Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2240–2244, March 1994
Plant Biology

Cloning of a 3-methyladenine-DNA glycosylase from *Arabidopsis thaliana*

(DNA repair)

ANNE SANTERRE AND ANNE B. BRITT*

Section of Plant Biology, University of California, Davis, CA 95616

Communicated by André T. Jagendorf, November 5, 1993 (received for review September 13, 1993)

ABSTRACT **We have isolated an *Arabidopsis thaliana* cDNA that complements the methyl methanesulfonate-sensitive phenotype of an *Escherichia coli* double mutant deficient in 3-methyladenine glycosylases (DNA-3-methyladenine glycosidases I and II, EC 3.2.2.20 and 3.2.2.21, respectively, encoded by *tag* and *alkA*). Expression of the *Arabidopsis* cDNA enhances the methyl methanesulfonate resistance of the *E. coli* double mutant by nearly four orders of magnitude. The cDNA corresponds to a single-copy, nuclearly encoded sequence which specifies a predicted 28.1-kDa protein with a charge of +8 at pH 7.0. Enzymatic analysis of extracts prepared from the transformed mutants indicates that the cDNA encodes a 3-methyladenine glycosylase. The predicted amino acid sequence of the *Arabidopsis* glycosylase has significant homology to other eukaryotic 3-methyladenine glycosylases.**

All living things are subject to DNA damage in a wide variety of forms. Environmental and man-made toxins such as UV light, chemical mutagens, fungal and bacterial toxins, and ionizing radiation result in the formation of cross-links, single- and double-stranded breaks, and chemical adducts in DNA. In addition, because the genome is located in a chemically reactive environment, some amount of DNA damage occurs as an inevitable result of the interaction of nucleic acids with cellular alkylating and oxidizing agents. These modified bases can, in the absence of repair, contribute substantially to the cell's spontaneous mutation rate (1–4). DNA repair mechanisms, which protect the cell from the cytotoxic and mutagenic effects of these lesions, have been demonstrated in a variety of organisms, from bacteriophage to humans. The various pathways for repair include both nucleotide excision repair mechanisms, which repair a broad spectrum of lesions, and more specialized base repair enzymes (largely, though not exclusively, glycosylases).

The study of DNA repair, recombination, and replication in plants has lagged far behind the study of these processes in microbial and mammalian systems. Molecular isolation of the genes involved in DNA enzymology in higher plants has been limited to a handful of clones. These include two genes cloned by taking advantage of homologous nucleotide sequences: the *Arabidopsis* gene for topoisomerase I (5) and an apparent homolog of the *Escherichia coli recA* gene, localized to the chloroplasts, from *Arabidopsis* and *Pisum sativum* (6). Hays and coworkers (41, 42) have isolated several *Arabidopsis* clones on the basis of their ability to partially complement the UV-sensitive phenotype of an *E. coli recA*$^-$, *phr*$^-$, *uvrC*$^-$ triple mutant. With the exception of the *TOP1* clone, the biochemical activities encoded by these cDNAs have not yet been identified.

Although alkylating agents are among the most efficient and commonly used mutagens of higher plants, very little is known about the repair of methylated bases in higher plants. Methylation of the four bases can occur as a result of exposure to man-made alkylating agents, such as methyl methanesulfonate (MMS) or methylnitrosourea, or as a result of exposure to naturally occurring methylating agents (7). Nearly every oxygen and nitrogen atom in the DNA molecule is a potential target for methylation, though the frequency of induction of each type of lesion varies and is dependent on the chemical nature of the methylating agent (8).

The cytotoxic and mutagenic effects of these alterations in DNA structure are determined both by the chemical nature of the lesion and by the repertoire of DNA repair activities available to the affected organism. Experiments in microbial and mammalian systems indicate that some base modifications, such as the production of 7-methylguanine (7MeGua) by methylating agents, are apparently harmless, as they neither inhibit nor alter normal base pairing (9). Other lesions, such as O^6-methylguanine, are directly mutagenic, as the altered guanine residue is capable of mispairing with thymine (10). A third class of lesions, including 3-methyladenine (3MeAde), act as blocks to replication and transcription (9) and so are cytotoxic. The toxicity of this lesion is substantially reduced by the presence of a repair enzyme, 3MeAde glycosylase, which specifically recognizes this lesion and cleaves the base at its *N*-glycosylic bond, producing an apurinic site which is subsequently repaired by the combined actions of apurinic/pyrimidinic endonucleases (AP endonucleases), exonucleases, repair polymerases, and DNA ligase. Some organisms, notably *E. coli* and *Saccharomyces cerevisiae*, have evolved DNA damage-inducible, error-prone mechanisms for the replicational bypass of lesions which normally act as blocks to replication (11, 12). The *umuCD*-dependent bypass of 3MeAde has been shown to be mutagenic in *E. coli* (13).

