Journal of Macromolecular Science, Part A
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597274

Probabilities of Specfied Dimers or Trimers in Random Linear Copolymers

J. A. D. Ewart[a]
[a] Flour Milling and Baking Research Association, United Kingdom

To cite this Article Ewart, J. A. D.(1985) 'Probabilities of Specfied Dimers or Trimers in Random Linear Copolymers', Journal of Macromolecular Science, Part A, 22: 11, 1509 — 1515
To link to this Article: DOI: 10.1080/00222338508063351
URL: http://dx.doi.org/10.1080/00222338508063351

J. MACROMOL. SCI.–CHEM., A22(11), pp. 1509-1515 (1985)

Probabilities of Specfied Dimers or Trimers in Random Linear Copolymers

J. A. D. EWART

Flour Milling and Baking Research Association
Chorleywood, Rickmansworth
Herts., WD3 5SH, United Kingdom

ABSTRACT

Elementary combinatorial analysis was used to derive formulas for the number of arrangements of a random linear copolymer that contained none of a specified sequence, AB, ABB, BBA, ABA, or ABC. This enabled the probabilities of finding none, or at least one, of the specified sequence to be calculated.

INTRODUCTION

The problem of calculating the probability of runs of identical elements of specified length in a random sequence has been solved by Mosteller [1]. Bateman [2] simplified the formula.

In this paper the probabilities of finding at least one sequence of two, or three, nonidentical elements have been calculated.

In each case a formula for the number of arrangements lacking the sequence is first obtained.

 0022-233X/85/2211-1509$3.50/0

DISTRIBUTION OF IDENTICAL BALLS IN A ROW OF DISTINGUISHABLE COMPARTMENTS

The number of ways in which r identical white balls and (u - 1) identical black balls can be arranged in line is ${}^{r+u-1}C_r$. The (u - 1) black balls may be regarded as partitioning the white balls into u compartments, i.e., the (u - 2) spaces between black balls plus the two end positions. Therefore, the number of ways (W) of distributing r identical objects among u distinguishable compartments when empty ones are allowed is given by

$$W = {}^{r+u-1}C_r \tag{1}$$

PROBABILITY OF AT LEAST ONE AB

A random copolymer chain contains a monomers of A, b of B, and x unspecified monomers X.

In any arrangement of A's and X's there are a + x + 1 places where B's can be put. To fulfill the condition that there are no AB's, the a positions on the RHS of A are forbidden. Therefore B's can be arranged in ${}^{b+x}C_b$ ways that have no AB's, using Eq. (1). Dividing by the total number of ways of distributing B's and subtracting the quotient from 1 gives the probability of finding at least one AB(P_{AB}):

$$P_{AB} = 1 - P_{\overline{AB}} = 1 - \frac{{}^{b+x}C_b}{{}^{a+b+x}C_b} \tag{2}$$

PROBABILITY OF AT LEAST ONE ABB

The arrangements of the monomers may be partitioned into sets, each member of a particular set containing exactly i AB's, where i runs from 0 to (lesser of a or b).

To make arrangements that contain exactly i AB's, any one of the ${}^{a+x}C_a$ arrangements of A's and X's is taken and i A's are chosen (in ${}^{a}C_i$ ways) and fitted with a B on their RHS. The other (b - i)B's are distributed anywhere except on the RHS of (a - i) free A's (to avoid making more than i AB's) and the RHS of i AB's (to avoid making (ABBs). Therefore,

$$\text{number of permitted places} = a + x + 1 - (a - i) - i = x + 1$$

NB: AB's are counted as single units since splitting AB with B would duplicate the sequences in which the B goes on the RHS of AB.

Hence, the number of ways of making arrangements containing exactly i AB's from a particular arrangement of a A's and x X's is

$${}^{a}C_{i} \cdot {}^{b+x-i}C_{b-i} \qquad \text{from Eq. (1)}$$

Therefore the total number of ways of forming the polymer chain without any ABB's is

$${}^{a+x}C_{a} \sum_{i=0}^{\text{lesser of a or b}} {}^{a}C_{i} \cdot {}^{b+x-i}C_{b-i}$$

Dividing by the total number of arrangements, $(a + b + x)!/a!b!x!$, gives the probability of no ABBs, $P_{\overline{ABB}}$.

$$P_{\overline{ABB}} = \frac{{}^{a+x}C_{a} \sum_{i=0}^{\text{lesser of a or b}} {}^{a}C_{i} \cdot {}^{b+x-i}C_{b-i}}{\dfrac{(a + b + x)!}{a!b!x!}} \tag{3}$$

Therefore, the probability of at least one ABB $(P_{ABB}) = 1 - P_{\overline{ABB}}$.

