A Species Barrier Between Bacteriophages T2 and T4: Exclusion, Join-Copy and Join-Cut-Copy Recombination and Mutagenesis in the dCTPase Genes

Todd P. Gary, Nancy E. Colowick and Gisela Mosig

Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235

ABSTRACT

Bacteriophage T2 alleles are excluded in crosses between T2 and T4 because of genetic isolation between these two virus species. The severity of exclusion varies in different genes, with gene *56*, encoding an essential dCT(D)Pase/dUT(D)Pase of these phages, being most strongly affected. To investigate reasons for such strong exclusion, we have (1) sequenced the T2 gene *56* and an adjacent region, (2) compared the sequence with the corresponding T4 DNA, (3) constructed chimeric phages in which T2 and T4 sequences of this region are recombined, and (4) tested complementation, recombination, and exclusion with gene *56* cloned in a plasmid and in the chimeric phages in *Escherichia coli* CR63, in which growth of wild-type T2 is not restricted by T4. Our results argue against a role of the dCTPase protein in this exclusion and implicate instead DNA sequence differences as major contributors to the apparent species barrier. This sequence divergence exhibits a remarkable pattern: a major heterologous sequence counterclockwise from gene *56* (and downstream of the gene *56* transcripts) replaces in T2 DNA the T4 gene *69*. Gene *56* base sequences bordering this substituted region are significantly different, whereas sequences of the *dam* genes, adjacent in the clockwise direction, are similar in T2 and in T4. The gene *56* sequence differences can best be explained by multiple compensating frameshifts and base substitutions, which result in T2 and T4 dCTPases whose amino acid sequences and functions remain similar. Based on these findings we propose a model for the evolution of multiple sequence differences concomitant with the substitution of an adjacent gene by foreign DNA: invasion by the single-stranded segments of foreign DNA, nucleated from a short DNA sequence that was complementary by chance, has triggered recombination-dependent replication by "join-copy" and "join-cut-copy" pathways that are known to operate in the T-even phages and are implicated in other organisms as well. This invasion, accompanied by heteroduplex formation between partially similar sequences, and perhaps subsequent partial heteroduplex repair, simultaneously substituted T4 gene *69* for foreign sequences and scrambled the sequence of the dCTPase gene *56*. We suggest that similar mechanisms can mobilize DNA segments for horizontal transfer without necessarily requiring transposase or site-specific recombination functions.

ACCUMULATION and fixation of multiple mutations resulting in sequence divergence can ultimately lead to the emergence of new species. Inhibition of recombination because of sequence divergence is a major factor contributing to genetic isolation, that is, species barriers (Radman 1991). SOS-inducible recombination and repair proteins play important opposing roles in such genetic isolation (Vulić *et al.* 1997; Zahrt and Maloy 1997) both in bacteria and in eukaryotes (de Wind *et al.* 1995; Foster *et al.* 1996; Harris *et al.* 1996; Datta *et al.* 1997, and references therein). The accumulation of multiple mutations generating sufficient divergence to establish species barriers is less well understood. It is thought to occur during rare physiological conditions of stress (Rosenberg 1994; Arber 1995; de Wind *et al.* 1995; Rosenberg *et al.* 1995, Hunter *et al.* 1996; LeClerc *et al.* 1996; Matic *et al.* 1997; Torkelson *et al.* 1997, and references therein).

Mutagenesis studies in phage T4, reviewed in detail by Drake and Ripley (1994), have elucidated multiple substrates, pathways, and enzymes that play important roles in generating the raw material for evolutionary changes. It is now obvious that even mutations belonging to a given class, for example, frameshift mutations, large deletions, or duplications, can be generated by more than one pathway. T4 DNA polymerase and its accessory proteins play important roles in the fidelity of all these processes. Potential contributions of recombination to T4 mutagenesis are more ambiguous.

Exquisite and insightful sequence analyses of mutations in genes *e* (lysozyme), *rII*, *ac*, and others have shown that many frameshift mutations, base substitutions, deletions, or duplications can be explained by Streisinger's slipped mispairing model (Streisinger *et al.* 1966) and its variations (Drake and Ripley 1994; Wang and Ripley 1998, accompanying article). Slipped mispairing that is potentially mutagenic can occur between simple or complicated repeats (Pribnow *et al.* 1981), in imperfect palindromes, or between nearly homologous sequences located at considerable distances, that is, so-called ectopic sequences (Ripley 1990). More-

Corresponding author: Gisela Mosig, Department of Molecular Biology, Vanderbilt University, Box 1820 Sta. B, Nashville, TN 37235. E-mail: mosigg@ctrvax.vanderbilt.edu

over, topoisomerase-dependent staggered cuts, with associated synthesis or degradation, can lead to additions or deletions of a few base pairs (Ripley *et al.* 1988; Ripley 1990; Brown *et al.* 1993). Nevertheless, certain frameshift mutations are still unexplained. Drake and Ripley (1994) have discussed the possibility that pairing of T4 DNA with ectopic foreign DNA sequences can generate mutations. This possibility is attractive because there is ever-increasing evidence for transfer of genes or gene segments between phages, resident plasmids, and host genomes, which may contain complete or defective prophages (Hershey 1962; Campbell 1988; Brüssow and Bruttin 1995; Hill *et al.* 1995; Brown *et al* 1996; Kutter *et al.* 1996; Waldor and Mekalanos 1996; Calendar *et al.* 1998; Nelson *et al.* 1997; Tyndall *et al.* 1997; Waldor *et al.* 1997). The evolution of tail fiber genes and the transfer of mobile intron DNA are particularly good examples of such exchanges in T4 (Clyman *et al*, 1994; Henning and Hashemol-Hosseini 1994; Tétart *et al.* 1996). Repoila *et al.* (1994) and Kutter *et al.* (1996) have summarized and discussed the similarities and differences of the many T4-related phages and the apparent horizontal transfer of genes or gene segments to generate their present-day mosaic genomes.

There are now numerous examples demonstrating that moderate sequence divergence poses barriers against genetic exchanges (Radman 1991). Kutter *et al.* (1996) point out that "current concepts of homologous recombination cannot account for the formation of such chimeric genes, and the recombination mechanisms responsible are not known."

Here we have investigated the interrelation of recombination, mutagenesis, and exclusion in the genetic region encompassing gene *56* of the two closely related bacteriophage species, T2 and T4. Because this region has been reported to be important for exclusion of phage T2 by T4 (Streisinger and Weigle 1956; Russell and Huskey 1974; Okker 1981; Okker *et al.* 1981), we suspected that such experiments might yield insights into how species barriers evolve and how they are maintained.

Gene *56* codes for a dCTPase, dCDPase/dUTPase, dUDPase (called dCTPase hereafter), an enzyme essential for normal development of all T-even phages (Kutter and Wiberg 1969). These phages replace deoxycytosine with deoxy-hydroxymethyl-cytosine in their DNA and then glucosylate these residues to various extents. These modifications protect T-even DNA from restriction enzymes that would degrade cytosine-containing DNA (for review see Carlson *et al.* 1994). Gp 56 prevents the incorporation of dC into phage DNA by hydrolyzing dCTP and dCDP; its associated dUTPase activity also prevents incorporation of U instead of T. It has been postulated that T4 dCTPase is responsible for the apparent restriction or exclusion of T2 development by co-infecting T4 phage (Okker 1981; Okker *et al.* 1981).

Electron microscopy of heteroduplexes has revealed a major heterology between T2 and T4 DNA in or near gene *56* (Kim and Davidson 1974; Yee and Marsh 1981). Our comparison of the T2 gene *56* and its adjacent regions with T4 DNA indicates that the major heterology is because of substitution of T4 gene *69* for a completely different sequence in T2. Sequences bordering this heterologous region are partially diverged. They contain, among other features, evolutionary evidence for multiple base substitutions and frameshifts suggestive of slipped mispairings during recombination. We propose a model to explain this unusual pattern based on mispairing of ectopic sequences (Streisinger *et al.* 1966) and recombination-dependent initiation of DNA replication (Luder and Mosig 1982; Mosig *et al.* 1991; Mosig 1994), combined with a strong selection for a functional dCTPase.

To distinguish potential effects of differences between the proteins from those of differences between the base sequences of their genes on present-day exclusion of T2 by T4, we have constructed and tested chimeric T4/T2 derivatives of gene *56.* We found that at least two mechanisms contribute to the apparent exclusion of T2 by T4, one of which depends on yet unknown host functions. Results of our experiments probing the host-independent mechanism, seen in CR63 bacteria, are consistent with the idea that sequence divergence now generates genetic barriers to recombination. Other results argue against a role of the T4 dCTPase protein in either the host-dependent or the host-independent exclusion of phage T2 by T4.

MATERIALS AND METHODS

Bacteriophages are listed in Table 1. The T2 gene *56am* mutation was obtained from S. Hattman in a T2 double mutant α*gt-56am* (Revel *et al.* 1965) that we crossed with wild-type T2 to isolate the single T2 *56am.*

Bacteria: *E. coli* B strains B, B/2, B/4, and S/6 (restrictive for *am* mutants) and the K strain CR63 (*supD*) (permissive for *am* mutants) were initially obtained from A. H. Doermann,

TABLE 1

Bacteriophage strains

Phage		Gene affected	Source
T4	T4D wild type		A. H. Doermann
	amE51	*56*	R. S. Edgar
	amC153	*56*	J. Wiberg
	tsA90	*56*	R. S. Edgar
	amN56	*17*	R. S. Edgar
	amNG178	*17*	F. Arisaka
T2	T2H wild type		H. Revel
	α*gt-56am*	α*gt* and *56*	S. Hattman
	56am	*56*	This work
	T2 (T4 *56+-69−*)	*56*	This work
	T2 (T4 *56+-69+*)	*56* and *69*	This work

then at Vanderbilt University, and have been maintained by us since 1965. B/2 is resistant to T2 and B/4 is resistant to T4. The K strain UT481 (*supD*), permissive for *am* mutants, was constructed and kindly sent by Cynthia Lark, University of Utah, in 1985. The K strain M5219 (λN^+, *cI857*, *sup^o*) and the plasmid vector pPLc2833 (Remaut *et al.* 1981) were kindly supplied by Walter Fiers, University of Ghent, Belgium. The expression plasmid pLAM71* was constructed by Michael Trupin in our lab by standard methods (Sambrook *et al.* 1989) by inserting a 1.9-kb T4 fragment with *in vitro* attached *Bam*HI linkers into the *Bam*HI site downstream of the lambda P_L promoter of plasmid pPLc2833. During or after construction of pLAM71*, T4 gene *69* acquired two mutations: one of five as preceding position 970 (Figure 1B) was deleted, and the C at position 1515 (between two T runs) was changed to T. Thus, pLAM71* contains the wild type T4 gene *56*, gene *69* with a frameshift mutation and parts of *dam* and *soc*. It was used to transform *E. coli* M5219.

Phage crosses were done in bacteria in H-broth at 37°, using equal multiplicities of 3 to 6 of each of two parents, and burst sizes and recombinant frequencies were determined as described (Mosig *et al.* 1977).

