Proc. Natl. Acad. Sci. USA
Vol. 73, No. 7, pp. 2299–2303, July 1976
Biochemistry

Kinetics of branch migration in double-stranded DNA

(recombination/bacteriophage G4)

BETTY J. THOMPSON*, MERRILL N. CAMIEN, AND ROBERT C. WARNER

Department of Molecular Biology and Bichemistry, University of California, Irvine, Calif. 92717

Communicated by Howard A. Schneiderman, April 26, 1976

ABSTRACT **The rate of branch migration in double-stranded DNA has been measured by the use of a unique substrate formed by the action of the *Eco*RI restriction endonuclease on the dimeric figure-8 configuration of the replicative form DNA of phage G4. The figure-8 and the X-form derived from it contain a junction of the kind postulated to occur in the Holliday structure and to be an essential feature of a number of models of recombination. In the X-form this junction can branch migrate to an irreversible terminal configuration consisting of two linear monomers. The disappearance of X-forms was measured by electron microscopy. A treatment of branch migration as a random walk process was developed to permit the determination of the rate of the intrinsic process, a step movement of the junction by a distance of one base pair. A value of about 6 kilobase pairs per sec at 37° was obtained.**

The figure-8 configuration of dimeric replicative form (RF) DNA of phages ϕX174 and S13 was shown by Thompson *et al.* (1) to comprise up to 7% of the dimers. It was proposed to be an intermediate in recombination in these phages chiefly because it was assumed to be able to engage in branch migration (1, 2). A junction of the type present in the figure-8 is an essential feature of the Holliday structure (3), and its migration is basic to a number of models of recombination described in recent reviews (4–6). The figure-8 is, in fact, a Holliday structure for the simple, circular genome of the RF of the small DNA phages.

Evidence that branch migration is a physically possible process was given by Lee *et al.* (7). They studied a repetitious DNA structure with single-stranded tails by electron microscopy and concluded from the distribution of tail lengths that branch migration had occurred. Kim *et al.* (8) provided similar evidence for double-stranded branch migration of the kind that occurs in the figure-8. Structures interpreted as involving both single- and double-stranded migration were observed by electron microscopy to result from recombination in phage T4 by Broker and Lehman (9, 10). They developed models showing the relevance of branch migration to recombination and suggested the figure-8, among other configurations, as an intermediate. Largely because of the reversibility of the direction of branch migration, the methods that have been used thus far cannot give information about the dynamics of the process or even provide more than indirect evidence that it can occur.

The RF of phage G4 has one site per monomer for hydrolysis by the *Eco*RI restriction endonuclease. We have used this enzyme to convert figure-8 forms from G4-RF to an X-shaped configuration. These X-forms are a new type of substrate for studying branch migration because the migration terminates in an irreversible step, the formation of two linear monomers. We report here a preliminary study of branch migration in the X-form, including measurements of its rate.

Abbreviations: RO, runoff, formation of linear monomers from X-forms; RF, replicative form DNA.

* Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, Calif. 94143.

MATERIALS AND METHODS

Phage and DNA. G4 phage and RF were prepared as previously described (11). Cultures of G4 phage were obtained from Dr. G. N. Godson. The dimer fractions of G4 RF DNA were prepared by agarose gel electrophoresis (11, 12). The form I dimer band was isolated from a 0.5% agarose gel, and the DNA was extracted from the gel by equilibrium sedimentation in ethidium bromide, KI gradients (13).

***Eco*RI Digestion.** The *Eco*RI restriction endonuclease was purified by the method of Greene *et al.* (14). The *Escherichia coli* strain RF13 used for this preparation was obtained from Dr. H. Boyer. The activity was monitored by agarose gel electrophoresis. The dimer fraction was digested at 12° for 30 min to minimize branch migration.

Electron Microscopy. Samples were prepared for electron microscopy as previously described (1). Some of the spreadings were done at 12° for 30 min. Temperatures below 12° resulted in poor contrast and annealing of the sticky ends produced by the action of *Eco*RI.

