Crystal Structure of Human Uracil-DNA Glycosylase in Complex with a Protein Inhibitor: Protein Mimicry of DNA

Clifford D. Mol,* Andrew S. Arvai,*
Russell J. Sanderson,† Geir Slupphaug,‡ Bodil Kavli,‡
Hans E. Krokan,‡ Dale W. Mosbaugh,†
and John A. Tainer*
*Department of Molecular Biology
The Scripps Research Institute
10666 North Torrey Pines Road
La Jolla, California 92037
†Departments of Agricultural Chemistry
  and Biochemistry and Biophysics
and the Environmental Health Sciences Center
Oregon State University
Corvallis, Oregon 97331
‡UNIGEN Center for Molecular Biology
University of Trondheim
N-7005 Trondheim
Norway

Summary

Uracil-DNA glycosylase inhibitor (Ugi) is a B. subtilis bacteriophage protein that protects the uracil-containing phage DNA by irreversibly inhibiting the key DNA repair enzyme uracil-DNA glycosylase (UDG). The 1.9 Å crystal structure of Ugi complexed to human UDG reveals that the Ugi structure, consisting of a twisted five-stranded antiparallel β sheet and two α helices, binds by inserting a β strand into the conserved DNA-binding groove of the enzyme without contacting the uracil specificity pocket. The resulting interface, which buries over 1200 Å² on Ugi and involves the entire β sheet and an α helix, is polar and contains 22 water molecules. Ugi binds the sequence-conserved DNA-binding groove of UDG via shape and electrostatic complementarity, specific charged hydrogen bonds, and hydrophobic packing enveloping Leu-272 from a protruding UDG loop. The apparent mimicry by Ugi of DNA interactions with UDG provides both a structural mechanism for UDG binding to DNA, including the enzyme-assisted expulsion of uracil from the DNA helix, and a crystallographic basis for the design of inhibitors with scientific and therapeutic applications.

Introduction

Damage to DNA arises continually throughout the cell cycle and must be recognized and repaired prior to the next round of replication to maintain the genomic integrity of the cell. DNA base damage can be recognized and excised by the ATP-dependent nucleotide excision repair systems (Aboussekhra and Wood, 1994) or by the base excision repair systems typified by the DNA glycosylases. Uracil-DNA glycosylase (UDG) specifically recognizes uracil in DNA and initiates base excision repair by hydrolyzing the N–C1′ glycosylic bond linking the uracil base to the deoxy-ribose sugar. This creates an abasic site that is removed by a 5′-acting apurinic/apyrimidinic (AP) endonuclease and a deoxyribophosphodiesterase, leaving a gap that is filled by DNA polymerase and closed by DNA ligase. This base excision repair pathway is the major cellular defense mechanism against spontaneous DNA damage (Lindahl, 1994).

Uracil arises in DNA either through incorporation of dUMP in place of dTMP (Tye et al., 1977) or through the continuous deamination of cytosine, which can lead to GC→AT transition mutations if left unrepaired (Frederico et al., 1990; Duncan and Miller, 1980). In vivo, UDGs specifically recognize uracil in DNA and cleave the glycosylic bond to initiate the uracil excision repair pathway. In vitro, UDGs can recognize and remove uracil from both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) substrates. UDGs are ubiquitous enzymes, conserved throughout evolution, with greater than 55% amino acid identity between the human and bacterial proteins (Olsen et al., 1989; Upton et al., 1993). The crystal structures of both the human enzyme (Mol et al., 1995) and the enzyme from herpes simplex virus type 1 (HSV-1) (Savva et al., 1995) have highly conserved active site regions. The architecture of the UDG active site suggests that the enzyme must bind uracil that is extrahelical, or "flipped out", from the DNA base stack.

The Bacillus subtilis bacteriophages PBS1 and PBS2 are novel in that their DNA genomes naturally contain uracil in place of thymine (Takahashi and Marmur, 1963). These bacteriophages produce UDG inhibitor (Ugi) proteins to protect their genomic DNA from host UDG repair enzymes. Thus, PBS2 DNA has evolved to contain an overall base composition of 72% AU and 28% GC (Takahashi and Marmur, 1963). The bacteriophage PBS2 ugi gene has been cloned and expressed in Escherichia coli (Wang and Mosbaugh, 1988). Ugi is an 84 residue 9474 Da acidic protein (pI 4.2) (Bennett and Mosbaugh, 1992), with α/β secondary structure as determined by nuclear magnetic resonance (NMR) (Balasubramanian et al., 1995). Ugi inactivates UDGs from a variety of diverse organisms including Micrococcus luteus, Saccharomyces cerevisiae, rats, and humans (Wang and Mosbaugh, 1989; Karran et al., 1981; Cone et al., 1977). Ugi inhibits E. coli UDG (Ung) by a two-step mechanism (Bennett et al., 1993): rapid pre-equilibrium with a K_d of 1.3 µM is followed by formation of a stable 1:1 complex that is essentially irreversible under physiological conditions (Bennett and Mosbaugh, 1992).

Here, we report inhibition of human UDG by Ugi, as well as the human UDG–Ugi complex crystal structure determined at 1.9 Å resolution. The UDG–Ugi interface is very polar and buries nearly one quarter of the solvent-accessible surface of Ugi, with the two interacting surfaces displaying both shape and electrostatic complementarity. Interactions observed at the UDG–Ugi interface suggest that Ugi achieves its tight binding to UDG by acting as a DNA mimic. This mimicry is most likely specific for a dsDNA

substrate containing an AP site created when the uracil base flips out of the dsDNA base stack. The apparent mimicry by Ugi of DNA interactions with UDG provides a structural basis for UDG binding to dsDNA and clarifies the roles played by key enzyme residues.

