Crystal structure of the *Lactococcus lactis* formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA

Laurence Serre, Karine Pereira de Jésus¹,
Serge Boiteux², Charles Zelwer¹ and
Bertrand Castaing^{1,3}

Institut de Biologie Structurale, CNRS-CEA, 41 av. Jules Horowitz, 38027 Grenoble cedex 01, ¹Centre de Biophysique Moléculaire UPR4301 affiliated to the University of Orléans, CNRS, rue Charles Sadron, 45071 Orléans cedex 02 and ²Laboratoire de Radiobiologie du DNA, UMR217, CNRS-CEA, Centre d'Etudes Nucléaires, BP6, 92265 Fontenay-Aux-Roses, France

³Corresponding author
e-mail: castaing@cnrs-orleans.fr

The formamidopyrimidine-DNA glycosylase (Fpg, MutM) is a bifunctional base excision repair enzyme (DNA glycosylase/AP lyase) that removes a wide range of oxidized purines, such as 8-oxoguanine and imidazole ring-opened purines, from oxidatively damaged DNA. The structure of a non-covalent complex between the *Lactococcus lactis* Fpg and a 1,3-propanediol (Pr) abasic site analogue-containing DNA has been solved. Through an asymmetric interaction along the damaged strand and the intercalation of the triad (M75/R109/F111), Fpg pushes out the Pr site from the DNA double helix, recognizing the cytosine opposite the lesion and inducing a 60° bend of the DNA. The specific recognition of this cytosine provides some structural basis for understanding the divergence between Fpg and its structural homologue endonuclease VIII towards their substrate specificities. In addition, the modelling of the 8-oxoguanine residue allows us to define an enzyme pocket that may accommodate the extrahelical oxidized base.

Keywords: AP site/base excision repair/crystal/Endo VIII/Fpg

Introduction

Reactive oxygen species represent a major source of spontaneous damage to DNA. DNA lesions interfere with both the efficiency and fidelity of DNA replication, inducing pathological cellular processes such as mutagenesis, carcinogenesis and ageing (Lindahl, 1993). The 7,8-dihydro-8-oxoguanine (8-oxoG) and imidazole ring-opened purines (Fapy) are among the most abundant oxidized purine products in DNA (Cadet *et al.*, 1997). To avoid the potentially mutagenic and lethal effects of these DNA lesions, prokaryote and eukaryote cells have devised specific DNA repair strategies (Lindahl, 1993).

The bacterial formamidopyrimidine-DNA glycosylase (Fpg or MutM) is one of the enzymes that initiates the base excision repair (BER) pathway. The *Escherichia coli* Fpg was identified initially as a DNA glycosylase that removes methylated Fapy-G residues in DNA (Chetsanga and

Lindahl, 1979) and it was characterized further as having the ability to excise a wide range of oxidized bases such as 8-oxoG (Tchou *et al.*, 1991; Duarte *et al.*, 2000). Fpg also displays AP lyase activity, promoting the successive DNA strand cleavages at both the 3'- and 5'-phosphodiester bonds of the abasic (AP) site by a $\beta\delta$ -elimination mechanism (O'Connor and Laval, 1989). Besides these main activities, Fpg is also endowed with a dRpase activity able to cleave a 5'-pre-incised AP site at the 3' side (Graves *et al.*, 1992). Therefore, Fpg belongs to the bifunctional DNA glycosylase family, distinguished from the monofunctional family by its ability to form a transient imino enzyme-DNA intermediate between the C1'-aldehyde of the AP site and the amino group of its N-terminal proline (Zharkov *et al.*, 1997).

The cloning of the *fpg* gene and the isolation of an *fpg*⁻ *E. coli* mutant have contributed significantly to the elucidation of the biological role of the enzyme (Boiteux *et al.*, 1987; Boiteux and Huisman, 1989; Michaels and Miller, 1992). The *fpg*⁻ mutant is unable to repair the lethal oxidative DNA damage induced by methylene blue plus visible light (Czeczot *et al.*, 1991) and is characterized by a spontaneous G→T mutator phenotype (Michaels *et al.*, 1991). Furthermore, G→T transversions are clearly associated with the formation of 8-oxoG residues in DNA (Tchou *et al.*, 1991). An important physiological function of Fpg is therefore to protect the cell from the mutagenic effects of 8-oxoG lesions in DNA.

The known sequences of Fpg display up to 35% identity, especially in the N-terminal extremity containing the catalytic P1 (PELPEVET motif) and in the C-terminal zinc finger motif required for DNA binding (Castaing *et al.*, 1993). Among DNA glycosylases, *E. coli* endonuclease VIII (Endo VIII) is the only one that displays sequence homology with Fpg (12% identity), although these two enzymes have distinct substrate specificities (Jiang *et al.*, 1997).

The crystal structure of the free Fpg protein from the extremophile *Thermus thermophilus* (TtFpg) was solved at 1.9 Å resolution (Sugahara *et al.*, 2000). TtFpg consists of two domains, the C-terminal domain including the zinc finger motif, and the H2TH motif related to the HhH motif of *E. coli* Endo III (Thayer *et al.*, 1995). The P1 of the active site emerges at the bottom of an electropositive cleft (Sugahara *et al.*, 2000). Recently, the crystal structure of the *E. coli* Endo VIII cross-linked to an AP site-containing DNA showed that the bound DNA is kinked sharply by enzyme residues that are intercalated at the lesion site (Zharkov *et al.*, 2002).

In previous work, we studied abortive complexes to investigate the structural and/or functional features of the target DNA required for DNA recognition by Fpg (Castaing *et al.*, 1999). The enzyme contacts 5–6 nucleotides (including the AP site) in the vicinity of the

Fig. 1. Two overviews of the Fpg–DNA complex. Secondary structures are colour-coded according to the definition of Sugahara *et al.* (2000). The N-terminal domain is coloured in dark blue, the four-helix bundle subdomain in red and orange, and the zinc finger subdomain in green. An additional secondary structure is represented in yellow. The zinc atom is in grey and the molecule of glycerol is represented by small cyan ball-and-stick. The figures were generated by Molscript (Kraulis *et al.*, 1991) and Raster3-D (Merritt and Murphy, 1994).

DNA damage and only the cytosine opposite the lesion (Castaing *et al.*, 1999). However, biochemical studies do not provide detailed information about the Fpg structural features involved in DNA binding.

Here, we report the 2.5 Å resolution crystal structure of a non-covalent complex between the *Lactococcus lactis* Fpg protein and the 1,3-propanediol (Pr) site-containing DNA duplex, a high affinity AP site analogue (Castaing *et al.*, 1999). This complex mimics an event which precedes the Schiff base formation and, subsequently, the AP lyase reaction.

Results and discussion

The mutant P1G-*Ll*Fpg was crystallized complexed to a 13mer DNA duplex containing the Pr abasic site analogue. As previously described, our choice of the catalytically defective mutant P1G-*Ll*Fpg was motivated by its higher stability towards proteolysis compared with the wild-type *Ll*Fpg or *Ec*Fpg (Pereira de Jésus *et al.*, 2002; B.Castaing, unpublished data). In addition, this mutant binds to DNA with the same affinity as does the wild-type protein and probably retains all structural determinants for recognition of the DNA lesion.

The crystals contain two essentially identical Fpg–DNA complexes per asymmetric unit (r.m.s.d. = 0.19 Å on 260 C α using the program O, and the LSQ option; Jones *et al.*, 1991). Only a few small differences, which may arise from crystal packing, are located in some protein loops. For the sake of clarity, we will focus our analysis on only one of the complexes.

