A QUANTITATIVE MODEL FOR NONRANDOM GENERALIZED TRANSDUCTION, APPLIED TO THE PHAGE P22–*SALMONELLA TYPHIMURIUM* SYSTEM

W. MANDECKI,[1] K. KRAJEWSKA-GRYNKIEWICZ AND T. KLOPOTOWSKI

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, ul. Rakowiecka 36, 02-532 Warsaw, Poland

Manuscript received July 15, 1985
Revised copy accepted July 7, 1986

ABSTRACT

A mathematical model for nonrandom generalized transduction is proposed and analyzed. The model takes into account the finite number of transducing particle classes for any given marker. The equations for estimation of the distance between markers from cotransduction frequency data are derived and standard errors of the estimates are given. The obtained relationships depend significantly on the number of classes of transducing fragments. The model was applied to estimate the number of transducing fragment classes for a given marker in transduction with phage P22 of *Salmonella typhimurium*. It was found that the literature data on frequencies of cotransduction in crosses with mutual substitution of selective and nonselective markers can be rationalized most accurately by assuming that the mean number of classes is equal to 2. An improved method for analysis of cotransduction data is proposed on the basis of our model and the results of calculation. The method relies on solving a set of algebraic equations for cotransduction frequencies of markers located within one phage length. The method allows a relatively precise determination of distances between markers, positions of transducing particle ends and deletion or insertion lengths. The approach is applied to the *trp-cysB-pyrF* and *aroC-hisT-purF-dhuA* regions of the *Salmonella typhimurium* chromosome.

THE fine structure of the bacterial chromosome has been investigated by various methods including conjugation, transduction, transformation, deletion analysis and others. Rapid progress has been observed especially in the application of physicochemical and recombinant DNA methods, such as heteroduplex analysis, restriction nuclease mapping and sequencing of DNA. Nevertheless, to determine gene distances comparable to the length of phage DNA the method of cotransduction is still most often applied, as indicated in recent editions of linkage maps of *Salmonella typhimurium* (SANDERSON and ROTH 1983) and *Escherichia coli* (BACHMANN 1983). However, this approach meets some difficulties since the values of cotransduction frequency for given markers depend on the donor strain, the phage strain (CHELALA and MARGO-

[1] Present address: Corporate Molecular Biology, Abbott Laboratories, Abbott Park, Illinois 60064.

LIN 1974, 1976; SCHMIEGER and BACKHAUS 1976; KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979) and the selective marker used.

It was recognized relatively early that the chromosome is cut by the phage endonuclease system in nonrandom specific sites. The evidence for phage P22 and *Salmonella typhimurium* comes from works of OZEKI (1959), SMITH-KEARY and DAWSON (1963), ENOMOTO (1967) and others. This specificity leads to significant differences in cotransduction frequencies in crosses with mutual substitution of selective and nonselective markers (see Table 2) due to nonsymmetric position of the markers with respect to the ends of transducing particles. Because of nonrandom composition of transducing fragments, the additivity of distances determined on the basis of cotransduction frequency is not satisfactory. Moreover, the cotransduction frequency depends on the presence of deletions or insertions in the donor chromosome even at a distance from the markers (CHELALA and MARGOLIN 1974, 1976; KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979).

Therefore, it seemed worthwhile to perform a theoretical analysis in which the effect of nonrandom composition of transducing particles on cotransduction frequencies would have been included. We investigated how the number of different transducing particle classes influences the relationship between the cotransduction frequency and the distance between markers. The number of transducing particle classes that carry a given marker was estimated from published data on phage P22 transduction in *S. typhimurium*. We found that the data can be explained the best assuming the mean number of classes is 2.

The apparently small number of transducing particle classes for the phage P22–*S. typhimurium* system creates the possibility of a much more precise determination of distances between markers, positions of particle ends and relative numbers (weights) of the particles. In order to perform such an analysis, it is necessary to obtain a large set of linkage values between different markers in a chromosome region of length comparable to the phage DNA length. Such data have been available for the *trp-cysB-pyrF* region (SCHMIEGER and BACKHAUS 1976) and are described here for the *aroC-hisT-purF-dhuA* region of the *S. typhimurium* chromosome. The analysis has been applied to these regions.

The estimation of distances is routinely based on the relationships derived by KEMPER (1974) or by WU (1966). The results of the work presented in this paper show that the use of one relationship for mapping (KEMPER 1974) is unfounded from the mathematical point of view, and the application of the other (WU 1966) should be limited to systems with a large number of the transducing classes.

The generally accepted mechanism of transducing particle formation involves sequential cutting of the chromosome from preferred starting points (CHELALA and MARGOLIN 1974). This mechanism explains why a deletion or insertion in one point of the chromosome may change the linkage value for markers located even at relatively long distances. Therefore, the ability to define positions of transducing particle ends using the method presented here creates the possibility of estimating the lengths of chromosomal rearrangements (such as deletions, insertions, duplications).

TABLE 1

Bacterial strains

Strain	Genotype	Source
TK1000	Wild type	N. D. ZINDER
TK200	*hisCBHAFIE3501 purF145*	KRAJEWSKA-GRYNKIEWICZ AND KLOPOTOWSKI
TK221	HfRK5 *aroC5 hisT1529 hisCBHAFIE3501*	W. WALCZAK
TK514	*his1504 purF145*	This paper
TK515	*aroC6 hisT1504*	This paper
TK559	*pi-dhu-1 hisDCBHAFIE712*	KRAJEWSKA-GRYNKIEWICZ AND KLOPTOWSKI
TK560	*dhuA69 hisDCBHAFIE712*	KRAJEWSKA-GRYNKIEWICZ AND KLOPTOWSKI
TK566	*pi-dhu-1*	KRAJEWSKA-GRYNKIEWICZ AND KLOPTOWSKI
TK567	*dhuA69*	This paper
TK1558	*hisT1504*	B. N. AMES
TK2750	*dhuA69 hisCBHAFIE3501*	This paper
TK2751	*dhuA69 hisCBHAFIE3501*	This paper
TK2752	*hisCBHAFIE3501*	This paper
TK2760	*dhuA69 gnd*::Tn*10*	This paper
NK114	*gnd*::Tn*10 edd*	N. KLECKNER
TR140	*aroC5 purF145*	J. R. ROTH

MATERIALS AND METHODS

The genotypes and sources of strains used in characterization of the *aroC-hisT-purF-dhuA* region are shown in Table 1. All strains are derivatives of *S. typhimurium* strain LT2. Strains TK514 and TK515 were derived by transducing the *hisT1504* mutation from strain TK1558 into strain TR140 and selecting wrinkled recombinants *aroC*$^+$ and *purF*$^+$, respectively. Strains TK2750 and TK2751 were derived by transducing the *dhuA69* mutation from strain TK560 into strain TK200, selecting *dhu*$^-$ *purF*$^+$ and *dhu*$^-$ *purF*$^-$ recombinants, respectively, on minimal medium supplemented with D-histidine and adenine plus thiamine. Strain TK2752 was obtained as a *purF*$^+$ recombinant of TK200 as recipient and TK1000 as donor strain. Strain TK2760 was obtained as a result of transduction of the Tn*10* transposon from NK114 to the TK567 recipient and selection of a clone resistant to tetracycline.

Phage P22 L4 *int*$^-$ (SMITH and LEVINE 1967) was used in transductions.

Media, cultures and transduction procedures have been previously described (KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979).

Numerical calculations were performed on a CDC6000 computer (courtesy of the Computer Center, Polish Academy of Sciences). Computer programs were written in FORTRAN.

To convert the phage length units into the base pair numbers, we use the value of 41 kb (1 kb = 10^3 base pairs) for the length of phage P22 DNA (SUSSKIND and BOTSTEIN 1978).

