MODELS FOR DNA MEDIATED BACTERIAL TRANSFORMATION

WALTER F. BODMER

STANFORD UNIVERSITY SCHOOL OF MEDICINE

1. Introduction

The process of DNA mediated bacterial transformation, originally described by Griffiths in 1928 and later clarified by Avery, Macleod, and McCarty [1] (for a recent review see [25]), provides the most direct evidence that DNA is the genetic material. It can be used to study the relation between the physical and chemical properties of DNA and its biological activity. Of particular interest are the biological effects of physical and chemical treatments of DNA and the study of the mechanism of genetic recombination at the molecular level. Transformation was first described in *Pheumococcus* and subsequently in a limited number of other bacterial species notably *Hemophilus influenzae* and *Bacillus subtilis*. In a typical transformation experiment specially prepared "competent" recipient bacterial cells of the strain x^-, requiring, say, a substance x, are mixed under suitable conditions with a donor DNA purified from a strain x^+ which does not require substance x. Among the DNA treated cells, a small proportion of "transformed" cells are found which no longer require substance x. It is now well established that this "transformation" of x^- to x^+ cells is mediated solely by the purified x^+ DNA.

The total amount of DNA per nucleus in, for example, *B. subtilis* corresponds to a molecule containing approximately 2×10^6 nucleotide pairs, or 2000 genes, assuming an average size of 1000 nucleotide pairs per gene. The DNA preparations used in transformation experiments generally contain molecules with an average size of 3×10^4 to 5×10^4 nucleotide pairs which corresponds to between one and three per cent of the complete bacterial chromosome. The procedures for preparing DNA break the chromosome, probably at random, into some 30 to 100 fragments. The size distribution of fragments within any given preparation may have quite a large variance. The lesion in the x^- cells, which prevents their growth in the absence of substance x, will generally be a genetic mutation affecting one (or possibly a few) nucleotide pairs at some defined point on the chromosome. Thus, only a fraction of the DNA molecules in a normal preparation will carry those nucleotide pairs involved in the x^-x^+ genetic difference.

It is now known that transformation involves the physical insertion of a piece of donor DNA, corresponding to a region of the donor chromosome containing the x^- mutation, in the appropriate position on the recipient chromosome [6], [11]. The insertion process is assumed to be somewhat analogous to genetic recombination as it occurs in higher organisms. Transformation thus provides a molecular model for genetic recombination. When recipient cells having a double requirement x^-y^- are mixed with DNA from x^+y^+ cells, double (x^+y^+) transformants occur with a frequency which is, in general, approximately proportional to the product of the frequencies of the two singly transformed classes, x^-y^+ and x^+y^-. Occasional pairs of requirements are, however, found where the frequency of doubles is quite comparable to that of the singles. Most pairs of genes, chosen at random, will be on different donor DNA molecules, since these only have a size corresponding to a few per cent of that of the whole chromosome. Thus, the generally low frequency of cotransformation for any two markers is explained if only one or a few DNA molecules, as isolated, are incorporated into any given recipient cell. Only pairs of markers sufficiently close together to be frequently on the same molecule will show cotransformation frequencies comparable to those for single marker transformation (see, for example, [18], [23]).

Little is known of the specific requirements for a cell to be competent, that is, transformable by added donor DNA. The competent cell is in a very special physiological state and usually only a minority of the cells in the culture to which DNA is added are competent. It has generally been assumed that the competent cell has many attachment sites through which any arbitrary DNA molecule can enter the cell (see, for example, [3]). Some recent evidence suggests, however, that this may not be the case, but rather that there may be only one or a few sites which can accept only molecules corresponding to certain restricted portions of the recipient chromosome [4]. Specifically, there is evidence to suggest that the attachment site may correspond to a membrane associated region of the recipient chromosome surrounding a growing point of DNA synthesis.

The whole process of transformation can be formally divided into at least five stages, as indicated in figure 1, not all of which are necessarily independent. These stages are:

(1) the *initial attachment* of donor DNA molecules to the recipient competent cell which must be followed by

(2) *entry* of the DNA into the cell and

(3) *synapsis* between the donor DNA and the recipient chromosome, that is, some process of alignment of the two molecules to mediate

(4) *recombination* (or integration) by which donor fragments are inserted into the recipient genome, and finally,

(5) *expression* of the transformed state, which is the main usual observation and which may follow recombination after an appreciable lag.

The growing point model suggests that entry and synapsis may be synonymous. There is also evidence that the initial attachment (and perhaps entry

and synapsis without recombination) may be a reversible process. Recombination is presumably irreversible and represents the final outcome of the integration of donor DNA into the recipient cell. Undoubtedly each of these stages can be affected by the state of the cell and of the DNA, and may be dependent on the time of contact between cells and DNA.

Population of recipient cells (a^-) and donor DNA molecules (a^+)

Initial attachment
1

Entry
2

Synapsis (or alignment of donor and recipient DNA)
3

Recombination
4

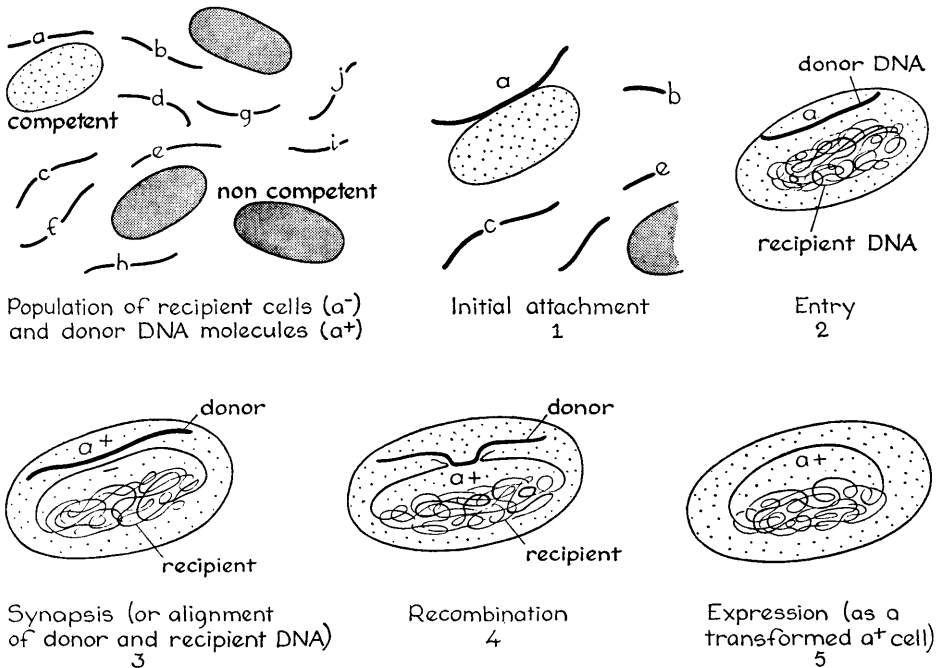Expression (as a transformed a^+ cell)
5

FIGURE 1

The transformation process.

The main aim of this paper is to outline some mathematical models for these various stages of the transformation process with a view to providing a theoretical basis for the quantitative interpretation of transformation experiments. A review of earlier work together with some new results will be presented. The models fall basically into two categories:

  (1) models for the initial attachment, entry and synapsis of donor DNA and
  (2) models for the recombination process.

2. Simple kinetic theory

2.1. *One marker.* Most attempts to provide a theoretical model for transformation have been based on the simple kinetic theory for bimolecular reactions. Thus, Thomas [29] in his original description of such a model assumed a scheme which may be depicted as shown in figure 2. Here B represents the noncompetent
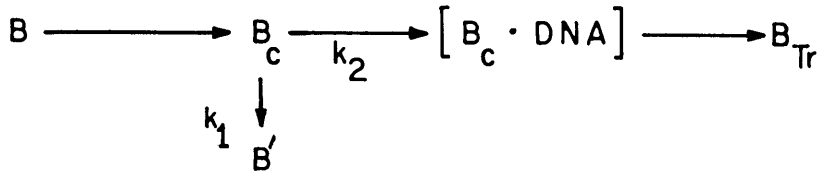
FIGURE 2

Scheme of theoretical model for transformation based on
simple kinetic theory for biomolecular reactions.

recipient bacterial cell, B_c the competent cell, B' a cell which has irreversibly
lost its competent state and $[B_c \cdot DNA]$ a complex between competent cells
and DNA which inevitably leads to the transformed state B_{Tr}. The irreversible
reaction rates $B_c \rightarrow B'$ and $B_c \rightarrow [B_c \cdot DNA]$ are k_1 and k_2, respectively. Atten-
tion is focused on transformation for some one particular genetic marker. This
scheme leads to the equations

