The effect of the length of direct repeats and the presence of palindromes on deletion between directly repeated DNA sequences in bacteriophage T7

James C. Pierce⁺, Daochun Kong and Warren Masker^{*}

Department of Biochemistry and Molecular Biology, Temple University School of Medicine, Philadelphia, PA 19140, USA

Received April 26, 1991; Revised and Accepted June 19, 1991

ABSTRACT

The frequency of genetic deletion between directly repeated DNA sequences in bacteriophage T7 was measured as a function of the length of the direct repeat. The non-essential ligase gene (gene 1.3) of bacteriophage T7 was interrupted with pieces of synthetic DNA bracketed by direct repeats of various lengths. Deletion of these 76 bp long inserts was too low to be measured when the direct repeats were less than 6 bp long. However, the frequency of deletion of inserts with longer direct repeats increased exponentially as the length of the repeats increased from 8 to 20 bp. When inverted repeats (palindromes) were designed in the midst of the insert there was essentially no increase in deletion frequency between 10 bp direct repeats. But, the same palindromic sequences increased the deletion frequency between 5 bp direct repeats by at least two orders of magnitude. Thus, in this system homology at the endpoints is a more important determinant of deletion frequency than is the presence of palindromes between the direct repeats.

INTRODUCTION

Direct repeats of DNA sequence play a major role in determining the probability of deletion events (1–7). Although it is not clear why short regions of homology are important to deletion, several models have been proposed to explain rearrangements of this type (8–13). One possibility is that recombinogenic exchange could take place between the direct repeats (3–6, 14–19). Or, slippage at the replication fork could realign a newly synthesized copy of one direct repeat with the downstream direct repeat (1, 20–21). The amount of homology between the direct repeats is likely to be a significant factor in determining deletion frequency. With this in mind, we have systematically investigated both how variation in the length of direct repeats and the presence of

palindromic sequences affected the likelihood of deletion at a specific site on the bacteriophage T7 genome.

Bacteriophage T7 offers numerous advantages as a system with which to measure genetic deletions. Foremost among these is the simplicity of the replication apparatus, which appears to consist of the phage gene 4 product (a helicase-primase), T7 DNA polymerase (associated with host thioredoxin), and T7 single strand binding protein (22,23). If deletion is closely coupled with the replication process the simplicity and relatively thorough characterization of the T7 DNA replication enzyme system may aid understanding deletion mechanisms. Also, the availability of good *in vitro* DNA replication, recombination, and packaging systems (24,25) add to the potential of the T7 system as a means for studying deletion.

To quantitatively measure deletion frequency at a specific site we interrupted a non essential T7 gene with a piece of synthetic DNA that was bracketed by direct repeats (20,26). Deletion between the direct repeats reveals itself by restoration of function to the gene that formerly harbored the insert. We used this system to monitor deletion frequency as a function of the length of the direct repeat in the range between 5 and 20 base pairs and found that the likelihood of deletion is exponentially dependent upon the extent of homology.

In other systems it is well established that palindromic sequences which could lead to cruciform formation increase deletion frequency (2,3,5,7,8,12,15, 21,27). In T7 it was more difficult to predict the effect of palindromes since in T7's linear chromosome the presence of palindromic sequences does not necessarily imply that cruciforms will form. We constructed inserts that contained palindromic sequences and had either 5 or 10 bp direct repeats at the ends of the insert. Secondary structure resulting from the palindromes could act as a retardant to replication fork progression. Or, T7 endonuclease I could make double strand breaks at the cruciform's base (28,29). As shown below, the presence of palindromic sequences caused a major increase in the frequency of deletion between 5 bp direct repeats, but had essentially no effect on deletion between 10 bp direct repeats.