3MeAde glycosylase (*MAG*) genes have been cloned and sequenced from a variety of microbial and mammalian sources, including *E. coli* (14), *S. cerevisiae* (15, 16), humans (17–19), mice (20), and rats (21). Most of the eukaryotic *MAG* cDNAs were cloned by the method of Chen *et al.* (15), which involves selecting for a cDNA which complements the MMS-sensitive phenotype of an *E. coli* 3MeAde glycosylase-defective strain (a *tag, alkA* double mutant).

We are interested in isolating a plant DNA repair gene to investigate its possible environmental and developmental regulation. Although there has been no evidence in the literature for or against the presence of 3MeAde glycosylase activity in any plant, we planned to determine whether an *Arabidopsis* cDNA could, like other eukaryotic cDNAs, restore MMS resistance to the *E. coli* double mutant. We have isolated a cDNA which, when expressed in the double mutant, increases the mutant's MMS resistance by as much

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Ap, ampicillin; Ap^r, Ap-resistant; 3MeAde, 3-methyladenine; 7MeGua, 7-methylguanine; MMS, methyl methanesulfonate.

*To whom reprint requests should be addressed.

as four orders of magnitude. We present genetic, biochemical, and molecular evidence that this *Arabidopsis* cDNA encodes a 3MeAde glycosylase.†

MATERIALS AND METHODS

Strains. *E. coli* AB1157 (F⁻, *thr-1, ara-14, leuB6, (gpt-proA)62, lacY1, tsx-33, supE44, galK2, hisG4, rfbD1, mgl-51, rpsL31, kdgK51, xyl-5, mtl-1, argE3, thi-1*) (22) was used as the wild type for alkylation sensitivity. *E. coli* MV1932 (F⁻, *alkA1, tag*) was a gift from Michael Volkert (University of Massachusetts). Although AB1157 is generally employed as a wild type for comparison of MMS sensitivity in MV1932 and its derivatives, it should be noted that the stocks are not isogenic and that a substantial (perhaps major) fraction of the MV1932 genome is derived from Hfr KL16 (23, 24). BNN132, used to convert the λYES phage library to a plasmid library, is a λ KC (*kan, cre*) lysogen of JM107 (25) and was a gift from R. W. Davis (Stanford University).

Plasmid Expression Library. The *Arabidopsis* cDNA expression library, derived from wild-type *Arabidopsis thaliana* (ecotype Columbia), was a gift of R. W. Davis. The λYES library (25) carries cDNAs cloned into an *Xho* I site flanked by two *Eco*RI sites on a 7.8-kb plasmid which is, in turn, embedded between two *lox* recombination sites in bacteriophage λ. The cDNA library was converted to a plasmid library by using the λ KC lysogen BNN132. The infected lysogen was plated onto LB plates (26) containing ampicillin (Ap; 100 μg/ml), and plasmid DNA was purified directly from the pooled Ap-resistant (Apʳ) colonies.

Screening Clones for MMS Resistance. All transformations were performed under nonselective (no MMS) conditions; Apʳ colonies were then rescreened for MMS resistance by streaking or replating on selective agar. Selective medium was prepared by adding MMS (1 mM) to LB Ap agar cooled to 50°C. Plates were used the same day as they were poured.

Measurement of Survival: Alkylation-Induced Cell Killing. Pelleted cells from 2 ml of an overnight culture in LB medium resuspended in 5 ml of M9 minimal medium (26), distributed into micro centrifuge tubes (0.5 ml per tube), and challenged with increasing doses of the mutagen MMS (immediately diluted from a newly opened vial) for 20 min at 37°C with agitation. After exposure to MMS, cells were promptly diluted in LB broth, plated on LB Ap plates, and incubated at 37°C for 20 hr to measure cell survival.

Southern Blot Hybridization. DNA was isolated from the aerial tissues of 3-week-old *Arabidopsis thaliana* (ecotype Columbia) essentially as described (27). DNA was purified on a CsCl gradient and dialyzed against 10 mM Tris·HCl/1 mM EDTA, pH 8. Restriction digestion and Southern blotting followed standard procedures (26). ^{32}P-labeled probes were prepared by using a random primer DNA labeling kit (GIBCO/BRL) according to the manufacturer's instructions. The final washes were at high stringency [0.1× SSPE (26)/0.1% SDS, 55°C].

