'/(1+a')]$ equal to $[a''/(1+a'')]^2$ is that the average length of "doublet" sequences in Y is identical with that of the "singlet" (that is, all) sequences in X . By expressing the equality as $\log[1+(1/a')] = 2\log[1+(1/a'')]$, it is seen that, when a' and a'' are large, the power terms in the expansion of the logarithms are negligibly small, and the relationship becomes, simply, $a'' \approx 2a'$. Moreover, since⁶ the fraction of the total number of sequences that contain m units of A is $(1/a)[a/(1+a)]^m$, the distribution of chain-lengths of the doublet sequences in Y is identical with that of the singlet sequences in X , and one half of the sequences in Y are doublet sequences. The other half are clearly those which contain an odd number of units of A.

For short chains, the corresponding formula for average sequence-lengths is $1/[1-(n-1)\omega/n]$, where ω has the same meaning as before.* When n is greater than 10, the effect of ignoring the factor $(n-1)/n$ is not, for present purposes, significant. If now, copolymer X has $n=n'$ and Y has $n=n''$, the consequence of setting $[a'/(1+a')]$ equal to $[a''/(1+a'')]^{n''/n'}$ is analogous to that just considered, and occurs when $a'/n' \approx a''/n''$. It can be concluded that, in physical terms, the effect of increasing a and b proportionately with n is, for large a and b , roughly equivalent to taking every individual repeating unit of a sequence, and replacing it with a larger group of (n''/n') repeating units of the same kind. The expanded copolymer so obtained will also contain other sequences of intermediate lengths, but when the blocks are long, the combined weight-fraction of all the expanded blocks whose lengths fall within an increment of one expanded repeating unit, is approximately the same as that of the corresponding single block in the original copolymer.

* The derivation of this formula is rather long, and will be reported in another paper, as it is not of central importance to the present work.

Table 4. Penultimate-unit model. Sets of values of a_1 , b_1 , a_2 and b_2 , and of n , with the corresponding values of the relative variance (V). In every case, $P_{AA}=P_{BB}=P_{AB}=P_{BA}=0.25$.

n	$a_1=b_1$	$a_2=b_2$	V
5	5	0.2000	0.396
6	6	0.1667	0.382
7	7	0.1430	0.382
8	8	0.1250	0.376
9	9	0.1111	0.376
10	10	0.1000	0.373
11	11	0.0909	0.374
12	12	0.0834	0.371
14	14	0.0714	0.371

Turning now to the penultimate-unit case, it is possible to recognise three independent average sequence-lengths: that of the homopolymeric A-blocks, which is determined by a_1 ; that of the homopolymeric B-blocks, which is determined by b_1 ; and that of the alternating $(AB)_m$ blocks, which is determined by both a_2 and b_2 . Given that a_1 and b_1 are large, and a_2 and b_2 small, a simultaneous and approximately proportional increase in the average lengths of the three kinds of block will occur when a_1 and b_1 are increased, and a_2 and b_2 are decreased, all by the same factor. Table 4 shows sets of values of n , and of the four constants which bear this relationship to each other, and gives in each case the relative variance¹⁴ (V), which was calculated directly from the composition-distribution printed out by the computer. It should be noted that it is possible to have other sets of constants giving the same variance, but a different distribution; this will occur when $a_2=a_1$ and $b_2=b_1$, that is, when the model reduces to the ultimate-unit case.

In the penultimate-unit model, it is possible to have, in addition to the homopolymeric blocks, two different kinds of heteropolymeric block in the same chain, whose lengths can be varied independently. These may consist either of repeating units of AAB and ABB, respectively, or of AABB and AB, respectively. For these cases, the same method of approximation should apply, but it will not be needed in this paper.