(2.1)
$$\frac{dB_c}{dt} = -k_1 B_c - k_2 D B_c,$$

$$\frac{dT}{dt} = k_2 D B_c,$$

where B_c is the concentration of competent bacteria, T the concentration of the
complex $[B_c \cdot DNA]$ (or equivalently of subsequently transformed cells B_{Tr}),
and D the concentration of DNA molecules. These quantities are all functions
of t, the time after the start of the reaction between cells and DNA. It is assumed
that the process starts with a given number of competent bacteria at a time
when the reaction $B \rightarrow B_c$ has ceased. We now assume that the concentration
of DNA molecules D is large compared with the initial concentration of com-
petent cells $B_c(0)$, so that D may be taken as approximately constant. Then
the equations (2.1) give the solution

(2.2)
$$T = \frac{k_2 D}{k_1 + k_2 D} B_c(0) \{1 - \exp[-(k_1 + k_2 D)]t\}.$$

The reaction between DNA and cells can be terminated after an arbitrary time
of contact t by the addition of the enzyme deoxyribonuclease (DNASE), which
very rapidly destroys any DNA which is not yet "fixed" in the irreversible
$(B_c \cdot DNA)$ complex. The DNASE does not, however, affect the reaction
$(B_c \cdot DNA) \rightarrow B_{Tr}$, so that the observed number of transformants corresponds,
under these assumptions, to $(B_c \cdot DNA) = T$ at the time of addition of the
DNASE. For small t, equation (2.2) approximates to

(2.3)
$$T = k_2 D t B_c(0),$$

while when $t \rightarrow \infty$ the saturating transformation level is given by

(2.4)
$$T_\infty = B_c(0) \frac{D}{(k_1/k_2) + D}.$$

Equation (2.3) predicts that the number of transformants, for small t, should vary linearly with time, DNA concentration, and the concentration of competent cells. This is in reasonably good agreement with the experimental data, except possibly for very small values of t [21]. When k_1/k_2 is very much larger than D, equation (2.4) predicts that the saturating level of transformants (T_∞) is proportional to the DNA concentration, whereas when k_1/k_2 is very much less than D, it approaches the initial concentration of competent bacteria $B_c(0)$.

This simple theory was modified by Fox and Hotchkiss [10] and Lerman and Tolmach [20], on the basis of their experimental results, to include the possibility that the reaction $B_c \leftrightarrow (B_c \cdot \text{DNA})$ was reversible. This gives rise to the scheme in figure 3, where k_2' is the reverse reaction rate, $(B_c \cdot \text{DNA}) \rightarrow B_c$ and
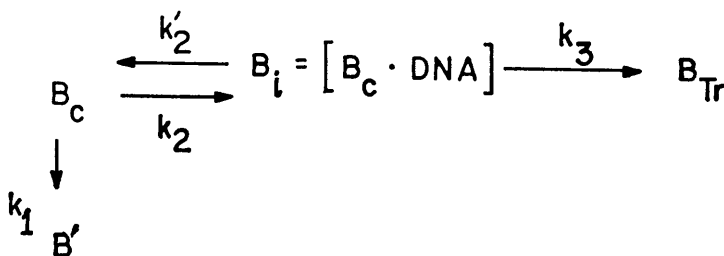

FIGURE 3

Scheme of theoretical model for transformation based on modification
by Fox and Hotchkiss [10] and Lerman and Tolmach [20].

k_3 the rate of the irreversible reaction $(B_c \cdot \text{DNA}) \rightarrow B_{Tr}$. The complete system is described by the set of equations

$$(2.5) \qquad \frac{dB_c}{dt} = -k_1 B_c - k_2 D B_c + k_2' B_i,$$

$$(2.6) \qquad \frac{dB_i}{dt} = k_2 D B_c - k_2' B_i - k_3 B_i,$$

and

$$(2.7) \qquad \frac{dT}{dt} = k_3 B_i,$$

where B_i represents the concentration of the complex $(B_c \cdot \text{DNA})$ and T that of the transformed cells B_{Tr}. If we neglect k_1 and assume that T is small relative to the initial concentration of competent cells $B_c(0)$, then we may write approximately

$$(2.8) \qquad B_i + B_c = \text{constant} = B_c(0).$$

Substituting for B_c in equation (2.6) gives the classical solution of enzyme kinetics

$$(2.9) \qquad B_i = \frac{D B_c(0)}{D + (k_2' + k_3)/k_2} \{1 - \exp[-(k_2' + k_3 + k_2 D)t]\}.$$

When t is large enough to neglect the exponential term, we have, approximately,

$$(2.10) \qquad T = \frac{k_3 t D B_c(0)}{(k_2' + k_3)/k_2 + D}.$$

These are the equations used by Fox and Hotchkiss [10] to estimate $K_m = (k_2' + k_3)/k_2$, the effective Michaelis constant, and $k_3 B_c(0)$ the effective V_{max}. Cavalieri and Rosenberg [8], assuming enzyme kinetic theory, namely, $B_c + B_i = $ constant, have extended equation (2.10) to include a term allowing for competing inactive DNA and a constant multiplier for the probability of expression of the incorporated DNA. A similar scheme has also been proposed (though not analyzed) for *Hemophilus* by Goodgal and Herriott [14].

More generally, the exact solutions of equations (2.5) and (2.6) can be written in the form

$$(2.11) \qquad B_c = \alpha e^{-\lambda_1 t} + \beta e^{-\lambda_2 t},$$

and

$$(2.12) \qquad B_i = \gamma(e^{-\lambda_1 t} - e^{-\lambda_2 t}), \qquad\qquad \lambda_2 > \lambda_1 > 0,$$

where $-\lambda_1$ and $-\lambda_2$, with $\lambda_2 > \lambda_1 > 0$, are the solutions of the quadratic equation

$$(2.13) \qquad \lambda^2 + \lambda(k_1 + k_2 D + k_2' + k_3) + k_1 k_2' + k_3(k_1 + k_2 D) = 0,$$

and α, β, γ are chosen to satisfy the initial conditions $B_c = B_c(0)$ and $B_i = 0$ when $t = 0$. Thus, from (2.6) and (2.12) when $t = 0$

$$(2.14) \qquad \frac{dB_i}{dt} = k_2 D B_c(0) = \gamma(\lambda_2 - \lambda_1).$$

From (2.7) and (2.12)

$$(2.15) \qquad T = k_3 \int_0^t B_i \, dt = k_3 \left[\gamma \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) - \gamma \left(\frac{e^{-\lambda_1 t}}{\lambda_1} - \frac{e^{-\lambda_2 t}}{\lambda_2} \right) \right],$$

so that when $t \to \infty$,

$$(2.16) \qquad T_\infty = k_3 \gamma \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) = \frac{k_3 \gamma(\lambda_2 - \lambda_1)}{\lambda_1 \lambda_2}.$$

Thus, from (2.13) and (2.14),

$$(2.17) \qquad T_\infty = \frac{k_3 k_2 D B_c(0)}{k_1 k_2' + k_3(k_1 + k_2 D)} = \frac{D B_c(0)}{D + \dfrac{k_1}{k_2}(k_2' + k_3)}.$$

Thus, as before, in equation (2.4), the saturating level of transformation is approximately equal to $B_c(0)$ when D is large and is proportional to D when D is small. We have further from equations (2.14) and (2.15) when t is very small,

$$(2.18) \qquad T = \frac{1}{2} k_3 k_2 D B_c(0) t^2 + O(t^3),$$

showing an initial quadratic dependence of the number of transformants on the time of contact. An initial lag in the appearance of transformants corresponding

to such quadratic variation has in fact been reported by Levine and Strauss [21]. However, if transformation is terminated in such a way as not to destroy the $(B_c \cdot DNA)$ complex (for example, by washing), the bacteria complexed at the time of termination would result in additional transformants according to equations (2.5), (2.6) and (2.7), with $D = 0$. The solution of these equations is then

$$(2.19) \qquad T = \frac{k_3 \overline{B}_i}{k_2' + k_3} \{1 - \exp[-(k_2' + k_3)\tau]\},$$

which tends to $(k_3 \overline{B}_i)/(k_2' + k_3)$ as $\tau \to \infty$, where $\overline{B}_i$ is the value of B_i at the time of termination and τ is the time since termination of transformation. When t is small, from equations (2.12) and (2.14), we have

$$(2.20) \qquad \overline{B}_i = \gamma(\lambda_2 - \lambda_1)t = k_2 D B_c(0)t$$

so that the resulting total frequency of transformants is given by

$$(2.21) \qquad T = \frac{k_3 D B_c(0)t}{(k_2' + k_3)/k_2} + O(t^2),$$

a result closely analogous to equation (2.10), when D is small.

