NH$_2$-terminal Proline Acts as a Nucleophile in the Glycosylase/AP-Lyase Reaction Catalyzed by *Escherichia coli* Formamidopyrimidine-DNA Glycosylase (Fpg) Protein*

(Received for publication, October 9, 1996, and in revised form, December 4, 1996)

Dmitry O. Zharkov, Robert A. Rieger, Charles R. Iden, and Arthur P. Grollman‡

From the Department of Pharmacological Sciences, The State University of New York, Stony Brook, New York 11794-8651

Formamidopyrimidine-DNA glycosylase (Fpg) protein plays a prominent role in the repair of oxidatively damaged DNA in *Escherichia coli*. The protein possesses three enzymatic activities, hydrolysis of the N-glycosidic bond (DNA glycosylase), β-elimination (AP lyase), and δ-elimination; these functions act in a concerted manner to excise oxidized deoxyxynucleosides from duplex DNA. Schiff base formation between the enzyme and substrate has been demonstrated (Tchou, J., and Grollman, A. P. (1995) J. Biol. Chem. 270, 11671–11677); this protein-DNA complex can be trapped by reduction with sodium borohydride. By digesting the stable, covalently linked intermediate with proteases and determining the accurate mass of the products by negative electrospray ionization-mass spectrometry, we show that the N-terminal proline of Fpg protein is linked to DNA and, therefore, is identified as the nucleophile that initiates the catalytic excision of oxidized bases from DNA. This experimental approach may be applicable to the analysis of other protein-DNA complexes.

Fpg, a DNA base excision repair enzyme with N-glycosylase and AP lyase activities (1–3), efficiently removes 8-oxoquaguanine (8-oxoGua) and formamidopyrimidines from oxidatively damaged DNA (4–7). We have shown previously (8) that this reaction involves an imino-enzyme-substrate (Schiff base) intermediate and that the amino group involved is located within a 72-amino acid fragment of Fpg protein containing the N terminus. The four-cysteine zinc finger motif located near the C terminus (9, 10) is utilized in binding oxidatively damaged DNA (10, 11). The 8-oxo moiety is a critical structural determinant by which Fpg protein recognizes duplex DNA substrates containing 8-oxoquaguanine (5).

It has been proposed that Schiff bases participate in the catalytic action of AP endonucleases (12) and DNA glycosylases that possess AP lyase activity (13–16). Sodium borohydride and cyanoborohydride have been used to trap Schiff base intermediates as covalently linked DNA-protein complexes in reactions catalyzed by Fpg protein (8, 16) and by bacteriophage T4 endonuclease V, *Micrococcus luteus* UV endonuclease and *Escherichia coli* endonuclease III (15, 16). A mechanism involving nucleophilic attack on C1' of the modified deoxyinosine targeted for excision (8, 15–17) has been proposed to explain the catalytic action of these enzymes. Nucleophilic attack, facilitated by protonation of the base, effects cleavage of the N-glycosidic bond. The AP lyase activity of Fpg protein is a concerted reaction involving β- and δ-elimination reactions (2, 18, 19), producing a gap in one strand of the duplex demarcated by 3'- and 5'-phosphate termini (2, 18).

Several nucleophiles capable of forming Schiff base intermediates (20) are located within the 72-residue N-terminal fragment of Fpg protein. Basic nitrogen functional groups are found in the N-terminal proline, the free amino group of Lys-56, and eight arginine residues. Following enzymatic digestion of the covalently linked complex, we used electrospray ionization-mass spectrometry (ESI-MS) to identify unambiguously the residual peptide-DNA fragment, leading to the conclusion that the N-terminal proline residue links the peptide to DNA. To our knowledge, Schiff base formation involving nucleophilic attack by secondary amines has not been reported for biological systems. The experimental approach used in this study is applicable to the structural analysis of protein-DNA complexes.