EXPERIMENTAL

The material investigated was the non-dialysable fraction, liberated into solution by hydrolysis of *Laminaria digitata* alginate in 0.3 N hydrochloric acid at 100° for 20 min, and described as "preparation A" in an earlier publication.² Application of the analytical methods cited² indicated that it had a number-average degree of polymerisation (P_n) of 28.4, and contained residues of mannuronic acid and guluronic acid only, in the ratio (M/G) of 1.64:1. Owing to a difference in starting material, these figures are a little different from those given previously.²

The material (A; 858 mg) was dissolved in water (85 ml), acidified with 1 N sulphuric acid (5.3 ml), and twice dialysed for 12 h against water (600 ml). This removed 225 mg of dialysable material. The non-diffusible residue (630.4 mg) had $P_n=20.6$ and an M/G ratio of 1.18; a portion of it (548 mg), in water (120 ml), was acidified to pH 2.5 with 1 N sulphuric acid, warmed 10 min at 50°, cooled, and dialysed as before. The nondiffusible material (483 mg) had $P_n=19.4$, and an M/G ratio of 1.07.

RESULTS AND DISCUSSION

When alginate is subjected to limited degradation by acid, it is readily cleaved into three kinds of polymeric fragment, which differ markedly in composition and electrophoretic mobility.¹⁻³ These can be easily separated on a preparative scale by making use of their very different solubilities in acid. The solubility—pH profile of whole alginate is unimodal, and clearly distinct from those of the degradation products, indicating that the latter are fragments of a single kind of macromolecule, rather than of separate parent molecules.¹⁵ When hydrolysis is carried out in M oxalic acid¹ or 0.3 N hydrochloric acid,^{2,3} two of the components are enhanced in yield because of their insolubility. Cleavage into similar fragments still occurs, however, when hydrolysis is effected in homogeneous solution² at pH 3.5, and it is also observed¹⁶ when degradation is brought about under homogeneous conditions by Fenton's reagent,¹⁷ by alginase,¹⁸ and by alkali.¹⁹ In the last three cases, the kinetics indicate random depolymerisation. The observed behaviour therefore can be attributed neither to the presence of weak linkages in the alginate molecule, nor to its physical condition in strong acid.

Compared with the starting material, one of the fragments (B) is much enriched in residues of mannuronic acid, while the second (C) is similarly enriched in guluronic acid. When isolated after very brief hydrolysis, both may contain up to 20 % of the other kind of uronic acid, but as degradation proceeds, the enrichment increases, until eventually, almost pure (approximately 95 %) polymannuronic and polyguluronic acids, with number-average degrees of polymerisation (P_n) of 15–20, are obtained.²

The third component (A) initially has the same composition as the original alginate ($M/G=1.6$). An experiment reported here, however, shows that, upon further degradation, this too changes gradually in composition, the M/G ratio tending towards unity. The excess of mannuronic acid residues is split off as dialysable material. An early examination^{1,2} of A indicated that it contained a high proportion of alternating mannuronic and guluronic acid residues, and a more recent study²⁰ has shown that, after further degradation until the M/G ratio is near to unity, it is almost entirely so constituted.

The three theoretical models just described are now examined for their capacity to explain these facts.

The ultimate-unit model. This case has been treated in detail elsewhere,¹² but must be re-examined briefly, to illustrate those respects in which it is inadequate to describe alginate. In this and the subsequent discussion, a unit of A is regarded as a residue of mannuronic acid, and a unit of B as a residue of guluronic acid. The constants a and b were set at 20 and 8, respectively, thus generating a kind of block copolymer, in which the average sequence-lengths of the A-blocks and B-blocks were 21 and 9, respectively, and F^A was 0.7, corresponding to an M/G ratio of 2.33. This quantitative composition (F^A) is fairly typical of mannuronic acid-rich alginates.

By using a formula arrived at analytically, the changes in composition-distribution occurring upon random depolymerisation were followed. As depolymerisation proceeded, the distribution became at first broader, and then divided, first into two, and then into three separate peaks. These corresponded

to one heteropolymeric, and two homopolymeric fragments, and were formed at degrees of scission corresponding to those giving rise to heterogeneity in alginate. It was concluded that it was unnecessary to invoke "weak linkage" hypotheses in order to explain the cleavage of alginate into polymeric fragments of different composition.

Beyond this point, however, the model failed. The composition-distribution of the heteropolymeric fragments was extremely broad, and the mode moved towards higher values of the M/G ratio with increasing degree of scission (Ref. 12, Fig. 1), and then disappeared. No peak corresponding to an M/G ratio of unity (50 % of B) appeared at any stage in the degradation.

