Crystal Structure of *Escherichia coli* Uracil DNA Glycosylase and Its Complexes With Uracil and Glycerol: Structure and Glycosylase Mechanism Revisited

Gaoyi Xiao, Maria Tordova, Jaya Jagadeesh, Alexander C. Drohat, James T. Stivers,* and Gary L. Gilliland*
Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute and the National Institute for Standards and Technology, Rockville, Maryland

ABSTRACT **The DNA repair enzyme uracil DNA glycosylase (UDG) catalyzes the hydrolysis of premutagenic uracil residues from single-stranded or duplex DNA, producing free uracil and abasic DNA. Here we report the high-resolution crystal structures of free UDG from *Escherichia coli* strain B (1.60 Å), its complex with uracil (1.50 Å), and a second active-site complex with glycerol (1.43 Å). These represent the first high-resolution structures of a prokaryotic UDG to be reported. The overall structure of the *E. coli* enzyme is more similar to the human UDG than the herpes virus enzyme. Significant differences between the bacterial and viral structures are seen in the side-chain positions of the putative general-acid (His187) and base (Asp64), similar to differences previously observed between the viral and human enzymes. In general, the active-site loop that contains His187 appears preorganized in comparison with the viral and human enzymes, requiring smaller substrate-induced conformational changes to bring active-site groups into catalytic position. These structural differences may be related to the large differences in the mechanism of uracil recognition used by the *E. coli* and viral enzymes. The pH dependence of k_{cat} for wild-type UDG and the D64N and H187Q mutant enzymes is consistent with general-base catalysis by Asp64, but provides no evidence for a general-acid catalyst. The catalytic mechanism of UDG is critically discussed with respect to these results. Proteins 1999;35:13–24.** © 1999 Wiley-Liss, Inc.

Key words: *E. coli* uracil DNA glycosylase; uracil complex; glycosylase mechanism; acid-base catalysis

INTRODUCTION

A uracil base in DNA may result from incorporation of deoxyuridine triphosphate instead of thymidine triphosphate during replication, or from deamination of cytosine in DNA. It can be estimated that ~10^4 uracil bases are incorporated into human DNA following S-phase DNA synthesis and that spontaneous cytosine deamination occurs at a rate of ~200 events per human cell per day.[1] The latter deamination pathway results in a premutagenic guanine-uracil mismatch that, unless repaired before the next round of replication, will result in a guanine-cytosine to adenine-thymine transition mutation. Thus, cytosine deamination is a major mutagenic force, and a multistep uracil-excision DNA repair pathway has been conserved in organisms as diverse as bacteria and humans to repair this type of DNA damage. The first step in this pathway is the hydrolysis of the *N*-glycosidic bond between uracil and the deoxyribose sugar catalyzed by UDG.

Although crystal structures of *h*UDG and *v*UDG, and their complexes with 6-aminouracil and uracil have been reported,[2,3] the high-resolution structure of a prokaryotic UDG has not. The structures of the free viral and human enzymes have been somewhat ambiguous in terms of a catalytic mechanism because of the differing orientations of an active-site aspartic acid and histidine residue, which led to confusion as to the catalytic roles of these completely conserved residues.[2,3] More recently, several crystal structures of the human enzyme in complex with the products uracil and duplex abasic DNA have been solved. These structures establish that UDG undergoes a conformational change, and “”flips” the uracil base out of the major groove of the DNA helix into the active site where essential interactions with the uracil base and deoxyribose can be made.[4,5] On the basis of these structures and mutagenesis experiments, it has been concluded that UDG uses a general acid-base reaction mechanism.[4] In this proposed mechanism, the conserved aspartic acid serves as the base to activate the incoming water nucleophile, and the con-

*Abbreviations: h*UDG, human uracil-DNA glycosylase; *v*UDG, herpes simplex virus type-1 uracil-DNA glycosylase; *e*UDG, *E. coli* strain B uracil-DNA glycosylase; Tris-HCl, tris(hydroxymethyl)aminomethane; HEPES, 4-(2-hydroxyethyl)-1-piperazineethane sulfonic acid; SDS, sodium dodecyl sulfate; EDTA, (ethylenedinitrilo)-tetraacetic acid, disodium salt; MALDI, matrix-assisted laser desorption-ionization; TMN, 10 mmol/L Tris-HCl (pH 8.0), 2.5 mmol/L $MgCl_2$, 25 mmol/L NaCl; **P**, 2-aminopurine; 4k PEG, polyethylene glycol with molecular weight of 4,000.

