

I. A THEORETICAL STUDY OF THE REPRESENTATION OF GENOMIC
SEQUENCES IN PARTIAL DIGEST LIBRARIES

II. STUDIES OF THE BACTERIOPHAGE T4 PROXIMAL HALF TAIL FIBER

Thesis by

Brian Seed

In Partial Fulfillment of the Requirements

For the Degree of
Doctor of Philosophy

California Institute of Technology

Pasadena, California

1981

(Submitted May 23, 1979)

Acknowledgement

Portions of this work were supported by the National Institutes of Health, and the Jean Weigle Memorial Fund. I would like to thank Bernita Larsh and Connie Katz for help in typing the thesis.

ABSTRACT

Part I of this thesis is a theoretical investigation of the fraction of a (suitably idealized) genomic DNA which can be found among restriction enzyme partial digest fragments having lengths between L and $L+r$ basepairs long.

Part II of this thesis summarizes some experimental investigations of the structure and evolution of the bacteriophage T4 proximal half tail fiber, its interactions with the distal half-fiber, and the variety of phenotypes exhibited by temperature-sensitive mutations in its structural gene.

TABLE OF CONTENTS

Page

CHAPTER 1

A Theoretical Study of the Fraction of a Genome That Can Be Incorporated in a Recombinant DNA Partial Digest Library	1
Introduction	2
Results and Discussion	3
References	44

CHAPTER 2

Homology Between T4 and Several Related Phages in the Structural Gene for the Proximal Half Tail Fiber	45
Introduction	46
Materials and Methods	47
Results	49
Discussion	85
References	87

CHAPTER 3

Investigation of the Interaction Between Bacteriophage T4 Proximal and Distal Half Tail Fibers Using a Novel Method of Pseudorevertant Analysis	89
Introduction	90
Materials and Methods	93
Results	94
Discussion	120
References	124

CHAPTER 4

Analysis of Temperature-Sensitive Mutations in Bacteriophage T4 Gene 34	125
Introduction	126
Materials and Methods	128
Results	131
Discussion	169
References	176

CHAPTER 1

A Theoretical Study Of The Fraction Of A Genome That
Can Be Incorporated In A Recombinant DNA
Partial Digest Library

Brian Seed and Richard C. Parker

Division of Biology
California Institute of Technology
Pasadena, California 91125

INTRODUCTION

Two approaches have been taken to degrade chromosomal DNA to an appropriate length for cloning in libraries of recombinant phage. The first, controlled shear, yields DNA fragments which must be subsequently adapted with either synthetic restriction sites or homopolymer tails to be cloned. At present, neither of these manipulations allow sheared fragments to be efficiently cloned (Maniatis et al.(1978), Meyerowitz et al. (1981)). A second approach, partial digestion with restriction endonucleases, yields chromosomal fragments which can be efficiently ligated and packaged in vitro. However sequences lacking appropriate flanking restriction sites cannot be cloned.

In the following paper we present a theoretical treatment of the fraction of a genome that can be incorporated in a partial digest library. The fraction clonable depends on the size range of DNA fragments to be cloned, and the frequency of occurrence of the restriction site to be cleaved in the genomic DNA. We will derive the fraction clonable by two complementary approaches, each based on a sequential analysis of the conditional probabilities of cloning an arbitrarily chosen nucleotide. The fraction clonable is the net probability that the chosen nucleotide can be cloned. Under and overestimates of the

fraction clonable will be formed under the two approaches, and the resulting estimates compared.

Elsewhere, we will use the equations derived below to calculate the fraction of a genome which can be expected in a finite sample of clones (Seed, Parker, and Davidson (1981)).

RESULTS AND DISCUSSION

The Basic Assumption

We will assume that the probability of occurrence of a restriction enzyme cleavage site is the same for all bonds. This requires that cleavage sites occasionally fall on adjacent bonds, even though the minimum distance between sites should be four to six basepairs in most cases. In the calculations presented below the contribution of terms due to sites at adjacent bonds has been neglected, on the assumption that the average distance between sites will be large. In an earlier study, Hamer and Thomas (1975) have compared the mass distributions of restriction enzyme cleaved DNAs with the theoretical distributions predicted by the random dispersion of restriction sites throughout a genome. They found the hypothesis of random site placement valid within experimental error for both *E. coli* DNA and a

non-refractory component of *D. melanogaster* DNA. A highly reiterated low complexity fraction of *D. melanogaster* DNA did not display a random cleavage pattern, however. Although randomly dispersed repeated sequences are not expected to significantly affect these calculations, the results presented below will not in general apply to tandemly repeated sequences.

Some Simple Results

Two conditions permit the clonable genome fraction to be easily computed: either the target DNA is digested to completion, or the vector accepts arbitrarily small inserts. The latter is a special case of the former. Following a complete digestion, the clonable genome fraction is the weight fraction of DNA in restriction fragments shorter than the maximum and larger than the minimum insert length, and can be written (Kuhn (1930); Hamer and Thomas (1975))

$$(pL+1)e^{-pL} - (p(L+r)+1)e^{-p(L+r)} \quad (1)$$

where $L+1$ and $L+r$ are the minimum and maximum insert lengths in basepairs, and p is the frequency of occurrence of the restriction enzyme cleavage site. When the vector accepts arbitrarily small inserts, the fraction clonable is

$$1 - (pr+1)e^{-pr} \quad (2)$$

We note here for future comparison that formula (2) is an approximation: the exact result is

$$1 - (pr+1)(1-p)^r \quad (2a)$$

which is negligibly different from (2) when p is small.

Partial digestion allows some of the fragments shorter than the minimum insert length to be cloned in a replacement vector (a vector for which $L > 0$). Fragments larger than the maximum insert length can never be cloned, and so

$$1 - (p(L+r)+1)e^{-p(L+r)} \quad (3)$$

is an overestimate of the fraction clonable in a replacement vector. We will return to this formula later in the paper.

Two Derivations of the Genome Capacity of a Replacement Vector

I. Background, Nomenclature, and Conventions

We have derived two exact expressions for the genome capacity of a replacement vector. The first is a product of recursively defined conditional probabilities, representing

the fraction of a genome which cannot be cloned. The second is a probability sum representing the union of all clonable sequences. The clonable genome fraction is not explicitly computable from either expression, but rigorous under and overestimates can be formed and computed from both formulas.

For convenience we will briefly review here the properties of conditional probabilities. If the probability of A given B is $P(A|B)$ and the probability of the intersection of A and B $P(A \cap B)$, then $P(A \cap B) = P(A|B)P(B)$. We shall frequently use this formula in the form $P(A|B) = P(A \cap B)/P(B)$. When A and B are independent events $P(A|B) = P(A)$, and $P(A \cap B) = P(A)P(B)$. The probability of the union of two events, $P(A \cup B)$ is the sum of the probabilities of the two individual events minus the probability of their intersection: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. The conditional probability analog of this equation is $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$.

As above, let $L+1$ and $L+r$ be the minimum and maximum lengths in basepairs of the inserted DNA, and let p be the mean frequency of occurrence of a restriction site. Arbitrarily choosing a particular (index) nucleotide, we proceed, in one direction, to calculate the successive probabilities that a clonable fragment containing the index nucleotide can be formed (or not) from each bond in

question, given that all previous attempts have failed. As a matter of terminology we call the successively enumerated bonds "adjacent bonds." To each adjacent bond there correspond r "window" bonds, which represent the acceptable endpoints for clonable partial digest fragments terminating at the adjacent bond.

Figure 1 is a schematic diagram of the indexing system we have adopted. The first adjacent bond falls immediately to the left of the index nucleotide. The first window consists of the r bonds which fall immediately to the right of the $L+1$ th nucleotide.

II. The Fraction of a Genome That Cannot be Cloned in a Replacement Vector

Our first approach will lead to a probability product of the form

$$\prod_{i=0}^{L+r} K(i, r, L)$$

representing an intersection of failures at every trial. $K(i, r, L)$ is the probability that a clonable partial digest fragment cannot arise at the i th adjacent bond, given $i-1$ previous unsuccessful attempts. $K(0, r, L)=1$ by definition.

When $L+1 > r$, there are three forms of recursion relation

Figure 1. A schematic diagram illustrating the labelling conventions of the text. Nucleotides are represented by the symbol n , internucleotide bonds by the symbol $?$. The first window contains r bonds $L+1$ nucleotides to the left of the first adjacent bond. Successive trials initiate at progressively more leftward adjacent bonds.

for the $K(i, r, L)$. The first is applicable when $1 \leq i \leq r$; the second when $r+1 \leq i \leq L+1$, and the third when $L+2 \leq i \leq L+r$. Similarly, for $L+1 \leq r$ there are three forms applicable: when $1 \leq i \leq L+1$, $L+2 \leq i \leq r$, and $r+1 \leq i \leq L+r$, respectively. For our first case we take $L+1 > r$.

The fraction not clonable when $L+1 > r$

In general $K(i, r, L)$ is the union of the conditional probabilities that either (1) no site will be found at the i th adjacent bond, given $i-1$ previous failures, or (2) no site will be found in the i th window, given $i-1$ previous failures. (1) is just $(1-p)$, and (2) we call $w(i)$. The union is the sum of both conditional probabilities minus their intersection, or

$$K(i, r, L) = 1 - p + w(i) - (1-p)w(i) = 1 - p + pw(i) \quad (4)$$

When $i=1$ there have been no previous failures so $w(1)$ is the unconditional probability that no sites fall in the first window, or $(1-p)^r$. Hence

$$K(1, r, L) = 1 - p + p(1-p)^r = 1 - p(1 - (1-p)^r) \quad (5)$$

On the second trial the window is displaced one bond to the left (Fig. 1). The probability that a site falls on

the leftmost bond of the second window shows no dependence on the distribution of sites in the first window, and is just $(1-p)$. Hence we can write $w(2)=(1-p)d(2)$, where $d(2)$ is the probability that no sites will be found among the $r-1$ rightmost bonds of the second window, given failure on the first trial. In general, for $i \leq L+2$

$$w(i) = (1-p)d(i) \quad (6)$$

and $d(i)$ is the conditional probability that there will be no sites in the $r-1$ rightmost bonds of the i th window, given $i-1$ previous failures. Since the $r-1$ rightmost bonds of the i th window are identical to the $r-1$ leftmost bonds of the $i-1$ th window, $d(i)$ can be written as

$$d(i) = s(i-1) / K(i-1, r, L)$$

where $s(i-1)$ is the probability that no sites fall among the $r-1$ leftmost bonds of the $i-1$ th window, given $i-2$ previous failures, and $K(i-1, r, L)$ is the probability that the $i-1$ th trial fails, given $i-2$ previous failures.

$s(i-1)$ can be written as the sum of two probabilities: $w(i-1)$, the probability that there are no sites in the $i-1$ th window, and $t(i-1)$, the probability that there is only one site, at the rightmost bond. Both $t(i-1)$ and $w(i-1)$ are conditional probabilities, given $i-2$ previous failures.

When $i=2$, $w(i-1) = (1-p)^r$, and $t(i-1) = p(1-p)^r$. In this case both probabilities are unconditional.

Using formula (4) we can always relate the $w(i-1)$ to the $K(i-1, r, L)$ and so obtain a recursion. $t(i-1)$ does not appear in the definition of $K(i-1, r, L)$, (4), and will be described below.

Because $t(i-1)$ is a conditional probability, given $i-2$ previous failures, there can be no sites at any of the preceding adjacent bonds whose windows overlap the rightmost bond of the i -th window. Thus on the third trial, $t(i-1) = t(2) = p(1-p)^{r+1} / K(1, r, L)$. The extra factor of $(1-p)$ represents the probability that no site has fallen on the adjacent bond of the first trial, and the factor of $1/K(1, r, L)$ converts the numerator into a conditional probability, given the failure of the first trial. Extending this reasoning, we find that in the region defined by $2 \leq i \leq r$,

$$t(i-1) = p(1-p)^{r-1} (1-p)^{i-1} / \prod_{j=0}^{i-2} K(j, r, L) \quad (7)$$

The factor of $(1-p)^{i-1}$ ensures that the $i-1$ preceding adjacent bonds will not give rise to clonable fragments terminating at the last bond of the i -th window. The product of $K(j, r, L)$ in the denominator converts the

numerator into a conditional probability, given $i-2$ previous failures. Collecting results, we find

$$K(i, r, L) = 1 - p + p(1-p)(w(i-1) + t(i-1)) / K(i-1, r, L) \quad (8)$$

and, since

$$w(i-1) = [K(i-1, r, L) - (1-p)] / p \quad (9)$$

$$K(i, r, L) = (1-p) \left[2 - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2 (1-p)^{r+i-2}}{\prod_{j=0}^{i-1} K(j, r, L)} \right] \quad (10)$$

in the region defined by $2 \leq i \leq r$.

When $r+1 \leq i \leq L+1$,

$$t(i-1) = p(1-p)^{2r-1} / \prod_{j=i-r}^{i-2} K(j, r, L) \quad (11)$$

and

$$K(i, r, L) = (1-p) \left[2 - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2 (1-p)^{2r-1}}{\prod_{j=i-r}^{i-1} K(j, r, L)} \right] \quad (12)$$

in a fairly simple extension of (10).

In the region defined by $L+2 \leq i \leq L+r$, the window contracts on successive trials without exposing any new bonds (Fig. 1). Thus

$$w(i) = d(i)$$

and the factor of $1-p$ in (6) which describes the absence of a site at the leftmost bond of the window is lost. To derive $t(i-1)$ here we note first that if $i = L+k$ there are $r-(k-1)$ bonds in the i th window, and the rightmost bond of

the $i-1$ th window is $r-k+2$ bonds distant from the index nucleotide. $p(1-p)^{r-k+1} = p(1-p)^{L+r+1-i}$ is the a priori probability that the sole site in the $i-1$ th window falls on the rightmost bond. Since the rightmost bond also falls in the $r-1$ preceding windows, we must include a factor of $(1-p)^r$ for the probability that none of these adjacent bonds contain restriction sites. Hence

$$t(i-1) = p(1-p)^{L+2r+1-i} / \prod_{j=i-r}^{i-2} K(j, r, L) \quad (13)$$

and

$$K(i, r, L) = 2 - p - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{L+2r+1-i}}{\prod_{j=i-r}^{i-1} K(j, r, L)} \quad (14)$$

$$L+2 \leq i \leq L+r$$

The fraction not clonable when $L+1 \leq r$

For $L+1 \leq r$ a different treatment is necessary. In the region defined by $L+2 \leq i \leq r+1$, successive windows contract, making $p(1-p)^{L+r+1-i}$ the a priori probability that the sole site in the $i-1$ th window falls on the rightmost bond; however there are only $i-1$ adjacent bonds which can be said to have no sites (with probability $(1-p)^{i-1}$) and so

$$t(i-1) = p(1-p)^{r+L} / \prod_{j=0}^{i-2} K(j, r, L) \quad (15)$$

$w(i) = s(i-1)/K(i-1)$ as in the previous analysis of a contracting window; thus

$$K(i, r, L) = 2 - p - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{r+L}}{\prod_{j=0}^{i-1} K(j, r, L)} \quad (16)$$

$$L+2 \leq i \leq r+1$$

For $i \leq L+1$ formula (10) holds, and for $i > r+1$, formula (12).

A novel derivation of formula (2)

The above method of analysis applies to the calculation of the fraction of a genome which cannot be cloned in an insertion vector ($L=0$). Here $K(i, r, 0) = 1 - p + pw(i)$ as before; and $w(i) = s(i-1)/K(i-1)$. There are $r-(i-2)$ bonds in the i -1th window, so the probability that the sole site in the i -1th window falls on the rightmost bond is $p(1-p)^{r-i+1}$. Only $i-1$ adjacent bonds have been considered, so

$$t(i-1) = p(1-p)^r / \prod_{j=0}^{i-2} K(j, r, 0) \quad (17)$$

and

$$K(i, r, 0) = 2 - p - \frac{(1-p)}{K(i-1, r, 0)} + \frac{p^2(1-p)^r}{\prod_{j=0}^{i-1} K(j, r, 0)} \quad (18)$$

We can now prove that

$$\prod_{j=0}^i K(j, r, 0) = (1-p)^i + p^i(1-p)^r \quad (19)$$

by induction. We note first that $K(1, r, 0) = 1 - p + p(1-p)^r$.

Under the hypothesis given,

$$K(i, r, 0) = \frac{(1-p)^i + p^i(1-p)^r}{(1-p)^{i-1} + p^{i-1}(1-p)^r} \quad (20)$$

Substituting in (18), we find

$$K(i+1, r, 0) = 2 - p - [(1-p)[(1-p)^{i-1} + p(i-1)(1-p)^i] + p^2(1-p)^i] / [(1-p)^i + p i (1-p)^i]$$

$$= [(1-p)^{i-1} + p(i+1)(1-p)^i] / [(1-p)^i + p i (1-p)^i]$$

as desired. Hence the fraction clonable is

$$1 - \prod_{j=0}^r K(j, r, 0) = 1 - (p^{r+1})(1-p)^r \approx 1 - (p^{r+1})e^{-p^r}$$

or formula (2).

The fraction not clonable when $r=1$

Another limit which can be computed explicitly is that in which the window consists of a single bond. All trials are then independent of each other, and the probability of never forming a clonable fragment should be the $L+1$ th power of $K(1, 1, L)$, or $(1 - p^2)^{L+1}$. We will show that the recursion formulas (10) and (12) yield this result. Substituting $r=1$ in (10) we get

$$K(2, 1, L) = (1-p) \left[2 - \frac{1-p}{1-p^2} + \frac{p^2(1-p)}{1-p^2} \right] = 1 - p^2$$

($i = 2$)

By hypothesis $K(i-1, 1, L) = 1 - p^2$, so (12) yields

$$K(i, 1, L) = (1-p) \left[2 - \frac{1}{1+p} + \frac{p^2}{1+p} \right] = 1 - p^2$$

as above. Equation (14) has no counterpart in the single bond limit.

III. The Fraction of a Genome That Can be Cloned in a Replacement Vector

The preceding formulas have described the fraction of a genome which can not be incorporated in a partial digest library. In the following we describe the complement of this fraction, the union of all clonable sequences. We will find it convenient to choose a set of mutually exclusive events to describe the desired union. The probability of the union will then be the sum of the probabilities of the individual events. To implement this, at the i th adjacent bond we derive the probability that the first $i-1$ trials fail, and the i th trial then succeeds. No two such combinations of different trial outcomes can ever simultaneously occur. The sum over all i of the probability that the i th trial yields the first success is the probability that at least one trial succeeds.

Formally we seek the sum

$$\sum_{i=1}^{L+C} p b(i) \quad (21)$$

where $p b(i)$ is the probability that a clonable fragment containing the index nucleotide will terminate at the i th adjacent bond, and that $i-1$ failures have already occurred. p is the probability that a site falls on the i th adjacent bond, and $b(i)$ is the probability that at least one site

falls in the i th window, and that $i-1$ failures have already occurred.

The fraction clonable when $L+1 > r$

For $i < L+2$ we follow a reasoning similar to that employed in the preceding section, and write

$$b(i) = p \prod_0^{i-1} K(j, r, L) + (1-p)c(i) \quad (22)$$

where $p \prod_0^{i-1} K(j, r, L)$ is the probability that the leftmost bond of the i th window contains a site, and that $i-1$ failures have occurred. $c(i)$ is the probability that the $r-1$ rightmost bonds of the i th window contain at least one site, and that $i-1$ failures have occurred. Because the definition of $c(i)$ incorporates the failure of all prior trials, we are required, for $i > 1$, to ensure that no prior successes arise from sites falling in the i th window. Thus whenever a restriction site falls in the i th window, there must be no sites at any of the adjacent bonds whose windows overlap that site. For example, when $i=2$, if a restriction site falls anywhere in the rightmost $r-1$ bonds of the second window, the first adjacent bond must not have borne a site. For $i > 2$, the situation is more complicated. For simplicity, we will derive $c(i)$ in the region defined by $r+1 \leq i \leq L+1$ first. Figure 2 is a schematic diagram illustrating our

Figure 2. A schematic diagram illustrating the derivation of $c(i)$, $r+1 \leq i \leq L+1$. Nucleotides are represented by the symbol n . Internucleotide bonds are represented by the symbol $+$ when the bond must bear a restriction enzyme cleavage site; by the symbol $-$ when the bond must not bear a site; and by the symbol $?$ when there is no information. When the rightmost site in the i th window falls on the n th bond from the right end, the absence of sites in the $n-1$ bonds to the right of the rightmost site affects the preceding $n-1$ windows. Similarly, the presence of the rightmost site in the i th window affects the $r-n$ preceding adjacent bonds.

reasoning.

The probability that the rightmost site in the i th window falls on the n th bond from the right end is $p(1-p)^{n-1}$. When this occurs the previous $r-n$ adjacent bonds must not have contained any restriction sites, with probability $(1-p)^{r-n}$. The probability that the rightmost site falls on the n th bond and the preceding $r-n$ trials have failed is then $p(1-p)^{n-1}(1-p)^{r-n} = p(1-p)^{r-1}$.

Now consider the $r+n$ th preceding trial. The window for this trial does not contain the site at the n th bond from the right end, but it does contain $n-1$ bonds at which we know there are no restriction sites. In fact, from the i -rth trial to the $i-r+n-1$ th, the number of possible bonds at which sites may be found decreases by one for each trial. The effective window for each trial thus decreases with each successive trial. We have treated this situation before, in deriving formula (14), which described the contracting window region $L+2 \leq i \leq L+r$. Replacing $K(i, r, L)$ in formula (14) by $K(j, r, i-r)$, we can say that the conditional probability that the j th trial fails, given $j-1$ failures, is, for $i-r+1 \leq j \leq i-r+n-1$

$$K(j, r, i-r) = 2-p - \frac{(1-p)}{K(j-1, r, i-r)} + \frac{p^2(1-p)^{r+i-j}}{\prod_{m=j-r}^{j-1} K(m, r, i-r)} \quad (23)$$

The complete probability that the rightmost site in the i th window falls on the n th bond from the right end, and that $i-1$ failures have already occurred is

$$\rho(1-\rho)^{n-1}(1-\rho)^{r-n} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \quad (24)$$

where for the first $i-r$ trials, $K(j, r, i-r) = K(j, r, L)$. Summing over n , $1 \leq n \leq r-1$, we get the probability that at least one site falls in the $r-1$ rightmost bonds of the i th window, and that $i-1$ failures have occurred;

$$\rho(1-\rho)^{r-1} \sum_{n=1}^{r-1} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \quad (25)$$

Using (21) and (22), the probability that the i th trial succeeds and all previous trials have failed is then

$$\rho b(i) = \rho^2 \prod_{j=0}^{i-1} K(j, r, L) + \rho^2(1-\rho)^r \sum_{n=1}^{r-1} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \quad (26)$$

In the region defined by $1 \leq i \leq r$, formula (22) still applies. The probability that the rightmost site in the i th window falls on the n th bond from the right end is $\rho(1-\rho)^{n-1}$. However it is not necessarily true that there must be $r-n$ previous adjacent bonds which cannot have borne sites, since there have only been $i-1$ prior trials (Fig. 3a). If $r-n \geq i-1$, the probability that the rightmost bond falls on the n th bond from the right end and that the previous $i-1$ trials have failed is $\rho(1-\rho)^{n-1}(1-\rho)^{i-1}$. If

Figure 3. a. A schematic diagram illustrating the derivation of $c(i)$, $1 \leq i \leq r$, $r-n \leq i-1$. Symbols are as described for figure 2, and only the first i windows are shown. For $r-n \leq i-1$ all prior windows contain the rightmost site of the i th window, and hence all prior adjacent bonds can not bear sites.

b. A schematic diagram illustrating the derivation of $c(i)$, $1 \leq i \leq r$, $r-n < i-1$. Symbols are as described for figure 2, and only the windows are shown. The first window contains only $i-1$ bonds which may contain restriction sites. The $i-r+n$ th window is the first to contain the rightmost site of the i th window.

$r-n < i-1$, the probability that the rightmost site falls on the n th bond from the right end and that the previous $r-n$ trials have failed is $p(1-p)^{r-1}(1-p)^{r-n}$ as in the case just treated (Fig. 3b). The trials preceding the $i-(r-n)$ th must be described by a new formula, however. The effective window for the first trial contains only $i-1$ bonds; for the second, $i-2$; and so on. There are $i-1-(r-n)$ windows which do not contain the site at the n th bond of the i th window. The probability of the successive failure of these $i-r+n-1$ trials is the same as the product of the first $i-r+n-1$ terms of the fraction not clonable in an insertion vector with capacity $i-1$ basepairs, for which

$$K(j, i-1, 0) = 2-p - \frac{(1-p)}{K(j-1, i-1, 0)} + \frac{p^2(1-p)^{i-1}}{\prod_{m=0}^{j-1} K(m, i-1, 0)} \quad (27)$$

is the probability that the j th trial fails ($1 \leq j \leq i-r+n-1$), given $j-1$ previous failures. Collecting results, we find, for $r-n < i-1$, the probability that the rightmost bond falls on the n th bond from the right end of the i th window, and that $i-1$ prior failures take place is (using (19))

$$p(1-p)^{r-1} \prod_{j=0}^{i-r+n-1} K(j, i-1, 0) = p(1-p)^{r-1} \left[(1-p)^{i-r+n-1} + p(i-r+n-1)(1-p)^{i-1} \right] \quad (28)$$

Taking into account the trials for which $r-n \geq i-1$, we find

$$c(i) = \sum_{n=1}^{r-i+1} p(1-p)^{n-1}(1-p)^{i-1} + \sum_{n=r-i+2}^{r-1} p(1-p)^{r-1} \prod_{j=0}^{i-r+n-1} K(j, i-1, 0) \quad (29)$$

and (using (22)),

$$p b(i) = p^2 \prod_{j=0}^{i-1} K(j, r, L) + p(1-p)^i [1 - (1-p)^{r-i+1}] \quad (30)$$

$$+ p(1-p)^{r+1} [1 - (1-p)^{i-2} + p^2(i-2)(i-1)(1-p)^{i-2}/2]$$

$$p b(1) = p(1 - (1-p)^r) \quad \text{as expected.}$$

Within the region defined by $L+2 \leq i \leq L+r$, the index n can only rise to $L+r+1-i$, the width of the window. Here we find

$$c(i) = p(1-p)^{r-1} \sum_{n=1}^{L+r+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \quad (31)$$

and

$$p b(i) = p^2(1-p)^{r-1} \sum_{n=1}^{L+r+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \quad (32)$$

There is no p^2 term because no new window bonds are unmasked.

The fraction clonable when $L+1 \leq r$

In the event that $L+1 \leq r$, a new treatment is needed for the region defined by $L+1 \leq i \leq r$. As in the preceding case, the index n can only rise to $L+r+1-i$. At the same time the effective window for the rightmost adjacent bond is only $i-1$ bonds wide. For $r-n \geq i-1$ all adjacent bonds are counted in the product $p(1-p)^{n-1}(1-p)^{i-1}$ as before. The sum over n gives

$$c(i) = \sum_{n=1}^{r-i+1} p(1-p)^{n-1}(1-p)^{i-1} + p(1-p)^{r-1} \sum_{n=r-i+2}^{L+r+1-i} \prod_{j=0}^{i-r+n-1} K(j, i-1, i-r) \quad (33)$$

and

$$p_b(i) = p(1-p)^i [1 - (1-p)^{r-i+1}] + p^2(1-p)^{r-1} \sum_{n=r-i+2}^{L+1-i} \prod_{j=0}^{i-r+n-1} K(j, i-1, i-r) \quad (34)$$

For $i \leq L+1$ formula (30) holds, and for $i > r+1$, formula (32).

Another derivation of formula (2)

The analysis above can easily be turned to insertion vector libraries ($L=0$). Since the window contracts with each trial, the formula for the region defined by $1 \leq i \leq r$ can be modified by the removal of all terms due to $r-n < i-1$. The sum over n gives

$$c(i) = \sum_{n=1}^{r-i+1} p(1-p)^{n-1}(1-p)^{i-1} = (1-p)^{i-1} [1 - (1-p)^{r-i+1}] \quad (35)$$

and

$$p_b(i) = p(1-p)^{i-1} [1 - (1-p)^{r-i+1}] \quad (36)$$

The complete union of successes gives

$$\sum_{i=1}^r p_b(i) = 1 - (1-p)^r - pr(1-p)^r \approx 1 - (pr+1)e^{-pr}$$

as desired.

A Derivation of Computable Under and Overestimates

An underestimate of the fraction not clonable, $L+1 > r$

In table 1 we have collected the results of the preceding derivations. On examining equations (10), (12), and (14), it should be clear that a recursion formula which eliminates or underestimates the terms derived from $t(i-1)$ must also underestimate $K(i, r, L)$. For example the approximation

$$K_a(i, r, L) = (1-p) [2 - (1-p)/K_a(i-1, r, L)] \quad (37)$$

$$2 \leq i \leq L+1$$

can be shown to yield

$$\prod_{i=1}^{L+1} K_a(i, r, L) = (1-p)^{L+1} [1 + (L+1)(1-p)^{r-1}] \quad (38)$$

when $K_a(1, r, L) = 1 - p(1 - (1-p)^r)$. We can form a better underestimate by replacing

$$\prod_{j=1}^{i-2} K(j, r, L)$$

in equation (10) by $(1-p)^{i-r-1}$, $2 \leq i \leq r$. Since $1 \leq (1-p)^{i-r-1}$ for all i in this region, and

$$\prod_{j=1}^{i-2} K(j, r, L) < 1$$

the replacement underestimates $K(i, r, L)$. Labeling the underestimate $K'(i, r, L)$ we find

$$K'(i, r, L) = (1-p) [2 - (1-p - p^2(1-p)^{2r-1})/K'(i-1, r, L)] \quad (39)$$

a tractable one-term recursion. Substituting $K'(1, r, L) = 1 - p(1 - (1-p)^r)$ we find $K'(2, r, L) = 1 - p(1 - (1-p)^r)$.