* To whom correspondence should be addressed

⁺ Present address: The E.I. DuPont Merck Pharmaceutical Co., Experimental Station, Wilmington, DE 19880-0328, USA

MATERIALS AND METHODS

Bacteria, Phage, and Growth Conditions

A unique XhoI restriction site was created in wild type bacteriophage T7 (from Dr. F.W. Studier) by using oligonucleotide mutagenesis at position 6663 in the ligase gene (1.3). The base alteration did not change the wild type amino acid sequence (30). This phage mutant was designated 'T7X'. Wild type *Escherichia coli* K-12 strain W3110 and the ligase deficient strain N2668 (*ligts7*) were used in this study. The temperature sensitive mutation in the bacterial ligase gene renders strain N2668 incapable of supporting the productive infection of T7 phage with deficiencies in gene 1.3 which encodes the phage ligase (31,32). Wild type T7 grows normally on a *ligts7* host. Bacteria were routinely grown on L broth or T broth media (33).

Construction of T7 Containing Inserts

The method for introduction of synthetic DNA sequences into gene 1.3 have been described (26). Synthetic oligonucleotides with complementary sequences were annealed to one another to generate a duplex DNA molecule with a unique BamHI restriction site and XhoI sticky ends at each terminus. DNA from T7X was digested with XhoI and ligated to the synthetic insert. After in vitro packaging ligase deficient phage were identified by their inability to grow on a *ligts7* host. In all cases the phage identity was confirmed by DNA sequencing.

Restriction Analysis, Cloning, and DNA Sequencing

Digestion with restriction enzymes (Bethesda Research Laboratories or New England Biolabs) was performed as per supplier's instructions. A 0.7 kb AluI fragment from positions 6191 and 6889 of T7 was cloned into the HincII site of M13-mp19 (34,35). DNA sequencing of single-stranded M13 templates was done using the dideoxy chain termination method (36).

Assays for Direct Repeat Mediated Deletions in T7

The assay for measurement of deletion frequency (26) is based on infection of bacterial cultures with one (or at most two) T7 phage. This essentially eliminates the possibility that any phage that has already undergone a deletion event is introduced into the cultures. Moreover, the number of phage recovered after lysis of the small cultures is equal to the total number of replications of the phage genome. T7 phage that contain an insert are diluted so that approximately 1.5 viable phage are added to a 1 ml culture of wild type *E. coli* growing exponentially at 37°C. After lysis, the phage titer is determined on lawns of wild type *E. coli* and on lawns of the *ligts7* host. The relative deletion frequency is calculated from the ratio of $\text{lig}^-/\text{lig}^+$ titers. This assay measures the fraction of phage that have either undergone a deletion event or have descended from phage that have undergone deletion. Deletions that occur early in the phage infection will influence this measurement of 'deletion frequency' disproportionately. Measurements of accumulation of ligase proficient phage derived from a single parent are typically higher than the true deletion frequency as measured by the far more laborious Luria-Delbruck fluctuations test (37). However, it has been found that as long as cultures of wild type strain (W3110) are used for replication of the single phage, the median value of a large number of measurements of deletion frequency made in this way is a reproducible and reliable way for comparing deletion frequencies involving different inserts (26). When this assay was used to measure how the distance between direct repeats affects the

deletion frequency the data obtained showed essentially the same pattern found (20) when fluctuations tests (37) were used instead (data not shown). To examine the very low deletion frequency of inserts with 5 and 6 bp direct repeats phage were added at an moi of $<10^{-3}$ to a 30 ml exponentially growing culture of wild type *E. coli*. After complete lysis of the culture, the phage were concentrated and their titer measured on wild type and *ligts7 E. coli*.

RESULTS

The Effect of the Length of Direct Repeats on the Frequency of Deletion

Synthetic 76 bp DNAs were designed to interrupt the T7 ligase gene and produce direct repeats at the ends of the insert (Figure 1). Most of the inserts will generate 5 bp direct repeats if they enter the T7 genome in one orientation; the other orientation will yield a longer (10 to 20 bp) repeat. One insert produces a 6 bp direct repeat in one orientation and a 8 base

T7X76⁺
 5'...ccggcacTCGATGGATCCGATACAACCTTAGTACGTTGCGATAGCCCTACACGTCAGTAGT
 GCAGTCACTAGTTAGCAAGTGCtcgagcacttaaacgggtttgat...3'