Thermocycle DNA sequencing was done without prior amplification as described (Mosig and Colowick 1995), using DNA released from phage and denatured by heating to 95° as template and oligonucleotides end-labeled with [^{32}P]- or [^{33}P]γ-ATP as primers. DNA Inspector and Gene Inspector programs (Textco, West Lebanon, NH) were used to analyze and align DNA sequences. National Center for Biotechnology Information databases were searched with Basic Local Alignment Study Tool and Position Specific Iterated-Basic Local Alignment Study Tool (Altschul *et al.* 1997).

RESULTS

The gene *56* DNA sequences are partially different, and sequences downstream of it are completely different, in T2 and T4: We determined phage T2 DNA sequences directly from virion DNA without prior amplification (Mosig and Colowick 1995). This sequence (Figure 1A) is compared with the corresponding T4 sequence and the deduced open reading frames (ORFs) starting at the −10 region of a middle promoter at position 16.819 of the T4 genome (Macdonald and Mosig 1984; Kutter *et al.* 1994; Figure 1B). Two T4 *am* mutations, and one T2 *am* mutation that we sequenced, confirm the reading frames of gene *56* in T4 and in T2. A T4 *ts* mutation (*A90*) affects a conserved amino acid of gp56.

Gene *69* of T4, located between genes *56* (dCTPase) and *soc* (small outer capsid protein), appears to be replaced by two completely different short ORFs (T2 *soc.1* and T2 *soc.2*) in T2 DNA. Genes *56* and *soc* are partially homologous in the two phages (Figure 2). The deduced amino acid sequences of T4 and T2 dCTPase are 66% identical (Figure 3), and a BLAST search (Altschul *et al.* 1997) shows them to be the closest relatives in the nonredundant GenBank database. Closer inspection of Figure 2 shows that the T2 and T4 sequences are almost identical between positions 1 and 91 (numbering of the T2 sequence), but they diverge considerably beyond this position in a remarkable pattern. All bases apparently

added in T2 (relative to T4) are matched by equal numbers deleted bases nearby, as in compensating frameshift mutations, except for a triplet that adds an entire codon in T2. In four internal segments of gene *56* (positions 140–160, 163–185, 319–349, and 491–519) T2 and T4 sequences are identical for at least 20 bp, the length of perfect homology required in yeast to initiate heteroduplex formation (Datta *et al.* 1997), but less than 50 bp, the apparent size limit for efficient T4 recombination (Goldberg 1966; Bautz and Bautz 1967; Drake 1967; Singer *et al.* 1982; Singer 1988). After the stop codon of gene *56* (position 559) the sequences diverge completely. They converge again at position 948, the first base of the late promoter of *soc* (Macdonald *et al.* 1984). There are several differences within *soc*, but fewer than within gene *56*: the addition of an A at position 962, upstream of the coding sequence; the addition of an A at position 1019 with a compensating deletion of an A at position 1027; and the deletion of GAA at position 1038 in T2 as compared with T4.

A cloned T4 gene *56* complements a T2 gene *56* am mutant without excluding it: Several independent investigators have reported that T4 phage exclude T2 and that different T2 genes are differentially excluded (Streisinger and Weigle 1956; Russell and Huskey 1974; Okker *et al.* 1981). The distinction between T2 and T4 is based on their differential adsorption to B/2 and B/4 bacteria, which are resistant to T2 and T4, respectively. These differences reflect differences in the tail fiber genes and in the molecules used as receptors in the cell wall of different bacterial strains (Henning and Hashemol-Hosseini 1994). Differential exclusion became apparent when T4 *am* mutants in different genes were crossed with T2 containing the corresponding wild-type allele, and the progeny phage were plated on B/2 or B/4 bacteria to distinguish T2 and T4. In this situation there appeared to be fewer T2 than T4 progeny. In general, the apparent exclusion was stronger when the *am* mutations were located in the clockwise direction from the tail fiber genes. This locus-related exclusion has been explained by assuming that splice recombination yields lethal combinations of certain genes whose products interact (Russell and Huskey 1974). However, the correlations are not strict. A few alleles, foremost among them those in gene *56*, appeared to be excluded even from patch recombinants. To explain this effect, Okker *et al.* (1981) have proposed that the dCTPases of T2 and T4 are incompatible and that this incompatibility is responsible for the exclusion.

To test whether the T4 and T2 dCTPase are functionally equivalent, we cloned the wild-type T4 gene *56* in an expression vector (Remaut *et al.* 1981) under control of a temperature-sensitive phage lambda *cI* repressor to give plasmid pLAM71*, and we asked whether it complements a T2 *56am* mutant. In this vector, T4 gene *56* is repressed at 28°, and it can be induced by raising the
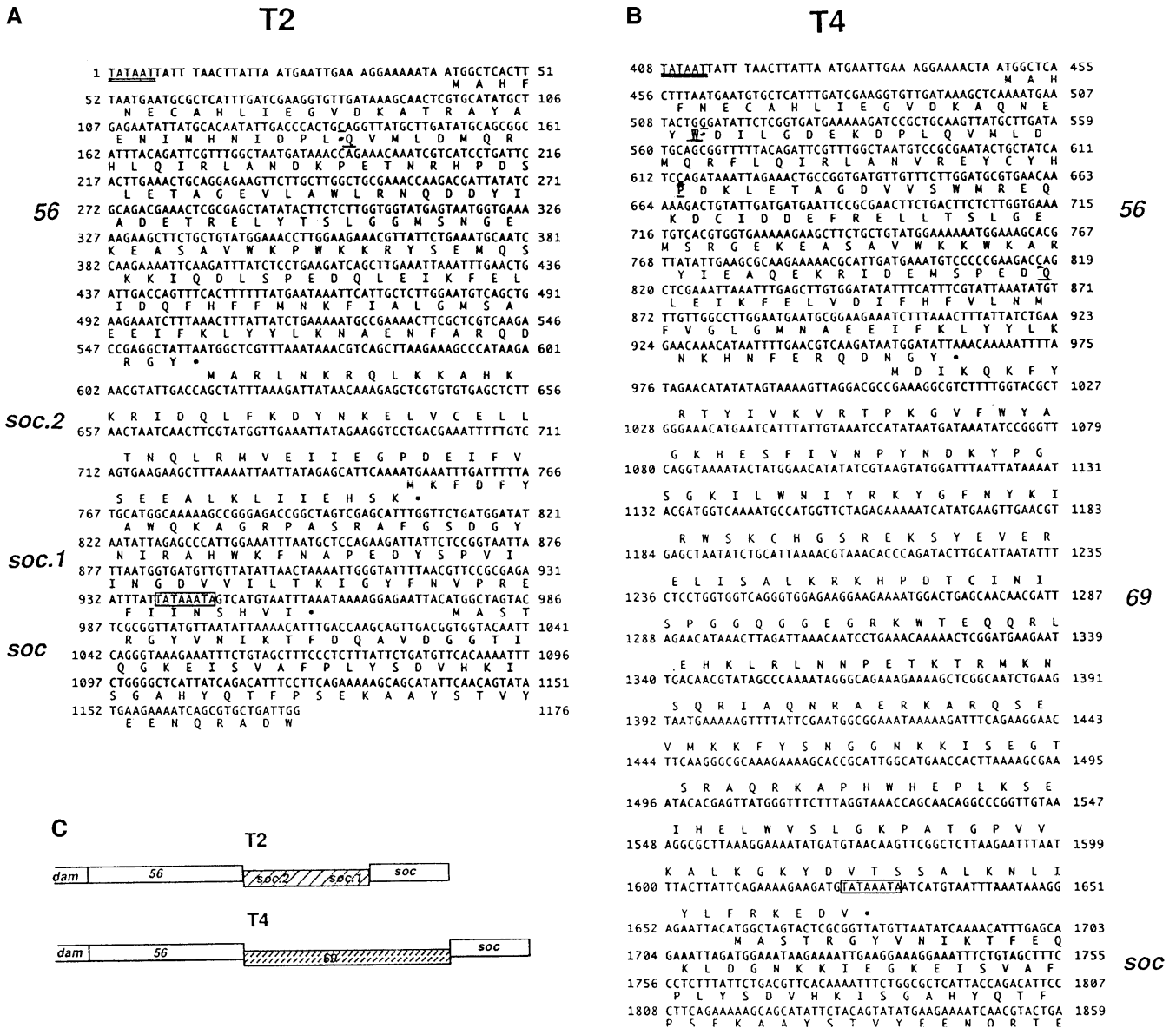
A **T2**

```
   1 TATAATTATT TAACTTATTA ATGAATTGAA AGGAAAAATA ATGGCTCACTT  51
                                                   M  A  H  F
  52 TAATGAATGCGCTCATTTGATCGAAGGTGTTGATAAAGCAACTCGTGCATATGCT 106
      N  E  C  A  H  L  I  E  G  V  D  K  A  T  R  A  Y  A
 107 GAGAATTATGCACAATATTGACCCACTGCAGGTTATGCTTGATATGCAGCGGC   161
      E  N  I  M  H  N  I  D  P  L  Q  V  M  L  D  M  Q  R
 162 ATTTACAGATTCGTTTGGCTAATGATAAACCAGAAACAAATCGTCATCCTGATTC 216
      H  L  Q  I  R  L  A  N  D  K  P  E  T  N  R  H  P  D  S
 217 ACTTGAAACTGCAGGAGAAGTTCTTGCTTGGCTGCGAAACCAAGACGATTATATC 271
      L  E  T  A  G  E  V  L  A  W  L  R  N  Q  D  D  Y  I
 272 GCAGACGAAACTCGCGAGCTATATACTTCTCTTGGTGGTATGAGTAATGGTGAAA 326
      A  D  E  T  R  E  L  Y  T  S  L  G  G  M  S  N  G  E
 327 AAGAAGCTTCTGCTGTATGGAAACCTTGGAAGAAACGTTATTCTGAAATGCAATC 381
      K  E  A  S  A  V  W  K  P  W  K  K  R  Y  S  E  M  Q  S
 382 CAAGAAAATTCAAGATTTATCTCCTGAAGATCAGCTTGAAATTAAATTTGAACTG 436
      K  K  I  Q  D  L  S  P  E  D  Q  L  E  I  K  F  E  L
 437 ATTGACCAGTTTCACTTTTTTATGAATAAATTCATTGCTCTTGGAATGTCAGCTG 491
      I  D  Q  F  H  F  F  M  N  K  F  I  A  L  G  M  S  A
 492 AAGAAATCTTTAAACTTTATTATCTGAAAAATGCCGAAAACTTCGCTCGTCAAGA 546
      E  E  I  F  K  L  Y  Y  L  K  N  A  E  N  F  A  R  Q  D
 547 CCGAGGCTATTAATGGCTCGTTTAAATAAACGTCAGCTTAAGAAAGCCCATAAGA 601
      R  G  Y  *              M  A  R  L  N  K  R  Q  L  K  K  A  H  K
 602 AACGTATTGACCAGCTATTTAAAGATTATAACAAAGAGCTCGTGTGTGAGCTCTT 656
      K  R  I  D  Q  L  F  K  D  Y  N  K  E  L  V  C  E  L  L
 657 AACTAATCAACTTCGTATGGTTGAAATTATAGAAGGTCCTGACGAAATTTTTGTC 711
      T  N  Q  L  R  M  V  E  I  I  E  G  P  D  E  I  F  V
 712 AGTGAAGAAGCTTTAAAATTAATTATAGAGCATTCAAAATGAAATTTGATTTTTA 766
      S  E  E  A  L  K  L  I  I  E  H  S  K  *
                      M  K  F  D  F  Y
 767 TGCATGGCAAAAAGCCGGGAGACCGGCTAGTCGAGCATTTGGTTCTGATGGATAT 821
      A  W  Q  K  A  G  R  P  A  S  R  A  F  G  S  D  G  Y
 822 AATATTAGAGCCCATTGGAAATTTAATGCTCCAGAAGATTATTCTCCGGTAATTA 876
      N  I  R  A  H  W  K  F  N  A  P  E  D  Y  S  P  V  I
 877 TTAATGGTGATGTTGTTATATTAACTAAAATTGGGTATTTTAACGTTCCGCGAGA 931
      I  N  G  D  V  V  I  L  T  K  I  G  Y  F  N  V  P  R  E
 932 ATTTATTTATAAATAGTCATGTAATTTAAATAAAAGGAGAATTACATGGCTAGTAC 986
      F  I  I  N  S  H  V  I  *           M  A  S  T
 987 TCGCGGTTATGTTAATATTAAAACATTTGACCAAGCAGTTGACGGTGGTACAATT 1041
      R  G  Y  V  N  I  K  T  F  D  Q  A  V  D  G  G  T  I
1042 CAGGGTAAAGAAATTTCTGTAGCTTTTCCCTCTTTATTCTGATGTTCACAAAATTT 1096
      Q  G  K  E  I  S  V  A  F  P  L  Y  S  D  V  H  K  I
1097 CTGGGGCTCATTATCAGACATTTCCTTCAGAAAAAGCAGCATATTCAACAGTATA 1151
      S  G  A  H  Y  Q  T  F  P  S  E  K  A  A  Y  S  T  V  Y
1152 TGAAGAAAATCAGCGTGCTGATTGG                              1176
      E  E  N  Q  R  A  D  W
```