RESULTS

In a limit digest of the dimer fraction of G4-RF, the X-form, derived from the figure-8, will be the only remaining dimeric species. All other forms of complex DNA will have been converted to linear monomers. A micrograph of such a preparation is shown in Fig. 1. To qualify as an X-form, a molecule must have the four arms occurring in paired lengths such that each pair sums to monomer length.

The process of branch migration in an X-form is shown diagrammatically in Fig. 2. Each step is reversible, except the last in which the X-form separates into two linear monomers. This final irreversible step has permitted observations to be made of the kinetics. The frequency of X-forms with respect to linear monomers was determined by electron microscopy in a dimer fraction of G4-RF that had been digested with *Eco*RI. The preparation was then exposed to conditions under which branch migration proceeded, the process was stopped by chilling, and the decrease in frequency of X-forms was determined. The data are given in Table 1. Branch migration is very slow at 4°. It increases with temperature, and a large fraction of the X-forms disappears in 15 min at 25° and 37°; none was detected after 15 min at 50°. There was no significant loss of X-forms during either condition of spreading for electron microscopy (lines 2 and 3). At 25° this may be because of the binding of cytochrome *c* to DNA in the monolayer. No significant loss of X-forms occurred in 60 min at 12° (line 8). This is in excess of the time used for the *Eco*RI digestion.

In evaluating the frequency of X-forms, only those meeting the criteria stated above, as shown by measurement of arm lengths, were included. The count of linear monomers was based on those estimated visually to be of monomer length. A small number of linear molecules shorter than monomer length were present, possibly as a result of nonspecific nuclease con-