DNA Sequence Analysis. The nucleotide sequence was determined by using the dideoxynucleotide-mediated chain-termination method (28) according to the instructions of the Sequenase Version 2.0 kit (United States Biochemical). The following oligonucleotide primers were used: 5′-AGCG-GATAACAATTTCACACAGGA-3′ (648) and 5′-CGC-CAGGGTTTTCCCAGTCACGAC-3′ (649). These primers are "universal" primers which are complementary to the *lacZ* fragment of pUC18 on each side and directed toward the polylinker. Three other primers (650, 651, and 652) were synthesized (PCR-Mate 391 DNA Synthesizer from Applied Biosystems) corresponding to regions within the cDNA sequence. These are underlined in Fig. 3*B*.

Computer Analysis of Peptide Sequence Homologies. The MAXSEGS locally optimal sequence alignment program was run on the Cray C90 at the Pittsburg Supercomputing Center (PSC). This program, developed at the PSC, is based on the dynamic programming algorithm developed by Waterman and Eggert (29). Alignments were determined using PAM 120 and 250 substitution matrixes (30) with default gap and new gap penalties. The entire Swiss-Prot data base, containing approximately 30,000 entries, was searched on April 9, 1993.

Preparation of Radiolabeled Substrate DNA. Salmon sperm DNA (20 mg) (Na salt, Sigma) was dissolved in 10 ml of 10 mM Tris·HCl, pH 7.5, sheared, precipitated with ethanol, and dissolved overnight in 6 ml of cacodylate buffer (250 mM sodium cacodylate/1 mM EDTA, pH 7.4) as described (31). A 1-ml portion of [^{3}H]dimethyl sulfate in hexane (DuPont/NEN), specific activity 3.6 Ci/mmol (1 Ci = 37 GBq), was added to the DNA and the mixture was incubated in the dark for 100 min at 37°C with agitation. The reaction was stopped by the addition of 2 vol of cold ethanol, and the precipitated DNA was washed essentially as described by Riazuddin and Lindahl (31) and redissolved in 5 ml of 10 mM Tris·HCl/1 mM EDTA, pH 8. The labeled DNA was then dialyzed overnight against several changes of 20 mM Tris·HCl, pH 8/1 mM EDTA at 4°C (32) and stored at −80°C in 100-μl aliquots (1 μg/μl). The specific activity of the [^{3}H]methylated salmon sperm DNA substrate was 2800 cpm/μg.

Preparation of Cell Extracts. Crude extracts from AB1157 pUC18, MV1932 pUC18, and MV1932 pMMS^R were prepared as follows: A 5-ml overnight culture in LB Ap was diluted 1:100 into fresh medium and grown at 37°C to an OD_{600} of 0.4. At this point isopropyl β-D-thiogalactoside was added to a final concentration of 0.5 mM and incubation was resumed for an additional 2.5 hr to fully induce expression of the cDNA. Cells were then collected by centrifugation, washed in 30 ml of chilled buffer A (50 mM Tris·HCl, pH 7.5/1 mM EDTA/0.1 M KCl), and pelleted again, and the total volume was brought up to 3–4 ml in buffer A. The cells, kept on ice, were then lysed by a single passage at 600 psi (4.1 MPa) in a French press. Insoluble debris and unlysed cells were removed by centrifugation at 10,000 × *g*, and the soluble protein concentration was estimated by the Bradford method (33). Enzymatic assays were performed immediately, as the *Arabidopsis* glycosylase was found to be unstable in crude extracts stored at 4°C, losing more than 50% of its activity over a 48-hr period.

Assay for Glycosylase Activity. DNA glycosylase activity in crude extracts from the three strains tested was quantitated according to Gallagher and Brent (34). Enzyme activity was determined by measuring the release of ethanol-soluble [^{3}H]methylated bases from [^{3}H]methylated salmon sperm DNA. The standard reaction mixture (100 μl) contained 20 μg of [^{3}H]DNA, 75 μg of crude protein extract, and buffer (50 mM Tris·HCl, pH 7.5/20 mM KCl/1 mM EDTA/1 mM 2-mercaptoethanol). The reaction was terminated by cooling on ice and the immediate addition of ⅓ vol of 3 M sodium acetate and 2 vol of ethanol. The DNA was then allowed to precipitate overnight at −20°C. After centrifugation for 15 min at 12,000 × *g* at 4°C, the supernatant was saved and concentrated to dryness in a Speed Vac concentrator (Savant). The residue was redissolved in 2.5 mM sodium phosphate buffer, pH 7.2, supplemented with unlabeled 3MeAde and 7MeGua (Sigma) as internal standards. The chemical nature and radioactivity of the [^{3}H]methylated bases in this fraction were then determined by HPLC using a reverse-phase C_{18} column [Apex ODS 5 μ (Jones Chromatography USA, Lakewood, CO); 250 mm × 4.6 mm] isocratically eluted at 1 ml/min. The mobile phase was 2.5 mM potassium phosphate buffer, pH 7.2/methanol (75:25, vol/vol). The

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. X76169).

authentic markers were detected by UV absorbance at 254 nm.