Certain commercial equipment, instruments, and materials are identified in this article to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the material or equipment identified is necessarily the best available for the purpose.

*Correspondence to: Gary L. Gilliland or James Stivers, CARB University of Maryland Biotechnology Institute and NIST, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: gary.gilliland@carb.nist.gov or stivers@carb.nist.gov

Received 28 September 1998; Accepted 20 November 1998

© 1999 WILEY-LISS, INC.

served histidine residue is the electrophilic catalyst facilitating leaving group departure by protonation of uracil O2. The importance of these residues is clearly established because they are completely conserved in all UDGs, and the human enzyme shows ~2,500-fold and 300-fold decreases in specific activity, respectively, when this aspartic acid and histidine are mutated.[2]

To establish the chemical mechanism and understand the specificity of this highly conserved DNA repair enzyme, we have initiated detailed structural, mechanistic, and nuclear magnetic resonance spectroscopy studies of *e*UDG.[6,7] The *E. coli* enzyme is ideal for such studies because it is small, soluble, overexpressed to very high levels, and gives high-resolution diffraction data (<1.60 Å) using an in-house X-ray source. Reported here are the high-resolution crystal structures of the free *E. coli* enzyme, its complexes with uracil, and the previously unreported inhibitor, glycerol. These represent the first high-resolution structures of a prokaryotic UDG, and establish the structure of the active site in the free enzyme and the binary uracil complex. The catalytic roles of the putative general base (Asp64) and general acid (His187) are discussed with respect to detailed studies of the pH dependence of the steady-state rates and mutagenesis studies of *e*UDG. The structural differences between *v*UDG and *e*UDG enzyme may contribute to the apparently different mechanisms of uracil recognition used by these enzymes.[7,8]

MATERIALS AND METHODS

Cloning and Purification of UDG and Site-Directed Mutagenesis

UDG from *E. coli* strain B was cloned from *E. coli* strain B genomic DNA (Ultrapure, Sigma, St. Louis, MO) using polymerase chain reaction (PCR) primers based on the 3′- and 5′-flanking regions of the published *ung* gene sequence from *E. coli* strain K12.[9] The full-length PCR product was ligated into the pET-21a expression vector (Novagen Inc., Madison, WI) to give the pET21a-UDG construct, which was transformed into the *E. coli* expression strain BL21(DE3)plysS (Novagen Inc.) for protein expression. The enzyme was purified to >99% homogeneity as described.[6] The concentration of the enzyme was determined using the relationship $\epsilon^{280} = 38{,}511\ (\text{mol/L})^{-1}\ \text{cm}^{-1}$.[10]

The site-directed mutant Y19H was obtained from a random PCR error during the cloning of the wild-type enzyme and was expressed using the same system as the wild-type enzyme. The active-site mutants D64N and H187Q were generated using the QuikChange double-stranded mutagenesis kit obtained from Stratagene (La Jolla, CA). After confirming the mutations in these constructs by DNA sequencing of both strands, the mutant genes were cloned into the HindIII and NdeI sites of the pET28(a) expression vector (Novagen, Madison, WI), which places a six histidine tag at the amino terminus. The pET28(H187Q) and pET28(D64N) constructs were then transformed into the recombination deficient host BLR(DE3)plysS (Novagen). The expressed mutant enzymes were purified using a nickel chelate resin according to the manufacturer's instructions (Qiagen, Chatsworth, CA). Using this expression system ensured that the mutant enzymes were not contaminated with wild-type activity from the chromosomal copy of the UDG gene [Control experiments established that the native UDG is not retained on the nickel resin (<0.0002% retention). Because expression of the chromosomal UDG copy is less than 1/1000 that of the mutant enzymes on the basis of activity measurements of soluble extracts, then the overall fractional contamination of the mutant UDG preparations by wild-type UDG is about $1/10^8 = 1/[(1{,}000)(10^5)]$. This level of contamination is 10^4–fold less than the damaging effects of these mutations on the activity of UDG, and therefore cannot contribute to the measured activity of the mutant enzymes. Controls also demonstrated that the histidine tag has no discernible effect on the steady-state activity of the wild-type or mutant enzymes.