**EXPERIMENTAL PROCEDURES**

Oligodeoxyxynucleotides—Oligodeoxynucleotides containing 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxo-dG) (21) and a carba analogue of 8-oxo-dG (2) (Fig. 1) were synthesized on an Applied Biosystems model 394 DNA Synthesizer using standard phosphoramidite chemistry. 5'-Dimethoxytrityl-protected phosphoramidite derivatives were used to prepare the 23-mer, 5'-CTCTCCCTTCXCTCTCTCTCT-3' (X = modified nucleotide). 2'-Deoxyctydine was positioned opposite the modified base in the complementary strand. The normal oligomer was deprotected and removed from the support by treatment with 28% ammonia at 55 °C for 16 h. For oligomers containing 8-oxo-dG, β-mercaptoethanol (final concentration, 0.1 mM) was added prior to the deprotection reaction. The resin was removed by filtration, and the solution was evaporated to dryness at room temperature in a SpeedVac (Savant). 5'-DMT-protected oligomers were purified by reverse phase HPLC (Waters) on a semipreparative PRP-1 column (Hamilton) and eluted over 20 min with 16–33% acetonitrile, 0.1 M triethylammonium acetate, pH 7.0, at a flow rate of 2 ml/min. Fractions containing the 5'-DMT-protected oligomer were combined and lyophilized; the 5'-DMT moiety was removed by treatment for 30 min with 80% acetic acid. Oligomers were further purified on a PRP-1 column, this time eluting over 15 min with 0–12% acetonitrile, 0.1 M triethylammonium acetate, pH 7.0, at a flow rate of 2 ml/min. The fraction containing the desired product was evaporated to dryness, dissolved in 1 ml of water, and stored frozen at −20 °C.

**Enzymes**—The general procedure for the purification of *E. coli* Fpg protein has been described elsewhere (10) and was used with minor modifications. Briefly, *E. coli* JM109 harboring the pKK233-3 overexpression plasmid with the fpg insert were grown at 37 °C to an *A$_{600}$* of 0.7. Cells were induced by adding isopropylthio-β-D-galactoside (1 mM), incubated at 37 °C for 3 additional h, and harvested by centrifugation.

---

* This work was supported by Grants CA17395 and ES04068 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1754 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 516-444-3080; Fax: 516-444-7641; E-mail: apg@pharm.som.sunysb.edu.

1 The abbreviations used are: Fpg, formamidopyrimidine-DNA glycosylase; AP, apurinic; carba-8-oxo-dG, an analogue of 8-oxo-7,8-dihydro-2'-deoxyguanosine in which the oxygen atom in the deoxyribose ring is replaced by CH$_2$; DMT, dimethoxytrityl; ESI-MS, electrospray ionization-mass spectrometry; HPLC, high performance liquid chromatography; 8-oxo-dG, 8-oxo-7,8-dihydro-2'-deoxyguanosine; 8-oxoGua, 8-oxoquaguanine; PAGE, polyacrylamide gel electrophoresis.

2 G. Dorman and F. Johnson, personal communication.
Cells were washed with 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, resuspended in a minimal volume of Buffer A (50 mM HEPES-KOH, pH 7.6, 1 mM EDTA, 1 mM dithiothreitol, 10% v/v glycerol) containing 200 mM NaCl and 1 mM phenylmethylsulfonyl fluoride, and subjected to ultrasonic disruption (Biosonik IV, VWR Scientific) for 5 min (1-min pulses, 5 min between pulses). Following centrifugation, the supernatant was diluted with Buffer A (A<sub>260</sub> = 10), and NaCl was added to bring the final salt concentration to 200 mM. This solution was applied to a Q Sepharose Fast Flow column. The flow-through fraction was loaded onto a S Sepharose Fast Flow column, and the protein was eluted by a 200–600 mM NaCl gradient. Fractions containing DNA glycosylase activity were pooled, precipitated with ammonium sulfate, dissolved in Buffer A containing 1 mM NaCl, and subjected to gel filtration on a Superdex 75 HR 10/30 column (Pharmacia Biotech Inc.). The main UV-absorbing peak was collected, dialyzed against 200 ml of Buffer A containing 250 mM NaCl and 50% v/v glycerol, and stored at −20°C. Coomassie Blue-stained SDS-PAGE analysis showed the protein to be at least 90% pure.

Trypsin and Pronase E were purchased from Sigma. Stock solutions were stored frozen at −20°C. Pronase E was found to be essentially free from contamination with DNase.