RESULTS

Screening of the *Arabidopsis* cDNA Library. Wild-type *E. coli* produces two 3MeAde glycosylases. 3MeAde glycosylase I (DNA-3-methyladenine glycosidase I, EC 3.2.2.20) is the constitutively expressed product of the wild-type *tag* gene and is highly specific for 3MeAde-containing DNA as a substrate, although it has recently been shown to excise 3-methylguanine with low efficiency (35). 3MeAde glycosylase II (EC 3.2.2.21), the product of the *alkA* gene, is induced by methylating agents and releases not only 3MeAde but also a variety of other methylated bases (36), including 7MeGua. MV1932 is a *tag, alkA1* double mutant of *E. coli* defective in the production of both these glycosylases and is extremely sensitive to the lethal effects of alkylating agents such as MMS. An *Arabidopsis* cDNA plasmid library derived from a λYES expression library (25) was screened for complementation of the MMS-sensitive phenotype of MV1932. MV1932 was transformed with 100 ng of the plasmid cDNA library by electroporation. The transformed bacteria were plated on LB Ap agar. The resulting colonies were scraped off the plates, resuspended in LB broth, and replated on selective LB Ap agar containing 1 mM MMS at a density of 10^8 to 10^9 colony-forming units per 9-cm plate. Approximately 1 in 10^7 cells survived this treatment to produce colonies. After 24 hr of incubation at 37°C, 2000 of the surviving colonies were individually streaked on the same medium; approximately half of colonies survived this second round of selection. To distinguish between MMS-resistant revertants and colonies which had gained an MMS-resistance factor from the transforming plasmid, plasmids were isolated from the 80 MMS-resistant colonies and retransformed into MV1932, selecting only for Ap resistance. One of these 80 minilysates yielded a plasmid which carried a heritable MMS-resistance factor. The approximately 1-kb *Eco*RI fragment containing the cDNA was subcloned into pUC18. This plasmid was designated pMMS^R.

Complementation of the MMS-Sensitive Phenotype. To determine the degree to which expression of the *Arabidopsis* cDNA complemented the MMS-sensitive phenotype of *E. coli* strains defective in repair of alkylation damage, we measured the lethal effects of a brief (20-min) exposure to increasing concentrations of MMS in unadapted strains with and without the cDNA. Fig. 1 illustrates the extreme MMS sensitivity of the double mutant MV1932 in comparison to its wild-type progenitor AB1157. While exposure to 20 mM MMS has no significant effect on the plating efficiency of the progenitor strain, only 1 out of 154,000 cells of the mutant strain survives exposure to 20 mM MMS. In contrast, 1 out of 25 cells of the identical strain transfected with the pMMS^R survives the same treatment, an improvement in survival of approximately 6000-fold.

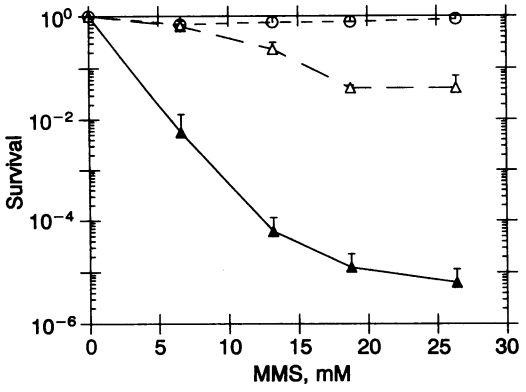

FIG. 1. Expression of the *Arabidopsis* cDNA enhances the MMS resistance of an *E. coli* strain defective in 3MeAde glycosylases I and II. Stationary-phase cells were exposed to MMS for 20 min, then plated on LB agar. ○, AB1157 pUC18; △, MV1932 pMMS^R; ▲, MV1932 pUC18.

Complementation of the Biochemical Defect. [^{3}H]Methylated DNA was incubated with a nonsaturating amount of freshly prepared cell extract (75 μg of protein and 20 μg of DNA) either from the wild-type strain AB1157 (harboring pUC18) or from the double-mutant strain MV1932 carrying pMMS^R and incubated at 37°C or room temperature (22°C). Aliquots were withdrawn at four intervals over a 45-min period, and the ethanol-soluble bases were extracted and analyzed as described in *Materials and Methods*. The results are presented in Fig. 2. In the absence of added extract, or in the presence of extract prepared from MV1932 pUC18, the amount of 3[^{3}H]MeAde released from the substrate DNA sugar-phosphate backbone was less than 2.5% of the total available substrate, and the concentration of free 3MeAde did not increase with time. In contrast, when the [^{3}H]methylated DNA was incubated in the presence of extracts from either AB1157 pUC18 or MV1932 pMMS^R, increased amounts of free 3MeAde were detectable after 5 min, and the concentration of this compound increased with time (Fig. 2*A*).