Biochemical Measurements of Wild-Type and Mutant UDG Enzymes

The synthesis and purification of the deoxyoligonucleotides used in the kinetic studies have been previously described.[6] The steady-state kinetics of uracil glycosidic bond cleavage in duplex DNA were monitored continuously at 25°C in TMN buffer. A continuous kinetic assay was used that monitored the fluorescence increase of a 2-aminopurine base (**P**) that was base paired to the excised uracil[6] (AUA/T**P**T, see sequence in Table I). The steady-state kinetic parameters were obtained from plots of the initial velocities against substrate concentration using a standard hyperbolic kinetic expression and the program Grafit.[11] Stopped-flow fluorescence single-turnover experiments were performed using a KinTek model SF-2001 apparatus (KinTek, Inc., University Park, PA) in the two-syringe mode.[7] The **P** base of the substrate was excited at 310 nm and fluorescence emission was monitored using a 70-nm bandpass centered at 400 nm. The data were fitted to a single exponential equation using the computer program Grafit.[11]

The pH dependence of k_{cat} for the wild-type UDG reaction was determined using the trinucleotide substrate 5′-pU**P**A-3′ which reacts by a rapid equilibrium mechanism, and the longer pentanucleotide substrate 5′-AU**P**AA-3′. The pH dependence of k_{cat} for the H187Q and D64N catalyzed reactions was determined using 5′-AU**P**AA-3′. The rates were measured using the fluorescence assay in the pH range 5.2–10.4 using a series of Good's buffers [T = 25°C, ionic strength = 0.03 M (NaCl)]. The k_{cat} data were fitted to the logarithmic form of equation $k_{cat} = k_{cat}\text{max}/(1 + [H^+]/K_a^{ES})$ to obtain pK_a^{ES}.[12] The dissociation constant for glycerol binding to UDG was determined by a displacement experiment in which the fluorescent trinucleotide substrate analog 5′-pU^F**P**A-3′ was displaced by glycerol (this analog contains a stable 2′-fluoro-2′deoxyuridine nucleotide[7] and binds to the UDG active site with a K_D = 1.2 μmol/L). The inhibition constant ($K_{0.5}$) for glycerol was estimated by using the program *Dynafit*, using the known dissociation constant for pU^F**P**A, and the chemical equations describing competitive binding of two ligands to a single site.[13]

TABLE I. Kinetic Parameters for Wild-Type, Y19H, D64N and H187Q eUDG[†]

Substrate or inhibitor	Parameter	Wild-type	Y19H	(Y19H/wt)[a]	D64N	(D64N/wt)[a]	H187Q	(H187Q/wt)[a]
AUA/T**P**T[b]	k_{cat} (s^{-1})	3.2	2.4	0.75	0.012	$10^{-2.4}$	0.02	$10^{-2.2}$
	k_{cat}/K_m (μmol^{-1} L s^{-1})	100	80	0.75	0.11	$10^{-3.0}$	0.58	$10^{-2.2}$
	K_m (μmol L^{-1})	0.03	0.03	1	0.11	3.6	0.03	1
	k_{max} (s^{-1})[c]	142	148	1	0.05	$10^{-3.5}$	0.019	$10^{-3.9}$
5′ AU**P**AA3′	k_{cat} (s^{-1})	25		—	0.019	$10^{-3.1}$	0.047	$10^{-2.7}$
	k_{cat}/K_m (μmol^{-1} L s^{-1})	62		—	0.021	$10^{-3.5}$	0.05	$10^{-3.1}$
	K_m (μmol L^{-1})	0.40		—	0.91	2.3	0.89	2.2
	k_{max} (s^{-1})[c]	145		—	0.018	$10^{-3.9}$	0.047	$10^{-3.5}$
	pK_a^{ES} [d]	6.2 ± 0.1		—	[e]		6.1 ± 0.2	
5′pU**P**A3′	k_{cat} (s^{-1})	0.19		—		—		—
	k_{cat}/K_m (μmol^{-1} L s^{-1})	0.15		—		—		—
	K_m (μmol L^{-1})	1.3		—		—		—
	k_{max} (s^{-1}[c]	0.2		—		—		—
	pK_a^{ES} [e]	6.6 ± 0.3		—		—		—