RESULTS

The model

The model is an extension of models discussed by KEMPER (1974) and WU (1966) on a finite number of transducing particle classes. It is assumed in the

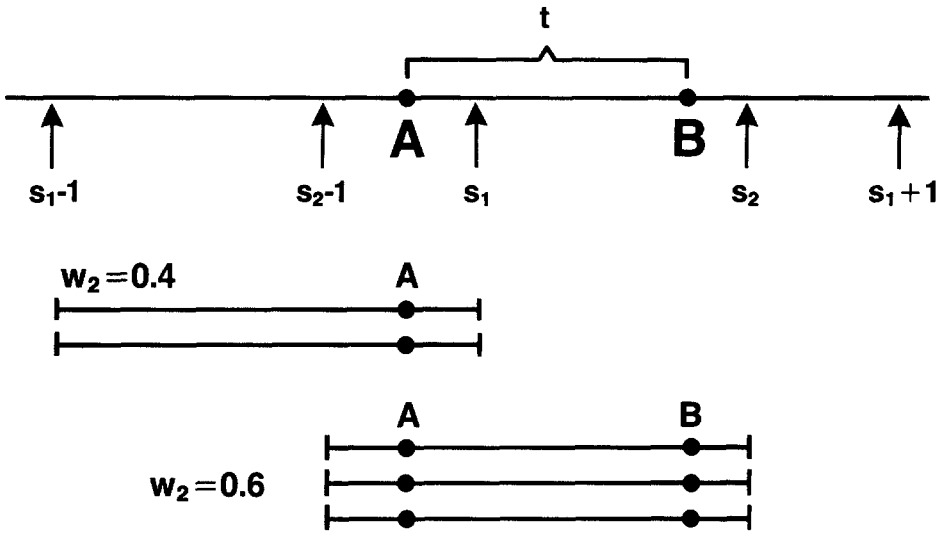

FIGURE 1.—Transducing particle classes. A hypothetical region of the bacterial chromosome carrying genetic markers *A* and *B* is shown. Marker *A* is carried on two classes of transducing particles ($N = 2$). t is the distance between the two markers. The different transducing particles are shown as segments of identical length, the ends of which are indicated by arrows. The particles are produced by sequential cuts of the chromosome proceeding from nuclease recognition sites located either to the left or to the right from the region shown. Neighbor particles resulting from the same sequence of cuts were omitted. The values of weights (w_i) are hypothetical.

model that endonucleolytic cleavages of chromosome are sequential and that the transducing particles generated are of equal length. The model takes into consideration the possibility that a given marker is carried by several classes of transducing particles. Different classes of particles are generated by different series of cleavages that originate at various chromosomal locations. To observe a nonzero cotransduction frequency of two markers, it is necessary and sufficient that both markers are carried by at least one class of transducing particles.

Let *A* and *B* be the point markers under consideration. The distance between the markers is denoted as t. The distance is expressed in units in which the length of the transducing particle is 1. The coordinates of markers *A* and *B* are 0 and t, respectively ($0 < t < 1$) (Figure 1).

Let N be the number of classes of transducing particles which contain marker *A*. If phage nuclease sites are given by coordinates $\cdots, s_1 - 1, s_2 - 1, \cdots; s_1, s_2, \cdots; s_1 + 1, s_2 + 1, \cdots$; and so on, then transducing particles carrrying marker *A* can be represented as intervals $(s_i, s_i + 1)$, where $0 < s_i < 1$, $i = 1, 2, \cdots, N$. Since the coordinate of marker *A* is 0, s_i can be considered also a distance between the cut site and marker *A*.

Different series of cuts may not be equivalent with respect to the frequency with which they occur in the cell; therefore, a term of weight of the transducing particle class is introduced. The weight w_i is defined as the ratio of the number of $(s_i, s_i + 1)$ transducing particles to the total number of the particles that carry a given marker. In an example in Figure 1, the two transducing

particle classes constitute 40 and 60% of total population of particles that carry marker A; therefore, their weights are 0.4 and 0.6, respectively.

In transduction, a part of the transducing fragment can be inherited. Let $P(A, B)$ be the probability that markers A and B will be coinherited. If A is the selected marker, the cotransduction frequency c_A is defined as the ratio of the mean number of events in which markers A and B are coinherited to the mean number of events in which marker A is inherited. Hence,

$$c_A(s_1, \cdots, s_N, t) = \frac{\sum_{i=1}^{N_{A,B}} w_i P_i(A, B)}{\sum_{i=1}^{N} w_i P_i(A)}, \tag{1}$$

where summation in the numerator is over transducing particle classes that contain both markers A and B. Summation in the denominator is over transducing particle classes that contain marker A; that is, over all N classes.

The above assumptions are confirmed by numerous experimental studies (for a review, see SUSSKIND and BOTSTEIN 1978). However, to establish a relation between the distance t and cotransduction frequency, further assumptions have to be adopted with respect to the explicit form of function $P(A, B)$, distribution of cut sites s_i, $i = 1, 2, \cdots, N$, and numerical values of weights w_i.

In previous quantitative analysis of cotransduction (OZEKI 1959; SMITH-KEARY and DAWSON 1963; WU 1966; KEMPER 1974), it was assumed that the probability of transfer of markers A and B from the transducing particle to the chromosome is proportional to the product of distances between markers and the ends of the particle. This corresponds to the model in which the transducing fragment first forms a duplex with the complementary region of the chromosome and then two crossovers occur at random sites. The above assumption is also applied in this analysis. If ends of the particle are s_i and $s_i + 1$, then

$$P_i(A, B) \sim (1 - s_i)(s_i - t). \tag{2}$$

It is unknown how the phage nuclease recognizes starting sites for series of cuts. It is assumed further that each position of cutting site s_i has equal probability of occurring in the cell. Finally, it is assumed for the purpose of evaluating the number of transducing particle classes that the weights of the cuts are the same, *i.e.*,

$$w_1 = w_2 = \cdots = w_N.$$

These assumptions and (1) lead to the following expression for c_A:

$$c_A(s_1, \cdots, s_N, t) = \frac{\sum_{i=1}^{N_{A,B}} (1 - s_i)(s_i - t)}{\sum_{i=1}^{N} s_i(1 - s_i)}. \tag{3}$$

One can also easily derive that

$$c_B(s_1, \cdots, s_N, t) = \frac{\sum_{i=1}^{N_{A,B}} (1 - s_i)(s_i - t)}{\sum_{i=1}^{N} |s_i - t|(1 - |s_i - t|)}. \tag{4}$$

Relationships between cotransducibility and distance

Let us consider the following problems:

1. What is the expected value (the mean) of cotransduction frequency when t, the distance between markers A and B, was determined by an independent method?
2. What is the expected value of distance between markers provided that c (*i.e.*, either c_A or c_B) is known? The experimentalist faces this problem most frequently.
3. What is the expected distance if both c_A and c_B are known?
4. What is the distribution of two-dimensional random variable (c_A, c_B)?
5. What are the random errors of estimates 1, 2 and 3?

The problems are solved below separately for different numbers of transducing particle classes. For two values of N, namely for $N = 1$ and $N = \infty$, an analytical solution can be easily obtained.

N = 1: From (3) and (4) one calculates that nonzero values of c_A and c_B are

$$c_A(s, t) = 1 - \frac{t}{s} \tag{5}$$

$$c_B(s, t) = \frac{1 - s}{1 - s - t} \tag{6}$$

If distance t is known, then, on basis of (5), the expected value of cotransduction frequency (the mean over s) is

$$\langle c \rangle = \frac{\int_t^1 \left(1 - \frac{t}{s}\right) \mathrm{ds}}{\int_0^1 \mathrm{ds}}$$

(symbol $\langle\ \rangle$ denotes the expected value). Hence

$$\langle c \rangle = 1 - t + t \ln t. \tag{7}$$

Thus, one arrives at the formula derived by KEMPER (1974). The plot of (7) is given in Figure 2. What has been calculated is the expected value of cotransduction frequency for $N = 1$. Expected value is used in the mathematical sense here, as the mean of a set of values. For any individual case with distance t between two markers, the actual c can lie between 0 and $1 - t$ [equation (5), $0 \leq t \leq s$], depending on the position of the ends of the transducing particle.

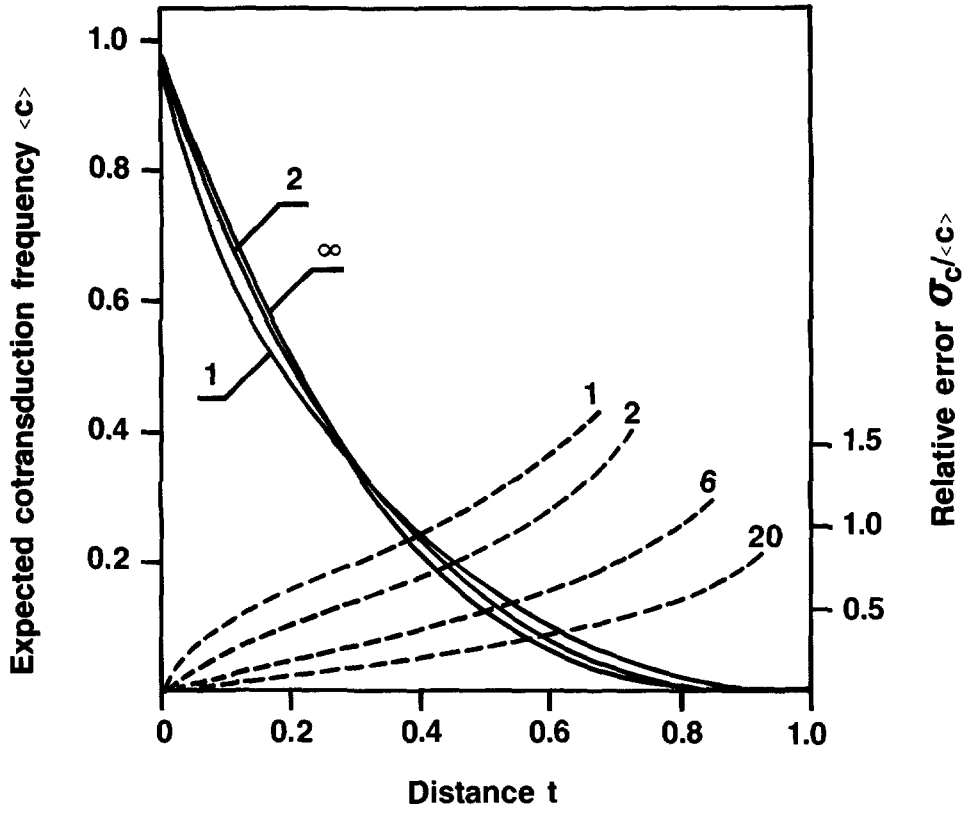

FIGURE 2.—Expected cotransduction frequency as a function of distance between markers. The numbers indicate the value of N for individual curves. Relative error of the estimate (broken line) is defined as the ratio between standard deviation and $\langle c \rangle$.