Table 1

COMPLETE FORMULAS FOR THE GENOME FRACTION THAT CAN OR CANNOT
BE INCORPORATED IN A PARTIAL DIGEST LIBRARY

The Fraction Not Clonable For $L+1 > r$

$$\begin{aligned} \prod_{i=1}^{L+r} K(i, r, L) &= \prod_{i=L+2}^{L+r} \left\{ 2 - p - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{L+2r+1-i}}{\prod_{j=i-r}^{i-1} K(j, r, L)} \right\} \\ &\cdot \prod_{i=r+1}^{L+1} \left\{ (1-p) \left[2 - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{2r-1}}{\prod_{j=i-r}^{i-1} K(j, r, L)} \right] \right\} \\ &\cdot \prod_{i=1}^r \left\{ (1-p) \left[2 - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{r+i-2}}{\prod_{j=0}^{i-1} K(j, r, L)} \right] \right\} \end{aligned}$$

The Fraction Not Clonable For $L+1 \leq r$

$$\begin{aligned} \prod_{i=1}^{L+r} K(i, r, L) &= \prod_{i=L+2}^{L+r} \left\{ 2 - p - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{L+2r+1-i}}{\prod_{j=i-r}^{i-1} K(j, r, L)} \right\} \\ &\cdot \prod_{i=L+2}^{r+1} \left\{ 2 - p - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{L+r}}{\prod_{j=0}^{i-1} K(j, r, L)} \right\} \\ &\cdot \prod_{i=1}^{L+1} \left\{ (1-p) \left[2 - \frac{(1-p)}{K(i-1, r, L)} + \frac{p^2(1-p)^{r+i-2}}{\prod_{j=0}^{i-1} K(j, r, L)} \right] \right\} \end{aligned}$$

The Fraction Clonable For $L+1 > r$

$$\begin{aligned} \sum_{i=1}^{L+r} p b(i) &= p^2 \sum_{i=1}^{L+1} \prod_{j=0}^{i-1} K(j, r, L) + \sum_{i=1}^r p(1-p)^i [1 - (1-p)^{r-i+1}] \\ &+ \sum_{i=2}^r p(1-p)^{r+1} [1 - (1-p)^{i-2} + \frac{p^2}{2} (i-1)(i-2)(1-p)^{i-2}] \\ &+ \sum_{i=r+1}^{L+1} p^2(1-p)^r \sum_{n=1}^{i-r} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \\ &+ \sum_{i=L+2}^{L+r} p^2(1-p)^{r-1} \sum_{n=1}^{L+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \end{aligned}$$

The Fraction Clonable For $L+1 \leq r$

$$\begin{aligned} \sum_{i=1}^{L+r} p b(i) &= p^2 \sum_{i=1}^{L+1} \prod_{j=0}^{i-1} K(j, r, L) + \sum_{i=1}^{L+1} p(1-p)^i [1 - (1-p)^{r-i+1}] \\ &+ p(1-p)^{r+1} \sum_{i=1}^{L+1} [1 - (1-p)^{i-2} + \frac{p^2}{2} (i-1)(i-2)(1-p)^{i-2}] \\ &+ \sum_{i=L+2}^r p(1-p)^i [1 - (1-p)^{r-i+1}] + \sum_{i=L+2}^r p^2(1-p)^{r-1} \sum_{n=i-r+2}^{L+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \\ &+ \sum_{i=r+1}^{L+1} p^2(1-p)^{r-1} \sum_{n=1}^{L+r+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \end{aligned}$$

Hence for all i , $1 \leq i \leq r$, $K'(i, r, L) = K'(1, r, L)$. We similarly modify (12) so that

$$\prod_{j=i-r}^{i-2} K(j, r, L)$$

is replaced by unity. Then

$$\prod_{j=0}^L K'(j, r, L) = (1 - \rho(1 - (1 - \rho)^r))^L \quad (40)$$

In the region defined by $L+2 \leq i \leq L+r$ we form an underestimate by replacing (14) with

$$K_0(i, r, L) = 2 - \rho - (1 - \rho)/K_0(i-1, r, L) \quad (41)$$

We will show

$$\prod_{i=0}^j K_0(n-i, r, L) = \frac{1}{\rho} (1 - (1 - \rho)^{j+1}) K_0(n-j, r, L) - \frac{(1 - \rho)}{\rho} (1 - (1 - \rho)^j) \quad (42)$$

by induction. First

$$\prod_{i=0}^1 K_0(n-i, r, L) = (2 - \rho) K_0(n-1, r, L) - (1 - \rho) = \frac{1}{\rho} (1 - (1 - \rho)^2) K_0(n-1, r, L) - (1 - \rho)$$

as desired; then

$$\prod_{i=0}^j K_0(n-i, r, L) = \frac{1}{\rho} (1 - (1 - \rho)^{j+1}) \left[2 - \rho - \frac{(1 - \rho)}{K_0(n-j-1, r, L)} \right] - \frac{(1 - \rho)}{\rho} (1 - (1 - \rho)^j)$$

so that

$$\prod_{i=0}^{j+1} = \frac{1}{\rho} (1 - (1 - \rho)^{j+2}) K_0(n-j-1, r, L) - \frac{(1 - \rho)}{\rho} (1 - (1 - \rho)^{j+1})$$

To construct the complete underestimate we want

$$\prod_{i=0}^{r-1} K_0(L+r+1-i, r, L) \prod_{j=0}^L K'(j, r, L) \quad (43)$$

Since $K'(j, r, L) = 1 - p(1 - (1-p)^r)$, $1 \leq j \leq L+1$, we find, after some algebra,

$$\begin{aligned} \prod_{i=0}^{r-1} K_0(L+r+1-i, r, L) \prod_{j=0}^L K'(j, r, L) &= [2(1-p)^r - (1-p)^{2r}] [1 - p(1 - (1-p)^r)]^L \\ &\approx (2e^{-p^r} - e^{-2p^r}) e^{-Lp(1-e^{-p^r})} \end{aligned} \quad (44)$$

An overestimate of the fraction not clonable, $L+1 > r$

Using the result just derived, we can now form an overestimate of the fraction not clonable by the simple expedient of substituting the underestimate $K'(i, r, L)$ for $K(i, r, L)$ in the formulas for the fraction clonable, (26), (30), and (32). For $r < L+1$

$$\begin{aligned} \sum_{i=1}^{L+r} p b(i) &= 1 - K(1, r, L) + p^2 \sum_{i=2}^{L+1} \prod_{j=0}^{i-1} K(j, r, L) \\ &+ \sum_{i=2}^r p(1-p)^i (1 - (1-p)^{r-i+1}) + \sum_{i=2}^r p(1-p)^{r+1} [1 - (1-p)^{i-2} - \frac{p^2(i-2)(i-1)}{2} (1-p)^{i-2}] \\ &+ \sum_{i=r+1}^{L+1} p^2(1-p)^r \sum_{n=1}^{r-1} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \\ &+ \sum_{i=L+2}^{L+r} p^2(1-p)^{r-1} \sum_{n=1}^{L-r+1-i} \prod_{j=0}^{i-r+n-1} K(j, r, i-r) \end{aligned} \quad (45)$$

Substituting $K_0(n-i, r, L)$ and $K'(j, r, L)$ where appropriate, we

$$\begin{aligned} \text{find } \sum_{i=1}^{L+r} p b'(i) &= 1 - K_1 + p^2 K_1 (1 - K_1^L) (1 - K_1) + (1-p)^2 (1 - (1-p)^{r-1}) \\ &+ p(1-p)^{r+2} (1 - r(1-p)^{r-1}) - p^2(r-1)r(1-p)^{2r}/2 - p(2-p)\lambda(1-p)^{r+1} (1 - (1-p)^{r-1}) \\ &+ p(r-1)(1-p)^{2r} (1 - K_1^{L-r}) \lambda (1 - (1-p)^r) + (1-p)^{r+1} (1 - (1-p)^{r-1}) (1 - K_1^{L-r}) \\ &+ (1-p)^r (K_1^{L+1-r} - K_1^L) + (1-p)(1 - (1-p)^r) ((1-p)^{r-1} K_1^{L+1-r} - K_1^L) \\ &+ (1-p)^{2r-1} (K_1^{L+1} - (rK_1 - r+1) K_1^{L+1-r}) \lambda (1 - (1-p)^r)^2 \end{aligned} \quad (46)$$

where $K_1 = 1 - p(1 - (1-p)^r)$. Neglecting terms of order p , factors of order $(1-p)^2$, and using the exponential approximation, we can simplify this to

$$\begin{aligned} \sum_{i=1}^{L+r} pb'(i) &\approx 1 - e^{-pr} - pr e^{-2pr} - (pr)^2 e^{-2pr}/2 \\ &+ [e^{-pr}(1 - e^{-pr}) + pr e^{-2pr}/(1 - e^{-pr})][1 - e^{-(L+r)p(1 - e^{-pr})}] \\ &+ [2e^{-pr} - e^{-2pr}(1 + (1 - pr(1 - e^{-pr}))/ (1 - e^{-pr})^2)] e^{-(L+r)p(1 - e^{-pr})} \\ &- [1 - e^{-2pr}/(1 - e^{-pr})^2] e^{-Lp(1 - e^{-pr})} \end{aligned} \quad (47)$$

An overestimate of the fraction not clonable, $L+1 \leq r$

Similar manipulations lead to

$$\begin{aligned} \sum_{i=1}^{L+r} pb'(i) &\approx 1 - (pr+1)e^{-(L+r)p} - (pr)^2 e^{-2pr}/2 \\ &+ e^{-prL} - e^{-prL}(1 - e^{-pr}) \\ &+ e^{-2pr} [e^{-Lp(1 - e^{-pr})} + Lp(1 - e^{-pr}) - 1]/(1 - e^{-pr})^2 \end{aligned} \quad (48)$$

when $L+1 \leq r$.

A second overestimate of the fraction not clonable, $L+1 > r$

Another approach to an overestimate of the fraction not clonable exploits the fact that $K(i, r, L)$ must be a monotone increasing function in i for $r > 1$. This allows us to conclude that

$$K_0'(i, r, L) = (1-p) \left[2 - \frac{(1-p)}{K_0'(i, r, L)} + \frac{p^2(1-p)^{2r-1}}{\prod_{i=r}^{i-1} K_0'(j, r, L)} \right] \quad (49)$$

is an overestimate for $K(i, r, L)$ since $(1-p)/K(i-1, r, L)$ in (12) has been replaced by the smaller $(1-p)/K(i, r, L)$.

Since

$$\prod_{j=i-r}^{i-1} K(j, r, L) > [1 - p(1 - (1-p)^r)]^r$$

$$K_0(i, r, L) = (1-p) \left[2 - \frac{(1-p)}{K_0(i, r, L)} + \frac{p^2(1-p)^{2r-1}}{[1-p(1-(1-p)^r)]^r} \right] \quad (50)$$

is also an overestimate. (50) can be solved for $K_0(i, r, L)$ taking the larger root of the quadratic equation (the smaller root yields $K_0 < (1-p)$). Thus

$$K_0(i, r, L) = 1-p + \frac{p^2(1-p)^{2r-1}}{2[1-p + p(1-p)^r]} + \frac{p(1-p)^r}{[1-p + p(1-p)^r]^{1/2}} \sqrt{1 + \frac{p^2(1-p)^{2r-2}}{4[1-p + p(1-p)^r]}} \quad (51)$$

and

$$K_0(i, r, L) < 1-p + \frac{p(1-p)^r}{[1-p + p(1-p)^r]^{1/2}} + \frac{p^2(1-p)^{2r-2}(5-4p)}{8[1-p + p(1-p)^r]} = K_0(i, L) \quad (52)$$

This treatment can be extended to $i > L+1$. We first note that

$$w(i) = (1-p)(w(i-1) + t(i-1))/K(i-1, r, L) \quad (53)$$

which may be written

$$w(i) = \frac{1}{1+x} (w(i-1) + t(i-1))$$

Since $w(i) \geq w(i-1)$, $t(i-1) \geq x w(i-1)$, or $t(i-1) \geq \frac{p}{1-p} (w(i-1))^2$. Then since

$$t(i) = \frac{p}{1-p} (1-p)^{2r} / \prod_{j=i-r+1}^{i-1} K(j, r, L) \quad (54)$$

and because it can be shown that

$$w(i) = (1-p)^r \prod_{j=0}^{i-1} K(j, r, i-r) / \prod_{j=0}^{i-1} K(j, r, L) \quad (55)$$

it follows that

$$\prod_{j=i-r+1}^{i-1} K(j, r, L) \geq \left[\prod_{j=i-r+1}^{i-1} K(j, r, i-r) \right]^2 \quad (56)$$

and so $[K_0(r, L)]^{L+r/2} \geq \prod_{i=0}^{L+r} K(i, r, L)$. For both mnemonic and computational convenience we will refer to a simpler form for $K_0(r, L)$ in which the term of order p^2 is ignored, and the exponential approximation is applied:

$$[K_0(r, L)]^{L+r/2} \approx e^{-p(L+r/2)(1-e^{-\frac{p}{2}(1+e^{-p^r})})} \quad (57)$$

For completeness we mention one additional approximation. This formula is similar to (44), and will be used in the derivation of the fraction of a genome incorporated in a finite partial digest library (Seed, Parker, and Davidson, 1981). Under the conditions to be described there

$$e^{-p(L+r/2)(1-e^{-p^r})} \quad (58)$$

is an underestimate of the fraction not clonable.

Comparison of the Various Estimates

We have compared the values returned by the various approximations listed in the text under a variety of different conditions. The underestimate (3) appears to be most valuable as the ratio r/L increases. For small r , formula (44) yields the better estimate. In figure 4 we have plotted three different approximations for the fraction not clonable in lambda phage and cosmid vectors as a function of the size range of acceptable inserts, r . For the lambda vector, the minimum length was chosen to be 14 kbp, and for the cosmid vector, 40 kbp. Figure 4 demonstrates that the best approximation varies with the dimensions of the inserted DNA. Curves 1 and 4 are overestimates derived from equation (57). Curves 2 and 5 are underestimates derived from (3), and curves 3 and 6 are underestimates from (44). In figure 5 we have plotted the best approximations describing the fraction of a genome which cannot be cloned as a function of the frequency of occurrence of the restriction site to be cleaved in the genomic DNA. Curve 1 describes the fraction not clonable in a vector similar to the lambda vehicle λ_{gt10} , and the curve shown is the overestimate of the fraction not clonable

Figure 4. The genome capacity of replacement vectors, $L+1 > r$. The logarithm of the fraction not clonable calculated from formulas (3), (44), or (57) of the text is plotted against the size range of the inserted DNA, r . The restriction site frequency is $1/(2.5 \text{ kbp})$, the observed frequency of occurrence of Eco RI sites in *Drosophila* (Hamer and Thomas, 1975) and mouse (Botchan et al, 1973) DNAs. The upper three curves represent a phage library with minimum insert length $L = 14 \text{ kbp}$. The lower three curves represent a cosmid library with minimum insert length $L = 40 \text{ kbp}$. Curves 1 and 4, 2 and 5, and 3 and 6 are derived from formulas (57), (3), and (44), respectively.

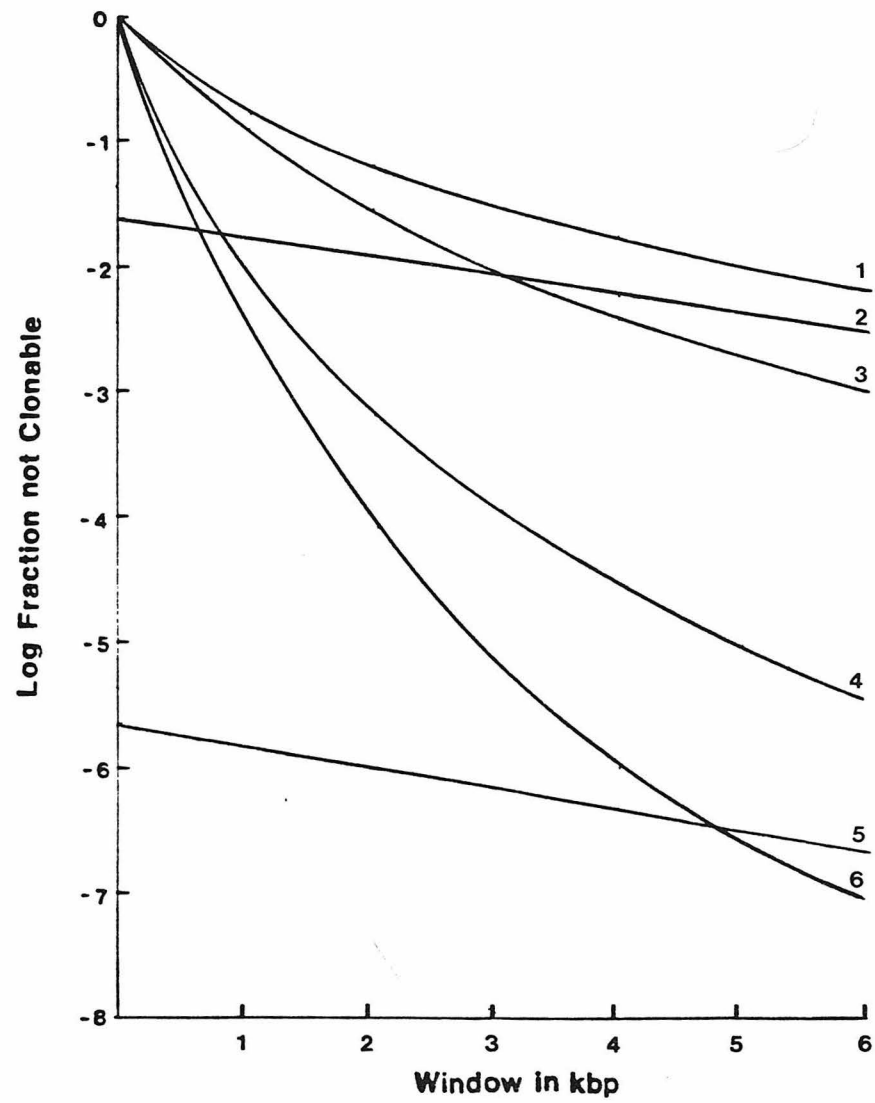
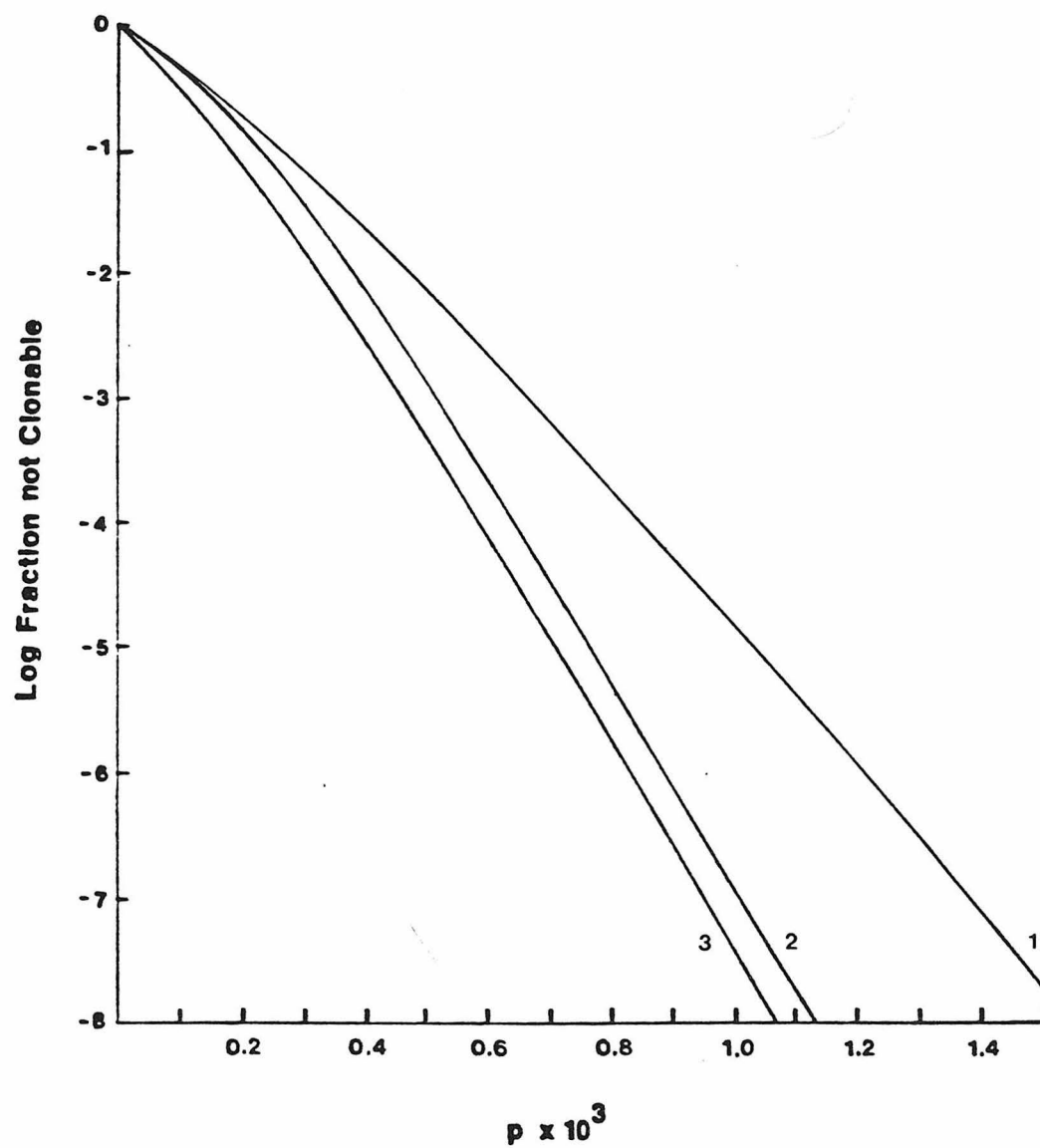


Figure 5. The effect of restriction site frequency on genome capacity. Curve 1 is a plot of the logarithm of the fraction not clonable in a vector accepting inserts ranging from 1.0 to 14.0 kbp. Over and underestimates can not be resolved in this curve. Curves 2 and 3 are the over and underestimates of the logarithm of the fraction not clonable in a vector accepting inserts ranging from 15 to 20 kbp. The abscissa is the frequency of occurrence of the restriction site cleaved in the genomic DNA.



(48). The sharpest underestimate in this case is given by (3), and is indistinguishable from the overestimate at the resolution of this graph. Curves 2 and 3 describe over and underestimates of the genome capacity of a vector accepting inserts ranging from 15 to 20 kbp in length. The overestimate is given by (57), and the underestimate by (3).

Figures 4 and 5 together satisfy the simple intuitive expectation that the fraction clonable should increase with increasing L , r , or p . Of practical interest are the bounds computed for the fraction clonable in currently popular vectors. In figure 4 we show that partial Eco RI digest fragments of mammalian DNAs, when inserted in high capacity lambda vectors, have a theoretical capacity of between 99.3% and 99.7% of all sequences. Similar cosmid-size fragments can accommodate all but 4×10^{-6} to 2×10^{-7} of a mammalian genome.

Figure 5, or the formulas (3) and (57), can be used to calculate the fraction clonable for other genomes or enzymes than modelled in figure 4. In principle the probability of any restriction enzyme recognition sequence can be calculated from the base compositions of the target DNA and the recognized sequence: for example the frequency of occurrence of the Eco RI recognition sequence GAATTC should be $p_A^4 p_G^2$ where p_A and p_G are the frequencies of

occurrence of A and G nucleotides in the target DNA. In practise this calculation may not accurately reflect the true prevalence of sites. A mild example of this inaccuracy is the discrepancy between the measured frequency of Eco RI cleavage sites in mammalian (Botchan et al., 1973) and *Drosophila* (Hamer and Thomas, 1975) DNAs, 1/(2.5 kbp), and the predicted frequency, 1/(3.2 kbp). The predicted mammalian frequency can be brought closer to the measured frequency by taking dinucleotide pair probabilities into account. With this correction the probability of any n-nucleotide sequence can be written

$$\frac{\prod_{i=1}^{n-1} D_i}{\prod_{j=2}^{n-1} P_j}$$

where D_i is the probability of the i th dinucleotide, and P_j is the probability of the j th nucleotide. For example the Eco RI frequency will be

$$\frac{D_{GA} D_{AA} D_{AT} D_{TT} D_{TC}}{P_A P_A P_T P_T}$$

and the Bam HI (GGATCC) frequency

$$\frac{D_{GG} D_{GA} D_{AT} D_{TC} D_{CC}}{P_G P_A P_T P_C}$$

Using these formulas, and the data of Swartz et al. (1962), the Eco RI frequency should be 1/(2.7 kbp), which is fairly close to the observed value. Thus with the appropriate precautions it should be possible in many cases to use

figure 5 to obtain a working estimate of the genome fraction clonable with different enzymes or DNAs.

REFERENCES

- Botchan, M., McKenna, G. and Sharp, P.A. (1973). Cleavage of mouse DNA by a restriction enzyme as a clue to the arrangement of genes. Cold Spring Harbor Symp. Quant. Biol. 38, 383-395.
- Hamer, D.H. and Thomas, C.A. (1975). The cleavage of *Drosophila melanogaster* DNA by restriction endonucleases. Chromosoma 49, 243-267.
- Kuhn, W. (1930). Kinetics of the destruction of high molecular weight chains. Ber. dtsh. chem. Ges. 63, 1503-1508.
- Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell 15, 687-701.
- Meyerowitz, E.M., Guild, G.M., Prestidge, L.S. and Hosness, D.S. (1981). A new high capacity cosmid vector and its use. Gene, in press.
- Seed, B., Parker, R.C. and Davidson, N. (1981). A theoretical treatment of the genome fraction contained in a finite partial digest library. Manuscript in preparation.
- Swartz, M.N., Trautner, T.A. and Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. J. Biol. Chem. 237, 1961-1967.

CHAPTER 2

Homology Between T4 and Several Related Phages in the Structural Gene for the Proximal Half Tail Fiber

Brian Seed and William B. Wood

Division of Biology

California Institute of Technology

Pasadena, California 91125

The bacteriophage T4 proximal half tail fiber activates the phage baseplate after one or several of the distal fiber tips have contacted the bacterial lipopolysaccharide phage receptor (Wilson et al., 1970; Yamamoto and Uchida, 1975; Crawford and Goldberg, 1977). The proximal half-fiber also mediates tail fiber extension and retraction in the presence of tryptophan or indole (Edgar et al., 1964). Not much is known of the mechanisms for either of these functions.

There is little evidence to suggest evolutionary divergence between the gene 34 homologs of the T-even bacteriophage family. The gene 34 products of phage T2, T4, and T6 are functionally interchangeable (Russell, 1967). Electron microscope heteroduplex studies have shown a single small nonhomology loop in the gene 34 region of each pairwise combination of T-even chromosomes (Kim and Davidson, 1974).

The frequency of recombination per base-pair rises sharply at the promoter-distal terminus of T4 gene 34, but is constant throughout T2 gene 34 (Beckendorf and Wilson, 1972). T6 gene 34 resembles T2 gene 34 in this respect (Russell, 1967). The selective advantages of the T4 gene 34 anomalous recombination gradient, if any, are not understood at present.

In 1964 Rosina Berry, an undergraduate research participant at the Cold Spring Harbor Laboratory for Quantitative Biology, isolated a group of bacteriophage serologically related to T4 from six Long Island sewage treatment plants. Some of these strains (collectively known as the RB phage) were partially characterized by Richard Russell in the course of his studies on the evolution and speciation of the T-even bacteriophage (Russell, 1967). We present here a detailed genetic and serological investigation of the proximal half tail fibers of a morphologically similar subset of the RB phage. We conclude this subset contains a family of bacteriophage evolutionarily intermediate between T2, T6, and T4.

Materials and Methods

Phage and bacterial strains. Phage strains derived from the wild-type T4D were obtained from the collections of R. S. Edgar and W. B. Wood and have been described elsewhere (Epstein et al., 1963; Edgar and Lielausis, 1965). Wild-type T2L and T6 strains and their derivatives were from the collection of R. L. Russell, as were all RB strains (Russell, 1967). Other strains were constructed by genetic crosses and are described in the text.

Escherichia coli. Strain CR63 was the permissive host for amber mutant phage strains. S/6/5, B/5, and G(λ) were the nonpermissive hosts. The permissive and restrictive temperatures for ts mutant phage were 25° and 42°C, respectively.

Media. H broth was used for phage and bacterial growth. EHA top and bottom agars for plating were prepared as described by Steinberg and Edgar (1962). Dilution buffer was prepared as described by King (1968). SM contains 0.1 M NaCl, 50 mM Tris-HCl, pH 7.5, 10 mM MgSO₄, and 0.01% gelatin.

Serum blocking assays. Serum blocking assays for proximal half-fiber antigen were performed with the antisera and protocol of Ward et al. (1970), as elaborated in the text. Some minor modifications were made as follows: the bacterial strain used as a plating indicator was S/6/5 rather than G(λ), and the standard phage contained amber mutations in genes 18 and 23 (sheath and capsid proteins) as well as the rII deletion *rdf41*. Plates measuring the residual titer of tester phage were occasionally incubated at 42°C to select against growth of temperature-sensitive phage being assayed. In some cases T4-RB hybrid phage bearing temperature-sensitive mutations were inactivated by ultraviolet irradiation to 20 to 40 lethal hits prior to serum neutralization. This procedure eliminated the background of hybrid phage and allowed the tester phage to be titered accurately. UV inactivation had no effect on the antigen titer of the standard phage. Viable phage were also eliminated by brief

incubation of samples at 90°–95°C. A pan of water was autoclaved and a rack of tubes containing samples to be blocked placed in the water immediately following removal from the autoclave. This procedure had no effect on the serum blocking profile of T4 or RB phage, and was easier to carry out when large numbers of phage were to be blocked.

Phage crosses. Initial phage crosses were carried out as described by Wilson and Kells (1972). Crosses between T4 and T6 strains were carried out by a modified protocol. B40 Su_I^+ was the usual amber permissive strain in these crosses. Bacteria were grown to mid-log phase at 30°C, harvested, and resuspended in H broth at a density of 2×10^9 cells/ml. Phage stocks were adjusted to a titer of 6×10^9 pfu/ml in SM, and equal volumes (usually 0.05 or 0.1 ml each) of the two phage to be crossed added to 15 x 85 mm tubes. Immediately before the cross the permissive bacteria were diluted with four volumes of SM to give a density of 4×10^8 cells/ml. A volume of bacteria equal to the combined volume of the two phage suspensions (net multiplicity of infection 15) was then added to each tube. After 6 min, one-tenth volume of anti-T4 serum diluted in H broth to a K of approximately 20/min was added. After another 6 min the phage were diluted either 1:400 once (for T4:T6 crosses) or 1:200 twice (for homologous crosses) in 15 x 85 mm tubes containing 2 ml of H broth. Chloroform was added 90 min after dilution into H broth. Progeny from crosses between T4 and T6 amber mutants were generally plated on P301 and B40. The temperature of incubation was not found to affect the titer in heterologous crosses. Thus few, if any, of the hybrid phage were significantly temperature- or cold-sensitive.

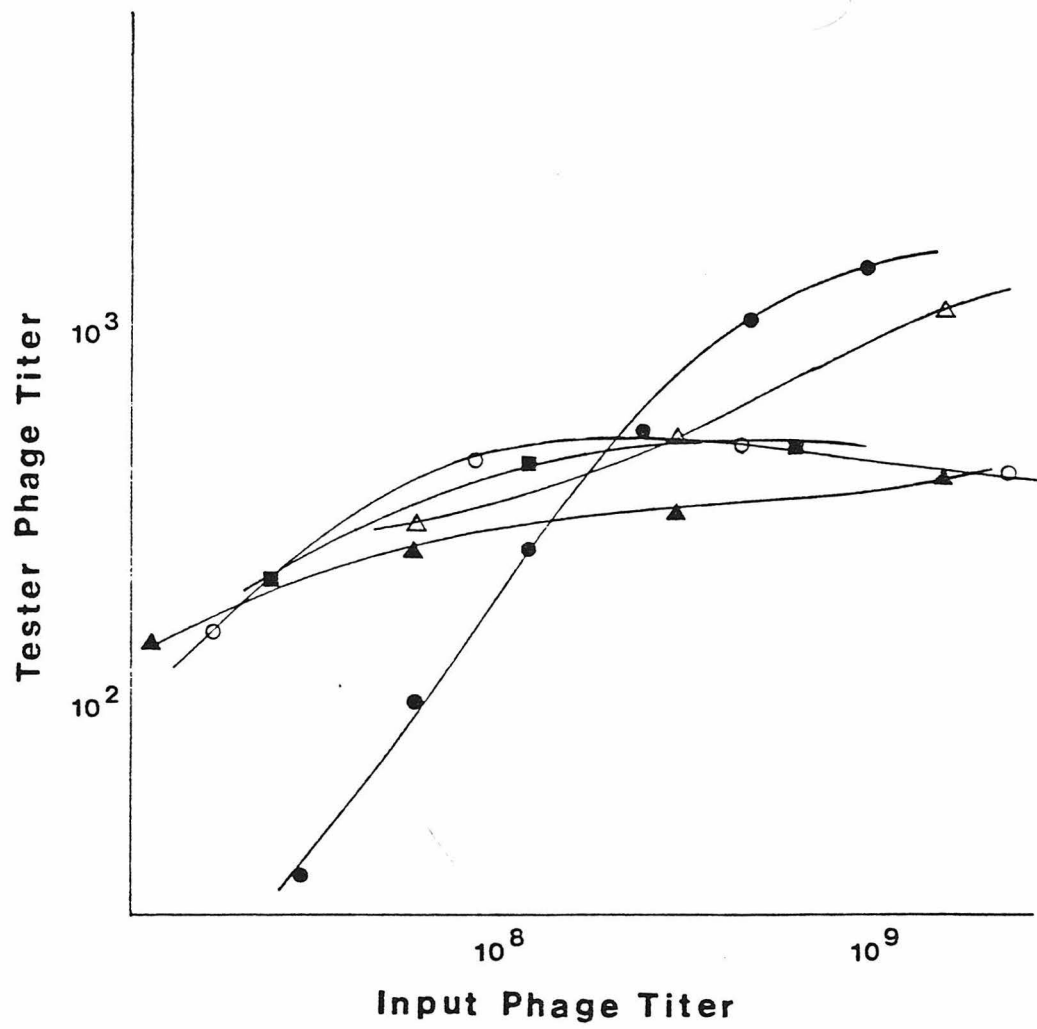
Results

Sixteen RB phage that exhibit T6 host range (Russell, unpublished observations) are susceptible to inactivation by antisera directed against T4 proximal half-fiber antigens. We assessed the quantitative differences in antibody-binding capacity of nine of these 16 phage by a standard antigen-dilution serum blocking assay. Plate stocks of the phage were serially diluted through buffer containing a constant small amount of half-fiber specific antiserum. Wild-type T4 phage were inactivated at an initial rate of 10% per minute in the presence of this diluted antiserum. Incubation at 45°C overnight allowed the unknown phage to irreversibly adsorb all antibodies for which they possessed antigens. The rate with which the remaining antibody inactivated a calibrated "tester" phage aliquot was then measured with a single-point kinetic assay. The unblocked serum is capable of inactivating 99% of the input phage after 46 minutes of incubation in this assay.

Some of the results of the preliminary screening are shown in Figure 1. Most of the phage, e.g., RB3, 6 and 8 are incapable of adsorbing out all the anti-half-fiber antibodies present in the serum and so leave a constant residual activity for T4 inactivation. By contrast RB2 appears to be able to adsorb out almost all of the anti-half-fiber antibody, but at a different concentration of plaque-forming units than the T4 phage. The higher titer of antigen per plaque-forming unit expressed by most of these phage suggests that many of the phage particles may be inactive.

To test this, a smaller subset of the above phage was selected for closer scrutiny. Gradient-purified stocks were made and titers of total particles estimated by absorbance at 265 nm. The optical density of the unknown phage was taken to be identical to that of T4. This assumption is not likely to be a major source of error, since all of the unknown phage are morphologically similar to the T-even phage, which share a common optical density. The serum blocking curves for the purified phage particles are shown in Figure 2, which also contains representative

Figure 1. Survey of proximal half-fiber antigen representation by RB phage. ●, T4 standard phage; X, RB2; ■, RB3; O, RB6; ▲, RB8.



curves for T6 and T2. With standardization by optical density, the curves appear quite uniform at high dilution, though not exactly superimposable with the T4 curve. The failure to superimpose may be a consequence of a lower titer of tail fibers on the unknown phage particles, or a decreased affinity for antibody of the unknown antigens. Figure 2 demonstrates the existence of at least three classes of bacteriophage serologically divergent from T4. T6 falls in the most divergent class of T2 and is indistinguishable from T4.

The degree of antigenic homology of the RB phage with T4 was determined by a more precise measurement of the residual killing power of RB phage-adsorbed serum. Adsorption of the diluted serum was carried out as before, but with a larger aliquot of serum. After adsorption the RB phage were removed by centrifugation and the supernatant antibody titered by kinetic assay of the rate of inactivation of the T4 standard phage. The fraction of killing power remaining is given in Table 1. 98% of the antibody killing power was removed by adsorption to T4, but only 36% by adsorption to RB8.

To determine the relative relationships of the various RB phage species an additivity matrix was constructed. A titer of phage sufficient to adsorb the serum to endpoint was mixed in pairwise fashion with similar titers of each of the other phage. Following overnight incubation the residual antibody was determined exactly as for the individual dilution curves. The results of one of several experiments performed in this way are presented in Table 2. It can be seen from these data that the most divergent phage are antigenically indistinguishable from one another, and from T6; and that the antigens not present on the least divergent phage form a subset of the antigens not present on the most divergent phage.

To assess the genetic localization of the divergent antigen regions we carried out a series of crosses between two representative members of the RB phage family and a set of five amber:temperature-sensitive double mutants of T4.

Figure 2. Serum blocking of gradient-purified phage. a. ●, T4 standard phage; ■, RB58; ▲, RB26; △, RB2; □, RB15; ○, RB8. b. ●, T4 standard phage; ▲, T2 standard phage. c. ●, T4 standard phage; △, RB2; ○, RB15; ▲, RB8; ■, T6.