[†]T = 25°C, TMN buffer (pH = 8.0). Errors are <10% for k_{cat} and k_{max}, and <20% for K_m.
[a]The damaging effect of the mutation relative to the wild-type UDG (i.e., H187Q/*wt* = H187Q activity/wild-type activity).
[b]The substrate was a 19-mer double-stranded oligonucleotide (sequence of the uracil strand was 5′GCGCCAAAUAAAAAGCGC3′).[5]
[c]The single-turnover k_{max} values are the concentration-independent maximal single-turnover rates for the enzyme determined in a stopped-flow fluorescence assay.[6]
[d]The pK_a value determined from the pH dependence of k_{cat}.
[e]No essential pK_a was seen in the pH range 4.6– 10 with the D64N mutant. However, a gradual decrease in activity below pH ~5.5 was observed (the activity at pH 4.6 was only fivefold less than the maximal value at neutral pH).

TABLE II. X-Ray Data Statistics

		*e*UDG	*e*UDG-Hg[a]	*e*UDG	*e*UDG-uracil	Y19H *e*UDG-glycerol
Space group		$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell	a(Å)	55.11	55.13	54.59	53.83	54.85
	b(Å)	61.53	61.32	59.52	59.19	58.93
	c(Å)	64.54	64.40	64.17	63.91	63.99
Temperature °K		298	298	100	100	100
Resolution (Å) with $I \geq 1.5\sigma$		2.20	2.59	1.60	1.50	1.43
Number of unique reflections		11,628	7,194	28,254	32,361	39,441
Completeness (%)		100.0	92.2	97.9	97.0	90.0
Redundancy		6.9	5.3	4.2	5.0	4.9
R_{merge} (%)[b]		11	6.5	9.5	8.5	7.3

[a]*e*UDG-HgCl2. R_{iso} (= $\Sigma|I_{der} - I_{nat}|/\Sigma I_{nat}$) = 15%; Data range used (Å): 10.0–3.0; phasing power at 3.0 Å was 2.8; figure of merit = 0.5.
[b]$R_{merge} = \Sigma|I_{obs} - I_{avg}|/\Sigma I_{obs}$ where the summation is over all reflections.

Crystallization and Data Collection

*e*UDG was crystallized at 20°C by vapor diffusion in hanging drops of equal volumes of 14.9 mg/ml protein solution and well solution (0.2 M sodium acetate, 30% polyethylene glycol 4,000 and 0.1 M Tris buffer, pH = 8.5). The crystals typically grew to dimensions of 300 × 200 × 100 µm in 5 days. They belong to space group $P2_12_12_1$ (a = 54.7 Å, b = 59.7 Å, c = 64.4 Å) and contain one molecule per asymmetric unit. The crystal V_m of 2.05 daltons/Å^3 indicates a 39.9% solvent content. Crystals of the $HgCl_2$ derivative were obtained by soaking the protein crystals in stabilizing buffer containing 1.0 mM $HgCl_2$ for 6 hr. *e*UDG-uracil complex crystals were prepared by soaking mutant Y19H crystals in stabilizing buffer containing 6.0 mM uracil for 2 hr. Crystals of the *e*UDG-glycerol complex were obtained after soaking the mutant Y19H crystals in cryoprotectant solution containing 25% (by volume) glycerol.