It was thought that this failure must have been due to the inability of the ultimate-unit model to account for either the length of, or the high proportion of alternating sequences in, the heteropolymeric segments of the alginate molecule. This inability arises because a high frequency of homogeneous addition or persistence automatically implies a low frequency of heterogeneous addition or transition. Thus, the formula given earlier shows that the average length of the alternating sequences, when $a=20$ and $b=8$, is 1.05 (AB) units, indicating that virtually only isolated (AB) units were present, as connections between longer sequences of A's and B's. Moreover, when the average degree of polymerisation of the whole molecule had been reduced to 30, nearly 60 % of all the heteropolymeric chains contained only one (AB) transition, so that most of them could be represented by the formula $(A)_p(B)_q$.

The penultimate-unit model. The foregoing objections disappear if it is assumed that units beyond the ultimate are able to influence the probabilities of persistence and transition. Again using kinetic terminology advisedly, it is readily seen that it is possible to have a high frequency of $AA\leftarrow A$ and $BB\leftarrow B$ additions, and yet a low frequency of $BA\leftarrow A$ and $AB\leftarrow B$ additions; the average sequence-length of the alternating parts is completely independent of those of the homopolymeric parts.

Fig. 1 shows a series of curves, representing various stages in the depolymerisation of a penultimate-unit copolymer. The curve for $n=10$ was obtained by choosing values of a_1 , b_1 , a_2 , and b_2 to give an average composition (F^A) typical of an alginate from *Laminaria digitata*^{2,3} and average sequence-lengths of the same order as those suggested by the experimental data for this kind of alginate.² The other curves were obtained by use of the method of approximation, with the figures for $n=10$ as the basis for extrapolation. The initiation frequencies were calculated in each case from eqns. 1-4 inclusive. The sets of figures generating the approximation for each value of n are shown in Table 5.

The curves are seen to depict, in a qualitative manner, the exact sequence of events that are observed experimentally when alginate is degraded, namely, the gradual liberation of the three kinds of block as discrete polymeric entities. A precise quantitative comparison cannot be attempted until more experimental data, obtained by degradation under conditions that are known to bring about random depolymerisation, are obtained. However, as already noted in the study of the ultimate-unit case,¹² there is close agreement between the degrees of scission at which the changes are observed experimentally, and those predicted by the theoretical model.

Table 5. Penultimate-unit model. Values of a_1 , b_1 , a_2 , and b_2 , and of the corresponding values of P_{AA} , P_{AB} , P_{BA} , and P_{BB} which, with $n=10$, give approximations for longer chains, all having $a_1=30$, $b_1=15$, $a_2=0.06$ and $b_2=0.03$, and $F^A=0.633\pm 0.020$.

Approx- imation for $n=$	a_1	b_1	a_2	b_2	P_{AA}	$P_{AB}=P_{BA}$	P_{BB}	$(F^A)^a$	$(F^A)^b$
20	15	7.5	0.12	0.06	0.410	0.238	0.114	0.64696	0.64769
30	10	5.0	0.18	0.09	0.403	0.240	0.117	0.64168	0.64247
40	7.5	3.75	0.24	0.12	0.396	0.241	0.122	0.63676	0.63690
60	5.0	2.5	0.36	0.18	0.386	0.243	0.128	0.62789	0.62836
100	3.0	1.5	0.60	0.30	0.368	0.245	0.142	0.61321	0.61314

^a Calculated from the copolymer composition equation ^{5,6} valid for long chains, and used to calculate the initiation frequencies with eqns. 1-4.

^b Obtained from the computer print-out (cf. Table 1, last column). The difference between these figures and those in the preceding column is due to correction to 3 significant figures in the manual calculation of the initiation frequencies.