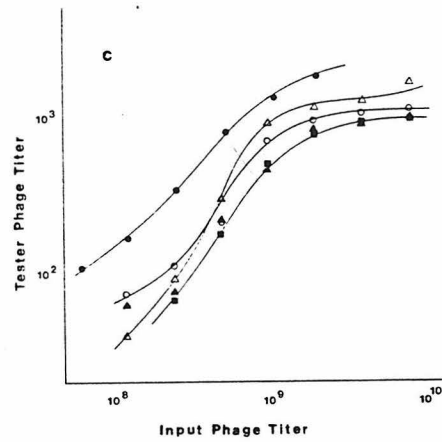
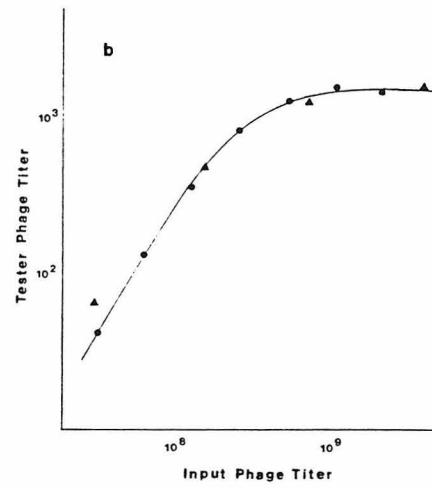
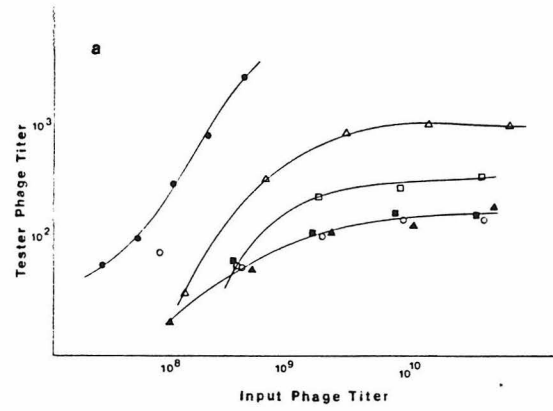


TABLE 1

% Killing Power Remaining after Preadsorption
with Various Phage

T4	0.02
RB2	0.25
RB8	0.64
RB15	0.51

TABLE 2

Additivity of Serum Blocking Power of RB Phage and T6. Titers
of Tester Phage Surviving Incubation with Serum Blocked by
Pairwise Combinations of Phage

	T6	RB2	RB8	RB15	RB26	RB58
T6	892	1900	842	1058	929	901
RB2		1979	2030	2190	1824	1874
8			864	1066	1044 ^a	940
15				970	1080	992
26					837	853
58						840

^aThe RB8, 26 entry in this experiment is in contradiction to the trend discussed in the text. However in two earlier experiments the 8, 26 entry was similar to the 26, 58, and 26, 26 entries.

Figure 3a contains a map of the T4 gene 34 with the various pairs of genetic markers used in these experiments (Seed et al., 1979). Figure 3b contains a diagram of the desired outcome of one of these crosses. Each RB phage was crossed with the complete set of T4 amber:temperature-sensitive double mutants.

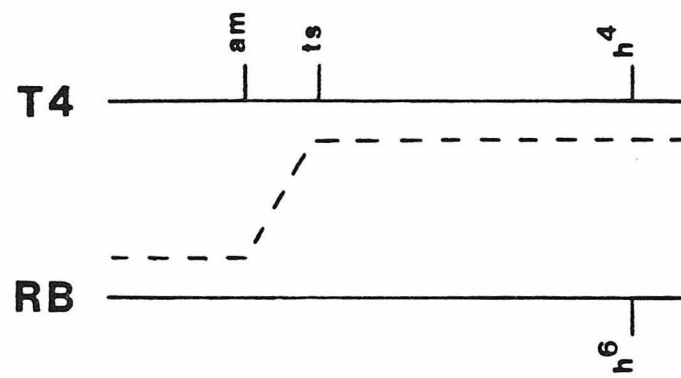
Since the RB phage have T6-like host range (h^6), the fraction of $am^+ : h^4$ progeny which are still temperature-sensitive should represent recombinant phage bearing a distal portion of the T4 gene 34, and a proximal portion of the RB gene. The crossover should have occurred between the two T4 markers (Figure 3b). Phage hybrids between T4 and T6 are not themselves significantly temperature- or cold-sensitive (Materials and Methods), so it is unlikely that the temperature-sensitive phenotype arose spontaneously in these crosses.

Not all pairwise crosses yielded temperature-sensitive recombinants; those that did are listed in Figure 3. Recombinant phage isolated from successful crosses were purified by cesium chloride gradient centrifugation and their relative antigen content assessed by the dilution-inactivation protocol described previously. The results of this experiment are shown in Figure 4. All of the recombinant phage tested exhibit the complete spectrum of T4 antigens.

A reciprocal cross selecting T4 proximal termini and RB8 distal termini was performed between $tsB3:amNG85$ phage and RB8. It was harder to obtain $am:ts^+$ recombinants from this cross, the frequency being about tenfold lower than for the reciprocal $am^+ : ts$ recombinants. However a recombinant of the desired phenotype was isolated and was shown to have serum adsorption characteristics identical to those of RB8 (Figure 5).

The data above support the interpretation that the divergent region of RB2 falls somewhere on the distal side of $tsA44$ and that the divergent region of RB8 falls somewhere on the distal side of $tsB22$. The failure of crosses with the two most distal pairs of markers to yield recombinants with RB8 suggests that the

Figure 3. a. Map locations of genetic markers used in T4-RB phage crosses. Markers are represented in the physical (amber fragment) map of gene 34 (Seed et al., 1979). amNG85 is near the promoter-proximal terminus of the gene, and tsA44 near the distal terminus. The locations of the amber and temperature-sensitive mutations used to select hybrid recombinants between RB and T4 phage are shown at the top of this figure. The RB phages capable of recombining with T4 in the interval between amber and temperature-sensitive mutations are listed separately beneath each pair of mutations. b. Schematic representation of T4-RB phage recombinants. The host range region, designated h^6 or h^4 (and specified by gene 37 in T4) falls 3' to the gene 34 coding sequence. The crossover event diagrammed gives rise to ts:am⁺ phage bearing T4 host range.



[34]

[37]

Figure 4. Proximal half-fiber antigen characteristics of representative hybrid phage bearing T4 ts mutations. ●, T4 standard phage; □, RB8/tsA20; ▲, RB8/tsB22; ○, RB8/tsB3; □, RB2/tsB33; Δ, RB2/tsA44.

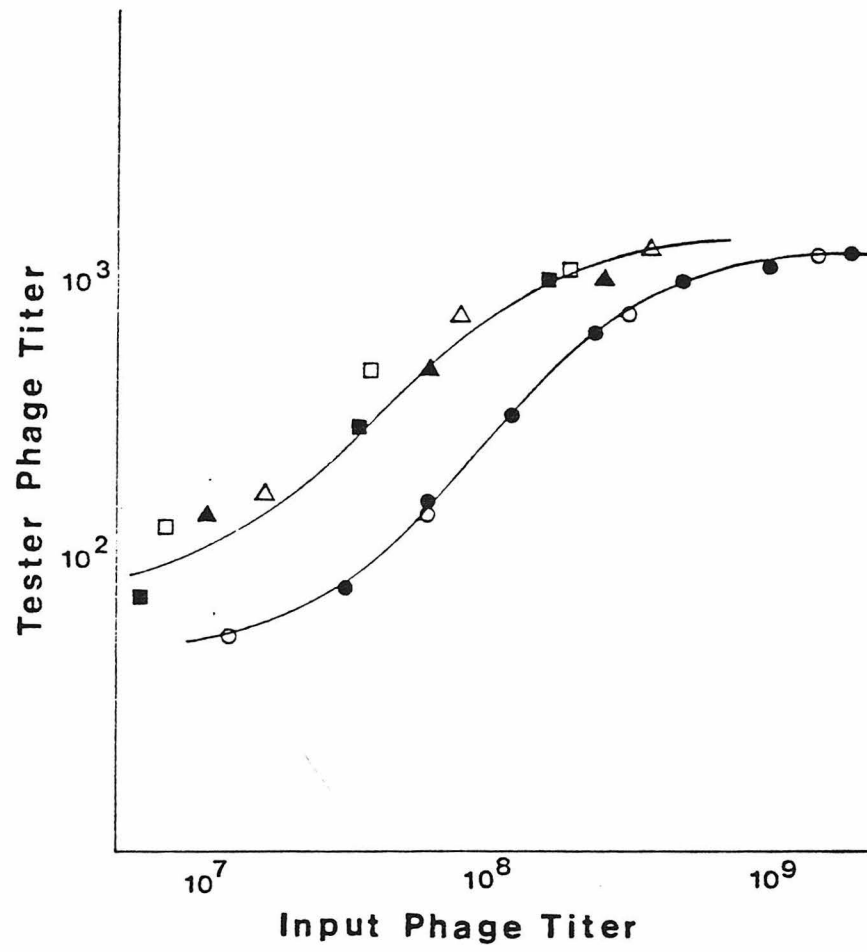
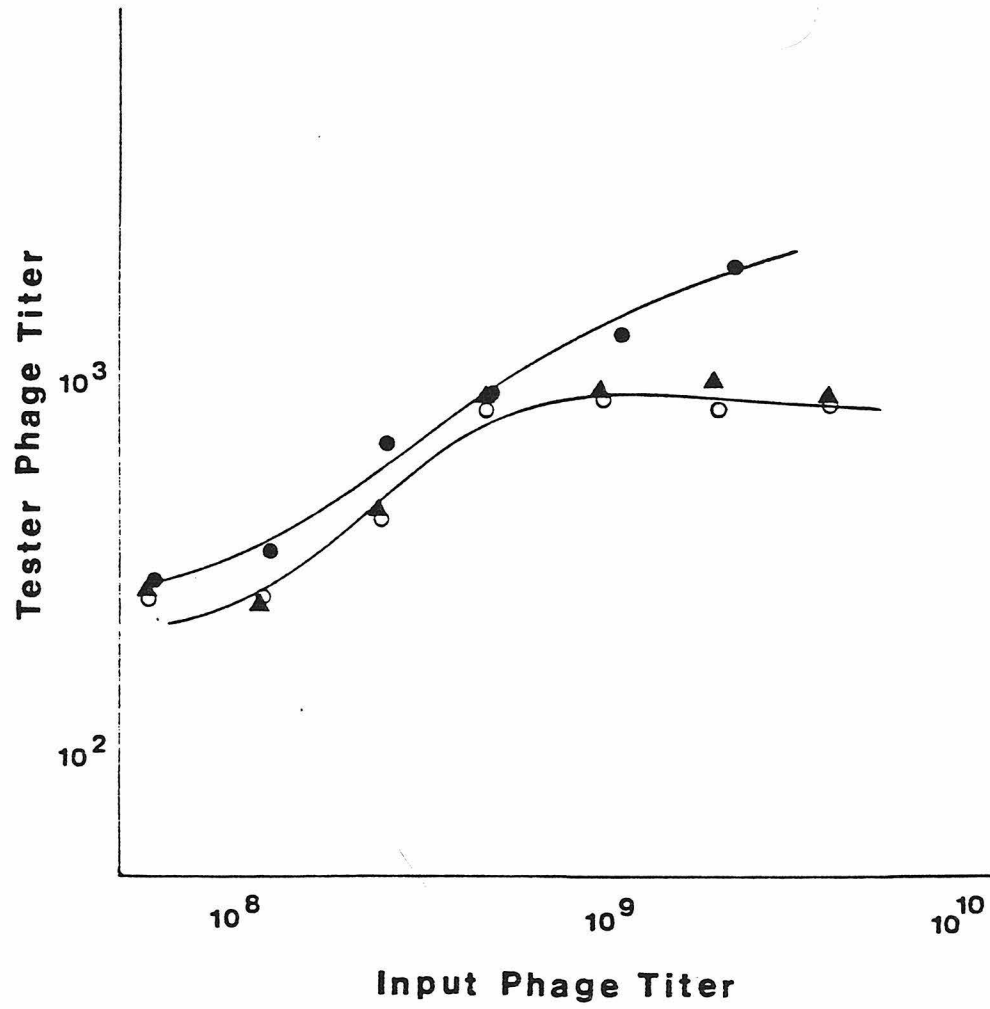


Figure 5. Serum blocking of a hybrid phage bearing T4 am mutation. ●, T4 standard phage; ○, RB8; ▲, RB8/amNG85.



divergent region of RB8 includes sequences nonhomologous to T4 in the regions of the markers tsB33 and tsA44.

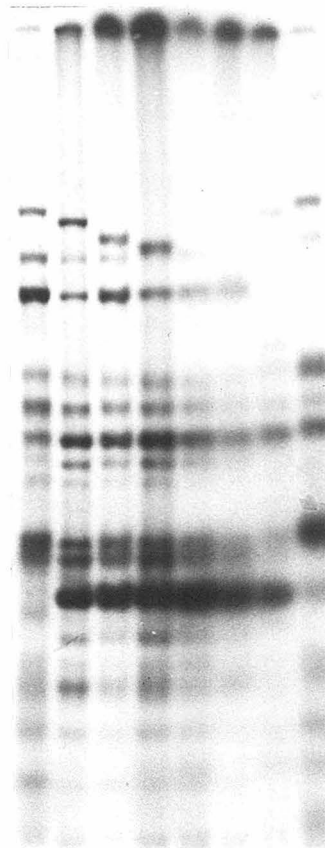
Another explanation is suggested, however, by the rapid rise in recombination frequency observed at the promoter-distal terminus of T4 gene 34 (Beckendorf and Wilson, 1972). If recombination between RB8 and T4 were to reflect only map distances in RB8, and if RB8, like T2 and T6, displayed constant recombination frequency throughout the gene, then recombination between T4 and RB8 at the distal end of gene 34 would take place with reduced frequency relative to that in the remainder of the gene.

To investigate this possibility we constructed hybrid phage from T4 and T6. T6 phage bearing amber mutations in gene 34, originally isolated by Russell (unpublished results), were ordered genetically. The most distal mutation, T6 am90, was then physically mapped by the technique of amber fragment sizing. Defective cell extracts labeled with ^{14}C amino acids were prepared on a nonsuppressing host and electrophoresed through a 7.5% polyacrylamide slab gel. Defective extracts of various T4 am mutant phage-infected cells were also prepared for comparison of amber fragment lengths. Figure 6 shows that the amber fragment generated by T6 am90 has a very slightly higher molecular weight than that generated by T4 amA455. The T6 am90 mutation thus lies just distal to the T4 amA455 mutation.

With such a map assignment, the T6 am90 site should fall in the region of nonhomology between T6 and T4. We predicted, therefore, that recombination between T6 and T4 would be inhibited throughout a region spanning the T4 amA455 and an unknown site proximal to the T4 amNG506 and distal to tsB22. To test this hypothesis it was necessary to devise a simple method for obtaining recombinants between T6 and T4.

T6 and T4 exhibit a phenomenon of mutual growth inhibition characterized by Russell (1967) as a "depressor effect": burst sizes following mixed infections

Figure 6. Determination of the length of T5 am90 and T4 gene 34 am nonsense-terminated polypeptides. Extracts labeled with ^{14}C amino acids were electrophoresed through polyacrylamide gels containing SDS. Lane a, T4 amB178 (gene 23); b, T4 amA455(34); c, T4 amB258(34); d, T4 amNG144(34); e, T4 amE4; f, T4 amNG506; g, T6 am90; h, T6 am126(37).



a b c d e f g h

are so severely reduced that detection of recombinant phage is impractical for closely linked markers. This difficulty was overcome by employing a novel cross protocol without cyanide (Materials and Methods) which made it possible to sample infected cells at very high density. Using this procedure (which stimulates recombination in T4 about twofold) we were able to obtain a linkage map between T6 am90 and various am mutations in T4. Figure 7 shows the results of that mapping. Recombinants could be obtained in the region of nonhomology, though few crossovers apparently took place in the region between T4 amNG144 and T4 amNG506.

Stocks of selected recombinant progeny from several of these crosses were grown and assayed for antigen titer in an endpoint blocking experiment. The results, presented in Figure 8, demonstrate that the observed recombinants had serum blocking characteristics intermediate between those of T6 and those of T4. The most distal T4 am mutation site giving rise to progeny bearing T4-like antigen was amN58. The proximal end of the region of nonhomology thus falls somewhere distal to amN58. No T4-like phage were observed among progeny from the cross to amNG506. These results support the interpretation that the nonhomology between T6 and T4 includes, and extends distally beyond, the site of amNG506. That the nonhomology between RB8 and T4 includes and extends beyond the NG506 site was suggested by the results of Figures 3, 4 and 5.

Antigenic nonhomology between T4 and the RB phage was physically mapped by electron microscopy. Anti-half-fiber serum was adsorbed to endpoint with RB2, 8, or 15, or with T4. The adsorbed sera were then incubated with T4 phage adsorbed to carbon-coated grids, and the sites of antibody binding visualized according to the protocol of Yanagida and Ahmad-Zadeh (1970). Figure 9 shows that the region of nonhomology with the RB phages lies at the distal end of the proximal half-fiber.

The reciprocal experiment, in which anti-half-fiber antibodies were visualized on various RB phage directly, was also performed. Figure 10 shows that in all cases

Figure 7. Genetic distances between T6 am90 and T4 gene 34 am mutations. Upper line, the physical (amber fragment) map of am90 and gene 34 am mutations. Lower lines give map distances between am90 and T4 gene 34 am mutations. Map distances between various T4 am sites (Beckendorf and Wilson, 1972) are also included.

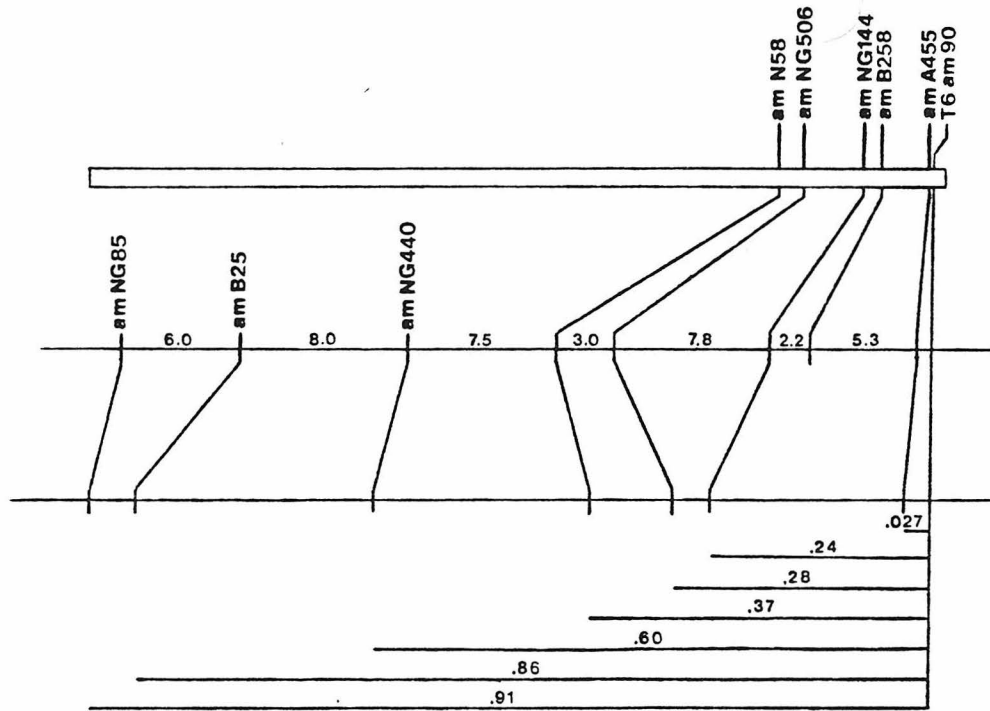
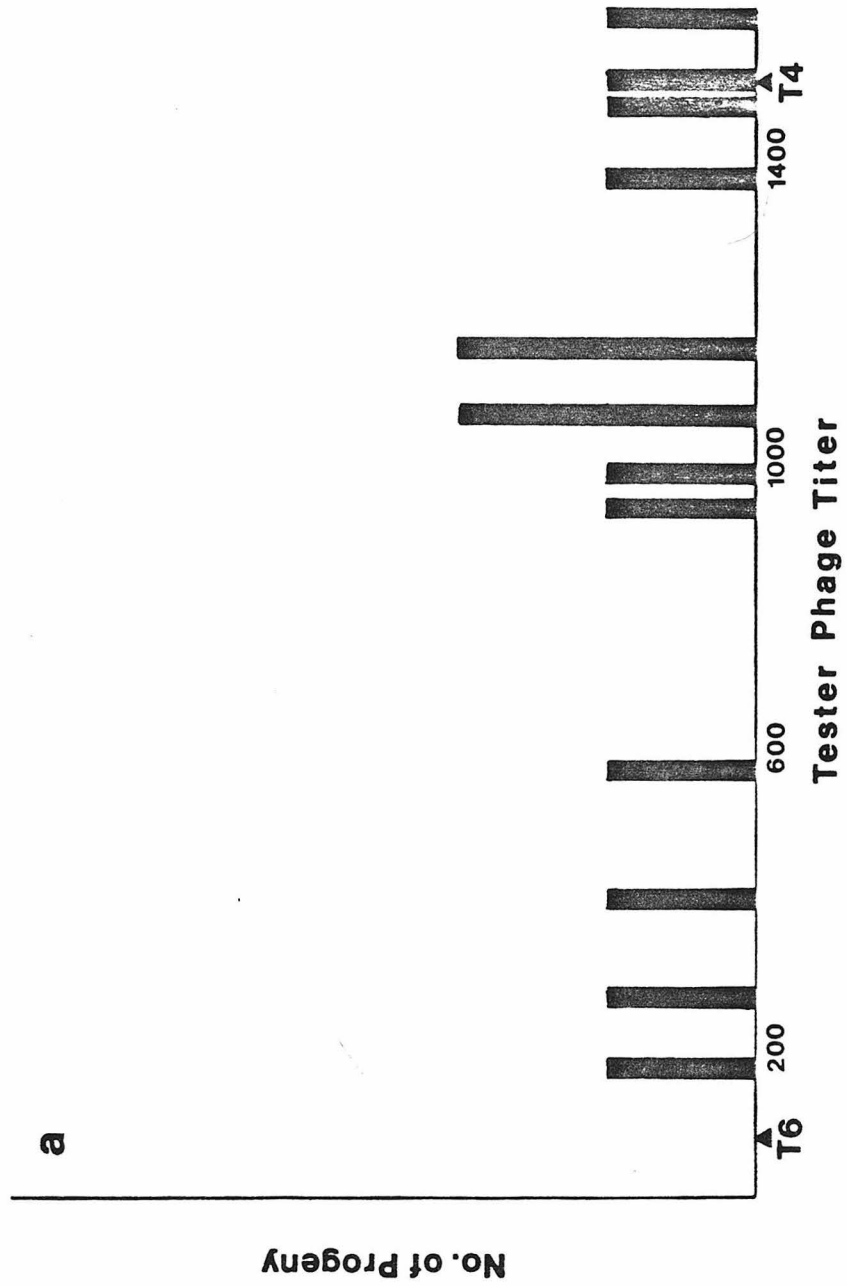
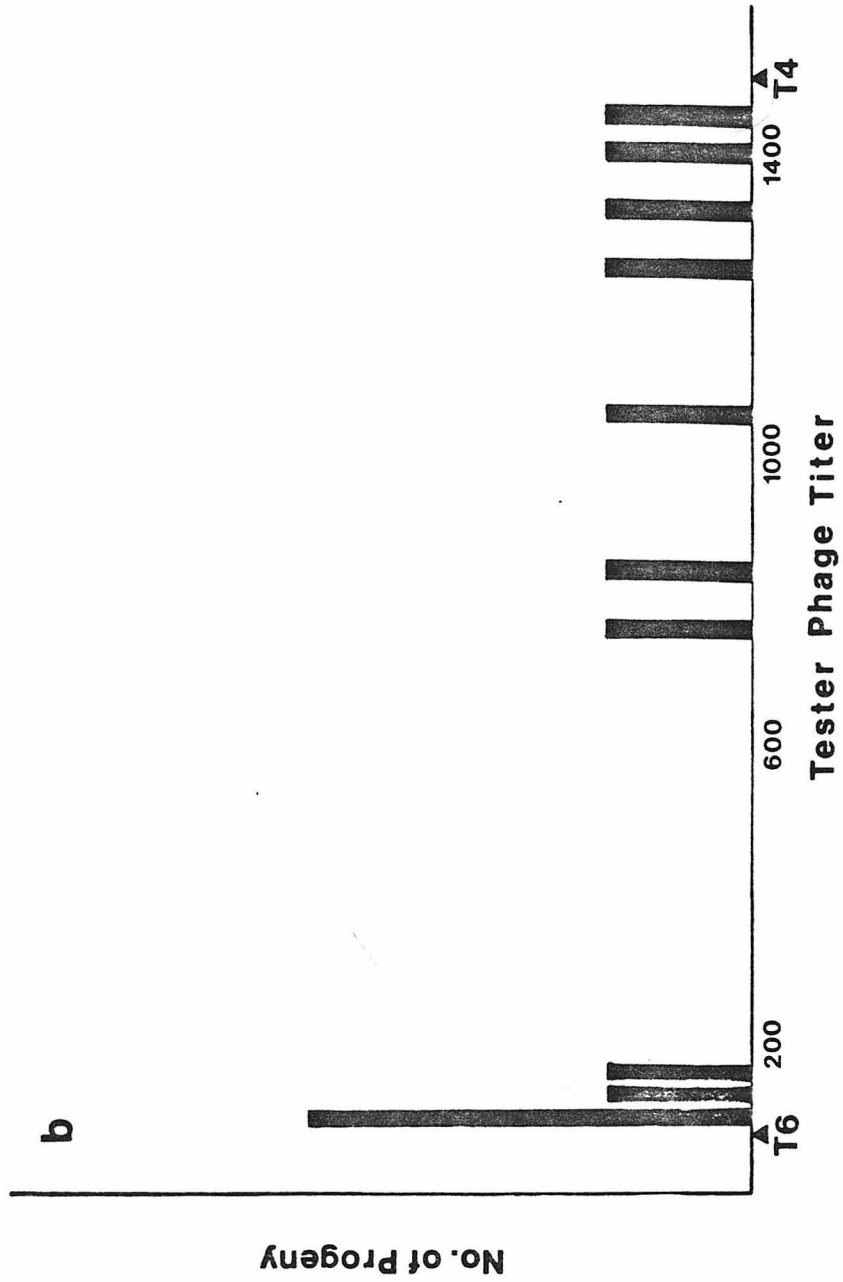
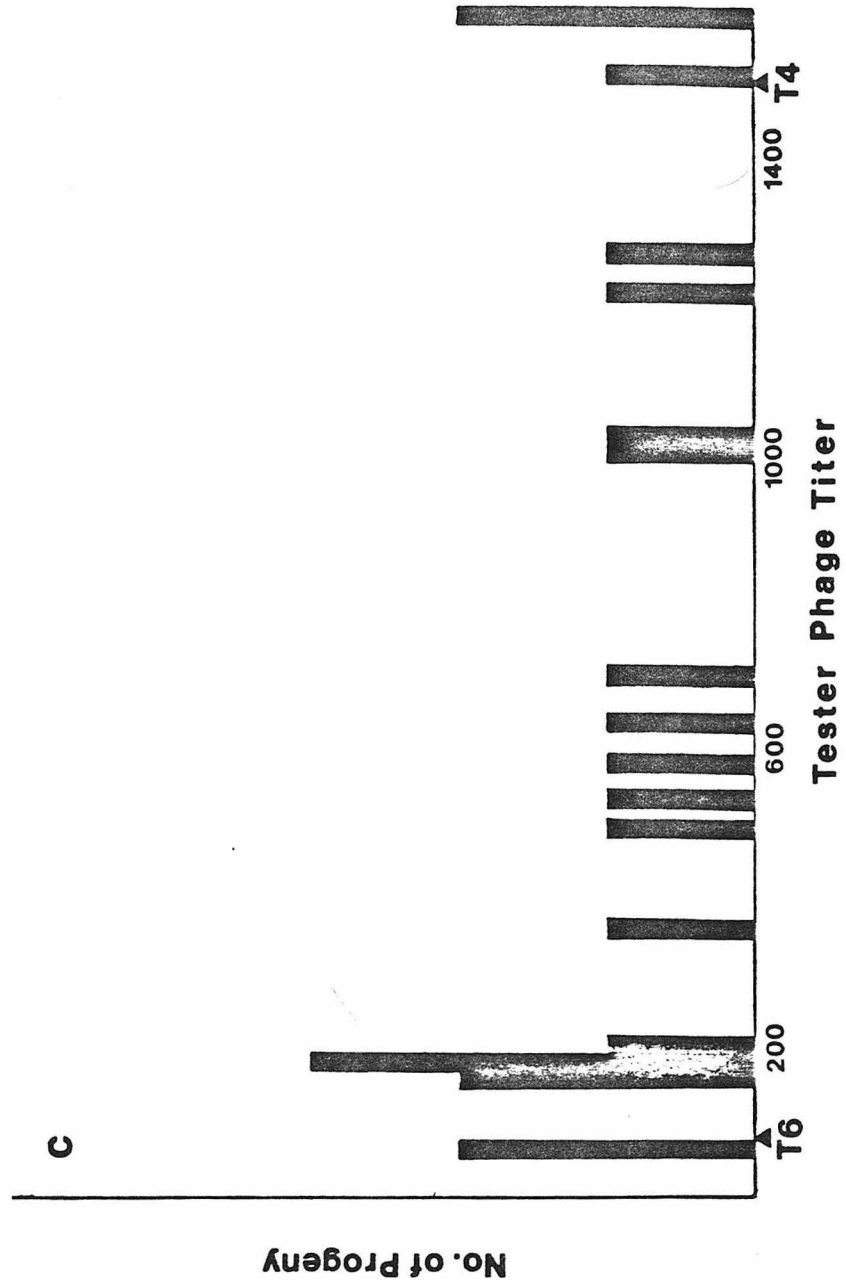
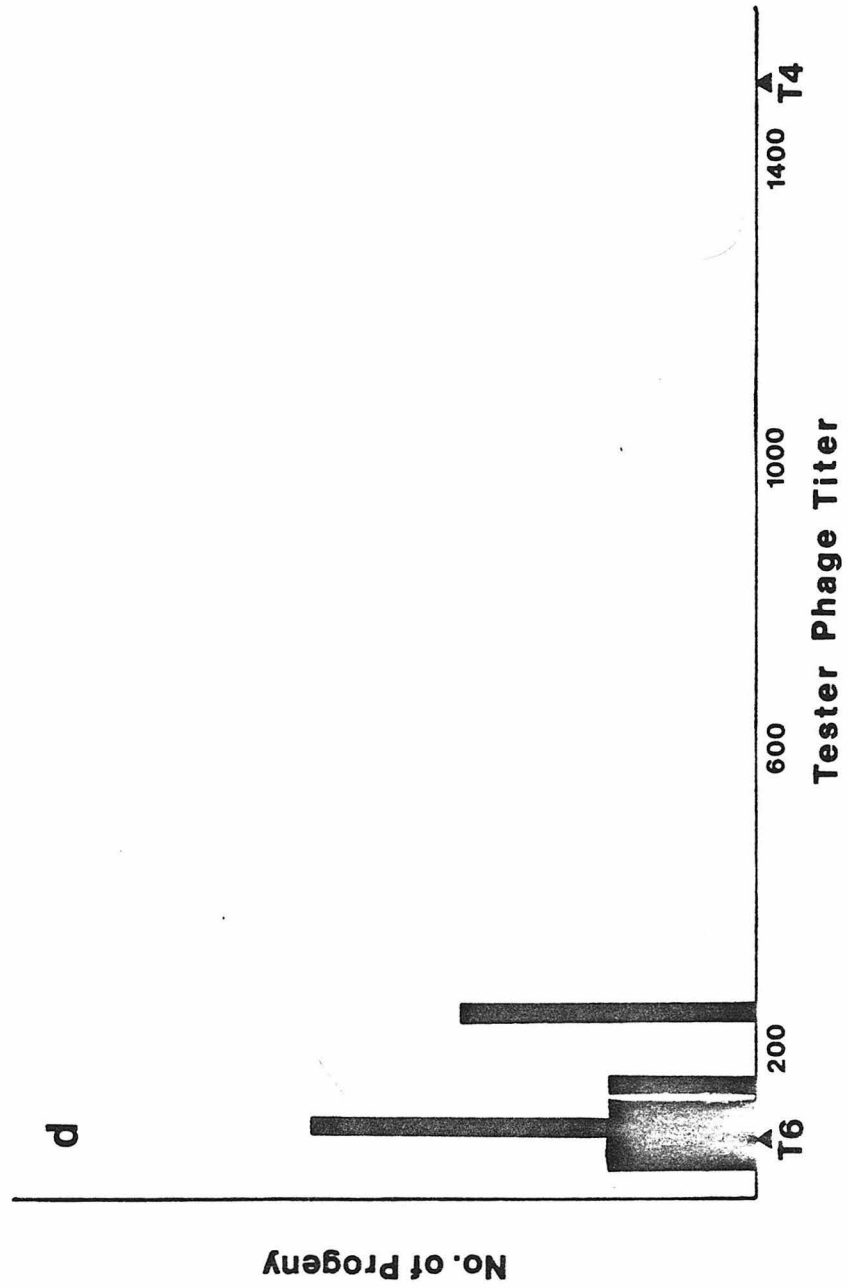


Figure 8. Antigen representation among progeny of crosses between T6 am90 phage and T4 gene 34 am phage. Each panel is a histogram of the endpoing blocking levels of selected am⁺ progeny from each cross. Panel a, progeny of T6 am90 x T4 amB25; b, progeny of am90 x amNG440; c, progeny of am90 x amN58; d, progeny of am90 x amNG506; e, progeny of am90 x amNG144.