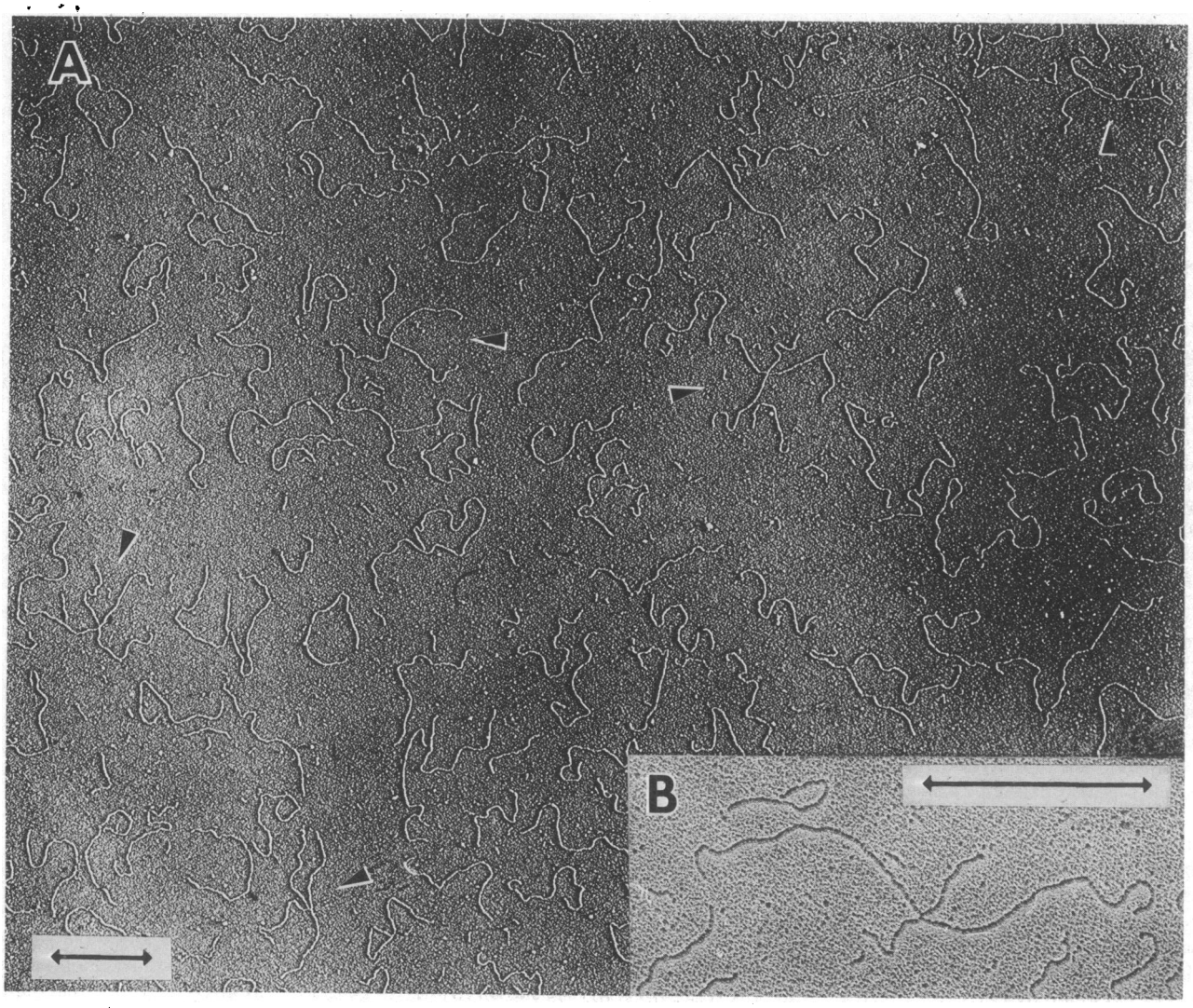

FIG. 1. Electron micrographs showing X-forms, indicated by arrows. The molecule in the inset shows clearly the paired lengths of arms that characterize this structure. The double-headed arrows indicate a distance of 1 μm.

tamination in the *Eco*RI preparation. If these were included in the monomer count, and putative X-forms with no more than one short arm were also counted, the frequencies of X-forms would be raised about 10%.

Formulation of the kinetics of branch migration

Branch migration in a figure-8 molecule can be described as the movement of the migrating junction with respect to some fixed position on the molecule. The problem is to relate this movement to the intrinsic process, which will be characterized by a rate constant k for a single step, taken as the migration of the junction by one nucleotide pair. The total number of steps will then be $N = kt$. The process is the well-known random walk problem for the one-dimensional, unrestricted case. At any position of the junction there is an equal probability or a step in either direction. The probability of finding the junction at a given position p after N random steps is described by the binomial distribution (15). As N becomes large, this probability approaches the normal frequency function (15) in the form

$$P(p) = (1/\sigma\sqrt{2\pi}) \exp(-p^2/\sigma^2) \quad [1]$$

where p and σ are measured in units of nucleotide pairs. The value of σ that characterizes the distribution can be related to the total number of steps N by equating the probability at $p = 0$, $P(0) = 1/\sigma\sqrt{2\pi}$ to the probability at the same point as formulated from the binomial distribution (ref. 15, pp. 179–184).

$$P(0) = \binom{N}{N/2} 2^{-N} \sim \sqrt{2/N\pi}$$

The result, $2\sigma = \sqrt{N}$, makes it possible to obtain σ from a

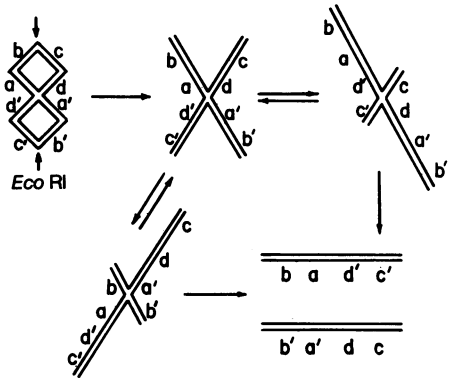

FIG. 2. Diagrammatic representation of the conversion of a figure-8 to an X-form by the action of *Eco*RI and of branch migration in the X-form. The terminal irreversible step is shown at the lower right.