Equation (7) gives the average value of c for all possible classes of transducing fragments.

Thus, (7) is not a mapping function at all. It can only be used to calculate the expectations of the cotransduction frequency for any given distance t between the markers. It seems that the original derivation of (7) (KEMPER 1974) was in fact an unintentional calculation of $\langle c \rangle$ as a function of t, which was not recognized at the time, and then was improperly used for mapping.

To illustrate the derivation of the mapping formula, the dependence of the cotransduction frequency c on distance t and position of the cut site s [function (5)] was plotted in Figure 3. The mean value of t is calculated as the average t in the area limited by projections of coutour lines $c = c_0$ and $c = c_0 + \Delta c$ on plane $c = 0$ (Figure 2). Then we find the limit of $\langle t \rangle$ as c approaches zero. Thus, the mean t is calculated for all pairs (s, t) which give c such that

$$c_0 \leq c \leq c_0 + \Delta c.$$

If dA is an elementary area on plane $c = 0$, then

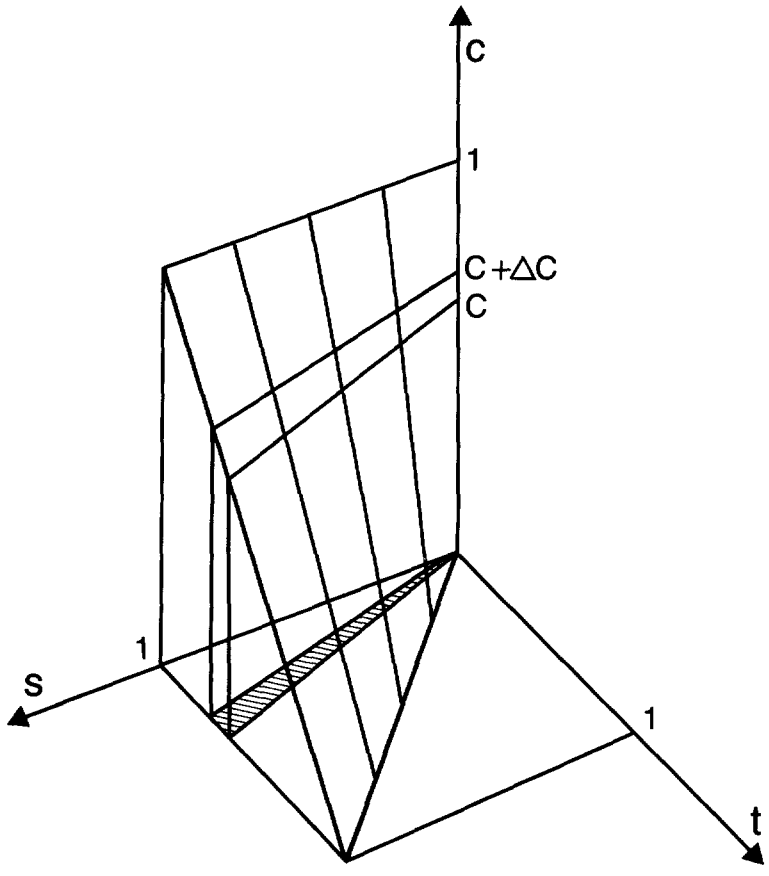

FIGURE 3.—Dependence of cotransduction frequency c on distance t between the markers and position of the cut site s ($N = 1$). Equation (5) was used to generate the plot. To obtain $\langle t \rangle$ as a function of c, the mean t is calculated over the shaded area.

$$\langle t \rangle = \frac{\int_A t dA}{\int_A dA},$$

where the integration is over the shaded area in Figure 3. By substituting t from (5), since $dA = s\Delta c$ ds, we obtain

$$\langle t \rangle = \frac{\int_0^1 (1 - c)s^2 \Delta c \text{ ds}}{\int_0^1 s\Delta c \text{ ds}}.$$

Hence

$$\langle t \rangle = \tfrac{2}{3}(1 - c). \tag{8}$$

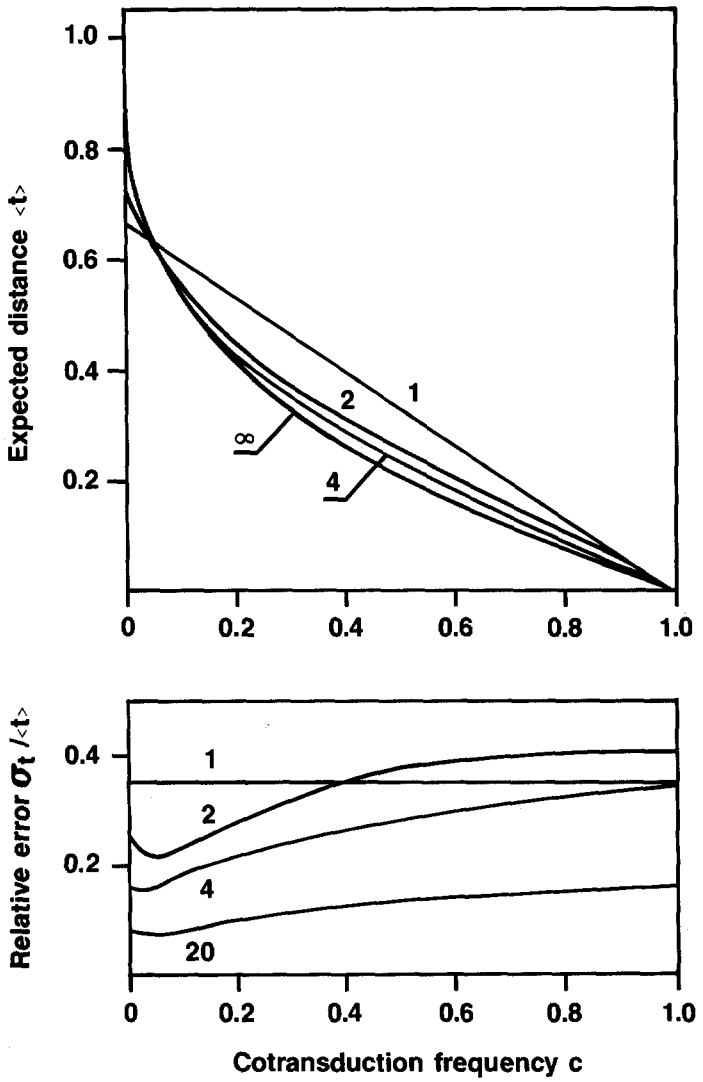

FIGURE 4.—Expected distance and the relative error of the estimate as a function of cotransduction frequency. The numbers indicate the value of N for individual curves.

It is not surprising that (8) (Figure 4) is not the inverse of (7). Obviously, when $t = 1$ then always $c = 0$, but if $c = 0$ then t need not to be 1, since zero cotransduction frequency is obtained also when the cut is situated between the markers.

Standard deviation of estimate (8) is

$$\sigma_t^2 = \frac{\int_A (t - \langle t \rangle)^2 \, dA}{\int_A dA}.$$

After substituting t and $\langle t \rangle$ from (5) and (8), we obtain

$$\sigma_t^2 = 1/18(1 - c)^2,$$

and the relative error of the distance estimation,

$$\frac{\sigma_t}{\langle t \rangle} = \frac{1}{2\sqrt{2}} \approx 0.35,$$

is constant (only for $N = 1$) and does not depend on experimentally obtained c.

