THE THEORY OF FORMAL PHAGE GENETICS FOR CIRCULAR MAPS[1]

FRANKLIN W. STAHL AND C. M. STEINBERG

Department of Biology and Institute of Molecular Biology, University of Oregon, Eugene and Biology Division, Oak Ridge National Laboratory,[2] Oak Ridge, Tennessee

Received April 16, 1964

STREISINGER, EDGAR, and HARRAR DENHARDT (1964) have shown that the genetic map of T4 is topologically equivalent to a circle. That is, any small region of the genome yields a conventional linear map, but the map as a whole has no ends. Thus, the possibility of genetically "silent" regions at the ends of the map is excluded, and estimates of the total map length of T4 can be unbiased estimates rather than minimal estimates. The purpose of this paper is to extend the theory of formal phage genetics to cover the case of circular maps while a companion paper (STAHL, EDGAR, and STEINBERG 1964) addresses itself more specifically to the problem of obtaining an estimate of total map length.

Before embarking upon algebra and arithmetic, it behooves us to state in nonmathematical terms what we intend to do. First, we will remain within the framework of the so-called "mating theories." Mating theories presume that recombination occurs during discrete, successive interactions ("matings") between or among entire (monoploid) phage genomes. The assumptions that the interactions are discrete and that there are no intermediates in the recombinational process other than entire phage genomes are the features of the mating theories which make the recombinational process easy to visualize and mathematically tractable. Whether or not there is any other justification for mating theories is moot. Second, as in previous treatments of mating theories (VISCONTI and DELBRÜCK 1953; STEINBERG and STAHL 1958), we will make no attempt to account for high negative interference (CHASE and DOERMANN 1958; EDGAR and STEINBERG 1958) or heterozygosis (HERSHEY and CHASE 1951).

Now the most straightforward interpretation of the circular genetic map is that the phage chromosome is a circle, at least at the time of mating. This interpretation of the circular map is also easy to treat within the above ground rules. We need only specify the probability of exchange as a function of distance for a single mating and then combine this with our previous results to obtain a mapping function. A mating between two circular chromosomes is limited to an even number of exchanges if only entire genomes are to emerge.

(If matings are by groups, the restrictions for a circular map are that *each genome* in the mating group must indulge in an even number of exchanges and that the spatial pattern of exchanges must be such as to effect a return to the

[1] Supported in part by a research grant (GB-294) from the National Science Foundation to F.S.
[2] Operated by Union Carbide Corporation for the United States Atomic Energy Commission.

original genome. We shall argue later that the number of exchange events per mating is small. The combined requirements of a small number of exchanges per mating and obligate return to the original genome robs the notion of group mating of most of its original meaning (STEINBERG and STAHL 1958; HERSHEY 1958). Our presentation here, therefore, will confine itself to a consideration of *pairwise* mating.)

Further reflection shows that the requirement for an even number of exchanges by itself leads to circularity of the map. In fact, if we assume that the chromosome is structurally a rod but nevertheless requires an even number of exchanges, we obtain formally identical results. We will refer to the above models as the "closed circle" and "rod," respectively.

While we cannot explicitly account for heterozygosis in our mating models, recent experiments with heterozygotes (STREISINGER, personal communication) have led to another interpretation of the circular genetic map. This interpretation, which we call the "open circle" model, postulates that the phage chromosome is structurally a rod, but the sequence of genes is not the same in all members of a population of chromosomes. One arrives at a circular map by requiring that the gene sequences within a population of chromosomes are all circular permutations of one another. One can construct such a population of permuted rods by making one randomly placed break in each member of a population of circles. The open circle model leads to results which are formally distinct from the closed circle-rod model.

After deriving mapping functions for the above models, we estimate numerical values for the parameters which appear in these functions. The models are then tested against previously published data with the conclusion that each of the two models provides a fairly adequate description of the basic observations of a phage cross.

DEVELOPMENT OF THE FUNCTIONS

Generalized mating theory: In this paper we employ the symbolism of HERSHEY (1958) and the generalized approach of STEINBERG and STAHL (1958).

