Quantitative Multiplex Sequence Analysis of Mutational Hot Spots. Frequency and Specificity of Mutations Induced by a Site-Specific Ethenocytosine in M13 Viral DNA†

Vaseem A. Palejwala, Robert W. Rzepka, Devendranath Simha, and M. Zafri Humayun*

Department of Microbiology and Molecular Genetics, UMD—New Jersey Medical School, 185 South Orange Avenue MSB F607, Newark, New Jersey 07103-2714

Received November 18, 1992; Revised Manuscript Received February 8, 1993

ABSTRACT: We describe an assay for determining the frequency and specificity of mutations occurring at hot spots within a population of DNA molecules. The procedure consists of (a) annealing the DNA population with a labeled oligonucleotide designed to prime DNA synthesis at the mutational hot spot; (b) DNA elongation in the presence of a single dideoxynucleoside triphosphate together with 1–3 deoxynucleoside triphosphates, and (c) quantitation of all limit elongation products by high-resolution gel electrophoresis followed by autoradiography and computing densitometry. Derivation of mutational frequency and specificity over a wide range of values is demonstrated for M13 viral DNA mixtures containing defined proportions of wild-type and mutant DNAs, as well as for M13 viral DNA populations obtained by transfection of DNA bearing a defined site-specific ethenocytosine lesion. The assay is shown to yield results similar to those obtained by laborious clone-by-clone sequencing of viral progeny. The method is not affected significantly by several tested variables and appears to be suitable for use as a quantitative assay for sequence microheterogeneity at defined positions within DNA populations. Application of the methodology demonstrates that ethenocytosine, an exocyclic DNA lesion induced by carcinogens such as vinyl chloride and urethane, is a highly efficient mutagenic lesion with a mutational specificity expected for noninstructive lesions.

Although mutagenesis is regarded as a random process, it is well established that spontaneous as well as mutagen-induced mutations cluster at sites called mutational hot spots (Benzer, 1961; Farabaugh et al., 1978; Coulondre & Miller, 1977). The origin of mutational hot spots is not understood beyond a few correlations such as the occurrence of addition and deletion mutations at repeated sequences (Streisinger & Owen, 1985) and the association of 5-methylcytosine with spontaneous base substitutions (Coulondre et al., 1978). The observations that many mutagen-induced hot spots do not coincide with spontaneous hot spots and that each mutagen appears to induce its own set of hot spots have been used to argue that mutational hot spots arise from the misreplication of DNA damage sites. An extension of this argument is that spontaneous hot spots arise from unrepaired spontaneous DNA damage.

Traditional methodology for analysis of mutational hot spots requires the isolation of a large number of individual mutants in a suitable marker gene and sequence definition of each mutant by genetic methods (e.g., Farabaugh et al. (1978)) or direct DNA sequencing (e.g., Kunkel (1984) and Refolo et al. (1987)). Although useful for the initial detection of hot spots, sequence analysis of individual mutants is cumbersome when it comes to a detailed analysis of the effects of variables on particular hot spots. To obtain a statistically significant body of data, a large number of individual mutants needs to be sequenced, a process that can be sufficiently tedious as to make it a rate-limiting step.

In this article, we describe a simple experimental procedure for analyzing both the mutational frequency and specificity of a mutational hot spot in bacteriophage M13 (Messing, 1991) viral DNA. The hot spot analyzed here is induced by an ethenocytosine (ϵC)[1] residue introduced at a single, predetermined locus in M13 DNA for the purpose of examining its genotoxic properties. Figure 1 shows the sequence of form II′$_{\epsilon C}$ DNA, the site specifically adducted M13 DNA construct used in this study.

There is evidence to suggest that ϵC may be a major contributor to the carcinogenic activities of chemical carcinogens such as vinyl chloride and urethane (Jacobsen et al., 1989; Palejwala et al., 1991). In a previous study, the transfection of the form II′$_{\epsilon C}$ DNA showed that ϵC was highly mutagenic in *Escherichia coli* (Palejwala et al., 1991). Since the experimental strategy had precluded phenotypic screening of mutant phage, both the mutational frequency and specificity were determined by direct sequence analysis of randomly picked individual progeny phage clones (plaques). Most mutations were found to be base substitutions targeted for the site of the ϵC residue. These data were generally in agreement with earlier results obtained from mutational spectra produced by chemical treatment of M13 DNA with a chemical mutagen capable of inducing etheno lesions at cytosine and adenine residues (Jacobsen et al., 1989; Jacobsen & Humayun, 1990). The data obtained from both approaches suggest that ϵC has novel mutagenic properties that make it a valuable model for dissecting the mechanisms of induced mutagenesis.

MATERIALS AND METHODS

Biochemicals. Deoxyribonucleoside triphosphates (dNTPs) were purchased as aqueous solutions from Pharmacia (lot

† This work was supported in part by grants from the NIH and the American Cancer Society.

* Corresponding author. Telephone: (201) 982-5217. Fax: (201) 982-3644.

[1] Abbreviations: ϵC, 3,N^4-ethenocytosine; RF-DNA, replicative-form DNA; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; bp, base pair(s); nt, nucleotide(s).