Results and Discussion

Inhibition of Human UDG by PBS2 Ugi

The effect of Ugi on UDG activity was determined by measuring the amount of uracil released from synthetic dsDNA substrates with various amounts of inhibitor present in the reaction mixture. Human UDG activity toward uracil-containing dsDNA was effectively inactivated by Ugi (Figure 1). The human enzyme was inactivated ~3 times more efficiently than the E. coli enzyme, since 50% inhibition was achieved at Ugi concentrations of 1.2×10^{-11} M and 3.6×10^{-11} M, respectively. This difference did not reflect a difference in the amount of active enzyme between the two preparations since 740 pmol of both the human and E. coli enzymes complexed with the same amount of [^{35}S] Ugi when placed under saturating conditions of inhibitor protein, indicating that both samples contained identical amounts of complexable enzyme.

Structure Determination

The crystal structure of the complex between human UDG and the Ugi protein from bacteriophage PBS2 was determined using a combination of molecular replacement and multiple isomorphous replacement (MIR) methods (Table 1). The crystals belong to the monoclinic spacegroup P2$_1$ (a = 48.3 Å, b = 64.5 Å, c = 55.2 Å, and β = 113.8°) and contain one UDG–Ugi complex per asymmetric unit. The orientation and position of the UDG in the complex was determined by molecular replacement using the structure of the uncomplexed enzyme refined at 1.57 Å resolution as a model and used to calculate phases that were combined with the MIR phases using the program σ_A (Read, 1986). The σ_A-weighted combined phases had an overall figure of merit of 0.86 for data to 2.5 Å resolution and were used to calculate electron density maps from which the structure of Ugi in the complex was traced (see Experimental Procedures).

The human UDG–Ugi complex structure has been refined at 1.9 Å resolution to a crystallographic R value of 0.197 with reasonable stereochemistry. All 224 residues of UDG and residues 3–84 of Ugi are clearly visible in $2F_o - F_c$ electron density maps (Figure 2). The UDG in the complex is a fully active recombinant protein and is nearly identical to the refined high resolution structure of the free enzyme (Mol et al., 1995). The root-mean-square (rms) deviation between their polypeptide backbone atoms is 0.47 Å, rising to 1.1 Å for all atoms, without any significant shifts in the UDG main chain at the UDG–Ugi interface.

Structure of the PBS2 Ugi Protein

The Ugi structure in the complex consists of a five-stranded antiparallel β sheet and two α helices that fold into a compact disk shape of approximate dimensions 35 Å x 35 Å x 25 Å (Figures 3A and 3B). An N-terminal
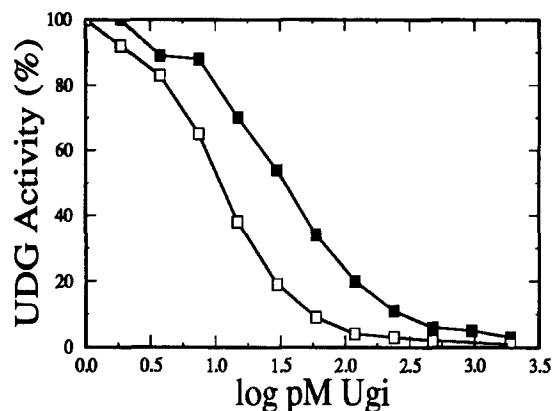

Figure 1. Inhibition of E. coli and Human UDG by Bacteriophage PBS2 Ugi Protein

Two standard UDG inhibitor reaction mixtures (900 μl) were prepared containing either 9 fmol of E. coli UDG (Ung, closed squares) or 9 fmol of human UDG (open squares). Aliquots (75 μl) of each reaction mixture were added to 25 μl samples of Ugi to contain various amounts of inhibitor protein. Each assay mixture, containing 0.07 U of E. coli or 0.11 U of human UDG, was incubated for 30 min at 37°C, and the amount of [^{3}H]uracil released from calf thymus [uracil-^{3}H]DNA (195 cpm/pmol) was determined as previously described (Bennett and Mosbaugh, 1992; Bennett et al., 1994). E. coli and human UDG concentrations were determined using the molar extinction coefficients $E_{280 nm} = 4.22 \times 10^4$ and 5.04×10^4, respectively.

helix, α1 (Ser-I5 to Lys-I14, residue numbers from Ugi are prefixed with the letter I), is followed by the first β strand, β1 (Glu-I20 to Met-I24), of the five-stranded antiparallel β sheet, the second α helix, α2 (Glu-I27 to Asn-I35), and the four remaining strands of the β sheet, β2 to β5 (Ile-I41 to Asp-I48, Glu-I53 to Ser-I60, Ala-I69 to Asp-I74, and Asn-I79 to Leu-I84, respectively). The β strands have nearest neighbor (+1) connectivity and the characteristic right-handed twist seen in other protein structures (Richardson and Richardson, 1989). The two α helices lie on opposite

Table 1. Structure Determination of Human UDG–Ugi Complex

Parameters	Native	EtHgP[a]	Hg$_3$P[b]
Concentration [mM]	—	1	1
Soaking time (hr)	—	24	24
Wavelength (Å)	1.08	1.5418	1.5418
Resolution (Å)	1.9	2.5	2.5
Completeness (%)	93	88	87
Observations	42,119	21,496	20,184
Unique reflections	22,406	9,911	9,745
R_{sym}[c] (%)	5.6	4.9	12.2
R_{iso}[d] (%)	—	23.6	30.3
R_{cullis}[e]	—	0.52	0.65
Rms F_h/E_{iso}[f]	—	2.06	1.50
Number of sites	—	3	1

[a] EtHgP, ethylmercuriphosphate.
[b] Hg$_3$P, mercuric phosphate, Hg$_3$(PO$_4$)$_2$.
[c] R_{sym} is the unweighted R value on I between symmetry mates.
[d] $R_{iso} = \Sigma(|F_{der}| - |F_{nat}|)/\Sigma|F_{nat}|$.
[e] $R_{cullis} = \Sigma[|F_h| - (|F_{ph}| - |F_p|)]/\Sigma|F_h|$, for centric reflections.
[f] Rms F_h/E_{iso}, rms heavy atom F/isomorphous lack of closure error; phasing power.