The broad, flat surfaces between the principal peaks in Fig. 1 are clearly similar in significance to that in the ultimate-unit case. That between the peaks at 0 % and 50 % of B, for example, must represent mainly chains with structures of the type $(A)_p(AB)_q$ and $(AB)_q(A)_p$. It is therefore readily understood why imperfections in the fractionation should, at very low degrees of scission, have resulted in contamination of the homopolymeric fragments with chains of this kind, and why the fractionation rapidly improves as the degree of scission increases. That the contaminating material is indeed of this kind has already been inferred from the experimental data alone.²

Accepting that the experimental facts seem to be accounted for very well by the penultimate-unit model, it is now pertinent to inquire whether an even simpler model could also explain them. One such model exists, namely, one which assumes dependence upon the penultimate unit alone, and not at all upon the ultimate unit. This model is merely a special case of the penultimate-unit model, and corresponds to the situation in which $a_1=1/b_2$ and $b_1=1/a_2$. Fig. 2 shows symmetrical distributions for a model of this kind, and illustrates a good possibility that it could account for the structure of alginate. It is clearly impossible in this case to vary the length of the alternating sequences independently of those of the homopolymeric sequences. Nevertheless, the authors' experience with a wide range of alginate samples suggests that the mean composition and structure of at least some of them could be accounted for on the basis of this simple model. In general, however, it appears unlikely that nearest neighbours will have no influence at all, and this question will not be considered further here.

The penpenultimate-unit model. Having shown that the assumption of penultimate-unit effects is sufficient to account for the structure of alginate, insofar as this is known, it is now pertinent to enquire whether this is a necessary condition. For this purpose two special cases of the penultimate-unit model are now considered. One of these is such that the persistence frequencies depend upon the identity of the penpenultimate unit alone. This situation obtains when $a_1=a_3=1/b_2=1/b_4$, and $b_1=b_3=1/a_2=1/a_4$. Referring to the

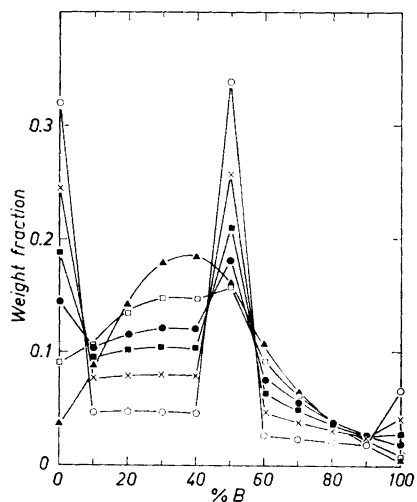


Fig. 1. Approximate representation of the changes in composition-distribution occurring upon random depolymerisation of a penultimate-unit copolymer having $F^A = 0.633 \pm 0.020$, $\bar{N}_2^A = 31$, $\bar{N}_2^B = 16$, and $\bar{N}_2^{AB} = 12$. \blacktriangle , $n=100$; \square , $n=60$; \bullet , $n=40$; \blacksquare , $n=30$; \times , $n=20$; \circ , $n=10$.

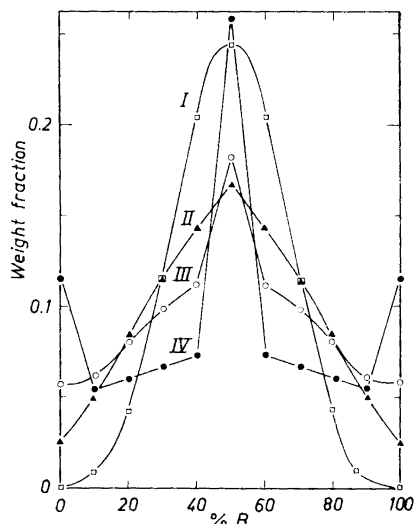


Fig. 2. Penultimate-unit model with influence of penultimate-unit only. $P_{AA} = P_{AB} = P_{BA} = P_{BB} = 0.25$, $F^A = 0.5$ and $n = 10$ throughout.

Curve I: $a_1 = b_1 = a_2 = b_2 = 1$, corresponding to a perfectly random copolymer.

Curve II: $a_1 = b_1 = 3$; $a_2 = b_2 = 0.33$.

Curve III: $a_1 = b_1 = 5$; $a_2 = b_2 = 0.2$.