56
soc.2
soc.1
soc

B **T4**

```
 408 TATAATTATT TAACTTATTA ATGAATTGAA AGGAAAACTA ATGGCTCA   455
                                                    M  A  H
 456 CTTTAATGAATGTGCTCATTTGATCGAAGGTGTTGATAAAGCTCAAAATGAA    507
      F  N  E  C  A  H  L  I  E  G  V  D  K  A  Q  N  E
 508 TACTGGGATATTCTCGGTGATGAAAAAGATCCGCTGCAAGTTATGCTTGATA    559
      Y  W  D  I  L  G  D  E  K  D  P  L  Q  V  M  L  D
 560 TGCAGCGGTTTTTACAGATTCGTTTGGCTAATGTCCGCGAATACTGCTATCA    611
      M  Q  R  F  L  Q  I  R  L  A  N  V  R  E  Y  C  Y  H
 612 TCCAGATAAATTAGAAACTGCCGGTGATGTTGTTTCTTGGATGCGTGAACAA    663
      P  D  K  L  E  T  A  G  D  V  V  S  W  M  R  E  Q
 664 AAAGACTGTATTGATGATGAATTCCGCGAACTTCTGACTTCTCTTGGTGAAA    715
      K  D  C  I  D  D  E  F  R  E  L  L  T  S  L  G  E
 716 TGTCACGTGGTGAAAAAGAAGCTTCTGCTGTATGGAAAAAATGGAAAGCACG    767
      M  S  R  G  E  K  E  A  S  A  V  W  K  K  W  K  A  R
 768 TTATATTGAAGCGCAAGAAAAACGCATTGATGAAATGTCCCCCGAAGACCAG    819
      Y  I  E  A  Q  E  K  R  I  D  E  M  S  P  E  D  Q
 820 CTCGAAATTAAATTTGAGCTTGTGGTATATATTTCATTTCGTATTAAAATATGT  871
      L  E  I  K  F  E  L  V  D  I  F  H  F  V  I  N  M
 872 TTGTTGGCCTTGGAATGAATGCGGAAGAAATCTTTAAACTTTATTATCTGAA   923
      F  V  G  L  G  M  N  A  E  E  I  F  K  L  Y  Y  L  K
 924 GAACAAACATAATTTTGAACGTCAAGATAATGGATATTAAACAAAAATTTTA   975
      N  K  H  N  F  E  R  Q  D  N  G  Y  *
                                        M  D  I  K  Q  K  F  Y
 976 TAGAACATATATAGTAAAAGTTAGGACGCCGAAAGGCGTCTTTTGGTACGCT  1027
      R  T  Y  I  V  K  V  R  T  P  G  V  F  W  Y  A
1028 GGGAAACATGAATCATTTATTGTAAATCCATATAATGATAAATATCCGGGTT  1079
      G  K  H  E  S  F  I  V  N  P  Y  N  D  K  Y  P  G
1080 CAGGTAAAATACTATGGAACATATATCGTAAGTATGGATTTAATTATAAAAT  1131
      S  G  K  I  L  W  N  I  Y  R  K  Y  G  F  N  Y  K  I
1132 ACGATGGTCAAAATGCCATGGTTCTAGAGAAAAATCATATGAAGTTGAACGT  1183
      R  W  S  K  C  H  G  S  R  E  K  S  Y  E  V  E  R
1184 GAGCTAATATCTGCATTAAAACGTAAACACCCAGATACTTGCATTAATATTT  1235
      E  L  I  S  A  L  K  R  K  H  P  D  T  C  I  N  I
1236 CTCCTGGTGGTCAGGGTGGAGAAGGAAGAAAATGGACTGAGCAACAACGATT  1287
      S  P  G  G  Q  G  G  E  G  R  K  W  T  E  Q  Q  R  L
1288 AGAACATAAACTTAGATTAAACAATCCTGAAACAAAAACTCGGATGAAGAAT  1339
      E  H  K  L  R  L  N  N  P  E  T  K  T  R  M  K  N
1340 TGCACAACGTATAGCCCAAAATAGGGCAGAAAGAAAAGCTCGGCAATCTGAAG  1391
      S  Q  R  I  A  Q  N  R  A  E  R  K  A  R  Q  S  E
1392 TAATGAAAAAGTTTTATTCGAATGGCGGAAATAAAAAGATTTCAGAAGGAAC  1443
      V  M  K  K  F  Y  S  N  G  G  N  K  K  I  S  E  G  T
1444 TTCAAGGGCGCAAAGAAAAGCACCGCATTGGCATGAACCACTTAAAAGCGAA  1495
      S  R  A  Q  R  K  A  P  H  W  H  E  P  L  K  S  E
1496 ATACACGAGTTATGGGTTTCTTTAGGTAAACCAGCAACAGGCCCGGTTGTAA  1547
      I  H  E  L  W  V  S  L  G  K  P  A  T  G  P  V  V
1548 AGGCGCTTAAAGGAAAATATGATGTAACAAGTTCGGCTCTTAAGAATTTAAT  1599
      K  A  L  K  G  K  Y  D  V  T  S  S  A  L  K  N  L  I
1600 TTACTTATTCAGAAAAGAAGATGTATAAAATAATCATGTAATTTAAATAAAGG  1651
      Y  L  F  R  K  E  D  V  *
                                  M  A  S  T  R  G  Y  V  N  I  K  T  F  E  Q
1652 AGAATTACATGGCTAGTACTCGCGGTTATGTTAATATCAAAACATTTGAGCA  1703
      K  L  D  G  N  K  K  I  E  G  K  E  I  S  V  A  F
1704 GAAATTAGATGGAAATAAGAAAATTGAAGGAAAGGAAATTTCTGTAGCTTTC  1755
      P  L  Y  S  D  V  H  K  I  S  G  A  H  Y  Q  T  F
1756 CCTCTTTATTCTGACGTTCACAAAATTTCTGGCGCTCATTACCAGACATTCC  1807
      P  S  F  K  A  A  Y  S  T  V  Y  F  E  N  Q  R  T  E
1808 CTTCAGAAAAAGCAGCATATTCTACAGTATATGAAGAAAATCAACGTACTGA  1859
```

56
69
soc

C

T2

```
dam |    56    |//soc.2////soc.1//|   soc   |
```

T4

```
dam |    56    |///////////69//////////|  soc  |
```

Figure 1.—Base sequences and predicted amino acid sequences of deduced proteins in the region between the middle promoter of gene *56* and the late gene *soc.* (A) T2. (B) T4. (C) A map comparing the gene arrangements in T2 and T4. The −10 regions of the middle promoters upstream of gene *56* are doubly underlined. The late promoter upstream of *soc* is boxed. The two possible T2 ORFs between genes *56* and *soc* are called *soc.1* and *soc.2*. The T2 *56am* mutation at position 137 of the T2 sequence, the T4 *amE51* mutation at position 512, the *tsA90* mutation at position 614 and the *amC153* mutation at position 817 of the T4 sequence are underlined. The numbering of the T4 sequence corresponds to the numbers in our Genbank submission with accession number 30001.

temperature, usually to 42°. Overproduction of gp56 after induction (Gary 1992, data not shown) required co-expression of the lambda *N* (antiterminator) gene, as expected from the regulation of gene *56* in T4-infected cells (Linder and Sköld 1980). The middle T4 promoter upstream of gene *56* is inactive in the plasmid-bearing bacteria. Transcripts that are initiated further upstream, either from an early T4 promoter (Macdonald and Mosig 1984) or from the lambda promoter in pLAM71*, are apparently subject to *rho*-dependent transcription termination (Linder and Sköld 1980)

and require antitermination factors or RNA-stabilizing factors (Mosig and Hall 1994; Stitt and Hinton 1994) to yield sufficient gp56 for phage growth.