Table 1. Branch migration of X-forms

Incubation conditions		Number						
Temp. (°)	Time (min)	X-forms	Linear monomers	Weight fraction as X-forms	RO/2*	No. of steps ($N \times 10^{-6}$)	Residual steps† ($N \times 10^{-6}$)	k (Kilobase pairs/sec)
—	—	—	—	0.019‡	0	0	—	—
—	—§	26	2916	0.018	0.03	0.03	—	—
—	—¶	13	1540	0.017	0.05	0.1	—	—
25	15¶	20	4203	0.0094	0.25	2.1	2.0	2.2
37	15¶	15	5878	0.0051	0.37	5.3	5.2	5.8
50	15¶	0	2124	0.0	—	—	—	—
4	5.4×10^4§	23	4708	0.0097	0.24	2.1	2.1	0.00065
4	5.4×10^4							
+12	60§	7	1420	0.0098	—	—	—	—

All determinations were made on a dimer fraction of G4-RF digested at 12° for 30 min with *Eco*RI in 100 mM Tris·HCl, 50 mM NaCl, 15 mM $MgCl_2$ at pH 7.5. Before digestion, the dimer fraction contained 0.019 weight fraction of DNA as figure-8 forms, determined by counting 32 figure-8 forms in a population of 3362 monomer equivalents of all species of DNA.

* Runoff from one end of the molecule, RO/2 = 0.5 (1 minus the fraction of X-forms remaining).

† Corrected for migration during spreading.

‡ Undigested sample, fraction shown is for figure-8 forms.

§ Spread for electron microscopy at 12° for 30 min.

¶ Spread for electron microscopy at 25° for 10 min.

knowledge of the rate constant k, and hence to use Eq. 1 in its integrated form to calculate aspects of branch migration.

This treatment is satisfactory for describing the progress of branch migration once k is known, but it cannot be applied to the determination of k from the rate of disappearance of the X-forms. For this it is desired to obtain the number of random steps that will be necessary to produce a measured formation of linear monomers from X-forms. This will be termed runoff (RO). This problem, formulated as the one-dimensional random walk in the presence of absorbing barriers, has been treated in theories of stochastic processes (15, 16), but no solution entirely satisfactory for our purposes appears to have been obtained.

The simplest case is that of the migration of a single junction, described for convenience as being situated at the center of the two monomeric lengths of an X-form and having symmetrically disposed absorbing barriers corresponding to the ends of the molecule at which runoff occurs. A DNA monomer of G4-RF consisting of N_0 nucleotide pairs will have a half-length $r = N_0/2$ and will contain $L = 2r - 1$ positions that a junction can occupy. Runoff can occur after any number of steps n, such that $r < n < N$. The probability of runoff for each value of n in this range must then be summed to give the expectation for the total runoff after a given time t at any specified condition. The runoff in a population of X-forms must also include the contributions from junctions initially located at different points on the molecule. The random walk of a junction, initially at some point other than the center, is described by the general case of the walk problem with asymmetric absorbing barriers. For such cases we will assume that all L possible initial positions are equally probable.

Feller (15) gives suitable combinatorial relations that could be used for small values of n and r, and also presents an approximation for symmetric barriers in terms of a normal distribution function that is approached at large n and r. This relation (ref. 15, p. 90, Eq. 7.7) is of limited use because it applies only to the symmetric case and because it fails rapidly as $\sqrt{N}/r$ increases above one. However, the indication from this equation that a limit relation may be a function of $\sqrt{N}/r$ only, and thus independent of N and r, has been used in obtaining our final relation. An analytic solution for the general asymmetric case is given by Cox and Miller (16). It is, however, so cumbersome for even small values of n that it would be much more difficult to program for computer calculation than the semi-empirical method we have adopted.

The random walk with symmetric absorbing barriers was programmed for the Sigma 7 computer by assigning a probability of one to the junction placed in the central position on the molecule and allowing it to perform a random walk with a probability of 0.5 for each forward and backward step. The probabilities were computed of finding it in each of the available positions for each increasing value of n. When $n > r$, runoff begins to occur. The probability of runoff for each n was accumulated to yield an expectation of runoff for a given value of N at that r. The result of the computation is shown in Fig. 3A as a plot of the runoff from one end, RO/2, against $\sqrt{N}/r$. The same method can be applied to a single, asymmetric random walk and extended to a series of such walks originating at equally spaced positions on the molecule. The results of this series can be integrated numerically to obtain the expectation of runoff for the general case, applicable to X-forms. We have made such computations and they give satisfactory results. However, we have devised a simpler algorithm that converts the series of asymmetric walks into a single symmetric process, which we will refer to as the multiple walk case.

