The aflatoxin B_1 formamidopyrimidine adduct plays a major role in causing the types of mutations observed in human hepatocellular carcinoma

Maryann E. Smela*, Michelle L. Hamm*†, Paul T. Henderson*, Constance M. Harris‡, Thomas M. Harris‡, and John M. Essigmann*§

*Biological Engineering Division and Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building 56, Room 669, Cambridge, MA 02139; and ‡Department of Chemistry, Vanderbilt University, Nashville, TN 37235

A G to T mutation has been observed at the third position of codon 249 of the p53 tumor-suppressor gene in over 50% of the hepatocellular carcinoma cases associated with high exposure to aflatoxin B_1 (AFB$_1$). Hypotheses have been put forth that AFB$_1$, in concert with hepatitis B virus (HBV), may play a role in the formation of, and/or the selection for, this mutation. The primary DNA adduct of AFB$_1$ is 8,9-dihydro-8-(N^7-guanyl)-9-hydroxyaflatoxin B_1 (AFB$_1$-N7-Gua), which is converted naturally to two secondary lesions, an apurinic site and an AFB$_1$-formamidopyrimidine (AFB$_1$-FAPY) adduct. AFB$_1$-FAPY is detected at near maximal levels in rat DNA days to weeks after AFB$_1$ exposure, underscoring its high persistence *in vivo*. The present study reveals two striking properties of this DNA adduct: (*i*) AFB$_1$-FAPY was found to cause a G to T mutation frequency in *Escherichia coli* approximately 6 times higher than that of AFB$_1$-N7-Gua, and (*ii*) one proposed rotamer of AFB$_1$-FAPY is a block to replication, even when the efficient bypass polymerase MucAB is used by the cell. Taken together, these characteristics make the FAPY adduct the prime candidate for both the genotoxicity of aflatoxin, because mammalian cells also have similar bypass mechanisms for combating DNA damage, and the mutagenicity that ultimately may lead to liver cancer.

Aflatoxin B_1 (AFB$_1$), one of the most potent known liver carcinogens, is produced by the common soil fungus *Aspergillus flavus*. Exposure to this toxin is high in regions of the world where certain foods are improperly stored (1). Hepatitis B virus (HBV) is also common in these regions, and epidemiological evidence indicates that there is a synergistic interaction between AFB$_1$ exposure and HBV infection on the induction of hepatocellular carcinoma (HCC). In over 50% of HCC cases studied in these areas, a characteristic G to T mutation is observed at the third position of codon 249 of the p53 tumor-suppressor gene (2, 3). Whether this specific sequence is an exceptional target for mutations caused by AFB$_1$ or whether the mutation is selected for once it occurs remains to be determined. However, each of these scenarios shares the fundamental early step involving generation of a G to T mutation.

There is substantial evidence that AFB$_1$-induced G to T mutations in cellular *ras* genes may also be a step in transformation of normal cells to malignant cells (4–6). In humans these mutations occur at the first and second positions of codon 12 in the Ha-*ras* protooncogene (7) and they are in sequence contexts similar, but not identical, to that of codon 249 in p53.

Many studies have defined the mutational spectrum produced after exposure of cells to either the epoxide, which is the toxicologically relevant natural metabolite of AFB$_1$ (8), or to other electrophilic derivatives that serve as models for the epoxide (9). The G to T mutation is predominantly observed (2, 3, 8–19). Studies of mutational landscapes, by their nature, do not elucidate which specific chemical form of AFB$_1$-DNA adduct is responsible for a given mutation (1). The epoxide reacts with DNA to form the primary AFB$_1$-DNA adduct, 8,9-dihydro-8-(N^7-guanyl)-9-hydroxyaflatoxin B_1 (AFB$_1$-N7-Gua) (20–26),

which can break down into two secondary lesions, the apurinic (AP) site or the AFB$_1$-formamidopyrimidine (FAPY) adduct (Fig. 1) (20). The mutational properties of the AFB$_1$-N7-Gua adduct have been studied in *Escherichia coli* (16), where it primarily causes G to T mutations, although at a very low frequency (4%). An interesting property of this adduct is that a significant number of the total mutations observed (13%) occurs 5′ to the site at which the adduct forms. NMR data (27) may explain this mutational asymmetry in that the AFB$_1$-position of the adduct is intercalated 5′ to the attached DNA base. The AFB$_1$-FAPY adduct is similar in structure to AFB$_1$-N7-Gua, but data show that these two adducts alter the structure of DNA differently (28, 29). These differences in DNA secondary structure may lead to differences in lethality, mutagenicity, and repair of the two chemical forms of the adduct.

The AFB$_1$-FAPY adduct itself consists of an equilibrium mixture of two presumably rotameric forms (FAPY mix) that are separable by HPLC into products termed FAPY major and FAPY minor (Fig. 1 *Inset*). Over time at 4°C or warmer, the equilibrium of this mixture shifts toward FAPY major. Stone and coworkers (29) attempted to resolve the NMR structure of FAPY and at first obtained complex results, presumably because of the presence of multiple species. After several days at 4°C, the NMR spectrum simplified to reveal presumably the structure referred to as FAPY major in this work. Although the exact chemical nature of FAPY minor remains elusive, a structure for one of the possible rotamers has been proposed (Fig. 1) based on data obtained for the imidazole ring-opened form of 7-methylguanine. This adduct forms two rotamers, indicating that structural duality may be a common characteristic of formamidopyrimidines (30–33). Under the conditions by which the AFB$_1$-FAPY oligonucleotides are made and stored, an equilibrium exists between the major and minor presumed rotamers at a ratio of 2:1, and these oligonucleotides can be used in studies of mutagenicity.