When both c_A and c_B are known, the direct solution of (5) and (6) permits determining the positions of the transducing particle ends and the exact distance between markers

$$\begin{aligned} s &= \frac{1 - c_B}{1 - c_A c_B} \\ t &= \frac{(1 - c_A)(1 - c_B)}{1 - c_A c_B}. \end{aligned} \tag{9}$$

Cotransduction frequencies c_A and c_B constitute a two-dimensional random variable $(s, t) \rightarrow (c_A, c_B)$. In further analysis, the knowledge of the distribution of this variable will be needed. To make the results directly applicable in this paper, we define the set of arguments of the variable as restricted to elements that are given nonzero cotransduction frequencies ($t < s$). After simple but arduous calculations, one can derive the equation for the probability-density function

$$f(c_A, c_B) = \frac{2(1 - c_A)(1 - c_B)}{(1 - c_A c_B)^3}. \tag{10}$$

The probability that we obtain cotransduction frequency from a given neighborhood of c_A and c_B is proportional to $f(c_A, c_B)$. From the definition of the probability-density function

$$\int_0^1 \int_0^1 f(c_A, c_B)\, dc_A dc_B = 1$$

N = ∞: There is a one-to-one relationship between t and c. To calculate this dependence, the sums in (3) should be replaced by integrals. Hence,

$$c = \frac{\int_t^1 (1 - s)(s - t)\mathrm{ds}}{\int_0^1 s(1 - s)\mathrm{ds}}$$

and

$$c = (1 - t)^3. \tag{11}$$

The above formula was originally developed by Wu (1966). Of course, in this

case ($N = \infty$) c_A is always equal to c_B. Whether this formula should be used for mapping is debatable. Its use is sound if there is a large number of transducing fragment classes for any given marker, but it will lead to significant errors (comparable to the standard deviation) for $N = 1$, 2 or 3 (APPENDIX 2). The function is graphically represented in Figure 4.

$2 < N < \infty$: The Monte Carlo computer simulation (HAMMERSLEY and HANDSCOMB 1964) was applied. The expected c for a given t was determined as follows. The pseudo-random numbers $s_1, s_2, \cdots, s_N$ uniformly distributed in interval (0, 1) were generated. For each sequence $s_1, s_2, \cdots, s_N$ and a given t, the cotransduction frequency c_A was computed according to (3). Number of repeats was 1000. Then, mean c_A and standard deviation were computed. The value of t was changed over the interval (0, 1), and the step was 0.01. The results are presented in Figure 2.

The results indicate that for all $N > 2$, good approximation (but only for expected cotransduction frequency) is given by WU's formula (11). Cotransduction frequency determined from this equation does not differ from the estimates obtained here by more than 10% of the standard deviation. Random errors of the estimation are approximately proportional to $1/\sqrt{N}$.

An experimentalist is usually much more interested in the estimate of distance between markers. To obtain the estimate, the following procedure was used. In elementary run the values $s_1, s_2, \cdots, t$, c_A and c_B could be established (see above). Altogether, there were 10^5 such sequences obtained. These sequences were grouped into a total of 28 classes with respect to the numerical value of c_A, ten classes for $c_A < 0.1$ and 18 classes for $0.1 \leq c_A \leq 1$, as given in APPENDIX 3. Within each class of sequences the mean t and standard deviation were computed. They are the estimates of the distance between markers and of the random errors, respectively.

Similarly, the expected value of t for known values of c_A and c_B could be determined by introducing classes with respect to the values of both c_A and c_B. To obtain the distribution of the probability-density function $f(c_A, c_B)$, the numbers of sequences in each class were normalized by dividing them by the total number of sequences (5000) and by the square of the linear dimension of the class (10^{-4} if $c_A, c_B < 0.1$; 0.025 if c_A or $c_B \geq 0.1$). The distribution is given in APPENDIX 3 for $N = 2$.

The results led to the following essential conclusions:

1. Dependence of the expected distance on cotransduction frequency measured is different for various N. The computed functions are displayed in Figure 4 and are tabulated in APPENDIX 2. It can be derived analytically that $\langle t \rangle = (N + 1)/(N + 2)$ when $c = 0$.

2. Relative error of the estimation, defined as the ratio of standard deviation and the mean, is the greater the greater is c (for $c \geq 0.1$). Only for $N = 1$ the relative error does not depend on c. For a given c, the error is roughly proportional to $1/\sqrt{N}$ (Figure 4).

3. When c_A and c_B are known, the accuracy of the estimation can be increased. For $N = 2$ the error may be reduced by about a one-half, as evident

from the comparison of numerical data in APPENDICES 2 and 3. For $N = 1$ the exact value of distance t can be determined.

Estimation of the number of transducing particle classes for phage P22–*Salmonella typhimurium* system

Usefulness of the conclusions presented above depends on the possibility of evaluating the number of transducing particle classes (N) for the chromosome region where genes A and B are located. Except for perhaps one region (OZEKI 1959), the number of classes has not been determined. However, there is a possibility of estimation of average N for the bacterial chromosome and the given phage, assuming that N is constant in various regions of the chromosome.

The data of the previous section—namely, the distribution of random variable (c_A, c_B)—show that the greater is N the less difference between c_A and c_B should be observed. Therefore, we can infer the N value from differences between c_A and c_B for various pairs of markers.

The reasoning is based on the method of maximum likelihood (*e.g.*, see POLLARD 1977). Let there be m independent observations [here: $(c_A, c_B)_i$, $i = 1, 2, \cdots, m$]. Let us formulate the hypothesis that the value of the estimated parameter is N ($N = 1, 2, \cdots$). For each hypothetical N we calculate the values of probability-density function $f(c_A, c_B)$. To find out which N is most likely, we define the likelihood function as the product of $f(c_A, c_B)$, *i.e.*,

$$L(N) = \prod_{i=1}^{N} f_i(c_A, c_B).$$

The value of N for which $L(N)$ assumes maximum (*i.e.*, for which the occurrence of the observations mentioned is the most likely) constitutes the required estimate.

Calculations were made for the phage P22–*Salmonella typhimurium* system. It was attempted to collect a possibly large set of published data on cotransduction frequencies in crosses with mutual substitution of selective and nonselective markers. The data for 17 crosses are listed in Table 2. The values of $f(c_A, c_B)$ were assigned to each cross and a given N ($N = 1, 2, 3, 4$) on basis of the computation results. The value of function L was subsequently calculated.

The results (Table 2) show that the maximum of L is assumed for $N = 2$, and that the values of L for other N are severalfold smaller. Therefore, it is concluded that the mean number of transducing particle classes is equal to 2.

Two main factors influence the accuracy of this estimate. The first is the assumption of the same number of classes (N) for various regions of the chromosome, which may not be fulfilled. Therefore, the result $N = 2$ does not exclude an alternative that for some regions the number of classes is 2, but for some others it is 3 or 1. Different cuts might also have different weights.

The second factor is that cotransduction data used in this analysis (Table 2) constitute a small sample with respect to distinguishing between two-dimensional distributions. However, this type of error seems to be less important since numerical values of $L(N)$ were severalfold smaller for N other than 2 as compared to the maximum value.

TABLE 2

Estimation of the mean number of transducing particle classes for the phage P22–*S. typhimurium* system

Markers		Frequency of cotransduction (%)			Value of the probability-density function for given N:			
A	*B*	c_A	c_B	Source	1	2	3	4
pyrA	*ilvS*	4.8	9.1	BLATT and UMBARGER (1972)	1.75	8.7	10.8	9.2
pyrA	*fol*	6.8	20	KEMPER (1974)	1.55	2.0	1.62	1.32
leuA	*fol*	0.7	4.9	KEMPER (1974)	1.89	4.5	2.3	1.42
gal	*nadA*	20	27	LANGLEY and GUEST (1974)	1.38	2.8	3.0	3.3
leuA	*araB*	37	44	SCHMIEGER and BACKHAUS (1976)	1.20	1.61	1.68	1.87
trpA	*cysB*	29	46	SCHMIEGER and BACKHAUS (1976)	1.18	1.17	1.25	1.15
trpA	*pyrF*	4.6	18	SCHMIEGER and BACKHAUS (1976)	1.60	2.0	1.38	0.93
cysB	*pyrF*	28	61	SCHMIEGER and BACKHAUS (1976)	0.98	0.57	0.34	0.17
hemA	*dadA*	2.5	3.7	WILD *et al.* (1974)	1.88	40	63	70
dhuA	*purF*	5	50	KRAJEWSKA-GRYNKIEWICZ, WALCZAK and KLOPOTOWSKI (1971)	1.02	0.13	0.02	<0.003
aroC	*purF*	6.8	14	KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI (1979)	1.65	5.1	5.4	5.1
guaA	*purI*	24	51	OZEKI (1959)	1.10	0.81	0.56	0.34
cysE	*pyrE*	1.2	2.6	SANDERSON and SAEED (1972)	1.92	39	56	60
argA	*thyA*	10	38	SANDERSON and SAEED (1972)	1.25	0.69	0.31	0.15
lys	*thy*	18	37	EISENSTARK, EISENSTARK and CUNNINGHAN (1968)	1.27	1.23	1.06	0.87
lys	*argA*	0.1	0.1	EISENSTARK, EISENSTARK and CUNNINGHAN (1968)	2.00	150	280	380
thr	*serB*	17	35	STUTTARD (1972)	1.30	1.35	1.21	0.92
				$L(N) \times 10^5 =$	0.004	680	28	<0.2

Determination of chromosomal positions of markers, transducing particle ends and lengths of deletions or insertions

***trp-cysB-pyrF region*:** SCHMIEGER and BACKHAUS (1976) measured cotransduction frequencies of *trpA*, *cysB* and *pyrF* markers and obtained a set of six values (Table 3). They also determined the respective cotransduction frequencies for the phage P22 mutant with increased transduction ability (HT mutant). Significant differences were observed between the two sets of cotransduction data. They were accounted for on the basis of different sites of cuts on host DNA. The measurements made it possible to apply the cotransduction model to calculate distances between the markers and to determine positions of transducing particle ends and particle weights.