The cloned T4 gene *56* produces dCTPase that is active after purification to homogeneity (Gary 1992). It synthesizes sufficient gp56 at 30° to partially complement and allow plaque formation of the T4 gene *56am* mutants *E51* and *C153*. Importantly, it complements the T2 gene *56am* mutant as well (Table 2), indicating that the dCTPases of the two phages are largely compatible with each other. This inference was confirmed by swap-

```
  1  TATAATTATTTAACTTATTAATGAATTGAAAGGAAAAA   38
408  TATAATTATTTAACTTATTAATGAATTGAAAGGAAAAC  445

 39  TAATGGCTCACTTTAATGAATGCGCTCATTTGATCGAA   76
446  TAATGGCTCACTTTAATGAATGTGCTCATTTGATCGAA  483

 77  GGTGTTGATAAAGC----AACTCGTGCATATGCTGAGA  110
484  GGTGTTGATAAAGCTCAAAA----TGAATA--CTG-GG  514

111  ATATT-------ATGCACAATATTGACCCACTGCAGGT  141
515  ATATTCTCGGTGATGAAAAA----GATCCGCTGCAAGT  548

142  TATGCTTGATATGCAGCGGCATTTACAGATTCGTTTGG  179
549  TATGCTTGATATGCAGCGGTTTTTACAGATTCGTTTGG  586

180  CTAATG-------ATAAACCAGAAACAAATCGTCATCC  210
587  CTAATGTCCGCGAATA---CTGCTA-------TCATCC  614

211  TGATTCACTTGAAACTGCAGGAGAAGTTCTTGCTTGGC  248
615  AGATAAATTAGAAACTGCCGGTGATGTTGTTTCTTGGA  652

249  TGCGAAACCAAGACGATTATATCGCAGACGAAACTCGC  286
653  TGCGTGAACAAAAAGACTGTATTGATGATGAATTCCGC  690

287  GAGCTATAT-ACTTCTCTTGGTGTATGAGTAA--TGG   321
691  GAACT-TCTGACTTCTCTTGGTGAAAT--GTCACGTGG  725

322  TGAAAAAGAAGCTTCTGCTGTATGGAAACCTTGGAAGA  359
726  TGAAAAAGAAGCTTCTGCTGTATGGAAAAAATGGAAAG  763

360  AACGTTATTCTGAAATGCAATCCAAGAAAA-----TTC  392
764  CACGTTATATTGAAGCG-----CAAGAAAAACGCATTG  796

393  AAGATTTATCTCCTGAAGATCAGCTTGAAATTAAATTT  430
797  ATGAAATGTCCCCCGAAGACCAGCTCGAAATTAAATTT  834

431  GAACTGATTGACCAGTTTCACTT--TTTTATGAATAAA  466
835  GAGCTTGTGGATATATTTCATTTCGTATTA--AATATG  870

467  TTCATTGCTCTTGGAATGTCAGCTGAAGAAATCTTTAA  504
871  TTTGTTGGCCTTGGAATGAATGCGGAAGAAATCTTTAA  908

505  ACTTTATTATCTGAAAAATGCCGAA---AACTTCGCTC  539
909  ACTTTATTATCTGAAGAA---CAAACATAATTTTGAAC  943

540  GTCAAGACCGAGGCTATTAA  ★  ★  ATAAATAGTC  948
944  GTCAAGATAATGGATATTAA        ATAAATAATC 1633

949  ATGTAATTTAAATAAAAGGAGAATTACATGGCTAGTAC  986
1634 ATGTAATTTAAAT-AAAGGAGAATTACATGGCTAGTAC 1670

987  TCGCGGTTATGTTAATATTAAAACATTTGACCAAGCAG 1024
1671 TCGCGGTTATGTTAATATCAAAACATTTGAGC-AGAAA 1707

1025 TT-GACGGTGGTAC---AATTCAGGGTAAAGAAATTTC 1058
1708 TTAGATGGAAATAAGAAAATTGAAGGAAAGGAAATTTC 1745

1059 TGTAGCTTTCCCTCTTTATTCTGATGTTCACAAAATTT 1096
1746 TGTAGCTTTCCCTCTTTATTCTGACGTTCACAAAATTT 1783

1097 CTGGGGCTCATTATCAGACATTTCCTTCAGAAAAAGCA 1134
1784 CTGGCGCTCATTACCAGACATTCCCTTCAGAAAAAGCA 1821

1135 GCATATTCAACAGTATATGAAGAAAATCAGCGTGCTGA 1172
1822 GCATATTCTACAGTATATGAAGAAAATCAACGTACTGA 1859
```

Figure 2.—Alignments of the DNA sequences of T2 (upper) and T4 (lower) in genes *56* and *soc* that are partially homologous in T2 and T4 DNA. The completely heterologous region between these two genes is omitted and indicated by two asterisks. Mismatches are printed red; unmatched bases are indicated by dashes. In the chimeric phage T2 (T4 *56+ 69−*) the sequence up to position 160 is replaced by T4 sequence.

```
  1  MAHFNECAHLIEGVDKAQNEYWDILGDEKDPLQVMLD   37
  1  MAHFNECAHLIEGVDKATRAYAENIMHNIDPLQVMLD   37

 38  MQRFLQIRLANVR-EYCYHPDKLETAGDVVSWMREQK   73
 38  MQRHLQIRLANDKPETNRHPDSLETAGEVLAWLRNQD   74

 74  DCIDDEFRELLTSLGEMSRGEKEASAVWKKWKARYIE  110
 75  DYIADETRELYTSLGGMSNGEKEASAVWKPWKKRYSE  111

111  AQEKRIDEMSPEDQLEIKFELVDIFHFVLNMFVGLGM  147
112  MQSKKIQDLSPEDQLEIKFELIDQFHFFMNKFIALGM  148

148  NAEEIFKLYYLKNKHNFERQDNGYÖ              172
149  SAEEIFKLYYLKNAENFARQDRGYÖ              173
```

Figure 3.—Alignments of the amino acid sequences of T2 (upper) and T4 (lower) dCTPases. Amino acids that differ are printed in red; an additional amino acid in T2 is matched with a dash in T4.

2 legend). In the other recombinant, T2(T4*56+*-*69+*), T4 genes *56* and *69* have replaced the T2 sequences. [Note that no other T4 sequences, except for some *dam* sequences and three codons of *soc*, could have recombined because they were not present in the plasmid. The T2 (T4*56+* -*69+*) recombinant contained the two gene *69* mutations of pLAM71* described in materials and methods, and it had acquired an additional A in the A run of an untranslated region at positions 1647–1649 (Figure 1B). All three mutations are consistent with Streisinger's model for frameshift mutations.] Both recombinants gave the same plating efficiencies as wild-type T2. We conclude that incompatibility of dCTPases is not the reason for exclusion of T2 by T4 phages.

Exclusion of wild-type T2 by wild-type T4 depends on the host bacteria: During the course of our experiments we found unexpectedly that wild-type T2 is not excluded by wild-type T4 in *E. coli* CR63, a K strain. In agreement with previous reports by other investigators, wild-type T2 is excluded by wild-type T4 in *E. coli* B, which is restrictive for T4 *am* mutants, as well as in the *am* suppressing B *su1* and the K strains UT481 and M 5219, that is, regardless of the presence of amber suppressors and regardless of whether they are B or K strains (Table 3, and data not shown). We surmise that two separate mechanisms contribute to the apparent exclusion of T2 sequences. Hereafter we distinguish them as "restriction" and "exclusion." We suspect that the first mechanism, restriction, depends on host enzymes that probably cooperate with T4 functions. This aspect requires further investigation, which is ongoing. Nevertheless, the lack of restriction in CR63 allowed us to investigate the second mechanism, allele-specific exclusion, without the complications due to restriction.

Sequence divergence excludes T2 gene *56* alleles from the progeny of crosses with T4: Recombination frequencies decline rapidly with decreasing numbers of identical base pairs below a certain minimal length. In T4 this minimal length is approximately 50 bp (Goldberg 1966; Bautz and Bautz 1967; Drake 1967; Singer *et al.* 1982). Figure 2 shows clearly that there is

ping the promoter-proximal and promoter-distal sequences of T2 and T4 gene *56* by recombination and showing that the chimeric gene product is functional *in vivo.* From the progeny of pLAM71* plasmid-bearing bacteria that were infected with the T2 gene *56am* mutant we isolated several T2 *am+* recombinants. Sequencing of the DNA revealed that in one of the recombinants, T2(T4 *56+*-*69−*), there was an exchange within the gene *56* sequence at or before position 160 of the T2 sequence, generating a chimeric dCTPase gene, without substituting T4 gene *69* for the T2 sequence (see Figure

TABLE 2

Burst sizes of T2 and T4 phages from *E. coli*, with or without a cloned T4 gene *56*, at 30°

E. coli	Phage			
	wt T2	wt T4	T4 *amC153*	T2 *56am*
M5219[pLAM 71*] (containing T4 gene *56*)	79	108	42	32
M5219[pLc 2833] (without T4 gene *56*)	80	110	0.1	0.2
B (without T4 gene *56*)	94	130	0.1	ND[a]

ND, not determined.

[a] Does not produce plaques on *E. coli* B.

no segment of 50 identical base pairs in T2 and T4 DNA bracketing a T4 gene *56* mutation. Therefore, we suspected, as proposed as one possible explanation by Russell and Huskey (1974), that exclusion, at least in part, is because of low probabilities of recombination between T4 and T2 sequences in the genetic segment carrying the gene *56am* mutation. The recombination hypothesis predicts that improving the degree of homology should enhance recombination and reduce exclusion. We tested this prediction by crossing T4 gene *56 am* mutants with three different strains: (1) wild-type T2; (2) the T2(T4 *56+69−*) hybrid, which has more homology than wild-type T2 with sequences upstream of, but not adjacent to, the *amber* mutations; and (3) the T2(T4 *56+69+*) hybrid, which has perfect homology around the *amber* mutations except for the mutated sites. These crosses were done in CR63 bacteria (to eliminate the contribution of host-dependent restriction). In support of our hypothesis, the proportions of T4 *am+* progeny (which must have recombined with T2) were 100 times higher in the third set of crosses, in which homology was restored, than in the two former sets (Table 4). In fact, the proportions of T4 *am+* progeny in the third set of crosses were similar to those in crosses with other genes in this area of the genome (Russell and Huskey 1974). The overall proportion of viable *am+* recombinants in crosses of T4 *am* mutants with wild-type T2 from any gene in this area is low, a finding that has been attributed to the inviability of splice recombinants in which incompatible T2 and T4 genes were joined to give lethal combinations (Russell and Huskey 1974). We consider this a satisfactory explanation.

TABLE 3

Percentage of T2 progeny produced in different *E. coli* strains after coinfection with wild-type T2 and T4

Bacteria	Percent T2	Burst Size
E. coli B[(8)], (*sup*[o], B strain)	1.2 ± 0.4	74
E. coli CR63[(3)], (*supD*, K strain)	50.0 ± 5.0	62
E. coli UT481[(2)], (*supD*, K strain)	1.2 ± 0.1	72

[( )], the number of times each cross was performed.

DISCUSSION

Our results can be summarized as follows:

1. A heterologous DNA segment differing in length (Kim and Davidson 1974) and sequence (our results) in T2 and T4 is bracketed by two genes that code for proteins (dCTPase and gp soc) that are similar in the two phages but show considerably more sequence divergence than the *dam* gene immediately upstream. In T4 the heterologous DNA segment encodes gene product 69, which has some resemblance to MobD, a member of a putative mobile endonuclease family of T4 (E. Kutter, personal communication). In T2 the heterologous segment contains two short possible ORFs (T2 *soc.1* and T2 *soc.2*, Figure 1) with no significant similarities with other proteins in the nonredundant Genbank database. We have not yet tested the expression of the two T2 ORFs.