**Trapping the Schiff Base Intermediate—**An oligonucleotide duplex containing a single 8-oxo-dG residue positioned opposite dC was mixed with Fpg protein in the presence of freshly dissolved NaBH<sub>4</sub>. The reaction mixture included 1 μM oligodeoxynucleotide, 25 mM potassium phosphate, pH 6.8, 100 mM NaCl, 100 mM NaBH<sub>4</sub>, and varying amounts of Fpg protein, added last to minimize substrate cleavage. For the Edman degradation experiments, a 1:2 ratio of Fpg:duplex was used in a final volume of 10–100 μl. For preparative scale synthesis, the ratio of Fpg:duplex was 5:1 in a reaction volume of 10 ml. To reduce foaming and protein denaturation, the preparative reaction mixture was allowed to equilibrate for 30 min before oligodeoxynucleotide and Fpg protein were added. All reactions were incubated for 1 h at 37°C.

**Tryptic Digestion of the Covalent Complex—**For analytical purposes, an aliquot (10 μl) of the reaction mixture was removed, dried in a SpeedVac, dissolved in 10 μl of 8 M urea, and heated for 10 min at 95°C. Tris-HCl, pH 7.5, urea, and trypsin were added, bringing the solution to a final concentration of 2 mM urea, 20 mM Tris-HCl, and 4% w/v trypsin/Fpg. The reaction mixture was incubated for 1 h at 37°C and then heated for 10 min at 95°C to terminate the reaction. In preparative scale reactions, the NaBH<sub>4</sub> reaction mixture was loaded on Centricon-10 concentration devices (Amicon); the volume was reduced to 200–300 μl, and the sample was washed three times with 2 ml of water. After the final wash, the sample was dried, dissolved in 50 μl of 8 M urea, and heated for 10 min at 95°C. Tris-HCl, pH 7.5, and urea were added to the solution; the final concentration was 2 M urea and 20 mM Tris-HCl. Trypsin was added (4% w/v trypsin/Fpg, assuming all of the Fpg protein had been recovered) following the Centricon concentration step. After 1 h incubation at 37°C, the sample was desalted by washing three times with 2 ml of water in a Centricon-3 device and dried in a SpeedVac.

**Pronase Digestion of the Covalent Intermediate—**After NaBH<sub>4</sub> reduction of the protein-DNA complex, glucose was added (final concentration, 400 mM), and the solution was incubated at 37°C for 3 h. Fpg (20 ng for the analytical reaction and 20 μg for the preparative reaction) was added after 2 h to cleave unreacted substrate. In the analytical reaction, Pronase E (final concentration, 2 mg/ml) was added directly to the reaction mixture. In the preparative reaction, the mixture was loaded onto a Q Sepharose Fast Flow column (10-ml bed volume) and eluted in a 100-ml gradient of 200–1000 mM NaCl with 20 mM Tris-HCl, pH 7.5. Protein samples were analyzed by SDS-PAGE. Peaks containing cross-linked material, which eluted at ~650 mM NaCl, were pooled, concentrated by evaporation to ~500 μl, and dialyzed twice against 1 liter of 10 mM Tris-HCl, pH 7.5, 10 mM Na-EDTA. Pronase E was added to the final concentration of 1 mg/ml. The reaction mixture was incubated at 37°C for 24 h, desalted using a SepPak reverse phase cartridge (Waters), and dried in a SpeedVac.

**PAGE and Reverse Phase HPLC—**Following tryptic or Pronase digestion, the samples were dried, dissolved in 10 μl of 8 M urea, heated for 5 min at 95°C, and loaded onto a 0.4-mm 20% polyacrylamide gel containing 8 M urea. Electrophoresis was conducted at 50 V/cm, maintaining constant voltage, with xylene cyanol loaded in a separate lane as a mobility marker. After the dye migrated approximately 65 cm, oligonucleotide-containing material was localized by UV shadowing; bands were cut from the gel, crushed with a spatula, and extracted twice with 1 ml of 1 mM Tris-HCl, pH 8.0, 0.1 mM EDTA for 2 h and the third time overnight. Extracts were concentrated to 1 ml, loaded on a NovaPak C<sub>18</sub> reverse phase column (Waters), and eluted with 0–15% acetonitrile, 0.1 M triethylammonium acetate buffer, pH 7.0, at 1 ml/min. Fractions were collected in Eppendorf tubes, dried in vacuo, and stored frozen.