0006-2960/93/0432-4105$04.00/0

M13AB28:
-ATGACCATGATTACGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTTGGCACTGGCC-

Form II':
-ATGACCATGATTACGAATTGTGGAGTTXTTATCGTGTGCAGCCAAGCTTGGCACTGGCC-
-TACTGGTACTAATGCTTAA ACGTCGGTTCGAACCGTGACCGG-

M13f2:
-ATGACCATGATTACGAATTGTGGAGTTCTTATCGTGTGCAGCCAAGCTTGGCACTGGCC-
AATAGCACACGTCGGTTCG-5' (Primer)

M13f2a:
-ATGACCATGATTACGAATTGTGGAGTTaTTATCGTGTGCAGCCAAGCTTGGCACTGGCC-
AATAGCACACGTCGGTTCG-5' (Primer)

M13f2t:
-ATGACCATGATTACGAATTGTGGAGTTtTTATCGTGTGCAGCCAAGCTTGGCACTGGCC-
AATAGCACACGTCGGTTCG-5' (Primer)

M13f2Δ1:

-ATGACCATGATTACGAATTGTGGAGTTTTATCGTGTGCAGCCAAGCTTGGCACTGGCC-
AATAGCACACGTCGGTTCG-5' (Primer)

FIGURE 1: DNA sequences of M13 derivatives and of site specifically modified DNA constructs used in this study. M13AB28 (Sambamurti et al., 1988) is the parent strain from which all constructs and variants used in this study are derived. The first base on the left is part of the initiating codon for the LacZ(α) gene segment (M13AB28 base 6217). The double-underlined sequence in the M13AB28 sequence (top) was replaced by a synthetic oligonucleotide to construct the form II' "gapped duplex" molecules, as well as to derive phages M13f2, M13f2a, M13f2t, and M13f2Δ1. In the form II'$_{\epsilon C}$ DNA construct, the central base X (in boldface), is ϵC; in the corresponding control construct (form II'$_C$), the central base X is normal cytosine. The construction and purification of these molecules have been described in detail elsewhere (Palejwala et al., 1991). The construction strategy placed the lesion (X) in the center of a 17-nt-long single-stranded region ("gap") of an otherwise duplex circular M13 DNA molecule. M13f2, M13f2a, M13f2t, and M13f2Δ1 were progeny phage recovered by transfection of form II'$_{\epsilon C}$ DNA into *E. coli*, as described, and they are used as authentic markers for wild-type progeny (M13f2), transversion mutant (M13f2a), transition mutant (M13f2t), and single-base deletion mutant (M13f2Δ1). The sequence and location of the 19-mer primer (in italics) used for multiplex sequence analysis are shown for M13f2, f2a, f2t, and f2Δ1.

no.: dATP, AE2050101; dTTP, AE2080101). Each lot was tested for interference from low-level cross-contamination with other dNTPs by carrying out trial elongations on standard DNAs. Concentrations were verified from UV absorption. Dideoxynucleoside triphosphates (ddNTPs) were purchased from U. S. Biochemicals (ddATP, lot no. 69723) or Pharmacia (ddGTP, lot no. 00414817). Dideoxynucleoside triphosphates from several other commercial sources proved to be unsatisfactory. Sequenase version 2.0 (a mutated version of phage T7 DNA polymerase devoid of 3′ → 5′ exonuclease activity) was purchased from U. S. Biochemicals. Although sequenase 2.0 is, in principle, devoid of 3′ → 5′ exonuclease activity, we have found it necessary to screen each enzyme lot for low levels of contaminating nonspecific nuclease activities. Primer degradation to shorter oligonucleotides was taken to indicate such contamination.

Bacterial and Phage Strains. Bacteriophage M13AB28 is essentially phage M13mp2 with the LacZ(α) gene segment, including the polylinker, derived from M13mp8. The *Escherichia coli* host KH2 provides ω complementation for β-galactosidase such that cells infected with M13AB28 derivatives produce β-galactosidase activity. Phage strain M13AB28 and *E. coli* strains KH2 (wild type) and KH2R (recA–) have been described elsewhere (Sambamurti et al., 1988; Palejwala et al., 1991, 1993). Phage M13f2 is a derivative of M13AB28 obtained by replacing 26 bp from within the polylinker sequence with a synthetic 25-mer. M13f2a (C→A), and M13f2t (C→T) are two mutants of M13f2 in which the same cytosine residue is substituted (see Figure 1). M13f2Δ1 is a single-base deletion affecting the same cytosine.

Construction of M13 DNA with a Site-Specific ϵC Lesion. Form II'$_{\epsilon C}$ DNA is M13 double-stranded replicative-form (RF) DNA bearing a single ϵC residue at the center of a 17-nt-long single-stranded region. Construction of this DNA has been described in detail elsewhere (Palejwala et al., 1991). Form II'$_C$ is a control DNA preparation containing normal cytosine in place of ϵC, which was constructed by procedures identical to those used for form II'$_{\epsilon C}$. Figure 1 shows the DNA sequence of the two constructs.

M13 DNA Preparations. M13 viral single-stranded (ss) DNAs used for the assays described here were prepared by the procedures of Refolo et al. (1987), except that the final ethanol precipitation step was replaced by two modified ethanol precipitation steps in which 2 M ammonium acetate replaced the usual 0.3 M sodium acetate. This modification is necessary to eliminate interference from the minute amounts of intracellular dNTPs that were found to copurify with phage during normal poly(ethylene glycol) precipitation procedures. M13 ssDNAs used for conventional sequence analysis, as well as RF-DNA used as a marker, were prepared as described (Refolo et al., 1987).

Transfection and Mass Culture of Progeny Phage. Competent cells were prepared and incubated with DNA at 4 °C for 30 min as described by Palejwala et al. (1991). For determining the number of infectious centers and the survival effects, aliquots of the transfection mix were plated on Xgal-IPTG plates as described previously. (It may be noted that Xgal-IPTG plates were used to monitor low levels of contaminating M13AB28 phage as described by Palejwala et al. (1991).) Under these transfection conditions, 10 ng of form II'$_{\epsilon C}$ DNA gave approximately 10 000 infectious centers, as did the form II'$_C$ DNA. (We have previously shown that ϵC does not significantly affect the survival of M13 DNA (Palejwala et al., 1991).) One-half of the transfection mix (0.5 mL; equivalent to 5 ng of transfected DNA or ~5000 infectious centers) and 0.1 mL of a fresh, saturated culture of the host *E. coli* strain were added to 5 mL of rich medium (LB), and the contents were incubated at 37 °C with vigorous aeration for 15 h. Cells and debris were removed by

centrifugation at 1500*g* for 10 min, and the progeny phage particles present in the supernatant were recovered by standard poly(ethylene glycol) precipitation procedures as outlined above.