Overall structure of the Fpg–DNA complex

The global structure of the *Ll*Fpg–DNA complex is shown in Figure 1. *Ll*Fpg consists of two globular domains: the N-terminal domain composed of an antiparallel β -sandwich core (β -strands 1–8 and the α -helices A and B) and the C-terminal domain constituted by two subdomains: an α -helix-rich domain (α -helices C–E) containing the H2TH motif (α -helices D and E) and the zinc finger domain (β 9– β 10 hairpin). Between the free *Tl*Fpg (Sugahara *et al.*, 2000) and the bound *Ll*Fpg, the r.m.s.d. calculated on 230 C α atoms is ~ 1.25 Å (Figure 2). If we dock the DNA of the complex with the free *Tl*Fpg, the fit is convincing without many protein rearrangements (Figure 3). The main atom displacements between the free and bound structures are clustered: (i) in the vicinity of the damaged site; (ii) in the conserved loop 256–261 of the zinc finger motif, shifted by 3 Å towards the surface in the liganded enzyme; and (iii) in the variable interdomain region 117–131 (Figure 2). This last region exhibits an α -helical conformation in the *Ll*Fpg–DNA structure (in yellow, Figure 1) and a random coiled structure in the free *Tl*Fpg. This local deviation could be related to sequence differences rather than to the DNA binding. Thus, this DNA–protein complex would be achieved without any special hinge effect between the two globular domains of Fpg (Figure 2). The DNA binding requires both globular domains of Fpg, the relative orientations of which being constrained by an internal hydrogen bond network (Sugahara *et al.*, 2000). In addition to the two C-terminal subdomains, i.e. the zinc finger and the H2TH (in green and orange, respectively, Figure 1) (O'Connor *et al.*, 1993; Sugahara *et al.*, 2000),

Fig. 2. Through DNA binding, the global structure of Fpg is not remodelled. Stereo views of the C_{α} backbone superposition of the free *TtFpg* monomers (in purple and cyan) (Sugahara *et al.*, 2000) and the bound *Lfpg* (in yellow). The arrow in magenta indicates the missing loop, and arrows in cyan indicate some rearrangements in the N-terminal domain and in the zinc finger motif. The figures were generated by Molscript (Kraulis *et al.*, 1991) and Raster3-D (Merritt and Murphy, 1994).

four loops of the N-terminal domain participate in the DNA binding (in blue, Figure 1). The *Lfpg* residues interacting with DNA nucleotides are K57, Y58, H72, R74, M75, H91, R109, K110, F111 and T113 located in the N-terminal domain, L161, Q163, G170 and N171 in the H2TH motif, and K254, K256 and R260 in the zinc finger motif (for details see below).

An earlier description of the Fpg surface pointed out a large positive patch able to neutralize the negatively charged phosphodiester backbone of DNA (Sugahara *et al.*, 2000). Actually, the 13mer Pr/C duplex fits nicely into the positively charged surface formed by the two Fpg globular domains, indicating that the recognition of DNA by the enzyme would be initiated by electrostatic interactions (Figure 3). This classical strategy has not been selected by the glycosylases of the Endo III superfamily, such as the human Ogg1 (Bruner *et al.*, 2000), *E.coli* AlkA (Hollis *et al.*, 2000) or human UDG (Parikh *et al.*, 1998), which used dipolar interactions. A positively charged DNA-binding cleft has also been described for *E.coli* Endo VIII, the Fpg structural homologue (Zharkov *et al.*, 2002).

The formation of the *Lfpg*-DNA complex leads to a large buried but accessible surface of 2000 Å² (calculated with GRASP), comparable to that of the hOGG1-DNA complex (2268 Å²) (Bruner *et al.*, 2000), but larger than those of the hANPG-DNA (1043 Å²) (Lau *et al.*, 1998), hUDG-DNA (750 Å²) (Parikh *et al.*, 1998) and *E.coli* AlkA-DNA (600 Å²) complexes (Hollis *et al.*, 2000). Interestingly, the crystal structure of *E.coli* Endo VIII covalently bound to an abasic site-containing DNA displays an interaction surface of only 860 Å² (Zharkov *et al.*, 2002).

Whereas there is no significant domain rearrangement between the free *TtFpg* and the bound *Lfpg* (this work),

Fig. 3. The Fpg-DNA complex occurs through electrostatic interactions. The distorted DNA fits nicely with the electropositive surface of the liganded *Lfpg* and of the free *TtFpg*. Each accessibility surface was created using GRASP (Nicholls *et al.*, 1993).

Fig. 4. Schematic representation of Fpg–DNA contacts. Amino acid residues are painted according to the colour code used in Figure 1. Asterisks mark the residues conserved between the Fpg and Endo VIII sequences. Underlined residues represent residues conserved among the Fpg sequences only. The prefix ‘mc’ stands for main chain atoms. Small circles represent the water-mediated interactions. The Pr site and its opposite cytosine (C20) are indicated in pink. The DNA backbone phosphates contacted by the enzyme are in yellow.

there are, however, slight differences in the relative orientations of their two domains between the free *Ti*Fpg and the covalently bound Endo VIII (Zharkov *et al.*, 2002). These global structural differences between Fpg and Endo VIII and the differences in their interaction surfaces with DNA may reflect the substrate specificity of each protein. However, it cannot be excluded that these differences in DNA binding are related to the type of protein–DNA complex that we compare. The non-covalent complex that we describe with *Ll*Fpg is prior to the formation of the Schiff base intermediate (see Introduction), which is mimicked by the covalent reduced Schiff base of the Endo VIII–DNA complex. Thus, the differences between Fpg and Endo VIII complexed with DNA could also result from the dynamics of the enzyme–DNA interface during the reaction.

DNA distortion: global description

The most remarkable feature of this complex is the sharp kink of the DNA towards the major groove at the Pr site level, with a distortion angle $\sim 60^\circ$ (Figure 1) (CURVES; Lavery and Sklenar, 1989). This DNA distortion results from extensive contacts with the lesion-containing strand distributed around the damaged site (from p^1 to p^{-3}) and involving essentially the two phosphates flanking the Pr site (p^0 and p^{-1} , respectively). Conversely, the Pr moiety does not make close contacts with the enzyme residues (Figure 4).

At the recognition site, the local DNA parameters are strongly altered with respect to those of B-DNA. The Pr site displays an extrahelical conformation resulting from rotations around the p^0 –O5' bond ($\alpha = -89^\circ$), the C5'–C4' bond ($\gamma = -52^\circ$) and the O3'– p^{-1} bond ($\zeta = -62^\circ$). Furthermore, the phosphate–phosphate distances clearly indicate a compression of the DNA at the lesion site (6.7 Å in the standard B-DNA, against 5.4–6.0 Å for nucleotides 6–9 of the lesion-containing strand, and 5.5–5.8 Å for nucleotides 18–20 of the complementary strand). The DNA bend in the Fpg–DNA complex is larger than that observed with human UDG or ANPG ($\sim 45^\circ$ and 22° , respectively; Lau *et al.*, 1998; Parikh *et al.*, 1998), but comparable to that observed with the human Ogg1 ($\sim 70^\circ$; Bruner *et al.*, 2000; Norman *et al.*, 2001), *E. coli* AlkA ($\sim 66^\circ$; Hollis *et al.*, 2000) and T4 Endo V ($\sim 60^\circ$; Vassilyev *et al.*, 1995). In the covalent complex of Endo VIII–DNA, the bend is only 45° and it seems that Endo VIII does not use a backbone compression mechanism as do Fpg and hOgg1 (Zharkov *et al.*, 2002).