**Electrospray Mass Spectrometry—**Samples were diluted to an approximate concentration of 20–50 μg using a solution of 80% acetonitrile/water containing 1% triethylamine and then injected into the electrospray source on a TRIO-2000 mass spectrometer (Micromass). The source was supplied with a flow of solvent, 60% acetonitrile/water, at a rate of 12 μl/min by means of a syringe pump (Harvard Apparatus). The mass spectrometer was operated in the negative ion spray mode with the probe voltage set to ~3.3 kV. Nitrogen was used as the drying gas (250 liters/h) and also as the nebulization gas (15 liters/h). Approximately eight scans were averaged; the base line was adjusted and filtered digitally to produce each mass spectrum. Protein Analysis Worksheet (PAWS, v6.1.1)<sup>3</sup> was used to identify amino acid sequences in the Fpg protein with masses equal to the observed mass of the peptide portion of the complex.

**Edman Degradation of the Covalent Intermediate—**The NaBH<sub>4</sub> reaction mixture was added to an equal amount of SDS reducing buffer (62.5 mM Tris-HCl, pH 6.8, 2% w/v SDS, 5% v/v β-mercaptoethanol, 10% v/v glycerol, and 0.05% w/v bromphenol blue), heated for 5 min at 95°C, and then subjected to 12% discontinuous SDS-PAGE (22) on a Mini-PROTEAN II Dual Slab Cell (Bio-Rad). Electrophoresis was conducted in a cold room at 300 V. Bands were transferred onto an Immobilon P membrane (Waters) using the Mini Trans-Blot electrophoretic transfer cell (Bio-Rad). The membrane was stained briefly with Coomassie Blue and destained with methanol:acetic acid:water 4:1.5 (v/v/v). Bands were cut, washed 5 times with 10% methanol and 10 times with deionized water, and then subjected to Edman degradation (23). Sequence analysis was conducted on an Applied Biosystems 475A Protein Sequencer.

**RESULTS**

**Formation of the Oligodeoxynucleotide-Peptide Complex—**When a 5-fold excess of Fpg protein was added to the reaction mixture, 90–95% of the 8-oxoguanine-modified oligonucleotide was covalently linked to Fpg in 1 h (Fig. 2A, lane 2). When the enzyme:substrate ratio was 1:2 (used in preparing a covalent complex for Edman degradation), approximately 80% of the protein was cross-linked in 1 h (Fig. 2B, lanes 1 and 2). No cleavage products were observed.

To exclude the possibility that a covalent bond was formed during the reduction step at a position other than C1 of 8-oxo-dG, a duplex oligodeoxynucleotide containing carba-8-oxo-dG (Fig. 1) was tested as a substrate for the cross-linking reaction. In this analogue, a methylene group replaces the ring oxygen in

![Fig. 1. Structures of the modified deoxynucleotides used in this study.](image-url)

<sup>3</sup> Protein Analysis Worksheet, Version 6.1.1, Copyright 1995, Ronald Beavis.
Mechanism of Action of Fpg Protein

the deoxyribose moiety of 8-oxo-dG; consequently, the glycosidic bond is stable to nucleophilic attack at C1. Duplexes containing the carbacyclic analogue bind Fpg protein tightly (Kd = 22.2 nM for carba-8-oxo-dG:dC and 8.9 nM for 8-oxo-dG:dC), but the duplex is not cleaved after 24 h incubation with a 20-fold excess of enzyme. As shown in Fig. 2A (lane 4), this modified oligomer does not generate a cross-linked product under conditions where more than 90% of the 8-oxo-dG-containing oligomer is covalently bound to Fpg.

The proteolytic digestion of the cross-linked complex is shown in Fig. 3. Following trypsin digestion for 1 h, the reduced enzyme-substrate complex was converted to lower molecular weight products (Fig. 3A, lane 3). Twenty-four-hour incubation was required to completely digest this complex with Pronase E (Fig. 3B, lane 3).

Isolation of the Cross-linked Complex—Following trypptic digestion, products of a preparative scale reaction were subjected to denaturing gel electrophoresis. Two clearly separated bands (data not shown) were excised and purified by reverse phase HPLC. The elution pattern of the lower mobility product (A) contained a single large peak with a retention time of 42 min (Fig. 4). The UV absorption spectrum of peak A (Fig. 4, inset) resembles that of a standard oligomer containing 8-oxo-dG (B); however, the shoulder absorbing at 300–320 nm, reflecting the presence of 8-oxo-dG, is not present in the spectrum of the reaction product. The higher mobility band co-migrated with the unmodified complementary strand during PAGE and HPLC.