Diffraction data were collected with a Bruker electronic area detector on a Bruker rotating anode X-ray generator. Two data sets were collected from separate crystals of the free *e*UDG at 293 K and at 100 K, respectively. The data for the $HgCl_2$ derivative was collected at 293 K. Both *e*UDG-glycerol and *e*UDG-uracil complex data were collected at 100 K. All data were indexed and processed using XENGEN suit of programs.[14] The diffraction data processing results are summarized in Table II.

Structure Solution and Refinement

Sequence alignments (Fig. 1) indicate that *e*UDG resembles *h*UDG more than *v*UDG. However, because only the coordinates for *v*UDG were available at the time, this structure was used as a probe using the molecular replacement method (MR). Although maps calculated with AMoRe[15] were reasonably good for regions with good sequence homology, attempts to solve the crystal structure

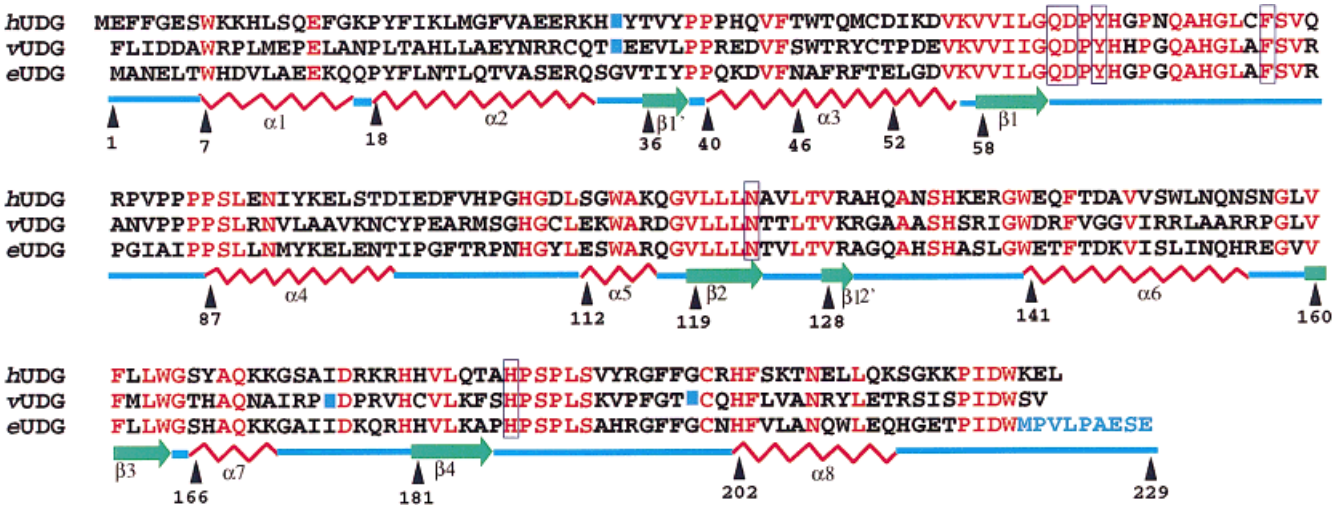

Fig. 1. Sequence alignments among *e*UDG, *v*UDG, and *h*UDG with major secondary structure elements indicated for *e*UDG (α–helices, zig–zag; strands, arrows). Residues in red are conserved among the three UDGs, and blue squares indicate amino acid insertion points. Conserved active-site residues are delineated by boxes in the sequence, and the residues comprising the C–terminal ""tail" in *e*UDG are indicated in blue.

of *e*UDG by the MR method using *v*UDG as a probe were not completely successful. This was because of missing electron density for the first 36 amino terminal residues, and uninterpretable electron density in regions that contained insertions or poor sequence homology. Nevertheless, a partial model was obtained by adjusting the main chains and building the side chains that were well defined in the electron density map using the program O.[16] This partial model was used to calculate the phases that were later combined with the phases derived from single isomorphous replacement with the anomalous scattering (SIRAS) method using the program PHASES.[17] The SIRAS phases were derived from a single $HgCl_2$ derivative. The heavy atom positions were located and refined using data between 10.0 and 3.0 Å resolution with PHASES. After incorporation of the phases from the anomalous scattering, the phasing power was 2.8 at 3.0-Å resolution with an overall figure of merit of 0.50. On one cycle of solvent flattening, most of the secondary structure was clear in the electron density map except for the first and second α helices.