To evaluate the mutagenic properties of AFB$_1$-FAPY, we took advantage of the fact that bacterial cells have mechanisms such as the SOS response for coping with DNA damage. After exposure of *E. coli* to UV light, the UmuDC operon is derepressed, allowing the error-prone bypass polymerase (polV) to be transcribed (34–38). MucAB, an analogous bypass polymerase in *Salmonella typhimurium*, is regulated in a similar manner. These bypass polymerases incorporate bases, often incorrectly, opposite the site of DNA damage, allowing for cell survival in a
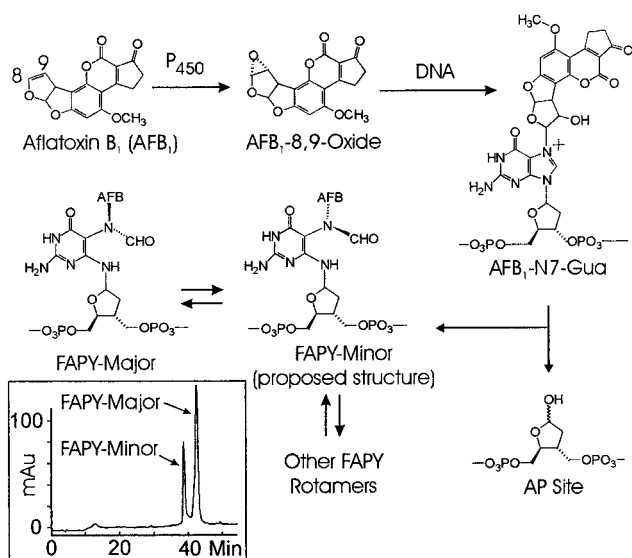
BIOCHEMISTRY

Fig. 1. Aflatoxin biochemistry. AFB$_1$ is activated to the reactive *exo*-8,9-epoxide form by cytochrome P$_{450}$ in the liver. The epoxide can react at the N7 position of guanine in DNA to form the primary AFB$_1$-N7-Gua adduct. This product can break down to form two secondary lesions, the AP site and the ring-opened AFB$_1$-FAPY adduct. The FAPY adduct itself consists of two proposed rotameric forms, FAPY major and FAPY minor. FAPY minor is formed first, subsequently equilibrating to a 2:1 ratio of FAPY major:FAPY minor. The FAPY major structure can be found in Stone and coworkers (29), and the FAPY minor structure is yet unresolved, but one possible structure is depicted here. (*Inset*) Separation of AFB$_1$-FAPY rotamers. HPLC was performed on a 13-base oligonucleotide containing one single AFB$_1$-FAPY adduct. Under the HPLC conditions described, FAPY major can be separated from FAPY minor. The equilibrium mixture shown here is roughly 2:1 major to minor. This mixture was maintained throughout the experiments described in this article and is referred to as FAPY mix. The major and minor peaks were collected and tested as indicated.

time of crisis. They can also be used as tools to determine which type of base is most often put opposite a specific DNA lesion. Because mammalian cells have similar mechanisms to cope with DNA damage, we used cells induced to express these bypass polymerases in this work.

Materials and Methods

Enzymes and Chemicals. *Eco*RI, *Hae*III, *Hinf*I, T4 polynucleotide kinase (ATP: 5′-dephosphopolynucleotide 5′phosphotransferase, 2.7.1.78), T4 DNA ligase [poly(deoxyribonucleotide):poly(deoxyribonucleotide) ligase (AMP forming), 6.5.11], exonuclease III (exodeoxyribonuclease III, 3.1.11.2), and uracil DNA glycolsylase (UDG) were from New England Biolabs. [γ-^{32}P]dATP and [α-^{35}S]dATP were from Perkin–Elmer. *Sfu*I and G-50 and G-25 Sephadex columns were from Roche Molecular Biochemicals. Cetricon-3 concentrators were from Amicon. ZipTip C18 pipet tips were from Millipore. Isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) were from Gold Biotechnology, St. Louis. *N,N*′-dimethylformamide (DMF) was from Fisher Scientific.

DNAs and Cell Strains. M13mp7L2 was obtained from C. W. Lawrence (39). A 13-base oligonucleotide, d(CCTC<u>TTC-GAA</u>CTC) (boldface G = site of adduct; underscore = *Sfu*I site), and a 53-base uraciled oligonucleotide scaffold, d(AAAACGACGGCCAGUGAAUU<u>GAGUUCGAAGAGG</u>-CACUGAAUCAUGGUCAUAGC) (the underlined sequence is complementary to the 13-base insert, and the 20 bases on either side of this sequence are complementary to the flanking

regions of the M13mp7L2 genome), as well as mass spectral internal standards, a 12-base oligonucleotide, d(ATTGGTA-AGTGC), and a 14-base oligonucleotide d(ATTGGTAAGTC-CGA), were obtained from GIBCO/BRL and purified with PAGE. *E. coli* strains used were DL7 (AB1157; *lac*ΔU169), DL7/pGW16 (DL7; pGW16 *mucAB*) (this laboratory), GW5100 (JM103, P1$^-$) (G. Walker, Massachusetts Institute of Technology), and NR9050 (F′ *prolacIZ*Δ*M15*, Δ*prolac*, *suB*) (R. Schaaper, National Institutes of Health).

