REACTIONS OF ADJACENT SUBSTITUENTS ON POLYMER CHAINS: COMBINATORIAL CONSEQUENCES OF THE MECHANISM

Fred GORNICK

Department of Chemistry, University of Maryland Baltimore County, Baltimore, Maryland 21228, USA

and

Richard W. FREEDMAN

Department of Information Systems Management, University of Maryland Baltimore County, Baltimore, Maryland 21228, USA

Abstract

This paper deals with a statistical problem arising from the pairwise reaction of immediately adjacent substituents along the backbone of a linear polymer chain. The possibility arises that a given substituent becomes unavailable for reaction as a result of the reaction of its nearest neighbors on either side. Previous treatments of this problem have not explicitly taken the consequences of the reaction mechanism into account. In this paper, it is shown that the expected number of unreacted substituents remaining after exhaustive pairwise reaction is a function of the reaction mechanism. For example, in the case of the chain model adopted here, which corresponds to a perfectly regular head-to-tail vinyl halide $(-CH_2-CHX-)$, we show that the fraction of halide atoms remaining after exhaustive removal is equal to 0.1233. The result is compared with results obtained from previous work.

1. Introduction

One of the distinguishing features of macromolecular chemistry is that mechanisms of polymerization are imprinted on the backbones of polymer chains. For example, the monomeric repeat units of vinyl polymers $(-CH_2-CHX-)$ may add to the end of a growing chain head-to-tail $(-CH_2-CHX-CH_2-CHX-)$ or head-to-head $(-CH_2-CHX-CHX-CH_2-)$. In the former case, the X substituents are separated by a single CH_2 group and are said to be in a 1–3 placement. In the latter case, the placement is obviously 1–2. Similarly, a 1–4 placement results when a head–head is immediately followed by a tail–tail addition. Now certain reactions are known to effect pairwise reaction of adjacent substituents provided they are separated by a less-than-critical

distance. In the case of polyvinyl methyl ketone ($X = -CO-CH_3$), it has been shown [1] that 1–2 and 1–3 placements are readily distinguishable. In this case, the substituents are not removed but are paired in a cyclization reaction. Studies on dehydration of polyvinyl alcohol ($X = -OH$) [2] lead to similar results. The reader is referred to the original papers for more detailed accounts of these results.

The general problem, exemplified by the aforementioned studies, may be stated as follows: Suppose that a chemical agent is capable of irreversibly removing or otherwise coupling pairs of adjacent substituents provided only that they are within some specified critical distance. In either case, we may consider that a substituent is effectively removed when it is no longer available for the reaction. If the removal process proceeds in a random fashion, a given substituent may become isolated as a result of the removal of substituents on either side (i.e. its nearest neighbors are then outside the critical distance. Given these circumstances, what is the expected number of isolated substituents remaining, after exhaustive pairwise removal in a chain of N substituents?

In this paper, we restrict ourselves to the simplest case in which the initial placement of substituents is uniform along the backbone of a polymer chain. Theoretical treatments of this problem have been given by Flory [3] and Wall [4,5], both authors having extended the treatment to more general situations in which various spacings are initially present. A purely mathematical treatment of the model in this paper has been presented by Ross [6] (who, no doubt, was unaware of the work of Flory and Wall, which appeared in the chemical literature more than forty years earlier). The treatments by all three of these authors lead to the same result (hereafter referred to as the FWR result), namely that, in the limit of very long chains, the fraction of isolated substituents is exp(-2), i.e. 0.1353.