By symmetry, Eq. (3) also gives the probability of no BBA's.

PROBABILITY OF AT LEAST ONE ABA

Let the set of random copolymers that have exactly i BA's each be partitioned into subsets such that each member of a given subset contains exactly (i - j) free BA's and j BBA's.

To make these, first distribute (i - j) BA's among the (x + 1) places provided by the X's, not more than one per place, in ${}^{x+1}C_{i-j}$ ways.

Among the x + i - j + 1 places for the j BBA's, (i - j) on the LHS of BA's are forbidden because they would give rise to ABA's. Therefore, using Eq. (1), the j BBA's can be distributed in x + 1 allowable places in ${}^{x+j}C_j$ ways.

In distributing the rest of the B's, (b - i - j) in number, places inside AB's or BBA's are not counted because they would lead to sequences identical to those where the B is put on the RHS of an AB or the LHS of a BBA. The (i - j) places on the LHS of BA's are forbidden because they would make more BBA's, thus putting the arrangement in a different subset. The number of places for the B's is therefore x + (i - j) + j + 1 - (i - j) = x + j + 1. By Eq. (1) the (b - i - j) B's can be distributed among them in ${}^{b+x-i}C_{b-i-j}$ ways.

Lastly, the remaining (a - i) A's must be placed. Positions inside BA or BBA are ignored for reasons already given. Of the x + (i - j) + j + (b - i - j) + 1 = (b + x - j + 1) places, (i - j) are forbidden on the LHS of free BA's (to avoid ABA's), and (b - (i - j) - 2j) on the RHS of free B's (to avoid making extra BA's). Hence the allowable places = (b + x - j + 1) - (i - j) - (b - i - j) = x + j + 1.
So by Eq. (1) the (a - i) A's can be placed in ${}^{a+x+j-i}C_{a-i}$ ways.

Therefore, the total number of arrangements, $n_{\overline{ABA}}$, without ABAs is obtained by summing over all sets:

$$n_{\overline{ABA}} = \sum_{i=0}^{\text{least of a, b or } \frac{(x+b+1)}{2}} \; \sum_{j = \text{greater of 0 or } (i-x-1)}^{\text{lesser of i or } (b-i)} {}^{x+1}C_{i-j} \cdot {}^{x+j}C_j \cdot {}^{b+x-i}C_{b-i-j} \cdot {}^{a+x+j-i}C_{a-i} \tag{4}$$

Therefore,

$$P_{\overline{ABA}} = \frac{n_{\overline{ABA}}}{\dfrac{(a+b+x)!}{a!b!x!}} \tag{5}$$

and the probability of at least one ABA = P_{ABA} is 1 - $P_{\overline{ABA}}$.

NB: The summation limits need explaining. The maximum value of i, the number of BA's, is the lesser of a or b, but if x $\ll$ a, b, then a or b $\gg$ 1 + x, so the BA's cannot all be accommodated, not more than one per place, among the X's unless the excess are converted to BBA's.

Let i_m be the maximum permitted limit for i. When there are i_m BA's and $(b - i_m)$ B's left to convert them to BBA's, there must be at least $i_m - (b - i_m)$ free BA's, which must $\leq (1 + x)$, hence $i_m \leq (x + b + 1)/2$.

The upper limit of i is therefore the least of a, b or $(x + b + 1)/2$.

If there are i BA's, there are $(b - i)$ B's left to form the j BBA's. Therefore, the upper limit of j will be the lesser of i or $(b - i)$.

The lower limit of j is governed by the fact that there cannot be more free BA's than there are spaces for them among the X's. Therefore

$$i - j \leq 1 + x$$

hence,

$$j \geq i - x - 1$$

Thus, j must be the greater of 0 or $(i - x - 1)$.

PROBABILITY OF AT LEAST ONE ABC

When c monomers of type C are added to the chain, all the arrangements,

$$\frac{(a + b + c + x)!}{a!b!c!x!}$$

in number, are partitioned into sets that have exactly i AB's per member, $0 \leq i \leq$ (lesser of a or b).