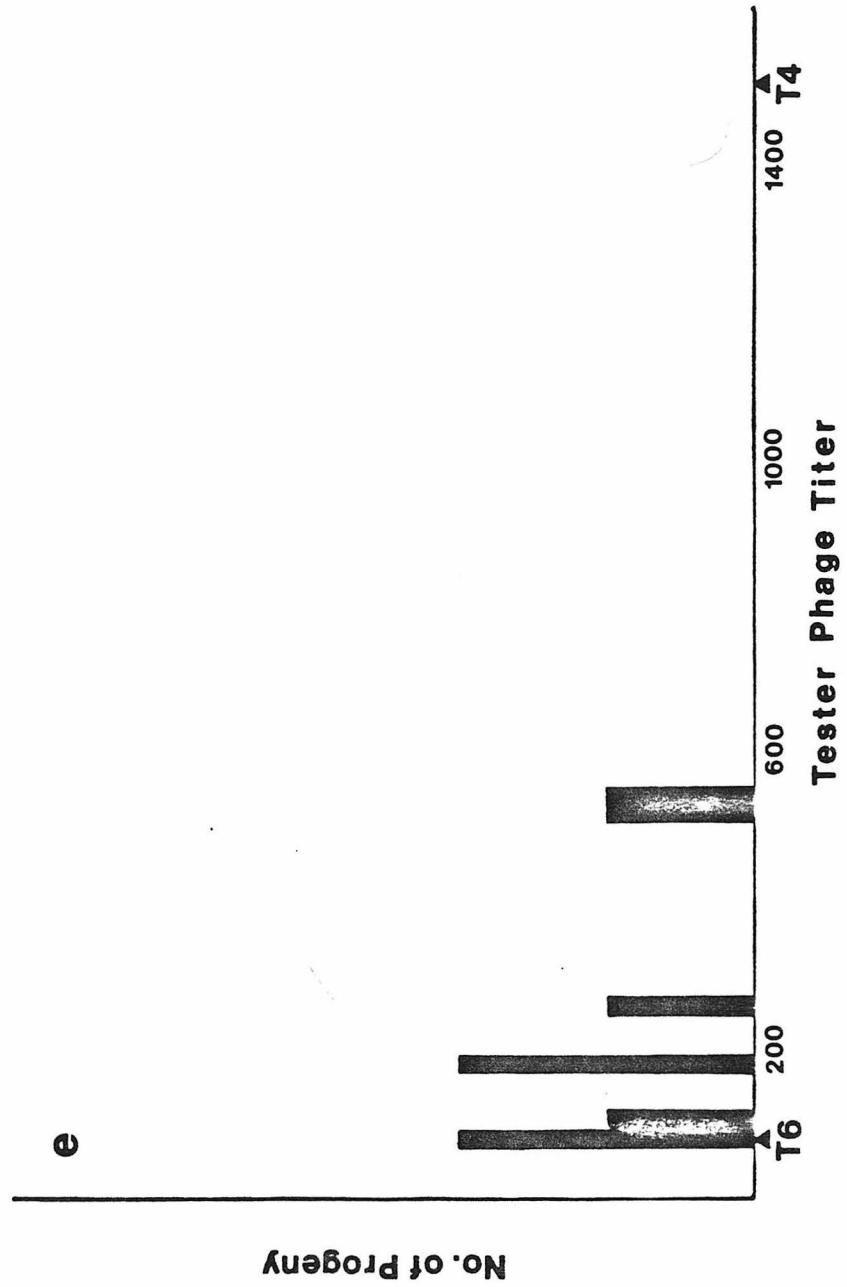
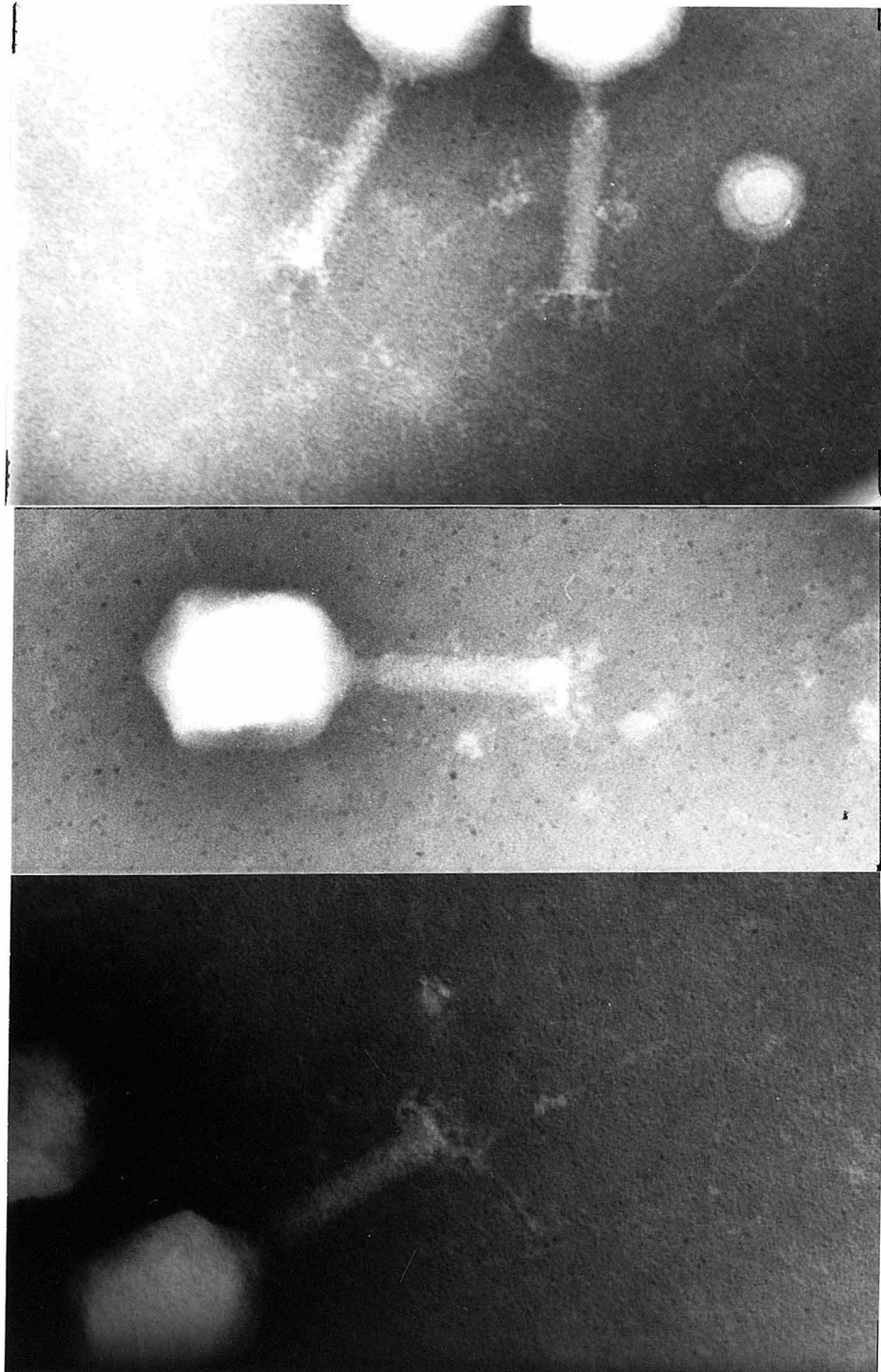


Figure 9. Electron microscopic visualization of regions of antigenic nonhomology between T4 and RB phages. Anti-T4 proximal half-fiber serum was adsorbed to endpoint with various RB phages and then incubated with T4 phage adsorbed to carbon-coated grids. Antibodies appear at the distal end of the proximal half-fiber. Panel a, T4 incubated with RB2-adsorbed serum. Panel b, T4 incubated with RB15-adsorbed serum. Panel c, T4 incubated with RB8-adsorbed serum.





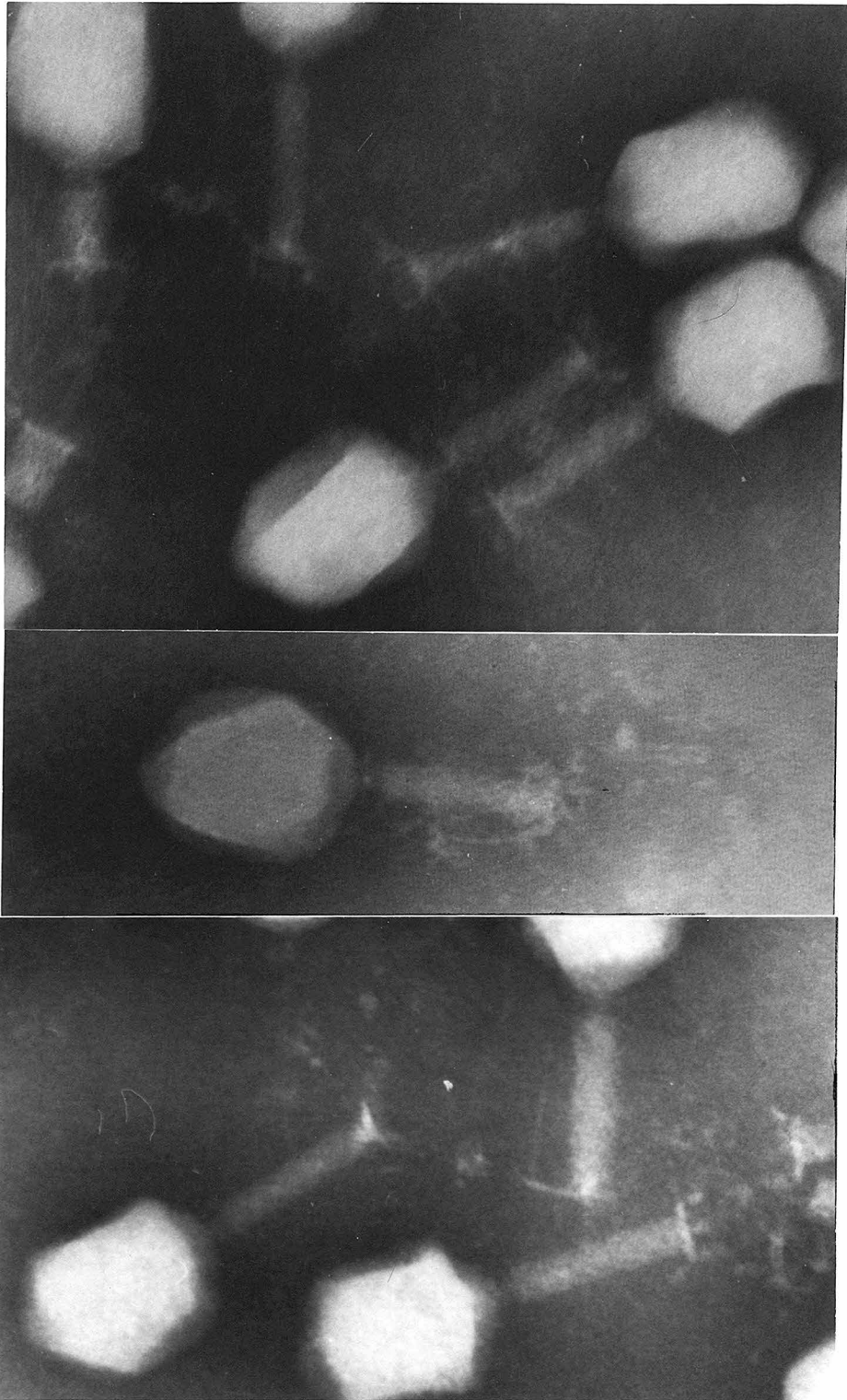
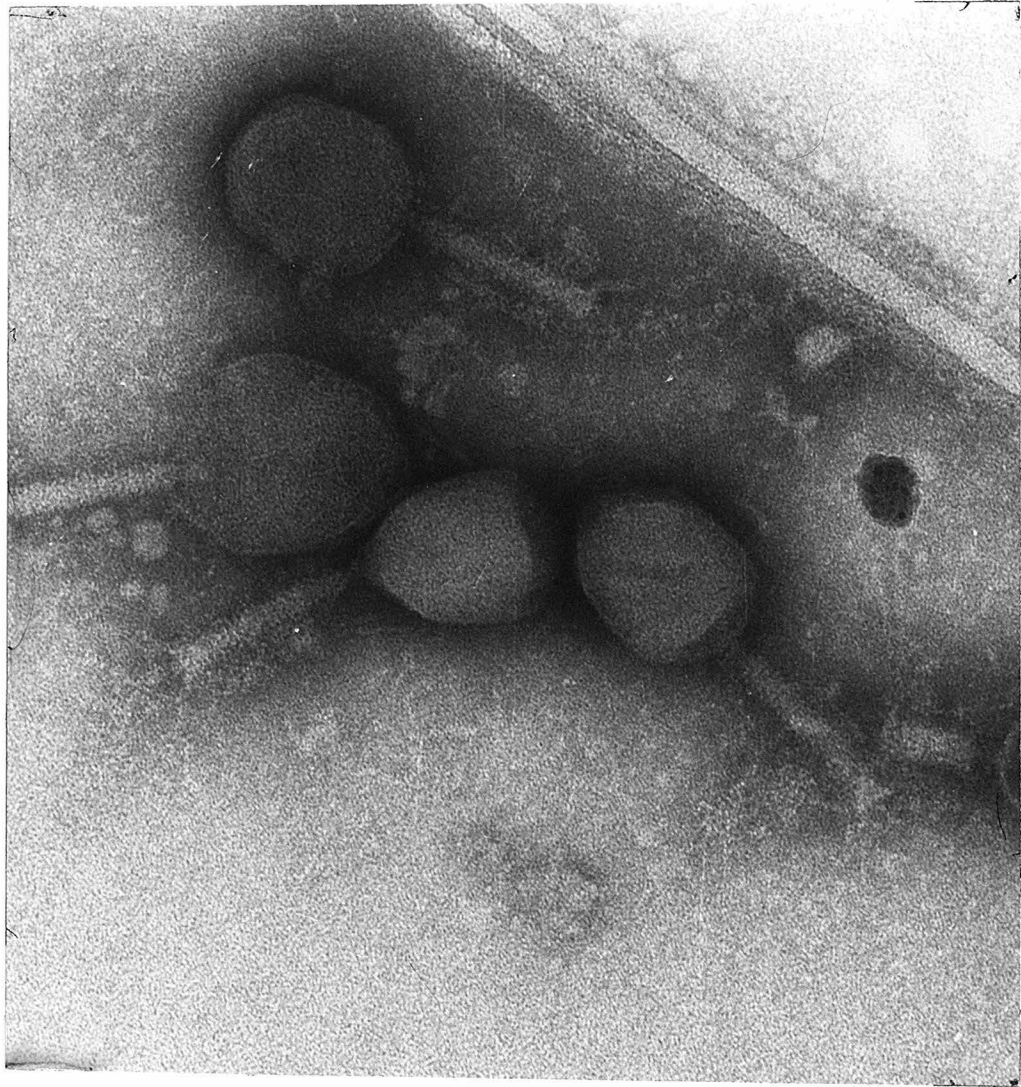
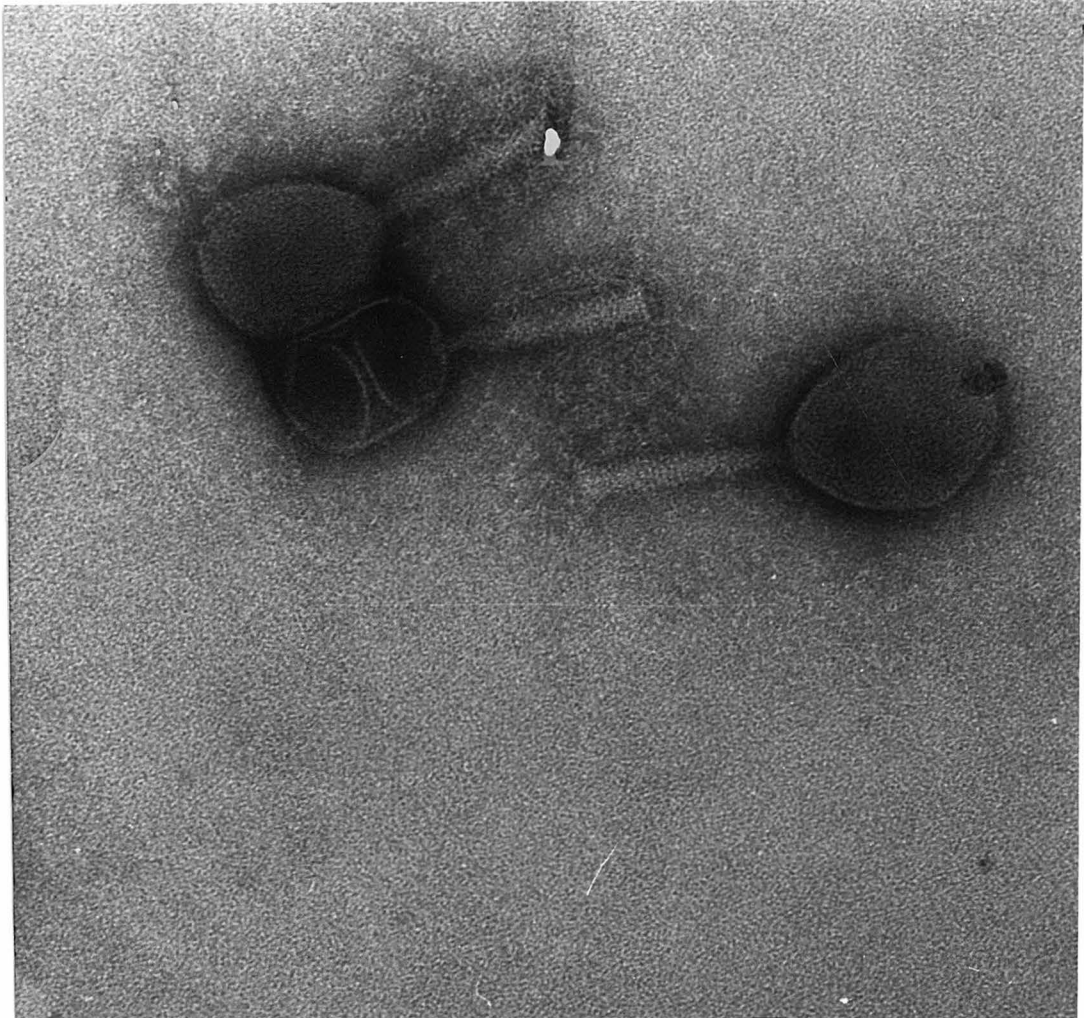
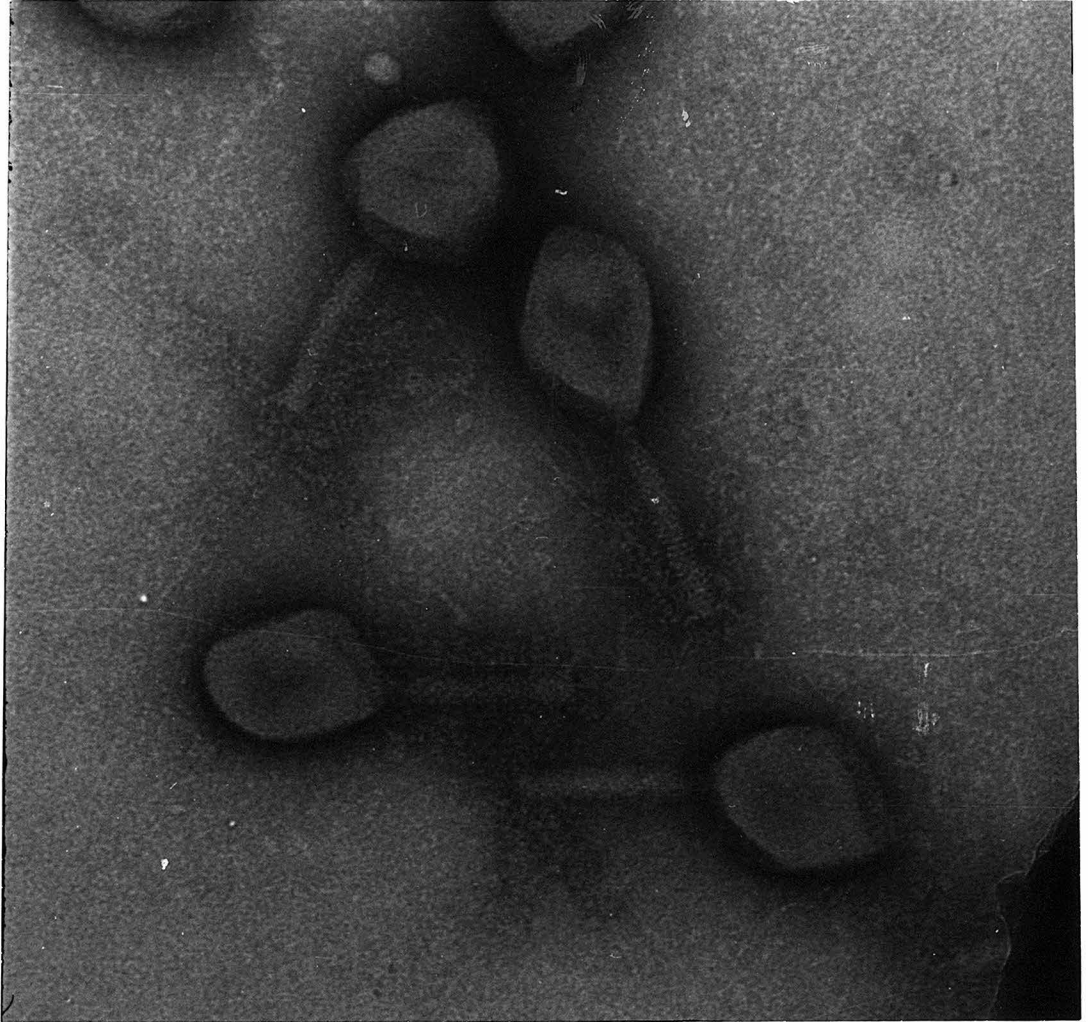


Figure 10. Visualization of anti-T4 proximal half-fiber antibodies directly adsorbed to RB phages. Antibodies bind throughout the half-fiber. Panel a, RB8; b, RB15; c, T4.







the antibody binds to the entire visible half-fiber. Thus some antigens remain conserved within regions in which other antigens have diverged. Antigens found to be conserved within the divergent region may also be repeated elsewhere in the half-fiber.

Discussion

The results of this survey have demonstrated the existence of a family of bacteriophage bearing T6 host range and partial homology to T4 proximal half-fiber antigens. The nonhomologous antigens of the least divergent phage form a subset of the nonhomologous antigens of the most divergent phage. Genetic localization of the nonhomology places that of the least divergent at the very distal end of the gene, and that of the most divergent over a more proximal region encompassing the least divergent.

Table 1 has shown that 25% of the T4-killing power remains in sera pre-adsorbed with the least divergent phage. The genetic and serological data of Table 2 and Figure 4 show that all of the divergent antigen of this phage is located somewhere distal to gene 34 tsA44. This ts mutation is located about 3% of the gene length from the distal terminus (Seed et al., 1979). Thus the antigenic determinants recognized by the neutralizing antibodies in the T4 proximal half-fiber antiserum are not homogeneously distributed with respect to length along the half-fiber. Half-fiber antigens at the carboxyl terminus of the polypeptide chain contribute far more to the susceptibility of the phage than do antigens elsewhere. Two explanations for this are likely: (1) that carboxy-terminal antigens are unusually effective immunogens, or (2) that such antigens represent phage sites of particular vulnerability to antibodies. In view of the antigenic divergence in this region, and the potent immunogenicity of T-even phage tail fibers in general, it may be interesting to examine the titers of anti-phage antibodies in the phage natural habitat, for example mammalian intestinal lumen.

The gene 34 products of phages RB2 and RB15 have diverged to lesser extents than those of RB8 or T6, but in both cases the divergence falls within a region extending to the distal terminus of the gene. The minimal divergence tree connecting T4 and T6 would thus place RB2 and RB15 intermediate between T4 and T6. If the divergent regions of RB2 and RB15 are homologous to T6, then either T4 or T6 may be the more ancestral phage. If, however, the divergent regions of RB2 or RB15 are not homologous to T6, T4 should be the ancestral phage. We cannot decide between these alternatives at present.

Since the region of divergence between T4 and T6 is coincident with the region of anomalous T4 recombination, we might also propose that the divergence is generated by errors arising during the course of the unusually frequent recombination events taking place at the distal terminus of an ancestral T6-like phage exhibiting the recombination anomaly. Such a simple picture is complicated by several factors.

Both DNA heteroduplex mapping (Kim and Davidson, 1974), and hybridization studies (Cowie et al., 1971) have shown that T2 and T6 are more closely related than T2 and T4 or T6 and T4. Neither T2 nor T6 exhibit the anomalously high recombination observed in T4. T4 excludes T2 and T6, as well as all RB phage examined in this survey (Russell, 1967). T2 and T6 exclude each other only weakly, and are capable of functional interchange of distal half-fiber gene products under circumstances in which the analogous T4 products cannot be exchanged with those of either T2 or T6 (Russell, 1967). So there is a sizeable body of evidence that T2 and T6 are closely related evolutionarily, and probably more primitive than T4, since T4 excludes them both. The puzzle cannot be resolved by merely ordering the gene 34 divergence so that T6 gives rise to RB15, RB2, and T4 successively, because T2 is then left close to T4; but in many respects such a history is the most attractive of the simple alternatives open to us. In such a scheme a recombination event would transfer the T4 gene 34 to T2.

References

- Beckendorf, S. K. and Wilson, J. H. (1972). A recombination gradient in bacteriophage T4 gene 34. *Virology* **50**, 315-321.
- Cowie, D. B., Avery, R. J. and Champe, S. P. (1971). DNA homology among the T-even bacteriophages. *Virology* **45**, 30-37.
- Crawford, J. T. and Goldberg, E. B. (1977). The effect of baseplate mutations on the requirement for tail-fiber binding for irreversible adsorption of bacteriophage T4. *J. Mol. Biol.* **111**, 305-313.
- Edgar, R. S., Denhardt, G. H. and Epstein, R. H. (1964). A comparative genetic study of conditional lethal mutations of bacteriophage T4D. *Genetics* **49**, 635.
- Edgar, R. S. and Lielausis, I. (1965). Serological studies with mutants of phage T4D defective in genes determining tail fiber structure. *Genetics* **52**, 1187-1200.
- Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, G. H. and Lielausis, A. (1963). Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 375-394.
- Kim, J.-S. and Davidson, N. (1974). Electron microscope heteroduplex study of sequence relations of T2, T4, and T6 bacteriophage DNAs. *Virology* **57**, 93-111.
- King, J. (1968). Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231-262.
- Russell, R. L. (1967). Speciation among the T-even phages. Ph.D. Thesis, California Institute of Technology, Pasadena, California.
- Seed, B., Del Zoppo, G. J. and Wood, W. B. (1979). Chapter 4 of this thesis.
- Steinberg, C. M. and Edgar, R. S. (1962). A critical test of a current theory of genetic recombination in bacteriophage. *Genetics* **47**, 187-208.
- Ward, S., Luftig, R. B., Wilson, J. H., Eddleman, H., Lyle, H. and Wood, W. B. (1970). Assembly of bacteriophage T4 tail fibers. II. Isolation and characterization of tail fiber precursors. *J. Mol. Biol.* **54**, 15-31.

- Wilson, J. H. and Kells, S. (1972). Bacteriophage T4 transfer RNA. I. Isolation and characterization of two phage-coded nonsense suppressors. *J. Mol. Biol.* **69**, 39-56.
- Wilson, J. H., Luftig, R. B. and Wood, W. B. (1970). Interaction of bacteriophage T4 tail fiber components with a lipopolysaccharide fraction from Escherichia coli. *J. Mol. Biol.* **51**, 423-434.
- Yamamoto, M. and Uchida, H. (1975). Organization and function of the tail of bacteriophage T4. II. Structural control of the tail contraction. *J. Mol. Biol.* **92**, 207-223.

CHAPTER 3

**Investigation of the Interaction Between Bacteriophage T4 Proximal and
Distal Half Tail Fibers Using a Novel Method of Pseudorevertant Analysis**

Division of Biology
California Institute of Technology
Pasadena, California 91125

Brian Seed, Jeffrey Slankard, Gregory J. Del Zoppo,
M. Patricia Conley and William B. Wood

Introduction

The proximal half tail fiber of bacteriophage T4 is a rod-like dimer of the polypeptide product of gene 34. The distal half-fiber is composed of two molecules each of the products of genes 36 and 37, and one molecule of the product of gene 35. Two nonstructural proteins are involved in the formation of half-fiber structures: the gene 38 product, required exclusively for the formation of the distal half-fiber, and the gene 57 product, required for the formation of both half-fibers. Figure 1 is a schematic diagram summarizing what is known about the assembly pathway for the formation of whole T4 tail fibers from these various gene products.

The isolation and analysis of second site revertants (pseudorevertants) of characterized bacteriophage mutations has been shown to be a useful genetic technique for the analysis of assembly mechanisms of phage T4 (Bishop and Wood, 1976). In this study we genetically define a site of interaction between the proximal and distal half-fibers using a novel method for the analysis of pseudorevertants.

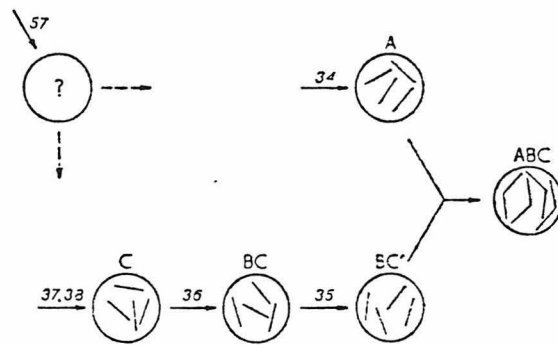
In principle, second site mutations compensating for a characterized genetic lesion (e.g., a ts or am mutation) can be identified and mapped, though phenotypically unselectable, by an appropriate choice of ancillary mapping markers. A simple application of this principle to the isolation of pseudorevertants of a gene 34 ts mutation is detailed in the text.

In outline, ts:am double mutants were constructed in which the am mutation was chosen from any of a number of structural or assembly protein genes participating in the tail fiber pathway. Revertants to ts⁺ were selected from these double mutants at nonpermissive temperature. Rapid gene assignments were then made by complementation testing.

Revertants in which the site of reversion falls in the same gene as the preselected am mutation should not express the revertant phenotype in a nonsuppressing host. ComPLEMENTING phage can supply the missing gene product, but not the revertant

Figure 1. The pathway of T4 tail fiber assembly. (Taken from Ward et al., 1970.)

Numbered arrows indicate the genes controlling the various steps; circled arrows represent the electron microscope morphology of the structures which accumulate when the pathway is blocked by mutations in the succeeding gene. The antigens associated with each structure are shown by the letters above the circles.



function. Thus at nonpermissive temperature the am mutant revertant phage should not be able to compensate for the ts mutant gene product.

The genetic locations of interesting revertant mutations can then be defined more precisely by backcrossing to the original ts mutant phage and scoring phenotypically ts⁺:am⁺ progeny. More sophisticated mapping experiments can be carried out as long as the mapping strains always carry the original ts mutation.

Although the procedures above entail more effort than those in which the revertant mutation is chosen to bear a selectable phenotype (Jarvik and Botstein, 1975; Bishop and Wood, 1976), the choice of phenotype can strongly bias the spectrum of revertants. Such a bias may lead to inapt conclusions or (as in the case below) too few revertants to be practical.

Materials and Methods

Phage and bacterial strains. Phage strains derived from the wild-type T4D were obtained from the Caltech collection and have been described elsewhere (Epstein et al., 1963; Edgar and Lielausis, 1965). Other strains were constructed by genetic crosses and are described in the text.

Escherichia coli. Strains CR63 and B40 Su_I⁺ were the permissive hosts for amber mutant phage. Strains B/5, S/6/5, and P301 (B40 Su⁻ arg⁻) were the non-permissive hosts. The permissive and restrictive temperatures for ts mutant phage were 25° and 42°C, respectively, on plates, and 30° and 39.5°C in liquid culture, unless otherwise noted.

Media and buffers. H broth for liquid culture of bacteria and bacteriophage, and EHA top and bottom agars for plating were prepared as described by Steinberg and Edgar (1962). Dilution medium was prepared as described by King (1968).

Genetic crosses. Phage crosses were carried out by the procedure of Steinberg and Edgar (1962) as modified by Wilson and Kells (1972).

Other methods. Defective extracts for in vitro complementation were prepared as described by Edgar and Wood (1966). Serum blocking assays were performed according to the endpoint protocol of Ward et al. (1970). Rapid complementation assays for screening mutants were performed by mixing one drop each of the phage to be tested (titer 2.5×10^6 /ml) in a well of a plastic tray and adding 1 ml of top agar which had been freshly mixed with plating bacteria (5 ml/100 ml top agar). Positive and negative controls were performed for each phage in the same tray.

Results

The temperature-sensitive mutation tsB45 is located near the promoter-distal terminus of gene 34, as shown in Figure 2. tsB45 shares a number of phenotypic features with other ts mutations in gene 34. At nonpermissive temperature wild-type amounts of gene 34 polypeptide are synthesized, but formation of the characteristic antigens of the proximal half-fiber (the A antigens) is partially inhibited, and the phage particles produced have few attached fibers (Seed et al., 1979). tsB45 phage grown at permissive temperature adsorb normally at nonpermissive temperature and are not thermolabile (Seed et al., 1979). One distinguishing feature of tsB45 phage, however, is their unusual sensitivity to SDS at 39.5°C (Figure 3). When SDS-inactivated tsB45 phage are centrifuged, the characteristic antigens of the distal half-fiber (BC' antigens) remain in the supernatant, while those of the proximal half-fiber pellet with the viral particle (Table 1). This result suggests that the polypeptide region altered by the tsB45 mutation plays an important role in the proximal-distal half-fiber interaction.

Figure 2. Placement of five ts mutations on the physical map of gene 34. Placement of ts mutations in the interval tsA44 to tsB33 is based on data of Beckendorf and Wilson (1972), Seed and Wood (1979), and Seed et al. (1979).