T7X76/6
 5'...ccggcacTCGAGGGATCCGATACAACCTTAGTACGTTGCGATAGCCCTACACGTCAGTAGT
 GCAGTCACTAGTTAGCGCATGctcgagcacttaaacgggtttgat...3'

T7X76/8
 5'...ccggcacTCGAGCATGCGCTAACTAGGTGACTGCACTACTGACGTGTAGGGCTATCGCAA
 CGTACTAAGTTGTATCGGATCCctcgagcacttaaacgggtttgat...3'

T7X76⁺
 5'...ccggcacTCGAGCACTTGCTAACTAGGTGACTGCACTACTGACGTGTAGGGCTATCGCAA
 CGTACTAAGTTGTATCGGATCCctcgagcacttaaacgggtttgat...3'

T7X76/13
 5'...ccggcacTCGAGCACTTAAAGTCTAGGTGACTGCACTACTGACGTGTAGGGCTATCGCAA
 CGTACTAAGTTGTATCGGATCCctcgagcacttaaacgggtttgat...3'

T7X76/16
 5'...ccggcacTCGAGCACTTAAACGGTTTAGTACTGCACTACTGACGTGTAGGGCTATCGCAA
 CGTACTAAGTTGTATCGGATCCctcgagcacttaaacgggtttgat...3'

T7X76/20
 5'...ccggcacTCGAGCACTTAAACGGGTTTAGTACTGCACTACTGACGTGTAGGGCTATCGCAA
 CGTACTAAGTTGTATCGGATCCGctcgagcacttaaacgggtttgat...3'

Figure 1. Inserts used to Disrupt the T7 Ligase Gene. The figure shows one the sequence of the non-template strand (i.e. the same sequence as the transcribed mRNA) of the pair of complementary oligonucleotides that were annealed together and then inserted into the XhoI restriction site in gene 1.3. Upper case letters indicate the sequence of the insert; lower case letters indicate the sequence in T7 wild type before insertion of the synthetic DNA. The lines above the sequence mark the location of direct repeats.

pair direct repeat in the other orientation. The sequences of the right ends of the inserts with 8–20 bp direct repeats are essentially the same (since the repeat itself is part of the original T7 genome). The DNA sequences of the inserts with 5 and 6 bp direct repeats is different due to the method by which they were placed in the T7 genome. All inserts introduce several stop codons in the normal reading frame and at least one stop codon in the other two, and all introduce a unique BamHI site in the T7 genome.

Table 1 shows that we found no example of wild type phage arising from excision of the insert with 5 bp direct repeats. Deletion of the X76/6 insert was barely detectable. In the range between 8 and 20 bp the deletion frequency increased exponentially with the length of the direct repeat (Figure 2).

The Effect of Potential Secondary Structure on Deletion Frequency

Figure 3 illustrates inserts with internal palindromic sequences which could form DNA secondary structure. The insert termed 'X76' was used as a control since it should not form hairpins. The X76EP (Extended Palindrome) insert can form a 25 bp stem and a 16 bp loop that would bring the 10 bp inserts into juxtaposition. If DNA replication proceeded across this junction

it would create a 1.3 gene with an out of frame transcript. But, a deletion event between the direct repeats could generate a functional ligase gene. When placed in the genome in one orientation the X76EP insert was bracketed by 10 bp direct repeats, but in the opposite orientation it was flanked by 5 bp direct repeats. Another insert, 'X76I', could form a perfect 14 bp palindrome internal to the 10 bp direct repeats.

Table 2 shows that the median deletion frequency for the X76⁺ and X76EP⁺ differed by no more than a factor of two. But, the palindromes caused a striking effect in deletion frequency when bracketed by 5 bp direct repeats. The excision of the X76EP[−] insert was at least two orders of magnitude higher than deletion of X76[−]. DNA from ligase positive revertants of phage that had carried insert X76EP[−] was examined for the presence of the unique BamHI restriction site present on the insert. A revertant from each of ten separate lysates was tested and in every case found to lack this restriction site, thus indicating that the insert had in fact been deleted from the genome. DNA from two revertants was sequenced, and in both cases the restoration of the ligase gene resulted from deletion between the direct repeats.