Such a distortion in the free DNA containing an AP site has never been observed in either the solution or the crystal structures (Kalnik *et al.*, 1989; Cuniassé *et al.*, 1990; Goljer *et al.*, 1995). Outside the immediate vicinity of the lesion, the conformational parameters of the DNA are close to those of the B-form (CURVES; Lavery and Sklenar, 1989). In the complex, the unusual distortion of the helical axis of the Pr-containing DNA appears as a result of its binding to Fpg in the absence of the unexpected nucleotide sequence and crystal packing effects.

Fpg structural determinants for damage-specific recognition

The specificity of the recognition is centred on the Pr-containing strand (nucleotides 5–9) and on the estranged cytosine (C20) (Figure 4). This observation correlates well with the footprinting experiments done with *Ec*Fpg and *Ll*Fpg bound to a 59mer DNA containing a unique Pr site within the same sequence context as that of the 13mer duplex used for crystallization (Castaing *et al.*, 1999).

The Fpg residues in contact with the DNA can be classified into two categories: (i) those strictly conserved within the Fpg superfamily (including Endo VIII), i.e. K57, H72, Q163, G170, N171 and R260, and less strictly, K254, probably related to the general binding mode and catalytic mechanism of this protein family; and (ii) those only conserved among the ‘true’ Fpgs, i.e. M75, H91, R109, F111 and L161. We may also add to this second category Y58 and K110, from the N-terminal domain, possibly related to the substrate specificity of Fpg (Figures 4 and 5).

A detailed view of the interactions at the Pr site is presented in Figure 6. Among the zinc finger residues that are in contact with the DNA, the conserved R260 plays a key role in stabilizing the Pr site in an extruded conformation. The R260 guanidinium group links the p^{-1} and p^0 phosphates by two hydrogen bonds. This double interaction would contrast with the electrostatic repulsion between the phosphates adjacent to the lesion. Interestingly, in the free enzyme, this guanidinium group (R253 in *Ti*Fpg) interacts with the side chain of N161 (N171 in *Ll*Fpg), another conserved residue of the H2TH

Fig. 5. Primary sequence of *L.lactis* Fpg and alignment with *E.coli* Fpg, *T.thermophilus* Fpg and *E.coli* Endo VIII. Residues that are identical in all the sequences are shown in white on a red background. Homologies are highlighted in red. Arrows and rectangles schematize β -strands and α -helices of the *Ll*Fpg structure, respectively. The colour code follows that of Figure 1. Alignment was performed by CLUSTALW (Thompson *et al.*, 1994) and is represented with ESPrit (Gouet *et al.*, 1999).

Fig. 6. Stereo view of Fpg contacts around the Pr abasic site analogue. Hydrogen bonds are indicated by dashed lines. DNA atoms are represented by orange ball-and-sticks, and mutagenesis targeted amino acids are underlined (see text for details). The figures were generated by Molscript (Kraulis *et al.*, 1991) and Raster3-D (Merritt and Murphy, 1994).

subdomain (Figure 5). In the Fpg–DNA complex, the N171:R260 contact is absent and the N171 side chain has swung towards the p^0 phosphate and K57 (Figure 6).

The conformation of the extruded Pr site and the recognition of the estranged C20 are related to the intercalation of three key residues of the N-terminal domain (M75, R109 and F111) inside the hole created by the exclusion of the Pr site (Figure 7). These residues are strictly invariant within ‘true’ Fpgs. In *E.coli* Endo VIII, Q69, L70 and Y71 play an equivalent intercalating role in DNA (Zharkov *et al.*, 2002). The major differences between the free (*Tl*Fpg) and bound enzymes (*Ll*Fpg–

DNA) are observed for the side chains of these three residues (Figures 7 and 8). The aromatic ring of F111 rotates by 90° and wedges its aryl ring into the space between C20 and A19; the R109 guanidinium group moves closer to the F111 ring and interacts with C20; and, in addition, the S-CH₃ group of M75 is pulled away, in front of F111 (Figure 8). The F111–M75 residue pair is stacked in the minor groove between the A19:T8 base pair and the ‘base–amino acid’ pair C20–R109 (Figure 8). Calculation of the molecular surface makes it even more obvious that these small rearrangements enable the enzyme structure to clamp tightly around the lesion and

maintain the bent conformation of the DNA. The intercalation of these three residues occurs by the minor groove side, shifting C20 slightly towards the major groove (Figure 8). They fill the space left by the absence of the base lesion opposite C20 (Figure 7). Intercalation of

Fig. 7. The intercalation of the Fpg triad fills the void between the expelled Pr site and the estranged C20. The accessibility surfaces clearly show that in the free enzyme (*Tt*Fpg) (A), the rearrangement of the conserved triad (M70/R99/F101) is required to accommodate the damaged DNA as in the *Lt*Fpg–DNA complex (B) (GRASP; Nicholls *et al.*, 1993).

aromatic or aliphatic amino acid residues in DNA duplexes has been observed for several proteins involved in base excision repair (Schärer and Jiricny, 2001). Thus, *E.coli* AlkA and human ANPG intercalate their L125 and Y162, respectively, between the base pairs flanking the flipped-out AP site ring (Lau *et al.*, 1998; Hollis *et al.*, 2000). hOGG1 also intercalates the aromatic residue Y203, in addition to N149 (Bruner *et al.*, 2000). On this structural basis, we propose that the intercalation of F111 and M75 in DNA by Fpg contributes to the sharp DNA bend. It pushes the AP site out of the DNA helix while R109 stabilizes the local DNA conformation by partially reconstituting a pseudo-Watson–Crick interaction with C20 (Figure 8). Another consequence of the R109–C20 interaction is that it keeps C20 intrahelical, thus avoiding the collapse of the DNA helix at the lesion site.

Recognition of the pyrimidine opposite the AP site

The R109–C20 interaction involves a weak hydrogen bond between the N η 2 of R109 and the N3 of C20 (distance 3.37 Å) and an essential hydrogen bond between the N ϵ of R109 and the O2 of C20 (distance 3.18 Å) (Figure 8). A similar hydrogen bond could occur between the N ϵ group of R109 and the carbonyl group of a thymine (T) as the opposite base, without affecting the environment around the lesion. These structural observations give new insights into the Fpg preference for a pyrimidine opposite an 8-oxoG or an AP site (Castaing *et al.*, 1993; Tchou *et al.*, 1994). As already demonstrated by biochemical studies, this structure predicted that the worst candidate for the opposite base would be a purine, since standard purines do not display a carbonyl group at position 2 on their Watson–Crick faces. Besides, replacing the estranged cytosine by a purine would induce a steric hindrance and prevent the proper intercalation of the Fpg triad even in the case of an AP site. Thus, we propose that a pyrimidine opposite the damaged base is a DNA structural determinant for the Fpg-specific recognition by contributing to the optimal conformation of the substrate. This kind of discrimination of the base opposite the lesion is not a general feature of glycosylases, although some of them adopt similar stringent strategies. hOgg1 exclusively specifies a cytosine opposite the lesion (Guibourt *et al.*, 2000), establishing five hydrogen bonds between the

Fig. 8. Recognition of the C20 opposite the Pr site by R109. Stereo view showing the intercalation of the Fpg triad by the minor groove and the pseudo-Watson–Crick interactions between R109 and C20. Hydrogen bonds are indicated by dashed lines. The atomic coordinates of the triad (M70/R99/F101) from the free *Tt*Fpg have been superposed and are represented by green ball-and-sticks. The figure was generated by Molscript (Kraulis *et al.*, 1991) and Raster3-D (Merritt and Murphy, 1994).