Primer. All multiplex assays described here were carried out with a 19-mer primer complementary to the viral (plus) strand of M13f2. As shown in Figure 1, the first nucleotide incorporation event primed by this oligonucleotide occurs at the template site that constitutes the mutational hot spot under investigation. The 19-mer was synthesized by conventional solid-phase oligonucleotide synthesis chemistry on an Applied Biosystems machine and was purified by electrophoresis on high-resolution denaturing polyacrylamide gels. The primer was labeled with ^{32}P at the 5′-end by incubating the primer (750 ng, ~120 pmol) for 60 min at 37 °C in 30 μL of buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine, and 0.1 mM EDTA, pH 7.6) in the presence of 150 μCi of [γ-^{32}P]ATP (>5000 Ci/mmol; Amersham) and 40 units of T4 polynucleotide kinase (New England Biolabs). The reaction was terminated by a phenol extraction followed by three ether extractions. Labeling efficiency was checked by analyzing a small aliquot by PEI–cellulose thin-layer chromatography (solvent: 0.5 M ammonium bicarbonate). Under these conditions, almost all of the radioactivity is found to be transferred to the primer.

Quantitative Multiplex Sequence Analysis. The M13 standard DNA mix or progeny viral ssDNA population obtained by mass culture (4 pmol, ~10 μg) and the 5′-end-labeled 19-mer primer (2 pmol) were annealed in 40 μL of buffer (40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, and 50 mM NaCl) through the following sequential incubations: 3 min at 90 °C, 3 min at 4 °C (ice water bath), 20 min at 60 °C, slow cooling to room temperature over a period of 2–3 h, and slow cooling to 4 °C overnight in a refrigerator. Each multiplex elongation reaction was carried out in 10 μL of buffer (40 mM Tris-HCl, pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl, and 10 mM dithiothreitol) containing 0.1 pmol of the annealed template–primer complex (i.e., 0.2 pmol of template and 0.1 pmol of primer), 0.1 vendor unit of T7 DNA polymerase (sequenase version 2.0, U. S. Biochemicals), 167 μg/mL bovine serum albumin, and the indicated concentrations of dideoxy- and deoxynucleoside triphosphates. Reactions for determining mutational frequency had 1 μM each dATP and dTTP and 10 μM ddGTP. Reactions for determining specificity had 0.05 μM each dTTP and ddATP. The elongation mixes were incubated for 15 min at 37 °C before termination by addition of 10 μL of a formamide–dye mix (0.025% each of bromphenol blue and xylene cyanole in formamide). The elongation products were fractionated on 16% polyacrylamide–8 M urea gels and quantitated by densitometry of the autoradiographs (Kodak XAR film) as described (Simha et al., 1991).

RESULTS

Multiplex Sequencing Strategy. Figure 2 summarizes the specific strategy used for the quantitative determination of sequence changes occurring at the defined site N (see top of Figure 2) in M13 ssDNA. Cytosine at this position (i.e., N = C) represents the wild-type sequence. Thymine represents a transition mutation, and adenine represents a transversion (i.e., N = T for a transition; N = A for transversion). In addition to these mutations, a single-nt deletion (Δ1) can occur at this site. The specific strategy described here ignores C→G mutations for the sake of simplicity and also because this transversion is at best a minor component of the mutations induced by εC. However, any point mutation can be analyzed by minor variations of the same strategy.