Pseudorevertants that Do Not Excise the Entire Insert

Phage that had contained the X76I insert show marked heterogeneity in plaque size when grown on the ligase deficient

Table 1. Frequency of Deletion of a 76 Base Pair Insert as a Function of the Length of the Direct Repeats at its Ends.

Phage	Length of direct repeat	Number of lysates tested	Median number of deletions per total phage population
X76 [−]	5	10	$< 10^{-10}$
X76/6	6	10	7.1×10^{-10}
X76/8	8	24	2.5×10^{-7}
X76 ⁺	10	55	1.3×10^{-6}
X76/13	13	16	4.5×10^{-6}
X76/16	16	21	4.8×10^{-5}
X76/20	20	23	7.1×10^{-4}

Figure 2. The Effect of Increased Homology on Deletion between Direct Repeats. The logarithm (base 10) of deletion frequency, measured as the accumulation of wild type phage in a bacterial culture infected with a single T7 phage particle that carried an insert in the ligase gene, is shown as a function of the length of direct repeat for 76 bp inserts with very similar DNA sequences.

Figure 3. Potential for Formation of Secondary Structure within the Inserts that Interrupt Gene 1.3. The figure shows the positions of 25 base pair inverted repeats near the ends of Insert X76EP and of 14 base pair inverted repeats near the center of Insert X76I. Also shown are single strands of cruciform structures that could be formed as a result of the palindromic sequences within the inserts. Asterisks mark the positions of stop codons. The dark solid areas mark the positions of direct repeats, which, as described in the text, can be either five or ten base pairs long.

Table 2. Frequency of Deletion of Inserts with and without the Potential for DNA Secondary Structure.^{a)}

Phage	Length of direct repeat (bp)	Length of potential palindrome (bp)	Number of lysates tested	Median number deletions per total phage population
X76 ⁺	10	none	55	1.3×10^{-6}
X76EP ⁺	10	25	23	3.0×10^{-6}
	10	14	20	1.5×10^{-6}
X76 [−]	5	none	5	$< 10^{-10}$
X76EP [−]	5	25	4	1.8×10^{-8}

^{a)} The relative number of phage that had undergone deletion of the indicated insert was measured in a series of small cultures each infected at very low multiplicity. Because of the low deletion frequency the phage with five base pair direct repeats at the ends of the inserts were measured using 30 ml cultures as described in Materials and Methods. The Table reports median values from several measurements.

Figure 4. Sequence of X76I Pseudorevertants. The figure shows that part of the sequence of one strand of Insert X76I that has potential for cruciform formation. The six base pair direct repeat that was fortuitously present within the insert is shown surrounded within a box. The sequence of two T7 'revertants', X76I-RS-1 and X76I-RS-2, that were able to grow on a ligase deficient host is also shown. The arrow shown under the sequence for X76I-RS-2 marks the position of a single base point mutation that accompanied the deletion event.

host. The larger size plaques typically outnumbered the smaller ones by a factor of 5 to 10. Several plaques of each type were picked, and their DNA was digested with BamHI. All of the DNA derived from the large plaques had lost the BamHI restriction site, whereas the DNA from small sized plaques was still sensitive to BamHI. The sequence of the DNA from the large plaque showed the expected deletion event between the direct repeats (data not shown). Part of the DNA sequence of samples from both small plaques is shown in Figure 4. In both cases the revertants arose when a 13 bp deletion occurred between two 6 bp direct repeats within the X76I insert. The deletion, which occurs at one end of the 14 bp palindrome in the X76I insert, removed two stop codons in the insert and reestablished the correct reading frame. Evidently the T7 ligase protein can tolerate at least 21 additional amino acids in this portion of the protein. A transition mutation (T to C) 5 bp past the deletion endpoint was noticed in one of two revertants (X76I-RS2) that was sequenced. The sequence change in this revertant was confirmed using a different M13 clone.