Suppose k_1 and k_3 are small compared with k_2 and k_2'. Then from equation (2.13), λ_1 is small compared with λ_2. If now t is such that $\lambda_1 t$ is still small enough for $e^{-\lambda_1 t} \sim 1$, but $\lambda_2 t$ is large enough for $e^{-\lambda_2 t}$ to be small, then from equation (2.12), $B_i \sim \gamma$ and is approximately constant. Now from equation (2.14), we have

$$(2.22) \qquad \gamma = \frac{k_2 D B_c(0)}{\lambda_2 - \lambda_1},$$

and from equation (2.13), whose roots are $-\lambda_2$ and $-\lambda_1$ for $\lambda_2 > \lambda_1 > 0$, we have

$$(2.23) \qquad \lambda_2 - \lambda_1 = k_1 + k_2 D + k_2' + k_3 - 2\lambda_1.$$

Thus, from equation (2.7), neglecting λ_1, we have

$$(2.24) \qquad T \sim k_3 \gamma t = \frac{k_3 D B_c(0)t}{D + (k_1 + k_2' + k_3)/2},$$

which is the same as equation (2.10), derived under the normal assumptions of enzyme kinetics but with $(k_1 + k_2' + k_3)/k_2$ replacing $(k_2' + k_3)/k_2$, and in contrast with equation (2.18), shows a linear dependence of the number of transformants on the time. This equation (2.24) applies to a slightly later time, when an approximate steady state in the number of $(B_c \cdot DNA)$ complexes has been reached.

It should be emphasized that these models effectively refer only to the initial attachment stage, and subsume under the single parameter k_3 the consequences of the four later stages of transformation.

2.2. *Two markers.* Following the simple approximate kinetic theory outlined above, the number of transformants for any given marker for a short time of exposure of cells to DNA and a relatively low DNA concentration is given approximately by

$$(2.25) \qquad\qquad T_1 = K_1 DB_c t.$$

Here the subscript 1 refers to a particular marker, D and B_c are the concentrations of DNA and competent bacteria, t is the time, and K_1 is an overall rate constant. If the interactions between DNA and cells are independent events, double transformations involving a second marker (subscript 2) not generally on the same molecule as the first, will be dependent on the concentration of transformants for the first marker T_1, that is,

$$(2.26) \qquad T_{12} = K_2 T_1 t D = K_1 T_2 t D = K_1 K_2 B_c D^2 t^2.$$

Thus, for small t, the number of double transformants for "independent" markers is proportional to the square of the time and of the DNA concentration. This has been verified by Goodgal [13], Kent and Hotchkiss [19] and others. In particular, Kent and Hotchkiss [19] have verified the linear dependence of T_{12} on time and DNA concentration given that T_1 has already reached a saturating level. From equations (2.25) and (2.26), we have

$$(2.27) \qquad\qquad \frac{T_1 T_2}{T_{12}} = B_c.$$

Thus, having observed T_1, T_2, T_{12} and N, the total number of cells in the recipient culture, the proportion of competent cells $p = B_c/N$ is estimated by $T_1 T_2/T_{12} N$. (See Goodgal [13], Nester and Stocker [24], Kent and Hotchkiss [19].)

When the two markers occur on the same molecule they are said to be linked and then only a single interaction between cells and DNA is needed to give rise to the double transformant. In this case

$$(2.28) \qquad T_1 = K_1 B_c Dt, \qquad T_2 = K_2 B_c Dt, \qquad \text{and} \qquad T_{12} = K_{12} B_c Dt,$$

where K_1, K_2, and K_{12} are the rates with which the attached molecules give rise to single transformants of types 1, 2 and double transformants of types 12, respectively. The proportion of double transformants now varies linearly with the time and the DNA concentration. This distinction between the kinetics of linked and unlinked double transformations provides the most clear cut criterion for linkage between markers in transformation (see, for example, Goodgal [13], Nester, Schafer, and Lederberg [23], Kent and Hotchkiss [19]). It should be emphasized that these analyses apply only to the initial rates, when t (and/or D) is small, and not to the saturation levels, when t (and/or D) is large.

3. Multiple site attachment model

The simple kinetic models described in the previous section made no attempt to specify the details of the reaction between cells and DNA. In particular, no specific assumptions were made as to the number of attachment sites per competent bacteria. However, as will be shown below, assuming exponential kinetics does imply certain assumptions as to the number of attachment sites and also as to the probabilities of their being filled. The only published attempt to work

out a more detailed model of the transformation process is due to Balassa and Prevost [3]. A closely analogous model was developed independently by Bodmer in unpublished notes in 1962.

We make the following assumptions.

(1) The competent bacterium has k receptor sites for donor DNA molecules which have no specificity with respect to type of DNA molecule.

(2) The initial association between sites and DNA is mediated by random diffusion and we assume irreversible attachment immediately after contact. This implies that the probability that a given site contacts a DNA molecule in the time interval dt is $\alpha D \, dt$ where α is a rate constant and D is the overall concentration of DNA molecules per competent cell. The probability of a site being contacted at least once in the time interval t, and so filled irreversibly, is then $1 - e^{-\alpha D t} = q$, say. Consider now a series of independent markers i where $i = 1, \cdots, n$ and the relative proportion of molecules carrying marker i is β_i. Then the probability that r_i sites are filled with molecules of type i and r_0 sites are empty is the multinomial

$$(3.1) \qquad \frac{k!}{\prod_{i=0}^{n} r_i!} (1 - q)^{r_0} \prod_{i=1}^{n} (q\beta_i)^{r_i},$$

where $\sum_{i=0}^{n} r_i = k$ and $\sum_{i=1}^{n} \beta_i = 1$, since the probability of a site being filled with marker i is $q\beta_i$ and of a site being empty is $1 - q$. We now assume further, for simplicity, that the probability that a cell will give rise to a transformant for marker i, given that it has r_i sites filled with the appropriate molecule, is $\gamma_i r_i$. The parameter γ_i then includes the probability of the correct recombinational event taking place and the probability of subsequent expression. We must have $r_i \gamma_i$ such that $\gamma_i r_i \leqq 1$. The total probability of transformation for marker 1 alone, say, is then

$$(3.2) \qquad \sum_{\substack{\text{all } r_i \text{ such} \\ \text{that } \sum_{i=0}^{n} r_i = k}} \frac{\gamma_1 r_1 k!}{\prod_{i=0}^{n} r!} (1 - q)^{r_0} \prod_{i=1}^{n} (q\beta_i)^i = \gamma_1 k q \beta_1.$$

The number of transformants is therefore

$$(3.3) \qquad T_1 = \gamma_1 \beta_1 k B_c (1 - e^{-\alpha D t}),$$

where B_c is the initial number of competent cells, which are assumed to retain their competence until they are transformed. Generally, β_i will be small (usually at most a few per cent and probably $<1/k$) so that γ_i can be nearly one, signifying a high efficiency of integration by recombination and of subsequent expression. Thus, $\gamma_1 \beta_1 k$ may be of order one, so that as $t \to \infty$ the proportion of transformants T_i for each marker tends to B_c. This, of course, implies a high frequency of cotransformation for unlinked markers at saturation, which is not generally found. When t is small,

$$(3.4) \qquad T_1 = \gamma_1 \beta_1 k \alpha D B_c t + O(t^2),$$

showing once again the linear dependence of the number of transformants on DNA concentration, concentration of competent cells, and time.

Assuming independent integration of markers on different molecules, we take the probability that a cell will be transformed for markers i and j, given that it has, respectively, r_i and r_j sites filled with molecules of type i and j, to be $\gamma_i\gamma_j r_i r_j$, so that the probability of the double transformation T_{12} is, for example,

$$(3.5) \qquad \sum_{\substack{\text{all } r_i \text{ such} \\ \text{that } \sum_{i=0}^{n} r_i = k}} \frac{k!}{\prod_{i=0}^{n} r_i!} \gamma_1\gamma_2 r_1 r_2 (1-q)^n \prod (q\beta_i)^{r_i} = \gamma_1\gamma_2\beta_1\beta_2 q^2 k(k-1).$$

Substituting for q, this becomes

$$(3.6) \qquad T_{12} = \gamma_1\beta_1\gamma_2\beta_2 k(k-1) B_c(1 - e^{-\alpha D t})^2.$$

If k is much larger than unity, we have

$$(3.7) \qquad \frac{T_1 T_2}{T_{12}} = B_c$$

as before. In this expression T_1 and T_2 include the double transformants, so that it is valid even when T_1 and T_2 are not large compared with T_{12}. The result does, however, depend on assuming that the number of sites k per bacterium is constant. (See Balassa and Prevost [3].) Note that when t is small, equation (3.6) gives

$$(3.8) \qquad T_{12} = \gamma_1\beta_1\gamma_2\beta_2 k(k-1) B_c D^2 t^2 + O(t^2),$$

showing again the quadratic dependence of unlinked (independent) double transformations on DNA concentration and time, when t is small. The linear dependence for linked double transformations follows from exactly the same arguments as given before. From the fact that if $q \to 1$ and $\gamma_i k \beta_i \to 1$ many competent cells would be transformed for a large number of markers, which does not seem to be the case, we must conclude that for this model to be valid either q always remains small and/or $k\beta_i\gamma_i$ is small for all i. Small q, that is, a low probability that a site will be filled, implies that random collision by diffusion between sites and molecules is a severely rate limiting step in the transformation process. Arguments developed in the next section suggest that this is unlikely. Small $k\beta_i\gamma_i$ could arise either from k being much less than $1/\beta_i$ or from the probability of integration and expression γ_i being small.