The multiple walk was initiated by placing a junction in each of the L positions and assigning it a probability of $1/L$. An interdependent walk was then programmed in a similar way to the single walk case, except that the probability of finding a junction in a given position included contributions from the stepwise movement from all initial positions. In this process runoff will begin at $n = 1$, and can be summed as for the single walk case to yield the expectation of runoff. Computations were made for values of r up to 700 at $\sqrt{N}/r = 0.1$ and up to 320 at higher $\sqrt{N}/r$. It was evident that at a constant $\sqrt{N}/r$, the value of RO/2 rapidly approached a limit as r was increased. The rate of approach to the limit is illustrated by three sample curves in Fig. 4. While the runoff computed for the largest values of r would be satisfactory for calculation of experimental data, we have analyzed the curves to estimate the limiting values. The solid lines in Fig. 4 are plots of the hyperbola, $RO/2 = -a/r + b$, fitted to the computer generated points. The fit is precise and is sensitive to the value of the limit (b in the above equation) to

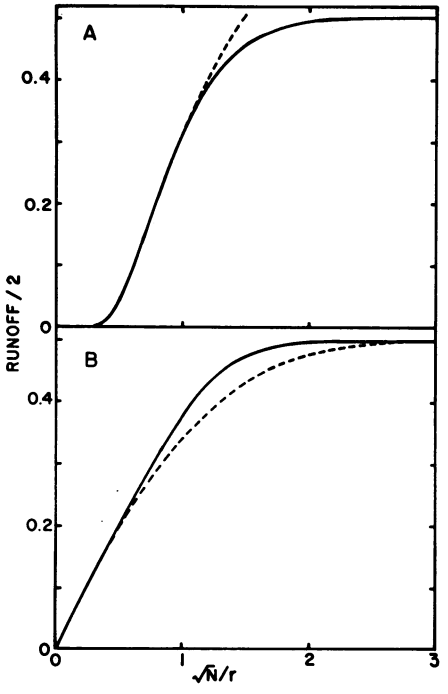

FIG. 3. Branch migration of DNA X-forms described as random walks. (A) Single walk case. The solid line is a plot of the fractional runoff from one end (RO/2) of an X-form as a function of $\sqrt{N}/r$ from the migration of a single junction initially in the center of the molecule. N is the number of steps in the walk and r is the half-length of the molecule. The curve shown is that computed for $r = 160$ as described in the *text*. It is within 10^{-5} RO/2 units of the limit curve at all values of $\sqrt{N}/r$. The dashed line is a plot of Feller's equation (15) for a random walk with symmetric absorbing barriers. It is coincident with the solid line when $\sqrt{N}/r < 1$. (B) Multiple walk case. The solid line is the computed curve for an initial state in which junctions occur with equal probability at any position on the molecule. The fitted equation described in the *text* using Eq. 3 is coincident with the computed curve with a much greater precision than can be shown graphically. The dashed line is a plot of the probability distribution function, Eq 2, between the limits of zero and $\sqrt{N}/r$.

at least 0.002% at any value of $\sqrt{N}/r$. The limit values obtained in this way are plotted to give the solid line in Fig. 3B. The runoff from the multiple walk represented by this curve can be viewed as a deviation from the distribution function

$$\mathrm{RO}/2 = (1/\sqrt{2\pi})\int_0^{\sqrt{N}/r} \exp(-y^2/2)\,dy \qquad [2]$$

which is plotted as the dashed line in Fig. 3B. It agrees with the computed runoff at low $\sqrt{N}/r$ and approaches the same value at high $\sqrt{N}/r$. The frequency function related to this distribution function is thus an approximation at high and low $\sqrt{N}/r$ to the probability of a junction reaching the location r after exactly n steps starting with the arrangement of initial positions defined for the multiple walk. The distribution function gives the summation of these probabilities for each of the n steps up to N. For the purposes of computation, RO/2 was fitted to a function of the form of Eq. 2 in which, for the upper limit of the integral, $\sqrt{N}/r$, $x = f(\sqrt{N}/r)$, was substituted, where f is a power series. The integral was evaluated from tables of the normal distribution function. The fifth order polynomial