Let us formulate the problem in mathematical terms. There are eight un-

TABLE 3

Cotransduction frequencies and their computed counterparts for the *trp-cysB-pyrF* region

Markers		*Cotransduction frequency (%)*			
		Wild-type phage (H5)		HT mutant phage (HT12/4)	
A	*B*	c_A	c_B	c_A	c_B
trpA	*cysB*	29.0	45.6	40.6	27.8
		28.6	*42.0*	*42.5*	*31.4*
cysB	*pyrF*	27.8	61.4	53.7	70.2
		26.1	*61.2*	*57.3*	*72.0*
trpA	*pyrF*	4.6	18.3	25.9	25.7
		5.3	*18.2*	*24.3*	*22.6*

Experimental data are taken from SCHMIEGER and BACKHAUS (1976). The donor strain was SU687. Theoretical values (italicized) were computed according to equations (1) and (2). c_A and c_B, index *A* or *B* indicates the selected marker.

known quantities; namely, *trpA-cysB* and *cysB-pyrF* distances, positions of transducing particle ends for the wild-type phage (one variable for each of the two particle classes) and for the HT mutant (also two variables) and particle weights (one variable for each phage). On the other hand, the region is described by 12 experimentally determined cotransduction frequencies. Each of them generates one equation. Thus, we get a system of 12 equations with eight unknown quantities similar to system (A3)–(A8) (APPENDIX 1). The system was solved numerically.

The following results were obtained:

1. The *trpA-cysB* and *cysB-pyrF* distances are equal to 17.1 and 6.8 kb, respectively.

2. The solution for the positions of ends and weights of transducing particle classes for the wild-type phage is $s_1 = 25$ kb, $s_2 = 12$ kb, $w_1 = 0.83$, $w_2 = 0.17$.

3. Parameters for the HT phage transducing particle classes are $s_1 = 33$ kb, $s_2 = 20$ kb, $w_1 = 0.87$, $w_2 = 0.13$. It is evident that the positions of the ends of the HT phage transducing particles can be obtained by shifting the positions of the ends (as determined for the wild-type phage) by 8 kb, whereas the weights are roughly the same for both solutions.

The above data are graphically presented in Figure 5.

When the distances between markers, positions of transducing particle ends and weights had been determined, it was possible to simulate the process of cotransduction and compute the theoretical cotransduction frequencies for the system according to (1) and (2). The computed values (Table 2) were in good agreement with experimental data. The difference between theoretical and experimental cotransduction frequencies were comparable to the mean error of the measurement (about 5% of the value measured; SCHMIEGER and BACKHAUS 1976). Despite the consistency between the experimental results and the

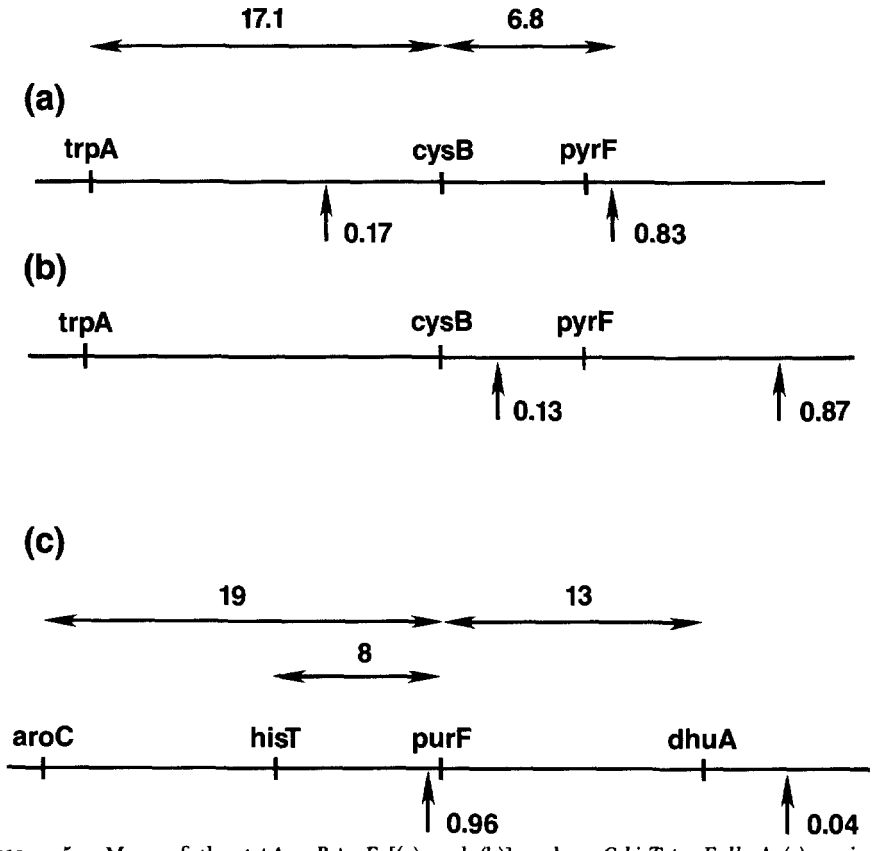

FIGURE 5.—Maps of the *trpA-cysB-pyrF* [(a) and (b)] and *aroC-hisT-purF-dhuA* (c) regions as computed from the model. Distances are shown in 10^3 base pair units (kb). Positions of transducing particle ends are indicated by arrows. Numbers next to the arrows are the weights of the transducing particle classes. Solutions of equations for phage H5 [(a) and (c)]; phage HT12/4 (b).

theoretical calculations, it is feasible that the changes in cotransduction frequencies are due to an increased number of transducing particle classes in transductions with the HT phage, in addition to changes in the positions of cut sites and weights.

***aroC-hisT-purF-dhuA* region:** To obtain a large set of data for the mathematical analysis, all possible linkages between the *aroC*, *purF* and *dhuA* markers were measured. As donors, two isogenic strains were used, differing only by insertion of the Tn*10* transposon into the *gnd* gene. Thus, 12 cotransduction frequencies (six for each donor) were obtained. They are collected in Table 4. It was found that Tn*10* insertion produced significant, severalfold changes of *aroC-purF*, *purF-aroC* and *purF-dhuA* linkages.

It is assumed further in that the Tn*10* insertion is between the phage nuclease start sites (which apparently are infrequent on the chromosome; see estimation in DISCUSSION) and the region being considered. Therefore, the insertion results in a shift of the cut sites in this region by a distance equal to the length of Tn*10*, *i.e.*, 9.2 kb (KLECKNER, ROTH and BOTSTEIN 1977). The

TABLE 4

Cotransdsuction frequencies and their computed counterparts for *aroC*, *purF* and *dhuA* markers

Recipient strain	Markers		Cotransduction frequency (%) for donor strains			
			TK567 (no insertion)		TK2760 (Tn*10* insertion)	
	A	*B*	c_A	c_B	c_A	c_B
TR140	*aroC*	*purF*	5.0 ± 0.6	10.8 ± 0.7	27.7 ± 0.7	38.4 ± 1.0
			0.9	*9.6*	*31.4*	*39.4*
TK200	*purF*	*dhuA*	52.2 ± 1.0	4.3 ± 0.6	22.2 ± 1.0	2.2 ± 0.3
			48.0	*5.2*	*1.5*	*2.8*
TK221	*aroC*	*dhuA*	0.5 ± 0.2	0.1 ± 0.1	1.7 ± 0.4	0.3 ± 0.1
			0.2	*0.3*	*0.3*	*0.6*

Computed cotransduction frequencies are italicized. c_A and c_B, index *A* or *B* indicates the selected marker.

insertion should not change the weights, since the likelihood of a nuclease start site on a 9.2-kb-long fragment is negligible.

The system of 12 equations was formulated for cotransduction frequencies of the donor with no insertions and the donor with Tn*10* (*gnd*) introduced (each value of the cotransduction frequency generates one equation). There are five unknown quantities in this system; namely, *aroC-purF* and *aroC-dhuA* distances, positions of transducing particle ends in the wild-type donor and weight. The following results were obtained by solving the system of equations using the approach outlined in APPENDIX 1:

1. The *aroC-purF* and *purF-dhuA* distances are equal to 19 and 13 kb, respectively.

2. Positions of the transducing particle ends and weights are s_1 = 18.5 kb, w_1 = 0.96 and s_2 = 36 kb, w_2 = 0.04 (Figure 5). The computed distance between particle end and *purF* is 0.5 kb for the more numerous particle class.

Based on the results presented above, the theoretical values of linkages were computed according to equations (1) and (2). The computed linkages (Table 4) generally fit well to experimental ones, except for *purF-dhuA* linkage in the strain with Tn*10* insertion. Nevertheless, the fit is worse than that obtained for the *trp-cysB-pyrF* region. Therefore, somewhat larger errors than before are expected for the distances between markers, positions of the transducing particle ends and weights. The relative error is estimated to be about ±15% of a given distance. Thus, *aroC-purF* distance is 19 ± 3 kb and *purF-dhuA* distance is 13 ± 2 kb.