2. No obvious significant inverted and direct repeats, such as those that exist at or near ends of most transposable elements, bracket the heterologous regions. However, there is a complex palindrome, possibly forming a pseudoknot (Figure 4A) near the gene T4 *56/69* junction, and a different palindromic sequence at the junction of *69* and *soc* (Figure 4B). In RNA the latter structure regulates expression of *soc* (Macdonald *et al.* 1984); the former structure is probably involved in ribosomal frameshifting to allow synthesis of a fusion protein combining dCTPase and gp69 of T4 (Mosig and Macdonald 1986; A. Chang, L. Davenport and G. Mosig, unpublished results).

3. In gene *56*, four segments each sharing more than 20 bp identity in T2 and T4 are interrupted by segments with many apparently mismatched or unpaired bases. Apparent additions of multiple bases in one sequence (relative to the other) are compensated by apparent deletions of the same numbers of bases nearby (Figure 2). Nevertheless, the deduced amino acid sequences of T2 and T4 dCTPases (Figure 3) are 66% identical, the genes complement each other, and the two proteins are the closest relatives in the current database.

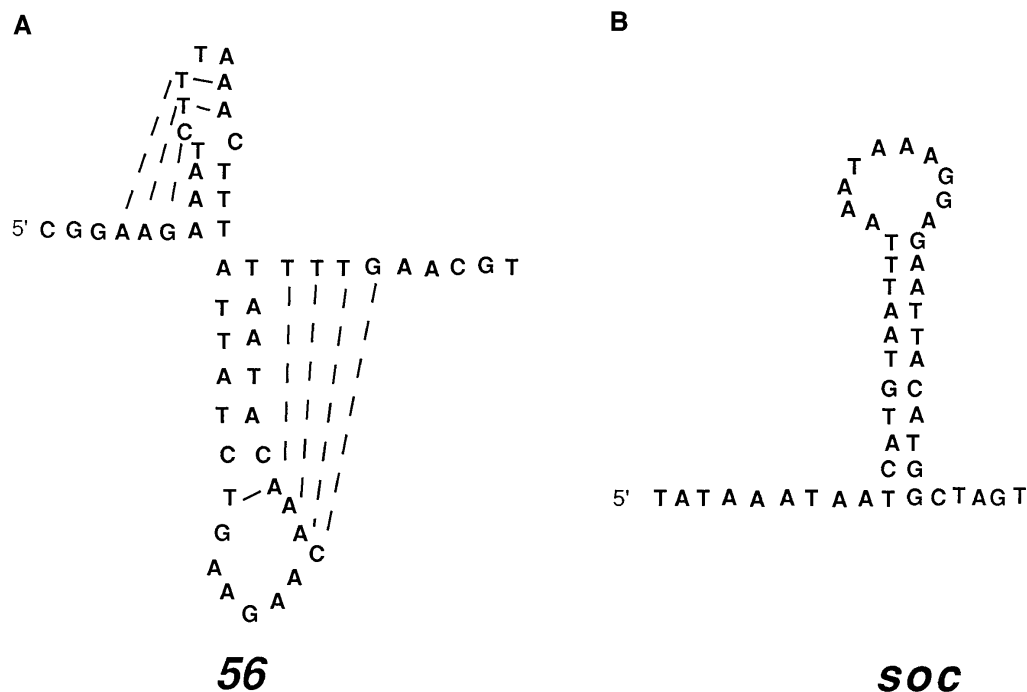4. The extreme exclusion of T2 *am+* alleles by T4 gene

A

```
          T A
        /T—A
       /|T~A
      / |C  C
     / /|T  A
    / / |A  T
   / /  |A  T
5' C G G A A G A T
          A T T T T G A A C G T
        T A    | | | | /
        T A    | | | | /
        A T    | | | | /
        T A    | | | | /
        C C I | | | /
        T‑A  A| | | /
       G     A 'I /
      A     A C
     A    A
      A  G A
       G A
```

56

B

```
        A A A
      T A     A
    A A       G
      A A    G
       A  A
       T   G
       T   A
       T   A
       A   T
       A   T
       T   A
       G   C
       T   A
       A   T
       C   G
5' T A T A A A T A A T G C T A G T
```

soc

Figure 4.—Possible folding into secondary structures of single-stranded nucleic acid of sequences preceding the stop codon of gene *56* (A) and adjacent to the late promoter of *soc* (B), that is, near the junctions of the substituted heterologous T2 and T4 sequences.

56 mutants is relieved ~100-fold (*i.e.*, 100 times more T2 *am⁺* alleles appear among viable T4 progeny) by improving the homology of the DNA surrounding the gene *56* mutations.

Exclusion of T2 gene *56* alleles by T4 is related to sequence divergence: Our results indicate that the severe exclusion of T2 gene *56* in crosses with T4 gene *56am* mutants is in part because of the sequence divergence of these genes. This is consistent with the reduction of recombination when the distance between mismatched base pairs is less than 50 base pairs (Goldberg 1966; Bautz and Bautz 1967; Drake 1967; Singer *et al.* 1982), because restoring sequence identity (except in the mutated site) restores the proportions of the T2 *56⁺* alleles, that is, wild-type recombinants, to the levels seen in neighboring genes.

This interpretation does not exclude the possibility

TABLE 4

Proportion of T4 *am⁺* progeny in crosses of T4 gene *56am* mutants with wild-type or chimeric T2

Phage strains	T4 *am⁺* among total *am⁺* progeny (%)
T4 *amC153* × T2[4]	0.009 ± 0.002
T4 *amC153* × T2(T4*56⁺-69⁻*)[2]	0.011 ± 0.001
T4 *amC153* × T2(T4*56⁺-69⁺*)[3]	1.50 ± 0.1
T4 *amE51* × T2[2]	0.019 ± 0.006
T4 *amE51* × T2(T4*56⁺-69⁻*)[2]	0.014 ± 0.006
T4 *amE51* × T2(T4*56⁺-69⁺*)[2]	2.12 ± 0.05

[], the number of times each cross was performed.

that the product of gene *69* may play additional roles in exclusion or restriction or both in other *E. coli* strains. For example, gp69 might activate a host nuclease or direct it to a T2 sequence. Gp69 does not contain one of the three consensus motifs found in other mobile endonucleases (for reviews see Belfort and Roberts 1997, and references therein), including those predicted for T4 (Gorbalenya 1994; Shub *et al.* 1994; E. Kutter, personal communication). Suspected interactions of gp69 with host functions and its possible role in exclusion require further investigation, which is in progress.

In other systems, recombination between diverged sequences is reduced by mismatch repair proteins (Radman 1991; Datta *et al.* 1997; Kirkpatrick and Petes 1997; Sugawara *et al.* 1997; Vulić *et al.* 1997; Zahrt and Maloy 1997). Recent results of Datta *et al.* (1997) indicate that in yeast ~20 bp of perfect homology are needed to initiate heteroduplex formation and that mismatches within 600 bp or more trigger the antirecombination activities of the mismatch repair enzymes. Our observation that gene *56* T2-T4 recombination is restored only when homology with both genes *56* and *69* is restored supports the notion that similar length constraints exist for T4 recombination.

How did T4 and T2 gene *56* sequences diverge? Electron microscopy of heteroduplexes made *in vitro* from T2 and T4 phage DNA has revealed several so-called substitution loops and many insertion loops (Homyk and Weil 1974; Kim and Davidson 1974; Yee and Marsh 1981). These hetroduplex loops were aligned with the genetic and physical maps by Kutter *et al.* (1996). Insertion loops are readily explained by dele-

tions, insertions, or transpositions of relatively large DNA segments. Acquisition or loss of mobile introns or endonuclease genes (for review see Clyman *et al.* 1994, and Belfort and Roberts 1997) and illegitimate recombination between short ectopic sequence repeats residing in the same genome (*e.g.*, Homyk and Weil 1974; Pribnow *et al.* 1981; Singer 1988; Wu *et al.* 1991; Drake and Ripley 1994; Tétart *et al.* 1996, for review) are prominent examples of sequence rearrangements evident as insertion loops.

The origins of heterologous regions revealed by substitution loops and the remarkable coincident sequence divergence in neighboring genes that we describe here require additional explanations. We surmise that this coincidence is related to the mechanisms that generated such sequence exchanges on an evolutionary scale. A current model of the interrelationship of T4 recombination, recombination-dependent DNA replication (Mosig 1994), and probably DNA packaging (Franklin and Mosig 1996) can be adapted to explain this striking pattern as depicted in Figure 5. For the present discussion, the most important aspects of this model are as follows: (1) the formation of D-loops (displacement loops) by invasion of partially single-stranded regions at ends of DNA molecules into homologous double-stranded DNA; (2) the initiation of DNA replication from an intermediate of recombination by two different modes, that is, from the 3′ end of a single-stranded DNA segment invading a duplex ("join-copy" recombination) to copy the invaded parent (Luder and Mosig 1982), and from a 3′-ended nick in the invaded DNA ("join-cut-copy" recombination) to copy the invading parent and thereby join it covalently to the invaded parent (Mosig *et al.* 1991; Mosig 1994); and (3) enlargements of heteroduplex regions by branch migration. Partial repair may then occur in the heteroduplex regions (Womack 1963; Shcherbakov and Plugina 1991; Shcherbakov *et al.* 1995).

To explain the unusual pattern of sequence divergence between T2 and T4 in the gene *56-69* region, we propose that in the dCTPase gene of an ancestor T-even phage, there was ectopic pairing with a foreign DNA element, which contained a gene with some sequence similarity to the T4 dCTPase gene *56* adjacent to an entirely different gene in the region now occupied by gene *69*. A single-stranded terminus of this element invaded this ancestral gene *56* or *dam* within a short segment of perfect homology. In one direction, for example, toward *dam*, which has nearly identical sequences in T2 and T4 (Miner and Hattman 1988), the 3′ end of the invading foreign element primed leading strand DNA replication, thereby stabilizing the hybrid. In the opposite direction, single-strand assimilation by branch migration generated a heteroduplex with numerous mismatches and bulges. Partial heteroduplex repair may have generated four short homoduplex regions separated by unrepaired heteroduplex segments

in the hybrid region evident in Figure 2. Alternatively, the homoduplex regions were identical in the two parental genes that recombined. We assume that strand assimilation was terminated by an endonuclease cut at the Y junction, and that this cut generated a 3′ end in the invaded DNA. Such a cut might have been triggered by the failure of base-pairing in the completely heterologous region or by the formation of a complex secondary structure, such as the one drawn in Figure 4A, in the displaced single-stranded DNA near the junction, or both. From this 3′ end in the invaded strand DNA synthesis was initiated, copying the unassimilated segment of the invading strand, and thereby covalently joining the ancestral T4 gene *56* sequence and a copy of the completely heterologous sequence of the invading element, for example, gene *69*. Similar series of recombinational events may have occurred on the other side of the heterologous region of the foreign element, that is, in the *soc* region (Figure 2). The recombination intermediate, still containing heteroduplex regions, was subsequently replicated, and the unrepaired different versions of the heteroduplex sequences segregated. The resulting sequence heterogeneity now contributes to the apparent present-day partial barrier between these two phage species. The initial pairing could have been initiated by the recombinogenic ends of T-even chromosomes (Doermann and Boehner 1963; Mosig 1963; Womack 1963; Mosig *et al.* 1971), whose ends become single-stranded as a result of replication (Luder and Mosig 1982; Dannenberg and Mosig 1983; Mosig 1987) or by endonucleolytic cuts in or near the homologous DNA (Clyman *et al.* 1994; Kreuzer and Morrical 1994; Mosig 1994; Kreuzer *et al.* 1995; George and Kreuzer 1996).