Fig. 9. Stereo view of the model of the 8-oxoG in *LIFpg* and *TIFpg*. The C_{α} backbones of *LIFpg* and *TIFpg* are represented in yellow and cyan, respectively. The lesion-containing strand is in red. 8-OxoG is modelled in an *anti* (red) or *syn* (green) conformation. The N-terminal amine of G1 (*LIFpg*) or P1 (*TIFpg*) is indicated by a yellow or cyan arrow, respectively. The missing loop of *LIFpg* is in the vicinity of the oxidized base. The intercalation residues in *LIFpg* are also indicated in the figure. L209 and Y214 belong to the *TIFpg* structure and are discussed in the text. The figures were generated by Molscript (Kraulis *et al.*, 1991) and Raster3-D (Merritt and Murphy, 1994).

enzyme residues (R154, R204 and R149) and the estranged cytosine (Bruner *et al.*, 2000). These strong interactions are strictly specific for a cytosine. We can also notice the case of the G:T/U mismatch-specific glycosylase (Mug) that discriminates a guanine opposite T or U (Schärer *et al.*, 1997). Mug establishes three hydrogen bonds with the widowed guanine defining the specific complementary strand interactions (Barrett *et al.*, 1998).

The substrate specificity of Fpg includes the specific recognition of the lesion and, as we have just shown, the recognition of the estranged base. Interestingly, its structural homologue Endo VIII does not have the same substrate specificity. This enzyme was first identified in *E.coli* as an Endo III-like enzyme that recognizes and removes oxidized pyrimidines (Jiang *et al.*, 1997). Thus, Endo VIII must accommodate, in its active site, a purine opposite the oxidized pyrimidine. This constraint implies the selection of a complementary strand discriminatory system, different from that of Fpg. In the crystal structure of Endo VIII bound covalently to DNA, the intercalated triad (Q69, L70, Y71) is formed by consecutive residues (Zharkov *et al.*, 2002). Conversely, in the Fpg–DNA complex, the intercalation involves two different regions of the N-terminal domain. Zharkov *et al.* (2002) mentioned that the opposite adenine is stabilized by a single weak hydrogen bond between the N3 and Q69. Therefore, in Endo VIII, there is no such discriminating element equivalent to the Fpg R109. In Endo VIII, any base opposite the oxidized pyrimidine is permitted without leading to unfavourable contacts with the protein. Recent work supporting this hypothesis (Hazra *et al.*, 2000) showed that the wild-type Endo VIII is significantly active in excising 8-oxoG paired with C, G or A, relative to the more canonical substrate dihydrouracil. Furthermore, quantitative analysis demonstrated that Endo VIII prefers a purine opposite 8-oxoG. As a matter of fact, a comparison of Endo VIII and Fpg models clearly illustrates the importance of the nature of the base opposite the lesion. In bacteria, the physiological significance of a second 8-oxoG–DNA glycosylase that prefers

removing 8-oxoG paired with G or A suggests that Endo VIII constitutes a fourth element of the GO system, already including Fpg, MutY and MutT (Michaels and Miller, 1992). As proposed by Hazra *et al.* (2000), the couple Fpg–Endo VIII would be the functional homologue of yOgg1–Ntg1(Ogg2) (Bruner *et al.*, 1998).

A model for base recognition

Recent evidence indicates that various DNA repair enzymes, including DNA methyl-transferases and DNA glycosylases, have devised an outstanding molecular mechanism to recognize and process their substrates (Roberts, 1995). This general mechanism suggests that after specific recognition inside the DNA, the altered base is flipped out of the DNA helix, stabilized inside an enzyme pocket and finally is processed (Schärer and Jiricny, 2001). In the Fpg–DNA crystals, the Pr site is extruded and makes van der Waals contacts with the strictly conserved residues G1 (or P1 for the wild-type Fpg), M75, I172, Y238 and one water molecule, w41 (Figure 6). Additional atoms can probably be identified in the binding site, as related to putative further interactions involved in the recognition of removable substrates (8-oxoG, Fapy or true AP sites). Using the model of 8-oxoG in a complex with hOgg1 (Bruner *et al.*, 2000), we superposed the deoxyribose linked to the oxidized base with the backbone (p^0 -O-C5'-C4'-C3'-O- p^{-1}) of the Pr site and we modelled the 8-oxoG in an *anti* conformation in the Fpg-binding pocket, according to the hOgg1 model (in red, Figure 9). Nevertheless, the size and the nature of this pocket also allow the 8-oxoG to be fit in a *syn* conformation (in green, Figure 9) requiring fewer rearrangements of the surrounding residue side chains. In *LIFpg*, this binding pocket would be lined by residues 1–5, 172–173 and 206–227, including the missing loop (I219–S227) (Figures 1 and 9) and a spontaneous cleavage site (B.Castaing, unpublished data). Previous electrophoretic analysis of the crystal contents showed that the crystallized enzyme was intact and, consequently, the disorder of the loop I219–S227 was related to intrinsic dynamics of this

region rather than to a spontaneous cleavage. These intrinsic dynamics could be explained by the presence of four glycine residues (G215, G216, G226 and G229) conserved among the Fpg sequences (Figure 5). This flexibility could allow the accommodation of various ligands of the enzyme. Interestingly, the homologous region (215–222) in the *E.coli* Endo VIII–DNA model is also missing (Zharkov *et al.*, 2002) and its sequence diverges significantly from those of the ‘true’ Fpgs, suggesting the implication of this loop in the substrate specificity of Fpg and Endo VIII. Among the Fpg sequences, few residues of this region are conserved, but two positions are always occupied by small aliphatic and aromatic residues, respectively (I219 and Y222 in *Ll*Fpg, L209 and Y214 in *Tl*Fpg, and L216 and F219 in *Ec*Fpg; Figure 5). The model of the free *Tl*Fpg provides a better resolved view of the active pocket region since the complete polypeptide chain was entirely built (Sugahara *et al.*, 2000). In the *Tl*Fpg structure superposed with the DNA of *Ll*Fpg–DNA, the modelled oxidized base would be stabilized between the aryl ring of Y214 (Y222 in *Ll*Fpg) and the aliphatic side chain of L209 (I219 in *Ll*Fpg) (Figure 9). Stabilization of the flipped out base by aromatic and hydrophobic side chains in the enzyme-binding pocket is common to glycosylases (Schärer and Jiricny, 2001).

Some proton donor/acceptor groups present inside the Fpg-binding pocket may also play a role in the specific recognition of the substrates, as illustrated in the hOgg1–DNA model, by the specific hydrogen bond between the carbonyl group of G42 and the N7 of 8-oxoG. Moreover, in hOgg1, the architecture of the lesion-binding pocket could be rearranged significantly by the binding of the substrate (Bruner *et al.*, 2000; Norman *et al.*, 2001). The size of the base recognition pocket of hOgg1 is greater when the enzyme is bound to an AP site than to an 8-oxoG (Norman *et al.*, 2001). In the structure of *Ll*Fpg–DNA, the flexible region 219–227 could correspond to a relaxed state of the active site pocket in the absence of the oxidized base.