The FWR result rests on the assumption that all *pairs* of available adjacent substituents are equally likely to be selected for removal. The results we obtain, while based upon an identical model of the initial polymer structure, differ from FWR in that the two-step mechanism of the removal process is explicitly taken into account. Accordingly, a reaction *site* is first randomly selected from among available sites. The first step then consists of establishing an initial attachment between the selected site and the removal agent. A second step then consists of selecting *either* an available site to the right or an available site to the left, provided only that they lie within the critical distance. Although the authors of the FWR treatment never suggested an element of directionality in their model, their result does in fact correspond to the situation in which our second step would consistently entail selection of a site either to the left or to the right. In consequence, it will be shown that the probability of the selection of terminal substituent pairs differs from that of non-terminal pairs. It will also be shown that this effect is not merely an "end effect" which becomes negligible in the limit of long chain length, but in fact leads to results which differ quantitatively from those of FWR. Moreover, our results are in precise agreement with the results of Monte Carlo computer simulations of the pair removal process we have described.

2.   Illustrative calculation

The initial state of the system is represented by a row of N equally spaced points, only adjacent points being within the critical distance for pairwise removal. Using this simplified model, we calculate the expected number of isolated points after exhaustive pairwise coupling.

Consider, for example, an array of four points representing four uniformly spaced chemical groups subject to pairwise removal. If the initial state of the system is represented by the diagram [* * * *], then the (only) possible final states are represented as follows:

State 1:   [− −] ;   two pairs removed, no points isolated.

State 2:   [* − *];   one pair removed, two points isolated.

We define S_N as the expected number of isolated sites in a chain of length N so that

$$S_N = \sum_{i=1}^{N} p(i)\, s(i), \tag{1}$$

where $p(i)$ is the probability that the removal process will result in final state (i), and $s(i)$ is the number of isolated points associated with that state.

To illustrate our method, we now proceed to calculate the probabilities of each of the two possible final states. For convenience, we indicate the location of each site in the chain from left to right (i.e. $1, 2, 3, 4$).

State 1 can be realized in only two ways: either pair 1–2 is formed, followed by the formation of pair 3–4, or the reverse order (3–4 followed by 1–2). By symmetry, the probabilities of either order are the same, so we will calculate only the former. We note that, if the pair 1–2 is formed first, the conditional probability of forming the pair 3–4 is unity. The probability of initially forming pair 1–2 is the sum of the probabilities of selecting site 1 in the first step (of the reaction mechanism) and then selecting site 2 in the second step, or the reverse (site 2 in step one, followed by site 1 in step two). If site 1 is selected, the conditional probability of forming pair 1–2 is clearly unity. However, if site 2 is selected, the probability of forming pair 1–2 is 1/2. Hence, the total probability of initially forming pair 1–2 is $(1/4)(1 + 1/2) = 3/8$. It then follows from the symmetry condition that

$$p(1) = (3/8) + (3/8) = 3/4.$$

As for state 2, it is apparent that the probability of its realization is simply the probability that the initial pair is 2–3. That is,

$$p(2) = (1/4)(1/2 + 1/2) = 1/4.$$

It then follows from eq. (1) that

$$S_4 = (3/4) (0) + (1/4) (2) = 1/2.$$ (OURS)

To compare our results with those of FWR, we note that, in the latter case, it is assumed that all possible pairs are equally probable. In a chain of length N, the initial pair selection occurs with equal probability among the $(N - 1)$ possible pairs, denoted as $(1, 2)$ through $(N - 1, N)$, which is to say that the probability that *any* particular pair is selected is $1/(N - 1)$. Specifically, terminal pairs (those located at the ends of the chain) have the same probability of forming as those located internally. In contrast, the method described in this paper ascribes a probability of $(1/N) (3/2)$ to the formation of each of the terminal pairs, and a probability of $(1/N)$ to each of the $(N - 3)$ internal pairs.

