CHAPTER 8

SINGLE-STRANDED GAPS AS LOCALIZED TARGETS FOR IN VITRO MUTAGENESIS

David Shortle and David Botstein

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

SUMMARY

Short single-stranded gaps in circular DNA molecules can be generated enzymatically, often at predetermined sites. These can serve as targets for in vitro mutagenesis procedures that result in alterations in nucleotide sequence within or very near the gap. Deamination of unpaired cytosine residues with sodium bisulfite has been used to induce mutations in the *BglI* restriction site of SV40 DNA and within defined regions of the *β-lactamase* gene on pBR322. A new method of induction of mutations at gaps, called "gap misrepair," has been developed; it was used to cause changes at the *HindIII* and *ClaI* restriction sites on pBR322 DNA. Gap misrepair reactions using DNA polymerase I of *Micrococcus luteus* in the presence of T4 DNA ligase and three of the four deoxynucleoside triphosphates yielded all three possible substitutions for adenine and cytosine residues in the DNA.

INTRODUCTION

A stretch of single-stranded DNA (a gap) in an otherwise duplex molecule has a number of structural features that render the nucleotide sequence in the gap specifically susceptible to mutagenesis in vitro by two methods. First, cytosine residues in the single-stranded segment become accessible to deamination by sodium bisulfite, a reaction resulting in C•G → T•A transition mutations (Shortle and Nathans, 1978). Second, a gap can serve as a primer-template for mutagenic "misrepair" reactions with DNA polymerase, either by incorporation of nucleotide analogues (Muller et al., 1978) or, as shown below, by misincorporation of the standard four

nucleotides. These two reactions can be used for site-specific mutagenesis by enzymatically generating a gap, the target of mutagenesis, at a predetermined site on a circular DNA molecule. In the following discussion, the enzymatic reactions used to generate gaps at specific sites will be briefly reviewed, and then some of the mutants that have been recovered specifically at single-stranded gaps will be described.

GENERATION OF SINGLE-STRANDED GAPS

The basic route by which a single-stranded gap is generated enzymatically in a duplex DNA molecule is outlined in Fig. 1. First, an endonuclease is used to introduce a break or nick in one of the two DNA strands; then, an exonuclease that can initiate hydrolysis at a nick is allowed to remove a few nucleotides, proceeding either in the 5' → 3' or 3' → 5' direction away from the nick. Once the site of the nick is specified in the first reaction, the direction and extent of hydrolysis by the exonuclease determine the final structure of the gap. To generate single-stranded gaps at uniquely defined positions, therefore, requires methods for nicking DNA at specific sites plus exonucleolytic reactions that can be controlled.

Nicking at Restriction Sites

When incubated with a negatively supercoiled circular DNA containing one or more restriction sites, type II restriction endonucleases cleave at (or near) their recognition sequences by generating two nicks, one in each strand. With some type II enzymes (e.g., BglI, HindIII, and ClaI), this cleavage reaction can be inhibited after the first nick has been induced by including in the reaction mixture the intercalating compound ethidium bromide (Parker et al., 1977). Using purified supercoiled DNA and ethidium bromide at a

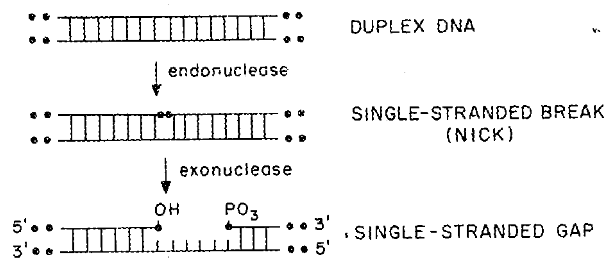

Fig. 1. The two enzyme-catalyzed reactions used in the generation of single-stranded gaps.

concentration determined to be optimal by titration, it is often possible to nick between 50 and 90% of the input DNA, yielding open circular molecules with a single nick located at the restriction enzyme's normal cleavage site (Shortle and Nathans, 1978). After conversion of the nick to a gap that exposes the restriction recognition sequence, such DNA molecules provide convenient substrates for assaying mutagenesis by reactions specific for single-stranded gaps, since mutations affecting the recognition sequence confer the easily scorable phenotype of loss of a particular restriction site.