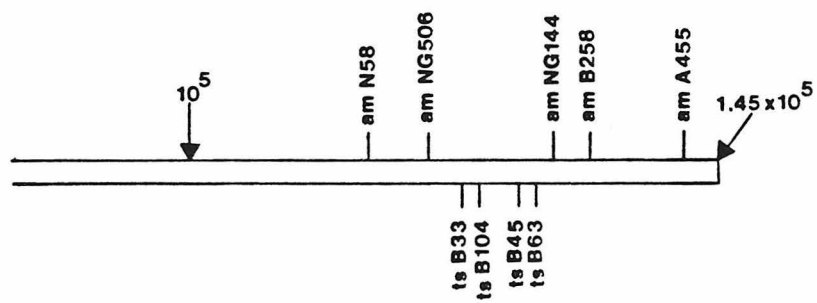


Figure 3. SDS sensitivity of tsB45 and selected gene 34 ts phage at 39.5°C. Phage were diluted in water with or without 1% SDS. The upper two curves represent phage inactivation in the absence of SDS. The remainder represent inactivation in the presence of SDS.

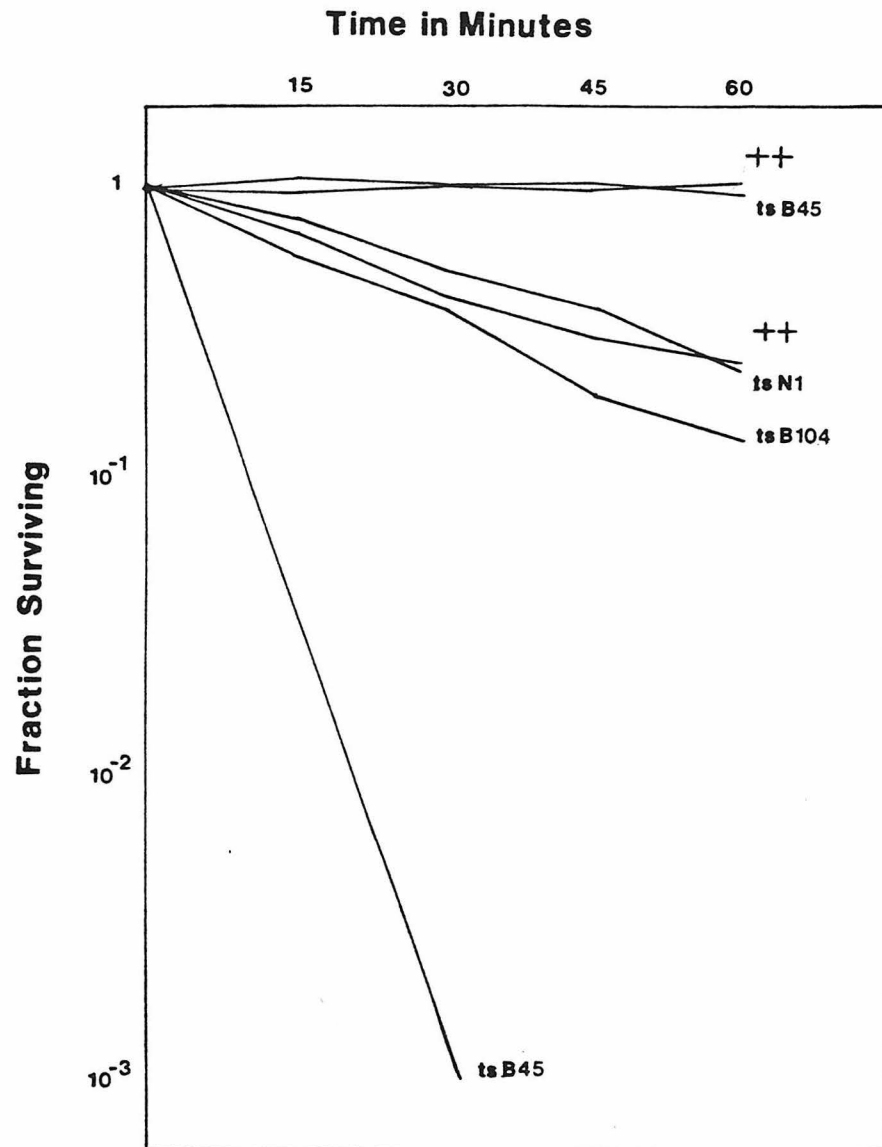


TABLE 1

Stability of tsB45 Phage in the Presence of 1% SDS^a

Phage/Fraction	+ 1% SDS		- SDS	
	A antigen ^b	BC' antigen ^b	A antigen	BC' antigen
<u>tsB45</u> supernatant	8 ± 1	72 ± 7	0 ^c	0
<u>tsB45</u> pellet	92 ± 9	28 ± 3	100 ± 10	100 ± 10
Recovery	<u>46%</u>	<u>60%</u>	<u>100%</u>	<u>100%</u>
<u>rdf</u> supernatant	6 ± 1	8 ± 1	0	0
<u>rdf</u> pellet	94 ± 9	92 ± 9	100 ± 10	100 ± 10
Recovery	<u>100%</u>	<u>97%</u>	<u>100%</u>	<u>100%</u>

^aSuspensions of 5×10^{10} phage were incubated for 60 min at 39.5°C with and without SDS. The suspensions were centrifuged, the pellets were resuspended, and the amounts of A and BC' antigen were determined in both fractions. rdf phage, used as controls, carry deletions of the rII A and B cistrons.

^bValues reported as percent of total antigen recovered.

^cValues $\leq 0.5\%$, background.

To isolate mutants affecting the interaction between the tsB45 site and the polypeptides of the distal half-fiber, double mutants containing tsB45 and one amber mutation from each of genes 34 through 38, and from gene 57, were constructed. The double mutants were plated at 42°C on a suppressing host to isolate temperature-insensitive revertants. The isolated revertants were then used directly to complement tsB45 or gene 34 amber phage on nonsuppressing host at high temperature. Phage that failed to complement in the initial screen were presumed to have the revertant site in the same gene as the amber mutation, and were replated and rescreened. The results of this preliminary survey are presented in Table 2. Pseudorevertants could apparently be isolated from each of the structural genes of the distal half-fiber, but could not be obtained from either gene 34 itself nor from either of the two genes coding for nonstructural proteins required for distal half-fiber assembly.

To further define the genetic locations of the pseudorevertants tentatively assigned to genes 35 through 37, we carried out a series of mapping crosses. Seven pseudorevertants in gene 37 were each crossed with both gene 34 tsB45 and gene 34 amNG144 phage. The recombinant progeny were scored on nonsuppressing host at 42°C. An equal multiplicity mixture of all 71 gene 37 revertants was also crossed with tsB45 and amNG144 to see if the majority of revertant phage would map at the same locus as the individually mapped strains.

When a revertant phage is crossed by tsB45, the percent recombinant progeny determines the absolute distance of the revertant site from its respective amber mutation; but it does not determine on which side of the amber mutation the revertant site lies. A cross with gene 34 amNG144, however, does define the relative position of the revertant site because a double crossover is required to produce phenotypically wild-type progeny when the revertant site lies distal to the amber site (Figure 4).

Figure 5 contains the results of the gene 37 revertant crosses. Data were normalized to align with the standard gene 37 map by performing control crosses

TABLE 2

Apparent Frequency of
Pseudorevertants by Gene

Gene	Frequency
34	0/200
35	16/200
36	17/200
37	71/100
38	0/100
57	0/100

Figure 4. Outcome of a hypothetical cross between gene 37 revertant and gene 34 amNG144 phage. Solid and dotted lines schematize the simplest genetic exchanges producing phenotypically wild-type phage. When the revertant site lies proximal to the tsB45 site (solid line), only one genetic exchange is necessary. When the revertant site falls distal to the tsB45 site (dotted line), two exchanges are required.

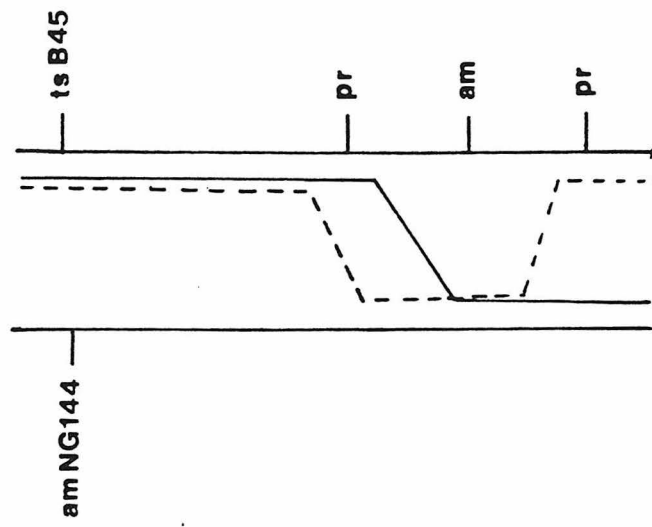
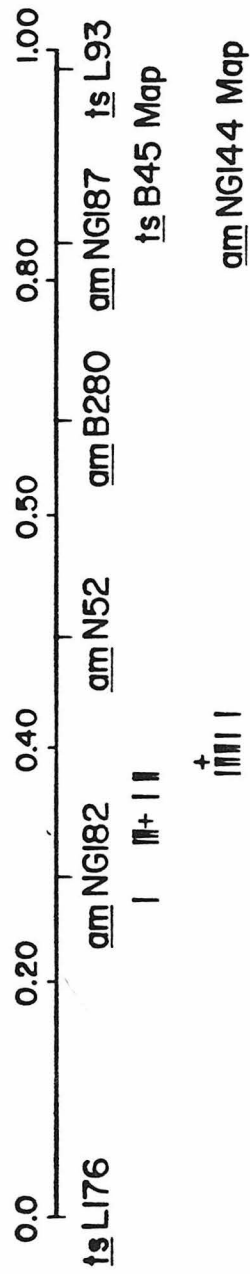


Figure 5. Map location of gene 37 pseudorevertants. Existing markers in gene 37 (Beckendorf et al., 1973) are included solely for reference to mutational sites discovered in this study. Vertical bars represent individual revertant site locations; crosses represent map locations of equal multiplicity mixture of revertant phage. Upper row, data from crosses with tsB45; lower row, data from crosses with amNG144.



identical to those performed by previous investigators (Beckendorf et al., 1973).

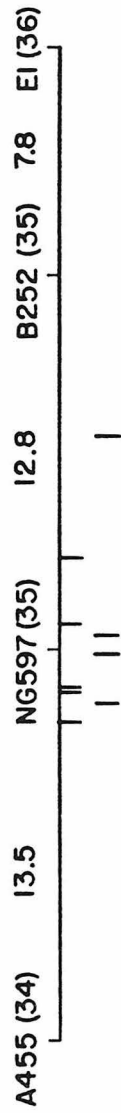
As can be seen from our map assignments, the gene 37 tsB45 revertants appear to form a cluster at about one-third the gene length from the promoter terminus.

The equal-weights mixture of revertants maps well within one standard deviation of the average map distance of the individually mapped revertants. Thus there appears to be no gross contamination of this population with revertants from other genes.

The extent of recombination between revertant phage and tsB45 phage was virtually the same as that between revertant phage and amNG144 phage. Thus a double crossover was not required in the latter case, and the revertant mutation can be placed C-terminal to amN52. Crosses to tsB45 do not yield exactly the same map locations as crosses to amNG144, however. One possible explanation for this effect might be that crossovers between the amNG144 site and the revertant site which recombine again between the revertant site and amN52 yield inviable progeny, whereas the analogous double crossovers involving tsB45 give rise to phenotypically wild-type phage.

Genes 35 and 36 revertants were mapped by recombination both with tsB45 and with a double mutant constructed from tsB45 and gene 37 amNG182. The genes 35 and 36 revertants gave large recombination frequencies with tsB45 and very low frequencies with the double mutant. We concluded from these results that a double crossover was required to produce phenotypically wild-type recombinants from a cross with the double mutant, and that consequently the revertant sites lay proximal to their respective amber mutations. The distances obtained from the tsB45 crosses placed the gene 36 revertant sites at nearly the same locations as those of the gene 35 isolates. Both appeared to fall in a cluster close to the most promoter proximal amber mutation of gene 35. These initial assignments, as well as map distances between various amber mutations of genes 35 and 36 (J. Wilson and B. Seed, unpublished results), are presented in Figure 6.

Figure 6. Preliminary map locations of genes 35 and 36 pseudorevertants. Upper row, revertants isolated from tsB45:gene 35 am phage; lower row, revertants isolated from tsB45:gene 36 am phage. Distances between am markers (J. Wilson, personal communication; Seed et al., 1979) are included for reference.



Because the map distances in these crosses were relatively large, and because the physical endpoint of gene 35 was not well defined genetically, we undertook a series of crosses to establish whether the genes 35 and 36 revertants lay in gene 34 or in gene 35. One each of the genes 35 and 36 revertant phage were crossed to tsB45 phage to isolate tsB45:gene 35 or 36 revertant phage lacking am mutations. Gene 34 am mutations were then introduced by crosses with tsB45:amNG506 or tsB45:amNG144. Map locations for these am mutations can be found in Figure 2. Phage with the ts⁺:am phenotype were picked and backcrossed to wild type to demonstrate the presence of the original ts mutation. The results are presented in Table 3. Neither of the crosses with tsB45:amNG144 yielded recombinant progeny, while both crosses with tsB45:amNG506 yielded the desired tsB45:amNG506:gene 35 or 36 revertant. One each of the latter phage was crossed to tsB45 and the linkage determined as previously. Figure 7 shows that both revertant sites can be closely linked to the amNG144 mutation, in good agreement with the results of the constructive crosses. Thus both revertants can be assigned to gene 34 by recombination.

To characterize the biochemical basis of the pseudorevertant phenotypes we examined a few phage for burst sizes, SDS sensitivity, and ability to assemble in vitro. Table 4 compares the burst sizes of selected pseudorevertant phage with tsB45 and wild-type phage at 25° and 42°C. At 42°C the pseudorevertants isolated with amber mutations in genes 35 and 36 displayed burst sizes 20 to 35 times larger than those of tsB45, but still markedly smaller (one- to two-hundredfold) than those of wild type. The pseudorevertant assigned to gene 37 had a burst size only twice that of tsB45 phage. Since the pseudorevertant was isolated by its ability to form a plaque at 42°C, we assume there must be substantial differences between the burst size of this phage in liquid culture and its burst size on plates.

Burst sizes of revertant phage were also examined at 25°C. Here the genes 35- and 36-derived pseudorevertants gave about fourfold higher bursts than tsB45,

TABLE 3Construction of tsB45:gene 34 am:revertant Phage

Cross	Number of progeny tested	<u>ts</u> ⁺ : <u>am</u> Phage bearing <u>ts</u> B45
<u>ts</u> B45: <u>am</u> NG144 x <u>ts</u> B45: <u>pr</u> S1(35)	133	0
<u>ts</u> B45: <u>am</u> NG144 x <u>ts</u> B45: <u>pr</u> S2(36)	89	0
<u>ts</u> B45: <u>am</u> NG506 x <u>ts</u> B45: <u>pr</u> S1(35)	83	1
<u>ts</u> B45: <u>am</u> NG506 x <u>ts</u> B45: <u>pr</u> S2(36)	79	3

Figure 7. Map locations in gene 34 of genes 35- or 36-derived revertant mutations. Revertant sites have been placed distal to tsB45 on the basis of the constructive cross results.

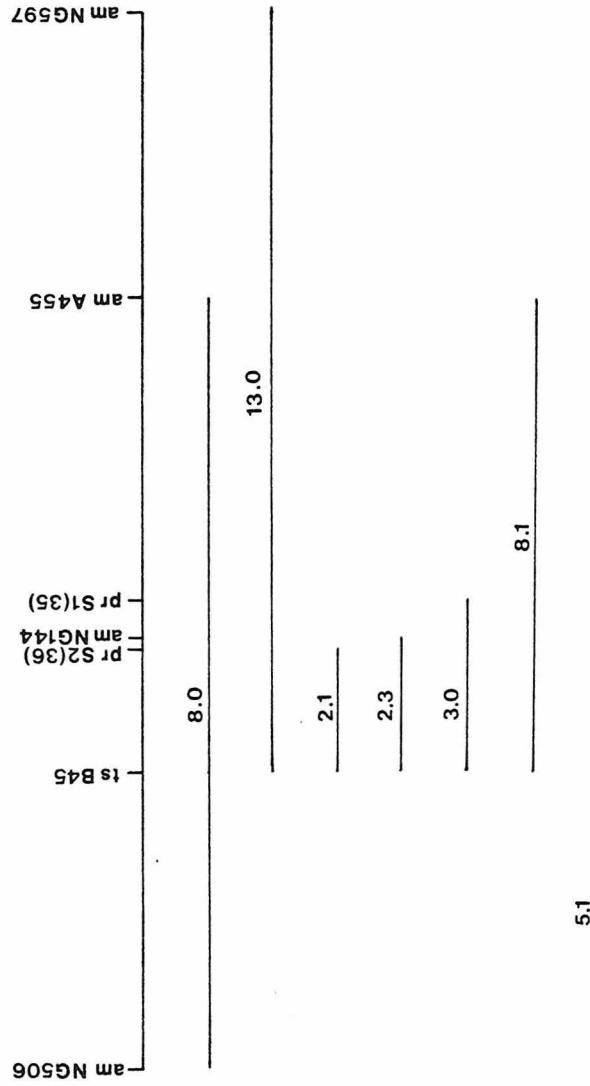


TABLE 4Burst Sizes of tsB45 Revertants at 25° and 42°C

Phage	25°C	42°C
<u>ts</u> B45	46.6	0.07
B45:R35	238	1.30
B45:R36	206	2.34
B45:R37	46.6	0.15
++	258	226

whereas the gene 37 pseudorevertant gave the same burst as tsB45. Thus at both temperatures tested, the genes 35- and 36-derived revertants produced higher bursts of intracellular phage than tsB45 alone, while the gene 37 revertant appeared identical to its temperature-sensitive parent.

The SDS sensitivities of the various pseudorevertants were determined by incubating phage in buffer containing 1% SDS at 42°C. As can be seen in Figure 8, wild-type phage suffer a 90% inactivation within 80 min of incubation, whereas tsB45 phage are nearly 99% inactivated under the same conditions. Neither gene 35- nor gene 36-derived revertants differ significantly from tsB45 in their rates of inactivation. The gene 37 revertant, however, is substantially more resistant to SDS at 42°C than tsB45, and is essentially equivalent to wild type under the same conditions. Discrimination between wild type and tsB45 resistances to SDS was reduced in these experiments, presumably because the original sensitivity experiments were performed in water, and the present experiments in buffer.

We examined all three types of revertant for production of proximal half-fiber antigens at 25° and 39.5°C. Table 5 shows that all three classes of revertant phage exhibit depressed antigen production at both temperatures, though the revertants isolated with genes 35 and 36 ambers exhibit higher titers of intracellular antigen than tsB45 phage. Since there was still a discrepancy between the burst sizes and titers of intracellular antigens, we felt there was a significant possibility that the rates or efficiencies of joining of ts half-fibers might be increased by the presence of revertant gene product. Accordingly, the revertant mutation from one of the gene 36-derived revertants was crossed into a tsB45:37amNG182 double mutant to give a phage which would be defective in the production of distal half-fiber, but which would leave the revertant mutation in a functional gene product. Extracts of nonsuppressing bacteria infected at 25°C with this phage were then used to complement a tsB45:gene 36amE1 phage extract in vitro. The results are shown in Figure 9.

Figure 8. SDS sensitivity of pseudorevertant phage. Phage were pipetted into dilution medium containing 1% SDS at 42°C and survivors titered after 0, 10, 20, 40, and 80 min of incubation.

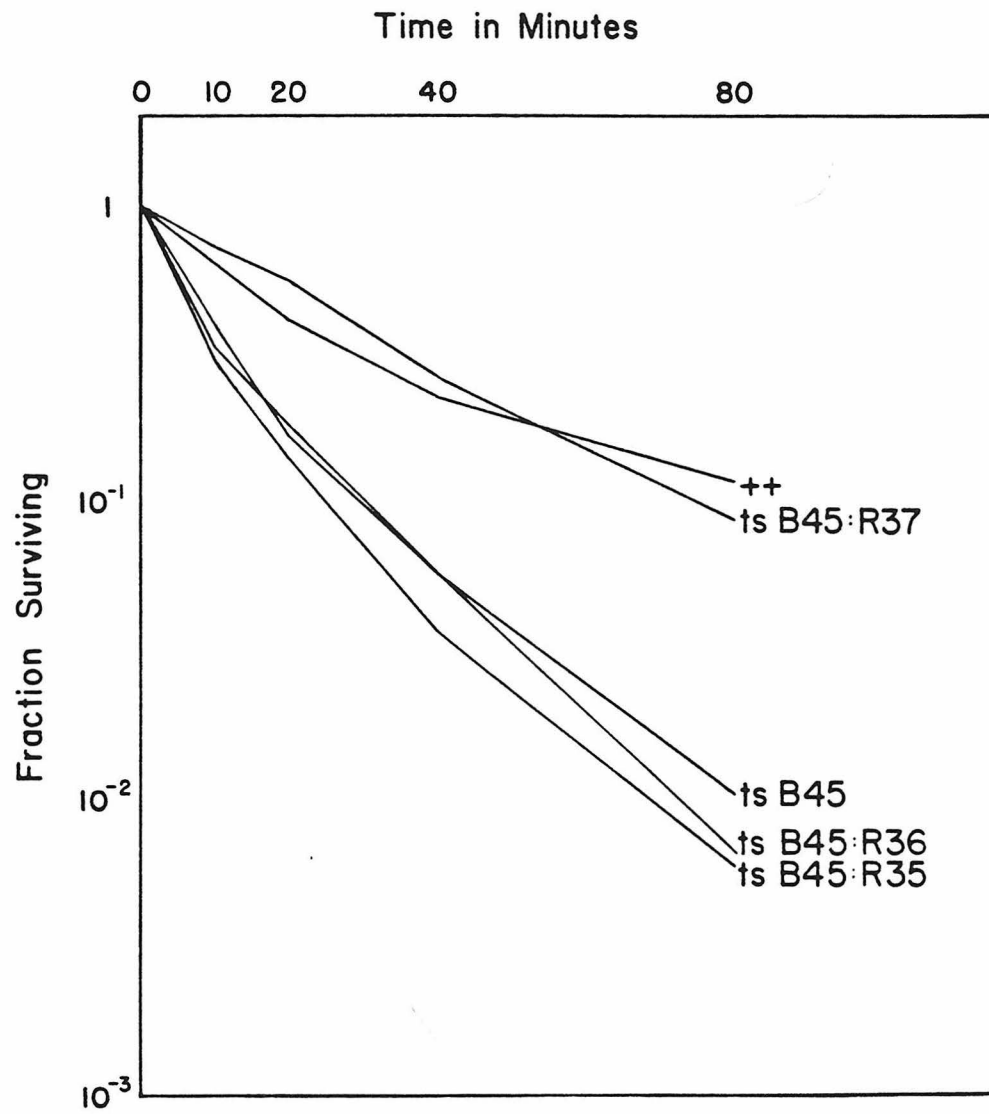
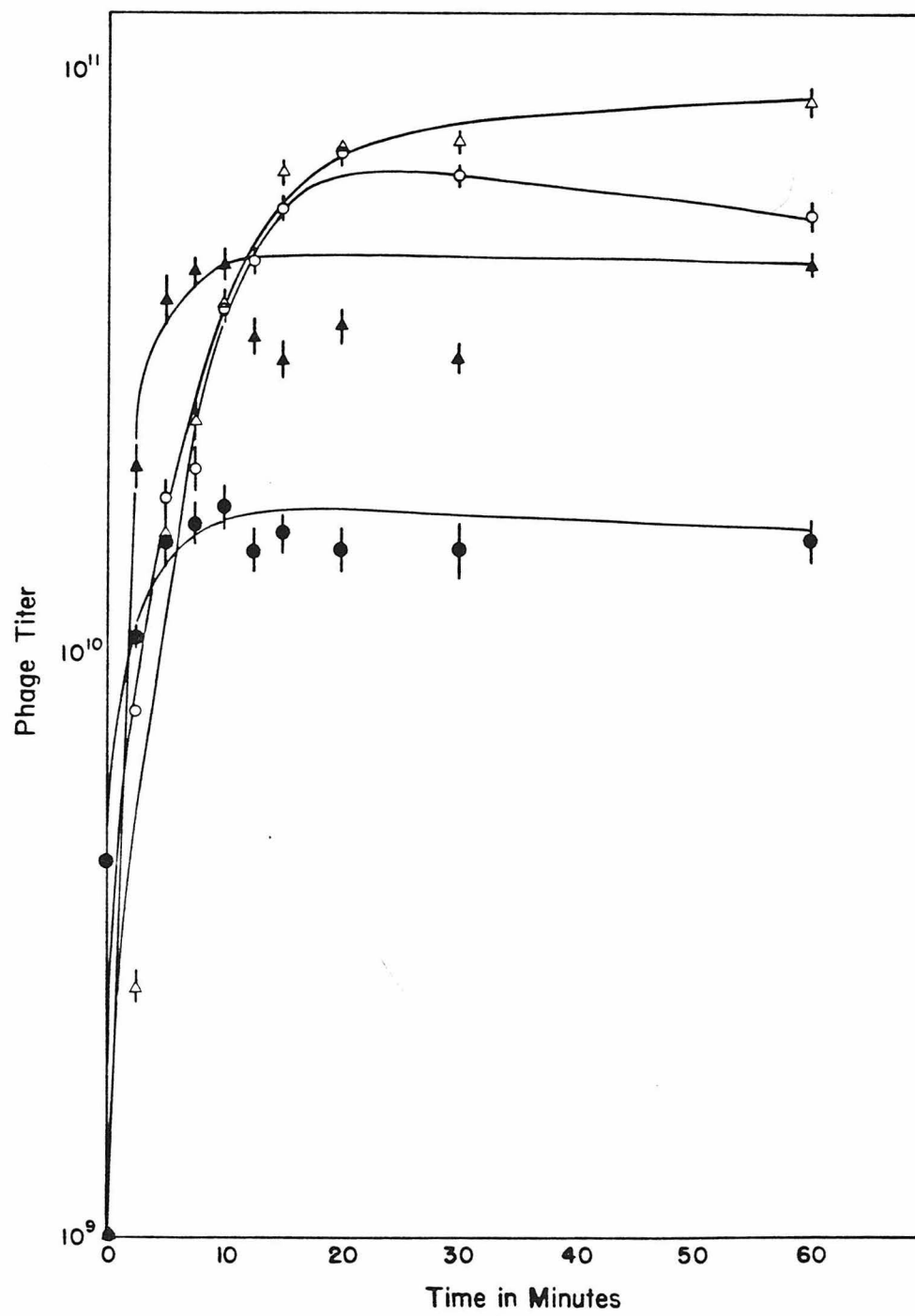


TABLE 5

A Antigen Production at Permissive and Nonpermissive
Temperature Phage-Equivalents of Antigen
per Infected Bacterium

	25°C	39.5°C
<u>ts</u> B45	58 $\pm$ 12.3	35 $\pm$ 4.2
++	557 $\pm$ 173	78 $\pm$ 11.3
R35	112 $\pm$ 9	56 $\pm$ 5.7
R36	154 $\pm$ 23	41 $\pm$ 1.4
R37	80	31 $\pm$ 4.2

Figure 9. In vitro complementation of a pseudorevertant located in gene 34. Extracts of a gene 36-derived revertant:tsB45:gene 37 am triple mutant or of a tsB45:gene 37 am double mutant were prepared at 25°C and mixed with extracts of a tsB45:gene 36 am double mutant, also prepared at 25°C. ▲, tsB45:amNG182: prS2(36) reaction at 37°C. Δ, same reaction at 25°C. ●, tsB45:amNG182 reaction at 37°C. O, same reaction at 25°C.



The rates and efficiencies of the complementation reaction with revertant gene product are significantly increased at 37°C relative to the homologous reactions with nonrevertant gene product. The rates and efficiencies of complementation at 25°C are not significantly different between the two reactions. It appears that at nonpermissive temperature in vitro assembly of tail fiber gene products can be significantly enhanced by revertant mutations.

To demonstrate that assembly efficiencies were altered by the revertant mutations in vivo we prepared lysates of an isogenic series of tsB45:amNG506:revertant phage at 39.5°C as a suppressing host. Phage particles were prepared from the lysates by CsCl gradient centrifugation, titered by optical density, and assayed for proximal half-fiber antigen by serum blocking. Table 6 shows that phage revertants isolated from genes 35 and 36 amber mutant phage exhibited nearly wild-type titers of tail fiber antigen per particle; gene 37 revertant phage exhibited only slightly higher titers than the tsB45 phage. Table 6 also shows that the fraction of particles capable of forming plaques was significantly enhanced in all three types of revertant. In these experiments the infectivity of wild-type phage was only 10% of that obtainable under optimal conditions.

Discussion

In this paper we have reported a new method of pseudorevertant analysis. Pseudorevertants of a temperature-sensitive mutation were isolated from double mutants containing both the temperature-sensitive mutation and one amber mutation selected from various genes of interest. Revertants were then classified by their failure to complement the original temperature-sensitive mutation at high temperature on a nonsuppressing host. Using this technique we uncovered two classes of pseudorevertants to the gene 34 tsB45 mutation, one falling in gene 37, and one in gene 34. A representative of the gene 37 revertants was less labile

TABLE 6

Half-Fiber Antigen and Phage Titer per Particle of tsB45 and
tsB45:Revertant Phage at Nonpermissive Temperature

Phage ^a	Antigen, fiber-equivalents/particle	% Infectious
++	4.0	4.8
<u>ts</u> B45	0.33	5.6×10^{-2}
<u>ts</u> B45: <u>pr</u> S1(35)	3.29	2.2
<u>ts</u> B45: <u>pr</u> S2(36)	2.72	0.64
<u>ts</u> B45: <u>pr</u> S3(37)	0.75	0.14

^aAll phage contained the gene 34 amNG506 mutation.

to SDS than the tsB45 phage, while representatives of the gene 34 revertants displayed increased efficiencies of assembly.

The most significant advantage of screening pseudorevertants by complementation is that a pseudorevertant need not have an additional phenotype, such as cold sensitivity, to be mapped. In our case, for example, a screen of over a thousand revertants of tsB45 turned up no cold-sensitive phage (B. Seed, unpublished results). The imposition of a dual phenotype thus can clearly limit the spectrum of revertants to be analyzed.

There are two major drawbacks to complementation screening for pseudorevertants. The first is that the genes to be examined for revertants must be known in advance. The second, and more serious, is that a failure to complement in the initial complementation test need not indicate that the revertant falls in the gene being tested. For example, a failure to complement will be observed whenever the revertant phenotype is recessive to wild type. Also, the presence of the parental amber mutation may have an effect upon the frequency of occurrence or dominance of the revertant mutation.

In this survey revertants appearing to belong in genes 35 and 36 by complementation screening were subsequently assigned to gene 34 by recombination. However, the revertants ultimately assigned to gene 34 only appeared in the genes 35 and 36 parental strains. If the genes 35 and 36-derived revertants were truly recessive, all parent strains should have yielded them. Hence the presence of genes 35 and 36 amber mutations in the complementation test may make the gene 34 revertant recessive. While it is also possible to argue that the gene 34 revertant mutation yields inviable phage when accompanied by an amber mutation in any gene but 35 or 36, since the revertant mutation can be successfully crossed into a tsB45:gene 37 amber strain this explanation is less likely.

No true revertants to tsB45 were revealed by our survey. Such revertants may be intrinsically unlikely compared to pseudorevertants. Alternatively, it may be that the detection of true revertants is hampered by the presence of an amber mutation in gene 34.

There are several interesting consequences of the pseudorevertant phenotypes we have observed. For example, because the gene 37 revertant had an identical burst size to tsB45 in liquid culture, we concluded that the burst size of these phage on plates must be significantly larger than in liquid culture. One consequence of this result is that we can no longer be sure that the observed plating efficiency of the tsB45 parental phage is due to the low burst size measured in liquid culture. Since gene 37 revertants increase the structural stability of the phage, the conditions for determining stability in vitro (detergent and high temperature) may have some analog in the life cycle of the phage on plates.

References

- Beckendorf, S. K., Kim, J. S. and Lielausis, I. (1973). Structure of bacteriophage T4 genes 37 and 38. *J. Mol. Biol.* **73**, 17-35.
- Beckendorf, S. K. and Wilson, J. H. (1972). A recombination gradient in bacteriophage T4 gene 34. *Virology* **50**, 315-321.
- Bishop, R. J. and Wood, W. B. (1976). Genetic analysis of T4 tail fiber assembly. I. A gene 37 mutation that allows bypass of gene 38 function. *Virology* **72**, 244-254.
- Edgar, R. S. and Lielausis, I. (1965). Serological studies with mutants of phage T4D defective in genes determining tail fiber structure. *Genetics* **52**, 1187-1200.
- Edgar, R. S. and Wood, W. B. (1966). Morphogenesis of bacteriophage T4 in extracts of mutant-infected cells. *Proc. Nat. Acad. Sci. USA* **55**, 498-505.
- Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, G. H. and Lielausis, A. (1963). Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 375-394.
- Jarvik, J. and Botstein, D. (1975). Conditional-lethal mutations that suppress genetic defects in morphogenesis by altering structural proteins. *Proc. Nat. Acad. Sci. USA* **72**, 2738-2742.
- King, J. (1968). Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231-262.
- Seed, B., Del Zoppo, G. J. and Wood, W. B. (1979). Chapter 4 of this thesis.
- Steinberg, C. M. and Edgar, R. S. (1962). A critical test of a current theory of genetic recombination in bacteriophage. *Genetics* **47**, 187-208.
- Ward, S., Luftig, R. B., Wilson, J. H., Eddleman, H., Lyle, H. and Wood, W. B. (1970). Assembly of bacteriophage T4 tail fibers. II. Isolation and characterization of tail fiber precursors. *J. Mol. Biol.* **54**, 15-31.
- Wilson, J. H. and Kells, S. (1972). Bacteriophage T4 transfer RNA. I. Isolation and characterization of two phage-coded nonsense suppressors. *J. Mol. Biol.* **69**, 39-56.