The AB's can be placed among the X's in ${}^{x+i}C_i$ ways.

The $(a - i)$ A's can then be distributed (counting AB's as indivisible, as explained before) in ${}^{a+x}C_{a-i}$ ways, by Eq. (1).

The $(b - i)$ B's may be distributed anywhere except inside AB's, or on the RHS of free A's (to prevent a change of set). Allowable positions $= x + i + (a - i) - (a - i) + 1 = (x + i + 1)$. Therefore, the number of ways is ${}^{b+x}C_{b-i}$.

When the C's are distributed, a difficulty arises. Clearly, to avoid an ABC, the RHS of AB is forbidden. If, however, an AB is split by a C, then not only will this not lead to duplication of a sequence but it will free the RHS of the AB to take one or more C's.

To deal with this, the set with i AB's is partitioned into subsets. In a given subset each member has exactly j AB's split by C's. These j AB's can be chosen in ${}^{i}C_{j}$ ways, leaving (c - j) C's to be distributed, subject only to restriction that the (i - j) places on the RHS of unsplit AB's are forbidden. The number of allowable places, including one inside each split AB which is now allowed to take further C's, is (1 + the numbers of X's, free A's, free B's, + twice the number of split AB's). Each indivisible unit adds one place, but the act of splitting the AB creates an extra place, hence the factor of 2.

Thus, the number of places = $1 + x + (a - i) + (b - i) + 2j = 1 + x + a + b - 2i + 2j$.

The (c - j) C's can therefore, by Eq. (1), be distributed in ${}^{a+b+c+x+j-2i}C_{c-j}$ ways.

If the number of ways of forming sequences with no ABC's is summed over all sets and divided by the total number of arrangements, the probability of no ABC's is obtained:

$$P_{\overline{ABC}} = \frac{\displaystyle\sum_{i=0}^{\text{lesser of } a \text{ or } b} {}^{x+i}C_{i} \cdot {}^{a+x}C_{a-i} \cdot {}^{b+x}C_{b-i} \sum_{j=0}^{\text{lesser of } i \text{ or } c} {}^{i}C_{j} \cdot {}^{a+b+c+x+j-2i}C_{c-j}}{\dfrac{(a+b+c+x)!}{a!\,b!\,c!\,x!}} \qquad (6)$$

therefore, the probability of at least one ABC (P_{ABC}) = $1 - P_{\overline{ABC}}$.

These formulas have been tested as follows:

Equation (2)	$a = b = x = 2$	$P_{\overline{AB}} =$	$\frac{2}{5}$
Equation (3)	$a = 2;\ b = x = 4$	$P_{\overline{ABB}} =$	$\frac{2\,325}{3\,150}$
	$a = 5;\ b = 7;\ x = 4$	$P_{\overline{ABB}} =$	$\frac{444\,780}{1\,441\,440}$
Equations (4) and (5)	$a = 4;\ b = 2;\ x = 2$	$P_{\overline{ABA}} =$	$\frac{258}{420}$
	$a = 5;\ b = 4;\ x = 3$	$P_{\overline{ABA}} =$	$\frac{14\,044}{27\,720}$
	$a = 5;\ b = 5;\ x = 3$	$P_{\overline{ABA}} =$	$\frac{35\,132}{72\,072}$

$a = 5; b = 7; x = 4 \quad P_{\overline{ABA}} = \frac{754\ 600}{1\ 441\ 440}$

Equation (6) $a = b = c = x = 2 \quad P_{\overline{ABC}} = \frac{2\ 166}{2\ 520}$

$a = 3 = x; b = 4; c = 5 \quad P_{\overline{ABC}} = \frac{9\ 281\ 160}{12\ 612\ 600}$

Results obtained by computer generation and counting of sequences agreed with those obtained by calculation from the formulas.

ACKNOWLEDGMENTS

This work forms part of a project sponsored by the UK Ministry of Agriculture, Fisheries and Food (MAFF) to whom thanks are due. The results of the research are the property of the MAFF and are Crown Copyright. The author thanks Dr T. Fearn for testing the formulas by computer.

REFERENCES

[1] F. Mosteller, Ann. Math. Stat., 12, 228 (1941).
[2] G. Bateman, Biometrika, 35, 97 (1948).

Accepted by editor June 18, 1984
Received for publication July 5, 1984