Electrospray Mass Spectrometer Analysis—The negative ion ESI mass spectrum of product A (Fig. 5) shows an excellent signal:noise ratio and reveals the presence of two components. The major product is distributed into nine ion peaks (A series) with charged states ranging from −7 to −15; the mass of this product is 7786.4 ± 3.7 Da. A series of peaks representing a minor product (B series) also is evident in the mass spectrum. The measured molecular mass of this substance is 7944.1 ± 4.0 Da.

Mass data were analyzed by PAWS to determine the mass of those amino acid sequences within Fpg protein that correspond to the measured mass of the peptide portion of the complex. Since the 8-oxoGua moiety is cleaved by Fpg protein, the mass of the depurinated oligomer (6631.3 Da) must be used to calculate the mass of residual Fpg protein sequences covalently bound to the oligomer. Thus, the mass difference between the apurinic oligomer and products A and B, corresponding to the mass of the residual peptide fragment, is 1155.1 and 1312.8 Da, respectively. Mass calculations from the PAWS program returned a total of 11 sequences in the Fpg protein (268 amino acids) for the major product with a calculated mass of 8756.1 Da (Table I). The range of masses permitted in the calculation was ±3.7 Da, equivalent to one standard deviation in the error of the mass measurement. The calculated mass is 1 Da more...
than the measured mass, since the program calculates the mass of the neutral peptide molecule rather than the mass of the peptide cleavage product obtained from the complex. Only one of these sequences is a tryptic fragment. It is located at the N terminus of Fpg and contains the first 10 amino acids, PELPEVETSR. The 10 remaining sequences are not expected to be products of trypsin cleavage (Table II). The calculated mass of the N-terminal peptide fragment is 1155.3 Da; the measured mass of the residual peptide component of the complex is 1155.1 Da.

A similar search was performed to determine the sequence of the minor product found in the mass spectrum. For Fpg protein, the PAWS program returned 17 possible sequences with mass 1313.8 ± 4.0 Da (Table I); however, only two of those represent trypsin-generated fragments (Table II). One sequence corresponds to the first 11 amino acids of the Fpg sequence, PELPEVETSR, with a calculated mass of 1312.8 Da. The second fragment (Thr-151 through Lys-161) is located within the internal sequence of the protein and has a mass of 1315.6 Da.

A similar experiment was conducted using Pronase E. The ESI mass spectrum of the product eluted from the HPLC column showed a good signal/noise ratio, yielding a molecular mass of 7213.2 ± 1.4 Da (Fig. 6). Signals from other components present in the spectrum are insufficient to provide accurate mass assignments. Analysis of the major product by PAWS returned a total of 15 sequences in Fpg protein with a calculated mass of 582.9 ± 1.4 Da for the peptide component (Table I). One of these sequences contains the first five amino acids, PELPE, and the PAWS program returned 17 possible sequences with

---

**TABLE I**

Fpg protein sequences with the indicated mass

| Trypsin digest (series A), 1156.1 ± 3.7 Da | Trypsin digest (series B), 1313.8 ± 4.0 Da | Pronase digest, 582.9 ± 1.4 Da |
|------------------------------------------|------------------------------------------|-------------------------------|
| 1PELPEVETSR                              | 2PELPEVETSR                              | 1PELPEVETSR                  |
| 2VRNGLRWP                                | 2VRNGLRWP                                | 58LLELP                      |
| 2VRNGLRWP                                | 2VRNGLRWP                                | 3PEELP                       |
| 46PVLSVQRRA                              | 46PVLSVQRRA                              | 3PEELP                       |
| 47PVLSVQRRA                              | 47PVLSVQRRA                              | 83ELPPE                      |
| 72GMSGLRLPE                              | 72GMSGLRLPE                              | 83ELPPE                      |
| 132PLSDFNGEY                             | 132PLSDFNGEY                             | 83ELPPE                      |
| 173ESLFAAGHPD                            | 173ESLFAAGHPD                            | 83ELPPE                      |
| 199RVIKVLRS                              | 199RVIKVLRS                              | 83ELPPE                      |
| 202KAVLLRSI EQ                          | 202KAVLLRSI EQ                          | 83ELPPE                      |
| 1PELPEVETSR                              | 2PELPEVETSR                              | 1PELPEVETSR                  |
| 2VRNGLRWP                                | 2VRNGLRWP                                | 58LLELP                      |
| 2VRNGLRWP                                | 2VRNGLRWP                                | 3PEELP                       |
| 46PVLSVQRRA                              | 46PVLSVQRRA                              | 3PEELP                       |
| 47PVLSVQRRA                              | 47PVLSVQRRA                              | 83ELPPE                      |
| 72GMSGLRLPE                              | 72GMSGLRLPE                              | 83ELPPE                      |
| 132PLSDFNGEY                             | 132PLSDFNGEY                             | 83ELPPE                      |
| 173ESLFAAGHPD                            | 173ESLFAAGHPD                            | 83ELPPE                      |
| 199RVIKVLRS                              | 199RVIKVLRS                              | 83ELPPE                      |
| 202KAVLLRSI EQ                          | 202KAVLLRSI EQ                          | 83ELPPE                      |