In a mirror image of this model, a T4 3′ single-stranded end would at first invade the promoter-distal end of the ancestral gene *56* and initiate DNA synthesis in the direction of the ancestral *soc.1* and *soc.2*. The latter possibility cannot be excluded but we consider it less likely, mainly because there is less sequence identity, necessary to initiate the invasion, between T2 and T4 at the promoter-distal segment than at the promoter-proximal segment of gene *56*, and also because the T2-T4 sequence differences in *soc* are larger than in *dam*, but smaller than in gene *56*.

We surmise that formation of mismatched heteroduplexes and partial repair, together with the selective pressure in the T-even phages for a functional dCTPase, have led to the remarkable divergence of the T2 and T4 dCTPase genes *56*.

More complicated models can be envisioned and cannot be excluded. However, the model described here provides the simplest interpretation within the framework of recombination and replication mechanisms that have been shown to operate in T-even phages and that are now implicated in other organisms as well (Haber 1995; Lloyd and Low 1996; Kogoma 1997). We suspect
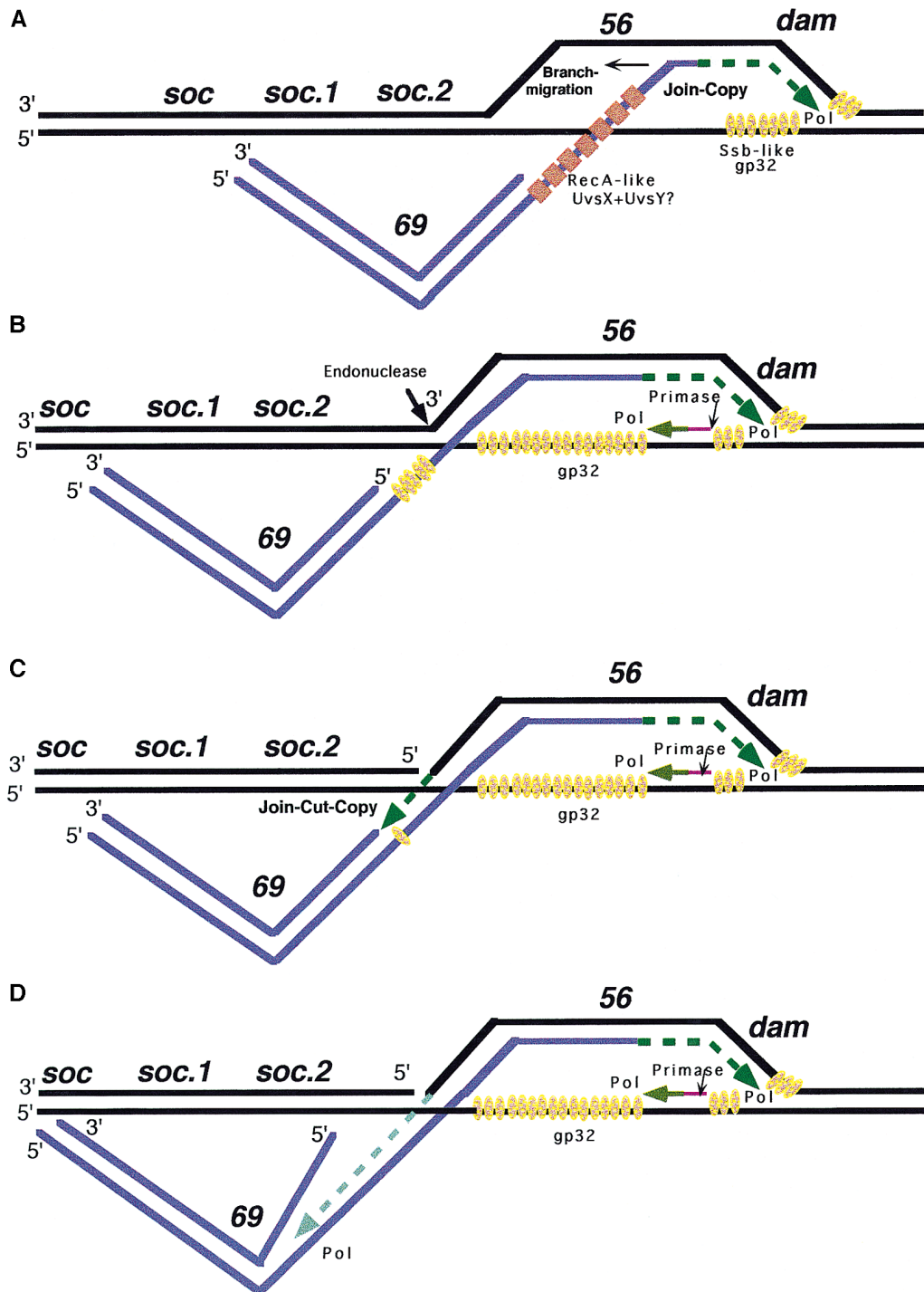
Figure 5.—A model to explain the sequence divergence of T2 and T4 genes *56* adjacent to the substituted regions and the conservation of the adjacent *dam* genes (A–D). A single-stranded DNA segment (generated by replication or by partial digestion), homologous to *dam* or to the beginning of gene *56*, invades homologous double-stranded DNA. For reasons of simplicity, we have drawn the invaded molecule (ancestor of T2) straight and the invading molecule containing gene *69* with a bend. (The reciprocal situation would give equivalent exchange patterns.) The initial pairing region can be short, because DNA replication initiated from the 3′ end of this invading single-stranded segment can stabilize the hybrid and can eventually copy all DNA in the rightward direction. Because the *dam* genes of T2 and T4 are nearly identical (Miner and Hattman 1988), we cannot determine exactly where the invasion occurred. In the leftward direction the invading single strand of the other DNA can be further assimilated into the duplex by branch migration, generating several mismatches. If there are extensive mismatches, partial heteroduplex repair may convert some gene *56* sequences to homoduplexes, leaving intermittent unrepaired sequences as heteroduplexes. Subsequent replication will fix these sequence alterations. In an essential gene, such as gene *56*, only those scrambled sequences that still encode a functional gene product can survive.

A ●●●●●●●●●●AC●AAAAAGTCCATCA
　●●●●●●●●●●TG●TTTTTCAGGTAGT

　●●●●●●●●●●AC●AAAAAGTCCATCA
　●●●●●●●●●●TGTTTT————➤

　●●●●●●●●●●AC●AAAAAGTCCATCA
　●●●●●●●●●●TGTTTTTTCAGGTAGT

— — — — — — — — — — — — — — — — — — — —

B ●●●●◄——ooACoAAAAAGTCCAooo
　●●●●●●●●●●TGTTTT————➤Tooo

　●●●●●●●●ooACoAAAAAGTCCAooo
　●●●●●●●●●TGTTTTTTCAGGTooo

— — — — — — — — — — — — — — — — — —

C ABCD   XYZAB
　[████████████████████]
　ABCD   XYZa'b

Figure 6.—Streisinger's model for frameshift mutations (from Streisinger *et al.* 1966). (A) Origin of a frameshift mutation at the end of a molecule. Line 1 shows the normal end of a molecule, line 2 shows an end in which one chain has been digested by an exonuclease followed by mispairing, and line 3 shows the appearance of the molecule after resynthesis of the digested chain. (B) Origin of a frameshift mutation in a heterozygous region. The lengths of the various regions of overlap are meant to be indicated schematically only. The contributions of the two parental DNA molecules to the heterozygote are distinguished by light *vs.* heavy print and newly synthesized material is indicated by smaller print. The heterozygote is shown as a joint molecule with a set of mispaired bases in the first line and, after synthesis, as a hybrid molecule in which a mutation has occurred, in the second line. (C) A terminal redundancy heterozygote of T4 that is also heteroduplex for a frameshift mutation and for fictitious markers at the same end of the molecule.

that different, apparently illegitimate, recombinations between different segments of ancient T-even genomes, evident from the present-day divergence (Kim and Davidson 1974; Henning and Hashemol-Hosseini 1994; Repoila *et al.* 1994; Kutter *et al.* 1996), may have occurred, perhaps in part at different times, by similar mechanisms.

The results of our T2-T4 crosses are consistent with evidence from other organisms (Radman 1991; Rayssiguier *et al.* 1991) that, although most mismatched heteroduplex regions *in vivo* are aborted, such barriers to recombination break down occasionally, allowing sequence divergence and evolution of new species barriers (Arber 1995).

Mutagenic consequences of slipped mispairing in recombinational heteroduplex regions: Streisinger's model revisited: The alignments of T2 and T4 genes *56* (Figure 2) suggest that one coding sequence for functioning dCTPase was converted into another by multiple com-

pensating frameshift mutations in addition to simple base substitutions. In other words, all additions or deletions of base pairs in T2 (as compared with T4) that are not multiples of three are compensated by nearby deletions or additions, respectively, of the same number of base pairs, restoring the reading frame. Our explanation for this pattern of mispairing of partially homologous sequences and partial heteroduplex repair of the apparently mispaired segments is based on an original proposal by Streisinger *et al.* (1966) to explain acridine-induced frameshift mutations (Figure 6). The diagrams in Streisinger's original paper (Streisinger *et al.* 1966), as well as my perhaps faded memory of conversations with George Streisinger, indicate that he assumed that both erroneous slipping of DNA polymerases during DNA replication and mispairing in recombinational intermediates could cause bulges in mispaired heteroduplexes as precursors of frameshift mutations. Streisinger's original model pictured bulging by misalignment of short simple repeats. Ripley *et al.* (1988), Ripley (1990), and Wang and Ripley (1998) (accompanying article) have demonstrated directly that misaligned bases in potential stem-loop structures with no direct repeats are also hot spots of frameshift mutations. Taken together, these results suggested that bulged mispairing of any single-stranded DNA with any complementary sequence, regardless of the reasons for single-strandedness, can generate mispaired heteroduplexes and thus be mutagenic (Drake and Ripley 1994). The keen insights of George Streisinger and Jan Drake have provided the framework for our current understanding of mutagenesis in general and in phage T4 specifically.

Because the T-even DNA polymerase is by far the major determinant of fidelity (Reha-Krantz 1994; Dressman *et al.* 1997), one might assume that most mutations arise as consequences of errors in replication. However, DNA polymerase binding to the 3′ ends of invading single strands (Luder and Mosig 1982) may also excise misaligned bases in recombinational heteroduplexes with its proofreading exonuclease activity (Bloom *et al.* 1994; Reha-Krantz 1994; Otto 1997).