Figure 2. A Stereoview Looking at the UDG–Ugi Interface at the Active Site of the Enzyme

The electron density (pink) is from a $2F_o - F_c$ map contoured at 1.0σ. The map is superimposed on the coordinates of the complex refined at 1.9 Å resolution. Residues from the UDG active site (yellow carbon bonds) interact closely with residues from the first Ugi β strand (green carbon bonds). The red crosses represent bound water molecules. Figure displayed with XtalView (McRee, 1992).

sides of the sheet roughly perpendicular to the β strands (Figure 3B). Although the topology of Ugi resembles the proteinase inhibitors cystatin (Bode et al., 1988) and stefin B (Stubbs et al., 1990), Ugi apparently binds UDG by a different recognition mechanism. In the proteinase inhibitors, the interactions with proteinase are provided by loops that extend from the β sheet, with the sheet itself oriented away from the enzyme. In the UDG–Ugi complex, the residues of the Ugi β sheet and α2 dominate the interface (Figures 3A and 3B).

The Human UDG–Ugi Interface

Ugi interacts with conserved UDG active site residues over a large interface providing the basis for its tight binding and broad biological specificity for UDG enzymes. Ugi inserts β1 into the active site groove (Figure 2) and envelops the conserved UDG Leu-272 active site loop (Figure 3A). The amino acid distribution at the interface is markedly asymmetric (Figures 3B and 3C). Hydrophobic residues cluster within a pocket between the Ugi β sheet and α2 and at the conserved, protruding UDG loop, with polar and charged amino acids and bound waters lining the interface periphery (Figures 3B and 3C). The interface is very polar, containing 18 hydrogen bonds, involving 12 enzyme side chains and 14 inhibitor side chains. There are 22 ordered water molecules, primarily rimming the interface and the central uracil recognition pocket, which is not bound by Ugi (Figure 4A). The interacting surfaces display electrostatic (Figure 4B) and shape (Figure 4A) complementarity. A negatively charged ridge on the edge of the Ugi β sheet nestles into the positively charged active site groove of UDG, while the Ugi hydrophobic pocket surrounds the protruding UDG Leu-272 loop (Figures 4B and 4C). The two-step kinetic mechanism for Ugi association with E. coli Ung

(Bennett et al., 1993) may derive from initial electrostatic interactions followed by the specific hydrogen bonding and packing interactions that dominate the interface.

The interface encompasses over 1200 Å^2 or nearly one quarter of the total Ugi solvent-accessible surface (Figure 4B); 22 Ugi residues and 22 UDG residues dominate the interface and together comprise 93% and 92%, respectively, of the total buried surface area for each protein (see Figures 3B and 3C). The UDG Leu-272 side chain is completely exposed in the free enzyme and completely buried in the complex (Figure 4C). Near the rim of the Ugi leucine-binding pocket, the side chains of three inhibitor residues (Asn-I54, Gln-I73, and Glu-I78) establish a hydrogen-bonded network with ordered solvent molecules and with the backbone atoms of UDG residues Pro-271 through Tyr-275.

For Ugi, the residues of β1 undergo the most drastic change in solvent accessibility upon complex formation. Conserved amino acids near the active site of the enzyme interact with six sequential Ugi residues (Gln-I19 to Met-I24; Figure 2). The side chains of conserved glutamine residues Gln-144 and Gln-152 form hydrogen bonds with the backbone amide and carbonyl groups of Leu-I23 and the backbone amide of Gln-I19, respectively. The His-268 imidazole packs against the Ile-I22 side chain, while Ser-I21 Oγ forms a strong hydrogen bond (~ 2.8 Å) with His-148 Nε2 (Figure 2). The Glu-I20 carboxylate forms a pair of hydrogen bonds to the Ser-169 backbone amide and side chain Oγ. The Glu-I28 carboxylate, near the N-terminus of Ugi α2, similarly forms hydrogen bonds to the Ser-247 amide and Oγ. Glu-I28 Oε1 also forms water-mediated hydrogen bonds with the His-268 backbone amide and Ser-273 Oγ. Additional interactions observed between Ugi and UDG involving nonconserved UDG resi-
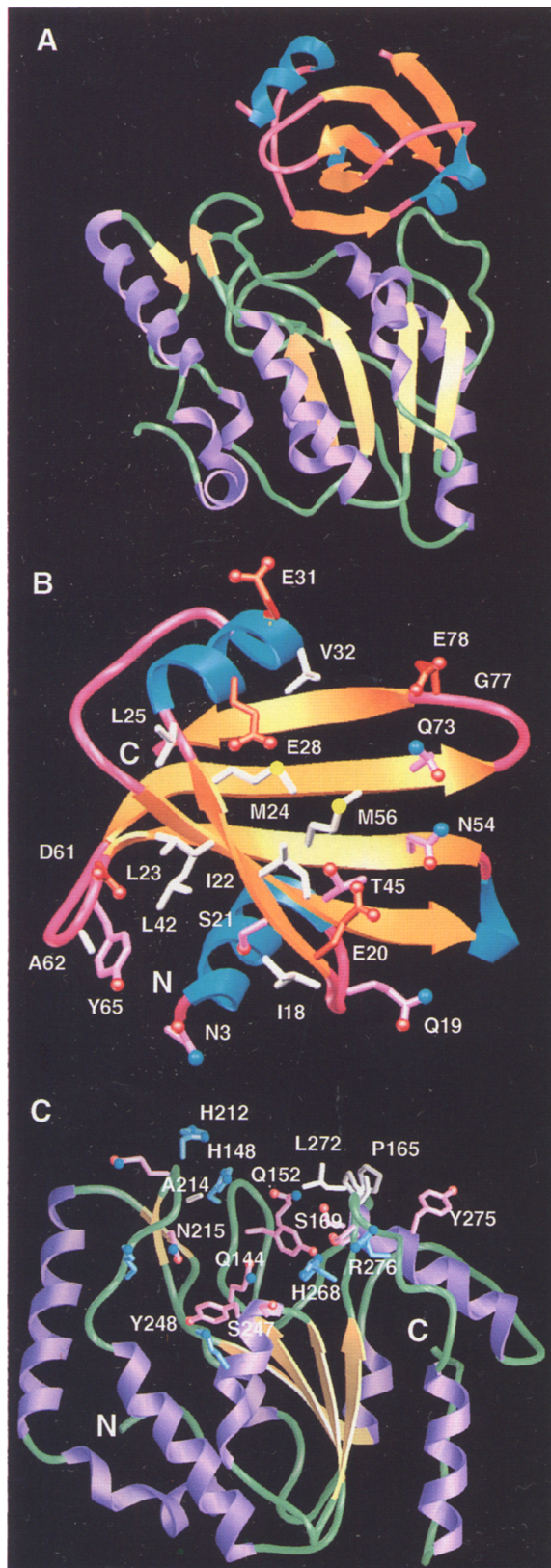
dues include salt bridges between Asp-I61 and Lys-218 (alanine in E. coli Ung) and between Glu-I31 and Arg-276.