To see the effect of the distinction of these two approaches, consider the calculation of S_4 by the FWR method. Using the probabilities obtained as a result of the FWR assumptions, it is found that $p(1) = 2/3$ and $p(2) = 1/3$. Hence, referring to eq. (1),

$$S_4 = (2/3) (0) + (1/3) (2) = 2/3.$$ (FWR)

3.   Recursion relationships

Consider a series of N equally spaced points, numbered 1 through N. There are therefore $N - 1$ pairs of adjacent points, denoted as $(i, i + 1)$, where i can be any value from 1 to $(N - 1)$. Let $p(i, N)$ be the probability that the pair $(i, i + 1)$ is the first pair selected for removal in a chain of N points. It follows from the assumption in our model that

$$p(i, N) = \begin{cases} 3/(2N) & \text{for } i = 1, N - 1; \\ (1/N) & \text{for } i = 2, \ldots, N - 2. \end{cases}$$ (2)

We observe that the selection of the initial pair $(i, i + 1)$ splits the chain into two subchains of lengths $(i - 1)$ and $(N - i - 1)$. The total number of isolated points in the chain will be the sum of the isolated points in each of the two subchains. Recall that S_N is the expected number of isolated points for a chain of length N. Let $S_{i,N}$ be the expected number of isolated points for a chain of length N, given only that the pair $(i, i + 1)$ has been selected. Then,

$$S_{i,N} = S_{i-1} + S_{N-i-1}.$$ (3)

Since any of the $(N - 1)$ possible pairs could have been initially selected,

$$S_N = \sum_{i=1}^{N-1} p(i, N) (S_{i,N}).$$ (4)

Combining eqs. (3) and (4),

$$S_N = \sum_{i=1}^{N-1} p(i, N) S_{i-1} + \sum_{i=1}^{N-1} p(i, N) S_{N-i-1} . \tag{5}$$

By symmetry, $p(i, N) = p(N - i, N)$. Introducing the index $j = N - i$, eq. (5) may be rewritten as

$$S_N = \sum_{i=1}^{N-1} p(i, N) S_{i-1} + \sum_{j=N-1}^{1} p(j, N) S_{j-1} . \tag{6}$$

The two sums on the right-hand side of eq. (6) are obviously equal. So,

$$S_N = (2) \sum_{i=1}^{N-1} p(i, N) S_{i-1} . \tag{7}$$

Substitution of eq. (2) into (7) yields

$$S_N = (2/N) \left[(3/2) S_0 + \{ S_1 + S_2 + \ldots + S_{N-3} \} + (3/2) S_{N-2} \right]. \tag{8}$$

Since S_0 equals zero, eq. (8) can be rewritten as

$$(N) S_N = (2)[S_1 + S_2 + \ldots + S_{N-2}] + S_{N-2}. \tag{9}$$

Subtracting $(N - 1) S_{N-1}$ from the above and rearranging, we obtain

$$S_N = (1/N)[(N - 1) S_{N-1} + (3) S_{N-2} - S_{N-3}]. \tag{10}$$

Equation (10) is a general recursive relationship for S_N. Since it is obvious that $S_1 = 1$, $S_2 = 0$, and $S_3 = 1$, we can insert these values into eq. (10) and obtain $S_4 = 0.5$, which we have previously shown to be the case.

We observe that the difference between the FWR method and ours resides entirely in the difference in the two expressions for $p(i, N)$. The FWR equation which corresponds to our eq. (2) is

$$p(i, N) = 1/(N - 1) \quad \text{for } i = 1, 2, \ldots, N - 1. \qquad \text{(FWR)} \qquad \text{(2A)}$$

Replacing our expression for $p(i, N)$ with the FWR expression (eq. (2A)) and, thereafter, following the same line of reasoning, we obtain the following recursion relationship for S_N:

$$S_N = [1/(N - 1)] [(N - 2) S_{N-1} + (2) S_{N-2}]. \qquad \text{(FWR)} \qquad \text{(10A)}$$

For obvious reasons, the model which we share with FWR requires agreement as far as the values of S_1, S_2, and S_3 are concerned. As we have previously shown, the FWR result for S_4 is 2/3, which also follows from eq. (10A).

To further compare the results of our calculation, we have tabulated the results of our recursion (eq. (10)) and those obtained by the FWR method (eq. (10A)) for several values of chain length N. These results (shown in table 1) were verified by a statistically large number of Monte Carlo simulations corresponding to, respectively, the two mechanisms presented in eqs. (10) and (10A).