Structural interpretations of the deleterious effects of the Fpg mutations

The first residues suspected to play a catalytic role are the six N-terminal residues (PELPEV) strictly conserved in the Fpg family and, more especially, P1, E2, P4 and E5, which are also conserved in Endo VIII (PEGPEI) (Figure 5) (Lavrukhin and Lloyd, 2000; Sidorkina and Laval, 2000; Burgess *et al.*, 2002; Zharkov *et al.*, 2002). Evidence for a covalent Schiff base intermediate mediated by the amine of P1 allowed us to locate, unambiguously, the N-terminal residues in the enzyme active site (Zharkov *et al.*, 1997). We solved the structure of the P1G-*Ll*Fpg bound to an AP site-containing DNA. Although P1G-*Ll*Fpg binds to DNA specifically with the same affinity as the wild-type enzyme (Table I), this mutant is strongly affected in catalysis, especially in the $\beta\delta$ -elimination process (cleavage assay, Figure 10A). The P1G-*Ll*Fpg mutant is still able to form a covalent Schiff base intermediate with the resulting AP site (trapping assay, Figure 10A). Nevertheless, the wild-type *Ll*Fpg cleaves 8-oxoG-containing DNA, 700- and 2000-fold more quickly than does the P1G-*Ll*Fpg mutant, at the 3' side and successively at the 3' and 5' sides of the resulting AP

Table I. Apparent dissociation constant (K_{Dapp}) of wild-type and mutant P1G-*Ll*Fpg for 1,3-propanediol (Pr)- and 8-oxoG (GO)-containing 13mer DNA duplexes

DNA probe	K_{Dapp} (nM) ^a	
	Wild-type <i>Ll</i> Fpg	P1G- <i>Ll</i> Fpg
13mer GO/C	6	3
13mer Pr/C	2	2

^aDetermined by EMSA as described by Castaing *et al.* (1993, 1999).

Fig. 10. Cleavage of 8-oxoguanine-containing DNA by wild-type and mutant P1G-*Ll*Fpg. (A) Qualitative analysis of the Fpg GO-DNA glycosylase/lyase. 5'-[³²P]-labelled 13mer GO/C-DNA was incubated for 20 min at 37°C alone (lane 1), or in the presence of 0.2 and 2 μ M of wild-type and P1G-*Ll*Fpg (lanes 2 and 3), respectively. DNA incubation products were analysed by 7 M urea/20% SDS-PAGE ('13mer GO cleavage assay'). The same samples were incubated in the presence of 0.1 M NaBH₄ and analysed by SDS-PAGE ('13mer GO trapping assay'). The 13mer GO was for the intact DNA substrate, and 6merAP (1) and 6mer (2) for the β - and $\beta\delta$ -elimination products, respectively. (B) Time course of the 13mer GO/C cleavage. The DNA duplex was incubated as described in (A) in the presence of the limited protein concentrations indicated. At various times, samples of each incubation mixture were analysed by 7 M urea/20% SDS-PAGE. After gel autoradiography, quantifications were performed as described by Castaing *et al.* (1993).

site, respectively (Figure 10B). Both binding and cleavage experiments suggest that P1 of Fpg is required for catalysis optimization but not for specific DNA recognition. In the Fpg–DNA structure, the C1' of the modelled 8-oxoG (Figure 9) would be positioned at 3.1 Å from the amino group of P1 in the complex modelled with *Tl*Fpg and at 2.4 Å from that of G1 in the P1G-*Ll*Fpg (cyan and

yellow arrows, respectively, Figure 9). Superposition of the wild-type and mutant structures does not give any hints relative to their activity differences. The catalytic amino groups of P1 and G1 are very close (Figure 9). Whereas proline is the unique amino acid with a secondary amine, glycine is the unique amino acid without an asymmetrical C α and which displays a wide spectrum of Φ and Ψ angles. We suppose that, in P1G-Fpg, the free rotation around the N–C α bond of the N-terminal amino group is partly responsible for the loss of activity. Statistically, the amino group of G1 would not be orientated properly for catalysis, which may explain why this mutant is altered in the rate of reactions but not in the substrate recognition.

Because acidic residues were suspected to play a key role in the acid–base mechanism of glycosylase (Schärer and Jiricny, 2001), they were also targeted systematically by mutagenesis (Lavrukhin and Lloyd, 2000). Among these, the *E.coli* E173Q and E2Q mutants (E176 and E2 in *Lfpg*) exhibit a severe deficiency in their glycosylase activity. In the structure, E176 (E173 of *EcFpg*) contributes towards maintaining the hinge formed by residues 235–238 in the vicinity of the active site. Indeed, this region is functionally important since Y238 interacts with the phosphate p⁰ at the 5' side of the lesion (Figure 6). Y238 and V237 are also located in the vicinity of the hydrophobic side of the putative oxidized base pocket (Figure 9). It is therefore plausible that disturbing this region would affect the binding of the damaged base and consequently the glycosylase activity. However, the special geometry of this region is not essential for the AP lyase activity (Lavrukhin and Lloyd, 2000). On the other hand, E2 does not interact with the phosphate backbone of DNA, but its side chain points towards the active site (Figure 6). The carboxylate oxygen can make two hydrogen bonds with the backbone amides of the conserved G170 and Y173, and two other hydrogen bonds with the backbone amide of the conserved I172 and one water molecule (w41, Figure 6). The introduction of two-proton donor groups (E2Q and E176Q mutants) in this environment would therefore certainly disrupt the local hydrogen bond network and consequently disturb the optimal orientation of the oxidized purine in its binding pocket. In addition, the recent crystal structure of a trapped complex Endo VIII–DNA (Zharkov *et al.*, 2002) showed that E2 and a water molecule make two hydrogen bonds with the C4'-OH of the ring-opened AP site covalently bound to the enzyme, suggesting that E2 is essential in stabilizing the structural conformation of the substrate during the catalytic process. This observation explains why the non-cyclic reduced AP (redAP) site analogue is the best ligand of Fpg (Castaing *et al.*, 1993) and confirms our previous hypothesis about Fpg recognition of the C4'-OH of the redAP site analogue (Castaing *et al.*, 1999). Surprisingly, only the glycosylase activity of the E2 mutants in Fpg and Endo VIII is strongly affected, suggesting that the recognition of the C4'-OH is not essential for the AP lyase process and that the opening of the AP site ring occurs before Schiff base formation and, consequently, before the glycosylase process (Lavrukhin and Lloyd, 2000; Burgess *et al.*, 2002; Zharkov *et al.*, 2002).

Targeted mutagenesis on the conserved basic residues of the *E.coli* enzyme has also been performed (Rabow and Kow, 1997; Sidorkina and Laval, 1998). Among these

mutations, K157A and K57G exhibited dramatically reduced 8-oxoG binding and 8-oxoG glycosylase activity. These two conserved lysines are located on both sides of the active site pocket of Fpg. Although K157 was proposed for a direct interaction with the O8 atom of 8-oxoG, it is difficult to suppose that it could interact directly with the damaged guanine. K157 is located far from the modelled 8-oxoG and makes hydrogen bonds with the N171 (O δ 1) of the H2TH motif, the G261 (O) and the T262 (O γ 1) of the zinc finger interacting with the phosphodiester backbone around the lesion. In our model, K57 interacts with the backbone carbonyl of L169 and with two adjacent phosphate groups (p⁻¹ and p⁻²) at the 3' side of the AP site (Figures 4 and 6). The role of K57 rather would be to help maintain the bend of the DNA in the lesion region. Comparison with the free enzyme shows that DNA binding does not affect the conformations of these two lysines. K157 and K57 therefore play a key role in maintaining the architecture of the DNA and that of the enzyme since they interconnect the different subdomains of Fpg (Sugahara *et al.*, 2000). Therefore, changing these charged residues to alanine or glycine certainly affects the position of the substrate as well as the relative positions of the Fpg domains, inducing critical changes in the binding and, consequently, in the catalysis.