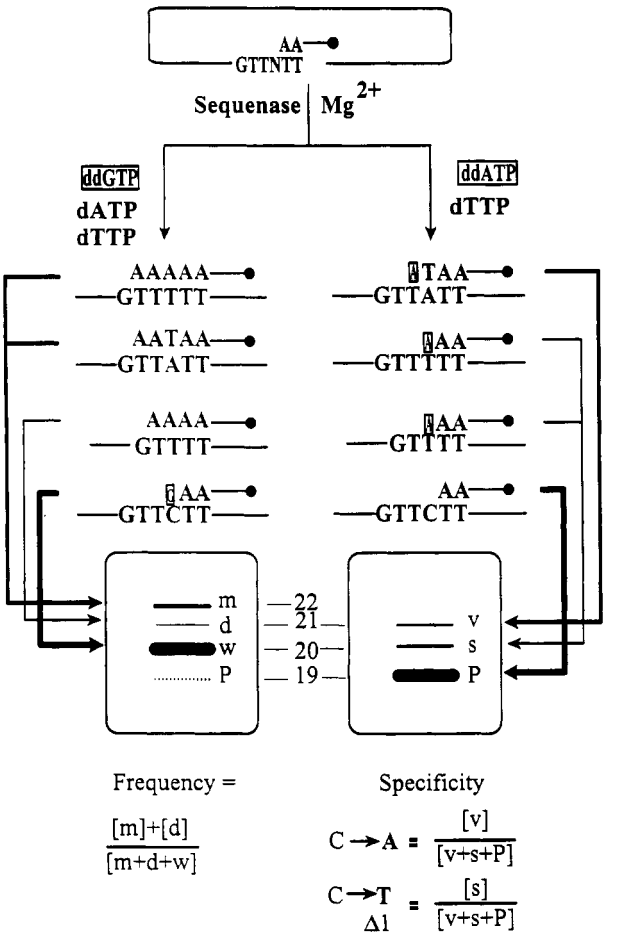

FIGURE 2: Quantitative multiplex sequence analysis of a mutational hot spot in M13 viral DNA. The first step consists of annealing a 19-mer primer (black dot denotes the 5′-^{32}P-labeled end) to a population of M13 viral DNA containing sequence microheterogeneity at the position N. The DNA population in the present example consists of four species: wild type (N = C), transition (N = T), transversion (N = A), and single-base deletion (Δ1). Mutational frequency is determined by primer elongation in the presence of ddGTP + dATP + dTTP (left). In the wild-type template fraction (N = C), the first nucleotide incorporated is ddGMP, giving rise to a chain-terminated (dead-end) 20-mer. In the transition template, the first nucleotide incorporated is dAMP, immediately followed by two more dAMPs to give a 22-mer limit elongation product. In the transversion template, the first nucleotide incorporated is dTMP followed by two dAMP residues, again yielding a 22-mer. In the Δ1 template, incorporation of two dAMP residues will yield a 21-mer. The elongation products are fractionated by electrophoresis on denaturing high-resolution polyacrylamide gels. Mutation frequency is determined (Figure 2, bottom left) as the ratio of the normalized densitometric signal present in the 22-mer and 21-mer bands (bands m and d) and in the 20-mer band (band w). In addition, the frequency of deletions is determined from the normalized ratio of the 21-mer with the wild-type and base substitution fractions. Mutational specificity, i.e., the relative proportions of transitions and transversions, is determined in a separate elongation reaction using the same template–primer complexes in the presence of ddATP + dTTP (right). Here, no elongation occurs in the wild-type template (N = C) because of the absence of dGTP or ddGTP. In the transition template (N = T), and in the Δ1 template, the incorporation of a ddAMP residue yields a dead-end 20-mer. In the transversion template, dTMP is incorporated followed by a ddAMP, yielding a terminal 21-mer. Mutational specificity is determined from the normalized densitometric signals present in the 21-mer for transversions (band v) and in the 20-mer (band s) for transitions. The transition frequency is corrected for contribution from the Δ1 template.

Figure 2 (left) shows that the products of primer elongation in the presence of ddGTP, dATP, and dTTP can be used to derive mutation frequency. Since the template base immediately upstream (5′) of the primer 3′-terminus is the site of

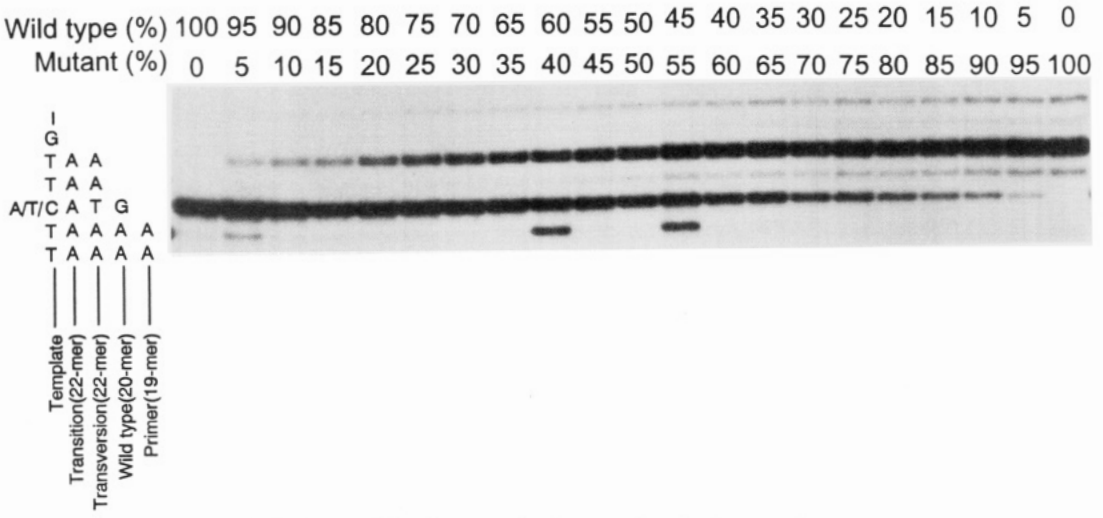

FIGURE 3: Multiplex sequence analysis for determining the mutation frequencies of mixtures of authentic wild-type and mutant DNAs. For this assay, two DNA preparations were prepared: (i) the wild-type fraction consisted of M13f2 DNA, and (ii) the mutant fraction consisted of a mix of three different mutant DNAs (M13f2a–M13f2t–M13f2Δ1 in a 50:43:7 ratio). The wild-type and mutant fractions were mixed in various proportions to simulate mutation frequencies ranging from 0% (lane 1, far left lane) to 100% (lane 21, far right lane) in increments of 5%. For example, in the lane labeled 30%, the proportions of the four DNA species are M13f2:M13f2a:M13f2t:M13f2Δ1 = 70:15:13:2. The DNA sequences of the templates, primers, and elongation products are shown at left. When the DNA is pure wild type (0% mutants, lane 1), the only product of DNA elongation is a 20-mer. In lane 2 (5% mutants), a faint band corresponding to the 22-mer product expected for the mutants can be seen. This band increases in intensity in proportion to the mutant fraction, and at the expense of the 20-mer product, until the 20-mer product completely disappears in lane 21, where all of the DNA is mutant. In addition to the expected 22-mer, traces of a 21-mer (from the deletion mutant) can also be seen at high mutant concentrations. At high mutant concentrations, bands with a lengths corresponding to 24-mer also accumulate. Since sequenase version 2 is devoid of 3′ → 5′ exonuclease activity, these longer products might represent the accumulation of slippage–elongation products (e.g., slippage at the TTTTT sequence in the transition template) because all four dNTPs are not present in the reaction. For the purposes of determining mutational frequency, all products above a 20-mer arise from the mutant fraction. It can be seen that in two lanes (corresponding to 40% and 55% mutants) significant amounts of the starting primer have remained unelongated. This is most likely the result of incomplete elongation (probably due to poor mixing of the particular test tubes at the start of the reaction), as other experiments do not show unelongated primer. Note that only elongated products are evaluated in this assay; residual unelongated primer is unlikely to cause interference except under special conditions.

the mutation, the first nucleotide incorporated is different for the wild-type and mutant templates.

For the template fraction constituting the wild-type progeny (N = C), the incorporation of ddG results in chain termination, resulting in a dead-end 20-mer (product w, bottom left). For the template fraction making up the transition mutation (N = T), the first base incorporated is adenine. In this case, chain elongation can continue for two more bases because the two bases upstream of the mutation site are also thymines, yielding a 22-mer as the product. (The absence of dCTP prevents further elongation.) Similarly, for the template fraction that constitutes the transversion mutation (N = A), the first base incorporated is thymine. Elongation continues for two more bases because of the availability of dATP, again yielding a 22-mer. In the template fraction making up deletions (Δ1), elongation occurs up to a 21-mer because of the presence of dATP.