The frequency of recombinants in the mating pool (and among phages withdrawn at any instant from the mating pool) is described by

$$R = 2abf_1f_2(1-e^{-mp}) \qquad (1)$$

(HERSHEY 1958)

where a and b are the frequencies of the two parental phages, f_1 is the "finite input" factor of LENNOX, LEVINTHAL, and SMITH (1953), f_2 is a measure of the degree to which the population is panmictic, m is the number of matings per lineage, and p is the probability per (heterozygous) mating of recombining a given pair of markers.

The frequency of recombinants among mature phage will be

$$\bar{R} = 2abf_1f_2\left(1 + \frac{e^{-m_2p} - e^{-m_1p}}{p\,(m_2-m_1)}\right) \qquad (2)$$

(HERSHEY 1958)

where m_1 and m_2 are the average number of rounds of mating for the first and last particles to mature respectively.

Mating theory for closed circles and rods: We define a distance d on a chromosome of unit length as follows: (a) Closed circles: $d =$ the shorter of the two circumferential distances between two markers. (b) Rods: $d =$ the distance between two markers.

In each of the models, the requirement for recombination in a mating act is that the paired exchanges be so disposed that an odd number of exchanges occur both in the distance d and in the distance $1 - d$. Therefore, p, the probability of recombination for two markers separated by distance d will be a function of d and $1 - d$.

If the matings involve exactly one pair of exchanges (which assumption seems justifiable for T4) (see below), then

$$p = 2d(1 - d) \tag{3}$$

if the exchange events are randomly disposed on the genome.

Substitution of equations (3) into equations (1) and (2) yields

$$R = 2abf_1f_2\,(1 - e^{-2md(1-d)}) \tag{4}$$

and

$$\bar{R} = 2abf_1f_2\left(1 + \frac{e^{-2m_2d(1-d)} - e^{-2m_1d(1-d)}}{2d(1-d)\ (m_2 - m_1)}\right) \tag{5}$$

Mating theory for open circles: We shall develop one model for mating between open circles. We suppose that the population of genomes is composed of two classes. In one class the distance between two markers is d; in the other class the distance is $1 - d$. If we assume that each "circle" is "open" at a randomly chosen point, the two classes will be represented with frequencies of $(1 - d)$ and d respectively.

For each class we suppose that exchanges in elementary acts occur with strict positive interference—recombination probabilities are proportional to distance. This assumption minimizes the number of exchanges per lineage (see below). The observed negative interference which characterizes the progeny from a phage cross (see VISCONTI and DELBRÜCK 1953) is supposed to arise in this model from finite input, spread in maturation, and from the circularity of the map.

Thus, the open-circle mapping function is composed of two terms corresponding to two classes of particles occuring with frequencies d and $(1 - d)$ respectively. The p's for the two terms are respectively $1 - d$ and d, so that

$$R = 2abf_1f_2\,[d(1 - e^{-m(1-d)}) + (1 - d)\ (1 - e^{-md})] \tag{6}$$

and

$$\bar{R} = 2abf_1f_2\left[d\left(1 + \frac{e^{-m_2(1-d)} - e^{-m_1(1-d)}}{(m_2 - m_1)\ (1-d)}\right) + (1-d)\left(1 + \frac{e^{-m_2d} - e^{-m_1d}}{(m_2 - m_1)d}\right)\right] \tag{7}$$

ESTIMATION OF PARAMETERS

The variables a and b are adjusted at ½ in standard equal-input crosses. As pointed out by HERSHEY (1958), panmixia must hold for T-even phage, so we take $f_2 = 1$. A value of 0.9 for f_1 is appropriate for crosses involving high multiplicities of each parent (LENNOX, LEVINTHAL, and SMITH 1953).

In order to evaluate m_1 it is necessary to identify a pair of markers for which d is known. The data of DOERMANN and HILL (1953, Table 2) suggest that the closely linked triplet tu_{43}, m_{41}, and tu_{45} is opposite the closely linked pair r_{47} and r_{51}. The $\bar{R}$ values observed in the six appropriate crosses involving these markers are among the nine largest $\bar{R}$ values observed by DOERMANN and HILL (1953). We assume that the distance d between the triplet and the pair $= \frac{1}{2}$. At $m = m_1$ and $d = \frac{1}{2}$ equations (4) and (6) both reduce to $R_{\max} = 0.45\ (1 - e^{-m_1/2})$.