Because *Arabidopsis* grows optimally at 22°C and poorly at 37°C, while *E. coli* grows rapidly at 37°C but slowly at 22°C, we measured the release of 3MeAde by the crude cell extracts under both of these conditions. Because the specific activity of our preparations of MV1932 pMMS^R varied from one preparation to the next, the activity at each temperature was

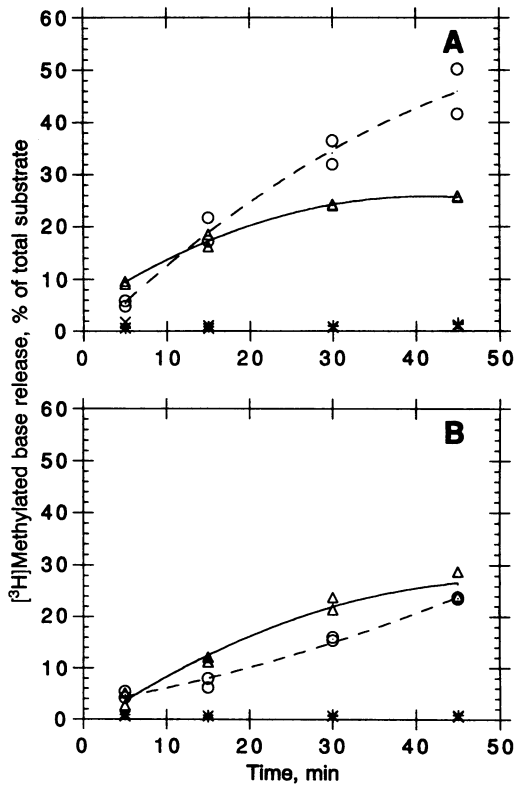

FIG. 2. Kinetics of release of [^{3}H]methylated bases by *E. coli* extracts. (*A*) Release of 3MeAde and 7MeGua at 37°C. ○, 3MeAde by AB1157 pUC18 extract; △, 3MeAde by MV1932 pMMS^R; +, 7MeGua by AB1157 pUC18; ×, 7MeGua by pMMS^R. Points represent each of two values obtained from duplicate reactions with a single extract preparation; curves are drawn through the average of those two values. (*B*) Release of 3MeAde and 7MeGua at 22°C, symbols as in *A*. The total radioactivity incorporated in the substrate DNA was 5000 cpm as 3MeAde and 28,400 cpm as 7MeGua.

measured using a single extract. The rate of release of 3MeAde by the *Arabidopsis* protein was unaffected by this temperature variation, while, in contrast, the rate of release of 3MeAde by the *E. coli* glycosylase was nearly halved by the drop in temperature (Fig. 2*B*).

No release of 7MeGua was observed in either the AB1157 pUC18 or the MV1932 pMMS^R preparations at either 37°C or 22°C (Fig. 2). While some free 7MeGua, corresponding to approximately 1% of the total available 7MeGua, is detectable in our reactions, this free base is also observed in the "no extract" controls, and the amount of free 7MeGua does not increase with time of incubation. In contrast, the wild-type *E. coli* extract released as much as 50% of the available 3MeAde during a 45-min incubation, while the MV1932 pMMS^R extract released 25% of the substrate. The lack of detectable activity specific for 7MeGua in the wild-type *E. coli* extract is not surprising, as the extracts were prepared from unadapted cells, and the *alkA* gene (which encodes the glycosylase that recognizes 7MeGua) is not induced under these conditions (24).

Sequence of the cDNA. Both strands of the entire *Eco*RI fragment containing the cDNA were sequenced after subcloning in pUC18. Fig. 3*A* presents a schematic of the sequencing strategy. The linker sequences did not match either of the two described for the λYES library (25) but were instead found to be (including the *Xho* I site) CTCGAGCTACGTCAGGG. The full-length cDNA is included in a 1037-bp *Eco*RI fragment. A single long open reading frame was found (762 bp), starting at an ATG 16 bases from the 5′ end of the cDNA. The *Arabidopsis* long open reading frame is not in frame with the *E. coli lacZ* fragment, and it terminates nearly 200 bases 5′ of the poly(A) tail. For this reason it is unlikely that the plasmid-encoded 3MeAde glycosylase is a LacZ fusion protein. A second in-frame ATG is present at nucleotide position 166; it is unlikely that this codon represents the start site in *Arabidopsis*, however, as the nucleotides immediately bordering the upstream ATG (TATAAUGAA) are a better match to the plant translational initiation consensus sequence (AACAAUGGC) (37). The predicted protein is 254 amino acids in length, with a molecular mass of 28.1 kDa and a charge of +8 at pH 7. This size is comparable to previously published 3MeAde glycosylase sequences for *E. coli* (Tag = 21.1 kDa, AlkA = 31.4 kDa), *S. cerevisiae* (34.3 kDa), humans [32.9 kDa (17) and 25.5 kDa (19)], rats (27.9 kDa), and mice (36.5 kDa).