Thus, in the wild type, chain elongation is terminated at the 20-mer level, whereas in the substitution and deletion mutations, the products are 22-mers and a 21-mer, respectively. The relative proportions of the 20-mer and of the longer products would therefore represent the wild-type and mutant fractions, respectively.

Figure 2 (right) shows that essentially the same strategy can be used derive the relative frequency of the transition and transversion mutations. Here primer elongation is carried out in the presence of ddATP and dTTP. On the wild-type template (N = C), no elongation can occur because dGTP (or ddGTP) is not available. On the transition template (N = T), as well as on the deletion template (Δ1), ddAMP incorporation generates a dead-end 20-mer. On the transversion template (N = A), the first nucleotide added is dTMP, followed by ddAMP to yield a dead-end 21-mer. Therefore, the relative proportions of the 20-mer (corrected for the Δ1 fraction) and of the 21-mer should be representative of the frequency of occurrence of the two base substitution mutations in the population.

Determination of Mutation Frequency. To test the mutation frequency assay, we individually prepared "standard" ssDNAs from M13f2 (wild type, N = C), M13f2t (transition, N = T), M13f2a (transversion, N = A), and M13f2Δ1 (deletion). The purified DNAs were mixed to yield a series of mutation frequency standards in which the proportion of the mutants was increased from 0% to 100% in increments of 5%. The mutant fraction was a mixture of DNAs derived from three phages in the ratio, f2a:f2t:f2Δ1 = 50:43:7. For example, the mix corresponding to 30% mutation frequency consisted of 70% M13f2 (wild type) DNA, 15% M13f2a DNA, 13% M13f2t DNA, and 2% M13f2Δ1 DNA. Figure 3 shows that when all of the DNA is of the wild type, elongation proceeds to the 20-mer, but no further. Products longer than the 20-mer (mostly the 22-mer) increase in intensity relative to the 20-mer in direct proportion to the mutant fraction, until complete disappearance of the 20-mer in the 100% mutant DNA.

Figure 4 is a plot of the experimentally determined mutation frequencies calculated from densitometric analyses of the data of the type shown in Figure 3. Figure 4 shows that the experimental strategy can in fact determine mutation frequency over a wide range of values with reasonable accuracy.

Determination of Mutational Specificity. Figure 5 shows an analysis of mutational specificity using DNA mixes. For these experiments, seven DNA mixes with a wild-type to mutant ratio of 70:30 were prepared. In this case, the proportion of the transition mutation (M13f2t) was varied from 0 to 30% in steps of 5%, with the balance made up of

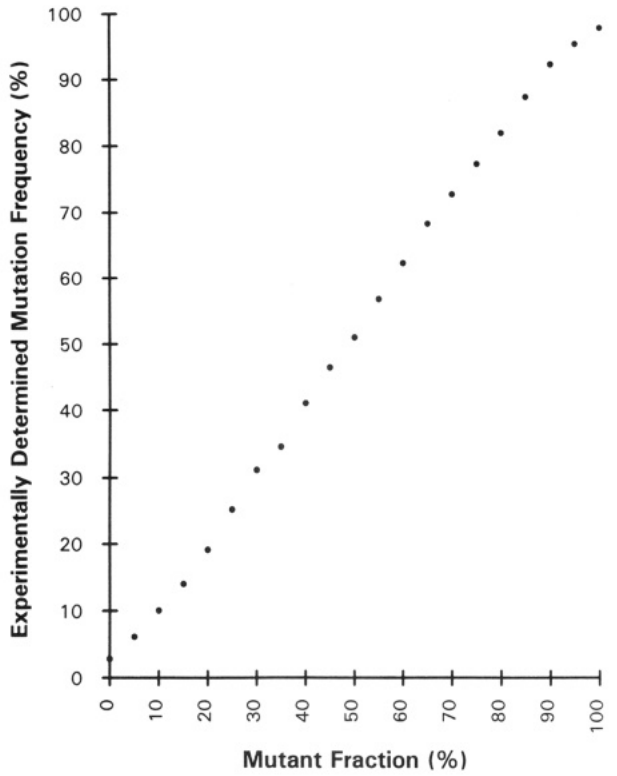

FIGURE 4: Comparison of the experimentally determined mutation frequencies with expected values. Autoradiographs such as the one shown in Figure 3 were analyzed using a Molecular Dynamics Model 300B computing densitometer as described (Simha et al., 1991). Mutation frequencies were calculated from the normalized total densitometric signal present in the 20-mer band and bands longer than the 20-mer as described in Figure 2 and in the legend for Figure 3. Average values obtained from three independent experiments are plotted.

the transversion mutation (M13f2a) such that the total mutant fraction was 30%. Lane 5 of Figure 5 (0% transitions and 30% transversions) shows that, when transversion is the only mutant present, the only elongation product is a 21-mer. The 20-mer product first appears in lane 6 (5% transition and 25% transversion) and increases in proportion to the transition mutant until it is the only product found in lane 11 (30% transition and 0% transversion). Table I summarizes results

Table I: Determination of Mutational Specificity by Multiplex Sequencing

expected ratio (transverion:transition)[a]	observed ratio (transverion:transition)[b]
30:0	31:1
25:5	25:4
20:10	20:8
15:15	15:13
10:20	9:19
5:25	4:24
0:30	1:27

[a] Transversion is C→A; transition is C→T. Seventy percent of the DNA in all mixes was of the wild type. [b] Observed ratios are the averages of four independent experiments.

of a densitometric analysis of data of the type shown in Figure 5 and shows that the method permits the determination of specificity with reasonable accuracy.