The phases from MR and SIRAS methods were combined to generate a new electron-density map using PHASES. Further cycles of solvent flattening, model building, and refinement were carried out. At this point, the backbone electron density was clearly defined for the first 36 amino-terminal residues. In the final electron density map at this stage, the first four residues at the amino-terminus were still absent, and the backbone of the last seven carboxyl-terminal residues (223–229) and side chains of several residues were still poorly defined.

The model was then refined at 1.43-Å resolution against the data collected at 100 K for the Y19H mutant in complex with glycerol using X-PLOR.[18] After simulated annealing and positional refinements, the R_{free}[19] was 0.30, whereas the R (working set) was 0.26. The model was subsequently refined using data between 20.00 and 1.43-Å resolution with SHELX-97.[20] After 10 cycles of refinement, the electron density map was very well defined for the side-chains of the seven C–terminal residues, which extend out from the body of the molecule as a tail. All other missing side-chains were added to the model except for the four N-terminal residues. At this stage, water molecules were identified as 3σ peaks in the F_o-F_c map and included into the model.

The final coordinates of the Y19H *e*UDG-glycerol complex were used as a starting model for the refinements of wild-type *e*UDG and Y19H *e*UDG-uracil complex, after deleting both glycerol molecules. The structures were refined with SHELX-97. In the Y19H *e*UDG-uracil complex, the uracil molecule was clearly identified in the difference electron density map. Like the Y19H *e*UDG mutant structure, no electron density was visible for the four amino-terminal residues in either *e*UDG or its uracil complex structures. The refinement results for all the native *e*UDG, Y19H *e*UDG-uracil, and mutant Y19H-glycerol complexes are summarized in Table III.

RESULTS AND DISCUSSION

Biochemical Properties of Wild-Type and Mutant *e*UDG Enzymes

The *ung* gene was cloned from *E. coli* strain B genomic DNA using PCR methods, and it was inserted into a plasmid under control of the strong T7 polymerase promoter for overexpression.[6] The B strain sequence is identical to that published for UDG obtained from *E. coli* strain K12,[9] with the exception of the substitution of histidine for Arg213. Inspection of the *ung* sequences of seven species

Fig. 2. Stereoview of the electron density for (**a**) the active-site residues in free *e*UDG ($2F_o$-F_c map contoured at 1.5σ), (**b**) uracil in the Y19H *e*UDG–uracil complex (omit map contoured at 3σ), and (**c**) glycerol in the Y19H *e*UDG–glycerol complex (omit map contoured at 3σ).