Ugi does not bind directly to residues of the UDG uracil-binding pocket (Tyr-147, Phe-158, Asn-204, and the Gln-144 and Asp-145 backbone) (Figure 4A). Thus, the tight binding of Ugi to UDG does not derive from targeting of the exquisite specificity of the enzyme for uracil, but apparently arises from its ability to functionally mimic substrate DNA binding. Several biochemical observations (Bennett and Mosbaugh, 1992; Bennett et al., 1994) and mutational analysis (Mol et al., 1995) suggest that Ugi binds to UDG at the DNA-binding site. First, the UDG–Ugi complex does not bind DNA. Second, UV-catalyzed cross-linking of UDG to single-stranded oligonucleotide dT_{20} abolished binding of Ugi. Third, the UDG–Ugi complex is refractory to UV cross-linking to dT_{20}. Fourth, five single-site UDG mutants in the UDG–Ugi interface (Gln-144, Tyr-147, His-148, His-268, and Ser-270) all show reduced binding to DNA (Mol et al., 1995). Taken together, these results suggest that binding of Ugi and DNA to the enzyme are mutually exclusive and are likely to involve the same enzyme residues. Moreover, the complex crystal structure suggests the remarkable Ugi inactivation of UDGs from diverse biological systems results from Ugi targeting the DNA-binding region of UDG. Apparently, Ugi has capitalized on the biologically required structural conservation of UDG regions that bind DNA, as no selective pressure exists to maintain the integrity of the Ugi-binding site. Ugi apparently mimics duplex DNA interactions with UDG by overall charge and shape

---

Figure 3. Structure of the UDG–Ugi Complex Shown as Ribbon Diagrams

(A) View showing Ugi (top) bound to UDG (bottom). β1 of the five-stranded antiparallel Ugi β sheet (orange arrows) inserts into the UDG active site groove approximately perpendicular to the C-terminal edge of the four-stranded parallel UDG β sheet (yellow arrows). The sequence-conserved (His-268, Pro-269, Ser-270, Pro-271, Leu-272, and Ser-273) UDG active site loop (green tube, right) protrudes upward into a hydrophobic pocket formed by α2 (blue coil, right) and the curved β sheet of Ugi.

(B) Ugi structure in the complex, shown from the viewpoint of the enzyme active site. Side chains of inhibitor residues that interact with UDG are labeled (with their one letter amino acid code and residue number) and colored according to their chemical nature (negatively charged, red; polar, purple; hydrophobic, white), with nitrogen (blue), oxygen (red), and sulfur (yellow) atoms displayed as spheres. Ugi residues in the UDG–Ugi interface include the following: Asn-3 (bottom) from the N-terminus (N); Ile-18, Gln-19 from the loop (pink tube, lower right) between α1 and β1; all residues (Glu-20, Ser-21, Ile-22, Leu-23, and Met-24) from β1 (yellow arrow, foreground, lower right to upper left); Glu-28, Glu-31, and Val-32 from α2 (blue coil, upper left); Thr-45 from β2; Asn-54 and Met-56 from β3 (central yellow arrow, horizontal); Asp-61, Ala-62, and Tyr-65 from the loop between β3 and β4 (pink tube, lower left); Gln-73 from β4; and Gly-77 and Glu-78 from the loop between β4 and β5 (pink tube, upper right).