CHAPTER 4

**Analysis of Temperature-Sensitive Mutations in
Bacteriophage T4 Gene 34**

Brian Seed, Gregory J. Del Zoppo and William B. Wood

Division of Biology
California Institute of Technology
Pasadena, California 91125

The proximal half tail fiber of bacteriophage T4 is a rod-shaped dimer of the 1.45×10^5 dalton polypeptide coded by gene 34 (Ward and Dickson, 1971; King and Laemmli, 1971; Vanderslice and Yegian, 1974). The dimerization of the gene 34 product requires a nonstructural assembly protein coded by gene 57 (gp57) (King and Wood, 1969; Dickson, 1973). The mechanism of action of gp57 is unknown.

Antiserum raised against purified proximal half-fibers does not cross-react with other phage components (Ward et al., 1970). Studies of the serologically related phages of the RB family have shown that the characteristic antigens of the proximal half-fiber (the A antigens) are not homogeneously repeated throughout the half-fiber (Seed and Wood, 1979). An antigenic divergence between the RB phages and T4 has been localized to the carboxy-terminal region of the gene 34 polypeptide. When assembled, this region forms the distal portion of the proximal half-fiber (Seed and Wood, 1979). The carboxy-terminal antigenic determinants of the polypeptide chain contribute significantly more to the susceptibility of the phage to antiserum than do the amino-terminal determinants (Seed and Wood, 1979).

The genetic map of T4 gene 34 exhibits a gradient of recombination frequency increasing toward the promoter-distal terminus of the gene (Beckendorf and Wilson, 1972). The region of anomalous recombination extends into gene 35, located distally to gene 34 (Stahl et al., 1970).

To date, both temperature- (Edgar and Lielausis, 1965) and cold- (J. Flatgaard, unpublished results) sensitive mutants of gene 34 have been isolated. Genetic characterization of the former has demonstrated that, with the exception of tsA44, all gene 34 ts mutants exhibit intragenic complementation in liquid culture (Bernstein et al., 1965). tsA44 complements other gene 34 ts mutants only weakly (Bernstein et al., 1965). A preliminary characterization of antigens present in extracts of bacterial cells infected with gene 34 ts mutant phage at nonpermissive temperature has already

been reported (Edgar and Lielausis, 1965).

In this paper the genetic and biochemical properties of gene 34 ts mutations are examined in greater detail. A contradiction between mapping results of previous investigators is shown to be a consequence of a high recombination anomaly appearing in crosses of ts mutations by ts mutations, but not in crosses of am by am mutations nor of am by ts mutations. The anomaly is confined to the distal terminus of the gene.

All gene 34 ts mutant phage-infected cells are found to produce wild-type amounts of gene 34 product (gp34) at nonpermissive temperature. Antigen production by these cells is significantly depressed by distal ts mutations, but only slightly depressed by proximal ts mutations. With one exception the antigen produced at nonpermissive temperature by ts phage-infected cells does not appreciably differ from wild-type antigen either by quantitative serum blocking assay or sucrose gradient sedimentation. The adsorption rates and thermolabilities of gene 34 ts mutant phage prepared at permissive temperature exhibit no differences from wild-type phage when examined at nonpermissive temperature. With three exceptions, phage particles isolated from ts mutant-infected cells bear significantly fewer tail fibers than do wild-type particles.

Materials and Methods

Bacteriophage and bacterial strains. Amber (am) and temperature-sensitive (ts) mutants of bacteriophage T4D were obtained from the collections of R. S. Edgar and W. B. Wood and have been described elsewhere (Epstein et al., 1963; Edgar and Lielausis, 1965). Multiple mutants were constructed by genetic crosses and are described in the text.

Escherichia coli strains CR63, CR63(λ), and B40 Su_I^+ were the amber permissive hosts used in this study. Strains B/5 and P301 (B40 $Su^- arg^-$) were the nonpermissive hosts. The plating indicators for serum neutralization experiments were S26(λ), G(λ) and B/5. The restrictive temperature for ts mutants was 39.5°C in liquid media and 42°C on plates (Edgar and Lielausis, 1965). The respective permissive temperatures were 30° and 25°C.

Media. H broth was used for phage and bacterial growth. EHA top and bottom agars were as described by Steinberg and Edgar (1962). Dilution buffer was prepared according to King (1968). Radioactively labeled infected-cell lysates were prepared from cultures grown in M9 medium supplemented with tryptophan. Lysates to be assayed for serum blocking were prepared from cultures grown on the same medium or on M9A.

Serum blocking assays. Serum blocking assays for proximal half-fiber (A) antigen were performed by the endpoint protocol described by Ward et al. (1970). When comparing the amounts of A antigen produced in lysates of mutant phage-infected cells to those produced by wild-type infected cells, the percentage of antigen was calculated as the concentration of wild-type lysate sufficient to block 50% of the serum divided by the equivalent concentration of mutant lysate.

Preparation of defective lysates at nonpermissive temperature. E. coli strain B15 was grown in M9 or M9A medium to a density of 4×10^8 cells/ml at 39.5°C.

Aliquots were transferred to bubbler tubes or small (50 ml) shaker flasks at 39.5°C. The cultures were infected at a multiplicity of infection (m.o.i.) of 4 to 5 phage/cell, usually with tryptophan added as an adsorption cofactor. Strains not containing the lysis delay mutation amtB5 were superinfected at 7 min after infection. Cultures were lysed with CHCl_3 45 to 60 min after infection. In some cases, DNase (Sigma DN-C) and egg white lysozyme (Sigma) were added to each culture to a concentration of 10 $\mu\text{g/ml}$ each. For experiments to test antigen formation lysates shifted to permissive temperature, cell debris was removed from lysates by centrifugation at 8000 g for 10 min in a rotor warmed to 40°C. In other experiments, the cell debris was not removed, and had no significant effect on antigen titers (see text).

Preparation of ^{14}C labeled infected-cell lysates at nonpermissive temperature.

Strain B5 was grown to 4×10^7 cells/ml at 39.5°C in M9 medium + L-tryptophan, centrifuged, and resuspended to 4×10^8 cells/ml in fresh medium. Two milliliter aliquots were dispensed into small Erlenmeyer flasks and warmed at 39.5°C prior to infection. The cultures were infected at a m.o.i. of 4-6 phage/cell, superinfected at 7 min after infection at the same m.o.i. and aerated by agitation on a rotary shaker. At 15 min after infection, 2 μCi of a ^{14}C labeled algal hydrolysate (Schwarz-Mann) was added to each culture. At 20 min, the cultures were centrifuged at 3000 g for 3 min in a rotor warmed to 40°C. The pellets were resuspended in 0.4 ml of a buffer containing 0.01 M Tris-HCl (pH 6.8), 0.001 M MgSO_4 , 10 $\mu\text{g/ml}$ DNase, and 5 $\mu\text{g/ml}$ lysozyme. These suspensions were incubated for 10-15 min, cleared by the addition of 20 μl of CHCl_2 , and incubated for 30 min at 39.5°C. Prior to electrophoresis, the lysates were adjusted to 0.065 M Tris-HCl (pH 6.8), 1% SDS, and 1% 2-mercaptoethanol.

SDS discontinuous gel electrophoresis. The procedures for the preparation and electrophoresis of phage proteins in a discontinuous buffer system containing SDS were those described by Laemmli (1970) as modified by Dickson (1973). This

method was adapted for use with 2 mm slabs of 5% polyacrylamide gel and is similar to that described by Studier (1972). Samples of labeled lysate containing $4-6 \times 10^4$ cpm were electrophoresed from each well. After electrophoresis, the gels were fixed in an aqueous solution of 50% (V/V) anhydrous methanol, 10% acetic acid for 2 hr and were soaked in 10% anhydrous methanol, 10% acetic acid for 12-14 hr. The fixed gels were dried directly onto sheets of No. 1 grade Whatman filter paper and autoradiographed on Kodak No-Screen X-ray film by the procedures of Fairbanks, Levinthal and Reeder (1965) and Dickson (1973).

Temperature shift experiments. *E. coli* strain B/5 was grown with rotary shaking in M9 medium at 39.5°C to a density of 4×10^8 cells/ml. DL-tryptophan and infecting phage were added to 50 µg/ml and a m.o.i. of 4, respectively. A freshly prepared 1 mg/ml solution of chloramphenicol (Sigma, D-threo) in 10 mM Tris-HCl, pH 8.0, was added at the times indicated in the text. Control cultures were treated with 10 mM Tris solution alone. Shift-down was accomplished by using a chilled 10 ml pipette to transfer culture at 39.5°C to flasks maintained at 22°C in a gyratory water bath. At selected times following shift-down, 0.25 ml aliquots of culture were pipetted into tubes containing a few drops of chloroform. Aliquots from these tubes were used directly to measure serum blocking titer.

Phage adsorption assay. *E. coli* strain B/5 was grown and prepared as for a cross (Wilson and Kells, 1972). Phage were mixed with bacteria at a m.o.i. of 15 and unadsorbed titer measured at various intervals thereafter by pipetting aliquots of the adsorption mixture into H broth containing chloroform.

Sucrose gradient sedimentation. 13 ml 5-20% linear sucrose gradients in phosphate buffer (King and Wood, 1969) containing 0.01% gelatin were poured in SW41 tubes. Aliquots of 0.1 to 0.2 ml of centrifuged ts mutant-infected cell lysates were layered on the gradients and spun at 37,000 rpm for 15 to 16 hr. Five drop fractions were collected, diluted as appropriate, and assayed by serum blocking.

CsCl density gradient centrifugation. CsCl step gradients were prepared as described by Dickson (1973). Phage particles were removed either by aspiration with a pipette or by puncture through the side of the tube.

Results

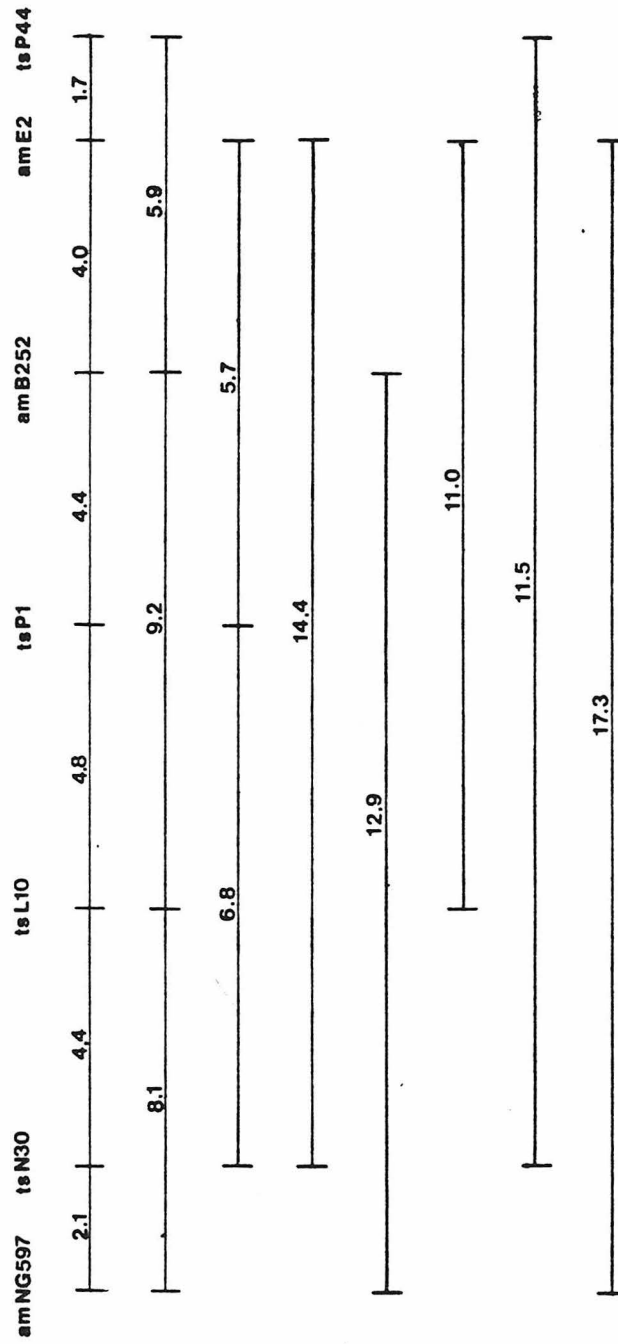
Two conflicting genetic maps have been reported for T4 gene 34. The first, constructed from ts by ts mutant crosses (Bernstein et al., 1965), is nearly 15% longer than the second, constructed from am by am mutant crosses (Beckendorf and Wilson, 1972). Our results, summarized in Figure 1, confirm this discrepancy. Two-factor crosses between am and ts markers demonstrate that the distal region of gene 34 has a genetic map expansion observable between ts markers, but not between am markers. Furthermore the ts by am map aligns more consistently with the am by am map than with the ts by ts map (Figure 1).

The promoter-distal region of T4 gene 34 exhibits a gradient of am by am recombination frequency increasing toward the distal end of the gene (Beckendorf and Wilson, 1972). This region of elevated recombination apparently also extends into gene 35, since gene 35 exhibits a genetic length far too large for the length of polypeptide chain it produces (Russell, 1967). However, the existence of an increased recombination frequency per unit length of gene 35 DNA has not been shown directly. The gene 34 ts by ts map expansion does not have a counterpart in gene 35. Figure 2 presents a composite map of gene 35 drawn from both our own results and unpublished work of John Wilson (J. Wilson, personal communication). There is no apparent ts-specific recombination anomaly.

One possible explanation for the ts by ts recombination anomaly is that parental heterozygotes of gene 34 ts mutant phage will produce viable phage by intragenic complementation. Heterozygotes of gene 35 ts phage should be incapable of producing viable phage by this mechanism because the gene 35 product exists

Figure 1. Genetic map of T4 gene 34. The upper line represents the physical length of polypeptide chain produced by gene 34; the positions of the am mutations N58 through A455 have been determined by measuring lengths of nonsense-terminated polypeptides (Beckendorf and Wilson, 1972). Alignment of other am mutations is based on a supposition of constant recombination frequency throughout the remainder of the gene. ts mutation sites have been aligned relative to am sites on the basis of the genetic data collected in map III (this study). Other genetic data are the am by am map of Beckendorf and Wilson (1972) (map I), and the ts by ts map of Bernstein et al. (1965) (map II). The interval tsB104 to tsA44 exhibits a map expansion found in ts by ts crosses, but not in ts by am crosses.

Figure 2. Genetic map of T4 gene 35. No polypeptide mapping data exist for this gene. However the results of the ts by ts crosses agree with those of the ts by am and am by am crosses.



as a monomer in the distal half-fiber. Table 1 demonstrates that gene 35 ts mutant phage do not exhibit intragenic complementation, as expected.

Production of proximal half-fiber antigens by gene 34 ts mutants was quantitated by serum blocking assay of lysates prepared from cells infected with double mutants carrying a ts mutation in gene 34 and the amN52 mutation in gene 37. The latter mutation eliminates production of distal half-fiber antigens. Figure 3 shows representative serum blocking curves for five different centrifuged gene 34 ts:gene 37 am defective lysates. Other defective ts lysates gave similar results. The curves recorded for lysates of phage producing low titers of antigen appear slightly broader than those for high titer lysates. Similar curves are observed, however, when wild-type antigen is diluted in buffer containing any of a variety of interfering agents, including sucrose, broth, and salt (B. Seed, unpublished observations). With the exception of this effect, no curve of Figure 3 has an appreciably different slope from the control. This result indicates that in a ts lysate prepared at nonpermissive temperature the relative abundances of the various antigens remain constant, irrespective of the total amount of antigen produced. Such a result differs from those obtained using adsorbed antisera prepared by treatment of whole anti-T4 serum with mutant extracts lacking gene 34 product (Edgar and Lielausis, 1965). We believe this difference reflects the presence in the adsorbed sera of antibodies directed against other phage structures.

The amount of proximal half-fiber antigens present in a defective lysate of a particular ts mutant appears to correlate with the position of the ts mutation in gene 34 (Figure 4). Mutations near the distal end of the gene (e.g., tsA44) produce 0 to 20% of wild-type levels of A antigens in defective lysates, whereas ts mutations near the proximal end of the gene produce 50 to 100%. Since these measurements, as well as those obtained previously (Edgar and Lielausis, 1965), have been made on centrifuged lysates, it is possible that the differing amounts of antigen measured

TABLE 1

Absence of Intragenic ts Complementation
in Gene 35

Phage pairs complemented	Burst size
<u>ts</u> N30(35) alone	1.81 $\pm$ 2.03
<u>ts</u> P1(35) alone	0.95 $\pm$ 0.72
<u>ts</u> P44(35) alone	0.60 $\pm$ 0.10
<u>ts</u> L10(35) alone	0.86 $\pm$ 0.16
<u>ts</u> N30 + <u>ts</u> P1	1.14 $\pm$ 0.75
<u>ts</u> N30 + <u>ts</u> P44	0.72 $\pm$ 0.11
<u>ts</u> P1 + <u>ts</u> P44	0.77 $\pm$ 0.38
<u>ts</u> N30 + <u>ts</u> L10	0.68 $\pm$ 0.22
<u>ts</u> P1 + <u>ts</u> L10	2.0 $\pm$ 2.60
<u>ts</u> B104(34) alone	0.39 $\pm$ 0.17
<u>ts</u> B111(34) alone	1.61 $\pm$ 0.87
<u>ts</u> B104 + <u>ts</u> B111	24.7 $\pm$ 19.0

Figure 3. Blocking curves of proximal half-fiber antiserum by representative defective ts lysates. Lysates were infected with amN52:gene 34 ts double mutants (amN52 prevents assembly of A antigen into tail fibers) and centrifuged. The blocking curves are identified by their respective ts mutations. ts mutant curves were not normalized for burst size.

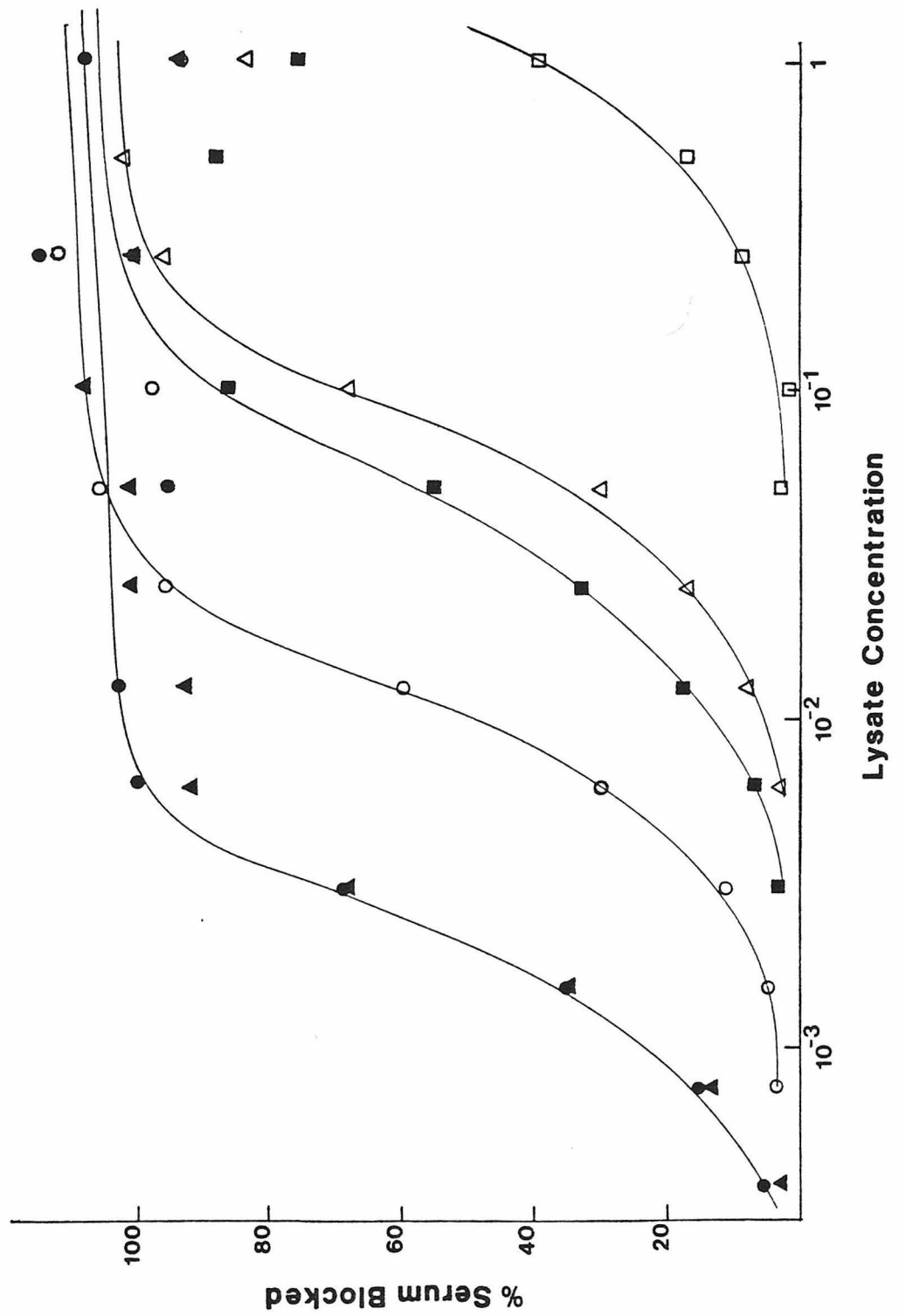
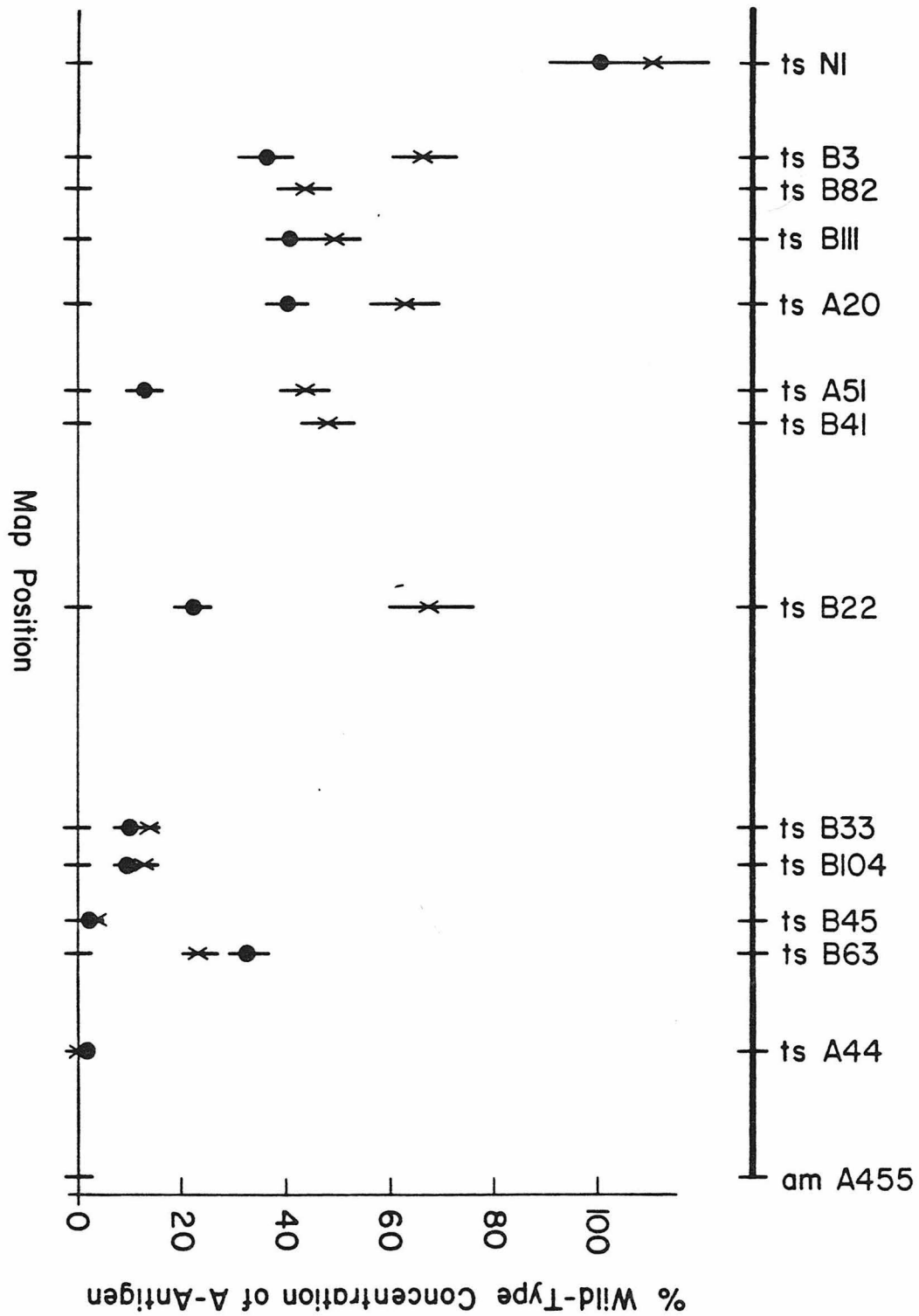


Figure 4. Relationship between the map locations of ts mutations and the amount of soluble A antigen formed in defective lysates. The values of the ordinate are calculated from the ratios of the concentration of wild-type lysate at the 50% blocking point, to the concentration of ts mutant lysate at the 50% blocking point. ts lysates were infected with amN52:gene 34 ts mutants and centrifuged. Soluble antigen was normalized by phage particle titer, measured by in vitro complementation. The two types of symbols represent results obtained in two different experiments.



reflect a selective sedimentation of the A¹⁴² antigens of different ts half-fibers. To test this possibility, A antigens were measured in lysates of tsA44:amN52, tsB104:amN52, tsB45:amN52 and tsN1:amN52, before and after centrifugation. The antigen measured in all centrifuged lysates represented 30 to 50% of the antigen in the corresponding uncentrifuged lysates, and the quantitative results obtained with uncentrifuged lysates confirmed those with centrifuged lysates (G. Del Zoppo, unpublished observations).

Earlier studies have shown that heating gene 37 defective lysates to 100°C in the presence of SDS destroys proximal half-fiber antigen and dissociates the half-fiber into gp34 monomers, whereas heating lysates to 37°C in the presence of SDS does not (Ward and Dickson, 1971; King and Laemmli, 1971). Uncentrifuged labeled defective lysates were prepared under nonpermissive conditions with gene 34 ts:amN52 mutants, treated with SDS, incubated at 100° or 20°C, and analyzed by electrophoresis on SDS gels. The results of these analyses are shown in Figure 5. The gel pattern of lysates incubated at 100°C shows no gross decrease in synthesis of gp34 by any of the ts mutants at nonpermissive temperature. The gel pattern of lysates incubated at 20°C shows that, with the exception of tsN1, all lysates prepared at nonpermissive temperature either have significant amounts of unassembled gp34 or have labile half-fibers. The former interpretation would be consistent with the serum blocking data shown earlier. The prominent band that might be expected from undissociated half-fibers is not seen, probably because these structures do not enter the gel.

Lysates of the gene 34 ts:amN52 mutants also were prepared at the permissive temperature and treated as described in the preceding paragraph. Only lysates of double mutants carrying tsB63, tsB45, and tsB104 produced bands on SDS gels of the same mobility as gp34 when treated at 20°C (data not shown). Thus with the exception of the three mutants mentioned, assembled ts half-fibers are not labile to SDS at 20°C. The gp34 bands of the three exceptional mutants are not necessarily due to half-fiber lability. For example tsB45-infected cells exhibit depressed antigen

Figure 5. SDS gels of defective amN52:gene 34 ts lysates at 20° and 100°C. Samples were prepared as described in Materials and Methods and electrophoresed on 5 or 6% polyacrylamide gels containing SDS. Gels were dried and autoradiographed. 20° and 100°C treated samples of a given mutant lysate were run in parallel. Panel a, 100°C samples. Panel b, 20°C samples.

144

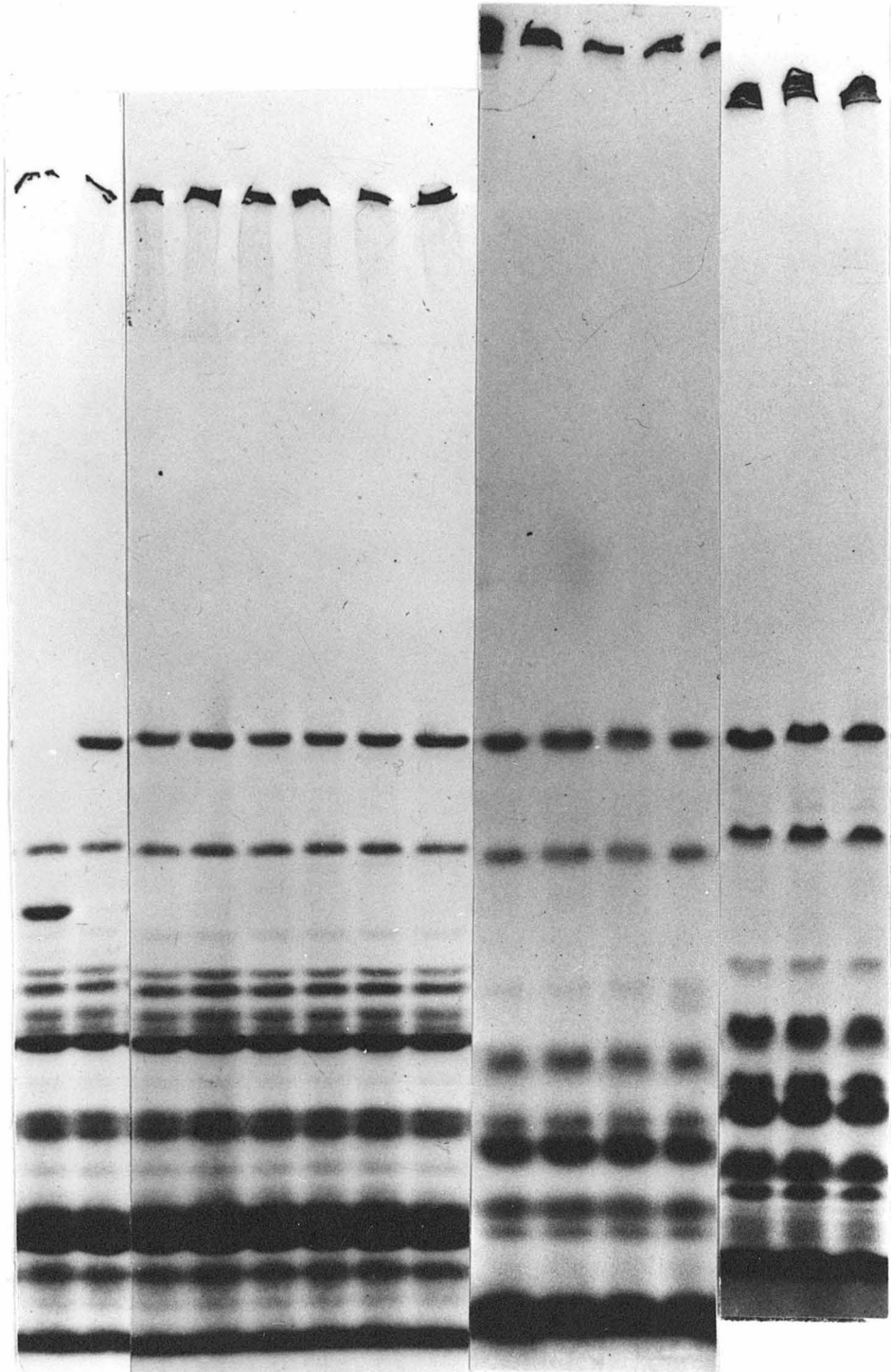
100°

gp34

gp7

gp37

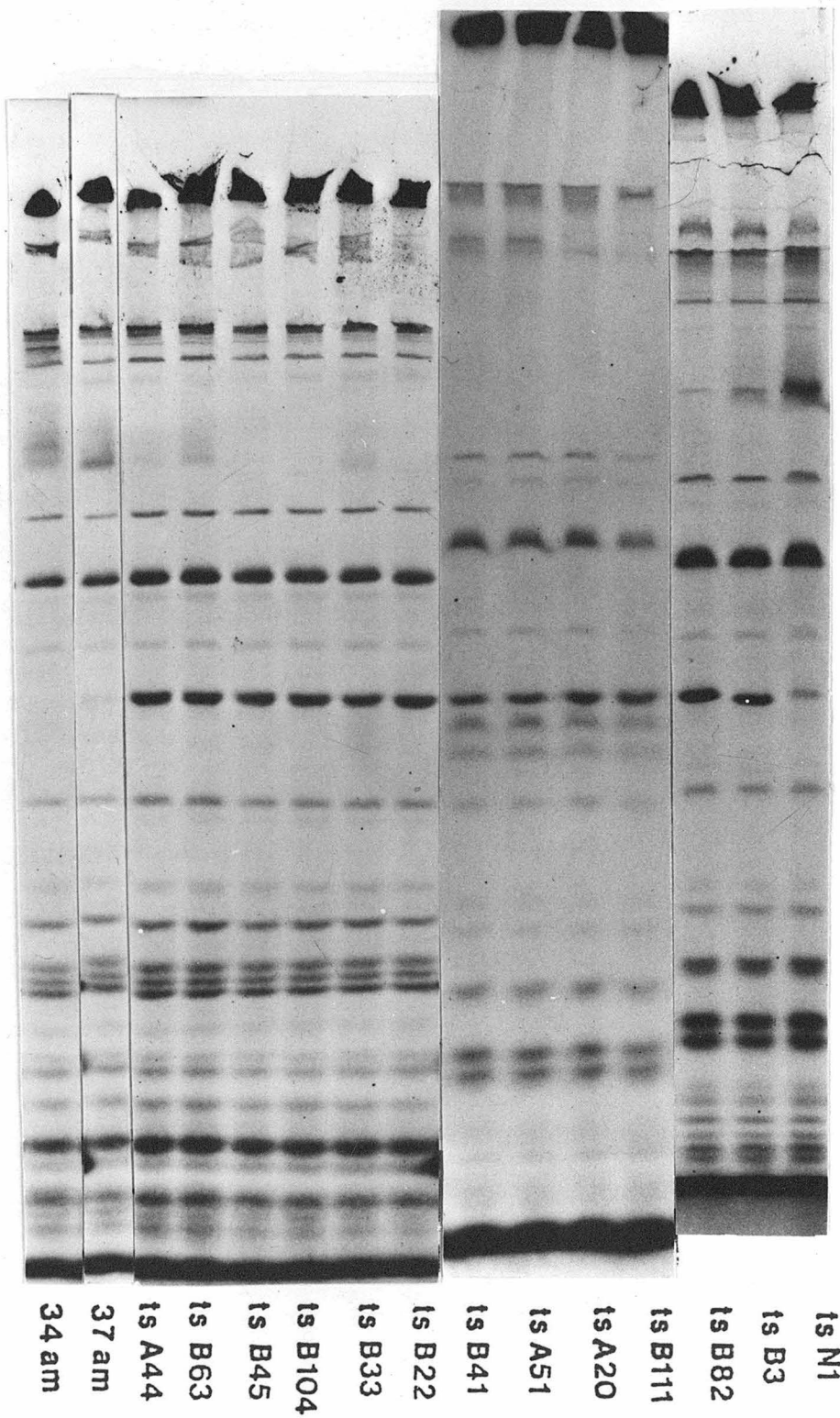
ts N1
ts B3
ts B82
ts B111
ts A20
ts A51
ts B41
ts B22
ts B33
ts B104
ts B45
ts B63
ts A44
37 am
34 am



145

20°

gp34



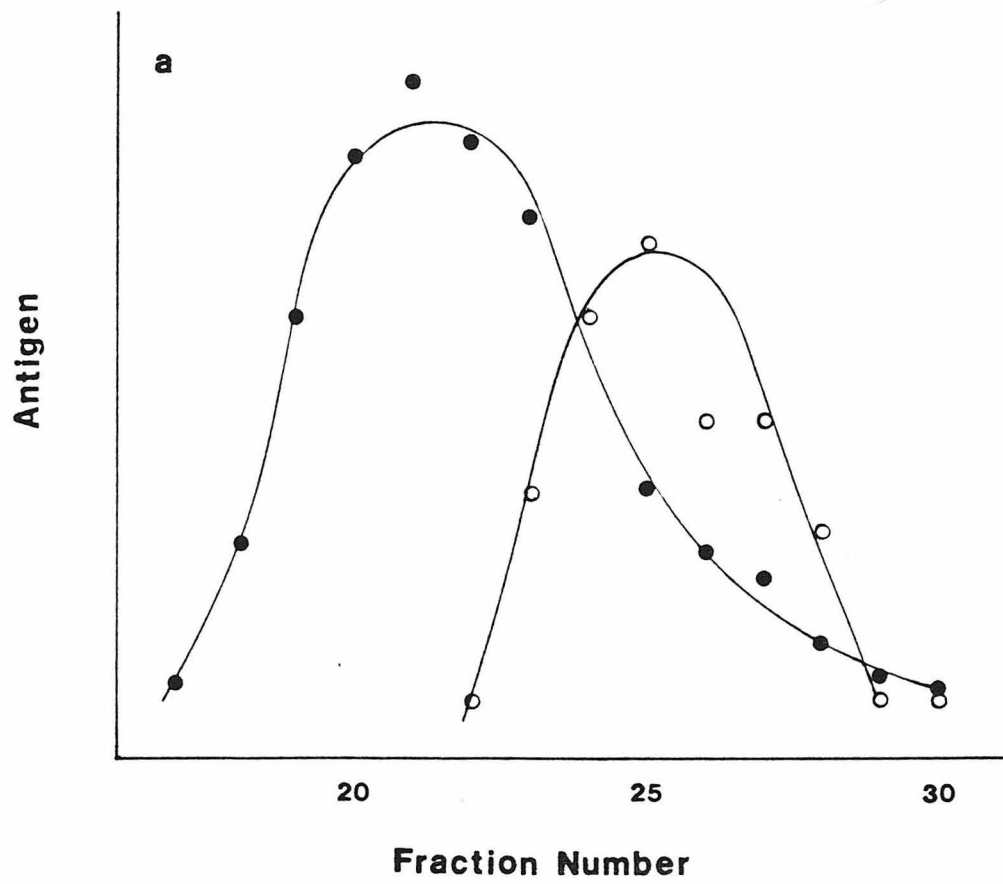
production even at permissive temperature (Seed et al., 1979). Thus the bands observed in lysates of these cells may be unassembled gp34.