Premature lysis experiments by DOERMANN (1953) employing the marker pair r_{47}–tu_{43} (for which we have presumed $d = \frac{1}{2}$) show that among the first particles to appear the recombinant frequency is 0.32. Therefore, $m_1 = 2.5$. The only parameter remaining to be determined is m_2. Two methods for its estimation are available— (1) the factor of increase in $\bar{R}$ for close markers from early to normal lysis and (2) the value of $\bar{R}_{\max}$.

Data of DOERMANN (1953, Table 4) show a twofold increase in $\bar{R}$ for close markers from early to normal lysis. This result sets $m_2 = 7.5$ (or $\bar{m} = 5$ where $\bar{m} = (m_1 + m_2)/2$) for the closed circle-rod model. For the open circle model $m_2 = 9.5$ ($\bar{m} = 6$).

The average recombinant frequency of the six appropriate crosses involving markers for which $d = \frac{1}{2}$ is $(0.455 + 0.436 + 0.432 + 0.417 + 0.413 + 0.393)/6 = 0.424 = \bar{R}_{\max}$ (DOERMANN and HILL 1953). This result sets $m_2 = 12.5$ ($\bar{m} = 7.5$) for both models. The discrepancy between the two estimates is not serious.

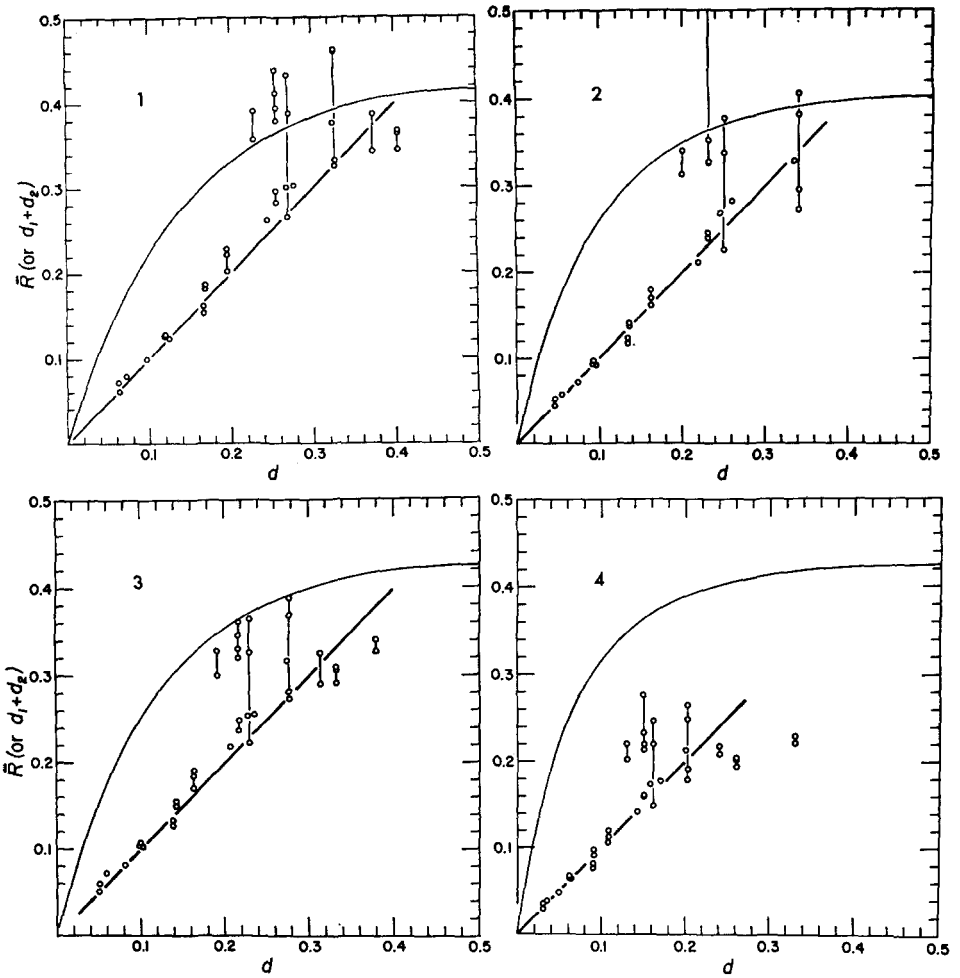
We have examined the two mapping functions (equations (5) and (7)) at each of the estimated m_2 values.