The *Arabidopsis* cDNA Is Homologous to Mammalian 3MeAde Glycosylases. The predicted 254 amino acid sequence of the *Arabidopsis* 3MeAde glycosylase contains several small regions homologous to sequences in the rat and human glycosylases, but it lacks an intuitively obvious overall homology to these proteins. To determine whether the observed similarity is significantly better than that expected for a comparison with a random protein found in nature, we used the MAXSEGS dynamic sequence analysis program for local alignments to determine whether the *Arabidopsis* sequence is more similar (i.e., can produce a better alignment score) to the previously cloned 3MeAde glycosylases than to the other approximately 30,000 entries in the Swiss-Prot data bank. The final score of the computer-generated pairwise alignment is determined by summing the score of each pair of amino acids. Each pairwise score comes from an empirically determined table, which is derived from the known frequency, for a given evolutionary distance, of mutation (and fixation in a population) of one amino acid into another. We searched the Swiss-Prot data base two times, using PAM 120 and 250 tables (corresponding to mutation frequencies of 120 mutations per 100 bases and 250 mutations per 100 bases). At each PAM value, we found that, of the 30,000 entries in the Swiss-Prot library, the rat 3MeAde glycosylase (the *APDG* gene) sequence had the highest alignment score with our *Arabidopsis* clone. Fig. 3*B* shows the partial alignment of the *Arabidopsis MAG* cDNA and rat 3MeAde-DNA glycosylase amino acid sequences.

The *Arabidopsis* Genome Carries a Single Copy of the 3MeAde Glycosylase Gene. Fig. 4 shows a Southern blot of *Xho* I (lane 2) and *Xba* I (lane 4) restriction digests of 2 μg of DNA from *Arabidopsis* (ecotype Columbia), while a molar equivalent (assuming a 120,000-kb genome) of a linearized 8.9-kb plasmid carrying the intact *Arabidopsis* cDNA was loaded in lanes 1 and 3. The blot was hybridized with the entire ^{32}P-labeled 1-kb *Arabidopsis MAG* cDNA. From the relative band intensities we concluded that the cDNA corresponds to a single-copy sequence.

DISCUSSION

3MeAde acts as a block to both transcription and replication, and so may be lethal if left unrepaired. This lesion, or the apurinic site resulting from its repair, is subject to error-prone "repair" in *E. coli* (13), and therefore it is premutagenic, as well as cytotoxic. The presence of a 3MeAde glycosylase that cleaves the modified base from the DNA backbone has been well documented in microbial and mammalian systems but

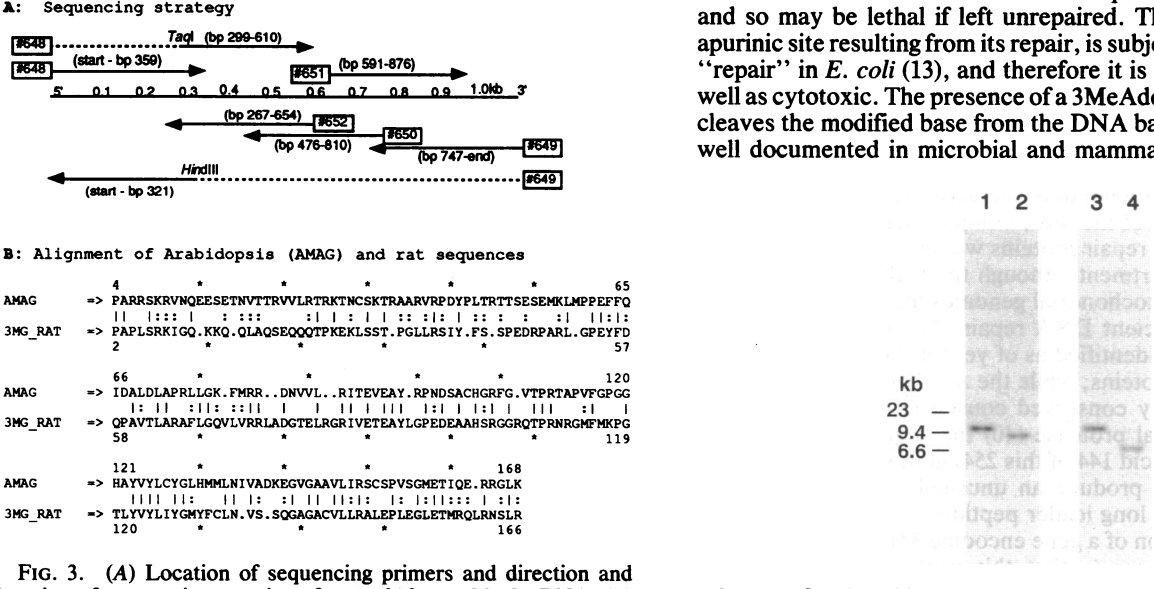

FIG. 3. (*A*) Location of sequencing primers and direction and location of sequencing reactions for *Arabidopsis MAG* cDNA. (*B*) Partial alignment of rat and *Arabidopsis* cDNA amino acid sequences produced by the MAXSEG program using a PAM 250 scoring matrix.