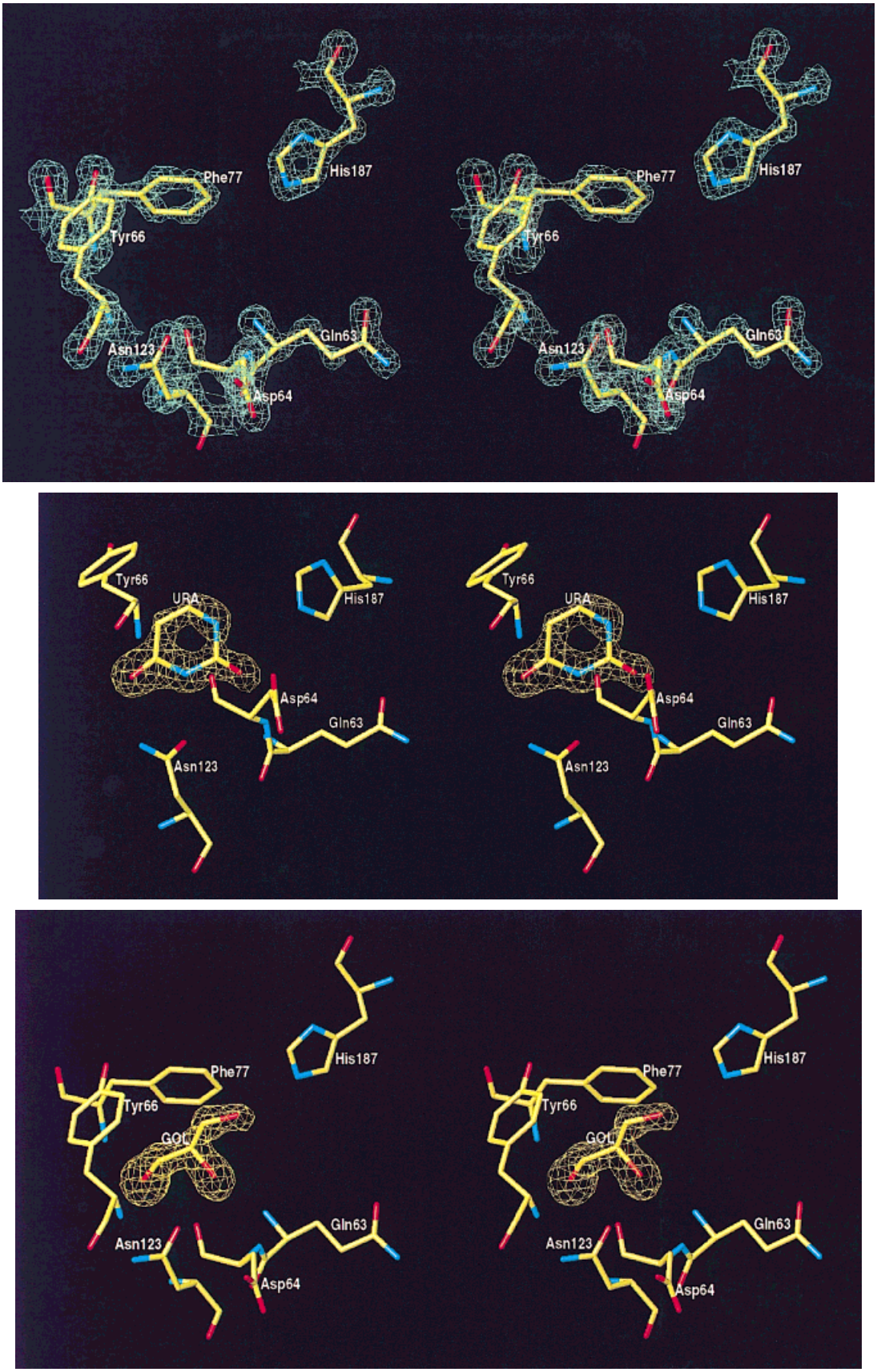
Phe77
His187
Tyr66
Asn123
Gln63
Asp64
Phe77
His187
Tyr66
Asn123
Gln63
Asp64
Tyr66
URA
His187
Asp64
Gln63
Asn123
Tyr66
URA
His187
Asp64
Gln63
Asn123
His187
Phe77
Tyr66
GOL
Asn123
Gln63
Asp64
His187
Phe77
Tyr66
GOL
Asn123
Gln63
Asp64

Figure 2.

TABLE III. Refinement Statistics

	*e*UDG (at 100 K)	Y19H *e*UDG-uracil (at 100 K)	Y19H *e*UDG-glycerol (at 100 K)
Resolution (Å)	1.60	1.50	1.43
Crystallographic R factor[a]	0.194	0.197	0.162
R_{free}	0.250	—	—
Number of residues	225	225	225
Water molecules	300	277	334
Substrate molecules	0	1 uracil	2 glycerol
RMS deviations (Å) from ideal			
Bond distances	0.008	0.008	0.009
Angle distances	0.024	0.025	0.026
Planarity	0.027	0.026	0.027

[a] $R = \Sigma|F_{obs} - F_{cal}| / \Sigma|F_{cal}|$, where F_{obs} and F_{cal} are the observed and calculated structure factors, respectively.

with a 33%–71% amino acid identity with the *E. coli* enzyme, showed no strict conservation of the residue corresponding to His213. Four of the seven sequences had a polar residue at this position (threonine, serine, lysine, or arginine), one had a methionine, and two had a histidine (*Haemophilus influenzae* and *Mycoplasma*). Thus, this sequence difference between the K12 and B strains is a naturally occurring variation at this position. Analysis by MALDI mass spectrometry revealed that the amino-terminal methionine was removed posttranslationally. The primary sequence of the wild-type UDG from *E. coli* strain B is compared with the human and viral enzymes in Figure 1.