Conclusion

The crystal structure of Fpg bound to a Pr-containing DNA highlights the Fpg and DNA structural features required to form a stable and non-covalent specific complex prior to the Schiff base formation. Based on biochemical and structural data, the DNA binding strategy selected by Fpg, taking place before the formation of the imino enzyme–DNA intermediate, can be subdivided into four steps: (i) through a non-specific complex, Fpg slides along the DNA examining the major groove; (ii) by an unknown mechanism, Fpg recognizes specifically in DNA the base modification emerging from the major groove; (iii) Fpg flips out the damaged base by the major groove, intercalating M75, R109 and F111 via the minor groove, while the R109 stabilizes the opposite cytosine in an intrahelical conformation; and, finally, (iv) Fpg locks the damaged base inside its active site and processes it. All these steps are essential for substrate recognition by Fpg. According to this Fpg binding scheme, the structure of the non-covalent complex that we describe here elucidates the third step, proposes a model for the fourth step and suggests a molecular mechanism for understanding the divergence between Fpg and its structural homologue Endo VIII. However, access to the complete scenario of the base excision reaction by Fpg (concerning especially the base recognition and catalytic process) requires further high-resolution structures in which the enzyme would be associated with other substrates or substrate analogues and transitory or final products.

Materials and methods

DNA and proteins

The wild-type *Lfpg* and the P1G-*Lfpg* mutant were purified as previously described (Pereira de Jésus *et al.*, 2002). The selenomethionylated (Se) P1G-*Lfpg* was prepared according to the procedure described by Van Duyne *et al.* (1993), except that immediately prior to

Table II. Data collection

Wavelength (Å)	0.9797	0.9798	0.9778
Resolution (Å)	2.55	2.54	2.54
Total observations	67 360	68 563	69 184
Unique reflections	21 902	21 938	22 140
Data completeness (%)	98.6	98.6	98.5
Average redundancy	3.1	3.1	3.1
Mean $I/\sigma(I)$	6.8	6.8	5.8
R_{sym}^a	0.052	0.053	0.056
Last resolution shell R_{sym}	0.156	0.190	0.210
R_{ano}^b	0.059	0.041	0.047

^a $R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I \rangle| / \sum_h \sum_i I_{h,i}$; where I_i is the intensity of the i th observation and $\langle I \rangle$ the mean intensity of the reflection.

^b R_{ano} is the measure of the mean relative anomalous difference between the Bijvoet pairs defined as: $\sum |I(+) - I(-)| / \sum [I(+) + I(-)]$.

the isopropyl-β-D-thiogalactopyranoside (IPTG) induction, 0.01 mM of zinc acetate was added. The (Se)PIG-*L*/Fpg was purified as the native enzyme (Pereira de Jésus *et al.*, 2002), concentrated to 6–10 mg/ml and stored at –80°C in 10 mM HEPES/NaOH pH 7.6, 100 mM NaCl, 5% glycerol, 1 mM TCEP, 0.1 mM phenylmethylsulfonyl fluoride (PMSF).

All the 13mer single-stranded oligonucleotides were purchased from Eurogentec (Belgium), except for the oligomer containing the GO lesion given by Drs D.Gasparutto and J.Cadet (CEA, Grenoble): CTCTTTXTTTCTC (13mer X) and GAGAAACAAAGAG (13mer C), with X = G (guanine), GO (8-oxoguanine) or Pr. After purification, 13mer X was annealed with the complementary strand, the 13mer C, to generate 13mer X/C duplexes. For glycosylase/lyase assays, trapping assays and electrophoretic mobility shift assays (EMSAs), the 13mer X was 5′-³²P-labeled before annealing as previously described (Castaing *et al.*, 1993, 1999).

Crystallization

The DNA–protein complex was prepared by mixing the 13mer Pr/C duplex in a 1.3 molar excess with (Se)PIG-*L*/Fpg. Crystals were obtained at room temperature using the hanging drop method under the following crystallization conditions: 28% PEG 4000, 0.2 M LiSO₄, 0.1 M Tris–HCl pH 7.7, 1 mM [Tris (carboxyethyl) phosphine] TCEP, and 2 mM spermidine. Small plate-shaped crystals appeared after a few days and were used for growing larger crystals by micro-seeding. These crystals are monoclinic and belong to the $P2_1$ space group (with unit cell dimensions of $a = 69.88$ Å, $b = 62.03$ Å, $c = 80.97$ Å, $\beta = 104.7^\circ$, assuming two molecules of the complex per asymmetric unit).

Data collection

Among these crystals, a few grew as single crystals, and selection of the good candidates depended on the quality of the diffraction pattern obtained on our laboratory rotating anode. Before exposure to X-rays, the crystals were soaked in 20% glycerol, 28% PEG 4000, 0.2 M LiSO₄, 0.1 M Tris–HCl pH 7.7, mounted on a cryo-loop and finally frozen directly in liquid nitrogen. High resolution experiments were conducted at 100 K and at three wavelengths ($\lambda = 0.9798$, 0.9796 and 0.9778 Å) using synchrotron radiation (Beamline BM30 at ESRF) and a MarCCD detector at 100 K (details on data collection are presented in Table II). Each data set was integrated with DENZO (Otwinowski and Minor, 1997) and reduced by SCALA (Evans, 1993). Scaling between the different data sets was performed according to the protocol written by Phil Evans, i.e. scaling factors were calculated previously from the three merged wavelength data, then applied to each data set. Conversion of I to F was performed with TRUNCATE, and each wavelength was scaled to every other by SCALEIT. Other calculations were carried out with the CCP4 suite of programs (Collaborative Computational Project 4, 1994).

Phasing

Anomalous signal. Inspection of the anomalous difference Patterson maps did not clearly show peaks corresponding to the expected eight selenium peaks. However, analysis of the anomalous signal at the selenium peak absorption ($\lambda = 0.9796$ Å) by SnB (Weeks *et al.*, 1994) led to a solution for eight selenium atoms. Using these positions, we calculated multiple amorphous displacement (MAD) phases with SOLVE (Terwilliger and Berendzen, 1999) to a 2.5 Å resolution and applied the electron density modifications implemented in DM (Cowtan and Main, 1993). However, it

was still not obvious to build a polypeptide chain in the resulting electron density map.

Molecular replacement. *Thermus thermophilus* Fpg (PDB code: 1EE8) was used as a research model, and data collected at the selenium anomalous peak were selected for the molecular replacement. MOLREP (Vagin and Teplyakov, 2000) gave a possible solution for two molecules of *Tt*Fpg per asymmetric unit (R -factor = 0.56, correlation = 0.25 for data between 25 and 3.0 Å). We calculated a set of phases from this molecular replacement solution and used these phases to calculate a difference anomalous Fourier map. This map clearly showed eight peaks located inside the two molecules. Comparison of the *T.thermophilus* and *L.lactis* Fpg primary sequences indicated that the location of these eight peaks was fitting the position of the expected methionines in the *L.lactis* enzyme and consequently that the molecular replacement solution was correct. Moreover, the eight selenium positions deduced from the molecular replacement solution were also correlated to the atomic positions given by SnB.