FIG. 4. Southern blot to determine copy number. Lanes 2 and 4, 2 μg of *Arabidopsis* DNA digested with *Xho* I and *Xba* I, respectively; lanes 1 and 3, a molar equivalent of a linearized 8.9-kb clone carrying the intact cDNA.

has not been described in plants. We have isolated an *Arabidopsis* cDNA clone, termed pMMS^R, which complements the MMS sensitivity of an *E. coli* strain defective in the repair of this lesion, increasing survival after treatment with MMS by nearly four orders of magnitude (Fig. 1). This substantial, though still incomplete, restoration of MMS resistance suggests that the *Arabidopsis* gene product is directly involved in the repair of 3MeAde.

The *E. coli tag*$^-$, *alkA1* mutant MV1932 has no detectable 3MeAde glycosylase activity. As shown in Fig. 2, this activity is restored in crude extracts prepared from MV1932 expressing the *Arabidopsis* clone, suggesting that the cDNA encodes a 3MeAde glycosylase. The yeast and mammalian 3MeAde glycosylases also release 7MeGua, although the cleavage at this base occurs at a substantially lower efficiency than at 3MeAde (16, 18–21, 38). However, the release of 7MeGua by the *Arabidopsis* enzyme, if it occurred at all, was below the level of detectability in our assays.

3MeAde glycosylases have been cloned from microbial (*E. coli* and *S. cerevisiae*) and mammalian (mouse, rat, and human) sources. A variety of amino acid sequence alignments have been produced which purport to show significant homology between these sequences. Although the homology between the mammalian sequences is intuitively obvious, the suggested trans-kingdom homology (and the homology between the two *E. coli* 3MeAde glycosylases), involving alignment of common amino acids over short regions, is subtle and so is subject to interpretation. To determine, qualitatively, the significance of observed similarities between the predicted *Arabidopsis* sequence and previously described mammalian sequences, we searched the entire contents of the Swiss-Prot data bank for the proteins that had the best local alignment to the *Arabidopsis* sequence. The alignment score of the rat 3MeAde glycosylase was higher than that of the other approximately 30,000 library entries. Because the *Arabidopsis* cDNA was cloned on the basis of its functional complementation of an *E. coli* mutant and sequence homology was not a criterion for the selection of this clone, the preferential alignment with the rat 3MeAde glycosylase sequence suggests that the *Arabidopsis* cDNA encodes an evolutionarily or functionally related protein.

Southern blot analysis using the intact cDNA as a probe and equimolar amounts of chromosomal and plasmid DNAs indicates that the *Arabidopsis MAG* gene is present as a single copy per genome sequence and that, unlike many other plant genes, there are no closely related sequences elsewhere in the genome. Because the plastid and mitochondrial genomes are observed in high copy number (several hundred copies per cell) in DNA preparations from the leaves of adult plants (39), the low copy number of the gene indicates that it is nuclearly encoded. The localization of the gene product (whether to the mitochondria, the plastids, the nucleus, or to some combination of the three compartments) remains to be determined. DNA repair proteins would seem to be required in all three compartments, though the high copy number of the plastid and mitochondrial genomes may reduce the need for rapid and efficient DNA repair. There are no strongly conserved motifs identified as of yet for the compartmentalization of plant proteins; while the *Arabidopsis* protein does include the weakly conserved consensus sequence for the chloroplast stromal protease (40) the sequence (VXAA) is located at amino acid 144 of this 254-amino acid protein and, if cleaved, would produce an unusually small glycosylase with an unusually long leader peptide.

The identification of a gene encoding 3MeAde glycosylase in *Arabidopsis* suggests that this protein plays a role in reducing the cytotoxic and potentially mutagenic effects of alkylated bases in plants. Determination of the biological significance of this protein will require further genetic, biochemical, and molecular analysis.

The authors thank the Plant Cell Biology Training Grant at University of California at Davis for the use of their facilities. We also thank Tim O'Connor (Institute Gustave-Roussy) and Tina Mok and Doug Nelson (University of California at Davis) for technical advice. This work was supported by National Science Foundation Grant DMB-90-19159 and in part by a grant from the Pittsburgh Computing Center through the National Institutes of Health National Center for Research Resources (Cooperative Agreement 1 P41 RR06009).