Construction and Characterization of AFB$_1$-N7-Gua- and AFB$_1$-FAPY-Containing Oligonucleotides. Oligonucleotides containing single AFB$_1$-N7-Gua and AFB$_1$-FAPY residues in the d(CCTC<u>TTC-GAA</u>CTC) context (boldface G = site of adduct; underscore = *Sfu*I site) were synthesized based on published procedures (40). First, the oligonucleotide was gel-purified, and 100–250 nmol were resuspended in 90 ml of buffer (10 mM NaHPO$_4$, pH 7/100 mM NaCl/50 mM EDTA). One milligram of AFB$_1$ epoxide (41) in 50 ml of CH$_2$Cl$_2$ was then added and the mixture was shaken for 10 min at 4°C. The same amount of AFB$_1$ epoxide was added once more, again shaking at 4°C for 15 min. The solution was then extracted twice with 100 ml of CH$_2$Cl$_2$ and desalted on a G-25 Sephadex column. Oligonucleotides containing AFB$_1$-N7-Gua adducts were separated from unreacted DNA by HPLC with a Beckman Ultrasphere C18 reverse phase column (part no. 235329). Buffer A was 0.1 M TEAAC (pH 7.5) and buffer B was CH$_3$CN. The gradient used went from 8 to 12% B from 0 to 40 min, and from 12 to 25% B from 40 to 60 min. Fractions were collected and desalted with Centricon-3 concentrators at 4°C. A portion of the AFB$_1$-N7-Gua lesions was converted to AFB$_1$-FAPY through incubation in 0.1 M NaHPO$_4$ (pH 10) at 37°C for 4 h. AFB$_1$-FAPY minor was formed first, subsequently reaching equilibrium with AFB$_1$-FAPY major, at a ratio of 2:1 major:minor. Oligonucleotides containing AFB$_1$-FAPY lesions were separated from those containing AFB$_1$-N7-Gua lesions with the HPLC conditions above. These conditions were also used to separate FAPY major from FAPY minor.

Matrix-Assisted Laser Desorption Ionization–Time-of-Flight (MALDI-TOF) MS. Thirteen-base oligonucleotides containing FAPY major, FAPY minor, or 2:1 mixture of the major:minor FAPY adducts (FAPY mix) were analyzed on a PerSeptive Biosystems Voyager Elite DE MALDI-TOF mass spectrometer. Twelve-base and 14-base unmodified oligonucleotides were used as internal calibration standards. Ten picomoles of the sample and each standard were mixed in a 10-μl volume and desalted with ZipTips. They were then resuspended in the MALDI matrix (2.5% anthranilic acid/1.1% nicotinic acid/45.5% CH$_3$CN/27.3% 100 mM ammonium citrate) and spotted onto the sample plate. Spectra were obtained in the reflector mode with a laser energy of 2,400, an accelerating voltage of 25,000 V, a grid voltage of 95%, a guide wire voltage of 0.28%, and a delay time of 120 ns. Each spectrum was an average of 128–256 laser shots.

Genome Construction. Genomes were constructed as in Bailey *et al.* (16). Four separate ligation reactions were done with each of the inserts: (*i*) AFB$_1$-N7-Gua, (*ii*) FAPY mix, (*iii*) FAPY major, and (*iv*) unmodified control. These reactions were carried out at 16°C to ensure the integrity of the adducts.

Transformation of AFB$_1$-Modified Genomes into Cell Strains. Once genomes were constructed and desalted, they were kept on ice and used immediately to transform DL7 (not induced for SOS), UV-irradiated DL7 (expresses UmuDC), DL7/pGW16 (expresses MucAB), or UV-irradiated DL7/pGW16 (expresses both MucAB and UmuDC) by using conditions in Bailey *et al.* (16). The only procedural difference was that the cells were irradiated in a Stratalinker UV crosslinker at 45 J/m^2. Approx-

Smela *et al.*

imately 0.1 pmol of genome in a volume of 10 μl was mixed with 190 μl of the transfection strain for each electroporation. Progeny phage were produced from each transformation and stored at 4°C until needed.

Pooling Progeny Phage. Between 8 and 40 parallel replicate transformations were carried out per sample per cell strain to generate a progeny phage pool that could be easily examined for mutants. A total of 20,800 infective centers from the eight most successful transformations made up each pool. Replicative form DNA was isolated with the Qiagen (Chatsworth, CA) Midi double-stranded plasmid procedure.

Enrichment for Mutant Genomes. Approximately 0.5 pmol of DNA isolated from pooled phage was subjected to cleavage by *Sfu*I, a restriction enzyme that cleaves at a unique site within the 13-base region of the inserted oligonucleotide. Digested products, as well as mock-digested controls, were analyzed on agarose gels to determine linearization. Approximately 0.1 pmol of digested or control DNA was desalted on a G-25 Sephadex column and electroporated as above, using DL7 as the transfection strain and NR9050 as the plating bacteria.

Sequencing. Sequencing was performed on enriched M13 DNA isolated from the cell strain that expressed both UmuDC and MucAB. Two hundred fifty-six blue and 100 clear plaques were sequenced.

Results

MALDI–Time-of-Flight (TOF) MS. To determine whether the peaks observed in the HPLC trace (Fig. 1 *Inset*) represented chemically unique structures, MALDI-TOF MS was performed on 13-base oligonucleotides containing FAPY major, FAPY minor, or a 2:1 mixture of the major:minor forms (FAPY mix). Twelve-base and 14-base oligonucleotides were used as internal standards, and in each case, only one peak was observed for the FAPY-containing oligonucleotide. This peak had m/z ratios of 4191.73, 4191.97, and 4191.99, respectively, for each of the three samples. This result indicates that FAPY major and FAPY minor share identical chemical compositions, implying that they are most likely rotamers of one another, as is the case for other formamidopyrimidines (33).