Effect of distant deletions and insertions on cotransduction linkages

KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI (1979) observed strong changes of *purF-aroC* and *aroC-purF* linkages caused by *his* operon deletions or by insertions located between the *his* operon and *dhuA* gene. The authors interpreted their results as being due to alterations in composition of transducing particle classes in the *aroC-purF* region. Knowledge of the characteris-

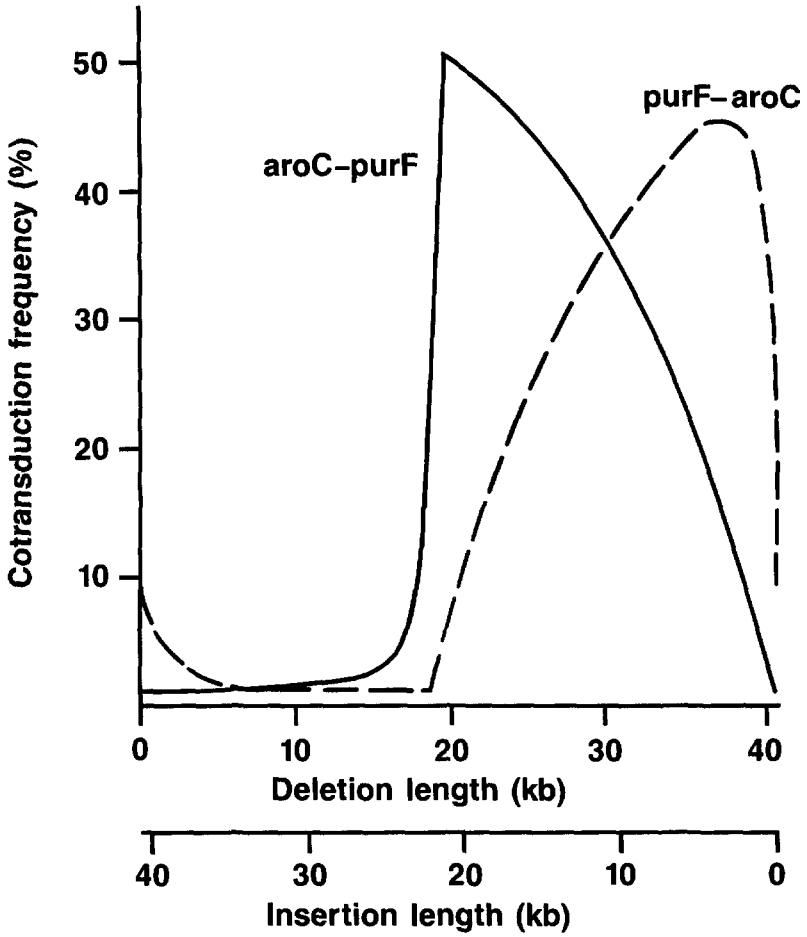

FIGURE 6.—Simulation of effects of deletions or insertions on the *aroC-purF* cotransduction frequency. Ends of two transducing particle classes are assumed to be shifted due to deletions or insertions. Point 0 on the abscissa (no deletion or insertion) corresponds to the background of TK567 strain.

tics of the region (see previous section) allowed us to simulate changes of contransduction frequency caused by a deletion or, equivalently, a shift of the transducing particle ends. The curves for *aroC-purF* and *purF-aroC* linkages computed from (1) and (2) are presented in Figure 6. Point 0 on the *x*-axis represents the donor strain with no chromosomal rearrangement (Table 4). If the deletion or insertion length is known, one can determine the expected cotransduction frequencies from results given in Figure 6, and vice versa, if cotransduction frequencies are known, the deletion or insertion length can be estimated. We can test the method using an insertion of known length, *F* factor HfrK5 (94.5 kb; OHTSUBO and OHTSUBO 1977), which increases the *purF-aroC* linkage to 39.4% (KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979). Distance of 92.5 kb is equivalent in this approach to 12.5 kb, since the difference

between them is equal to the multiple phage DNA length. The expected value as read from Figure 6 for an insertion length of 12.5 kb is close to the experimental value and is equal to 33%.

Let us consider now the effect of deletions on the *purF-aroC* and *aroC-purF* linkages. The conclusions drawn from theoretical considerations (Figure 6) are compared with the experimental results (KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979) and are presented below.

Prediction	Experiment
Deletions up to 19 kb decrease *purF-aroC* linkage. Deletions 5- to 19-kb long given similar and low linkage values.	Nine of eleven deletions tested decrease *purF-aroC* linkage. Exceptions are *his-57* (see below) and *his-645*. Long deletions (more than six genes deleted) give values of cotransduction frequencies of 4–7%.
Deletions 20- to 40-kb long strongly increase (up to 45%) *purF-aroC* linkage	*his-57* deletion gives *purF-aroC* linkage of 32.4%.
purF-aroC cotransducibility is greater than *aroC-purF* linkage value for small deletions. Proportions become reverse as deletion length increases.	For two small deletions (*his-129*, *his-153*) and two longer ones (*his-203*, *his-712*) *purF-aroC* linkage is greater than *aroC-purF*. For two long deletions (*his-640*, *his-538*) the reverse is true. The data are available for seven deletions. Exceptional behavior of *his-645* is noted.

The agreement of the predictions based on the model with the experimental evidence confirms the usefulness of the model and justifies an attempt to estimate the length of some *his* operon deletions. By comparing the relative changes of linkage values caused by deletions (KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979) with the theoretical predictions (Figure 6) we get lengths of less than 10 kb for *his-203* and *his-712* deletions, 10–20 kb for *his-538* and *his-640* deletions and 30–40 kb for *his-57* deletion.

The *aroC-purF* and *purF-aroC* linkages were also measured (K. KRAJEWSKA-GRYNKIEWICZ and T. KLOPOTOWSKI, unpublished results) for a donor strain in which the *pi-dhu-1* duplication is introduced. The values were 32.6% ± 1.5% and 26.6% ± 1.0%, respectively, for TK566 × TR140 cross. On the basis of the model (Figure 6), the length of 12 ± 4 kb was assigned to this insertion.

DISCUSSION

The model we propose in this paper for the generalized transduction allows one to determine how the relationship between cotransduction frequency and distance depends on the degree of randomness of the relevant transducing

particle population. The degree of randomness is expressed in our calculations as number of cuts, N, per unit chromosome length corresponding to the length of one transducing particle. On the basis of our deductions and published data we estimated the average value of N for the phage P22–*Salmonella typhimurium* transduction system as being equal to 2. Because series of cuts during bacterial chromosome fragmentation by phage-coded nuclease system reach regions that are distant by 5–10% of the chromosome length from a nuclease starting site (KRAJEWSKA-GRYNKIEWICZ and KLOPOTOWSKI 1979) and transducing particles comprise 1% of the chromosome, the number of P22 nuclease starting sites should be about 20–40 per chromosome of *S. typhimurium*.

The model is based on the generally accepted mechanism of transducing particle formation (CHELALA and MARGOLIN 1974; SUSSKIND and BOTSTEIN 1978) and the assumption that the probability of recombinational transfer of a pair of markers from a transducing particle to the chromosome is proportional to the product of distances between the markers and the respective ends of the transducing particle [relationship (2)]. The same assumption was made by WU (1966) and KEMPER (1974). Adequacy of the assumption is indicated by a degree of additivity of distances calculated from cotransduction frequencies on the basis of their equations (BACHMANN 1983; SANDERSON and ROTH 1983).

Equations (1) and (2) show that cotransduction frequency depends not only on the degree of randomness of transducing fragment population but also on the weights of its constituents. The term "weight" represents the fraction of a transducing particle class among all those that carry a given marker. There is a certain probability for each class that the marker will be recombined into the recipient chromosome. Therefore, the frequency of cotransduction is influenced by the transducing fragment composition. Numerical values of weights are usually unknown. However, it was shown that the transduction analysis of certain chromosome regions can be programmed to provide data for calculating weights of characterized classes of transducing fragments. It should also be noted that ENOMOTO (1967) experimentally determined proportions of transducing fragments, the ends of which are located between two given markers, and introduced a notion close to the definition of weight.

The results obtained in this study allow one to recommend a more precise procedure for estimating the distances between genetic markers from P22 transduction data in *S. typhimurium*. When the cotransduction frequency of a given pair of markers can be determined by using only one of them for selection, the expected distance should be read from APPENDIX 2 under $N = 2$. The function is also displayed in Figure 4. For example, when $c = 0.6$, one reads that $\langle t \rangle = 0.21 \pm 0.08$. The value of t equal to 1 corresponds to a distance that is about the length of a P22 transducing particle or phage P22 DNA (SUSSKIND and BOTSTEIN 1978). If one assumes that the length is 41 kb, or approximately 1% of *S. typhimurium* chromosome, one distance unit is equal to one map unit in this organism (SANDERSON and ROTH 1983).