**TABLE II**

Tryptic peptides generated from Fpg protein

Superscripts denote the position of the first amino acid in the fragment. Peptide bonds between lysine or arginine and proline are resistant to trypsin.

1PELPEVETSR13R13GIEPHLVGATILHAVVR20N29R29L29P29V29S29E29K29H29D29L29M29N29K29H

FIG. 5. Negative ion ESI mass spectrum of the HPLC-purified product of trypsin digestion of the Fpg protein-oligonucleotide complex.
Edman Degradation of the Covalent Complex—The reduced covalent complex was fully resistant to 10 cycles of Edman degradation (Table III), indicating that its N terminus was modified. A control reaction mixture containing NaBH₄ but lacking the oligonucleotide substrate was subjected to similar treatment. This reaction did not lead to blockage of the N terminus, and the eight amino acids in the N-terminal sequence of Fpg protein were identified.

**DISCUSSION**

In this study, negative electrospray ionization-mass spectrometry was used to determine the mass of a peptide-oligodeoxynucleotide complex created by reductive cross-linking of Fpg protein to a duplex oligodeoxynucleotide containing a single 8-oxo-dG residue. The complex was partially digested by treatment with trypsin or Pronase E and denatured, leaving a residual peptide bound to a single strand oligomer. Accurate masses for the peptide component, determined from the ESI-MS data, were used to identify amino acid sequences within Fpg. A single peptide was identified as the major product of the trypptic digest; its calculated mass was consistent with the measured mass of a trypptic fragment forming the peptide portion of the complex. This peptide, PELPEVETSR, representing the N terminus, contains 10 amino acids. The mass of a minor product also was determined from the same digest. Fpg contained two sequences consistent with the data. The first, PELPEVETSRR, includes the sequence of the major product with an additional Arg present at the C terminus. This peptide may be the result of incomplete trypptic digestion. The second sequence, Thr-151 to Lys-161, is located internally and is not part of the 72-amino acid N-terminal fragment of Fpg protein to which the active site has been mapped (8). The central role of the N-terminal sequence was confirmed by analysis of a peptide-DNA complex formed by digestion with Pronase E. Several sequences corresponding to the measured mass were identified (Table I); however, only two of these were located within 72 amino acids of the N terminus. One sequence, PELPE, represents the first five amino acids of the N terminus and completely overlaps both trypsin-generated fragments. The second sequence (LLELP, Leu-59–Pro-63) does not contain an amino group capable of Schiff base formation or share common elements with any of the trypsin-generated fragments. Other sequences compatible with the mass data are found in the internal sequence of Fpg protein and have no overlap with the trypsin-generated fragments. Most of these cannot form Schiff bases due to lack of a suitable functional group.