Construction of Genomes Containing AFB₁ Adducts. Single-stranded oligonucleotides containing AFB$_1$-FAPY adducts were analyzed by HPLC after being exposed to the conditions of genome construction. Stability studies were performed on oligonucleotides containing the FAPY major adduct, the FAPY minor adduct, or FAPY mix. Although the FAPY major adduct maintained its integrity through all steps of the procedure, there was, however, some conversion of FAPY minor to the major form. Accordingly, genetic studies were performed with only oligonucleotides containing either FAPY major or FAPY mix. Bailey *et al.* (40) demonstrated the stability of AFB$_1$-N7-Gua oligonucleotides under these conditions.

After the ligation of the oligonucleotide inserts (either modified or unmodified, as a control), it was noted that the yield of intact modified genomes was low compared with that of unmodified controls, indicating that steric or other factors may inhibit efficient annealing and ligation of the AFB$_1$-FAPY- or AFB$_1$-N7-Gua-containing inserts. The amount of DNA used in each transformation was adjusted based on the percentage of circular ligation product observed with agarose gel analysis.

Cellular Tolerance of AFB₁ DNA Adducts. Genomes containing an unmodified insert or an insert containing the FAPY major, FAPY mix, or AFB$_1$-N7-Gua adduct were electroporated into *E. coli* (Fig. 2*A*). Several cell strains were used in these experiments
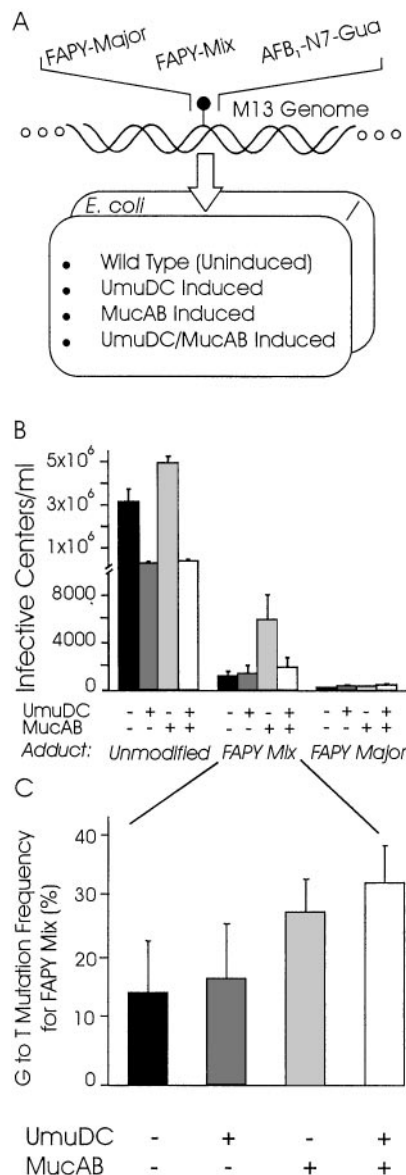

Fig. 2. Lethality and mutagenicity of AFB$_1$-FAPY. (*A*) M13 genomes contained a single aflatoxin adduct and were replicated in *E. coli* expressing different bypass polymerases. (*B*) Cell strains challenged with FAPY mix, FAPY major, or an unmodified control exhibited different levels of survival. The *x* axis depicts the adduct with which the cell was challenged and the polymerase status of the cell, whereas the *y* axis is the number of infective centers per milliliter. (*C*) The G to T mutation frequency for cell strains challenged with FAPY mix is represented here. Cell strains expressed the bypass polymerases indicated by the *x* axis. The G to T mutation frequency was calculated by counting the number of light blue plaques and dividing by the total number of blue plaques observed. Error bars for both graphs reflect the SEM for four sets of data.

to probe the genetic requirements for toxicity. Cells not induced for the SOS response (DL7) or cells that expressed the bypass polymerases UmuDC (UV-irradiated DL7), MucAB (DL7/pGW16), or both UmuDC and MucAB (UV irradiated DL7/pGW16) were used in the electroporations. G to T mutations are scored as light blue plaques, whereas wild-type sequences, other point mutations, and in-frame deletions (provided they are small enough) yield dark blue plaques (16). Clear plaques would result from frame shifts, large deletions, or genomes without inserts.

Each adduct-containing genome was electroporated into each

cell strain 8–40 separate times to yield one set of data. Every graph shown represents four sets of data, each set generated from a separate genome construction event. After transformation, the total number of blue plaques was scored as the number of infective centers. Each infective center represents a single transformation event by one M13 genome that was successfully replicated in the host cell, yielding a wild-type or mutant outcome. Thus, we can determine how well or poorly the replication machinery of the cell can handle the challenge of a single AFB$_1$-FAPY adduct. Fig. 2B compares the numbers of infective centers per milliliter of transformed cells for genomes containing an unmodified, FAPY major, or FAPY mix oligonucleotide insert in each of the genetic backgrounds described above. We found that unmodified inserts were tolerated an average of 2–3 orders of magnitude better than those containing a modified insert. When cells were challenged with the FAPY major adduct alone, the genome survived less well than a comparable genome containing FAPY mix. Cells tolerated the AFB$_1$-N7-Gua adduct 1–2 orders of magnitude better than the FAPY mix adduct, as inferred from the number of infective centers obtained (data not shown). These data indicate that the AFB$_1$-N7-Gua adduct is far less effective as a replication block than the AFB$_1$-FAPY adduct. In our hands the FAPY major adduct is the most lethal known aflatoxin-induced replication block *in vivo*. FAPY mix is not as lethal a block to replication, therefore it is possible that cells that survive the challenge with aflatoxin contain mostly the FAPY minor component of this mixture and, therefore, most of the mutations that arise may be a result of this form of the adduct. It is for this reason that we chose to evaluate the FAPY mix adduct in our mutagenicity studies.