$$x = a(\sqrt{N}/r)^5 + b(\sqrt{N}/r)^4 + c(\sqrt{N}/r)^3 + d(\sqrt{N}/r)^2 + e(\sqrt{N}/r) \qquad [3]$$

where $a = 0.03901$, $b = -0.24398$, $c = 0.46687$, $d = -0.083$, and $e = 1.00362$, fits the runoff with a maximum deviation of

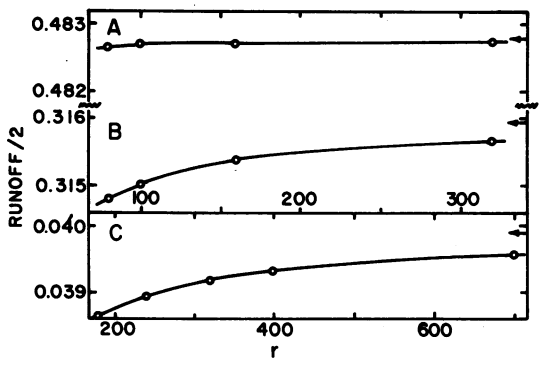

FIG. 4. Extrapolation to obtain limit value of runoff. Fractional runoff from one end of the molecule, RO/2, is plotted against r for three values of constant $\sqrt{N}/r$. The points are computer values for the multiple walk case. The solid lines are hyperbolic curves, fitted to the points, which yield the limit values at $r = \infty$ indicated by the arrows. (A) $\sqrt{N}/r = 1.6$; (B) $\sqrt{N}/r = 0.8$; (C) $\sqrt{N}/r = 0.1$.

0.0005 RO/2 units over the entire range of $\sqrt{N}/r$. This deviation is not perceptible on the scale used in Fig. 3B, and the limit curve shown is coincident with that constructed with the aid of Eq. 3. Determination of $\sqrt{N}/r$ from an observed RO/2 has been made from a large-scale graph of Fig. 3B. For G4, r = 2400 nucleotide pairs has been assumed in order to calculate N.

Rate of branch migration

The results of calculations are shown in Table 1. The treatment outlined above yields a value for N, the number of steps or nucleotide pair migrations required to produce the measured runoff. This is converted to a rate constant $k = N/t$ for the fundamental process. Each calculation of the number of steps from the experimental runoff must be based on the number of X-forms present in the initial digest (line 1). A correction for loss of X-forms, for example, by migration during spreading, can then be made by subtracting the number of steps leading to that loss from the total number. This correction has been made in Table 1 even though it is small. The net number of steps on which the rate measurement is based is entered as "Residual steps."

If the precision of k is estimated by taking the standard deviation of each observation of X-forms as the square root of the number, the value of k at 37° may be stated as 6 ± 2 kilobase pairs/sec or as a jump time per nucleotide pair of $\tau = 170 \pm 50$ μsec.

The energy of activation calculated between 25° and 37° is 15 kcal (63 kJ); that between 4° and 25° is 64 kcal (268 kJ). The precision of these values is low.

DISCUSSION

Branch migration has been shown by observation of the X-forms obtained from G4-RF to be a physically occurring process capable, in principle, of supporting the role attributed to it in recent models of genetic recombination. The migrating junction in the X-forms is of the type termed a fused junction by Clayton *et al.* (17), who prepared a figure-8 structure containing it from circular monomeric and linear dimeric single strands of mitochondrial DNA. Its properties have been studied with molecular models by Sigal and Alberts (18) and by Sobell (19). Migration of this junction has been referred to as double-stranded branch migration (8, 9) to distinguish it from strand displacement or single-stranded branch migration (7).

The observed rate at 37° of 6 kilobase pairs/sec is about 0.3% of the rate that can be calculated from the theoretical treatment

of Meselson (20) for a DNA structure the size of a G4 X-form. His result was based on the assumption that rotary diffusion is the rate-limiting step. We conclude that for a molecule of this size, the rate-limiting steps are those occurring in the junction rather than rotary diffusion.