4. Growing point attachment model

Recent evidence on the involvement of DNA synthesis in transformation [4] suggests that the donor DNA is integrated at a stationary growing point of DNA synthesis. The growing point is probably associated with the cell wall membrane [12] in such a way as to make it readily accessible to incoming donor DNA. This model for the initial stages of transformation has two features which clearly distinguish it from the other models we have so far considered. First, it

implies that there is only one attachment site (or at most perhaps two or three) for donor DNA on the recipient cell. Second, it implies a restriction on the donor DNA molecules which are acceptable for transformation at any given attachment site. Only those molecules containing regions homologous to the portion of the recipient chromosome surrounding the growing point associated with any given attachment site can be integrated at that site. We shall now describe a simple model encompassing these features.

We consider the fate of a given attachment site under the following assumptions:

(1) molecules arrive at the site according to a Poisson process parameter μ, that is, the probability of an arrival in time dt is $\mu\,dt$ where dt is small;

(2) the proportion of acceptable molecules is β;

(3) unacceptable molecules stay at the site for a random length of time ξ with probability density function $g(\xi)$. While at the site they prevent any further attachments;

(4) acceptable molecules "fix" the site, preventing any further attachments.

The probability that a site becomes fixed in the time interval $(t, t + dt)$ is then given by the integral equation

$$(4.1) \qquad f(t) = \beta\mu e^{-\mu t} + \int_0^t \int_0^{t-\tau} (1 - \beta)\mu e^{-\mu\tau}\, g(\xi)\, f(t - \tau - \xi)\, d\xi\, d\tau.$$

The first term is the probability that the first particle arrives at time t and is acceptable, while the second term is the convolution of the probabilities that the first particle arriving at time τ is unacceptable, stays for time ξ and then afterwards an acceptable particle arrives in the time interval $(t, t + dt)$. If we take Laplace transforms, multiplying each side by e^{-st} and integrating over t from 0 to ∞ then equation (4.1) becomes

$$(4.2) \qquad \varphi(s) = \frac{\beta\mu}{s + \mu} + \frac{(1 - \beta)\mu}{s + \mu}\, \psi(s)\, \varphi(s),$$

where $\varphi(s)$ and $\psi(s)$ are the Laplace transforms of $f(t)$ and $g(t)$, respectively. Solving equation (4.2) for $\varphi(s)$ gives

$$(4.3) \qquad \varphi(s) = \frac{\beta\mu}{(s + \mu) - (1 - \beta)\mu\,\psi(s)}.$$

If we assume, that $g(t)$ is an exponential distribution $\lambda e^{-\lambda t}$, then $\psi(s) = \lambda/(s + \lambda)$ and

$$(4.4) \qquad \varphi(s) = \frac{\beta\mu(s + \lambda)}{(s + \mu)(s + \lambda) - (1 - \beta)\mu\lambda}$$

$$= \frac{\beta\mu(s_1 + \lambda)}{(s_1 - s_2)(s - s_1)} + \frac{\beta\mu(s_2 + \lambda)}{(s_2 - s_1)(s - s_2)},$$

giving, on inversion,

$$(4.5) \qquad f(t) = \frac{\beta\mu}{s_1 - s_2} [(s_1 + \lambda)e^{s_1 t} - (s_2 + \lambda)e^{s_2 t}],$$

where s_1, s_2 are the roots of the quadratic

$$(4.6) \qquad s^2 + s(\lambda + \mu) + \mu\beta\lambda = 0.$$

The probability of a site not being fixed by time t is therefore

$$(4.7) \qquad F(t) = \int_t^\infty f(\tau) \, d\tau = \frac{\beta\mu}{s_2 - s_1} \left[\left(1 + \frac{\lambda}{s_1}\right) e^{s_1 t} - \left(1 + \frac{\lambda}{s_2}\right) e^{s_2 t} \right].$$

Since, from equation (4.6), s_1 and s_2 are both negative, clearly as $t \to \infty$, $F(t) \to 0$, leaving no unfixed cells. When t is small

$$(4.8) \qquad F(t) = \frac{\beta\mu}{s_2 - s_1} \left[\frac{(s_1 + \lambda)}{s_1} \cdot (1 - s_1 t) - \frac{(s_2 + \lambda)}{s_2} (1 - s_2 t) \right] + O(t^2),$$

so that the proportion of transformed sites is

$$(4.9) \qquad 1 - F(t) = \mu\beta t + O(t^2),$$

since from equation (4.6), $s_1 s_2 = \mu\beta\lambda$. The above equation for $F(t)$ can also be obtained by the standard "infinitesimal" arguments, yielding two simultaneous differential equations for the probability a site is empty and the probability a site is transiently occupied. This derivation depends on Poisson arrival distributions and exponential waiting time distributions for unacceptable molecules, and so is not as general as the result given by the integral equation (4.1).

The parameter μ, which represents the rate at which contact is established between DNA and cells, will, for simple diffusion theory, be proportional to the DNA concentration. Thus, assuming one attachment site per cell, equation (4.9) is analogous to the equations (2.3), (2.10), (2.21), (2.24), (2.25), and (3.4) for the expected number of transformations when t is small. These equations all reflect the same assumptions of independence between cells, and random collision between DNA and cells as mediated by diffusion.

If there are two attachment sites on a cell, designated by subscripts 1 and 2, the probability that both are fixed by time t with their appropriate molecules is $(1 - F_1)(1 - F_2)$, assuming they are independent. When t is small this gives $\beta_1 \beta_2 \mu^2 t^2 + O(t^3)$, predicting, as before, quadratic variation of the number of unlinked double transformations with respect to time and DNA concentration. However, when $t \to \infty$, F_1 and $F_2 \to 0$ so that the probability that a cell has all its sites fixed with the appropriate DNA molecules tends to unity. The saturating level of cotransformation for unlinked markers depends therefore, with this hypothesis, on the pairwise chromosomal distributions of the attachment sites on single cells. Only if the acceptable regions for pairs of sites are randomly distributed along the chromosome will the limiting level of cotransformation for unlinked markers be the product of the levels for the two constituent markers. A change in the specificity of the attachment site with time by slow movement of the growing points will lead to a somewhat different picture of the kinetics of cotransformation for unlinked markers. Cotransformation of linked markers on the same molecule can be interpreted in exactly the same way as has been

done in previous sections. The observed relatively low frequencies of cotransformation for *linked* markers situated on *different* molecules (obtained by mixing DNA's prepared from two singly marked bacterial strains) are explained by a slight relaxation of the assumption that once a molecule is fixed at the attachment site no other molecules are acceptable.

When t is large, $F(t)$ will be dominated by the term involving $e^{s_1 t}$, where $|s_1| < |s_2|$. If $\beta \ll 1$, which will usually be the case, then from equation (4.6), we have

$$(4.10) \qquad s_1 = -\frac{\mu \beta \lambda}{\lambda + \mu} + O(\beta^2)$$

so that, from (4.7), we have

$$(4.11) \qquad F(t) \sim \exp\left(-\frac{\lambda}{\lambda + \mu}\mu \beta t\right).$$

The effect of the finite mean time of attachment of unacceptable molecules $1/\lambda$ is represented by the factor $\lambda/(\lambda + \mu)$ in the exponent, which tends to 1 as $\lambda \to \infty$ and the mean attachment time tends to zero.