(C) UDG structure in the complex shown with the active site groove and Ugi interface at the top and β sheet (yellow arrows) approximately on edge (rotated ~90° around a vertical axis from [A]). The protein is colored according to secondary structure with β sheets (yellow), α helices (purple), and loop regions (green). Coloring and labeling of the amino acid side chains that interact with UDG is as described for Ugi with positively charged residues colored blue. Key UDG residues in the UDG–Ugi interface include the following: Gln-144, His-148, Gln-152, Pro-165, Ser-169, His-212, Ala-214, Asn-215, Ser-247, Tyr-248, His-268, Leu-272, Tyr-275, and Arg-276. Figure displayed using standard and local modules of the AVS graphics system (AVS Incorporated, Waltham, Massachusetts).
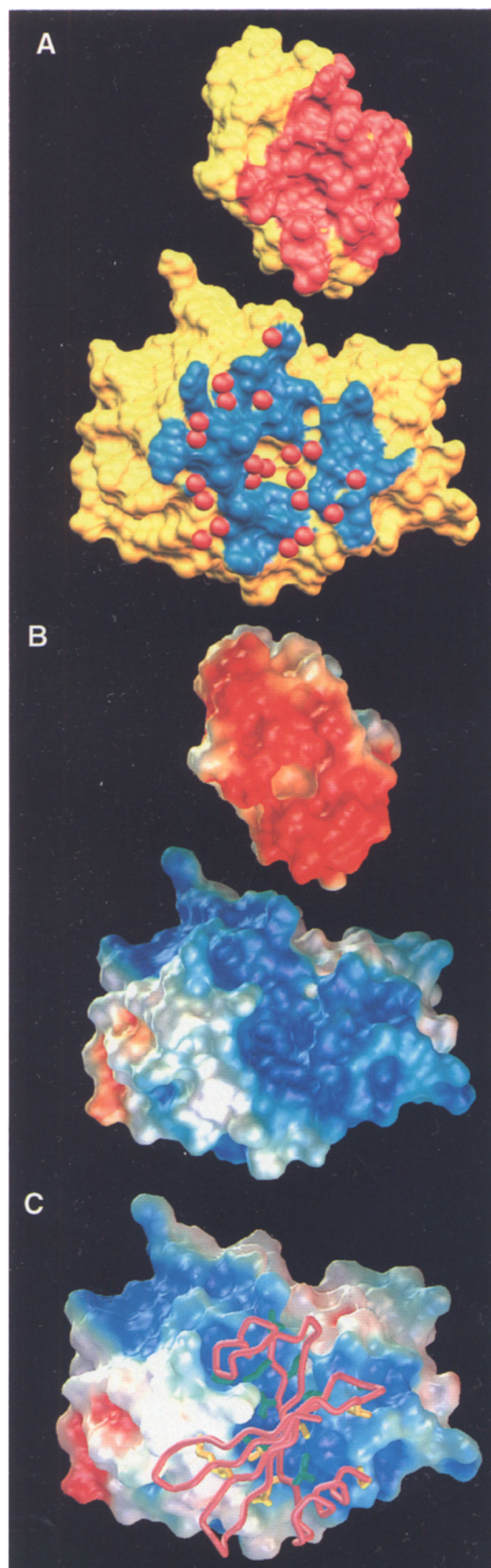
complementarity through charged hydrogen bonds to conserved active site residues and by insertion of the protruding Leu-272 of UDG into the hydrophobic pocket of Ugi. However, Ugi does not inhibit other DNA glycosylases that are specific for 3-methyladenine, 2,6-diamino-4-hydroxy-5-(N-methylformamido) pyrimidine or hypoxanthine (Karran et al., 1981), or other DNA metabolizing enzymes, such as E. coli DNA polymerase I, HeLa DNA polymerase α, Novikoff hepatoma DNA polymerase β, or porcine liver DNA polymerase γ, which also recognize DNA (Wang and Mosbaugh, 1989). Thus, the DNA mimicry by Ugi is specific for the DNA-binding region of UDGs.

Implications for UDG Recognition of Uracil-Containing dsDNA

These structural results and those from the human (Mol et al., 1995) and HSV-1 (Savva et al., 1995) UDG crystal structures suggest a specific model for UDG–DNA interaction. The structure of the human UDG–Ugi complex highlights conserved residues that are likely to play key roles in the in vivo activity of UDG. Glu-I20 and Glu-I28 appear to mimic DNA phosphate groups. Glu-I20 interacts with the equivalent residues in human UDG that are identified as the sulfate- and phosphate-binding sites in the HSV-1 enzyme (Savva et al., 1995). Assuming Leu-272 pene-

Figure 4. Solvent-Accessible Surfaces of Human UDG and PBS2 Ugi

View rotated ~90° around a horizontal axis from Figure 3A looking down into the active site groove of UDG (bottom). The Ugi protein (top) has been rotated ~180° to expose the interacting surface that contacts the enzyme.

(A) The surface is colored according to the buried surface area of the UDG–Ugi complex. Residues of UDG that are inaccessible to a 1.6 Å solvent probe when Ugi binds are colored blue, while the buried residues of Ugi are colored red. An area of approximately 1200 Å², out of a total of 4800 Å², of the solvent-accessible surface of Ugi is buried. The corresponding buried surface area of UDG in the complex is about 1275 Å² out of a total surface area of 9880 Å². Solvent molecules that occur at the interface (12 buried waters and a further 10 partially buried) are shown as red spheres and mapped onto the surface of UDG. The yellow area in the center of the UDG buried surface corresponds to the uracil-recognition pocket, which is filled with solvent and does not interact directly with the inhibitor.

(B) The surface is colored according to the electrostatic potential (−3 kT/e to +3 kT/e) with regions of positive potential colored blue, neutral white, and negative red. Surface potential was calculated with the program DELPHI (Biosym Technologies, Incorporated) using AMBER partial charges at pH 5.0 and an ionic strength of 145 mM and a radius of 1.6 Å for the solvent, with the solute and solvent dielectric constants set to 2.0 and 80.0, respectively. The side chain of Leu-272 projects out from the surface of UDG (white, center left) and inserts into the hydrophobic pocket on the otherwise negatively charged surface of Ugi (pink, center). This leucine-binding pocket is lined by a number of hydrophobic residues, including Met-I24, Val-I32, and Met-I56.

(C) The surface of the UDG is colored according to the electrostatic potential (−3 kT/e to +3 kT/e) as in (B). Ugi is shown as a red tube in the bound orientation observed in the human UDG–Ugi complex crystal structure. Key side chains of Ugi that interact with UDG are also shown (tubes colored green for hydrophobic and yellow for polar residues). The insertion of the Leu-272 side chain (white, center left) into the hydrophobic Ugi leucine-binding pocket and the extensive interactions that Ugi makes as it encompasses the UDG Leu-272 loop can be clearly seen. Figure displayed using standard and local modules of the AVS graphics system.

trates the DNA base stack and a flipped-out uracil base reaches into the uracil-binding pocket, DNA from the HhaI (cytosine-5)-methyltransferase–DNA complex crystal structure (Klimasauskas et al., 1994) can be docked onto UDG (Figure 5). In this model complex, as proposed for HSV-1 UDG (Savva et al., 1995), the long axis of the DNA helix is oriented essentially parallel to the positively charged UDG active site groove. The Leu-272 loop inserts into the DNA major groove, while the α helix connecting the third and fourth β strands of UDG directs its helix dipole toward the DNA minor groove. This model UDG–DNA complex, and the UDG–Ugi complex structure, suggest that the Ugi β sheet and α2 structurally mimic the sugar–phosphate backbones of the DNA strands on either side of the UDG Leu-272 loop. With DNA bound in this orientation, the His-268 environment is not perturbed by DNA phosphate groups, and the extrahelical base from the methyltransferase–DNA complex fits into the UDG active site. In fact, the uracil 3′-phosphate is in the same position as the Ugi Glu-I20 carboxylate, and the uracil 5′-phosphate interacts with the Gln-144 side chain.