Nicking Within Specific DNA Segments

The number of sites at which single-stranded gaps can be generated via restriction enzyme-catalyzed nicking is, of course, limited by the availability of usable restriction sites. When the objective is to construct mutations within a defined nucleotide sequence interval carried on a circular DNA, a two-step, segment-specific nicking procedure can be used (Shortle et al., 1980). In principle, with this procedure a nick can be introduced into any segment of a circular DNA, provided that a single-stranded fragment corresponding to that segment can be isolated. With ATP serving as cofactor, the recA protein is used to catalyze the annealing of a unique, single-stranded fragment to the complementary sequence on a covalently closed circular DNA (McEntee et al., 1979; Shibata et al., 1979). The annealed fragment displaces one strand of the circular DNA, creating a single-stranded D-loop. In a second reaction, this D-loop structure becomes a substrate for the single-strand specific endonuclease S1. Once the first nick is induced by S1 nuclease attack on the displaced strand, the D-loop rapidly breaks down by spontaneous displacement of the fragment, yielding an open circular DNA molecule with a nick located within the DNA segment corresponding to the single-stranded fragment.

When negatively supercoiled plasmid pBR322 DNA (4.3 kilobases in length) was subjected to nicking with this procedure, nicks were efficiently induced in defined segments as short as 130 base pairs in length (Shortle et al., 1980). Before use of this nicked DNA as a substrate for gapping and mutagenesis, the specificity of the nicking reaction was assayed by incorporation of α-^{32}P-labeled nucleotides at the site of the nick in a limited nick translation, followed by determination of the site of the incorporated label by restriction analysis and gel electrophoresis. From this biochemical analysis and from fine-structure genetic mapping of the mutants subsequently constructed by using this DNA as substrate for mutagenesis, it was concluded that greater than 90% of nicks were localized to the 130 base-pair segment specified by the single-stranded fragment.

Exonuclease Reactions

DNA polymerase I from *Micrococcus luteus* has both the 3' → 5' and 5' → 3' exonuclease activities typical of this class of bacterial polymerase (Miller and Wells, 1972). The intrinsic activity of both exonuclease functions is relatively low in comparison to *Escherichia coli* DNA polymerase I. Consequently, this enzyme (available free of endonuclease activity from commercial sources) can be used in a time-controlled reaction to remove an average of approximately five to six nucleotides predominantly in the 5' → 3' direction (Shortle and Nathans, 1978), although gaps as long as 20 nucleotides may be generated. Alternatively, the exonuclease activity of T4 DNA polymerase can be used for hydrolysis in the 3' → 5' direction (Rawlins and Muzyczka, 1980). This enzyme has the advantage that the extent of hydrolysis is controllable by the addition of one or more deoxyribonucleotide triphosphates in the reaction mixture (Englund et al., 1974).

BISULFITE MUTAGENESIS

Sodium bisulfite catalyzes the deamination of cytosine under mild conditions of temperature and pH to form uracil. Although cytosine residues in single-stranded polynucleotides react at nearly the same rate as the free mononucleotide, cytosine residues within the duplex structure of double-stranded DNA are essentially inert to bisulfite deamination (for discussion, see Hayatsu, 1976). Therefore, sodium bisulfite is in effect a single-strand specific mutagen and can be used to specifically deaminate cytosine residues exposed in a gap. On filling in with DNA polymerase, bisulfite-treated gapped DNA molecules which have suffered C•G → U•A changes would be expected to result in C•G → T•A transition mutations on replication and/or repair.

The *BglI* Restriction Site of SV40

That sodium bisulfite can be used for site-specific mutagenesis of single-stranded gaps was demonstrated at the single *BglI* restriction site on SV40 DNA (Shortle and Nathans, 1978). As shown in Fig. 2, *BglI*-induced nicks can occur on either strand, and thus gapping in the 5' → 3' direction with the *M. luteus* DNA polymerase generates two types of gapped molecules. After incubation of *BglI* site-gapped DNA with sodium bisulfite under conditions that yield approximately 30% deamination of accessible cytosine residues, the gap was filled in with DNA polymerase in vitro. Molecules that lost the *BglI* site were recovered by digesting the modified DNA with *BglI* and isolating the circular DNA resistant to cleavage. This *BglI*-resistant DNA was transfected onto permissive tissue culture cells to obtain individual SV40 plaques. Analysis of 23 independent SV40