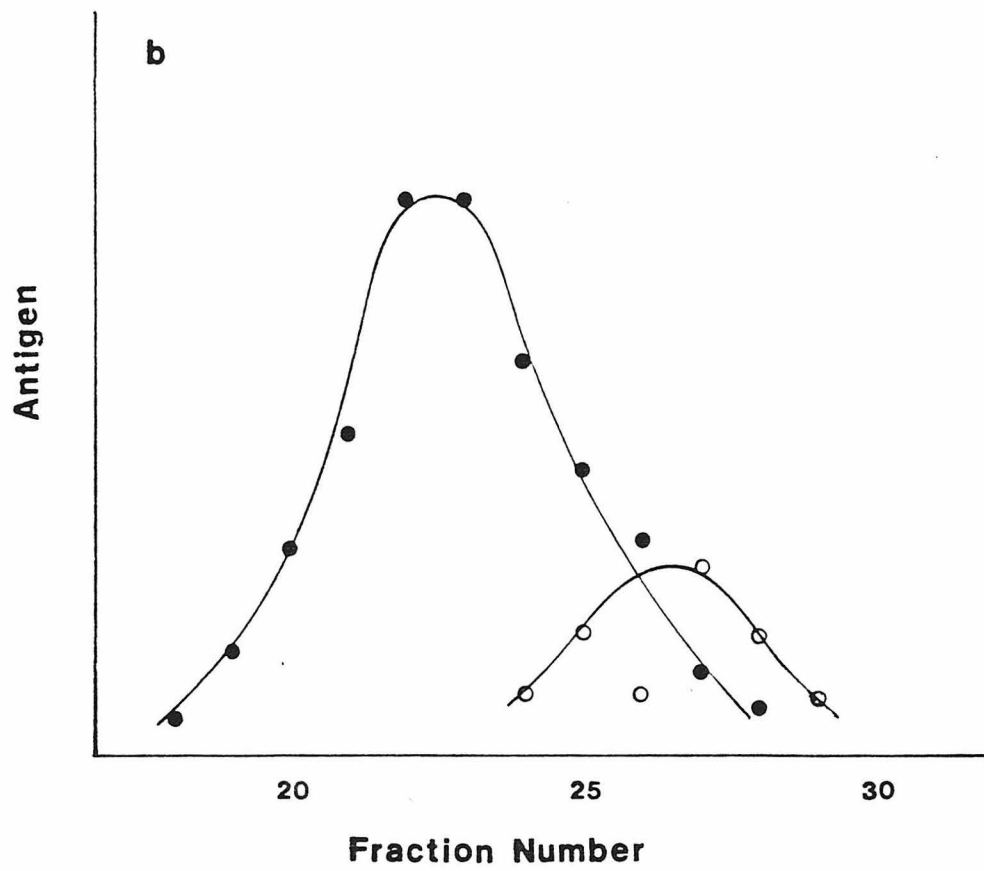
Lysates of selected gene 34 ts mutant phage-infected cells were prepared at nonpermissive temperature and sedimented through 5-20% sucrose gradients (Figure 6). Serum blocking of the gradients revealed no gross alteration in sedimentation behavior of the ts mutant antigens, with one exception. tsB33 lysates were observed to contain a small but reproducible amount of a rapidly sedimenting antigen specie. This specie may represent a proximal half-fiber dimer similar to that observed in vitro by Takata and Tsugita (1970).

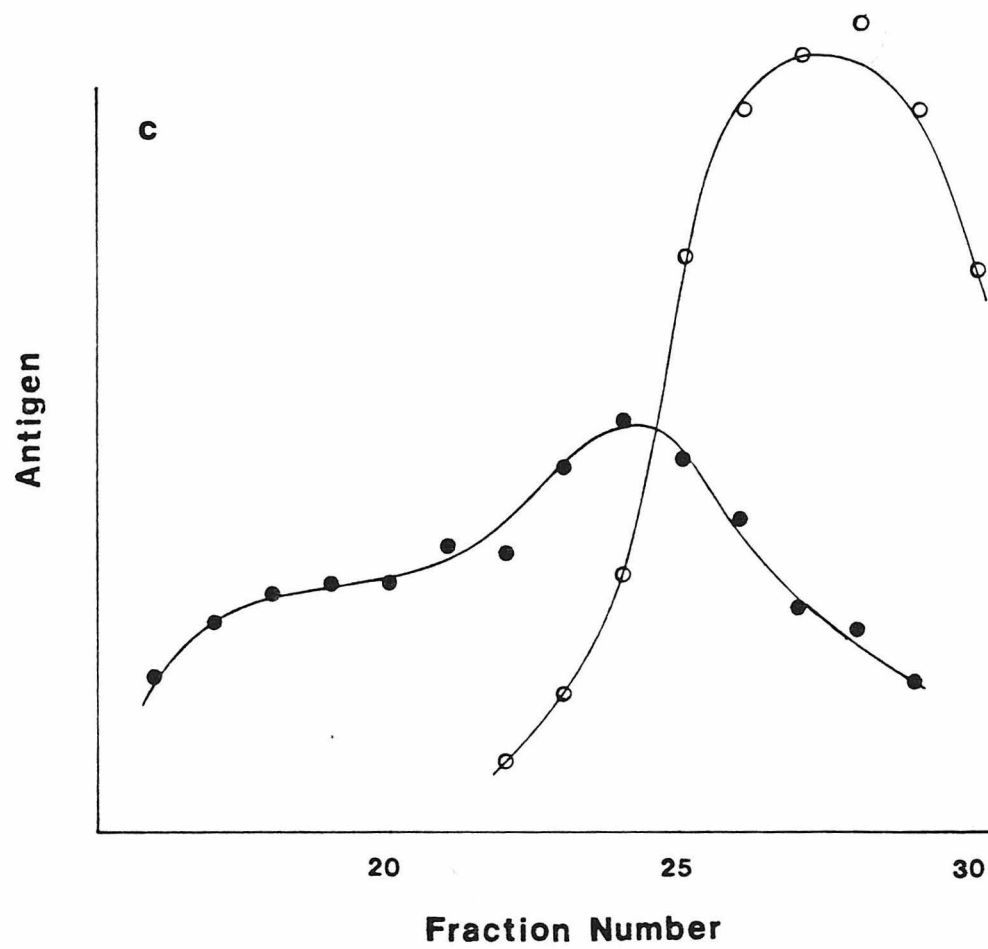
Cells infected with gene 34 ts mutant phage produce substantial amounts of half-fiber antigens when shifted from nonpermissive to permissive temperature at any time up to but not after lysis. Figure 7 shows the result of shift-down experiments with several mutants. Cultures were infected with various gene 34 ts mutant phage at 39.5°C and aliquots shifted to permissive temperature at selected intervals following infection. Superinfected cultures begin to lyse at about 30 min after infection at 39.5°C under these conditions. At about the same time the net antigen production (measured at 90 min) of the temperature-shifted cultures begins to decline. This suggests that production is limited by lysis. To test this hypothesis a shift-down experiment with the multiple mutant tsA44:amN52:amB5, capable of extending the lysis period, was performed (Josslin, 1971). As seen in Figure 6 a decrease in antigen yield occurs around 45 min after infection. Lysis begins at 40 min in cultures infected with this mutant.

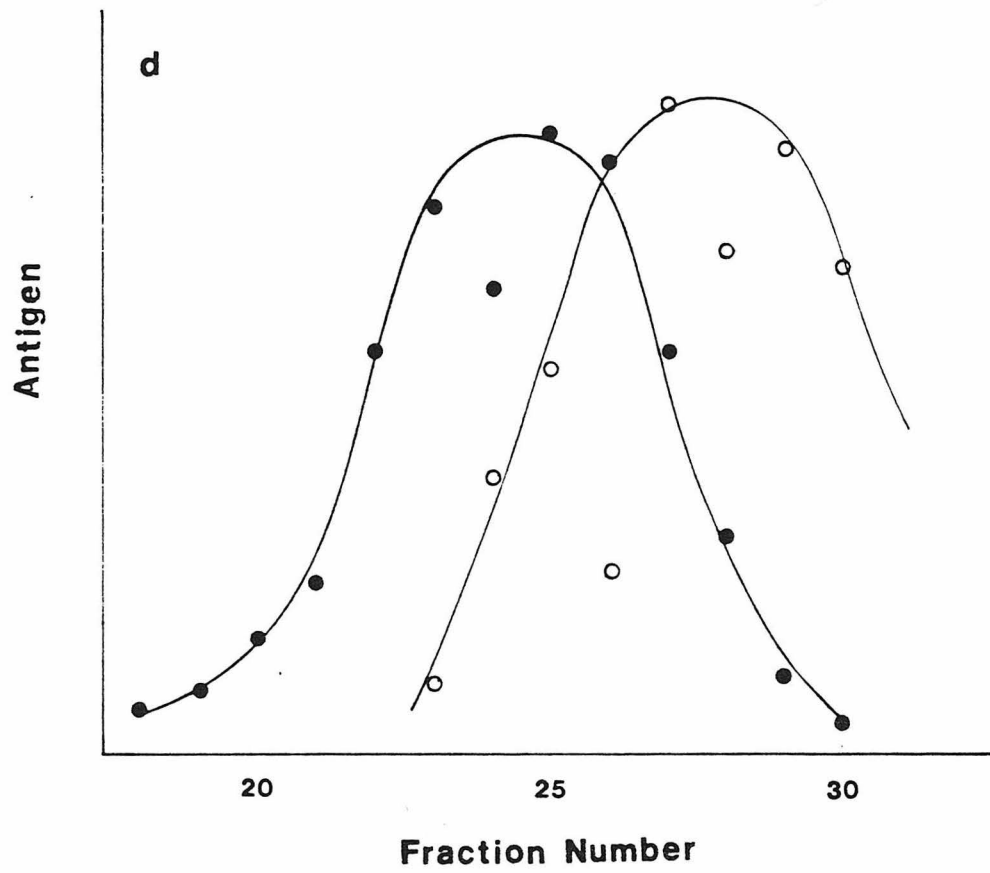
The above experiments suggest that mutant gp34 cannot be converted to A antigen by a shift to permissive temperature after lysis. This observation was confirmed for lysates of several ts mutants (Table 2). A antigen made by preparing lysates at permissive temperature was found to be stable to a shift to nonpermissive temperature (Table 2).

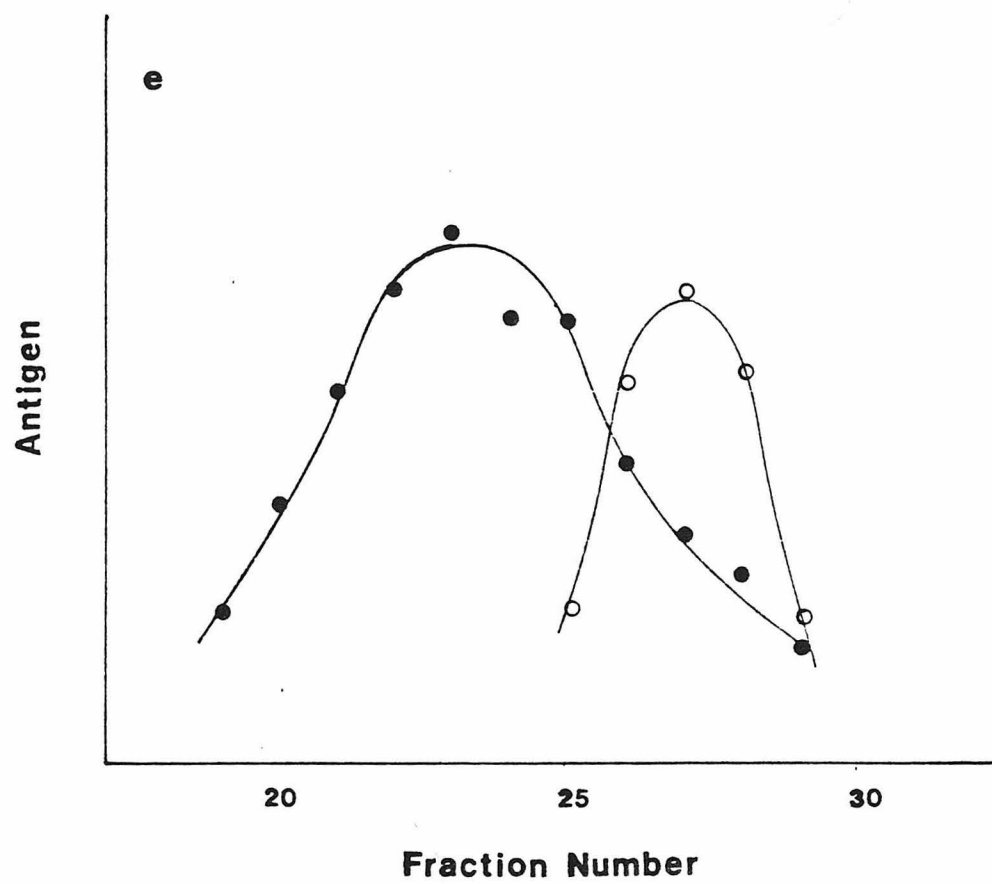
Figure 6. Sucrose gradient sedimentation of ts mutant A antigen produced at non-permissive temperature. ts lysate samples were centrifuged through 5 to 20% sucrose gradients. The direction of sedimentation is from right to left. Distal half-fiber antigens were also measured for an internal standard. The ordinate is expressed in arbitrary units of antigen/ml, and the same scale does not apply to both proximal and distal half-fiber antigens. ●, proximal half-fiber antigen; O, distal half-fiber antigen. Panel a, tsB111; panel b, tsN1; panel c, tsB33; panel d, tsA20; panel e, tsB41; and panel f, wild type.











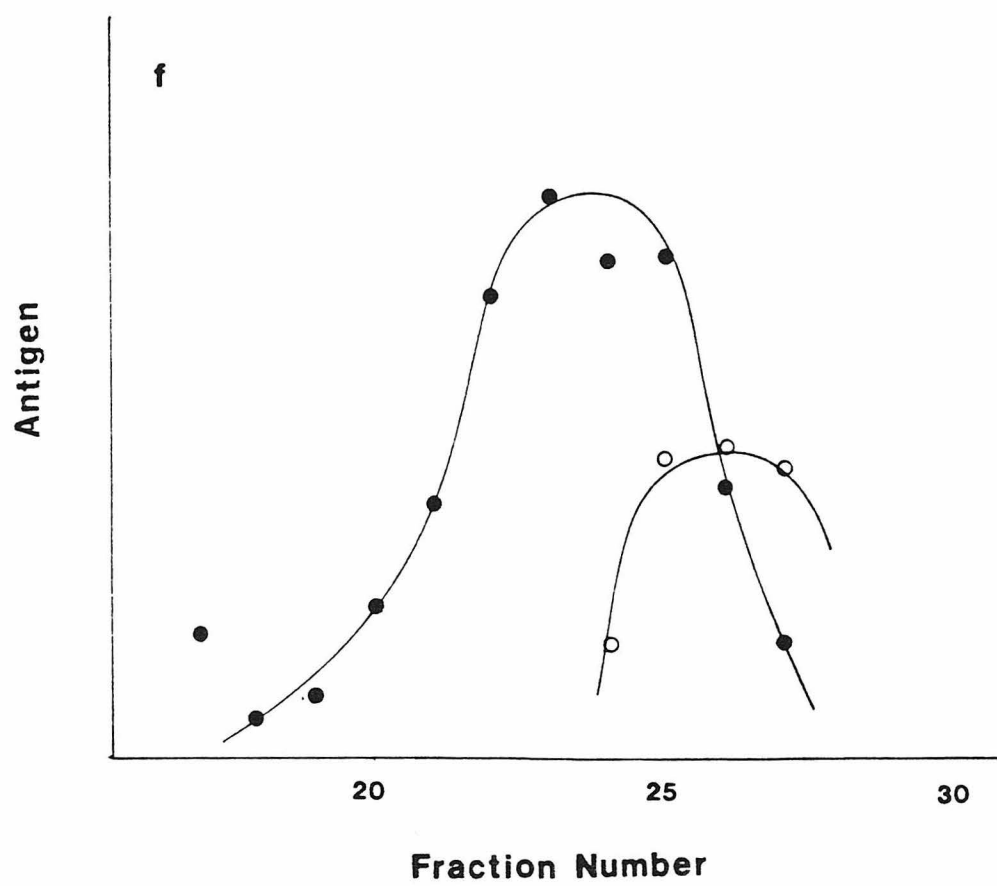


Figure 7. Lysis limitation of A antigen in vivo. Temperature shift from 39.5° to 30°C. Cultures of E. coli B/5 at 39.5°C were infected and superinfected with the appropriate ts mutant or ts:am double mutant. At various times after infection aliquots of each culture were shifted to 30°C under aeration. Each subculture was lysed with CHCl_3 and 10 $\mu\text{g/ml}$ lysozyme at 90 min after infection. A antigen yields are plotted as (concentration of lysate at 50% blocking point)⁻¹. , , tsA44; O, tsA44:amN52; , tsB45:amN52; X, tsA44:amN52:amB5.

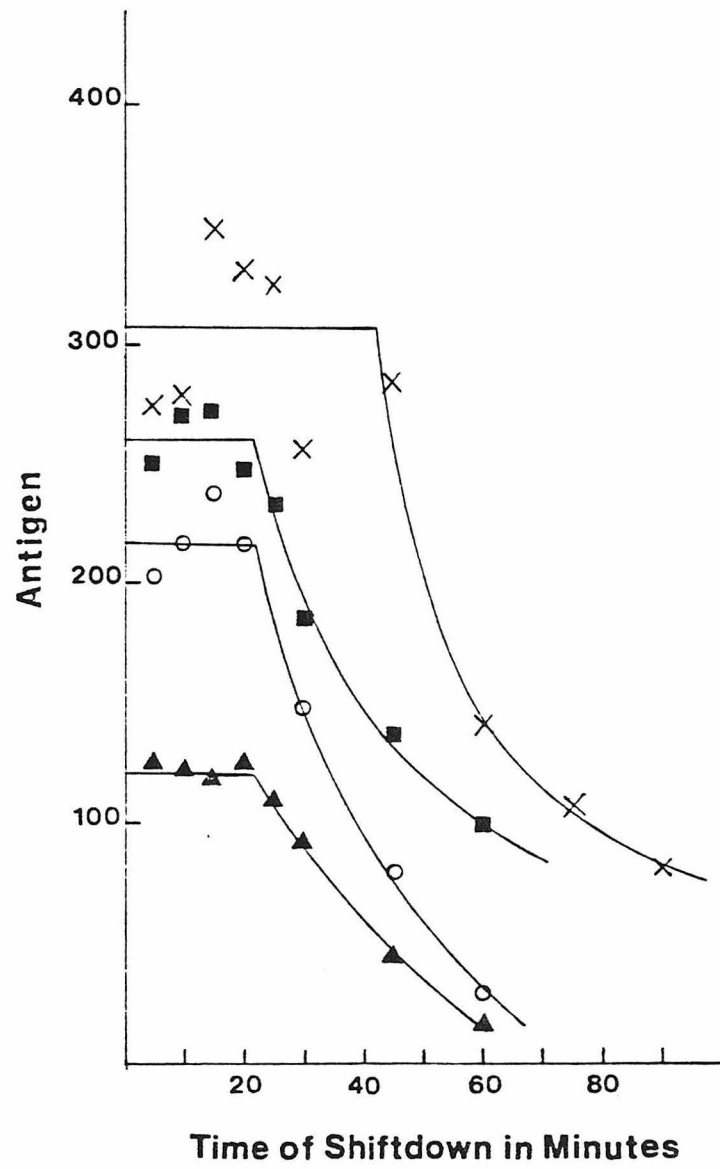


TABLE 2

Stability of A Antigen in some Lysates of Gene 34 ts Mutants

Mutant	A antigen in lysate ^a			
	Lysate prepared at 39.5°C	Shift to 30°C	Lysate prepared at 30°C	Shift to 39.5°C
<u>tsA44:amN52</u>	(0.77 ± 0.1) ^b	(0.77 ± 0.1) ^b	222 ± 22	244 ± 24
<u>tsB63:amN52</u>	91 ± 9	100 ± 10	156 ± 16	170 ± 17
<u>tsB45:amN52</u>	12 ± 1	14 ± 1	232 ± 23	232 ± 23
<u>tsB104:amN52</u>	18 ± 2	23 ± 2	172 ± 17	167 ± 17
<u>tsN1:amN52</u>	42 ± 4	48 ± 5	384 ± 38	345 ± 35

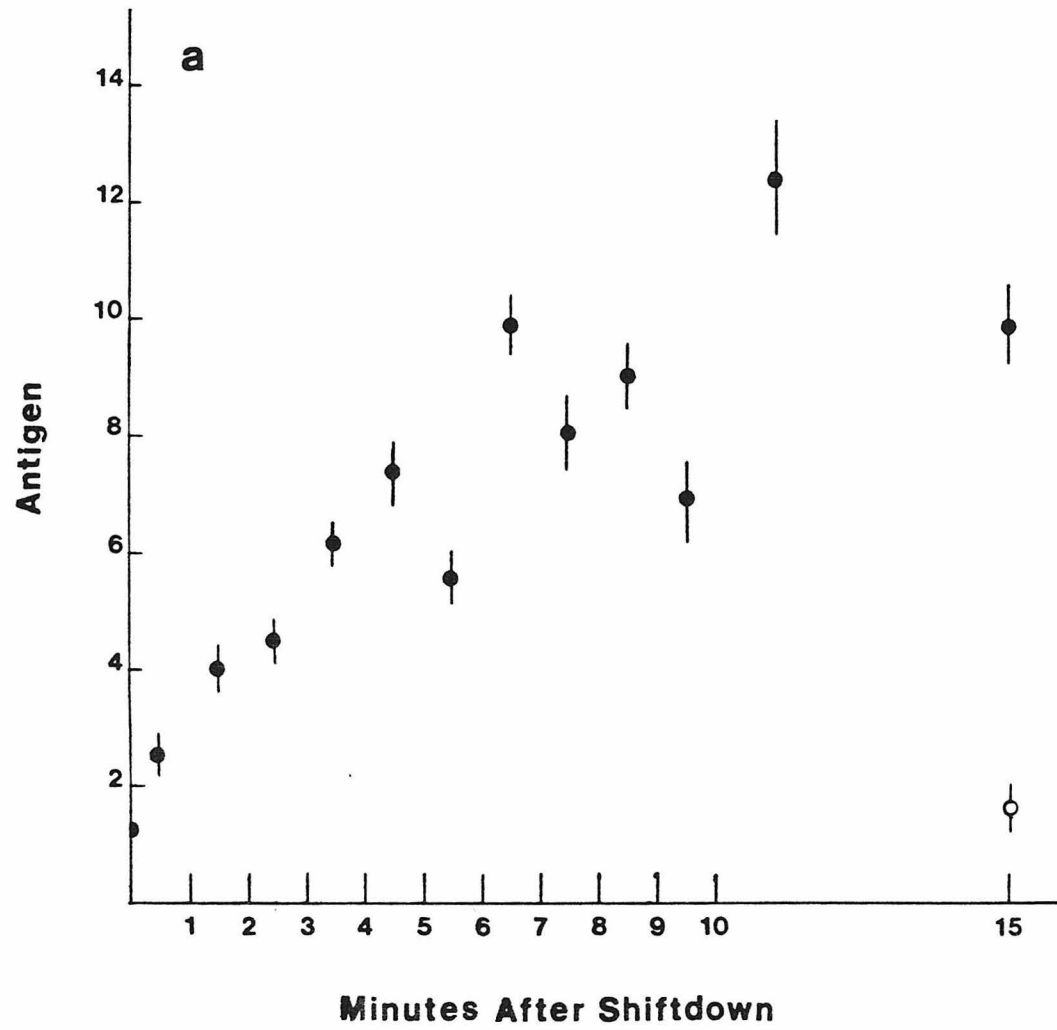
^aLysates were prepared at 39.5° or 30°C. An aliquot of each lysate was shifted to the appropriate temperature and incubated for 60 min. Serum blocking assays were performed on the lysate before and after shift. Values of A antigen are reported as (concentration of lysate at 50% blocking point)⁻¹.

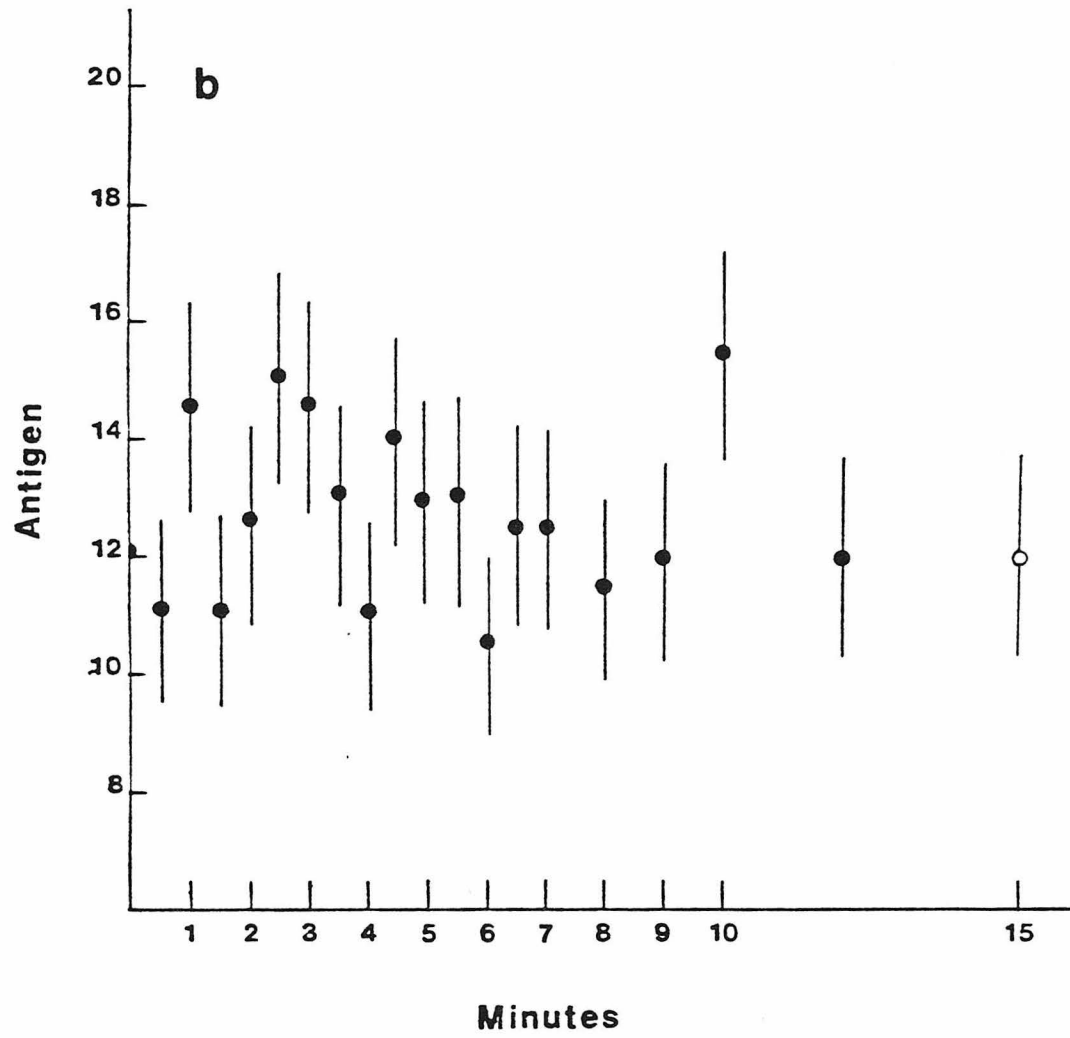
^bParentheses indicate extrapolated values.