In this model, the Leu-272 side chain penetrates the DNA base stack from the major groove, either passively occupying the space vacated by an extrahelical uracil or actively guiding the uracil into the uracil recognition pocket of the enzyme. Thus, the uracil could only flip out on the minor groove side of the DNA, as suggested for the HhaI methyltransferase (Klimasauskas et al., 1994). If actual "flipping out" of the uracil base is rate limiting, as suggested by the fact that the human, E. coli, and HSV-1 enzymes all excise uracil more efficiently from UG mispairs than from UA pairs (Bennett et al., 1995; Verri et al., 1992), then the DNA sequence surrounding the uracil should influence the cleavage rate. Uracil accessibility to the active site pocket of the enzyme should decrease as the local melting of the DNA base stack decreases with increasing GC content. This is found for uracil excision by human UDG under conditions in which enzyme concentration is rate limiting. Uracil excision after 5 min incubation ranges from 100% for the 5′ to 3′ sequence CAUAA, to 50% for CGUAA, falling to 25% for TGUGA, and finally only 5% for GGUGG (Slupphaug et al., 1995). Combined with the kinetics for UDG–Ugi complex formation (Bennett et al., 1993) and the structural information provided here, these sequence activity data suggest that UDG binding to both uracil-containing dsDNA and Ugi may be governed by the particular interactions surrounding Leu-272, either penetration into the DNA base stack or insertion into the leucine-binding pocket of Ugi. This may explain the specific nature of the Ugi mimicry of the DNA substrate of UDG.

Uracil in DNA is a rather subtle lesion, closely related to the naturally occurring DNA bases thymine, cytosine, and 5-methylcytosine. The recognition system developed for uracil detection may require actively probing the DNA by facilitating base expulsion rather than fortuitously binding to a preexisting extrahelical uracil. A productive UDG–dsDNA complex can only be formed, however, if the flipped-out base is a uracil. No other naturally occurring base will fit into the recognition pocket of the enzyme (Mol
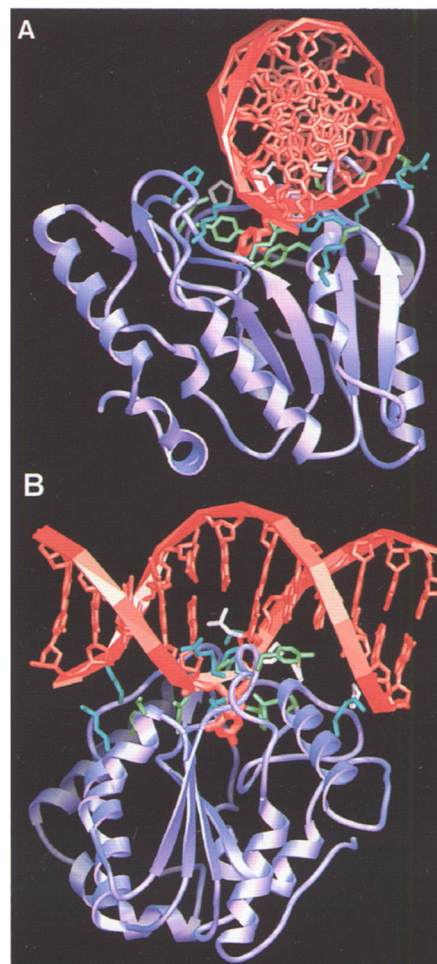

Figure 5. Model for UDG Binding to Uracil-Containing dsDNA

The UDG is shown as a Cα ribbon (purple) with key residue side chains from the UDG–Ugi interface depicted as colored tubes with positive (blue), polar (green), and hydrophobic (white). The DNA (red) is shown as a phosphate backbone ribbon with the individual nucleotides (tubes) and the extrahelical uracil base (thicker tubes). The DNA has been slightly bent about its long axis using the local DNABENDER module (developed by T. J. Macke) for the AVS graphics system. (A) View of the proposed UDG–dsDNA complex with UDG in the same orientation as in Figure 3A. The DNA interacts with the same regions of UDG that interact with Ugi in the complex. (B) View rotated ~90°, illustrating the penetration of the side chain of Leu-272 into the DNA base stack via the major groove.

et al., 1995; Savva et al., 1995). Thus, the side chain of Leu-272 will fully insert into the dsDNA base stack only when a uracil lesion has been detected. Full insertion of Leu-272 into the abasic site created by an extrahelical uracil can be expected to cause a certain amount of DNA bending, and the curvature and distribution of positively charged side chains around the UDG active site groove is able to accommodate this induced DNA bending (Figures 4B and 5B). The resultant widening of the bent DNA major groove would allow deeper Leu-272 penetration and its movement toward the active site of the enzyme. Such movement of Leu-272 toward the active site is rigidly coupled to the proposed catalytic His-268 via the highly con-

served intervening residues Pro-269, Ser-270, and Pro-271. Thus, a productively bound flipped-out uracil base and a fully inserted Leu-272 side chain into the base stack of bent dsDNA may allow His-268 to move into catalytic position for direct or general base attack.

Taken together, the structural results from the UDG–Ugi complex reveal that Ugi acts as a DNA mimic, suggest a more detailed model for UDG–DNA interactions, and provide the structural basis for the inhibition of an important DNA repair enzyme. The design of new specific DNA repair inhibitors is of paramount importance for improving the efficacy and selectivity of current chemo- and radiotherapies as well as for developing new antiviral strategies.