Building of the model. The positions of the selenium atoms were used to calculate MAD phases with SOLVE (figure of merit: 0.43 for all the data). Both sources of phases (MAD and molecular replacement) were combined with SIGMAA (Read, 1986) until 2.5 Å resolution. The quality of the combined phases was improved using the options implemented in DM (density averaging, solvent flattening and histogram matching). At this stage, the resulting map clearly showed the density of the missing DNA for each molecule of Fpg. Building of the whole 13mer Pr/C and of the polypeptide chains was performed on the SGI-octane with the O program (Jones *et al.*, 1991).

Refinement

The refinement of the structure was performed by CNS (Brünger *et al.*, 1998) with all the data between 25 and 2.5 Å. Ten per cent of the data were selected randomly and excluded from the refinement procedure (Brünger, 1992). It was possible to build most of the polypeptide chain using the MAD/molecular replacement map. This first model was then submitted to a whole cycle of simulating annealing (at 2500 K) followed by some minimization energy cycles, considering strong non-crystallographic symmetry (NCS) restraints between the two Fpg molecules (R -factor = 0.347 and $R_{\text{free}} = 0.395$). One zinc atom was then added to each enzyme and the whole DNA was built step by step and also included in the NCS restraints. The duplexes were refined by simple energy minimization followed by B-group refinement (two B-values per residue or nucleotide describing the backbone and the side chain atoms, respectively). Two glycerol molecules and 59 water molecules were added progressively to the model, and the NCS restraints between the two complexes were reduced to the core of the proteins. Finally, the individual isotropic B -factors were refined. The crystallographic R -factors of the final model are R -factor = 0.251 and $R_{\text{free}} = 0.285$ for the data between 25 and 2.5 Å. The final model consists of two polypeptide chains corresponding to residues 1–218 and 227–261 (loop 219–227 is missing in both molecules), two 13mer DNA duplexes, 59 solvent atoms, one zinc atom per polypeptide chain and one glycerol molecule per complex. Detailed statistics on the quality of the model are presented in Table II. Analysis of the Ramachandran plot shows that 89% of the residues are located in the most favoured zones (PROCHECK; Laskowski *et al.*, 1993), the outsider residues being located in the regions corresponding to the fuzzy electron densities. R.m.s.d.s are 0.007 Å and 1.40° on bonds and angles, respectively.

Accession numbers

The atomic coordinates of the *Lt*Fpg–DNA complex have been deposited in the Protein Database under the accession code 1KFV.

Acknowledgements

We are greatly indebted to M.Pirrochi, P.Carpentier and J.-L.Ferrer (Beamline 30, ESRF-Grenoble, France) for their help in data collection; Dr S.Kuramitsu (Osaka University, Japan) for the kind gift of free *Tt*Fpg coordinates before their release from the Protein Database; Dr R.Lavery for his help using the CURVES program; F.Coste (CRC, Dundee, UK), J.-B.Charbonnier (CEA, Saclay, France) and K.Hinsen (CBM, CNRS, Orléans, France) for critical discussions; and N.Hervouet and N.Bureau for their technical assistance. This work was supported by Electricité de France (EDF), by the Association pour la recherche contre le cancer

(ARC) and the Chemistry/Biology interface Program of the CNRS. K.P. is supported by a doctoral fellowship from the Région Centre (France).

References

- Barrett, T.E., Savva, R., Panayoto, G., Brown, T., Jiricny, J. and Pearl, L.H. (1998) Structure of a DNA base-excision product resembling a cisplatin inter-strand adduct. *Cell*, **92**, 117–129.
- Boiteux, S. and Huisman, O. (1989) Isolation of a formamidopyrimidine-DNA glycosylase (*fpg*) mutant of *Escherichia coli* K12. *Mol. Gen. Genet.*, **215**, 300–305.
- Boiteux, S., O'Connor, T.R. and Laval, J. (1987) Formamidopyrimidine-DNA glycosylase of *Escherichia coli*: cloning and sequencing of the *fpg* structural gene and overproduction of the protein. *EMBO J.*, **6**, 3177–3183.
- Bruner, S.D., Nash, H.M., Lane, W.S. and Verdine, G.L. (1998) Repair of oxidatively damaged guanine in *Saccharomyces cerevisiae* by an alternative pathway. *Curr. Biol.*, **8**, 393–403.
- Bruner, S.D., Norman, D.P.G. and Verdine, G.L. (2000) Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature*, **403**, 859–866.
- Brünger, A. (1992) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Brünger, A.T. et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Burgess, S., Jaruga, P., Dodson, M.L., Dizdaroglu, M. and Lloyd, R.S. (2002) Determination of active site residues in *Escherichia coli* endonuclease VIII. *J. Biol. Chem.*, **277**, 2938–2944.
- Cadet, J., Berger, M., Douki, T. and Ravanat, J.L. (1997) Oxidative damage to DNA: formation, measurement and biological significance. *Rev. Physiol. Biochem. Pharmacol.*, **131**, 1–87.
- Castaing, B., Geiger, A., Seliger, H., Nehls, P., Laval, J., Zelwer, C. and Boiteux, S. (1993) Cleavage and binding of a DNA fragment containing a single 8-oxoguanine by wild type and mutant FPG proteins. *Nucleic Acids Res.*, **21**, 2899–2905.
- Castaing, B., Fourrey, J.L., Hervouet, N., Thomas, M., Boiteux, S. and Zelwer, C. (1999) AP site structural determinants for Fpg specific recognition. *Nucleic Acids Res.*, **27**, 608–615.
- Chetsanga, C.J. and Lindahl, T. (1979) Release of 7-methylguanine residues whose imidazole rings have been opened from damaged DNA by a DNA glycosylase from *Escherichia coli*. *Nucleic Acids Res.*, **6**, 3673–3683.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- Cowtan, K.D. and Main, P. (1993) Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr. D*, **49**, 148–157.
- Cuniasse, P., Fazakerley, G.V., Guschelbauer, W., Kaplan, B.E. and Sowers, L.C. (1990) The abasic site as a challenge to DNA polymerase. A nuclear magnetic resonance study of G, C and T opposite a model abasic site. *J. Mol. Biol.*, **213**, 303–314.
- Czczot, H., Tudek, B., Lambert, B., Laval, J. and Boiteux, S. (1991) *Escherichia coli* Fpg protein and UvrABC endonuclease repair DNA damage induced by methylene blue plus visible light *in vivo* and *in vitro*. *J. Bacteriol.*, **173**, 3419–3424.
- Duarte, V., Gasparutto, D., Jaquinod, M. and Cadet, J. (2000) *In vitro* DNA synthesis opposite oxazolone and repair of this DNA damage using modified oligonucleotides. *Nucleic Acids Res.*, **28**, 1555–1563.
- Evans, P.R. (1993) Data reduction. In Sawyer, L., Isaacs, N. and Bailey, S. (eds), *Proceedings of CCP4 Study Weekend on Data Collection and Processing*. Daresbury Laboratory, Warrington, UK, pp. 114–122.
- Goljer, I., Kumar, S. and Bolton, P.H. (1995) Refined solution structure of a DNA heteroduplex containing an aldehydic abasic site. *J. Biol. Chem.*, **270**, 22981–22987.
- Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **4**, 305–308.
- Graves, R.J., Felzenswalb, I., Laval, J. and O'Connor, T.R. (1992) Excision of 5'-terminal deoxyribose phosphate from damaged DNA is catalyzed by the Fpg protein of *Escherichia coli*. *J. Biol. Chem.*, **267**, 14429–14435.
- Guibourt, N., Castaing, B., Auffret van der Kemp, P. and Boiteux, S. (2000) Catalytic and DNA binding properties of the Ogg1 protein of *Saccharomyces cerevisiae*: comparison between the wild type and the K241R and K241Q active-site mutant proteins. *Biochemistry*, **39**, 1716–1724.
- Hazra, T.K., Izumi, T., Venkataraman, R., Kow, Y.W., Dizdaroglu, M. and Mitra, S. (2000) Characterization of a novel 8-oxoguanine-DNA glycosylase activity in *Escherichia coli* and identification of the enzyme as endonuclease VIII. *J. Biol. Chem.*, **275**, 27762–27767.
- Hollis, T., Ichikawa, Y. and Ellenberger, T. (2000) DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J.*, **19**, 758–766.
- Jiang, D., Hatahet, Z., Blaisdell, J., Melamed, R.J. and Wallace, S.S. (1997) *Escherichia coli* endonuclease VIII: cloning, sequencing and overexpression of the *nei* structural gene and characterization of *nei* and *nei nth* mutants. *J. Bacteriol.*, **179**, 3773–3782.
- Jones, T., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **97**, 110–119.
- Kalnik, M.W., Chang, C.N., Johnson, F., Grollman, A.P. and Patel, D.J. (1989) NMR studies of abasic sites in DNA duplexes: deoxyadenosine stacks into the helix opposite acyclic lesions. *Biochemistry*, **28**, 3373–3383.
- Kraulis, P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lau, A.Y., Schärer, O.D., Samson, L., Verdine, G.L. and Ellenberger, T. (1998) Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell*, **95**, 249–258.
- Lavery, R. and Sklenar, H. (1989) Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.*, **6**, 655–667.
- Lavrukhin, O.V. and Lloyd, R.S. (2000) Involvement of phylogenetically conserved acidic amino acid residues in catalysis by oxidative DNA damage enzyme formamidopyrimidine glycosylase. *Biochemistry*, **39**, 15266–15271.
- Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
- Meritt, E.A. and Murphy, M.E.P. (1994) Raster3D version 2.0, a program for photorealistic molecular graphics. *Acta Crystallogr. D*, **50**, 869–873.
- Michaels, M.L. and Miller, J.H. (1992) The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.*, **174**, 6321–6325.
- Michaels, M.L., Pham, L., Cruz, C. and Miller, J.H. (1991) MutM, a protein that prevents G-C→T-A transversions, is formamidopyrimidine-DNA glycosylase. *Nucleic Acids Res.*, **16**, 3629–3632.
- Nicholls, A., Sharp, K. and Honig, B. (1993) GRASP: graphical representation and analysis of surface properties. *Biophys. J.*, **64**, 166–169.
- Norman, D.P.G., Bruner, S.D. and Verdine, G.L. (2001) Coupling of substrate recognition and catalysis by a human base-excision DNA repair protein. *J. Am. Chem. Soc.*, **123**, 359–360.
- O'Connor, T.R. and Laval, J. (1989) Physical association of the 2,6-diamino-4-hydroxy-5N-formamidopyrimidine-DNA glycosylase of *Escherichia coli* and an activity nicking DNA at apurinic/aprimidinic sites. *Proc. Natl Acad. Sci. USA*, **86**, 5222–5226.
- O'Connor, T.R., Graves, R.J., de Murcia, G., Castaing, B. and Laval, J. (1993) Fpg protein of *Escherichia coli* is a zinc finger protein whose cysteine residues have a structural and/or functional role. *J. Biol. Chem.*, **268**, 9063–9070.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E. and Tainer, J.A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.*, **17**, 5214–5226.
- Pereira de Jésus, K., Serre, L., Hervouet, N., Bouckson-Castaing, V., Zelwer, C. and Castaing, B. (2002) Crystallization and preliminary X-ray crystallographic studies of a complex between the *Lactococcus lactis* Fpg DNA repair enzyme and an abasic site-containing DNA. *Acta Crystallogr. D*, **58**, 679–682.
- Rabow, L.E. and Kow, Y.W. (1997) Mechanism of action of base release