PROPERTIES OF THE FUNCTIONS

(a) *Graphs of the functions:* Figures 1 to 4 show plots of $\bar{R}$ vs d for the two models at the estimated values for m_1 and m_2.

(b) *Exchanges per lineage:* In the closed circle model we have assumed two exchanges per mating so that the average number of exchanges per lineage is $2(\bar{m})$. The open circle model assumes one exchange per mating. In addition one "event" per lineage determines the opening point of the circle. This event "unlinks" markers on either side of it and thus acts like an exchange point so that the total number of "exchanges" per lineage is about $(\bar{m} + 1)$. The total exchanges range from 7 (for open circles at $\bar{m} = 6$) to 15 (for closed circles at $\bar{m} = 7.5$). Examination of other properties of the two mapping functions indicates that the actual total number of exchanges probably lies within this range (see *Circular additivity test*, below).

(c) $\bar{R}_{\max}$: An $\bar{m}$ value of 7.5 was selected for both models in order to fit the experimental $\bar{R}_{\max}$ value of 0.424. At $\bar{m} = 5$ for the closed- and 6 for the open-circle functions we get 0.403 and 0.414 respectively for $\bar{R}_{\max}$. It seems probable that all of these values are within the range of experimental uncertainty in the data of DOERMANN and HILL (1953).



FIGURES 1–4.—The curves are plots of $\bar{R}$ versus d as described by equations (4) and (6). $\bar{R}$ values from laboratories of A. H. DOERMANN and R. S. EDGAR were converted to d values by use of the curves. The d values for adjacent regions were then added together ($d_1 + d_2$) and plotted versus the d value for the outside markers. Sets of points which share a common estimate of d are connected by a vertical line. The straight line of unit slope is the expectation for additivity of d values. Reading from left to right, the first three points in each figure are from EDGAR (1958) and EDGAR (personal communication). The next 11 points are from DOERMANN and HILL (1953); they involve crosses within "linkage groups II and III." The remaining points are from DOERMANN and HILL (1953) and involve crosses of r_{48} with markers in "linkage groups II and III," crosses of tu_{41} with markers in "linkage group III," and crosses of tu_{42} with markers in "linkage group II." FIGURE 1.—Open circle model evaluated at $m_1 = 2.5$; $m_2 = 9.5$ ($\overline{m} = 6$). FIGURE 2.—Closed circle model evaluated at $m_1 = 2.5$; $m_2 = 7.5$ ($\overline{m} = 5$). FIGURE 3.—Open circle model evaluated at $m_1 = 2.5$; $m_2 = 12.5$ ($\overline{m} = 7.5$). FIGURE 4.—Closed circle model evaluated at $m_1 = 2.5$; $m_2 = 12.5$ ($\overline{m} = 7.5$).

(d) *Drift in recombination frequency:* The increase in recombination frequency from early to normal lysis (with close markers) observed by DOERMANN was $2x$. The models tested here predict drift factors ranging from $2x$ to $3x$ and appear to be within reasonable limits of uncertainty.

(e) *Total map length:* The total map lengths for each model can be determined by extrapolation to $d = 1$ of the approximate solution for $\bar{R}$ at small values of d. These solutions are for closed circles $\bar{R} \cong 0.45d(2\bar{m})$, and for open circles $\bar{R} \cong 0.45d(\bar{m} + 1 - e^{-\bar{m}})$; $\bar{R} \cong .45d(\bar{m} + 1)$. The values obtained range from 315 to 675.

These values are to be considered illustrative rather than definitive. In the companion paper, map-length estimates are presented which are more meaningful (because they acknowledge high negative interference) and more accurate (because they are based on more extensive data.)