Mutagenicity of the AFB$_1$-FAPY Adduct. The experimental system allowed the facile scoring of G to T mutations by using the convenient light blue/dark blue color selection described above. Fig. 2C shows the G to T mutation frequency of the FAPY mix adduct in each of the four genetic backgrounds used in this study. Cells not induced for bypass polymerases showed the lowest number of mutations, about 15%. Those expressing UmuDC showed a 17% mutation frequency, whereas the MucAB-expressing cells showed a 27% mutation frequency. Finally, cells expressing both UmuDC and MucAB showed a 32% mutation frequency, which is less than additive, implying that both polymerases may compete to bypass the same adduct. The same trend observed for survival is not observed for mutation frequency and, in fact, the only significant difference discovered here is that cells expressing MucAB, whether or not UmuDC is present, have a slightly higher G to T mutation frequency than the cell strains that do not express the MucAB bypass polymerase. To simplify our analysis, we therefore decided to evaluate the frequency of other types of mutations only in the cell strain expressing both UmuDC and MucAB.

The G to T mutation frequency observed by Bailey *et al.* (16) for the AFB$_1$-N7-Gua adduct was 4% in the cell strain expressing MucAB. After repeating this experiment for all four cell strains used in these studies, the mutation frequency was found to be between 2–6%, which agrees well with prior results and allows for a direct comparison of the two studies. The modest mutagenicity of AFB$_1$-N7-Gua is striking by contrast with AFB$_1$-FAPY. Indeed, FAPY adducts are at least 6 times as likely to cause G to T mutations as their molecular predecessor, AFB$_1$-N7-Gua.

Enrichment for Non-G to T Mutations. The non-G to T mutations were expected to be infrequent; therefore, we used a method that would help to enrich the phage pool for mutants by eliminating the majority of wild-type DNA. Progeny phage obtained from the eight transformation events that gave the highest number of

Table 1. Multiple mutations induced by AFB$_1$-FAPY mix

Mutation type	Sequence (5′–3′)	Contribution*, %
Wild type	CCTCTTC<u>G</u>AACTC	
Quadruple	CCTCT**AAA**<u>G</u>ACTC	16
Tandem	**CTGC**TT**G**<u>A</u>AACTC	10
	CCTCTT**G**<u>A</u>AACTC	3
5′ or 3′	CCTCTT**T**<u>G</u>AACTC	10
	CCTCTT**G**<u>G</u>AACTC	3
	CTCC**ATT**<u>G</u>AA**G**TC	8
Deletions	**CGA**CTTC(Δ<u>G</u>AA)CTC	2
	CGTCTTC(Δ<u>G</u>AA)CTC	8
Other	**CA**TCTTC<u>T</u>AACTC	8

Underlined base is target site of modification. Boldface indicates mutated bases. Parentheses (Δ) indicate deletions.

*Contribution of the type of mutation to the total number of non-G to T mutations observed.

transformants in each cell strain were pooled, the population was expanded in cell culture, and the double-stranded (replicative form) M13 DNA was isolated. The oligonucleotide inserts were engineered such that they contained a unique *Sfu*I restriction site that encompassed the target base for evaluation of mutagenesis. If this and the surrounding positions retained their wild-type sequence, the restriction enzyme would successfully cut this DNA, rendering it unable to transform *E. coli*. However, if a mutation occurred within the restriction site, the DNA would be refractory to digestion and therefore able to transform *E. coli* successfully.

DNA from progeny phage that were a result of AFB$_1$-FAPY-modified genomes transfected into cells expressing both UmuDC and MucAB was subjected to treatment with *Sfu*I or was mock-treated. The number of dark blue plaques in the digested sample divided by the number of dark blue plaques in the mock-treated sample is termed the restriction-resistant fraction (RRF) and is a first approximation of the mutation frequency. Sequencing the dark blue plaques in the digested samples and multiplying their values by the RRF yields the true frequency of each type of non-G to T mutation present. It should be noted that all of the dark blue plaques did not consist of wild-type or mutant sequences; some were the result of the genetic engineering procedures used in the experiment. For example, some sequences were M13 DNA that had been ligated back together without an insert, containing only a few extra random bases to put the gene back in frame. Others were small deletions that presumably result from errant nuclease action. These species are included in the total used to calculate the fraction of each type of mutation observed. The following *overall* mutation frequencies were observed: G to A = 1%, total 5′ mutations (C to G and C to T) = 1%, a specific quadruple mutation = 1%, and other multiple mutations = 3%. For comparison, under these genetic conditions the G to T mutation frequency was 32%.

When looking only at the dark blue plaques that were mutants, it can be determined how each specific type of mutation contributed to the total. The following percentages were observed: G to A mutations = 18%, total 5′ mutations (C to T and C to G) = 16%, a specific quadruple mutation = 16%, and other multiple mutations = 49%. Table 1 shows the different multiple mutations and calculates the contribution of each mutation to the total. One surprising result was the presence of a specific quadruple mutation. This sequence was 5′-CCTCT**AAA**-<u>G</u>ACTC-3′. The bold bases are mutations, and the underlined base is the originally modified site. The mutation at the target site was a G to A, and these were not included in the total number of G to A mutations. The two 5′ transversions are T to A and C to A, respectively, and the 3′ transition is A to G, none of which

were included in the total number of 5′ or 3′ mutations. We believe that the multiple mutations are completely unique to the AFB$_1$-FAPY adduct, specifically the FAPY minor form, because of the fact that FAPY major is a lethal replication block in our system. We also hypothesize that the FAPY minor adduct, before converting to the low energy FAPY major structure, may cause some conformational flexibility of the DNA helix that may interfere with the replication of several bases that are adjacent to the modified site.