If we include in the model a probability $\nu\, dt$ that any unfixed site loses its competence in the small time interval dt, then the new probability that a site will become fixed in the time interval $(t, t + dt)$ is $f^*(t) = f(t)\, e^{-\nu t}$, where $f(t)$ is defined as before. This simply includes the extra term $e^{-\nu t}$ for the probability that the site is still competent at time t. If we allow, further, for an influx of competent sites at a rate $\bar{\nu}\, dt$, then the probability that a site becomes fixed in the time interval $(t, t + dt)$ is now

$$(4.12) \qquad f^{**}(t) = \int_0^t \bar{\nu}\, e^{-\bar{\nu}\tau_1}\, e^{-\nu(t - \tau_1)} f(t - \tau_1)\, d\tau_1.$$

This is the convolution of the probability that a site becomes competent at time τ_1, remains competent for a time $t - \tau_1$ and given that it remains competent, becomes fixed at a time $t - \tau_1$ after becoming competent. The Laplace transform of $f^{**}(t)$ is thus given by $[\bar{\nu}/(s + \bar{\nu})]\, \varphi(s + \nu)$ where as before $\varphi(s)$ is the transform of $f(t)$. If at time $t = 0$ there are B_c competent cells and a population B of cells can become competent with probability $\bar{\nu}\, dt$ in the small time interval dt, then assuming one site per cell the total number of cells that become fixed in $(t, t + dt)$ is

$$(4.13) \qquad B_c f^*(t) + Bf^{**}(t),$$

which has a Laplace transform

$$(4.14) \qquad \varphi(s + \nu)\left(B_c + \frac{B\bar{\nu}}{s + \bar{\nu}}\right).$$

So long as $\bar{\nu}$ and ν are small the asymptotic properties of this model are similar to those already discussed above.

5. The application of diffusion theory to the growing point model

The kinetics of absorption of virus particles to bacterial cells was explained by Schlesinger [26] in terms of a two body collison model involving freely diffusing virus particles and bacterial cells. The model he used was based on a theory of the kinetics of coagulation developed by Smoluchowsky. The theory predicts that the quantity of a solute $J\ dt$ which diffuses onto a sphere of radius R in a small interval of time dt is given by

$$(5.1) \qquad\qquad J\ dt = 4\pi\ DRc\ dt,$$

where D is the diffusion coefficient of the solute particles, and c their concentration. If the diffusion of the bacterial cells is ignored, then equation (5.1) is applicable to transformation on the assumptions that the solute is DNA, that R is the mean radius of a sphere whose volume is approximately equal to that of the bacterial cell, and that a contact anywhere on the surface of the cell leads to the location of the attachment site. Since the average DNA molecule will be long compared to R, this latter assumption is perhaps not too unreasonable. Following the growing point model developed in the previous section, we should then have

$$(5.2) \qquad\qquad \mu = 4\pi\ DRc,$$

where J is measured in numbers of DNA molecules, since $\mu\ dt$ is the probability of a molecule contacting a competent cell in the small time interval dt. A typical concentration of DNA used in a transformation experiment is 1 μg/ml or $(10^{-6}/2 \times 10^{7}) \times 6 \times 10^{23} = 3 \times 10^{10}$ molecules/ml where we assume an average molecular weight of 2×10^{7}. The radius R may be taken as approximately 5×10^{-5} cm and D as approximately 10^{-8} cm^{-2} sec^{-1} (see, for example, Tanford [28]). These values give $\mu = 1.9 \times 10^{-1}$ corresponding to a mean interval of about five secs between collisions between bacteria and DNA molecules. From our knowledge of the average size of the DNA molecules we can take $\beta = .01$. If now we assume the asymptotic formula (4.11) for the probability that a cell is not fixed by time t, the approximate time needed to produce half the saturating level of transformants is given by $(t\lambda\mu\beta/\lambda + \mu) = \log_e 2$ which leads to the following estimate for λ

$$(5.3) \qquad\qquad \lambda = \frac{\mu \log_e 2}{t\mu\beta - \log_e 2}.$$

A reasonable value for t is 15 min or 900 sec, which with the previous estimates of μ and β gives $\lambda = 0.082$ corresponding to a mean time of about 12 sec for the length of time an unacceptable molecule remains attached to an attachment site. This may be compared with the average time taken to replicate a portion of the genome corresponding to the size of the donor molecule, which is about 30 sec.

Assuming the specificities of attachment sites are random with respect to chromosomal location, the proportion of the whole genome available at an

attachment site u will be related to the proportion of transformants for a given marker p and the proportion b of the recipient cells which have an attachment site (that is, are competent) by the formula

$$(5.4) \qquad\qquad ub = p.$$

Saturating transformation levels of one to two per cent are frequently attained in the *B. Subtilis* transformation system, on which most of this discussion is based. Thus, since $b \leq 1$, we must have $u \geq 1/50$ to $1/1000$. From experimental considerations it seems unlikely that b is greater than about 0.2, suggesting a value for u of from 0.05 to 0.1.

While the calculations described in this section are undoubtedly very rough they do not indicate any major discrepancy between the requirements of the growing point model and experimental observations.

6. Recombination theory

The probability that a single marker will be integrated once the appropriate molecule has been fixed will simply be a further constant multiplier relating the frequency of transformants to the frequency of competent cells fixed for the appropriate DNA molecules. This frequency will, of course, depend in some way on the properties of the DNA molecules on which the marker is situated. However, to relate the relative frequencies with which single and double transformants are produced for markers which can be on the same molecule (that is, are linked) requires some assumptions concerning the nature of the recombination process leading to integration. A simple recombination theory will be described in this section which is similar to that discussed by Balassa and Prevost [3].

Following classical theories of recombination as developed originally by Haldane [15] (see also Bailey [2]), we assume that integration is mediated by a series of random switches in the association of donor DNA with the recipient chromosome. Thus, in figure 4, OT represents a donor fragment temporarily
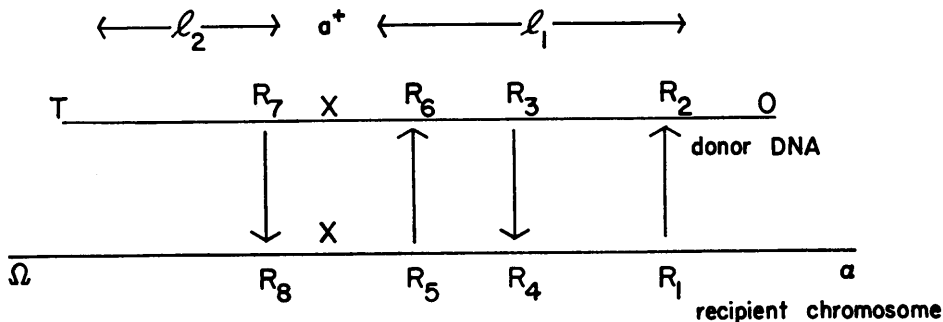

FIGURE 4

Schematic representation of integration by recombination.

aligned with its homologous region on the recipient chromosome $\alpha\Omega$, and X the position of the mutant lesion in the recipient chromosome (a^-) distinguishing it from the donor (a^+). The points R_i represent homologous positions on the donor and recipient where an exchange in the associations of donor and recipient material has occurred. Thus, for the particular configuration of figure 4, integration gives rise to a recipient chromosome $\alpha R_1 R_2 R_3 R_4 R_5 R_6 R_7 R_8 \Omega$ which has donor fragments $R_2 R_3$ and $R_6 R_7$ replacing the homologous recipient regions $(R_1 R_4$ and $R_5 R_8$, respectively). Transformation will be achieved only if a region including X is integrated into the recipient chromosome. In terms of DNA structure it seems probable (for example, see Bodmer and Ganesan [6]), that $0T$ and $\alpha\Omega$ represent single strands of a DNA molecule having the same polarity and that the exchange process is mediated by an exchange in pairing of donor and recipient strands with their complementary strands, followed by breakage and rejoining. A viable transformation will only be produced if the number of exchanges within the interval $0T$ is even, so that both the α and Ω ends of the recipient chromosome are included in the final product. There are two distinct possibilities, either (1) all products of an odd number of exchanges die or are simply not observed or (2) there is a mechanism which forces an even number of exchanges. Experimental observations on the efficiency with which integrated donor DNA gives rise to transformants, as well as general theoretical considerations, favor the second of these two possibilities. For example, switching may be a directed process, starting from α and proceeding toward Ω, which is such that if an odd number of switches has occurred by the time T is reached, a further switch back to the recipient chromosome is forced at the terminal point T.

Assume now that the probability of the switch occurring in a small increment of length dx is simply dx for all positions on the donor fragment. The number of switches occurring in the interval x will then have a Poisson distribution. This implies that switches occur independently at any position and that length is measured by a recombination metric which is such that an average of one switch occurs per unit length. It is clear that this metric will, in general, be monotonically related to the actual physical distance along the DNA molecule as measured in nucleotide pairs. Recombination data from many organisms suggests that the relationship may in fact be one of simple proportionality. The probability that an odd number of switches occurs in the length x is $(1 - e^{-2x})/2$, which is Haldane's [15] classic formula relating recombination frequency and "map" length. Given a particular molecule with the lengths $0X = \ell_1$, $XT = \ell_2$, and $\ell_1 + \ell_2 = L$, as in figure 4, there are at least three possible expressions for the probability that the marker a^+ will be integrated, depending on the assumed model.