Table 1

Comparison of results

N	Present (eq. (10))		FWR (eq. (10A))	
	S_N	S_N/N	S_N	S_N/N
1	1.000	1.0000	1.000	1.0000
2	0.000	0.0000	0.000	0.0000
3	1.000	0.3333	1.000	0.3333
4	0.500	0.1250	0.667	0.1667
5	1.000	0.2000	1.000	0.2000
6	0.917	0.1528	1.067	0.1778
7	1.143	0.1633	1.222	0.1746
8	1.219	0.1523	1.352	0.1690
9	1.362	0.1514	1.489	0.1654
10	1.478	0.1478	1.624	0.1624
20	2.713	0.1357	2.977	0.1489

It is apparent from table 1 that our results differ from those of FWR and, as we shall show in the next section, these differences persist even for large N. We have previously noted that the FWR results correspond to the assumption that the removal mechanism is a two-step process in which a randomly selected point is paired with the one immediately to its right, if the latter point has not been previously selected. In contrast, our calculation allows for equally probable pairing either to the right *or* to the left. In our view, this corresponds more closely to an actual chemical reaction mechanism.

4.   General formulation of the problem

We now consider the implications of the recursion relationships, eqs. (10) and (10A). The quantity of greatest physical interest is, in general, the fraction of isolated substituents following exhaustive pairwise removal, S_N/N. As shown in the appendix, the FWR formulation leads directly to an analytical solution for the limiting value of S_N/N, as N increases without bound, i.e. the limit for FWR is $\exp(-2)$, which is equal to 0.1353.

Unfortunately, the mechanistic formulation of the problem, which we have presented, does not lend itself to a closed form solution. As shown in table 2, our calculations demonstrate that in the limit of large N, the fraction of isolated substituents is 0.1233, corresponding to a somewhat higher efficiency of substituent removal.

Table 2

Asymptotic values of S_N/N

N	Present (eq. (10))		FWR (eq. (10A))	
	S_N	S_N/N	S_N	S_N/N
10	1.478	0.1478	1.624	0.1624
100	12.579	0.1258	13.804	0.1380
1,000	123.565	0.1236	135.606	0.1356
10,000	1233.433	0.1233	1353.627	0.1354
100,000	12332.240	0.1233	13533.971	0.1353

While the two results are not vastly different numerically, it is of interest to note that the difference is significant from a theoretical point of view.

5.   Conclusions

In this paper, we have demonstrated that the limiting number of isolated substituents in a chain of length N, subject to pairwise removal by a two-step mechanism, differs from that previously reported. The difference between these two results arises from the explicit recognition that the probability of removal of terminal pairs in a sequence of contiguous constituents differs from that of a non-terminal pair (eq. (2)). However, this should not be viewed as a typical end effect that becomes negligible in the limit of large N. For example, in a chain of 108 units, the random removal of four pairs from the chain could result in as many as five subchains. In the original chain, the proportion of terminal units is 2/108 (approximately 0.02), whereas in the subdivided chain, the proportion of terminal units could be as large as five times that in the original chain. The point to be emphasized here is that, regardless of the length of the original chain, new terminal units are generated as a result of the division of the chain into an increasing number of smaller subchains. Thus, the terminal-pair effects exert increasing impact as the selection process proceeds.

Appendix

In tables 1 and 2, the results of numerical computation of S_N and S_N/N are shown for our method of computation and for that of FWR. We reiterate that the same physical model is employed in both cases. As seen in table 1, the values of S_0, S_1, S_2, S_3, and S_5

are identical in both cases but all other terms are different. The recursive relationships governing S_N are given by eq. (10) (our method) and eq. (10A) (FWR's method). They are rewritten below after minor rearrangement.