Experimental Procedures

UDG–Ugi Complex Formation

The UDG–Ugi complex was formed by combining 0.5 μmol of human UDG with 1.0 μmol of Ugi and incubated at room temperature for 10 min and at 4°C for an additional 20 min. UDG–Ugi complex was purified from excess Ugi by DE-52 cellulose column chromatography as previously described (Bennett and Mosbaugh, 1992), except that the equilibration buffer contained 30 mM Tris–HCl (pH 7.4), 1 mM EDTA, 1 mM DTT, 50 mM NaCl, and 5% (w/v) glycerol. Complex was eluted stepwise by the addition of 120 ml of equilibration buffer containing 125 mM NaCl. Fractions (4 ml) were analyzed by both 18% nondenaturing and 20% SDS–polyacrylamide gel electrophoresis to verify that complex was resolved from free UDG and Ugi (Bennett and Mosbaugh, 1992). Peak fractions of UDG–Ugi were pooled, concentrated ~32-fold to ~1 ml, and diafiltrated into equilibration buffer without NaCl using an Amicon stirred cell (YM10 membrane). Following the concentration step, the UDG–Ugi complex ($A_{280\ nm}$ = 10.8) was determined to contain 170 μM protein by absorbance spectroscopy using the molar extinction coefficient $E_{280\ nm}$ = 6.3 × 10^4 calculated from the sum of the extinction coefficients of UDG (Figure 1) and Ugi ($E_{280\ nm}$ = 1.2 × 10^4).

Crystallization and Data Collection

The cloning, expression, and purification of human UDG have been described previously (Olsen et al., 1989; Slupphaug et al., 1995; Mol et al., 1995). UDG refers to the enzyme from human and nonbacterial systems. The human UDG used here represents an active recombinant protein (UNGΔ84), in which the N-terminal 84 amino acids encoded by the human *UNG* gene have been replaced by three amino acids from the vector. The E. coli UDG counterpart (Ung) is encoded by the *ung* gene. The cloning and expression of the bacteriophage PBS2 Ugi protein are described elsewhere (Wang and Mosbaugh, 1988; Wang et al. 1991).

Crystals of the human UDG–Ugi complex were grown by the hanging drop vapor diffusion technique by mixing equal volumes of the protein solution (15 mg/ml) with reservoir solutions containing 20% polyethylene glycol 4000 (Sigma) and 100 mM imidazole/maleate buffer (pH 5.0). The crystals are monoclinic P2$_1$ (a = 48.3 Å, b = 64.5 Å, c = 55.2 Å, and β = 113.8°) and contain one UDG–Ugi complex per crystallographic asymmetric unit. These crystals have a V_M (Matthews, 1968) of 2.3, which is within the range observed for other protein crystals.

X-ray diffraction data for all crystals were collected at room temperature using a 30 cm MAR research imaging plate and an oscillation width of 1.0° per frame. Native data were collected at a wavelength of 1.08 Å on beamline 7-1 of the Stanford Synchrotron Radiation Laboratory (SSRL) and were processed using the program DENZO (Otwinowski, 1993). Heavy atom derivative data were collected at a wavelength of 1.54 Å using a Rigaku rotating anode X-ray generator operating at 50 kV and 100 mA with a typical exposure time of 15 min per frame. These data were processed using the program MARXDS (Kabsch, 1988). Heavy atom positions were determined from isomorphous difference Patterson maps and cross-difference Fourier maps using the XtalView program suite (McRee, 1992), and by analogy to the same

heavy atom–binding sites determined for the structure of uncomplexed human UDG (Mol et al., 1995). The heavy atom positions and occupancies were refined and MIR map coefficients calculated using XtalView.

Molecular Replacement

Molecular replacement calculations were performed with X-PLOR (Brünger et al., 1987) version 3.1, using as a model the structure of uncomplexed UDG (Mol et al., 1995) refined to 1.57 Å resolution with a crystallographic R value of 0.197 (30,530 reflections, R_{free} = 0.255) with rms deviations of the bond lengths and angles of 0.020 Å and 3.1°, respectively. Rotation function calculations were performed using all data (5,369 reflections) in the resolution range 3.0–6.0 Å. The uncomplexed UDG search model was placed in a P1 box of dimensions 128.0 × 128.0 × 128.0 Å, and the maximum Patterson search vector was set to 35.0 Å. The correct rotation peak was at a height of 14.6σ above the mean with the highest background peak at a height of 3.9σ above the mean. This rotation peak gave a correlation factor of 0.31 after Patterson correlation refinement (Brünger et al., 1987). The translation search was carried out in the Y = 0.0 plane using the oriented model and data between 3.0 and 6.0 Å resolution. The correct translation peak was 9.8σ above the mean, with the highest background peak at a height of 2.4σ above the mean.

Model Building and Refinement

The correctly oriented and positioned UDG model was subjected to Powell conjugate-gradient energy minimization using X-PLOR (Brünger et al., 1987) yielding an initial crystallographic R value of 0.281 for 9,480 reflections in the resolution range 2.5–6.0 Å. The value of R_{free} (Brünger, 1992) for 5% of reflections against which the model was not refined was 0.422. This model was used to calculate phases that were combined with the MIR phases using the program σ$_A$ (Read, 1986) to produce combined phase maps at a resolution of 2.5 Å. Electron density for portions of Ugi that were close to UDG, notably β1 and α2, were clearly visible in these phase-combined electron density maps, and the amino acid sequence was fit to the density. This partial model of Ugi was then refined along with the model for UDG and used to calculate new phases. Cycles of model building, refinement, and phase combination with σ$_A$ were used until the entire sequence of Ugi from residues 3–84 had been placed.