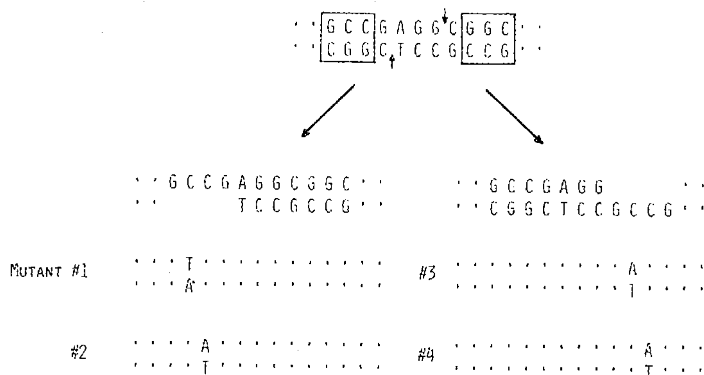

Fig. 2. The nucleotide sequences of the BglI restriction site (the recognition sequence is enclosed in boxes) in SV40 DNA and four viral mutants induced with bisulfite. The probable structures of the two types of single-stranded gaps generated by nicking with BglI plus ethidium bromide and gapping with the exonuclease activity of *M. luteus* DNA polymerase I are shown in the second line.

isolates revealed that 19 contained viral DNA that had lost the BglI site. These BglI-resistant SV40 mutants could be grouped into four phenotypic classes on the basis of plaque size at three different temperatures, and the nucleotide sequence surrounding the missing BglI site determined for one member of each class is shown in Fig. 2 (Shortle and Mathans, 1979). Three mutants had the base substitution expected of bisulfite-induced deamination events - a C·G → T·A transition, and a fourth SV40 mutant had a C·G → A·T transversion mutation. A second C·G → A·T transversion appeared in one or two other SV40 mutants induced with bisulfite at a short gap generated by very limited exonuclease III hydrolysis from the BglI nick (Shortle and Nathans, 1979). The molecular mechanism responsible for these "nonstandard" bisulfite-induced base substitutions is not known.

The β-lactamase Gene: Codons 1 to 46

To obtain point mutants in the "signal sequence" of the β-lactamase (*bla*) gene, which could be used for biochemical analysis of transport of this enzyme into the periplasmic space, specific nicks were induced in a segment of the gene (carried on the plasmid pBR322) that spans codon 1 to codon 46 (Shortle et al., 1980). After converting the nick into a gap with the *M. luteus* polymerase, reacting with sodium bisulfite, and filling in the gap in vitro, the modified plasmid DNA was used to transform an *E. coli* strain by selection for

the tetracycline resistance marker. Of the 22 bla$^-$ mutants recovered from 800 transformants, 14 mapped genetically within the signal sequence (codons 1 to 23) and the remaining eight mutants mapped within the amino terminal end of the mature protein (codons 24 to 129). Nucleotide sequence analysis of some of these bla$^-$ mutants and others induced in this same DNA segment in subsequent experiments has identified the four single C•G → T•A transition mutations shown in Fig. 3; the mutants in codon 4 and in codon 20 have been independently isolated three times each (Shortle, Grisafi, Koshland, and Botstein, unpublished data). A majority of the remaining mutants have, in addition to one or a few tightly clustered C•G → T•A transition mutations, frameshift mutations consisting of either a single T residue added to a run of four or more T residues (+1 frameshift) or a deletion (-1 frameshift) of a single C residue within a run of three C residues.

Experiments to determine the mechanism responsible for these frameshift mutations are in progress; at this point one can only speculate. One reasonable hypothesis on the origin of the +1 frameshifts is that the extra nucleotide was inserted in vitro during the gapping reaction. In this reaction, dATP was included at high concentrations to stimulate the 5' → 3' exonuclease activity of the *M. luteus* DNA polymerase (Miller and Wells, 1972). If the initial

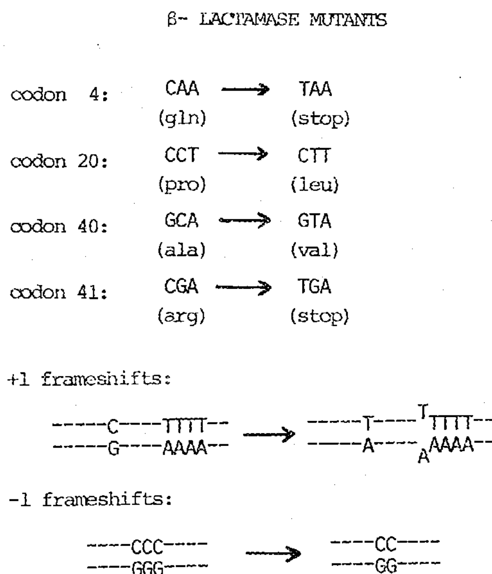

Fig. 3. The nucleotide sequences of β-lactamase mutants recovered after segment-directed mutagenesis of codons 1 to 46.