$$(N)\,(S_N) \qquad = (N-1)\,(S_{N-1}) + (3)\,(S_{N-2}) - S_{N-3}, \tag{10'}$$

$$(N-1)\,(S_N) = (N-2)\,(S_{N-1}) + (2)\,(S_{N-2}). \tag{10A'}$$

Adopting Flory's approach [3], we define first and second differences of the series $S_0, S_1, S_2, \ldots$ as δ_j and μ_j, respectively, where

$$\delta_j = S_j - S_{j-1}; \qquad j = 1, 2, 3, \ldots \tag{A1}$$

and

$$\mu_j = \delta_j - \delta_{j-1}; \qquad j = 2, 3, 4, \ldots . \tag{A2}$$

Summing successive terms in (A1) and (A2), we obtain

$$\sum_{k=2}^{N} \delta_k = S_N - S_1 , \tag{A3}$$

$$\sum_{j=2}^{k} \mu_j = \delta_k - \delta_1 ; \qquad k > 1. \tag{A4}$$

Since, in all cases, $S_0 = 0$ and $S_1 = 1$, then $\delta_1 = 1$. It follows that

$$
\begin{aligned}
S_N = 1 &+ \sum_{k=2}^{N} \left(1 + \sum_{j=2}^{k} \mu_j\right) \\
&= 1 \\
&+ 1 + \mu_2 \\
&+ 1 + \mu_2 + \mu_3 \\
&+ \ldots \\
&+ 1 + \mu_2 + \mu_3 + \ldots + \mu_N \\
&= (N)\,(1 + \sum_{j=2}^{N} \mu_j) - \sum_{j=2}^{N} (j-1)(\mu_j).
\end{aligned}
\tag{A5}
$$

The problem is thus reduced to finding suitable expressions for μ_j. These are obtained for the two cases under consideration by taking second differences of successive terms in eqs. (10) and (10A). The resulting recursion relationships are, for the present case,

$$\mu_j = (-2/j)(\mu_{j-1}) + (1/j)(\mu_{j-2}) \qquad \text{(OURS)} \qquad \text{(A6)}$$

and, for FWR's method

$$\mu_j = [-2/(j-1)](\mu_{j-1}). \qquad \text{(FWR)} \qquad \text{(A7)}$$

Considering the FWR result first and recalling that $\mu_2 = (S_2 - S_1) - (S_1 - S_0) = -2$, we obtain the results

$$\mu_j = [(-2)^{(j-1)}]/[(j-1)!]. \qquad \text{(FWR)} \qquad \text{(A8)}$$

Substitution of (A7) and (A8) into (A5) then leads to

$$S_N = (N+2)\sum_{j=0}^{N-1}[(-2)^j]/(j!) + \{(-2)^N/[(N-1)!]\} \quad \text{(FWR)} \qquad \text{(A9)}$$

or

$$S_N/N = (1+2/N)\sum_{j=0}^{N-1}[(-2)^j]/(j!) + [(-2)^N/(N!)], \quad \text{(FWR)} \qquad \text{(A9')}$$

from which we obtain FWR's result

$$\lim_{N\to\infty}(S_N/N)_{\text{FWR}} = \sum_{j=0}^{\infty}[(-2)^j]/(j!) = \exp(-2) = 0.1353. \qquad \text{(A10)}$$

Returning to eq. (10'), we note that the recursion expression of S_N does not lead to a closed form expression for μ_j. We have accordingly resorted to direct numerical computation of S_N, the resulting ratio (S_N/N) being

$$\lim_{N\to\infty}(S_N/N)_{\text{present case}} = 0.1233.$$

Acknowledgements

References

[1] C.S. Marvel and C.L. Levesque, J. Amer. Chem. Soc. 60(1938)280.
[2] C.S. Marvel and C.E. Denoon, J. Amer. Chem. Soc. 60(1938)1045.
[3] P.J. Flory, J. Amer. Chem. Soc. 61(1939)1518.
[4] F.T. Wall, J. Amer. Chem. Soc. 62(1940)803.
[5] F.T. Wall, J. Amer. Chem. Soc. 63(1941)821.
[6] S.M. Ross, *Stochastic Processes* (Wiley, New York, 1983), pp. 20–22.