(f) *Additivity test:* $\bar{R}$ values are converted to d values with the aid of the plots in Figures 1–4. These d values are tested for additivity in all possible pairwise combinations. The test cannot be applied to regions with large $\bar{R}$ values since in that range small errors in the estimate of $\bar{R}$ lead to large errors in the estimate of d. These errors result in a high degree of scatter in all the plots of $d_1 + d_2$ vs d (Figures 1–4). An examination of these plots makes it appear unlikely that any of the sets of assumptions can be rejected at present by application of this additivity test.

(g) *Circular additivity test:* The circularity of the T4 map permits an extension of the additivity test to large values of d. The $\bar{R}$ values for a set of adjacent regions circumscribing the map can be (graphically) converted to d values. The derived d values should sum to one if a mapping function is a good one. The genetic regions listed below have been used in this test. The $\bar{R}$ values are from DOERMANN and HILL (1953).

$tu_{41} - r_{47}$	$\bar{R} = 25.0$
$r_{47} - r_{48}$	$\bar{R} = 37.7$
$r_{48} - tu_{45}$	$\bar{R} = 32.9$
$tu_{45} - tu_{42}$	$\bar{R} = 32.6$
$tu_{42} - tu_{41}$	$\bar{R} = 36.9$

The summed d values range from 0.61 (closed circles at $\bar{m} = 7.5$) to 1.06 (open circles at $\bar{m} = 6$). In Table 1, the properties described above are summarized for the two functions.

SUMMARY

The population framework of VISCONTI and DELBRÜCK (1953) for recombination in bacteriophage can be adjusted to accommodate the recent finding of a circular linkage map in T4. Two models for the origin of the circularity were considered; neither can be ruled out by the published linkage data which we used to test the models.

TABLE 1

Comparison of the properties of the two models at the selected values for $\overline{m}$. *In all cases,* $m_1 = 2.5$

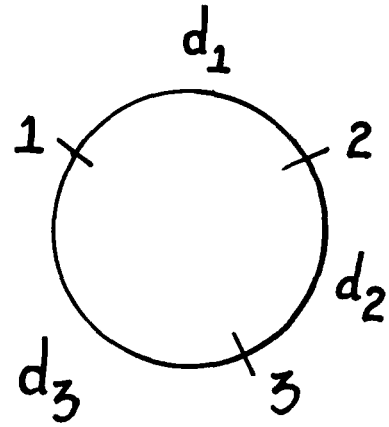
	Closed circle and rod		Open circle	
	$\overline{m}=5$	$\overline{m}=7.5$	$\overline{m}=6$	$\overline{m}=7.5$
$\overline{R}_{max}$	0.403	0.424	0.414	0.424
Rise in recombinant frequency for close markers	$2x$	$3x$	$2x$	$2.4x$
Σd in circular additivity test	0.93	0.61	1.06	0.90
Total map length	450	675	315	383
Total average number of "exchanges" per lineage	10	15	7	8.5

The authors are deeply indebted to DR. G. STREISINGER and MRS. H. FOSS; many of the ideas embodied here originated with them. DR. R. S. EDGAR tried to keep us honest.

APPENDIX: THE FORMALISM OF 3-FACTOR CROSSES INVOLVING CIRCULAR MAPS

The theory of VISCONTI and DELBRÜCK described the frequencies of the eight possible genotypes arising in a three-factor cross as a function of linkage distances and number of rounds of mating. That theory was generalized by STEINBERG and STAHL (1958) to which the reader wishing to use this appendix is referred.

Consider loci 1, 2, and 3 and distances on the chromosome of unit length of d_1, d_2, and d_3 as shown in the diagram. "Descent," "conversion," "mating" and all symbols are used exactly as defined by STEINBERG and STAHL (1958).