Effect of Mass Phage Culture on Mutational Frequency and Specificity. Since a major advantage of the experimental strategies described here is their applicability to progeny phage derived from mass liquid culture, it is necessary to determine whether these culture conditions can distort the mutation frequency and specificity initially produced by mutation fixation. For example, if the wild-type progeny has a replicative advantage over the mutants, then after many generations the mutants will be underrepresented in the progeny pool. To examine this issue, we transfected DNA mixes consisting of various proportions of M13f2, M13f2t, and M13f2a into competent cells and diverted one-half of the transfection mix into liquid culture as described in Materials and Methods. Progeny phage were harvested at 5-, 10-, and 20-h intervals and were analyzed for mutation frequency and specificity. The results, as summarized in Table II, showed that no distortions in frequency or specificity occurred for these mutants as a result of mass culture up to 20 h after transfection.

Analysis of a Mutational Hot Spot Induced by a Site-Specific εC Lesion Contained in M13 DNA. Form II′$_{\epsilon C}$ DNA containing a site-specific εC residue (Figure 1) and form II′$_C$, the corresponding control construct with normal C in the place of εC, were transfected into *E. coli* KH2 (recA$^+$) and KH2R (recA−) cells as described (Palejwala et al., 1991). Aliquots

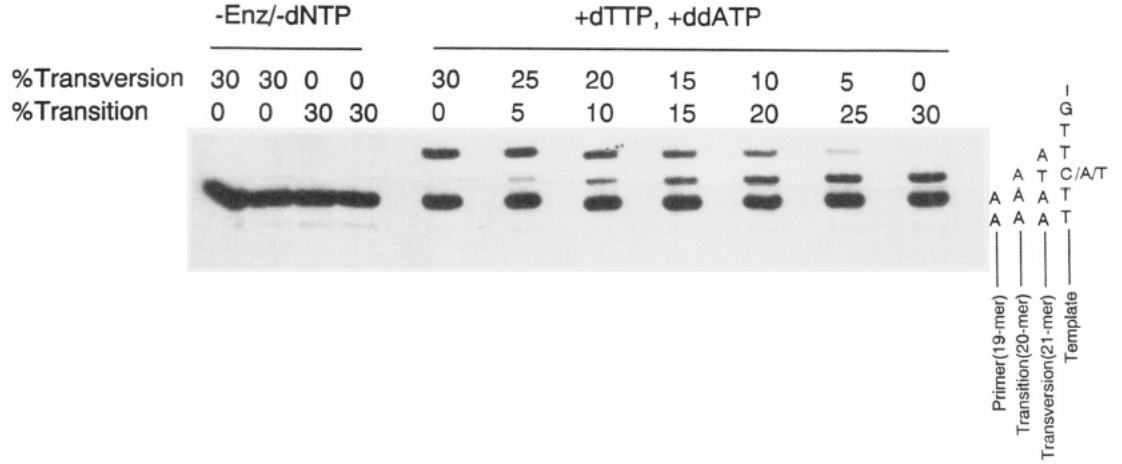

FIGURE 5: Quantitative multiplex sequencing assay for determining mutational specificity. Wild-type (M13f2), transition (M13f2t), and transversion (M13f2a; see Figure 1) DNAs were mixed to create seven DNA mixes in which a constant 70% of the DNA was of the wild type and the remainder was made up of one or both of the mutants as indicated. The four lanes labeled –Enz/–dNTP (lanes 1–4 from left) are controls showing lack of primer elongation in the absence of polymerase or nucleotide precursors. Lanes 5–11 show DNA mixes with the proportion of transitions ranging from 0% (lane 5) through 30% (lane 11). In the absence of transitions (lane 5), as expected (see Figure 2, right), the only elongation product is a 21-mer. An increasing proportion of the transition mutant results in increasing levels of the 20-mer, until it becomes the only product in lane 11.

Table II: Effect of Replication on Mutation Frequency and Specificity[a]

expected frequency (%)	observed mutation frequency (%) after replication (h)			
	0	5	10	20
	A. All Mutations			
20	18 ± 1	21 ± 0	22 ± 2	22 ± 0
40	40 ± 2	39 ± 1	44 ± 2	44 ± 1
60	61 ± 2	62 ± 0	62 ± 1	63 ± 2
80	81 ± 2	80 ± 1	83 ± 3	81 ± 2
100	99 ± 0	99 ± 0	99 ± 0	99 ± 0
	B. C→A Transversions			
10	11 ± 1	11 ± 1	10 ± 1	10 ± 1
20	21 ± 2	17 ± 1	18 ± 0	17 ± 1
30	29 ± 0	31 ± 0	31 ± 0	31 ± 0
40	37 ± 0	39 ± 1	39 ± 0	39 ± 0
50	48 ± 0	46 ± 0	46 ± 0	46 ± 0
	C. C→T Transitions			
10	11 ± 1	9 ± 1	8 ± 1	8 ± 1
20	21 ± 1	23 ± 1	24 ± 1	23 ± 1
30	30 ± 0	29 ± 0	29 ± 0	29 ± 0
40	39 ± 1	38 ± 1	39 ± 0	39 ± 1
50	49 ± 0	50 ± 1	51 ± 0	50 ± 1

[a] Values shown are averages (±standard deviation; numbers rounded off to nearest integer) of three independent elongation reactions. Five DNA mixes containing the indicated proportions of transition (C→T), transversion (C→A), or wild-type DNA were used in this study. In the mixes, the ratios for wild-type:transitions:transversions were as follows: 80:10:10; 60:20:20; 40:30:30; 20:40:40; 0:50:50. Ten nanograms of each DNA mix was transfected into *E. coli* KH2 cells, and one-half of the transfection mix was diverted for mass culture as described in Materials and Methods, except that the incubation time at 37 °C in rich medium was varied from 0 to 20 h. The progeny DNA were harvested and subjected to multiplex analyses as described.

of the transfected cells were plated on agar to determine survival effects, and one-half of the transfection mix was diverted for mass liquid culturing of progeny phage as described in Materials and Methods. Figure 6 shows the results of multiplex sequence analysis of the frequency and specificity of mutation in progeny DNAs obtained from these experiments. Table III summarizes data obtained by conventional clone-by-clone sequencing methods, as well as data obtained by the multiplex sequencing methods. Although a majority of mutations induced by εC are base substitutions, the data here show that an appreciable fraction of the mutations are simple deletions (Δ1). The data here confirm that εC is a highly efficient, recA-independent mutagenic lesion, and they show that multiplex sequencing yields mutational frequency data that are close to those obtained by the much more laborious clone-by-clone sequencing approach. Similarly, the specificity of εC-induced mutations (i.e., predominantly C→T mutations) is very similar to that reported from the analysis of a large number of individual mutants obtained by chloroacetaldehyde treatment to induce etheno lesions in M13 DNA at random (Jacobsen et al., 1989; Jacobsen & Humayun, 1990).