Model 1. If all products of an odd number of exchanges in $0T$ are lost, the integration probability is $\{[1 - \exp(-2\ell_1)]/2\}\{[1 - \exp(-2\ell_2)]/2\}$, this being the probability that an odd number of switches occurs in $0X$ and in XT.

Model 2. Switching occurs sequentially, at random, along the donor frag-

ment starting at 0 and proceeding toward T. If there is still an odd number of switches by the time T is reached, then a further switch at T is forced. The integration probability is now simply $(1 - e^{-2\ell_1})/2$, since if an odd number of switches has occurred in $0X$, the model forces an odd number in XT and so an even number in $0T$.

Model 3. As in model 2, except that the switching process can start with equal probability, either at the end 0 or at the end T. The integration probability is now $[1 - \exp(-2\ell_1) + 1 - \exp[(-2\ell_2)]/4$.

Model 2 is the only one giving a polarity effect, such that the probability of integration is not the same when $0X = \ell_1$ and $XT = \ell_2$ as when $0X = \ell_2$ and $XT = \ell_1$ where $\ell_1 \neq \ell_2$. For a given molecule, model 2 gives a probability of integration which increases monotonically as the distance $0X$ increases. Models 1 and 3 give a maximum probability of integration when $\ell_1 = \ell_2 = L/2$ for molecules of fixed length $\ell_1 + \ell_2 = L$. Only models 2 and 3 will be considered for the rest of this section.

In the introduction it was pointed out the DNA preparations used for transformation are probably equivalent to a random chopping of the bacterial chromosome into some hundred fragments. We cannot, therefore, assume that the marker we are interested in has a defined position on all molecules with given length. We must assume a probability distribution for ℓ_1 and ℓ_2 given that $\ell_1 + \ell_2 = L$. Most models simply assuming random "chopping"; we make the assumption that ℓ_1 and ℓ_2 are distributed uniformly in the interval $(0, L)$. This gives an average integration probability

$$(6.1) \qquad r(L) = \frac{1}{2}\left[1 - \frac{1}{2L}(1 - e^{-2L})\right], \qquad \bullet$$

for both models 2 and 3. When L is small

$$(6.2) \qquad r(L) \sim \frac{L}{2} + O(L^2)$$

and as $L \to \infty$, $r(L) \to 1/2$. Thus, with this simple model, the maximum integration probability is $1/2$. We have

$$(6.3) \qquad \frac{dr}{dL} = \frac{1}{4L^2} - \frac{1}{4L^2}e^{-2L} - \frac{1}{2L}e^{-2L}$$

$$= \frac{e^{-2L}}{4L^2}\left[e^{2L} - (1 + 2L)\right] \geqq 0$$

for all values of L. Thus, $r(L)$ increases monotonically from 0 to $1/2$ as L increases from 0 to ∞.

We shall now consider recombination for two markers X_1, X_2 whose distance apart ℓ is less than the length L of donor molecules. Then, as indicated in figure 5, there are three classes of molecules of length L which carry one or both of the markers X_1, X_2. The first class has its right end 0 falling in the interval X_1X_2 on the chromosome and carries only marker X_1, the second with its left end T in
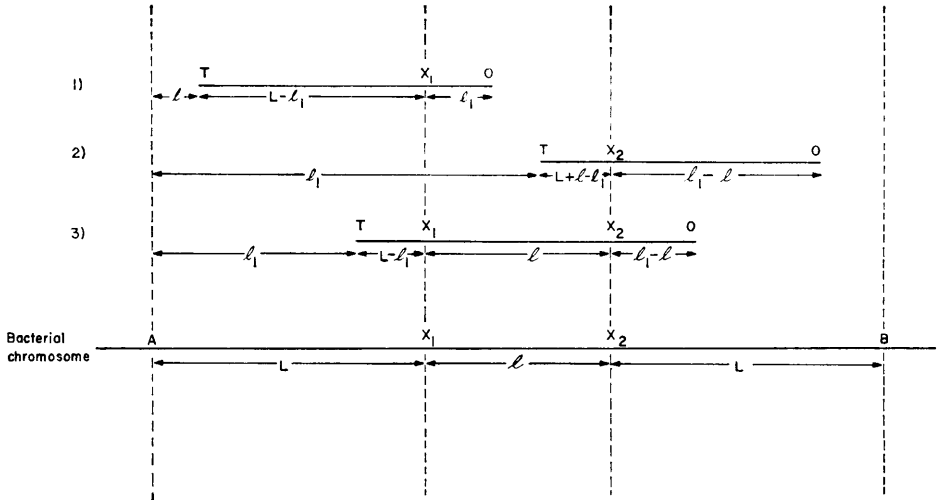
FIGURE 5

Recombination scheme for a pair of linked markers.

X_1X_2 carries only marker X_2, while the third with both ends outside the interval X_1X_2 carries both markers. If we assume random chopping, the position of the end T is uniformly distributed in the interval AX_2 on the recipient chromosome. Preferential breakage points would, of course, give rise to different distributions of the position of T. Assuming random chopping the relative proportions of the three types of molecules are $\ell/(L + \ell)$, $\ell/(L + \ell)$, and $(L - \ell)/(L + \ell)$. When $\ell \ll L$, molecules of type 3, containing both markers, predominate. Assuming recombination model 2, the probability that both markers X_1 and X_2 are integrated from molecules of type 3 is

$$(6.4) \qquad r_{33} = \frac{1}{2}(1 + e^{-2\ell})\frac{1}{2}[1 - e^{-2(\ell_1 - \ell)}],$$

since this requires an odd number of switches in $0X_2$ and an even number in X_1X_2. The probabilities of integrating, singly, the markers X_1 and X_2 from molecules of type 3 are, analogously

$$r_{31} = \frac{1}{2}(1 - e^{-2\ell})\frac{1}{2}[1 + e^{-2(\ell_1 - \ell)}],$$

$$(6.5)$$

$$r_{32} = \frac{1}{2}(1 - e^{-2\ell})\frac{1}{2}[1 - e^{-2(\ell_1 - \ell)}],$$

respectively. The three corresponding probabilities for model 3 are clearly

$$\frac{1}{8}(1 + e^{-2\ell})[1 - e^{-2(\ell_1 - \ell)} + 1 - e^{-2(L - \ell_1)}],$$

$$(6.6) \qquad \frac{1}{8}(1 - e^{-2\ell})[1 + e^{-2(\ell_1 - \ell)} + 1 - e^{-2(L - \ell_1)}],$$

$$\frac{1}{8}(1 - e^{-2\ell})[1 - e^{-2(\ell_1 - \ell)} + 1 + e^{-2(L - \ell_1)}],$$

respectively. The integration probabilities r_1, for marker X_1 from type 1 molecules, and r_2, for marker X_2 from type 2 molecules, are as given before for single markers.

If now we wish to distinguish single and double transformation events, we have to average over distributions of ℓ_1 which are appropriate for the three different classes of molecules. Again, assuming random chopping, ℓ_1 will in all cases be uniformly distributed, but over different intervals. For the three types of molecules the intervals will be $(0, \ell)$, $(L, L + \ell)$, and (ℓ, L), respectively, where $L \geqq \ell$. The average values of all the integration probabilities under these assumptions are given in table I. When $L < \ell$,

$$(6.7) \quad r_{31} = r_{32} = r_{33} = 0 \quad \text{and} \quad r_1 = r_2 = \frac{1}{2} \left[1 - \frac{1}{2L} (1 - e^{-2L}) \right]$$

as before. There is polarity with respect to the frequency of integration of single markers when the different molecular types are considered separately. However, since

$$(6.8) \quad \frac{\ell}{L + \ell} (r_2 - r_1) = \frac{(1 - e^{-2\ell})}{4(L + \ell)} [1 - e^{-2(L-\ell)}] = \frac{-(L - \ell)}{L + \ell} (r_{31} - r_{32}),$$

the polarity effects cancel each other. For given L ($\geqq \ell$), the average frequency of single transformants (either marker) is

$$(6.9) \quad T_S = \frac{\ell}{L + \ell} (r_1 + r_2) + \frac{L - \ell}{L + \ell} (r_{31} + r_{32}),$$

and of double transformants

$$(6.10) \quad T_D = \frac{L - \ell}{L + \ell} r_{33} = \frac{1}{4} (1 + e^{-2\ell}) \frac{L - \ell}{L + \ell} \left\{ 1 - \frac{1}{2(L - \ell)} [1 - e^{-2(L-\ell)}] \right\}$$

for both models. Substituting for r_1, and so forth, in equation (6.9) we have, for both models,

$$(6.11) \quad T_S = \frac{\ell}{2(L + \ell)} \left[2 - \frac{1}{2\ell} (1 - e^{-2\ell}) - \frac{e^{-2L}}{2\ell} (e^{2\ell} - 1) \right]$$

$$+ \frac{L - \ell}{2(L + \ell)} (1 - e^{-2\ell})$$

$$= \frac{\ell}{L + \ell} + \frac{(1 - e^{-2\ell})}{4(L + \ell)} [2(L - \ell) - 1 - e^{-2(L-\ell)}].$$