This complete model for the human UDG–Ugi complex was refined using stereochemically restrained simulated annealing and positional refinement with X-PLOR. Rounds of refinement, gradually including the high resolution data to 1.9 Å resolution, were interspersed with manual inspection and rebuilding of the structure using the interactive graphics program XFIT (McRee, 1992). The model was inspected using σ$_A$-weighted $F_o - F_c$ and $2F_o - F_c$ electron density maps, as well as simulated–annealed omit maps in which 20 residue stretches of the model had been omitted from the phase calculation. Ordered solvent molecules were manually placed at the appropriate positions of $F_o - F_c$ electron density maps and included in the refinement at higher resolutions. Individual atomic temperature values were applied to all atoms, and an overall anisotropic temperature factor correction was applied to the native data set. Progress was monitored throughout the course of refinement and manual rebuilding of the model by monitoring the consistent decrease in the values of both the conventional crystallographic R value and the free R value.

The human UDG–Ugi complex structure contains 2455 nonhydrogen protein atoms and 267 ordered solvent molecules and has been refined to an R value of 0.197 for 19,407 reflections in the resolution range 8.0–1.9 Å with F > 3.0 σ$_F$. The corresponding value for the free R is 0.282 for 5% of reflections against which the model was not refined. The rms deviations from ideality of the bond lengths is 0.018 Å, the bond angles 3.47°, the dihedral angles 25.7°, and the improper angles 1.56°.

References

Aboussekhra, A., and Wood, R. D. (1994). Repair of UV-damaged DNA by mammalian cells and *Saccharomyces cerevisiae*. Curr. Opin. Genet. Dev. *4*, 212–220.

Balasubramanian, S., Beger, R. D., Bennett, S. E., Mosbaugh, D. W., and Bolton, P. H. (1995). Secondary structure of uracil-DNA glycosylase inhibitor protein. J. Biol. Chem. *270*, 296–303.

Bennett, S. E., and Mosbaugh, D. W. (1992). Characterization of the *Escherichia coli* uracil-DNA glycosylase:inhibitor protein complex. J. Biol. Chem. *267*, 22512–22521.

Bennett, S. E., Schimerlik, M. I., and Mosbaugh, D. W. (1993). Kinetics of the uracil-DNA glycosylase/inhibitor protein association. J. Biol. Chem. *268*, 26879–26885.

Bennett, S. E., Jensen, O. N., Barofsky, D. F., and Mosbaugh, D. W. (1994). UV-catalyzed crosslinking of *Escherichia coli* uracil-DNA glycosylase to DNA: identification of amino acid residues in the single-stranded DNA binding site. J. Biol. Chem. *269*, 21870–21879.

Bennett, S. E., Sanderson, R. J., and Mosbaugh, D. W. (1995) Processivity of *Escherichia coli* and rat liver mitochondrial uracil-DNA glycosylase is affected by NaCl concentration. Biochemistry *34*, 6109–6119.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988). The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. *7*, 2593–2599.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brünger, A. T., Kuriyan, J., and Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. Science *235*, 458–460.

Cone, R., Duncan, J., Hamilton, L., and Friedberg, E. C. (1977). Partial purification and characterization of a uracil DNA N-glycosidase from *Bacillus subtilis*. Biochemistry *16*, 3194–3201.

Duncan, B. K., and Miller, J. H. (1980). Mutagenic deamination of cytosine residues in DNA. Nature *287*, 560–561.

Frederico, L. A., Kunkel, T. A., and Shaw, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry *29*, 2532–2537.

Kabsch, W. J. (1988). Evaluation of single crystal X-ray diffraction data from a position sensitive detector. J. Appl. Crystallogr. *21*, 916–924.

Karran, P., Cone, R., and Friedberg, E. C. (1981). Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase. Biochemistry *20*, 6092–6096.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.

Lindahl, T. (1994). Instability and decay of the primary structure of DNA. Nature *362*, 709–715.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491–497.

McRee, D. E. (1992). XtalView: a visual protein crystallographic software system for X11/XView. J. Mol. Graphics *10*, 44–47.

Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., and Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell *80*, 869–878.

Olsen, L. C., Aasland, R., Wittwer, C. U., Krokan, H. E., and Helland, D. E. (1989). Molecular cloning of human uracil-DNA glycosylase. EMBO J. *8*, 3121–3125.

Otwinowski, Z. (1993). Oscillation data reduction program. In Data Collection and Processing, L. Sawyer, N. Isaacs, and S. Bailey, eds.

(Warrington, England: Science and Engineering Research Council/ Daresbury Laboratory), 56–62.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallogr. *A42*, 140–149.

Richardson, J. S., and Richardson, D. C. (1989). Principles and patterns of protein conformation. In Prediction of Protein Structure and the Principles of Protein Conformation, G. D. Fasman, ed. (New York: Plenum). pp. 2–98.

Savva, R., McAuley-Hecht, K., Brown, T., and Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature *373*, 487–493.

Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T., Levine, D. W., and Krokan, H. E. (1995). Properties of a recombinant human uracil-DNA glycosylase from the UNG-gene and evidence that UNG encodes the major uracil-DNA glycosylase. Biochemistry *34*, 128–138.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990). The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J. *9*, 1939–1947.

Takahashi, I., and Marmur, J. (1963). Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. Nature *197*, 794–795.

Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S., and Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. Proc. Natl. Acad. Sci. USA *74*, 154–157.

Upton, C., Stuart, D. T., and McFadden, G. (1993). Identification of a poxvirus gene encoding a uracil DNA-glycosylase. Proc. Natl. Acad. Sci. USA *90*, 4518–4522.

Verri, A., Mazzarello, P., Spadari, S., and Focher, F. (1992). Uracil-DNA glycosylases preferentially excise mispaired uracil. Biochem. J. *287*, 1007–1010.

Wang, Z., and Mosbaugh, D. W. (1988). Uracil-DNA glycosylase inhibitor of bacteriophage PBS2: cloning and effects of expression of the inhibitor gene in *Escherichia coli*. J. Bacteriol. *170*, 1082–1091.

Wang, Z., and Mosbaugh, D. W. (1989). Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. *264*, 1163–1171.

Wang, Z., Smith, D. G., and Mosbaugh, D. W. (1991) Overproduction and characterization of uracil-DNA glycosylase inhibitor of bacteriophage PBS2. Gene *99*, 31–37.