More precise estimation of map distance is feasible when cotransduction frequencies of a given pair of markers can be determined by using each of

them for selecting recombinants. In this case the expected distance should be read from APPENDIX 3. For example, if $c_A = 0.5$ and $c_B = 0.7$, the expected distance between *A* and *B* taken from Table 3 is $\langle t \rangle = 0.19 \pm 0.04$.

The comparison of distance estimates obtained from WU's equation ($N = \infty$) and from the proposed method shows that the former overestimates the distance for small $c (c < 0.03)$. The difference is up to 30% of the mean for $N = 2$. For $c > 0.03$, WU's method gives the distances too short by up to 30% of the computed mean for $N = 2$. If KEMPER's equation is used, the discrepancies are even greater.

Further increase of accuracy of distance estimation is possible provided that more cotransduction data are available. The presented method of analysis was applied to two regions of the *S. typhimurium* chromosome (*trp-cysB-pyrF* and *aroC-hisT-purF-dhuA* regions). We used data from transductions mediated with wild-type P22 phages and an HT mutant, and with donor strains harboring different deletions or insertions at distant sites. Determination of the characteristics of these regions allowed another application of the method. The lengths of two *trp* and five *his* deletions and one insertion were estimated. It was also possible to define quantitatively the total deletion or insertion length between the nuclease starting site and *trp* operon in one strain relative to the other strain. Such chromosomal changes can be silent and may not noticeably change the phenotype of the strains.

Perhaps one generalization is noteworthy here. It was found that, for all phages and regions considered, computed weights for two transducing particle classes were strongly differentiated. The weights of more numerous classes were fourfold to 25-fold greater than that of the other class. The results of ENOMOTO (1967) on composition of transducing fragments in *flaAII-H1-flaK* region located on unit 41 of *S. typhimurium* map (SANDERSON and ROTH 1983)—*i.e.* close to the *trp* operon (unit 34) and *purF* gene (unit 49)—show that 94% of transducing particles active in transduction had ends located between *H1* and *flaK* genes, and only 6% between *FlaAII* and *H1* genes. These results indicate again a strong differentiation in weights. OZEKI (1959) suggested that only one particle class is active in transduction of *purI* or *guaA* markers (unit 54 of the linkage map). This may be an exceptional situation, because usually more than one particle class is active in transduction of a given marker (PEARCE and STOCKER 1965; ROTH and HARTMAN 1965; KEMPER 1974), or again, a strong differentiation of weights did not allow the detection of more than one transducing particle class. The above data support the idea presented and applied in this paper of taking into account two transducing particle classes. Even if the third class existed, its weight should be small (perhaps of order 10^{-2}), and the class would not influence the cotransduction frequencies observed.

It has to be noted that the efficiency of the transduction of markers by P22, which can vary by three orders of magnitude depending on the region of the chromosome, does not play a role in the model. This is because the cotransduction frequency depends in the model on the relative composition of transducing particles (weights) and not on their number. Other factors, however,

such as the presence of hot spots for recombination in relevant regions and chi sites can certainly affect the accuracy of predictions.

The present approach can be helpful in the quantitative characterization of phage mutant strains with increased or decreased transduction ability (HT and LT mutants, respectively). An interesting feature of some HT mutants is that they give different cotransduction linkage for a given pair of markers (CHELALA and MARGOLIN 1976; SCHMIEGER and BACKHAUS 1976) due to altered recognition of starting sites by the phage nuclease system. One HT mutant (HT12/4) was analyzed in this paper. The cotransduction data were available for the *trp-cysB-purF* region. The data allowed us to conclude that the HT12/4 starting site producing the most numerous particle class active in transduction of the *cysB* marker is different from the wild-type phage site. A decrease of recognition specificity (*i.e.*, equalization of weights of transducing particle classes) was not noted for the mutant phage.

Although this paper deals with the phage P22–*S. typhimurium* system, the approach is valid for other generalized transduction phages with affinity to other bacterial species, and, in particular, phage P1 of *E. coli*.

We thank J. L. ROSNER for valuable comments during preparation of this manuscript. This work was supported by the Polish Academy of Sciences within Project 09.7.

LITERATURE CITED

BACHMANN, B. J., 1983 Linkage map of *Escherichia coli* K12, Ed. 7. Bacteriol. Rev. **47:** 180–230.

BLATT, J. M. and H. E. UMBARGER, 1972 On the role of isoleucyl-tRNA synthetase in multivalent repression. Biochem. Genet. **6:** 99–118.

CHELALA, C. A. and P. MARGOLIN, 1974 Effects of deletions on cotransduction linkage in *Salmonella typhimurium*: evidence that bacterial chromosome deletions affect the formation of transducing DNA fragments. Mol. Gen. Genet. **131:** 97–112.

CHELALA, C. A. and P. MARGOLIN, 1976 Evidence that HT mutant strains of bacteriophage P22 retain an altered form of substrate specificity in the formation of transducing particles in *Salmonella typhimurium*. Genet. Res. **27:** 315–322.

EISENSTARK, A., R. EISENSTARK and S. CUNNINGHAM, 1968 Genetic analysis of thymineless (*thy*$^-$) mutants in *Salmonella typhimurium*. Genetics **58:** 493–506.

ENOMOTO, M., 1967 Composition of chromosome fragments participating in phage P22-mediated transduction of *Salmonella typhimurium*. Virology **33:** 474–482.

HAMMERSLEY, J. M. and D. C. HANDSCOMB, 1964 *Monte Carlo Methods*. Menthuen, London.

KEMPER, J., 1974 Gene order and co-transduction in the *leu-ara-fol-pyrA* region of the *Salmonella typhimurium* linkage map. J. Bacteriol. **117:** 94–99.

KLECKNER, N., J. ROTH and D. BOTSTEIN, 1977 Genetic engineering *in vivo* using translocatable drug resistance elements. New methods in bacterial genetics. J. Mol. Biol **117:** 125–159.

KRAJEWSKA-GRYNKIEWICZ, K. and T. KLOPOTOWSKI, 1979 Altered linkage values in phage P22-mediated transduction caused by distant deletions or insertions in donor chromosomes. Mol. Gen. Genet. **176:** 87–93.

KRAJEWSKA-GRYNKIEWICZ, K., W. WALCZAK and T. KLOPOTOWSKI, 1971 Mutants of *Salmonella typhimurium* able to utilize D-histidine as a source of L-histidine. J. Bacteriol. **105:** 28–37.

LANGLEY, D. and J. R. GUEST, 1974 Biochemical and genetic characteristics of deletion and other

mutant strains of *Salmonella typhimurium* LT2 lacking α-keto acid dehydrogenase complex activities. J. Gen. Microbiol. **82:** 319–335.

OHTSUBO, H. and E. OHTSUBO, 1977 Repeated DNA sequences in plasmids, phages and bacterial chromosomes. pp. 49–63. In: *DNA Insertion Elements, Plasmids and Episomes*, Edited by A. I. BUKHARI, J. A. SHAPIRO and S. L. ADHYA. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

OZEKI, H., 1959 Chromosome fragments participating in transduction in *Salmonella typhimurium*. Genetics **44:** 147–470.

PEARCE, U. and B. A. D. STOCKER, 1965 Variation in composition of chromosome fragments transduced by phage P22. Virology **27:** 290–296.

POLLARD, J. H., 1977 Handbook of numerical and statistical techniques. Cambridge University Press, Cambridge.

ROTH, J. R. and P. E. HARTMAN, 1965 Heterogeneity in P22 transducing particles. Virology **27:** 297–307.

SANDERSON, K. E. and J. R. ROTH, 1983 Linkage map of *Salmonella typhimurium*, Ed. 6. Microbiol. Rev. **47:** 410–453.

SANDERSON, K. E. and Y. A. SAEED, 1972 P22 mediated transduction analysis of the rough A (rfa) region of the chromosome of *Salmonella typhimurium*. J. Bacteriol. **112:** 58–63.

SCHMIEGER, H. and H. BACKHAUS, 1976 Altered cotransduction frequencies exhibited by HT-mutants of *Salmonella*-phage P22. Mol. Gen. Genet. **143:** 307–309.

SMITH, H. O. and M. LEVINE, 1967 A phage gene controlling integration of prophage. Virology **31:** 207–216.

SMITH-KEARY, P. E. and G. W. P. DAWSON, 1963 Transduction analysis using leucine requiring mutants of *Salmonella typhimurium*. Genet. Res. **4:** 427–440.

STUTTARD, C., 1972 Location of *trpR* mutations in the *serB-thr* region of *Salmonella typhimurium*. J. Bacteriol. **111:** 368–374.

SUSSKIND, M. M. and D. BOTSTEIN, 1978 Molecular genetics of bacteriophage P22. Microbiol. Rev. **42:** 385–413.

WILD, J., W. WALCZAK, K. KRAJEWSKA-GRYNKIEWICZ and T. KLOPOTOWSKI, 1974 D-amino acid dehydrogenase: the enzyme of the first step of D-histidine and D-methionine racemization in *Salmonella typhimurium*. Mol. Gen. Genet. **128:** 131–146.