DISCUSSION

The assay described here was developed to analyze the characteristics of mutational hot spots induced by site-specific mutagenic DNA adducts. The mutagenic efficiency of site-specific adducts has a broad range, depending on the type of adduct as well as the genetic background of the cell. εC, the particular lesion investigated here, has a high mutagenic efficiency (~30%) such that the progeny phage can be directly analyzed without any selection or enrichment for mutants. Our unpublished data indicate that the assay gives reasonably accurate results for mutation frequencies as low as 2–3%. For quantitation of lower mutation frequencies, one will have to incorporate a mutant enrichment regime.

Although the specific procedures described here are designed to quantitate the two base substitution mutations and a single-base deletion mutation provoked by εC in this experimental system, the principles can be readily extended to devise assays for any point mutation. For example, to detect C→G transversions, an additional elongation reaction similar to the one for mutational frequency (Figure 2, left) can be carried out with ddCTP in place of ddGTP. In this case, the C→G

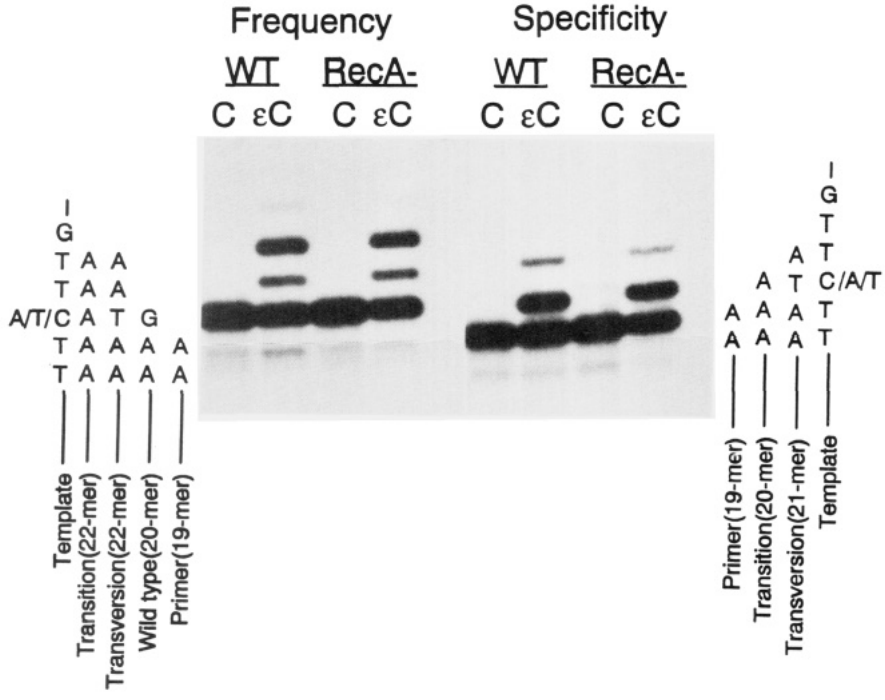

FIGURE 6: Frequency and specificity of mutations induced by a site-specific εC residue carried on M13 DNA. Form II′$_{εC}$ (lanes marked εC) or form II′$_C$ DNA (lanes marked C) was transfected into *E. coli* KH2 (wild type; lanes marked WT) or KH2R (recA–) cells and progeny phage harvested after mass culture, as described in Materials and Methods. The pooled DNA from each transfection was subjected to the mutational frequency (left) and specificity (right) assays as described in Figure 2 and in Materials and Methods.

Table III: DNA Sequence Analysis of Progeny Phage Obtained by Transfection of Form II′$_{\epsilon C}$ DNA into *E. coli* KH2 (Wild-Type) Cells

expt[a]	sequencing method	number of phage sequenced	mutation frequency (% (±SD))[e]	mutational specificity (% (±SD))[f] C→A	C→T	Δ1
1a	individual[b]	53	32	19	11	1.9
1b	individual	53	30	13	13	1.9
2a	individual	55	20	3.6	16	0
2b	individual	55	13	5.5	5.5	1.8
3	multiplex[c]	~5000[d]	36 (±2)	3 (±2)	22 (±2)	8 (±1)
4	multiplex	~5000	35 (±2)	4 (±1)	19 (±1)	8 (±1)