- by *Escherichia coli* Fpg protein: role of lysine 155 in catalysis. *Biochemistry*, **36**, 5084–5096.
- Read, J.R. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A*, **42**, 849–861.
- Roberts, R.J. (1995) On base flipping. *Cell*, **82**, 9–12.
- Schärer, O.D. and Jiricny, J. (2001) Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *BioEssays*, **23**, 270–281.
- Schärer, O.D., Kawate, T., Gallinary, P., Jiricny, J. and Verdine, G.L. (1997) Investigation of the mechanisms of DNA binding of the human G/T glycosylase using designed inhibitors. *Proc. Natl Acad. Sci. USA*, **94**, 4878–4883.
- Sidorkina, O. and Laval, J. (1998) Role of lysine-57 in the catalytic activities of *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg protein). *Nucleic Acids Res.*, **26**, 5351–5357.
- Sidorkina, O. and Laval, J. (2000) Role of the N-terminal proline residue in the catalytic activities of the *Escherichia coli* Fpg protein. *J. Biol. Chem.*, **275**, 9924–9929.
- Sugahara, M., Mikawa, T., Kuramitsu, T., Yamamoto, M., Kato, R., Fukuyama, K., Inoue, Y. and Kuramitsu, S. (2000) Crystal structure of a repair enzyme of oxidatively damaged DNA, MutM (Fpg), from an extreme thermophile, *Thermus thermophilus* HB8. *EMBO J.*, **19**, 3857–3869.
- Tchou, J., Kasai, H., Shibutani, S., Chung, M.H., Laval, J., Grollman, A.P. and Nishimura, S. (1991) 8-Oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity. *Proc. Natl Acad. Sci. USA*, **88**, 4690–4696.
- Tchou, J., Bodepudi, V., Shibutani, S., Antoshechkin, I., Miller, J., Grollman, A.P. and Johnson, F. (1994) Substrate specificity of Fpg protein. Recognition and cleavage of oxidatively damaged DNA. *J. Biol. Chem.*, **269**, 15318–15324.
- Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D*, **55**, 849–861.
- Thayer, M.M., Ahern, H., Xing, D., Cunningham, R.P. and Tainer, T.A. (1995) Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure. *EMBO J.*, **14**, 4108–4120.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Acta Crystallogr. D*, **55**, 849–861.
- Vagin, A. and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D*, **56**, 1622–1624.
- Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.*, **229**, 105–124.
- Vassilyev, D.G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E. and Morikawa, K. (1995) Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell*, **83**, 773–782.
- Weeks, C.M., DeTitta, G.T., Hauptman, H.A., Thuman, P. and Miller, R. (1994) Structure solution by minimal-function phase refinement and Fourier filtering. II. Implementation and applications. *Acta Crystallogr. A*, **50**, 210–220.
- Zharkov, D.O., Rieger, R.A., Iden, C.R. and Grollman, A.P. (1997) NH₂-terminal proline acts as a nucleophile in the glycosylase/AP-lyase reaction catalyzed by *Escherichia coli* formamidopyrimidine-DNA glycosylase (Fpg) protein. *J. Biol. Chem.*, **272**, 5335–5341.
- Zharkov, D.O., Golan, G., Gilboa, R., Fernandes, A.S., Gerchman, S.E., Kycia, J.H., Rieger, R.A., Grollman, A.P. and Shoham, G. (2002) Structure analysis of an *Escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J.*, **21**, 789–800.

Received February 21, 2002; revised and accepted April 24, 2002