When $L - \ell$ is small

$$(6.12) \quad T_D = (L - \ell)^2 \frac{(1 + e^{-2\ell})}{4(L + \ell)} + O(L - \ell)^3,$$

while

$$(6.13) \quad T_S = \frac{\ell}{L + \ell} + (1 - e^{-2\ell}) \frac{L - \ell}{L + \ell} + O(L - \ell)^2.$$

Thus, for markers sufficiently far apart, it should be possible to find conditions where a reduction in the average length of donor molecules L causes a drop in the frequencies of double transformants but hardly affects the frequency of single transformants. Knowing the value of L, in physical terms, at which this

TABLE I

INTEGRATION PROBABILITIES FOR TWO MARKERS CORRESPONDING TO THE
SCHEME GIVEN IN FIGURE 5

In each case the first term is the probability given ℓ_1 and the second is the probability averaging over ℓ_1 in the appropriate interval.

Molecule Type	Model 2 (polar)	Model 3 (nonpolar)
Type 1. Proportion $\ell/(L+\ell)$ r_1 = probability of integrating only X_1 ℓ_1 uniform in $(0,\ell)$	$\dfrac{1}{2}(1 - e^{-2\ell_1})$ $\dfrac{1}{2}\left[1 - \dfrac{1}{2\ell}(1 - e^{-2\ell})\right]$	$\dfrac{1}{4}[2 - e^{-2(L-\ell_1)}) - e^{-2\ell_1}]$ $\dfrac{1}{4}\left[2 - \dfrac{1}{2\ell}(1 - e^{-2\ell}) - \dfrac{e^{-2L}}{2\ell}\right.$ $\left. (e^{2\ell} - 1)\right]$
Type 2. Proportion $\ell/(L+\ell)$ r_2 = probability of integrating only X_2 ℓ_1 uniform in $(L, L+\ell)$	$\dfrac{1}{2}[1 - e^{-2(\ell_1-\ell)}]$ $\dfrac{1}{2}\left[1 - \dfrac{e^{-2L}}{2\ell}(e^{2\ell} - 1)\right]$	same as r_1
Type 3. Proportion $(L-\ell)/(L+\ell)$ ℓ_1 uniform in $[\ell, L)$ r_{31} = probability of integrating only X_1	$\dfrac{1}{4}(1 - e^{-2\ell})[1 + e^{-2(\ell_1-\ell)}]$ $\dfrac{1}{4}(1 - e^{-2\ell})\left\{1 + \dfrac{1}{2(L-\ell)}\right.$ $\left. [1 - e^{-2(L-\ell)}]\right\}$	$\dfrac{1}{8}(1 - e^{-2\ell})[2 + e^{-2(\ell_1-\ell)}$ $- e^{-2(L-\ell_1)}]$ $\dfrac{1}{4}(1 - e^{-2\ell})$
r_{32} = probability of integrating only X_2	$\dfrac{1}{4}(1 - e^{-2\ell})[1 - e^{-2(\ell_1-\ell)}]$ $\dfrac{1}{4}(1 - e^{-2\ell})$ $\left\{1 - \dfrac{1}{2(L-\ell)}[1 - e^{-2(L-\ell)}]\right\}$	$\dfrac{1}{8}(1 - e^{-2\ell})$ $[2 - e^{-2(\ell_1-\ell)} + e^{-2(L-\ell_1)}]$ same as r_{31}
r_{33} = probability of integrating only X_1 and X_2	$\dfrac{1}{4}(1 + e^{-2\ell})[1 - e^{-2(\ell_1-\ell)}]$ $\dfrac{1}{4}(1 + e^{-2\ell})$ $\left\{1 - \dfrac{1}{2(L-\ell)}[1 - e^{-2(L-\ell)}]\right\}$	$\dfrac{1}{8}(1 + e^{-2\ell})$ $[2 - e^{-2(\ell_1-\ell)} - e^{-2(L-\ell_1)}]$ same as for model 2

change in the kinetics of double versus single transformation occurs might be one way of relating directly the genetic and physical length metrics. The frequency T_D increases monotonically from 0 as $L - \ell$ increases or equivalently for fixed L, as ℓ decreases, and so is directly related to the distance between the markers. We require, however, an estimate for this distance ℓ, which is independent of L. A commonly used parameter is the cotransfer index [22]

$$(6.14) \qquad C = \frac{T_D}{T_S + T_D}.$$

From equations (6.10) and (6.11), we have that

$$(6.15) \quad 1 - C = \frac{8\ell + 2(1 - e^{-2\ell})[2(L - \ell) - 1 - e^{-2(L-\ell)}]}{8\ell + (3 - e^{-2\ell})[2(L - \ell) - 1) - (1 - 3e^{-2\ell}]e^{-2(L-\ell)}}.$$

When ℓ is small, this reduces to

$$(6.16) \quad 1 - C = 2\ell \frac{2L + 1 - e^{-2L}}{2L - 1 + e^{-2L}} + O(\ell^2) \qquad \text{or} \qquad 1 - C \sim 2\ell \text{ if } L \gg 1.$$

Thus, for markers which are very close together, $1 - C$ is a reasonable measure of the genetic distance between them. When ℓ is small

$$(6.17) \qquad T_D \sim \frac{1}{2}\left[1 - \frac{1}{2L}(1 - e^{-2L})\right],$$

which is the usual single marker transformation frequency, and $T_S/T_D \sim (1 - C) \sim 2\ell$, since T_S is small for closely linked markers. T_S can often be measured rather more easily than T_D using appropriate selective techniques. However, T_D can then be estimated from the transformation frequency, determined in the same experiment, for some other unlinked marker. This approach to estimating small distances in transformation experiments by using the transformation frequency of an unlinked marker to standardize the frequency of recombinant, single transformants for pairs of closely linked members has been used, for example, by Sicard and Ephrussi-Taylor [27].

7. Discussion

The complete transformation process is the product of a number of constituent processes which, as discussed in previous sections, fall into two categories. The first is the combination of attachment, entry and synapsis and the second recombination and expression. Attachment models account for the main features of transformation for single markers. However, for joint transformation of linked markers, the interpretation of the recombination mechanisms becomes of some interest. It is important to emphasize that the parameters for each stage of the process are functions of the integrity of the DNA preparation used as well as of the competent state of the recipient cells. Variations in the size of DNA molecules (before and after denaturation) are particularly significant in this respect [5]. Moreover, while all our discussions on recombination focused

on molecules of fixed size L, there is, as pointed out already, considerable heterogeneity with respect to the size of molecules in the standard preparations of DNA used for transformation experiments. In summary, then, the complete transformation process may be described, for molecules of given size L, by the product

(7.1) $T(L) = [1 - F(t, \lambda, \mu, \beta)] \, r(L) \, m(L) \, e,$

where the first factor is the probability of fixation of a favorable molecule, the second is the probability of the appropriate recombinational event, and the third and fourth are probabilities of expression. The quantity L basically refers to the single stranded size of the DNA, which is measured by the size after denaturation. The parameters λ, μ, β may all be functions of L, while $m(L)$ is that part of the probability of the expression which is a function of L. The ultimate probability of transformation will then be $\bar{T} = \int_0^\infty T(L)\rho(L) \, dL$, where $\rho(L)$ is the probability density for molecules of size L in the donor DNA preparation. Thus, for example, to describe the effect of a given treatment of the DNA on transformation requires the specification of the effect of the treatment on $\rho(L)$ and hence on $\bar{T}$. The effects on transformation of deoxyribonuclease treatment of the DNA have been extensively studied (see, for example, [5]). Some deoxyribonucleases create breaks in single strands of the DNA molecule and so effectively reduce L. It has been shown experimentally that such treatment causes a reduction in transformation frequencies which is directly correlated with the reduction in the effective value of L, in this case the distance between breaks on one strand of the DNA molecule. This reduction has been shown to be mediated through an effect on F as well as m and presumably r. A description of the change in $\rho(L)$ when scision is at random goes back to Charlesby [9], though Bodmer [5] and Bresler, Kolinin, and Perumov [7] have made some further generalizations. The problem is closely analogous to the analysis of neutron cascades (see, for example, T. E. Harris [16]). The complete quantitative interpretation of these effects is, however, likely to be very complex and would probably add very little to their understanding.