[a] Experiments 1a, 1b, 2a, and 2b: Plaques were picked at random from plates as described previously (Palejwala et al., 1991). Experiments 1a and 1b are two sets of phage clones (53 clones per set) from the same transfection experiment. Experiments 2a and 2b are similarly two sets (55 clones each) from a second transfection experiment. Data shown for experiment 1a are from Palejwala et al. (1991), except that a resequencing of the set revealed one of the mutants (previously classified as a C→T event) to be a single-base deletion; all other data were obtained as a part of this study. [b] Plaques were picked at random from indicator plates, and the DNA sequence of phage derived from each plaque was determined by a conventional dideoxy chain termination method as described previously (Palejwala et al., 1991). [c] Multiplex sequence analysis was carried out as described in Figure 2 and in Materials and Methods. [d] The number of phage clones sequenced in the multiplex assay are estimates based on the average number of infectious centers (~10 000) obtained by transfecting 10 ng of form II′$_{\epsilon C}$ DNA by procedures described by Palejwala et al. (1991). One-half of the transfection mix, equivalent to approximately 5000 infectious centers, was used for mass culture and subsequent preparation of progeny pool DNA. [e] Mutation frequency for experiments 1a, 1b, 2a, and 2b is the percentage representing mutants over the total number of progeny phage sequenced. Mutation frequency for experiments 3 and 4 was derived by the multiplex frequency assay described in Figure 2 (left), in Figure 3, and in Materials and Methods. Values shown are averages (±standard deviation; numbers rounded to nearest integer) of three multiplex frequency assays carried out on the progeny phage DNA pool from each experiment. [f] Mutational specificity for experiments 1a, 1b, 2a, and 2b is expressed as percentages representing the number of each of the three mutants over the total number of progeny phage sequenced. Mutational specificity for experiments 3 and 4 was determined by the specificity assay as described in Figure 2 (right), in Figure 5, and in Materials and Methods. The Δ1 mutation refers to progeny with a TTTT sequence in place of the TTεCTT sequence present in transfected DNA. This mutation could have arisen from a polymerase skip of εC in vivo. However, a C→T mutation followed by strand slippage (e.g., Refolo et al. (1987)) in the resulting TTTTT sequence can also be a source for this mutation. In addition to the three types of mutations shown, the following rare atypical mutations were also seen in experiments 1b and 2a: a single C→G mutation found in experiment 1b; a single untargeted base substitution (T→G) found in association with a C→T mutation (5′-εCTTATCGTGTGCA to 5′-TTTATCGTGGGCA), also in experiment 1b; a large deletion, resulting in loss of almost all of the *E. coli* Lac sequence present in M13, in experiment 2a. (The large deletion is believed to be an untargeted background event occasionally observed upon transfection of control DNA as well as εC-containing DNA (Palejwala et al., 1991).) It may be noted that summation of the frequencies for each of the three mutants gives a mutation frequency figure that is slight lower (up to 5% lower) than that obtained by the mutation frequency assay. For example, in experiment 3, the mutation frequency assay yields the figure shown (36%), whereas the figure obtained by summation of the frequencies of the three mutants is 33%. This difference is most likely due to technical reasons.

transversions will yield a 20-mer, whereas C→A and C→T mutations will both yield 23-mers and the Δ1-bp mutation will yield a 22-mer. We have in fact used such an assay, but we did not detect significant induction of C→G transversions by εC (data not shown).

The major advantage of the multiplex assay lies in its ability to rapidly and quantitatively analyze a selected set of changes known to occur at specific sites within large DNA populations. In the particular set of experiments described here, the progeny phage pools were derived from about 5000 infectious centers in each experiment, a number that can be readily increased by transfecting more DNA. Therefore, the data are in principle equivalent to those obtainable by sequence analysis of a very large number of individual clones. The assay thus offers the advantage of much greater potential accuracy made possible by analyzing large populations. The specificity of base substitutions induced by εC (i.e., predominantly C→T events) is in good agreement with the results obtained by sequencing a large number of mutants isolated from randomly adducted M13AB28 DNA (Jacobsen et al., 1989; Jacobsen & Humayun, 1990). In contrast, there are significant fluctuations in the specificity data obtained by clone-by-clone sequencing of progeny obtained by transfecting site specifically modified DNA (Table III, experiments 1a, 1b, 2a, and 2b) which are most likely attributable to the relatively small number of progeny clones sequenced in each set. When sequence data are limited, specificity deduced from mutational spectra (i.e., random adduction approach) may be more reliable than that deduced from single-site experiments.

The multiplex assay described here should make it possible to address questions that were essentially inaccessible because of the time and effort involved in clone-by-clone sequencing. As described in the following article in this issue (Palejwala et al., 1993), we have used this technology to describe the modulation of εC mutagenesis by UV in wild-type as well as in recA– cells. Although the focus in our laboratory has been on mutational mechanisms, the assay lends itself to any situation where sequence microheterogeneity within DNA populations needs to be quantitated.

The results presented in this article confirm that εC is a highly efficient premutagenic lesion and that the specificity of mutations induced by this lesion is that expected for noninstructive adducts, as previously deduced from random adduction experiments (Jacobsen et al., 1989; Jacobsen & Humayun, 1990). Although the random adduction experiments did not reveal significant frameshift mutagenesis attributable to εC, the application of multiplex sequencing technology as described here suggests that single-base deletions can also be induced by εC in some sequence contexts.

ACKNOWLEDGMENT

We thank Dr. A. K. Basu for drawing our attention to ethenocytosine-mediated single-base deletions.

REFERENCES

Benzer, S. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 403–415.
Coulondre, C., & Miller, J. H. (1977) *J. Mol. Biol. 117*, 577–606.
Coulondre, C., Miller, J., Farabaugh, P., & Gilbert, W. (1978) *Nature 274*, 775–780.
Farabaugh, P. J., Schmeissner, U., Hofer, M., & Miller, J. H. (1978) *J. Mol. Biol. 126*, 847–857.
Jacobsen, J. S., & Humayun, M. Z. (1990) *Biochemistry 29*, 496–504.
Jacobsen, J. S., Perkins, C. P., Callahan, J. T., Sambamurti, K., & Humayun, M. Z. (1989) *Genetics 121*, 213–222.
Kunkel, T. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1494–1498.
Messing, J. (1991) *Gene 100*, 3–12.
Palejwala, V. A., Simha, D., & Humayun, M. Z. (1991) *Biochemistry 30*, 8736–8743.
Palejwala, V. A., Rzepka, R. W., & Humayun, M. Z. (1993) *Biochemistry* (following article in this issue).
Refolo, L. M., Bennett, C. B., & Humayun, M. Z. (1987) *J. Mol. Biol. 193*, 609–630.
Sambamurti, K., Perkins, C. P., Callahan, J. T., Luo, X., Jacobsen, J. S., & Humayun, M. Z. (1988) *Genetics 120*, 863–873.
Simha, D., Palejwala, V. A., & Humayun, M. Z. (1991) *Biochemistry 30*, 8727–8735.
Streisinger, G., & Owen, J. (1985) *Genetics 109*, 633–659.