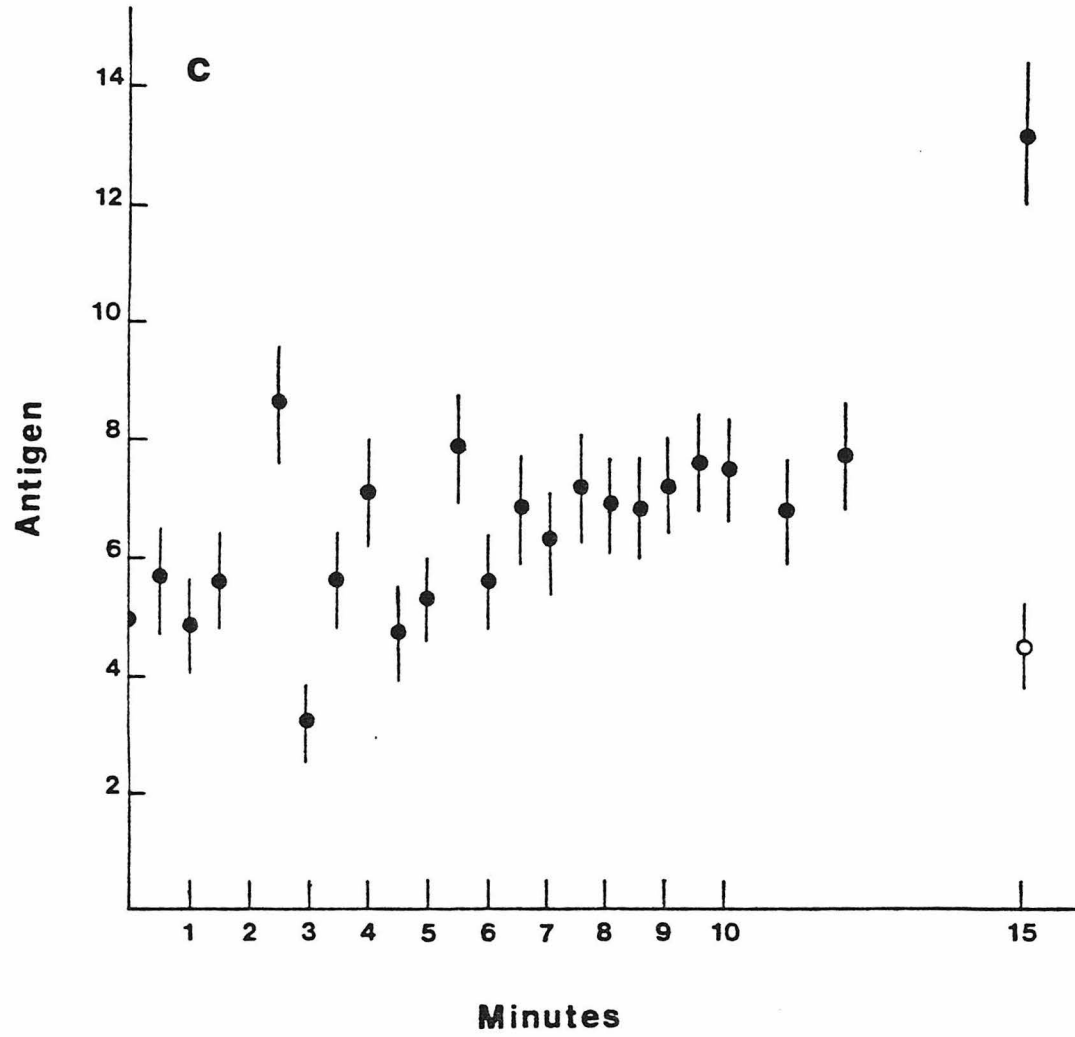
The A antigen produced on shift-down to permissive temperature could be produced either by assembly of gp34 newly synthesized at permissive temperature, or by assembly of gp34 already synthesized at nonpermissive temperature. To discriminate between these possibilities cultures infected with gene 34 ts:amB5 multiple mutants were grown for 20 min at 39.5°C, incubated with chloramphenicol for 2 min, and shifted to 22°C. The formation of A antigen with time at permissive temperature was measured by pipetting aliquots of the temperature-shifted cultures into chloroform. Experiments in which the temperature was measured by thermistor probe showed that the shift to the lower temperature was over 95% complete in 7 sec under these conditions. Pulse labeling followed by electrophoretic analysis showed that no detectable gp34 was synthesized after 2 min incubation with chloramphenicol under these conditions. Figure 8 shows that following chloramphenicol treatment and shift-down, A antigens are produced by tsA44 mutants, but not by tsB3, tsB63 or tsB41 mutants. However the latter display high levels of A antigens at all times examined. To test whether antigen formation was truly blocked in these cultures, or merely occurring at a rate too fast to be measured, we compared the amount of A antigens produced by cultures shifted-down and chloroform-treated with the amount produced by cultures lysed directly at 39.5°C. Table 3 shows there is a slight but not significant difference between these two treatments for tsB63:amB5 phage.

In order to determine whether the tsA44 chloramphenicol-insensitive antigen production was energy-dependent, we incubated infected cells at 39.5°C for 12.5 or 20 min, added chloramphenicol, and then split the cultures in two aliquots, one made 4 mM in KCN and the other untreated. After 2.5 min (cells incubated 12.5 min) or 2 min (cells incubated 20 min) the infected cultures were shifted to 22°C and A antigen production measured as a function of time as before. The results, shown in Figure 9, indicate that A antigen production at 12.5 min is insensitive to cyanide,

Figure 8. Measurement of in vivo rate of A antigen production following shiftdown. Bacteria growing at nonpermissive temperature were infected with selected gene 34 ts:gene t amB5 double mutants. After addition of chloramphenicol to suppress de novo synthesis of gene 34 polypeptide, cultures were shifted to permissive temperature and amount of antigen as a function of time assayed by serum blocking. O, antigen present in unshifted cells. Bars represent standard error of measurement. Panel a, tsA44; panel b, tsB3; panel c, tsB63; panel d, tsB41.







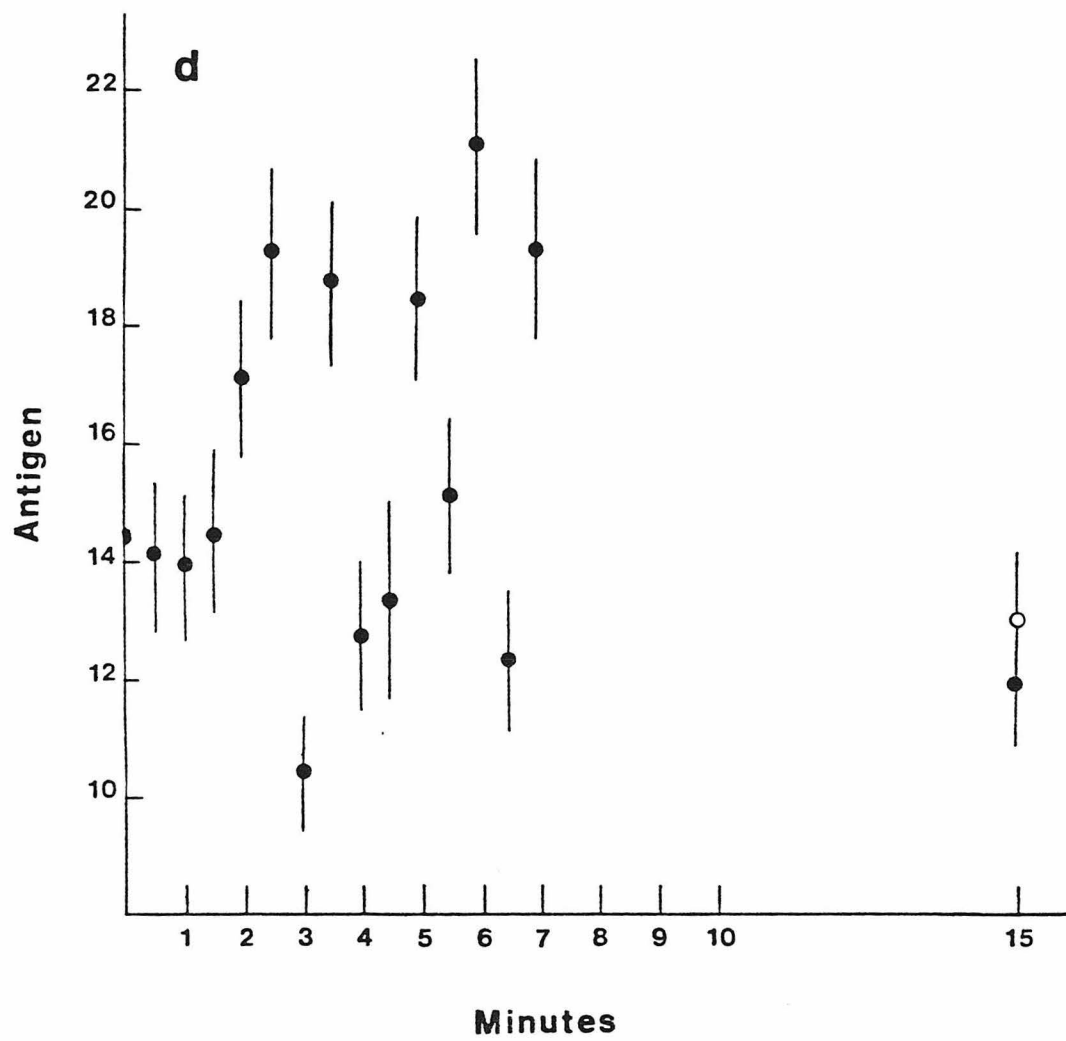
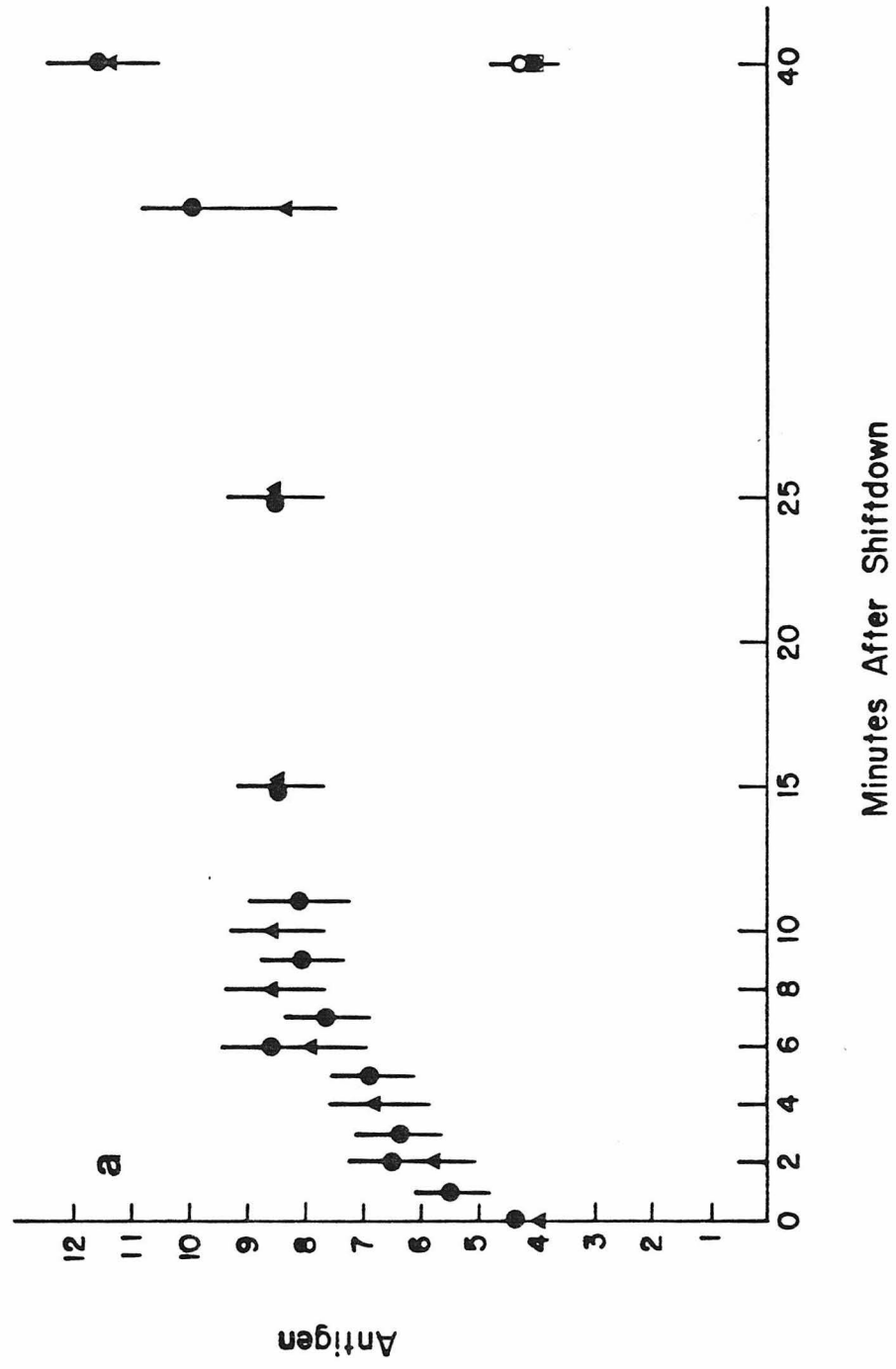


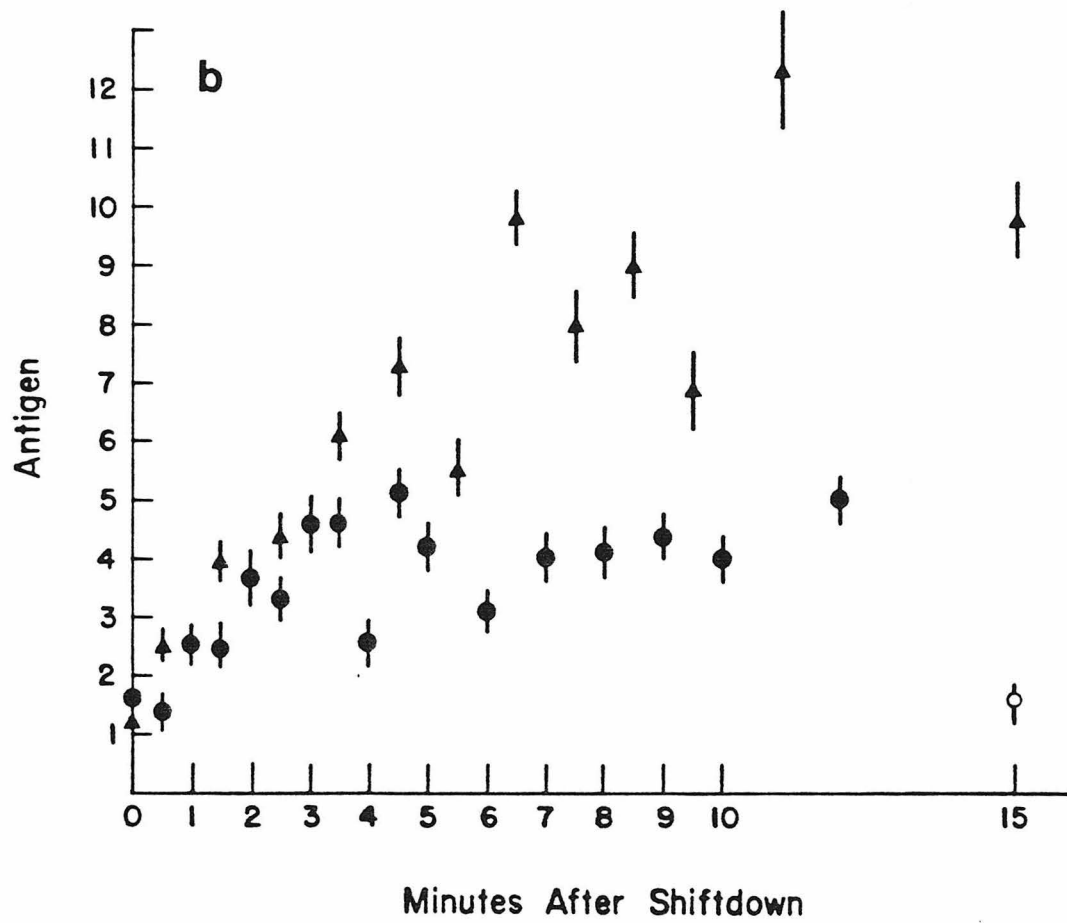
TABLE 3

Measurement of Antigen Formed During the Process of Shiftdown

	Accumulated A antigen after shiftdown	CHCl ₃ -insensitive A antigen	Difference antigen
0'	11.2 $\pm$ 0.8	8.1 $\pm$ 0.7	1.6 $\pm$ 1.5
5'	12.4 $\pm$ 0.8	11.3 $\pm$ 0.8	0.6 $\pm$ 1.6
10'	14.8 $\pm$ 0.9	11.2 $\pm$ 0.8	2.1 $\pm$ 1.7
15'	13.0 $\pm$ 0.8	11.5 $\pm$ 0.8	0.9 $\pm$ 1.6

Figure 9. Cyanide sensitivity of tsA44 antigen formation at 12.5 and 20 min post-infection. tsA44:amB5 infected cells were treated with chloramphenicol alone, or with chloramphenicol followed by cyanide. Production of A antigen following shiftdown was measured by serum blocking. ●, cyanide treated cells; ▲, untreated cells; O, antigen present in unshifted cells exposed to cyanide; □, antigen present in unshifted cells without cyanide. Bars represent standard error of measurement. Panel a, 12.5 min postinfection. Panel b, 20 min.





whereas that at 20 min is at least partially sensitive. To measure the time-course of development of cyanide sensitivity at 39.5°C, infected cells were treated with chloramphenicol at 20 min, with KCN at 21 min, and shifted-down either immediately or at 1 min intervals thereafter. Net antigen production after shift-down was measured by lysing cells 15 min after shift-down. After 3 min of incubation with cyanide at 39.5°C the infected cells were no longer capable of making antigen on shift-down (data not shown). These results resemble cyanide-induced lysis of infected cells. At 30°C such a phenomenon is observed in cells infected by wild-type T4, but not in cells infected by phage containing am mutations in gene t (Josslin, 1971). However gene t am mutations are known to be leaky at elevated temperature (Josslin, 1971). Similar experiments have shown that at 39.5°C, the gene t am mutation is only partially capable of protecting the infected bacterial cell from cyanide-induced lysis (data not shown).

Lysates of gene 34 ts and gene 34 ts:amB5 mutant phage-infected cells were prepared at 39.5°C, centrifuged to eliminate bacterial debris, and sedimented through CsCl density gradients. Phage particles were isolated and titered by optical density. Proximal half-fiber antigen was assayed by serum blocking, and the concentration of infectious particles determined at permissive temperature. The results are shown in Table 4. Lysis inhibition promoted by the gene t amB5 mutation raises both the half-fiber antigen per particle, and the average infectivity. tsB3, tsB63, and tsN1 particles were found to have high titers of proximal half-fiber antigens, but low infectivity. Particles of these phage may be defective in host adsorption when assembled at nonpermissive temperature. Previous results have indicated that tsN1 phage do not attach tail fibers at nonpermissive temperature (King and Wood, 1969). The data presented in Table 4 were obtained from two different tsN1 stocks, one of which was used to obtain the mapping data shown in this paper. The previous work did not measure tail fibers per particle, but rather the presence or absence of antigen sedimenting at the same rate as phage.

TABLE 4

Titers of Proximal Half-Fiber Antigen and Infectious Units per Phage Particle
of Gene 34 ts Mutant Phage at Nonpermissive Temperature

Phage	Proximal half-fibers/particle	% Infectious
<u>ts</u> A44	0.72 $\pm$ 0.12	0.11
<u>ts</u> B63	1.89 $\pm$ 0.13	2.0
<u>ts</u> B63: <u>am</u> B5	2.9 $\pm$ 0.35	8.0
<u>ts</u> B45	0.49 $\pm$ 0.25	.01
<u>ts</u> B45: <u>am</u> B5	2.7 $\pm$ 0.60	1.8
<u>ts</u> B104	1.8 $\pm$ 0.26	7.5
<u>ts</u> B33	0.90 $\pm$ 0.71	0.5
<u>ts</u> B22	0.99 $\pm$ 0.14	0.6
<u>ts</u> B41	0.49 $\pm$ 0.23	0.5
<u>ts</u> B41: <u>am</u> B5	2.7 $\pm$ 0.54	8.5
<u>ts</u> A51	0.62 $\pm$ 0.12	0.14
<u>ts</u> A20	0.87 $\pm$ 0.50	0.7
<u>ts</u> A20: <u>am</u> B5	3.9 $\pm$ 0.35	17
<u>ts</u> B111	1.22 $\pm$ 0.29	1.3
<u>ts</u> B3	2.36 $\pm$ 1.89	3.0
<u>ts</u> B3: <u>am</u> B5	2.4 $\pm$ 0.35	6.5
<u>ts</u> N1	3.38 $\pm$ 1.17	2.5
++	4.66 $\pm$ 0.64	16
<u>am</u> B5	6.9 $\pm$ 0.80	33
<u>am</u> N58	0.18 $\pm$ 0.04	-

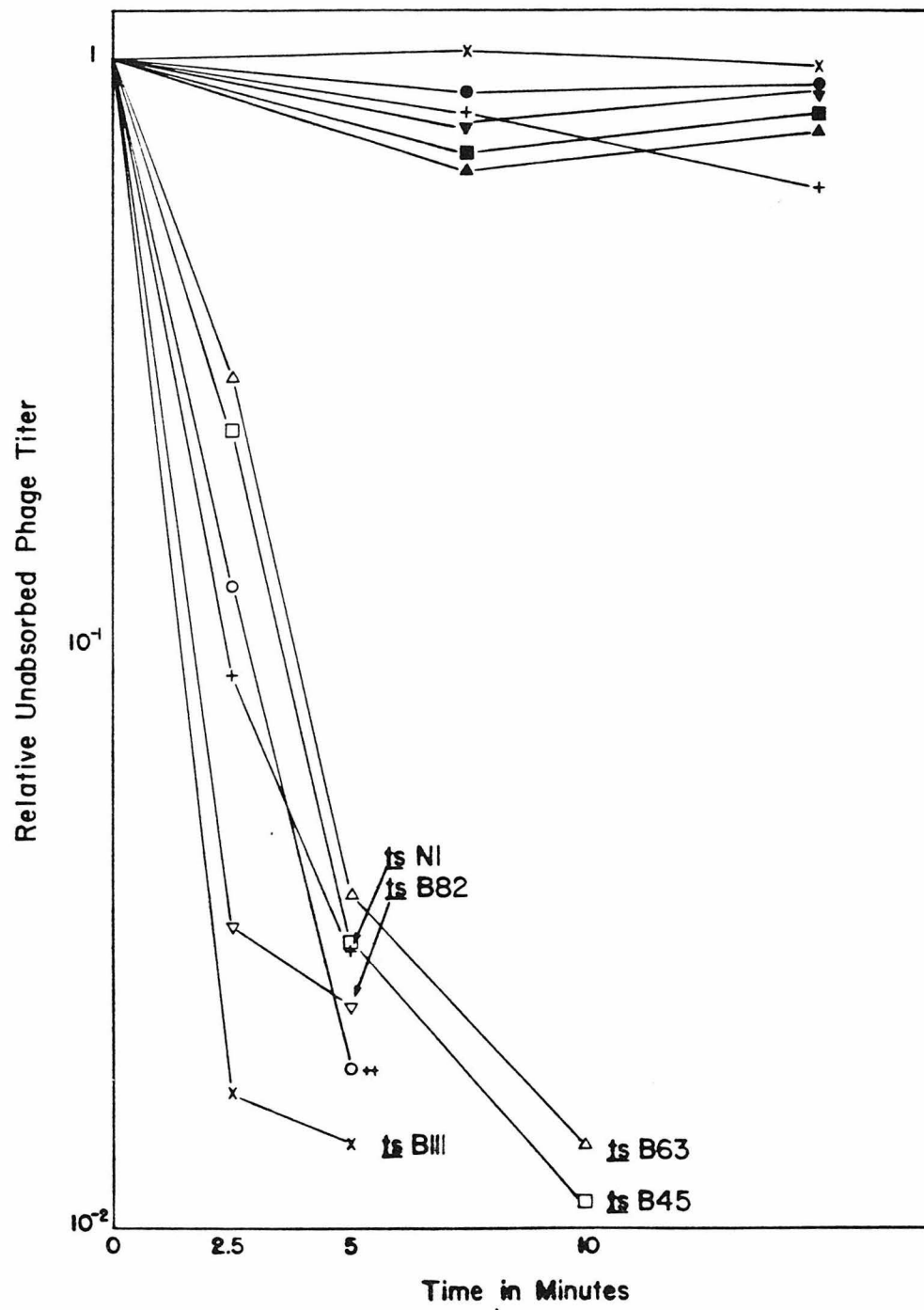
Infectious particles produced at permissive temperature by all gene 34 ts phage tested displayed wild-type host adsorption kinetics at nonpermissive temperature. Selected adsorption curves are presented in Figure 10.

Discussion

The genetic map of bacteriophage T4 differs substantially from that of phages T2 and T6 in the region encompassing genes 34 and 35 (Russell, 1967). The genetic lengths of the T4 genes 34 and 35 are much greater than those of T2 and T6 (Russell, 1967). A map of T4 based upon the physical length of DNA encapsidated in petite phage particles has shown that the genetic lengths of genes 34 and 35 are substantially greater than their physical lengths (Mosig, 1968). In gene 34, a direct comparison of the genetic map distances with the lengths of nonsense-terminated polypeptides has shown that the frequency of recombination per unit length of DNA rises sharply at the distal terminus of the gene (Beckendorf and Wilson, 1972). In this report we characterize a new facet of the genes 34 and 35 site-specific recombination, that of an enhanced recombination frequency observed between pairs of ts mutations, but not between pairs of am mutations, nor between am mutations and ts mutations. Because this phenomenon is confined to gene 34, and because gene 34 ts mutations exhibit intragenic complementation (Bernstein *et al.*, 1965), we suggest that the predominant source of extra ts by ts recombinants is the recombination or repair of parental heterozygotes which are viable by intragenic complementation. The greatest obstacle to this explanation is the low complementation exhibited in liquid culture by phage bearing the tsA44 mutation. The tsA44 mutation is unambiguously involved in ts by ts map expansion. We suggest that low levels of tsA44 complementation in liquid may still support sequential burst (and subsequent plaque) formation on plates. This proposal is not inconsistent with our prior experience with gene 34 ts phenotypes (Seed *et al.*, 1979), particularly in light of the fact that burst sizes

Figure 10. Adsorption of gene 34 ts mutant phage at nonpermissive temperature.

Chloroform resistant titers were measured following incubation of phage with susceptible bacteria at 39.5°C (open symbols). Control titers were measured under identical conditions in the absence of sensitive bacteria (closed symbols).



of liquid complementations involving gene 34 ts's and tsA44 are generally of order unity (Bernstein et al., 1965).

Some of the properties of parental (nonrecombinant) heterozygotes in T4 that have been deduced from studies on recombination in the rIIA cistron (Edgar, 1961) can be applied to the hypothesis above. For example, Edgar (1961) has calculated that in a single round of infection 4.6% of the nonrecombinant heterozygotes created between markers 0.016 map units apart will give rise to recombinant progeny. This figure assumes a mean heterozygosity per locus of 2%. If we assume that the probability of formation of recombinant phage from heterozygotes is, over small distances, a linear function of the distance between markers, then for markers greater than 0.4 map units apart, nearly all heterozygotes bearing both markers should yield recombinant progeny. This map separation is not large enough to significantly decrease the fraction of parental heterozygotes bearing both markers, since the mean length of a heterozygote is estimated to be 5 to 10 map units (Levinthal, 1954). In the rIIA system, 1.6% of all bacteria jointly infected with both parents gave rise to recombinant progeny. This is about one-third the frequency produced by single-infecting parental heterozygotes. It is possible the difference is due to mechanisms of recombination which are independent of length (e.g., excision repair) acting on the heterozygotes. Thus we might not expect all heterozygotes to give rise to recombinant progeny until the distance between markers were three times larger than quoted above, or about 1.2 map units distant. When the parental markers are this far apart, the percent heterozygosis will begin to show the effects of increasing marker separation. As a rough estimate, this may be calculated as $(1-M/L)$, where L is the average length of a heterozygote, and M the distance between markers. For markers separated by 1.2 map units and a mean heterozygote length of 6 map units, this formula predicts a 20% reduction in frequency of parental heterozygotes encompassing both markers.

The distance between the am markers which flank the region of anomalous ts recombination in gene 34 is about 10 map units (Beckendorf and Wilson, 1972). The frequency of recombination over this interval averages about threefold higher than the average for other genes in T4 (Beckendorf and Wilson, 1972). Hence the normalized map distance is about 3.3 units. This is large enough to allow essentially unit efficiency of recombination between markers on infecting parental heterozygotes.

The observed ts map expansion is confined to the promoter-distal region of gene 34. Heterozygotes throughout the gene should be viable by intragenic recombination, and thus contribute to the recombination frequency. Hence it is critical to our explanation that the frequency of formation of parental heterozygotes rise faster than the rate of am recombination in this region. Such a phenomenon could be the consequence of the postulated high efficiency of chromosome pairing at the genes 34 and 35 junction (Beckendorf and Wilson, 1972).

Work with the serologically related phages of the RB family has shown that the A antigen is not a homogeneous repeating structure. In this study we have shown that gene 34 ts mutants synthesize normal amounts of mutant polypeptide chain at nonpermissive temperature. We have also shown that almost all gene 34 ts mutants display a decreased production of A antigen at nonpermissive temperature. The relative proportion of antigens produced remains constant while the net amount of antigen declines. Taken together, the above results indicate that the synthesis of stable A antigen takes place by an all-or-none mechanism.

At present very little can be definitively concluded about the mechanism of A antigen production by wild-type phage. Antigen production appears to be limited by bacterial membrane lysis. Experiments with the mutant tsA44 have shown that early in infection A antigen can be produced on shift-down to permissive temperature despite metabolic inhibition by cyanide. tsA44 antigen production late in infection is apparently inhibited by cyanide, but this is probably attributable to a cyanide-induced lysis of the bacterial cell membrane (Josslin, 1971).

The rate of production of A antigen by tsA44-infected cells upon shift-down to permissive temperature is very slow, probably too slow to be characteristic of an intramolecular isomerization. One logical candidate for the rate-limiting step for antigen formation under these conditions would be a bi- or multi-molecular collision. Such a step could be the initial joining of the gp34 polypeptide chains, or, alternatively, an interaction between gp34 and gp57, the nonstructural assembly protein for gp34.

ts mutants other than tsA44 do not exhibit chloramphenicol-independent production of A antigen upon shift-down to permissive temperature. Since there are large pools of unassembled gp34 in cells infected with these mutants at non-permissive temperature, either some of the mutant gp34 polypeptide chains must be incapable of subsequent antigen formation, or antigen formation must be limited by the synthesis of another protein. The latter case suggests the involvement of a stoichiometric assembly protein utilized less efficiently by ts mutant gp34 than by wild-type gp34. The former suggests an irreversible denaturation of mutant polypeptide chains, perhaps subsequent to defective interaction with an assembly protein.

Other properties than antigen formation are affected by the gene 34 ts mutant phenotypes. For example tsB45 has been shown to confer an abnormal sensitivity to SDS on phage bearing that mutation (Seed *et al.*, 1979). When exposed to 1% SDS at 37°C, the distal half-fibers of tsB45 mutant phage detach from the particle, while the proximal half-fibers remain. tsB33 mutant-infected cell lysates contain a small amount of a rapidly sedimenting antigen specie not found in wild-type lysates.

Virtually all gene 34 ts mutants exhibit decreased levels of proximal half-fiber antigen per phage particle, either in addition to, or in consequence of the defects in antigen formation discussed above. Three exceptional phage appear to be able to attach some proximal half-fiber antigen to phage particles at nonpermissive temperature. It is not known whether these phage also attach distal half-fiber antigen. No evidence

suggests that infectious particles produced at permissive temperature have any defect in adsorption at nonpermissive temperature.

Lysis delay promoted by the gene *t* amB5 mutation raises the mean number of tail fibers per ts phage particle produced at nonpermissive temperature. This may indicate that tail fiber production eventually outstrips particle production late in infection, or that the gene *t* product has a pleiotropic effect altering the relative efficiencies of particle and tail fiber assembly. Raising the number of tail fibers per particle strikingly increases the infectivity of the phage.

References

- Beckendorf, S. K. and Wilson, J. H. (1972). A recombination gradient in bacteriophage T4 gene 34. *Virology* **50**, 315-321.
- Bernstein, H., Edgar, R. S. and Denhardt, D. H. (1965). Intragenic complementation among temperature-sensitive mutants of bacteriophage T4D. *Genetics* **51**, 987-1002.
- Dickson, R. C. (1973). Assembly of bacteriophage T4 tail fibers. IV. Subunit composition of tail fibers and fiber precursors. *J. Mol. Biol.* **79**, 633-647.
- Edgar, R. S. (1961). High negative interference and heterozygosis: A study of the mechanism of recombination in bacteriophage T4. *Virology* **13**, 1-12.
- Edgar, R. S. and Lielausis, I. (1965). Serological studies with mutants of phage T4D defective in genes determining tail fiber structure. *Genetics* **52**, 1187-1200.
- Epstein, R. H., Bolle, A., Steinberg, C. M., Kellenberger, E., Boy de la Tour, E., Chevalley, R., Edgar, R. S., Susman, M., Denhardt, D. H. and Lielausis, A. (1963). Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 375-394.
- Fairbanks, G., Jr., Levinthal, C. and Reeder, R. H. (1965). Analysis of ^{14}C -labeled proteins by disc electrophoresis. *Biochem. Biophys. Res. Comm.* **20**, 393-399.
- Josslin, R. (1971). Physiological studies of the t gene defect in T4-infected Escherichia coli. *Virology* **44**, 101-107.
- King, J. (1968). Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231-262.
- King, J. and Laemmli, U. K. (1971). Polypeptides of the tail fibers of bacteriophage T4. *J. Mol. Biol.* **62**, 465-477.
- King, J. and Wood, W. B. (1969). Assembly of bacteriophage T4 tail fibers: The sequence of gene product interaction. *J. Mol. Biol.* **39**, 583-601.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680-685.

- Levinthal, C. (1954). Recombination in phage T2: Its relationship to heterozygosis and growth. *Genetics* **39**, 169-184.
- Mosig, G. (1968). A map of distances along the DNA molecule of T4. *Genetics* **59**, 137-151.
- Seed, B., Slankard, J., Del Zoppo, G. J., Conley, M. P. and Wood, W. B. (1979). Chapter 3 of this thesis.
- Seed, B. and Wood, W. B. (1979). Chapter 2 of this thesis.
- Steinberg, C. M. and Edgar, R. S. (1962). A critical test of a current theory of genetic recombination in bacteriophage. *Genetics* **47**, 187-208.
- Studier, F. W. (1972). Bacteriophage T7. *Science* **176**, 367-376.
- Takata, R. and Tsugita, A. (1970). Morphogenesis of the tail fibre of bacteriophage T4: Isolation and characterization of a component (the gene 34 product). *J. Mol. Biol.* **54**, 45-57.
- Ward, S. and Dickson, R. C. (1971). Assembly of bacteriophage T4 tail fibers. III. Genetic control of the major tail fiber polypeptides. *J. Mol. Biol.* **62**, 479-492.
- Ward, S., Luftig, R. B., Wilson, J. H., Eddleman, H., Lyle, H. and Wood, W. B. (1970). Assembly of bacteriophage T4 tail fibers. II. Isolation and characterization of tail fiber precursors. Appendix: Analysis of endpoint fiber attachment reactions. *J. Mol. Biol.* **54**, 14-31.
- Wilson, J. H. and Kells, S. (1972). Bacteriophage T4 transfer RNA. I. Isolation of characterization of two phage-coded nonsense suppressors. *J. Mol. Biol.* **69**, 39-56.