Drake's pioneering experiments on frameshift mutations in T4 (Drake 1966) contributed significantly to formulating Streisinger's model. Moreover, results obtained by Drake and his collaborators (Drake 1966; Lindstrom and Drake 1970) showed clearly that in T4 most acridine-induced frameshift mutations are first evident as heterozygotes (Figure 6C) and that they do not occur at random sites, but preferentially near tips of T4 chromosomes, which are randomly permuted over the T4 map (Mosig 1963, 1968; Streisinger 1966; Mosig *et al.* 1971). To explain otherwise paradoxical results, particularly the differences in mutagenesis spectra of different acridine derivatives in different organisms, Jan Drake had already suggested at that time that certain enzymes might actively participate in acridine-induced mutagenesis (Lindstrom and Drake 1970). This sug-

gestion is now generally accepted and has been confirmed for T4 by two lines of evidence:

1. Amsacrine-resistant T4 mutations occur in a T4 topoisomerase gene (Huff *et al.* 1990), and certain hot spots of frameshift mutations correspond to topoisomerase cutting sites (Ripley *et al.* 1988; Brown *et al.* 1993). Moreover, certain enzyme-specific acridines that affect the cleavable intermediates are specific frameshift mutagens (Howard *et al.* 1994).

2. The existence of acridine-resistant gene *17* mutants (Piechowski and Susman 1967) indicates that the large subunit of the T4 terminase (gp17) is an additional target of acridines. This finding may be related to the preferred generation of frameshift mutations as heteroduplexes near chromosomal tips (Figure 6C), (Lindstrom and Drake 1970), because our recent results indicate that the largest of the several products of gene *17* binds preferentially to single-stranded DNA segments at junctions with double-stranded DNA, junctions that occur in recombinational intermediates (Figure 5) (Franklin and Mosig 1996; Franklin *et al.* 1998). If unrepaired heteroduplexes in the double-stranded segments of these intermediates contain bulged mismatches, perhaps stabilized by acridines, they would be packaged as terminal redundancy heteroduplex heterozygotes (Figure 6C), exactly as found by Lindstrom and Drake (1970).

CONCLUSIONS

Our results have shown that the partial species barrier between T2 and T4 genes *56*, seen even in hosts in which wild-type T2 is not restricted by T4, is because of DNA sequence divergence, which limits recombination between these genes. In spite of the DNA sequence divergence and corresponding multiple amino acid changes, the products of gene *56* of T2 and T4 (dCTPases) are functionally interchangeable. Our results show that evolutionary exchanges of large heterologous segments are correlated with unusual patterns of sequence divergence in adjacent genes. Most remarkably, this divergence does not extend into the promoter-proximal segment of gene *56* or into the upstream *dam* gene (Miner and Hattman 1988). These results have led us to propose a model in which evolution of such extensive DNA sequence divergence is a direct consequence of illegitimate pairing between heterologous DNA segments containing only a few complementary bases, and initiation of DNA replication from such a recombinational intermediate. This model can explain the mosaic arrangements of other genes in different phage genomes mentioned earlier as well as in bacterial genomes (Falkow 1996; LeClerque *et al.* 1996; Milkman 1996; Nelson *et al.* 1997; Tyndall *et al.* 1997). It suggests one general and potentially mutagenic way to accomplish horizontal

gene transfer, independent of transposons, by a variation on a theme of homologous recombination. Incidentally, our results and interpretations have obvious consequences for constructions of phylogenetic trees based on sequence divergence, and the apparent different tempos of evolution of different genes in the same organisms.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang et al., 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**: 3389–3402.

Arber, W., 1995 The generation of variation in bacterial genomes. J. Mol. Evol. **40**: 7–12.

Bautz, F. A., and E. K. F. Bautz, 1967 Transformation in phage T4: minimal recognition length between donor and recipient DNA. Genetics **57**: 887–895.

Belfort, M., and R. J. Roberts, 1997 Homing endonucleases: keeping the house in order. Nucleic Acids Res. **25**: 3379–3388.

Bloom, L. H., M. R. Otto, R. Eritja, L. J. Reha-Krantz, M. F. Goodman et al., 1994 Pre-steady state kinetic analysis of sequence dependent nucleotide excision by the 3′-exonuclease activity of bacteriophage T4 DNA polymerase. Biochemistry **33**: 7576–7586.

Brown, H. J., H. W. Stokes and R. M. Hall, 1996 The integrons In0, In2, and In5 are defective transposon derivatives. J. Bacteriol. **178**: 4429–4437.

Brown, M. D., L. S. Ripley and D. H. Hall, 1993 A proflavin-induced frameshift hotspot in the thymidylate synthase gene of bacteriophage T4. Mutat. Res. **286**: 189–197.

Brüssow, H., and A. Bruttin, 1995 Characterization of a temperate *Streptococcus thermophilus* bacteriophage and its genetic relationship with lytic phages. Virology **212**: 632–640.

Calendar, R., S. Yu, H. Myung, V. Barreiro, R. Odegrip et al., 1998 The lysogenic conversion genes of coliphage P2 have unusually high AT content, pp. 241–252 in *Symposium on Horizontal Gene Transfer*, edited by M. Syvaanen. Chapman Hall, London.

Campbell, A. C., 1988 Phage evolution and speciation, pp. 1–14 in *The Bacteriophages*, edited by R. Calendar. Plenum Press, New York.

Carlson, K., E. Raleigh and S. Hattman, 1994 Restriction and modification, pp. 369–381 in *Molecular Biology of Bacteriophage T4*, edited by J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Clyman, J., S. Quirk and M. Belfort, 1994 Mobile introns in the T-even phages, pp. 83–88 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Dannenberg, R., and G. Mosig, 1983 Early intermediates in bacteriophage T4 DNA replication and recombination. J. Virol. **45**: 813–831.

Datta, A., M. Hendrix, M. Lipsitch and S. Jinks-Robertson, 1997 Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc. Natl. Acad. Sci. USA **94**: 9757–9762.

de Wind, N., M. Dekker, A. Berns, M. Radman and H. te Riele, 1995 Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell **82**: 321–330.

Doermann, A. H., and L. Boehner, 1963 An experimental analysis of bacteriophage T4 heterozygotes. I. Mottled plaques from crosses involving six *rII* loci. Virology **21**: 551–567.

Drake, J. W., 1966 Heteroduplex heterozygotes in bacteriophage T4 involving mutations of various dimensions. Proc. Natl. Acad. Sci. USA **55**: 506–512.

Drake, J. W., 1967 The length of the homologous pairing region for genetic recombination in phage T4. Proc. Natl. Acad. Sci. USA **58**: 962–966.

Drake, J. W., and L. S. Ripley, 1994 Mutagenesis, pp. 98–124 in *Molecular Biology of Bacteriophage T4*, edited by J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Dressman, H. K., C.-C. Wang, J. D. Karam and J. W. Drake, 1997 Retention of replication fidelity by a DNA polymerase functioning in a distantly related environment. Proc. Natl. Acad. Sci. USA **94**: 8042–8046.

Falkow, S., 1996 The evolution of pathogenicity in *Escherichia, Shigella, and Salmonella*, pp. 2723–2729 in *Escherichia coli and Salmonella*, edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press, Washington, D.C.

Foster, P. L., J. M. Trimarchi and R. A. Maurer, 1996 Two enzymes, both of which process recombination intermediates, have opposite effects on adaptive mutation in *Escherichia coli*. Genetics **142**: 25–37.

Franklin, J. G., and G. Mosig, 1996 Expression of the bacteriophage T4 DNA terminase genes *16* and *17* yields multiple proteins. Gene **177**: 179–189.

Franklin, J. L., D. Haseltine, L. Davenport, and G. Mosig, 1998 The largest (70kDa) product of the bacteriophage T4 DNA terminase gene *17* binds to single-stranded DNA segments and digests them towards junctions with double-stranded DNA. J. Mol. Biol. **277**: (in press).

Gary, T. P., 1992 The dCTPase/dUTPase genes of bacteriophages T2 and T4 and a test of their role in the exclusion of T2 by T4. Ph.D. Dissertation, Vanderbilt University, Nashville, TN.

George, J. W., and K. N. Kreuzer, 1996 Repair of double-strand breaks in bacteriophage T4 by a mechanism that involves extensive DNA repair. Genetics **143**: 1507–1520.

Goldberg, E. B., 1966 The amount of DNA between genetic markers in phage T4. Proc. Natl. Acad. Sci. USA **56**: 1457–1463.

Gorbalenya, A. E., 1994 Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. Protein Science **3**: 1117–1120.

Haber, J. E., 1995 In vivo biochemistry: physical monitoring of recombination induced by site specific endonucleases. Bioessays **17**: 609–620.

Harris, R. S., K. J. Ross and S. M. Rosenberg, 1996 Opposing roles of the Holliday junction processing systems of *Escherichia coli* in recombination-dependent adaptive mutation. Genetics **142**: 681–691.

Henning, U., and S. Hashemol-Hosseini, 1994 Receptor recognition by T-even-type coliphages, pp. 291–298 in *Molecular Biology of Bacteriophage T4*, edited by J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Hill, C. W., G. Feulner, M. S. Brody, S. Zhao, A. B. Sadosky et al., 1995 Correlation of *Rhs* elements with *Escherichia coli* population structure. Genetics **141**: 15–24.

Homyk, T., and J. Weil, 1974 Deletion analysis of two nonessential regions of the T4 genome. Virology **61**: 505–523.

Howard, M. T., S. H. Neece, S. W. Matson and K. N. Kreuzer, 1994 Disruption of a topoisomerase-DNA cleavage complex by a DNA helicase. Proc. Natl. Acad. Sci. USA **91**: 12031–12035.

Huff, A. C., R. E. Ward and K. N. Kreuzer, 1990 Mutational alteration of the breakage/resealing subunit of bacteriophage T4 DNA topoisomerase confers resistance to antitumor agent m-AMSA. Mol. Gen. Genet. **221**: 27–32.

Hunter, N., S. R. Chambers, E. J. Louis and R. H. Borts, 1996 The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. **15**: 1726–1733.

Kim, J.-S., and N. Davidson, 1974 Electron microscope heteroduplex study of sequence relations of T2, T4, and T6 bacteriophage DNAs. Virology **57**: 93–111.

Kirkpatrick, D. T., and T. D. Petes, 1997 Repair of DNA loops involves DNA mismatch and nucleotide-excision repair proteins. Nature **387**: 929–931.

Kogoma, T., 1997 Stable DNA replication: interplay between DNA replication, homologous recombination and transcription. Microbiology and Molecular Biology Reviews **61**: 212–238

Kreuzer, K. N., and S. W. Morrical, 1994 Initiation of DNA replication, pp. 28–42 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Kreuzer, K. N., M. Saunders, L. J. Weislo and H. W. E. Kreuzer, 1995 Recombination-dependent DNA replication stimulated by double-strand breaks in bacteriophage T4. J. Bacteriol. **177**: 6844–6853.