The rate of branch migration at 37° appears to be just rapid enough, even if it is not a catalyzed reaction *in vivo*, to account for recombination in G4 and similar phages. In a normal infection by ϕX174, phage production becomes significant after 10 min (21), and no parental RF is available before 2 min (22). Thus, a few minutes is the maximum time available. The step rate of 6 kilobase pairs/sec can be calculated from Eq. 1 to yield a distribution of the extent of branch migration starting at the point of initiation of figure-8, with a σ of 0.06 of a genome in 1 min or of 0.18 genome in 8 min. There is thus a probability of 0.32 that the junction will be found at a region of the genome more than 300 nucleotide pairs from the origin in either direction in 1 min, or more than 850 nucleotide pairs in 8 min. Any region between the junction and its origin could contain a repairable heteroduplex introduced by the migration. This is sufficiently rapid so that recombination in these phages could be mediated entirely by the processes of branch migration and mismatch repair by the sequence of events previously outlined (1).

The data presented here are of limited precision because of the relatively small number of X-forms that can be observed by electron microscopy and because of uncertainty concerning branch migration during the spreading period. An assay that avoids these difficulties has been devised and will be reported on elsewhere.

This investigation was supported by U.S. Public Health Service Grant CA 12627 from the National Cancer Institute.

1. Thompson, B. J., Escarmis, C., Parker, B., Slater, W. C., Doniger, J., Tessman, I. & Warner, R. C. (1975) *J. Mol. Biol.* **91**, 409–419.
2. Doniger, J., Warner, R. C. & Tessman, I. (1973) *Nature New Biol.* **242**, 9–12.
3. Holliday, R. (1964) *Genet. Res.* **5**, 282–304.
4. Sobell, H. M. (1973) *Adv. Genet.* **17**, 411–490.
5. Radding, C. M. (1973) *Annu. Rev. Genet.* **7**, 87–111.
6. Hotchkiss, R. D. (1974) *Annu. Rev. Microbiol.* **28**, 445–468.
7. Lee, C. S., Davis, R. W. & Davidson, N. (1970) *J. Mol. Biol.* **48**, 1–22.
8. Kim, J-S., Sharp, P. A. & Davidson, N. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1948–1952.
9. Broker, T. R. & Lehman, I. R. (1971) *J. Mol. Biol.* **60**, 131–149.
10. Broker, T. R. (1973) *J. Mol. Biol.* **81**, 1–16.
11. Rush, M. G. & Warner, R. C. (1970) *J. Biol. Chem.* **245**, 2704–2708.
12. Hayward, G. S. & Smith, M. G. (1972) *J. Mol. Biol.* **63**, 383–395.
13. Blin, N., Gabain, A. V. & Bujard, H. (1975) *FEBS Lett.* **53**, 84–86.
14. Greene, P. J., Betlach, M. C., Boyer, H. W. & Goodman, H. M. (1974) in *Methods in Molecular Biology*, ed. Wickner, R. B. (Marcel Dekker, New York), Vol. 7, pp. 87–111.
15. Feller, W. (1968) *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, New York), 3rd ed.
16. Cox, D. R. & Miller, H. D. (1965) in *The Theory of Stochastic Processes* (John Wiley & Sons, New York), pp. 32–33.
17. Clayton, D. A., Davis, R. W. & Vinograd, J. (1970) *J. Mol. Biol.* **47**, 137–153.
18. Sigal, N. & Alberts, B. (1972) *J. Mol. Biol.* **71**, 789–793.
19. Sobell, H. M. (1974) in *Mechanisms in Recombination*, ed. Grell, R. F. (Plenum Press, New York), pp. 433–438.
20. Meselson, M. (1972) *J. Mol. Biol.* **71**, 795–798.
21. Denhardt, D. T. & Sinsheimer, R. L. (1965) *J. Mol. Biol.* **12**, 641–646.
22. Denhardt, D. T. & Sinsheimer, R. L. (1965) *J. Mol. Biol.* **12**, 647–662.