nick occurred in a run of A residues, the polymerase would be expected to nick translate to the end of such a run and stop. Slippage of the primer strand, however, might then allow the polymerase to add on the extra A residue. Tending to support this hypothesis is the failure to find this type of frameshift among mutants that were induced with bisulfite reaction on pBR322 DNA molecules gapped in this same segment of the bla gene but without dATP present and from which the in vitro gap filling was omitted. The other type of frameshift mutation (CCC → CC), however, has been recovered regardless of the presence of dATP or the subsequent gap filling. This lesion might arise at a low frequency in vivo after transformation at sites of uracil residues in single-stranded DNA. Noteworthy in this regard is that the E. coli strain used for all transformations has an active DNA-uracil N-glycosylase.

NUCLEOTIDE MISINCORPORATION: THE GAP MISREPAIR REACTION

When polymerizing a complementary strand on a primer-template in vitro, purified DNA polymerases are observed to incorporate noncomplementary nucleotides at frequencies orders of magnitude higher than the spontaneous mutation rate measured in vivo. Furthermore, the frequency of misincorporation can be substantially increased by addition of manganese (II) ion to the reaction mixture or by creating large imbalances in the ratio of the four deoxyribonucleoside triphosphates (for discussion and references, see Loeb et al., 1979). Since a short, single-stranded gap in a duplex DNA molecule is a good substrate for many DNA polymerases, nucleotide misincorporation during repair (misrepair) of such a gap in vitro would generate a base substitution in the newly synthesized strand. If DNA synthesis is terminated by ligation of the nick once the polymerase has completely filled in the gap, this mutagenic reaction will be confined to those nucleotides within the single-stranded gap; i.e., the short stretch of single-stranded DNA becomes the target for mutagenesis. Ligation also serves to trap misrepaired molecules, since closed circular DNA is no longer a substrate for DNA polymerase's repair functions. In the absence of mismatch repair, the misincorporated nucleotide should segregate from the wild-type sequence at the first round of replication after transformation into an appropriate host cell, resulting in a mutant carrying a base substitution.

In preliminary experiments to examine the feasibility of in vitro mutagenesis by gap misrepair, short gaps have been constructed in pBR322 DNA at the single cleavage sites for the restriction enzymes HindIII and ClaI (Shortle and Botstein, in preparation). When purified circular DNA molecules gapped at the HindIII site were incubated with DNA polymerase I from M. luteus in the presence of all four deoxyribonucleoside triphosphates plus T4 ligase and ATP, the gap was repaired in greater than 90% of molecules as measured by

conversion of the DNA from an open to a covalently closed circular form. Restriction enzyme analysis of the plasmid progeny appearing after transformation of this repaired DNA revealed that all 40 independent plasmid isolates screened retained the HindIII restriction site. When the first nucleotide required to repair the gap (dATP) was omitted from the reaction mixture and manganese (II) ion was added in addition to magnesium ion (see Fig. 4), again greater than 90% of the circular DNA molecules were eventually converted to a covalently closed form. However, five out of 36 plasmid isolates (14%) recovered on transformation with this DNA were resistant to cleavage by the HindIII enzyme. The nucleotide sequence changes identified in these five HindIII site mutants (one of which has two additional mutations adjoining the restriction site) are consistent with a noncomplementary nucleotide being incorporated in place of the missing nucleotide dATP during gap filling [A → G(5); A → C(1); A → T(1)].

A higher rate of mutagenesis was observed at the ClaI site when the gap misrepair reaction was carried out in the absence of dCTP. Twelve of 36 plasmid isolates (33%) in this experiment had lost the ClaI restriction site, and the nucleotide sequences of six of these mutants have been determined. Again several isolates had additional base-substitution mutations flanking the restriction site, but in each case the pattern of base substitution [C → T(6); C → G(4); C → A(3)] is consistent with misincorporation at sites where the missing nucleotide (dCTP) would have been expected.