Several small deletions consisting of the codon containing the modified G were also observed. They made up about 10% of the non-G to T mutations and may be a result of the complete bypass of the damaged area by the polymerase. One other occurrence was the point mutation of one or two bases located about five to six positions upstream of the modified G. Most often these mutations were C to T or C to G. It is possible that these mutations are the results of localized infidelity of the error-prone polymerase distal to the damaged site. Such mutations would taper off once the normal replicative DNA polymerase resumed its function.

Discussion

Epidemiological studies indicate that aflatoxin B$_1$ plays a synergistic role with hepatitis B virus in the formation of many human hepatocellular carcinomas (1). It is believed that the carcinogenicity of the toxin is mediated in part by the mutagenic properties of one or more of the AFB$_1$ DNA adducts. Although exposure to HBV alone increases the risk of HCC, the risk is much higher in an individual who is positive for HBV and is at risk for high levels of aflatoxin exposure. Studies indicate that people who live in areas where potential AFB$_1$ exposure is high are 3 times more likely to develop HCC. Those who test positive for HBV are about 7 times more likely to develop HCC, and when both these criteria are met, people are 60 times more likely to develop this disease.

The work presented here evaluated the contributions to lethality and mutagenicity of the aflatoxin formamidopyrimidine adduct, a DNA lesion that is highly persistent in liver *in vivo* (25). The G to T mutation at the third position of codon 249 of the p53 tumor-suppressor gene has been observed in over 50% of HCC cases as well as in human hepatocytes treated with AFB$_1$ in culture (2, 3). Additionally, activation of human *ras* genes by specific G to T mutations presumed to be caused by aflatoxin exposure may play a significant role in malignant transformation (4–6). This mutation was observed for the AFB$_1$-N7-Gua adduct in the mutation analysis system used in this study. AFB$_1$-N7-Gua, however, was only weakly mutagenic, with a mutation frequency of 2–6%. By contrast, we have established that AFB$_1$-FAPY produced the same mutation with a frequency of up to 32%; because this adduct is also highly persistent in the genome, it has great potential to play a role in hepatocarcinogenesis.

It would be of interest to know to what extent the structures of the different types of aflatoxin DNA adducts influence their biological properties. The NMR structures of AFB$_1$-DNA adducts provide insight into this issue (27). The AFB$_1$-FAPY adduct stabilizes the DNA duplex by unwinding it by about 15° and increasing stacking interactions between the intercalated aflatoxin moiety and the neighboring bases. In addition there is a possible hydrogen bond between the formyl proton of what is thought to be FAPY major and the exocyclic amino group on a neighboring adenine. The NMR solution structure of AFB$_1$-N7-Gua also reveals an increased melting temperature of the DNA duplex, but only by about 3–5°C (42). Additionally, this form of the adduct bends slightly the DNA helix (28). NMR data to establish the FAPY minor structure remain elusive.

In principle, repair enzymes may recognize AFB$_1$-N7-Gua better than the FAPY analog owing to the greater distortive effect of AFB$_1$-N7-Gua on DNA structure. In support of this

view, excision repair systems preferentially remove AFB$_1$-N7-Gua from the DNA of human fibroblasts (43). The FAPY adduct, having a more subtle effect on DNA architecture and a more profound effect on the local melting temperature of the DNA duplex than AFB$_1$-N7-Gua (29, 41, 42), may evade repair, which is in line with its persistence *in vivo*. The longevity of the FAPY adduct, combined with its high mutagenic potential, makes it a dangerous lesion.

Unfortunately, the sequences used in the two NMR studies of AFB$_1$-N7-Gua and AFB$_1$-FAPY (27, 29) were not identical, which raises the question of whether sequence context may have an effect on the secondary structures these adducts form within DNA. Sequence context may in addition affect whether the adducts are blocks to replication, as has been observed for both the AFB$_1$-N7-Gua and AFB$_1$-FAPY adducts in many *in vitro* systems (12, 44–48) and/or what frequencies and types of mutations they cause. It is possible that different sequence contexts allow for faster or slower conversion of FAPY minor to FAPY major, if the stability of the adduct depends on hydrogen bonding with neighboring bases.

There are clear differences in the survival of phage DNA containing an AFB$_1$-FAPY adduct when that DNA is replicated in cells expressing different bypass polymerases. UmuDC and MucAB are analogous bypass polymerases (polV) found in *E. coli* and *S. typhimurium*, respectively (38, 49–51). Phage DNA harboring the AFB$_1$-FAPY adduct survives best when replicated in cells expressing MucAB, a polymerase that is currently thought to be the most proficient at bypass of bulky lesions (Fig. 2*B*) (17, 50, 52–58). This study has established that the FAPY major adduct is the strongest block to replication of all of the aflatoxin adducts studied here, irrespective of which bypass polymerase is used by the cell to help it overcome the DNA damage. Thus, this lesion is a prime candidate as a contributor to the extreme toxicity of AFB$_1$. This conclusion may be extrapolated to humans, because mammalian cells have similar lesion bypass mechanisms (59, 60). In contrast, phage DNA containing the AFB$_1$-N7-Gua adduct generates about 2 orders of magnitude more infective centers than that containing the AFB$_1$-FAPY adducts, indicating that the AFB$_1$-N7-Gua adduct is a less formidable threat to the replication machinery of the cell.