DISCUSSION

The most direct interpretation of our data suggests a marked increase in deletion frequency as the length of the direct repeat increases from 6 to 8 bp. Because inserts with 6 and 8 bp repeats were selected by choosing phage in which a single piece of synthetic DNA had been inserted in each of two possible orientations, the sequence of the X76/6 and X76/8 inserts is not the same. If partial matches of homology contribute to excision of the insert (38) the X76/8 insert could be affected more than the X76/6 insert. Also, other authors have reported major differences in deletion frequency due to small sequence changes outside of the direct repeats (21). It must be kept in mind that our study focuses on one site in the T7 genome and that proteins involved in deletion may react to subtle features of DNA sequence. Nonetheless, the difference in deletion frequency between 6 and 8 bp repeats could reflect fundamentally different mechanisms by which the direct repeats participate in deletion

events. The inserts with 8 to 20 bp direct repeats have sequences that are basically the same except for the repeats themselves. In this range the deletion frequency is an exponential function of the length of direct repeat (Fig 2). Thus, the degree of homology between the endpoints of a potential deletion event is a major factor in determining deletion frequency.

Cruciform structures located between directly repeated sequences are likely to produce three effects that might alter the frequency of deletion between the direct repeats. First, the direct repeats would be brought physically closer together by cruciforms. But, our earlier work (26) showed that reduction of the distance between direct repeats had only a modest effect on deletion frequency when the insert was 76 bp or less. Secondly, since T7 endonuclease I attacks cruciform structures at their bases, double strand breaks might accumulate at palindromic sequences and these could promote recombinational exchanges nearby. Intermolecular recombination does not appear to be a major factor in deletion between 10 bp direct repeats in T7 (20). But, that work was done in the absence of double strand breaks and recombination increases when breaks are introduced (W.Masker, to be published). A third effect of cruciforms is to slow replication fork progression. This might force the replication enzymes to remain longer in the vicinity of two nearby homologous sequences and thereby increase the probability of slipped mispairing. By default, we consider this the most likely reason that palindromes increase deletion between the short (5 bp) repeats in our system.

The most direct interpretation of our data is that homology is a much more important determinant of deletion frequency than is the presence of palindromes internal to the direct repeats. While palindromes might contribute a little to the probability of deletion the enhancement of deletion frequency afforded by the presence of palindromes was undetectable unless the homology at the endpoints was so small as to make the background of deletion in the absence of palindromes below 10^{-9} . In fact, the substantial *proportional* enhancement of deletion frequency seen when the inserts were bracketed by 5 bp direct repeats provides the best evidence that cruciform structures do indeed form in the linear duplex T7 chromosome.

Other workers have reported (21,27) limits on the effects of palindromy on deletion frequency in *E. coli* plasmids that had inserts with endpoints where 13 out of 14 bases were direct repeats and 4 bp of the direct repeat was also part of the palindrome (21). Deletion of the inserts was enhanced by palindromes greater than 26 bp in length (21,27). Our finding that a 25 bp palindrome affects deletion between direct repeats only when the direct repeats are short and the deletion frequency is therefore low is also in accord with an inverse relationship between the length of the direct repeat and the length of the palindrome. Thus, there may be competition between the stability of complementary base pairing within the direct repeat and the stability of putative cruciform structure.

ACKNOWLEDGEMENTS

We thank Ms. Kathleen Howley and Mary Ann Crissey for technical assistance and Marie Searce for helpful discussions. Nicholas Tiariis was responsible for cloning the insert with 13 bp direct repeats into T7. This work was supported by Public Health Service research grant GM-34614 from the National Institutes of Health.