WU, T. T., 1966 A model for three-point analysis of random general transduction. Genetics **54:** 405–410.

Communicating editor: J. ROTH

APPENDIX 1: A METHOD FOR CALCULATION OF DISTANCES BETWEEN MARKERS FROM MULTIPLE POINT CROSSES

Let us consider a chromosomal region of less than phage DNA length where three markers are located. Three markers have been chosen for clarity of presentation, but the approach can be easily generalized for a greater number of markers. Let A, B and C be point markers with coordinates 1, $1 + t$ and $1 + u$. Let us assume that there are two classes of transducing particles active in transduction of each marker. The particle ends in interval (0, 1) are s_1 and

s_2. The weights (relative numbers of the two classes) are $w_1 = w$ and $w_2 = 1 - w$. Cotransduction frequency of markers B and C when B is the selected marker can be shown to equal

$$c_1(t, u, s_1, s_2, w) = \frac{\sum_{i=1}^{2} w_i[1 - \delta_{t,u}(s_i)][\delta_{u,1}(s_i) + t - s_i][\delta_{0,t}(s_i) + s_i - u]}{\sum_{i=1}^{2} w_i |t - s_i| (1 - |t - s_i|)} \quad \text{(A1)}$$

Index 1 (c_1) indicates that the first marker in a given pair is the selected one, index 2 (c_2) indicates that the second marker is the selected one. One should bear in mind that a simple model of the recombination process is inherent to this equation; namely, that the transducing fragment forms first a duplex with the complementary region of the chromosome and then two crossovers occur at random sites.

Similarly, when C is the selected marker,

$$c_2(t, u, s_1, s_2, w) = \frac{\sum_{i=1}^{2} w_i[1 - \delta_{t,u}(s_i)][\delta_{u,1}(s_i) + t - s_i][\delta_{0,t}(s_i) + s_i - u]}{\sum_{i=1}^{2} w_i |u - s_i| (1 - |u - s_i|)} \quad \text{(A2)}$$

Function $\delta_{a,b}$ is defined as

$$\delta_{a,b}(s_i) = \begin{cases} 1 & \text{if } a \leq s_i \leq b \\ 0 & \text{if } s_i < a \text{ or } s_i > b \end{cases}$$

Let us consider the situation when each one of three markers can be selected. Thus, six cotransduction data are available for the following pairs (first marker is the selected one): (A, B), (B, A), (B, C), (C, B), (A, C) and (C, A). Let us denote these values by a_i, $i = 1, 2, \cdots, 6$, and compare the computed and experimental values. By applying formulas (A1) and (A2) we get the following set of six algebraic nonlinear equations:

$$\begin{aligned} c_1(1, t, s_1, s_2, w) &= a_1 \\ c_2(1, t, s_1, s_2, w) &= a_2 \\ c_1(t, u, s_1, s_2, w) &= a_3 \\ c_2(t, u, s_1, s_2, w) &= a_4 \\ c_1(1, u, s_1, s_2, w) &= a_5 \\ c_2(1, u, s_1, s_2, w) &= a_6 \end{aligned} \quad \text{(A3)–(A8)}$$

with five unknown quantities; namely, t, u, s_1, s_2 and w.

Sets of nonlinear equations are usually solved numerically through seeking

such values of the unknown variables that give the best fit between computed and experimental values (here: cotransduction frequencies). The criterion of fitness is commonly the minimization of a least-square-type function. In this paper the function

$$F = \sum_{i=1}^{b} \frac{(a_i - a_i^{\text{comp}})^2}{a_i}$$

was applied [a_i^{comp} is the computed counterpart of a_i; see equations (1) and (2)]. Coefficients $1/a_i$ were chosen to render contributions of each cotransducibility value approximately equal. [The occurrence of transductant colonies is governed by the binomial distribution with cotransduction frequency being the probability of success and the number N of colonies tested being the number of trials. Standard deviation of the mean c (and the estimate of experimental random errors) is $\sigma = \sqrt{c(1-c)/N}$. Hence the ratio $\left(\text{compare to } \frac{(\alpha_i - a_i^{\text{comp}})^2}{a_i}\right)$ only weakly depends on c if c is much less than 1, and contributions of each component of function F are approximately equal].

A numerical program was developed on the CD 6000 computer to seek the approximate solution, defined as the set of t, u, s_1, s_2 and w values, that minimizes function F. Standard minimization subroutine from the CERN computer program library was used. One should have in mind, however, that function F may have not only one but several local minima. The values of unknown variables for these minima are recognized by the computer program as minimizing function F. One has to be extremely careful to perform the computations in such a way as to detect all local minima. The global minimum only yields the required solution of (A3)–(A8).

APPENDIX 2

Expected distance between markers and standard deviation of the estimate as a function of cotransduction frequency

Cotransduction frequency	$N = 1$		$N = 2$		$N = 3$		$N = 4$		$N = \infty$	
c	$\langle t \rangle$	σ_t	$\langle t \rangle$	σ_t	$\langle t \rangle$	σ_t	$\langle t \rangle$	σ_t	$\langle t \rangle$	σ_t
0.0	0.667	0.236	0.750	0.192	0.800	0.160	0.833	0.135	1.000	
0.05	0.633	0.224	0.644	0.139	0.627	0.111	0.619	0.098	0.632	
0.1	0.600	0.212	0.553	0.128	0.535	0.110	0.531	0.100	0.536	
0.15	0.567	0.200	0.491	0.125	0.474	0.109	0.472	0.098	0.469	0.000
0.2	0.533	0.189	0.441	0.124	0.424	0.105	0.424	0.094	0.415	for
0.3	0.467	0.165	0.373	0.120	0.355	0.098	0.346	0.084	0.331	all
0.4	0.400	0.141	0.307	0.110	0.294	0.088	0.286	0.076	0.263	c
0.5	0.333	0.118	0.257	0.097	0.239	0.076	0.228	0.063	0.206	
0.6	0.267	0.094	0.206	0.080	0.187	0.063	0.178	0.052	0.157	
0.7	0.200	0.071	0.154	0.061	0.138	0.048	0.131	0.041	0.112	
0.8	0.133	0.047	0.103	0.041	0.092	0.034	0.085	0.028	0.072	
0.9	0.067	0.024	0.052	0.021	0.046	0.017	0.042	0.015	0.035	
1.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

APPENDIX 3

Distance estimates, standard deviation and probability-density function for crosses with mutual substitution of selective and nonselective markers; $N = 2$

	c_A								
	0.05	0.1	0.15	0.2	0.3	0.4	0.5	0.7	0.9
c_B									
0.05	0.66								
	0.11								
	0.35								
0.1	0.63	0.56							
	0.13	0.13							
	10	16							
0.15	0.58	0.54	0.48						
	0.14	0.11	0.10						
	3.1	6.0	6.4						
0.2	0.57	0.52	0.47	0.43					
	0.15	0.12	0.10	0.11					
	1.4	3.1	4.3	4.3					
0.3	0.52	0.50	0.47	0.42	0.35				
	0.14	0.13	0.11	0.10	0.11				
	0.60	1.2	1.7	2.3	2.4				
0.4	0.46	0.46	0.43	0.40	0.34	0.29			
	0.13	0.11	0.10	0.10	0.09	0.10			
	0.23	0.57	0.77	1.2	1.6	1.9			
0.5	0.42	0.39	0.39	0.36	0.32	0.28	0.23		
	0.08	0.09	0.08	0.08	0.07	0.08	0.08		
	0.13	0.31	0.56	0.64	1.1	1.4	1.4		
0.7	0.27	0.27	0.25	0.25	0.24	0.22	0.19	0.13	
	0.04	0.04	0.06	0.04	0.03	0.04	0.04	0.04	
	0.03	0.09	0.13	0.25	0.41	0.61	0.91	1.4	
0.9	—	—	—	—	0.093	0.093	0.087	0.074	0.041
	—	—	—	—	0.014	0.014	0.013	0.012	0.012
	—	—	—	—	0.04	0.09	0.17	0.69	2.6

The following parameters are given in the table for each pair (c_A, c_B): expected distance between markers (upper number), standard deviation of the estimate (middle number) and the value of probability-density function (lower number). Values of c_A and c_B were grouped in classes defined as follows: for $c_A < 0.1$, $c_B < 0.1$ the classes were $(0.01k \pm 0.0005, 0.01l \pm 0.005)$, $k, l = 1, 2, \cdots, 10$; for other c_A and c_B the classes were $(0.05k \pm 0.025, 0.05l \pm 0.025)$, $k, l = 3, 4, \cdots, 20$. The computed matrices are symmetric with respect to the diagonal; therefore, the data are given only for $c_B > c_A$. The data are not presented if less than ten sequences $s_1, s_2, \cdots, s_N$, t, c_A, c_B were assigned to the class as indicated by dashes.

For $N = 1$, to get the exact values of distance and the probability-density function, one should use equations (9) and (10), respectively.