AFB$_1$-FAPY induces a G to T mutation frequency that is roughly 6 times that observed for AFB$_1$-N7-Gua. The G to A mutation frequency is similar for the two adducts, the G to C mutation frequency is higher for the AFB$_1$-N7-Gua adduct, and there are also a similar number of 5′ mutations, supporting the notion that it is the bulkiness of the AFB$_1$ adducts that causes adjacent mutations. Additionally, the FAPY minor adduct may induce a zone of conformational flexibility in the vicinity of the adduct. This zone possibly interferes with faithful replication beyond the adduct, causing multiple mistakes to be made by the polymerase, and resulting in several mutations in the immediate area. One particularly interesting case is that of a quadruple mutant that makes up 16% of the non-G to T mutations observed (1% of all mutations caused AFB$_1$-FAPY), demonstrating that multiple mutations are not necessarily random. The lesion itself may determine the type of mutations observed at sites other than the target site, which is consistent with the fact that these mutations are not observed for the AFB$_1$-N7-Gua adduct. One aspect to note regarding the quadruple mutants is that all four of the bases inserted opposite the lesion are pyrimidines, possibly because they are less sterically hindered than their bulkier purine counterparts, and therefore more easily placed opposite a lesion.

In most studies of the mutations observed in DNA globally modified with AFB$_1$, there are several instances in which mutations occur 5′ or 3′ to a G, and sometimes even several bases away. When these mutations occur at a C, they have been commonly attributed to a modified G in the opposite strand. This may or may not be the case, and when the mutations occur at

bases other than C, these mutations may be the results of an adduct a few bases away. We have closely evaluated several reports in the literature (61, 62) and have noted that the sequence context used in our studies appears several times. In these sequence contexts there exist mutations that mirror the multiple mutations that we have observed. These mutations could be a result of the adduct itself or a function of a bypass/repair mechanism of the cell. The notion that one DNA adduct can cause more than one mutation, even a few bases away from the adduct site, is supported by the work of others (63), who observe this phenomenon not only for AFB$_1$, but also for benzo(a)pyrene and UV light. All of the aforementioned studies support the observation that mutational hotspots do not always correspond with hotspots for adduct formation (1).

This work, in summary, leads to three main conclusions. The first is that the highly persistent AFB$_1$-FAPY adduct is also highly mutagenic, which poses a long-term threat to the genetic integrity of cells harboring the adduct. Second, the FAPY major form of the adduct is a very strong block to replication, even when cells use bypass polymerases. Similar mammalian replicative and DNA repair systems may have difficulty dealing with this adduct *in vivo*, leading to toxic effects. Finally, the AFB$_1$-FAPY adduct is responsible for mutations that occur at sites other than the site of initial damage.