REFERENCES

1. Albertini, A. M., Hofer, M., Calos, M. P. and Miller, J.H. (1982) *Cell*, **29**, 319–328.
2. DasGupta, U., Weston-Hafer, K. and Berg, D.E. (1987) *Genetics*, **115**, 41–49.
3. Ehrlich, S. D. (1989) In Berg D.E. and Howe M.M. (eds.) *Mobile DNA*. American Society for Microbiology, Washington, D. C., pp. 799–832.
4. Farabaugh, P. J., Schmeissner, U., Hofer, M. and Miller, J.H. (1978) *J. Mol. Biol.*, **126**, 847–863.
5. Singer, B. S. and Westlye, J. (1988) *J. Mol. Biol.*, **202**, 233–243.
6. Studier, F. W., Rosenberg, A. H., Simon, M. N. and Dunn, J.J. (1979) *J. Mol. Biol.*, **135**, 917–935.
7. Weston-Hafer, K. and Berg, D.E. (1989) *Plasmid*, **21**, 251–253.
8. Glickman, B.W. and Ripley, L.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 512–516.
9. Ikeda, H. Kazuko, A., and Natito, A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 3724–3728.
10. Ikeda, H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 922–926.
11. Marvo, S. L., King, S. R. and Jaskunas, S.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2452–2456.
12. Ripley, L.S. and Glickman, B.W. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 851–861.
13. Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terghi, and Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 77–84.
14. Anderson, P. (1987) *Genetics*, **15**, 581–584.
15. Balbinder, E., Mac Vean, C. and Williams, R.E. (1989) *Mut. Res.*, **214**, 233–252.
16. Bedinger, P., M. Munn, and Alberts, B.A. (1989) *J. Biol. Chem.*, **264**, 16880–16886.
17. Pribnow, D., Sigurdson, D., Gold, L., Singer, B. S., Napoli, C., Brosius, J., Dull, T. J. and Noller, H.F. (1981) *J. Mol. Biol.*, **149**, 337–376.
18. Primrose, S. B. and Ehrlich, S.D. (1981) *Plasmid*, **6**, 193–201.
19. Schaaper, R. M., Danforth, B. N., and Glickman, B.W. (1986) *J. Mol. Biol.*, **189**, 273–284.
20. Searce, L.M., Pierce, J.C., McInroy, B. and Masker, W. (1991) *J. Bacteriol.*, **173**, 869–878.
21. Weston-Hafer, K. and Berg, D.E. (1991) *J. Bacteriol.*, **173**, 315–318.
22. Richardson, C. C. (1983) *Cell*, **33**, 315–317.
23. Nakai, H., Beauchamp, B. B., Bernstein, J., Huber, H. E., Tabor, S. and Richardson, C.C. (1988) In Moses, R.E. and Summers W.C. (eds.) *DNA Replication and Mutagenesis*, American Society for Microbiology, Washington, D. C. pp. 85–87.
24. Masker, W.E., Kuemmerle, N.B., and Allison, D.P. (1978) *J. Virol.*, **27**, 149–163.
25. Masker, W.E. and Kuemmerle, N.B. (1980) *J. Virol.*, **33**, 330–339.
26. Pierce, J. C. and Masker, W. (1989) *Mol. Gen. Genet.*, **217**, 215–222.
27. Weston-Hafer, K. and Berg, D.E. (1989) *Genetics*, **121**, 651–658.
28. deMassy, B., Weisberg, R.A., and Studier, F.W. (1987) *J. Mol. Biol.*, **193**, 359–376.
29. Muller, B., James, C., and West, S. (1990) *Nuc. Acids Res.*, **18**, 5633–5636.
30. Dunn, J. J. and Studier, F.W. (1983) *J. Mol. Biol.*, **166**, 477–535.
31. Masamune, Y., Frenkel, G. D. and Richardson, C.C. (1971) *J. Biol. Chem.*, **246**, 6874–6879.
32. Studier, F. W. (1973) *J. Mol. Biol.*, **79**, 227–236.
33. Miller, J. H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
34. Messing, J. (1983) *Methods Enzymol.*, **101**, 20–78.
35. Yanisch-Perron, C. J. Vieira, and Messing, J. (1985) *Gene*, **33**, 103–119.
36. Sanger, F., S. Nicklen, and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
37. Luria, S. E. and Delbruck, M. (1943) *Genetics*, **28**, 491–51.
38. Ripley, L.S. (1990) *Annu. Rev. Genet.*, **24**, 189–213.