There are undoubtedly numerous possible extensions and generalizations of the models we have presented. Thus the possibility of interference between adjacent switches has not been taken into account in the recombination theory. This could presumably be included in a manner analogous to the theory which has been developed by Fisher and Owen (see Bailey [2]). Balassa and Prevost [3] following Hotchkiss [17] have considered the effect on integration and recombination of assuming the genetic marker has a finite length rather than being a point. However, their model for recombination, as well as that sketched by Bresler, Kolinin, and Perumov [7] does not take into account the fact that the DNA molecules probably do not have defined ends. The whole process of transformation is sufficiently complex that a complete quantitative description, even if the basis for constructing it were fully understood, would involve too many

unmeasurable parameters and would be quite unwieldy. The value of model building is in formalizing and clarifying ideas on the mechanisms involved and in highlighting the expected qualitative features of the observed quantitative experimental observations.

8. Summary

The transformation process can be divided into at least five stages: (1) initial attachment; (2) entry; (3) synapsis; (4) recombination (or integration); and (5) expression. Mathematical models fall basically into two categories: (1) models for the initial attachment, entry and synapsis of donor DNA and (2) models for the recombination process.

Simple models for the attachment process based on enzyme kinetic theory are discussed first. These illustrate the initial linear dependence of the number of transformants for a single marker on the time of contact, except possibly for very short times, and also on the DNA concentration. The number of double transformants for two unlinked markers is, in the initial stages of contact, proportional to the square of the time and of the DNA concentration.

A more specific probabilistic model is next discussed which involves random collisions of DNA molecules of different types with a given number of receptor sites on the recipient bacterium. This gives essentially the same results as the simpler kinetic theory.

Recent evidence on the involvement of DNA synthesis in transformation suggests that donor DNA is integrated at a stationary growing point of DNA synthesis. This implies that there is only one, or at most a few, attachment sites on the recipient cell and that only certain donor molecules are accepted by any given attachment site. A probabilistic model incorporating these features is described and analyzed. Application of diffusion theory to the model suggests rate constants for the attachment process which are not incompatible with the observed time kinetics of transformation.

A recombination theory is outlined in which it is assumed that integration is mediated by a series of random switches in the association of donor DNA with the recipient chromosome. Expressions are given for the dependence of single and double transformation frequencies for linked markers on the distance between the markers and on the average length of the molecules. The overall quantitative description of the transformation process, combining these various models, is discussed.

REFERENCES

[1] O. T. AVERY, C. M. MACLEOD, and M. McCARTY, "Studies on the chemical nature of the substance inducing transformation of Pneumococcal types. Induction of transformation by a deoxyribonucleic acid fraction isolated from Pneumococcus type. III," *J. Exp. Med.*, Vol. 79 (1944), pp. 137–158.

[2] N. T. J. BAILEY, *Introduction to the Mathematical Theory of Genetic Linkage*, Oxford, Oxford University Press, 1961.

[3] G. BALASSA and G. PREVOST, "Etude théorique de la transformation bactérienne," *J. Theor. Biol.*, Vol. 3 (1962), pp. 315–344.

[4] W. BODMER, "Recombination and integration in *Bacillus subtilis:* involvement of DNA synthesis," *J. Molec. Biol.*, Vol. 14 (1965), pp. 534–557.

[5] ———, "Integration of deoxyribonuclease treated DNA in *Bacillus subtilis* transformation," *J. Gen. Physiol.*, Vol. 49 (1966), pp. 233–258.

[6] W. F. BODMER and A. T. GANESAN, "Biochemical and genetic studies of integration and recombination in *Bacillus subtilis* transformation," *Genetics*, Vol. 50 (1964), pp. 717–728.

[7] S. E. BRESLER, V. L. KOLININ, and D. A. PERUMOV, "Theory of inactivation and reactivation of transforming DNA," *Biopolymers*, Vol. 2 (1964), pp. 135–146.

[8] L. F. CAVALIERI and B. H. ROSENBERG, "The interpretation of bacterial transformation data," *Proc. Nat. Acad. Sci., U.S.A.*, Vol. 44 (1958), pp. 853–857.

[9] A. CHARLESBY, "Molecular-weight changes in the degradation of long-chain polymers," *Proc. Roy. Soc. Ser. A*, Vol. 224 (1954), pp. 120–128.

[10] M. S. FOX and R. D. HOTCHKISS, "Initiation of bacterial transformation," *Nature*, Vol. 179 (1957), pp. 1322–1325.

[11] M. S. FOX and M. K. ALLEN, "On the mechanism of deoxyribonucleate integration in Pneumococcal transformation," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 52 (1964), pp. 412–419.

[12] A. T. GANESAN and J. LEDERBERG, "A cell-membrane bound fraction of bacterial DNA," *Biochem. Biophys. Res. Commun.*, Vol. 18 (1965), pp. 824–835.

[13] S. H. GOODGAL, "Studies on transformations of *Hemophilus influenzae*. IV. Linked and unlinked transformations," *J. Gen. Physiol.*, Vol. 45 (1961), pp. 205–228.

[14] S. H. GOODGAL, and R. M. HERRIOTT, "Studies on transformations of *Hemophilus influenzae*. I. Competence," *J. Gen. Physiol.*, Vol. 44 (1961), pp. 1201–1227.

[15] J. B. S. HALDANE, "The combination of linkage values and the calculation of distance between the loci of linked factors," *J. Genet.*, Vol. 8 (1919), pp. 299–309.

[16] T. HARRIS, *The Theory of Branching Processes*, Berlin, Springer-Verlag, 1963.

[17] R. D. HOTCHKISS, "Size limitations governing the incorporation of genetic material in the bacterial transformations and other non-reciprocal recombinations," *The Biological Replication of Macromolecules*, Symposium of the Society for Experimental Biology, 1958, No. 12, pp. 49–59.

[18] R. D. HOTCHKISS and J. MARMUR, "Double marker transformations as evidence of linked factors in deoxyribonucleate transforming agents," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 40 (1954), pp. 55–60.

[19] J. L. KENT and R. D. HOTCHKISS, "Kinetic analysis of multiple linked recombinations in Pneumococcal transformation," *J. Molec. Biol.*, Vol. 9 (1964), pp. 308–322.

[20] L. S. LERMAN and L. J. TOLMACH, "Genetic transformation I. Cellular incorporation of DNA accompanying transformation in Pneumococcus," *Biochim. Biophys. Acta*, Vol. 26 (1957), pp. 68–82.

[21] J. S. LEVINE and N. STRAUSS, "Lag period characterizing the entry of transforming deoxyribonucleic acid into *Bacillus subtilis*," *J. Bact.*, Vol. 89 (1965), pp. 281–287.

[22] E. W. NESTER and J. LEDERBERG, "Linkage of genetic units of *Bacillus subtilis* in DNA transformation," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 47 (1961), pp. 52–55.

[23] E. W. NESTER, M. SCHAFER, and J. LEDERBERG, "Gene linkage in DNA transfer: a

cluster of genes concerned with aromatic biosynthesis in *Bacillus subtilis*," *Genetics*, Vol. 48 (1963), pp. 529–551.

[24] E. W. NESTER and B. A. D. STOCKER, "Biosynthetic latency in early stages of deoxyribonucleic acid transformation in *Bacillus subtilis*," *J. Bact.*, Vol. 86 (1963), pp. 785–796.

[25] P. SCHAEFFER, "Transformation," *The Bacteria, Volume V: Heredity* (edited by I. Gunsalus and R. Stanier), New York, Academic Press, 1964, pp. 87–154.

[26] M. SCHLESINGER, "Adsorption of bacteriophages to homologous bacteria. II. Quantitative investigations of adsorption velocity and saturation. Estimation of particle size of the bacteriophage," *Z. Hygenie Immunitaets Forschung*, Vol. 114 (1932), pp. 149–160 (in German).

[27] A. M. SICARD and H. EPHRUSSI-TAYLOR, "Genetic recombination in DNA-induced transformation of Pneumococcus II. Mapping the ami A region," *Genetics*, Vol. 52 (1965), pp. 1207–1227.

[28] C. TANFORD, *Physical Chemistry of Macromolecules*, New York, Wiley, 1961.

[29] R. THOMAS, "Recherches sur la cinétique des transformations bactériennes," *Biochim. Biophys. Acta*, Vol. 18 (1955), pp. 467–481.