1. Xiao, W. & Samson, L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2117–2121.
2. Rebeck, G. W. & Samson, L. (1991) *J. Bacteriol.* **173**, 2068–2076.
3. Sargentini, N. J. & Smith, K. C. (1981) *Carcinogenesis* **2**, 863–872.
4. Smith, K. C. (1992) *Mutat. Res.* **277**, 139–162.
5. Kieber, J. J., Tissier, A. F. & Signer, E. (1992) *Plant Physiol.* **99**, 1493–1501.
6. Cerutti, H., Osman, M., Grandoni, P. & Jagendorf, A. T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8068–8072.
7. Vaughan, P., Sedgwick, B., Hall, J., Gannon, J. & Lindahl, T. (1991) *Carcinogenesis* **12**, 263–268.
8. Roberts, J. J. (1978) in *Advances in Radiation Biology*, eds. Lett, J. T. & Adler, H. (Academic, New York), Vol. 7, pp. 211–436.
9. Larson, K., Sham, J., Shenkar, R. & Strauss, B. (1985) *Mutat. Res.* **150**, 77–84.
10. Loechler, E. L., Green, C. L. & Essigman, J. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6271–6275.
11. Bagg, A., Kenyon, C. J. & Walker, G. C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5749–5753.
12. Singhal, R. K., Hinkle, D. C. & Lawrence, C. W. (1992) *Mol. Gen. Genet.* **236**, 17–24.
13. Foster, P. L. & Eisenstadt, E. (1985) *J. Bacteriol.* **163**, 213–220.
14. Clarke, N. D., Kvaal, M. & Seeberg, E. (1984) *Mol. Gen. Genet.* **197**, 368–372.
15. Chen, J., Derfler, B., Maskati, A. & Samson, L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7961–7965.
16. Berdal, K. G., Bjøräs, M., Bjelland, S. & Seeberg, E. (1990) *EMBO J.* **9**, 4563–4568.
17. Samson, L., Derfler, B., Boosalis, M. & Call, K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9127–9131.
18. Chakravarti, D., Ibeanu, G. C., Tano, K. & Mitra, S. (1991) *J. Biol. Chem.* **266**, 15710–15715.
19. O'Connor, T. R. & Laval, J. (1991) *Biochem. Biophys. Res. Commun.* **176**, 1170–1177.
20. Engelward, B. P., Boosalis, M. S., Chen, B. J., Deng, Z., Siciliano, M. J. & Samson, L. D. (1993) *Carcinogenesis* **14**, 175–181.
21. O'Connor, T. & Laval, F. (1990) *EMBO J.* **9**, 3337–3342.
22. Bachmann, B. J. (1987) in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, eds. Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M. & Umbarger, H. E. (Am. Soc. Microbiol., Washington, DC), pp. 1190–1219.
23. Karran, P. & Lindahl, T. (1980) *J. Mol. Biol.* **140**, 101–127.
24. Evensen, G. & Seeberg, E. (1982) *Nature (London)* **296**, 773–775.
25. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1731–1735.
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
27. Shure, M., Wessler, S. & Fedoroff, N. (1983) *Cell* **35**, 225–233.
28. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
29. Waterman, M. S. & Eggert, M. (1987) *J. Mol. Biol.* **197**, 723–728.
30. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Structure and Sequence*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–352.
31. Riazuddin, S. & Lindahl, T. (1978) *Biochemistry* **17**, 2110–2118.
32. Male, R. N., Nes, I. & Kleppe, K. (1981) *Eur. J. Biochem.* **121**, 243–248.
33. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254.
34. Gallagher, P. E. & Brent, T. P. (1982) *Biochemistry* **21**, 6404–6409.
35. Bjelland, S., Bjøräs, M. & Seeberg, E. (1993) *Nucleic Acids Res.* **21**, 2045–2049.
36. Sakumi, K. & Sekiguchi, M. (1990) *Mutat. Res.* **236**, 161–172.
37. Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. & Scheele, G. A. (1987) *EMBO J.* **9**, 43–48.
38. Male, R., Helland, D. E. & Kleppe, K. (1985) *J. Biol. Chem.* **260**, 1623–1629.
39. Pruitt, R. E. & Meyerowitz, E. M. (1986) *J. Mol. Biol.* **187**, 169–183.
40. Gavel, Y. & von Heijne, G. (1990) *FEBS Lett.* **261**, 455–458.
41. Pang, Q., Hays, J. B. & Rajagopal, I. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8073–8077.
42. Pang, Q. Hays, J. B., Rajagopal, I. & Schaefer, T. S. (1993) *Plant Mol. Biol.* **22**, 411–426.