Closed circles and rods: If we impose the condition of pairwise mating with only two exchanges per mating, we can write the following relations between c's ("conversion probabilities") and the d values: $c_1=2d_1d_3$; $c_2=2d_1d_2$; $c_3=2d_2d_3$. For a Poisson distribution of mating acts among lineages, then,

$$P_{000} = e^{-m2(d_1d_3 + d_1d_2 + d_2d_3)}$$
$$P_{100} = (1-e^{-m2d_1d_3})\, e^{-m2(d_1d_2 + d_2d_3)}$$
$$P_{010} = (1-e^{-m2d_1d_2})\, e^{-m2(d_2d_3 + d_2d_3)}$$
$$P_{001} = (1-e^{-m2d_2d_3})\, e^{-m2(d_1d_3 + d_1d_2)}$$

and

$$b_{ijk} = a_{ijk}P_{000} + a_{i..}a_{.jk}P_{100} + a_{.j.}a_{i.k}P_{010} + a_{..k}a_{ij.}P_{001} + a_{i..}a_{.j.}a_{..k}(1-P_{000}-P_{100}-P_{010}-P_{001}).$$

Open circles: We consider the case of pairwise mating with one exchange per mating as defined earlier. Three classes of particles exist according to the sector of the circle in which the genome is open. The relations between the c's and d's for each class of particles must be written separately. For instance, for the fraction (d_1) of the particles open in sector 1: $c_1=d_3$; $c_2=d_2$; $c_3=0$. For a Poisson distribution of mating acts among lineages, then,

$$P_{000} = d_1e^{-m(d_2+d_3)} + d_2e^{-m(d_1+d_3)} + d_3e^{-m(d_1+d_2)}$$
$$P_{100} = d_1e^{-md_2}(1-e^{-md_3}) + d_3e^{-md_2}(1-e^{-md_1})$$
$$P_{010} = d_1e^{-md_3}(1-e^{-md_2}) + d_2e^{-md_3}(1-e^{-md_1})$$
$$P_{001} = d_2e^{-md_1}(1-e^{-md_3}) - d_3e^{-md_1}(1-e^{-md_2})$$

and, as usual,

$$b_{ijk} = a_{ijk}P_{000} + a_{i..}a_{.jk}P_{100} + a_{.j.}a_{i.k}P_{010} + a_{..k}a_{ij.}P_{001} + a_{i..}a_{.j.}a_{..k}(1-P_{000}-P_{100}-P_{010}-P_{001}).$$

LITERATURE CITED

CHASE, M., and A. H. DOERMANN, 1958 High negative interference over short segments of the genetic structure of bacteriophage T4. Genetics **43:** 332–353.

DOERMANN, A. H., 1953 The vegetative state in the life cycle of bacteriophage: evidence for its occurrence and its genetic characterization. Cold Spring Harbor Symp. Quant. Biol. **18:** 3–11.

DOERMANN, A. H., and M. HILL, 1953 Genetic structure of bacteriophage T4 as described by recombination studies of factors influencing plaque morphology. Genetics **38:** 79–90.

EDGAR, R. S., 1958 Mapping experiments with *rII* and *h* mutants of bacteriophage T4D. Virology **6:** 215–225.

EDGAR, R. S., and C. M. STEINBERG, 1958 On the origin of high negative interference over short segments of the genetic structure of bacteriophage T4. Virology **6:** 115–128.

HERSHEY, A. D., 1958 The production of recombinants in phage crosses. Cold Spring Harbor Sym. Quant. Biol. **23:** 19–46.

HERSHEY, A. D., and M. CHASE, 1951 Genetic recombination and heterozygosis in bacteriophage. Cold Spring Harbor Symp. Quant. Biol. **16:** 471–479.

LENNOX, E. S., C. LEVINTHAL, and F. SMITH, 1953 The effect of finite input in reducing recombinant frequency. Genetics **38:** 508–511.

STAHL, F. W., R. S. EDGAR, and J. STEINBERG, The linkage map of bacteriophage T4. Genetics **50:** 000–000.

STEINBERG, C., and F. STAHL, 1958 The theory of formal phage genetics. Cold Spring Harbor Symp. Quant. Biol. **23:** 42–46.

STREISINGER, G., R. S. EDGAR, and G. HARRAR DENHARDT, 1964 Chromosome structure in phage T4, I. Circularity of the linkage map. Proc. Natl. Acad. Sci. U.S. **51:** 775–779.

VISCONTI, N., and M. DELBRÜCK, 1953 The mechanism of genetic recombination in phage. Genetics **38:** 5–33.