From these preliminary results, it would appear that the in vitro misrepair of specifically placed single-stranded gaps can be used for efficient site-specific mutagenesis. Since in the experiments described above all three base substitutions for A and for C have been induced, it is possible that, with the proper ratios of

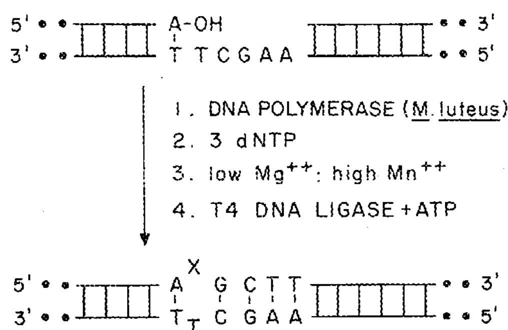

Fig. 4. Schematic diagram of the gap misrepair reaction. The DNA substrate is a short gap at the HindIII site in pBR322 DNA.

the three added deoxynucleoside triphosphates, the noncomplementary nucleotide that is inserted in lieu of the missing nucleotide can be controlled. The use of other DNA polymerases, such as the error-prone avian myeloblastosis virus reverse transcriptase, in this gap misrepair reaction may provide another device for controlling the pattern of misincorporation.

CONCLUSIONS

Single-stranded gaps constitute local disruptions of the uniform, highly ordered structure of DNA. Gaps can be enzymatically constructed in regions specified in advance. Two mutagenic reactions that depend on structural features unique to single-stranded gaps can thus be used for site-specific mutagenesis. In the future it seems reasonable to expect that additional, single-strand specific mutagenic reactions will be developed. As methods for precisely placing single-stranded gaps are further refined, it may become possible to routinely construct virtually any desired base substitution or frameshift mutation.

ACKNOWLEDGMENTS

This work was supported by grants from the American Cancer Society (MV-90) and from the National Institutes of Health (GM18973 and GM21253). D.S. was supported by a Postdoctoral Research Fellowship from the Helen Hay Whitney Foundation.

REFERENCES

Englund, P. T., Price, S. S., and Weigel, P. N., 1974, The use of the T4 DNA polymerase in identification of 3' terminal nucleotide sequences of duplex DNA, Methods Enzymol., 29:273.

Hayatsu, H., 1976, Bisulfite modification of nucleic acids and their constituents, Prog. Nucleic Acid Res. Mol. Biol., 16:75.

Loeb, L. A., Weymouth, L. A., Kunkel, T. A., Gopinathan, K. P., Beckman, R. A., and Dube, D. K., 1979, On the fidelity of DNA replication, Cold Spring Harbor Symp. Quant. Biol., 43:921.

McEntee, K., Weinstock, G. M., and Lehman, I. R., 1979, Initiation of general recombination catalyzed in vitro by the recA protein of Escherichia coli, Proc. Natl. Acad. Sci. U.S.A., 76:2615.

Miller, L. K., and Wells, R. D., 1972, Properties of the exonucleolytic activities of the Micrococcus luteus deoxyribonucleic acid polymerase, J. Biol. Chem., 247:2667.

Muller, W., Weber, H., Meyer, F., and Weissmann, C., 1978, Site-directed mutagenesis in DNA: Generation of point mutations in cloned β globin complementary DNA at the positions corresponding to amino acids 121 to 123, J. Mol. Biol., 124:343.

Parker, R. C., Watson, R. M., and Vinograd, J., 1977, Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis, Proc. Natl. Acad. Sci. U.S.A., 74:851.

Rawlins, D. R., and Muzyczka, N., 1980, Construction of a specific amber codon in the simian virus 40 T-antigen gene by site-directed mutagenesis, J. Virol., 36:611.

Shibata, T., DasGupta, C., Cunningham, R. P., and Radding, C. M., 1979, Purified *Escherichia coli* recA protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments, Proc. Natl. Acad. Sci. U.S.A., 76:1638.

Shortle, D., Koshland, D., Weinstock, G. M., and Botstein, D., 1980, Segment-directed mutagenesis: Construction in vitro of point mutations limited to a small predetermined region of a circular DNA molecule, Proc. Natl. Acad. Sci. U.S.A., 77:5375.

Shortle, D., and Nathans, D., 1978, Local mutagenesis: A method for generating viral mutants with base substitutions in preselected regions of the viral genome, Proc. Natl. Acad. Sci. U.S.A., 75:2170.

Shortle, D., and Nathans, D., 1979, Regulatory mutants of simian virus 40: Constructed mutants with base substitutions at the origin of DNA replication, J. Mol. Biol., 131:801.