The three peptide-DNA complexes identified by ESI-MS have common sequence elements in the peptide portion of the molecule; in each case, N-terminal residues of Fpg protein, beginning with Pro-1, are covalently bound to the oligonucleotide. This N-terminal sequence is conserved in all known Fpg proteins (24). The ESI-MS technique does not identify the precise amino acid involved in the formation of the complex; however, the nature of the peptide sequences identified by this method indicates that only the N-terminal proline is capable of reacting with 8-oxo-dG in the substrate. Strong support for participation of the N terminus as an active site residue comes from experiments in which the covalently trapped intermediate was subjected to Edman degradation. Formation of the complex completely blocked this reaction, indicating that the N terminus was modified. It seems unlikely that steric inhibition by the oligonucleotide prevented access of low molecular weight reagents used in the Edman procedure to the N terminus. Combined with the ESI-MS data, resistance to Edman degradation permits us to identify unambiguously the N-terminal

**FIG. 6.** Negative ion ESI mass spectrum of the HPLC-purified product of Pronase E digestion of the Fpg protein-oligodeoxynucleotide complex.

**TABLE III**

| Fpg                  | PELPEVET<sup>a</sup> |
|----------------------|-----------------------|
|                      |                       |
| Fpg treated with NaBH₄, no substrate | PELPEVET<sup>b</sup> |
| Covalent complex     | N terminus blocked    |

<sup>a</sup> Data from GenBank.  
<sup>b</sup> Edman degradation.
proline as the nucleophile involved in base elimination reaction by Fpg protein.

A mechanistic scheme illustrating the involvement of Pro-1 in the glycosylase and AP lyase activities of Fpg-catalyzed excision is shown in Fig. 7. The nitrogen atom of proline attacks at C1 and, assisted by protonation at O6 of 8-oxoGua (5), cleaves the glycosidic bond leading to expulsion of the damaged base. Removal of a proton from the positively charged proline nitrogen, mediated by solvent or the enzyme, is expected to follow. Protonation of the sugar oxygen leads subsequently to opening of the deoxyribose ring and Schiff base formation. The positively charged nitrogen can facilitate hydrogen abstraction at C2', and β-elimination (syn) of the C3'-phosphate leads to DNA strand breakage. At this point the complex may undergo elimination of the C5'-phosphate, mechanically similar and vinylogous to the one already described. In the final step, the enzyme is regenerated by attack of hydroxyl ion on C1' with release of 4-oxo-2-pentenal.

The N-terminal residue of another DNA glycosylase/AP lyase, bacteriophage T4 endonuclease V, has been shown to be involved in Schiff base formation (14, 25–26). The reaction mechanism proposed above for Fpg is similar to that suggested for endonuclease V (14) and other DNA glycosylase/AP lyases (16) with respect to Schiff base formation. In endonuclease V, the nucleophile is the α-NH$_2$ group of Thr-1; when this residue is replaced by proline, the activity of the enzyme is reduced to negligible values (25). The facts that Pro-1 is conserved in all Fpg proteins (24) and replacement with glycine abolishes cleavage but not binding (8) indicate that proline must play an important role in Fpg catalysis.

The finding that the N-terminal proline of Fpg protein acts as a nucleophile is unique. As a strong base, proline can act as a general acid/base catalyst and has been shown to abstract a proton from a substrate molecule bound to 4-oxalocrotonate tautomerase (27). Our study presents evidence for the involvement of an N-terminal proline in nucleophilic catalysis with formation of a Schiff base. Formation of an imine with a positive charge on the nitrogen atom may be rendered more favorable by coupling to a charge-delocalizing system, as in pyridoxal-containing enzymes (20) or stabilized by some other means. In this case, the positive charge on nitrogen undoubtedly facilitates AP lyase activity (β- and δ-elimination) through abstraction of the C2'- and C4'-hydrogens. In addition, introduction of a positive charge between the highly negatively charged phosphodiester backbone and the negatively charged N-terminal region, which contains three glutamic acid residues in the first seven amino acids, may play a role in spatial organization of the active center or in stabilizing the intermediate complex.

The ESI-MS approach used in this study to identify amino acid residues involved in nucleophilic attack on a DNA substrate may be used to analyze other DNA glycosylases and DNA-protein complexes and may, in some cases, replace more time-consuming methods of identifying the peptide portions of these complexes (28). A stable covalent bond, linking protein to DNA, is required for ESI-MS analysis; two new ionization techniques, electrospray and matrix-assisted laser desorption, provide accurate mass assignments for large complexes of this type. Negative ion ESI is especially useful for molecules containing multiple negative charges and was used in this case because it was perceived that the charged state of the covalent complex would be controlled by the phosphates of the oligodeoxyribose. With pure samples, ESI-MS data yield mass accuracies of 0.01%; our measurements of more complex samples in low concentrations yielded results in the range 0.02--
Enzymatic degradation of protein and/or DNA was not required for this analysis; however, partial degradation by proteases was helpful in defining the position of the covalent bond and in restricting the number of possible amino acid sequences.