Curve IV: $a_1 = b_1 = 10$; $a_2 = b_2 = 0.1$.

formulae given earlier for average sequence-lengths, it is immediately clear that if with such a model, the condition of long homopolymeric blocks is imposed (a_1 and b_1 large), it is impossible at the same time to have long alternating sequences (a_3 and b_3 small). Imposition of this condition will automatically generate two different kinds of long heteropolymeric sequence, namely, those composed of repeating units of AAB and ABB, respectively. The form of the composition-distribution to be expected upon depolymerisation of such a macromolecule is shown in Fig. 3. The two sharp peaks corresponding to compositions of 33 % of B and 67 % B, respectively, are clearly at variance with the experimental facts; the methods of fractionation used^{2,3} were capable of separating molecules differing in composition by much less than 33 %.

The other special case of interest is that in which the persistence frequencies depend upon the identities of both the ultimate and penultimate units, but not the penultimate unit. This situation obtains when $a_1 = a_3$, $a_2 = a_4$, $b_1 = b_3$, and $b_2 = b_4$. Here again, it is impossible to have a_1 and b_1 large, and a_3 and b_3 small, and the only essential difference between this and the previous example is that it is now possible to have a range of different average lengths of the $(AAB)_m$ and $(ABB)_m$ blocks for given lengths of the homopolymeric blocks. An example of this case is also shown in Fig. 3.

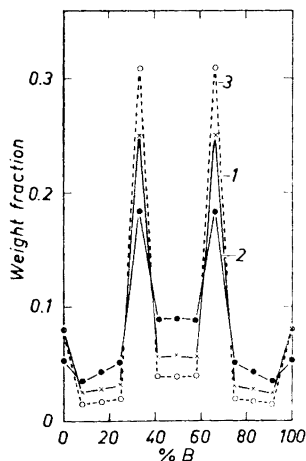


Fig. 3. Penultimate-unit model. The initiation frequencies are all 0.125, $F^A = 0.5$, and $n = 12$ throughout. Influence of penultimate unit only: \times , $a_1 = a_3 = b_1 = b_3 = 20$; $a_2 = a_4 = b_2 = b_4 = 0.05$; \bullet , $a_1 = a_3 = b_1 = b_3 = 10$; $a_2 = a_4 = b_2 = b_4 = 0.1$. Influence of ultimate and penultimate units only: \circ , $a_1 = a_3 = b_1 = b_3 = 20$; $a_2 = a_4 = b_2 = b_4 = 0.01$.

Application of a similar method of reasoning to still more complex models, which assume that the persistence frequencies are determined by units beyond the penultimate, again leads to the same conclusion; no unit or combination of units, not including the penultimate, is able to influence the persistence frequencies in such a way as to give a structure resembling that of alginate.

Experimental facts not explained by the penultimate-unit model. A prominent and unexplained feature of the experimental study was that two different guluronic acid-rich peaks were observed during the course of the degradation. One of these, which was formed first, was of considerably higher molecular weight than the other, and upon further degradation, it disappeared, being converted into the other.

These results were obtained upon hydrolysis of alginate in strongly acidic solution, under which conditions it was insoluble. It has not so far been possible to observe the same effects when alginate is degraded in homogeneous solution. It is quite likely, therefore, that they are connected in some way with the physical condition of alginic acid in the solid state, and are not a consequence of its primary structure. It is also necessary to recognise that separation of these higher molecular-weight intermediates from the more degraded materials rested upon their higher affinity for calcium ions, and the exact dependence of this affinity upon both molecular weight and fine structure are not yet clearly understood. Both of these matters are under further investigation.

Conclusions. From their studies of alginate biosynthesis in *Fucus gardneri* (Silva), Lin and Hassid⁸ concluded that guanosine-diphosphate-D-mannuronate and guanosine-diphosphate-L-guluronate were the precursors, and that chain growth occurred by stepwise transfer of the uronic acid residues on to an acceptor molecule. There are a number of different, detailed mechanisms that are consistent with these findings. The monomer units may be added directly to an acceptor of high molecular weight, or they may first form activated dimers or oligomers, or even long blocks. However, regardless of

the exact sequence of events, it seems impossible to avoid the conclusion that at least some of the enzymes in the pathway must be able to recognise the identity of the unit penultimate to the end of an acceptor chain.