1. Smela, M. E., Currier, S. S., Bailey, E. A. & Essigmann, J. M. (2001) *Carcinogenesis* **22,** 535–545.
2. Hsu, I. C., Metcalf, R. A., Sun, T., Welsh, J. A., Wang, N. J. & Harris, C. C. (1991) *Nature (London)* **350,** 427–428.
3. Bressac, B., Kew, M., Wands, J. & Ozturk, M. (1991) *Nature (London)* **350,** 429–431.
4. Bos, J. L., Fearon, E. R., Hamilton, S. R., Verlaan-de Vries, M., van Boom, J. H., van der Eb, A. J. & Vogelstein, B. (1987) *Nature (London)* **327,** 293–297.
5. Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. & Bos, J. L. (1988) *N. Engl. J. Med.* **319,** 525–532.
6. McMahon, G., Hanson, L., Lee, J. J. & Wogan, G. N. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9418–9422.
7. Riley, J., Mandel, H. G., Sinha, S., Judah, D. J. & Neal, G. E. (1997) *Carcinogenesis* **18,** 905–910.
8. Benasutti, M., Ejadi, S., Whitlow, M. D. & Loechler, E. L. (1988) *Biochemistry* **27,** 472–481.
9. Denissenko, M. F., Cahill, J., Koudriakova, T. B., Gerber, N. & Pfeifer, G. P. (1999) *Mutat. Res.* **425,** 205–211.
10. Basu, A. K. & Essigmann, J. M. (1990) *Mutat. Res.* **233,** 189–201.
11. Kamiya, H., Sakaguchi, T., Murata, N., Fujimuro, M., Miura, H., Ishikawa, H., Shimizu, M., Inoue, H., Nishimura, S., Matsukage, A., *et al.* (1992) *Chem. Pharm. Bull.* **40,** 2792–2795.
12. Levy, D. D., Groopman, J. D., Lim, S. E., Seidman, M. M. & Kraemer, K. H. (1992) *Cancer Res.* **52,** 5668–5673.
13. Puisieux, A., Lim, S., Groopman, J. & Ozturk, M. (1991) *Cancer Res.* **51,** 6185–6189.
14. Foster, P. L., Eisenstadt, E. & Miller, J. H. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2695–2698.
15. Lawrence, C. W., Borden, A., Banerjee, S. K. & LeClerc, J. E. (1990) *Nucleic Acids Res.* **18,** 2153–2157.
16. Bailey, E. A., Iyer, R. S., Stone, M. P., Harris, T. M. & Essigmann, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 1535–1539.
17. Foster, P. L., Groopman, J. D. & Eisenstadt, E. (1988) *J. Bacteriol.* **170,** 3415–3420.
18. Muench, K. F., Misra, R. P. & Humayun, M. Z. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6–10.
19. Misra, R. P., Muench, K. F. & Humayun, M. Z. (1983) *Biochemistry* **22,** 3351–3359.
20. Busby, W. F., Jr., & Wogan, G. N. (1984) in *Chemical Carcinogens*, ed. Searle, C. (Am. Chem. Soc., Washington, DC), pp. 945–1136.
21. Martin, C. N. & Garner, R. C. (1977) *Nature (London)* **267,** 863–865.
22. Lin, J. K., Miller, J. A. & Miller, E. C. (1977) *Cancer Res.* **37,** 4430–4438.
23. Groopman, J. D., Croy, R. G. & Wogan, G. N. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5445–5449.
24. Croy, R. G., Essigmann, J. M., Reinhold, V. N. & Wogan, G. N. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 1745–1749.
25. Croy, R. G. & Wogan, G. N. (1981) *J. Natl. Cancer Inst.* **66,** 761–768.
26. Croy, R. G. & Wogan, G. N. (1981) *Cancer Res.* **41,** 197–203.
27. Johnston, D. S. & Stone, M. P. (1995) *Biochemistry* **34,** 14037–14050.
28. Gopalakrishnan, S., Harris, T. M. & Stone, M. P. (1990) *Biochemistry* **29,** 10438–10448.
29. Mao, H., Deng, Z., Wang, F., Harris, T. M. & Stone, M. P. (1998) *Biochemistry* **37,** 4374–4387.
30. Boiteux, S., Belleney, J., Roques, B. P. & Laval, J. (1984) *Nucleic Acids Res.* **12,** 5429–5439.
31. Chetsanga, C. J. & Makaroff, C. (1982) *Chem. Biol. Interact.* **41,** 235–249.
32. Chetsanga, C. J., Bearie, B. & Makaroff, C. (1982) *Chem. Biol. Interact.* **41,** 217–233.
33. Tomasz, M., Lipman, R., Lee, M. S., Verdine, G. L. & Nakanishi, K. (1987) *Biochemistry* **26,** 2010–2027.
34. Witkin, E. M. (1976) *Bacteriol. Rev.* **40,** 869–907.
35. Friedberg, E. C., Walker, G. C. & Siede, W. (2001) in *DNA Repair and Mutagenesis* (Am. Soc. Microbiol. Press, Washington, DC), pp. 407–464.
36. Woodgate, R., Levine, A. S., Koch, W. H., Cebula, T. A. & Eisenstadt, E. (1991) *Mol. Gen. Genet.* **229,** 81–85.
37. Sommer, S., Bailone, A. & Devoret, R. (1993) *Mol. Microbiol.* **10,** 963–971.
38. Tang, M., Shen, X., Frank, E. G., O'Donnell, M., Woodgate, R. & Goodman, M. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 8919–8924.
39. Banerjee, S. K., Borden, A., Christensen, R. B., LeClerc, J. E. & Lawrence, C. W. (1990) *J. Bacteriol.* **172,** 2105–2112.
40. Bailey, E. A., Iyer, R. S., Harris, T. M. & Essigmann, J. M. (1996) *Nucleic Acids Res.* **24,** 2821–2828.
41. Gopalakrishnan, S., Byrd, S., Stone, M. P. & Harris, T. M. (1989) *Biochemistry* **28,** 726–734.
42. Jones, W. R., Johnston, D. S. & Stone, M. P. (1998) *Chem. Res. Toxicol.* **11,** 873–881.
43. Leadon, S. A., Tyrrell, R. M. & Cerutti, P. A. (1981) *Cancer Res.* **41,** 5125–5129.
44. Yu, F. L. (1983) *Carcinogenesis* **4,** 889–893.
45. Marien, K., Moyer, R., Loveland, P., Van Holde, K. & Bailey, G. (1987) *J. Biol. Chem.* **262,** 7455–7462.
46. Yu, F. L., Bender, W. & Geronimo, I. H. (1990) *Carcinogenesis* **11,** 475–478.
47. Refolo, L. M., Conley, M. P., Sambamurti, K., Jacobsen, J. S. & Humayun, M. Z. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3096–3100.
48. Johnston, D. S. & Stone, M. P. (2000) *Chem. Res. Toxicol.* **13,** 1158–1164.
49. Sutton, M. D., Smith, B. T., Godoy, V. G. & Walker, G. C. (2000) *Annu. Rev. Genet.* **34,** 479–497.
50. Perry, K. L., Elledge, S. J., Mitchell, B. B., Marsh, L. & Walker, G. C. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4331–4335.
51. Goldsmith, M., Sarov-Blat, L. & Livneh, Z. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 11227–11231.
52. Kulaeva, O. I., Wootton, J. C., Levine, A. S. & Woodgate, R. (1995) *J. Bacteriol.* **177,** 2737–2743.
53. Hauser, J., Levine, A. S., Ennis, D. G., Chumakov, K. M. & Woodgate, R. (1992) *J. Bacteriol.* **174,** 6844–6851.
54. Woodgate, R. (1992) *Mutat. Res.* **281,** 221–225.
55. O'Grady, P. I., Borden, A., Vandewiele, D., Ozgenc, A., Woodgate, R. & Lawrence, C. W. (2000) *J. Bacteriol.* **182,** 2285–2291.
56. Watanabe, M., Nohmi, T. & Ohta, T. (1994) *Mutat. Res.* **314,** 27–37.
57. Bennett, C. B., Luo, X., Refolo, L. M. & Humayun, M. Z. (1988) *Mutat. Res.* **202,** 223–234.
58. Szekeres, E. S., Jr., Woodgate, R. & Lawrence, C. W. (1996) *J. Bacteriol.* **178,** 2559–2563.
59. Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K. & Hanaoka, F. (1999) *Nature (London)* **399,** 700–704.
60. Gerlach, V. L., Aravind, L., Gotway, G., Schultz, R. A., Koonin, E. V. & Friedberg, E. C. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11922–11927.
61. Aguilar, F., Hussain, S. P. & Cerutti, P. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8586–8590.
62. Sambamurti, K., Callahan, J., Luo, X., Perkins, C. P., Jacobsen, J. S. & Humayun, M. Z. (1988) *Genetics* **120,** 863–873.
63. Courtemanche, C. & Anderson, A. (1999) *Mutat. Res.* **430,** 23–36.

Smela *et al.*