Acknowledgments—We are greatly indebted to Dr. Francis Johnson for many illuminating discussions regarding the mechanism of action of Fpg protein and for his guidance in the preparation of modified oligodeoxynucleotides; to Thomas Fischer for peptide sequence analysis, and to Dr. Nick Bulychev for helpful advice with laboratory procedures.

REFERENCES
1. Chetsanga, C. J., and Lindahl, T. (1979) *Nucleic Acids Res.* 6, 3673–3684
2. O'Connor, T. R., and Laval, J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* 86, 5222–5226
3. Friedberg, E. C., Walker, G. C., and Siede, W. (1995) *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, D. C.
4. Tchou, J., Kasai, H., Shibutani, S., Chung, M.-H., Laval, J., Grollman, A. P., and Nishimura, S. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 4690–4694
5. Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A. P., and Johnson, F. (1994) *J. Biol. Chem.* 269, 15318–15324
6. Boiteux, S., Gajewski, E., Laval, J., and Dizdaroglu, M. (1992) *Biochemistry* 31, 106–110
7. Karakaya, A., Jaruga, P., Bohr, V. A., Grollman, A. P., and Dizdaroglu, M. (1996) *Nucleic Acids Res.*, in press
8. Tchou, J., and Grollman, A. P. (1995) *J. Biol. Chem.* 270, 11671–11677
9. Boiteux, S., O'Connor, T. R., Lederer, F., Gouyette, A., and Laval, J. (1990) *J. Biol. Chem.* 265, 3916–3922
10. Tchou, J., Michaels, M. L., Miller, J. H., and Grollman, A. P. (1993) *J. Biol. Chem.* 268, 26738–26744
11. O'Connor, T. R., Graves, R. J., de Murcia, G., Castaigne, B., and Laval, J. (1993) *J. Biol. Chem.* 268, 9063–9070
12. Weiss, B., and Grossman, L. (1987) *Adv. Enzymol. Relat Areas Mol. Biol.* 60, 1–34
13. Kow, Y. W., and Wallace, S. S. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 82, 8354–8358
14. Dodson, M. L., Schrock, R. D., III, and Lloyd, R. S. (1993) *Biochemistry* 32, 8384–8290
15. Dodson, M. L., Michaels, M. L., and Lloyd, R. S. (1994) *J. Biol. Chem.* 269, 32709–32712
16. Sun, B., Latham, K. A., Dodson, M. L., and Lloyd, R. S. (1995) *J. Biol. Chem.* 270, 19501–19508
17. Grollman, A. P., Johnson, F., Tchou, J., and Eisenberg, M. (1994) *Ann. N. Y. Acad. Sci.* 726, 208–214
18. Bailly, V., Verly, W. G., O'Connor, T. R., and Laval, J. (1989) *Biochem. J.* 262, 581–589
19. Bhagwat, M., and Gerlt, J. A. (1996) *Biochemistry* 35, 659–665
20. Fersht, A. (ed) *Enzyme Structure and Mechanism*, 2nd Ed., pp. 69–75, W. H. Freeman & Co., New York
21. Bodepudi, V., Shibutani, S., and Johnson, F. (1992) *Chem. Res. Toxicol.* 5, 608–617
22. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Edman, P., and Begg, G. (1967) *Eur. J. Biochem.* 1, 80–91
24. Duwat, P., de Oliveira, E., Ehrlich, S. D., and Boiteux, S. (1995) *Microbiology* 141, 411–417
25. Schroock, R. D., III, and Lloyd, R. S. (1993) *J. Biol. Chem.* 268, 880–886
26. Schroock, R. D., III, and Lloyd, R. S. (1991) *J. Biol. Chem.* 266, 17631–17639
27. Stivers, J. T., Abeygunawardana, C., Mildvan, A. S., Hajipour, G., Whitman, C. P., and Chen, L. H. (1996) *Biochemistry* 35, 803–813
28. Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S., and Verdine, G. L. (1991) *Biochemistry* 30, 11018–11025