For example, if the blocks are synthesised separately, it is not necessary that the enzymes responsible for this be able to recognise the penultimate unit. Those that join the blocks together, however, must have this ability; otherwise, there would be a high frequency of joining homopolymeric guluronic acid-containing blocks directly on to homopolymeric mannuronic acid-containing blocks, and vice-versa. The experimental data suggest that the fraction of such unions in the molecule is small.²⁰

Brief consideration may also be given to an alternative pathway of biosynthesis, which could possibly operate in addition to that examined by Lin and Hassid. Although epimerisation of monosaccharide units in a long polysaccharide chain is probably not an important general pathway in heteropolysaccharide biosynthesis, it is possibly a little more plausible in the case of polyuronides, since the carboxyl groups, especially if esterified with the enzyme, would facilitate abstraction of a proton at C(5). Here again, however, it would still be necessary for the epimerases to recognise the neighbour next-nearest to the unit undergoing epimerisation, in order to give rise to the correct structure.

The immediate impression given by the block-copolymeric structure of alginate is that it portends a very complex mechanism of biosynthesis. However, it is now clear that it can be accounted for on the basis of a very simple postulate, which is in complete harmony with the known facts about the binding of polymeric carbohydrates to globular proteins. What seems new and surprising is that the penultimate unit should be of such overriding importance, compared to the ultimate unit, but on the other hand, the action pattern of enzymes such as β -amylase is a similar kind of phenomenon, and it may be that a similar principle is involved.

Finally, although alginate appears to be very similar to a penultimate-unit copolymer, the true extent of the similarity cannot be assessed until more is known about the sequence-length distribution in alginate. This problem will be discussed in a later paper.

One of the authors (T.J.P.) is indebted to the *Royal Norwegian Council for Scientific and Industrial Research*, for the award of a postdoctoral fellowship.

REFERENCES

1. Haug, A., Larsen, B. and Smidsrød, O. *Acta Chem. Scand.* **20** (1966) 183.
2. Haug, A., Larsen, B. and Smidsrød, O. *Acta Chem. Scand.* **21** (1967) 691.
3. Haug, A., Myklestad, S., Larsen, B. and Smidsrød, O. *Acta Chem. Scand.* **21** (1967) 768.
4. Painter, T. J. *J. Chem. Soc. C* **1967** 922.
5. Alfrey, T., Bohrer, J. J. and Mark, H. *Copolymerisation*, Interscience, New York and London 1952.
6. Ham, G. E. *Copolymerisation*, Interscience, New York, London and Sydney 1964.
7. Robbins, P. W., Bray, D., Dankert, M. and Wright, A. *Science* **158** (1967) 1536.
8. Lin, Tsau-Yen and Hassid, W. Z. *J. Biol. Chem.* **241** (1966) 3283, 5284.

9. Alfrey, T. and Goldfinger, G. *J. Chem. Phys.* **12** (1944) 205; Mayo, F. R. and Lewis, F. M. *J. Am. Chem. Soc.* **66** (1944) 1594.
10. Price, F. P. *J. Chem. Phys.* **36** (1962) 209.
11. Ito, K. and Yamashita, Y. *J. Polymer Sci A* **3** (1965) 2165.
12. Painter, T. J., Smidsrød, O., Larsen, B. and Haug, A. *Acta Chem. Scand.* **22** (1968) 1637.
13. Stockmayer, W. H. *J. Chem. Phys.* **13** (1945) 199.
14. Frensdorff, H. K. and Pariser, R. *J. Chem. Phys.* **39** (1963) 2303.
15. Haug, A. and Smidsrød, O. *Solution Properties of Natural Polymers* (Special publication No. 23), The Chemical Society, London 1968, p. 273.
16. Haug, A., Larsen, B., Smidsrød, O. and Painter, T. J. *Acta Chem. Scand.* **23** (1969). *In press.*
17. Smidsrød, O., Haug, A. and Larsen, B. *Acta Chem. Scand.* **19** (1965) 143.
18. Haug, A. and Larsen, B. *Acta Chem. Scand.* **17** (1963) 1653.
19. Haug, A., Larsen, B. and Smidsrød, O. *Acta Chem. Scand.* **21** (1967) 2859.
20. Larsen, B., Smidsrød, O., Haug, A. and Painter, T. J. *Acta Chem. Scand.* **23** (1969). *In press.*

Received June 14, 1968.