Kutter, E., and J. S. Wiberg, 1969 Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4. J. Virol. **4**: 439–453.

Kutter, E., T. Stidham, B. Guttman, E. Kutter, D. Batts et al., 1994 Genomic map of bacteriophage T4, pp. 491–519 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. AMS Press, Washington, DC.

Kutter, E., K. Gachechiladze, A. Poglazov, E. Marusich, M. Shneider et al., 1996 Evolution of T4-related phages. Virus Genes **11**: 213–225.

LeClerc, J. E., B. Li, W. L. Payne and T. A. Cebula, 1996 High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274**: 1208–1211.

Linder, C. H., and O. Sköld, 1980 Control of early gene expression of bacteriophage T4: involvement of the host *rho* factor and the *mot* gene of the bacteriophage. J. Virol. **33**: 724–732.

Lindstrom, D. M., and J. W. Drake, 1970 The mechanics of frameshift mutagenesis in bacteriophage T4: the role of chromosome tips. Proc. Natl. Acad. Sci. USA **65**: 617–624.

Lloyd, R. G., and K. B. Low, 1996 Homologous recombination, pp. 2236–2255 in *Escherichia coli and Salmonella*, edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press, Washington, DC.

Luder, A., and G. Mosig, 1982 Two alternative mechanisms for initiation of DNA replication forks in bacteriophage T4: priming by RNA polymerase and by recombination. Proc. Natl. Acad. Sci. USA **79**: 1101–1105.

Macdonald, P. M., E. Kutter and G. Mosig, 1984 Regulation of a bacteriophage T4 late gene, *soc*, which maps in an early region. Genetics **106**: 17–27.

Macdonald, P. M., and G. Mosig, 1984 Regulation of a new bacteriophage T4 gene, *69*, that spans an origin of DNA replication. EMBO J. **3**: 2863–2871.

Matic, I., M. Radman, F. Taddei, B. Picard, C. Doit et al., 1997 Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. Science **277**: 1833–1834.

Milkman, R., 1996 Recombinational exchange among clonal populations, pp. 2663–2684 in *Escherichia coli and Salmonella*, edited by F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. ASM Press, Washington, DC.

Miner, Z., and S. Hattman, 1988 Molecular cloning, sequencing, and mapping of the bacteriophage T2 *dam* gene. J. Bacteriol. **170**: 5177–5184.

Mosig, G., 1963 Genetic recombination in bacteriophage T4 during replication of DNA fragments. Cold Spring Harbor Symp. Quant. Biol. **28**: 35–42.

Mosig, G., 1968 A map of distances along the DNA molecule of phage T4. Genetics **59**: 137–151.

Mosig, G., 1987 The essential role of recombination in phage T4 growth. Annu. Rev. Genet. **21**: 347–371.

Mosig, G., 1994 Homologous recombination, pp. 54–82 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Mosig, G., and P. M. Macdonald, 1986 A new membrane-associated DNA replication protein, the gene *69* product of bacteriophage T4, shares a patch of homology with the *Escherichia coli* dnaA protein. J. Mol. Biol. **189:** 243–248.

Mosig, G., and D. H. Hall, 1994 Gene expression: a paradigm of integrated circuits, pp. 127–131 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Mosig, G., and N. Colowick, 1995 DNA replication of bacteriophage T4 *in vivo*, pp. 587–604 in *Methods in Enzymology*, Vol. 262, edited by J. Campbell. Academic Press, San Diego.

Mosig, G., R. Ehring, W. Schliewen and S. Bock, 1971 The patterns of recombination and segregation in terminal regions of T4 DNA molecules. Mol. Gen. Genet. **113:** 51–91.

Mosig, G., W. Berquist and S. Bock, 1977 Multiple interactions of a DNA-binding protein *in vivo*. III. Phage T4 gene-*32* mutations differentially affect insertion-type recombination and membrane properties. Genetics **86:** 5–23.

Mosig, G., A. Luder, A. Ernst and N. Canan, 1991 Bypass of a primase requirement for bacteriophage T4 DNA replication *in vivo* by a recombination enzyme, endonuclease VII. The New Biol. **3:** 1195–1205.

Nelson, K., F.-S. Wang, E. F. Boyd and R. S. Selander, 1997 Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphatase gene (*aceK*) and flanking regions in *Salmonella enterica* and *Escherichia coli*. Genetics **147:** 1509–1520.

Okker, R. J. H., 1981 Partial exclusion of bacteriophage T2 by bacteriophage T4: induction of early enzymes by excluded T2. J. Gen. Virol. **56:** 267–274.

Okker, R. J. H., E. Pees and V. Bom, 1981 Partial exclusion of bacteriophage T2 by bacteriophage T4: an exclusion resistant mutation in gene *56* of T2. J. Gen. Virol. **53:** 13–19.

Otto, M. R., 1997 Examination of the kinetic intermediates during bacteriophage T4 DNA polymerase proofreading by fluorescence spectroscopy. Ph. D. Dissertation, Vanderbilt University, Nashville TN.

Piechowski, M. M., and M. Susman, 1967 Acridine resistance in phage T4D. Genetics **56:** 133–148.

Pribnow, D., D. C. Sigurdson, L. Gold, B. S. Singer, C. Napoli *et al.*, 1981 *rII* cistrons of bacteriophage T4. DNA sequence around the intercistronic divide and positions of genetic landmarks. J. Mol. Biol. **149:** 337–376.

Radman, M., 1991 Avoidance of inter-repeat recombination by sequence divergence and a mechanism of neutral evolution. Biochimie **73:** 357–361.

Rayssiguier, C., C. Dohet and M. Radman, 1991 Interspecific recombination between *Escherichia coli* and *Salmonella typhimurium* occurs by the RecABCD pathway. Biochimie **73:** 371–374.

Reha-Krantz, L. J., 1994 Genetic dissection of T4 DNA polymerase structure-function relationship, pp. 307–312 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. ASM Press, Washington, DC.

Remaut, E., P. Stanssens and W. Fiers, 1981 Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. Gene **15:** 81–93.

Repoila, F., F. Tétart, J.-Y. Bouet and H. M. Krisch, 1994 Genomic polymorphism in the T-even bacteriophages. EMBO J. **13:** 4181–4192.

Ripley, L. S., 1990 Frameshift mutations—determinants of specificity. Annu. Rev. Genet. **24:** 189–213.

Ripley, L. S., J. S. Dubins, J. G. deBoer, D. M. DeMarini, A. M. Bogerd *et al.*, 1988 Hotspot sites for acridine-induced frameshift mutations in bacteriophage T4 correspond to sites of action of the T4 type II topoisomerase. J. Mol. Biol. **200:** 665–680.

Rosenberg, S. M., 1994 In pursuit of a molecular mechanism for adaptive mutation. Genome **37:** 893–899.

Rosenberg, S. M., R. S. Harris, S. Longerich and A. M. Galloway, 1995 Recombination-dependent mutation in non-dividing cells. Mutat. Res. **350:** 69–76.

Russell, R. L., and R. J. Huskey, 1974 Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism. Genetics **78:** 989–1014.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shcherbakov, V. P., and L. A. Plugina, 1991 Marker-dependent recombination in T4 bacteriophage. III. Structural prerequisites for marker discrimination. Genetics **131:** 673–685.

Shcherbakov, V. P., L. A. Plugina and E. A. Kudryashova, 1995 Marker-dependent recombination in T4 bacteriophage. IV. Recombinational effects of antimutator T4 DNA polymerase. Genetics **140:** 13–25.

Shub, D. A., H. Goodrich-Blair and S. R. Eddy, 1994 Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. TIBS **19:** 402–404.

Singer, B. S., L. Gold, P. Gauss and D. H. Doherty, 1982 Determination of the amount of homology required for recombination in bacteriophage T4. Cell **31:** 25–33.

Singer, B. W., 1988 On the role of homologous sequences in chromosome rearrangements. Genes Dev. **2:** 1800–1811.

Stitt, B., and D. Hinton, 1994 Regulation of middle-mode transcription, pp. 142–160 in *Molecular Biology of Bacteriophage T4*, edited by J. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson and E. S. Miller. AMS Press, Washington, DC.

Streisinger, G., 1966 Terminal redundancy, or all's well that ends well, pp. 335–340 in *Phage and the Origins of Molecular Biology*, edited by J. Cairns, G. S. Stent and J. D. Watson. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Streisinger, G., and J. Weigle, 1956 Properties of bacteriophages T2 and T4 with unusual inheritance. Proc. Natl. Acad. Sci. USA **42:** 504–510.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita *et al.*, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

Sugawara, N., F. Pâques, M. Colaiácovo, and J. Haber, 1997 Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand-break-induced recombination. Proc. Natl. Acad. Sci. USA **94:** 9214–9219.

Tétart, F., C. Monod and H. M. Krisch, 1996 Bacteriophage T4 host range is expanded by duplications of a small domain of the tail fiber adhesin. J. Mol. Biol. **258:** 726–731.

Torkelson, J., R. S. Harris, M.-J. Lombardo, J. Nagendran, C. Thulin, and S. M. Rosenberg, 1997 Genome-wide hypermutation in a subpopulation of stationary-phase cells underlies recombination-dependent adaptive mutation. EMBO J. **16:** 3303–3311.

Tyndall, C., H. Lehnherr, U. Sandmeier, E. Kulik and T. A. Bickle, 1997 The type IC *hsd* loci of the enterobacteria are flanked by DNA with high homology to the phage P1 genome: implications for the evolution and spread of DNA restriction systems. Molecular Microbiology **23:** 729–736.

Vulić, M., F. Dionisio, F. Taddei and M. Radman, 1997 Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA **94:** 9763–9767.

Waldor, M. K., and J. J. Mekalanos, 1996 Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272:** 1910–1914.

Waldor, M. K., E. J. Rubin, G. D. N. Pearson, H. Kimsey and J. J. Mekalanos, 1997 Regulation, replication, and integration functions of the *Vibrio cholerae* CTXϕ are encoded by region RS2. Molecular Microbiology **24:** 917–926.

Wang, F.J., and L. S. Ripley, 1998 The spectrum of acridine resistant mutants of bacteriophage T4 reveals cryptic effects of the *tsL141* DNA polymerase allele on spontaneous mutagenesis. Genetics (in press).

Womack, F. C., 1963 An analysis of single-burst progeny of bacteria singly infected with a bacteriophage heterozygote. Virology **21:** 232–241.

Wu, D. G., C. H. Wu and L. W. Black, 1991 Reiterated gene amplifications at specific short homology sequences in phage T4 produce *hp17* mutants. J. Mol. Biol. **218:** 705–721.

Yee, J. K., and R. C. Marsh, 1981 Alignment of a restriction map with the genetic map of bacteriophage T4. J. Virol. **38:** 115–124.

Zahrt, T. C., and S. Maloy, 1997 Barriers to recombination between closely related bacteria: MutS and RecBCD inhibit recombination between *Salmonella typhimurium* and *Salmonella typhi*. Proc. Natl. Acad. Sci. USA **94:** 9786–9791.