The Y19H enzyme was generated by a random PCR error during the cloning of the wild-type enzyme and was overexpressed and purified as described for the wild-type enzyme. As described below, the highest-resolution diffraction data were obtained from crystals of Y19H UDG with glycerol bound in the active site. As shown in Table I, the Y19H and wild-type enzymes have indistinguishable kinetic properties in the steady-state assay, and a single-turnover stopped-flow fluorescence assay for glycosidic bond cleavage.[6,7]

The pH dependence of k_{cat} shows the requirement for a single basic group with a pKa in the range 6.2 to 6.6 depending on the substrate (Table I). This basic group is also apparent in the pH dependence of k_{cat} for the H187Q UDG, but is *absent* for the D64N enzyme (Table I). This result strongly suggests that the essential pK_a in the ES complex is due to general-base catalysis by Asp64, and not a catalytic histidine general-base (i.e., His187) as previously suggested for the human enzyme.[2] The pH dependence of k_{cat} provides no evidence for a general-acid catalyst (i. e., there is no descending limb at high pH values indicating the presence of an essential protonated group). This result indicates that either general-acid catalysis is not occurring, the pK_a of the general-acid is outside the pH range studied, or that the general-acid group is not in equilibrium with solvent protons during the lifetime of the enzyme-bound complexes.

The D64N and H187Q mutations show as much as $10^{-3.1}$ and $10^{-2.7}$-fold decreases in k_{cat}, respectively, but little effect on K_m, indicating that these groups stabilize the transition state (Table I). Surprisingly, much larger damaging effects of $10^{-3.9}$ and $10^{-3.5}$-fold are seen for the D64N and H187Q mutants when the single-turnover glycosidic bond cleavage reactions are compared with that of the wild-type UDG (see k_{max} values in Table I). These larger effects are caused by a change in rate-limiting step for the wild-type enzyme, from product dissociation in the k_{cat} measurements, to the glycosidic bond cleavage step in the single-turnover reaction. In contrast, the rates for the mutant enzymes are limited by glycosidic bond cleavage in *both* the steady-state k_{cat} and single-turnover measurements (Table I). These results emphasize the importance of making single-turnover measurements when evaluating the damaging effects of mutations.

The Overall Structure

The crystal structures of the wild-type *e*UDG, Y19H *e*UDG-uracil, and Y19H *e*UDG- glycerol complexes have been determined and refined at 1.60 Å, 1.50 Å, and 1.43 Å resolution, respectively. The final coordinates of all three structures consist of 225 amino acid residues and lack the four amino-terminal residues. The Ramachandran plots of the three structures indicate that 91.5% of the residues are in its most favored region, with none in the disallowed region. The 2 F_o-F_c electron density maps show good agreement with all three structures. Representative maps corresponding to the active site of the free *e*UDG, Y19H *e*UDG-uracil, and Y19H *e*UDG-glycerol complex structures are shown in Figure 2. The coordinates and structure factors have been deposited for free *e*UDG, its Y19H *e*UDG-uracil, and Y19H *e*UDG-glycerol complexes, respectively, in Protein Data Bank[21] as entries 1EUG, 2EUG, and 3EUG.

Structural Comparisons of the *E. coli*, Human, and Viral UDG Enzymes

The secondary structure of *e*UDG is defined according to Kabsch and Sander,[22] and is shown as a RASTER 3D model[23] in Figure 3a. The main secondary structural features are summarized in Figure 1.

The root mean square deviation (RMSD) values obtained from aligning the C_α positions of the *e*UDG-*v*UDG and *e*UDG-*h*UDG structures are 1.3 Å and 0.9 Å, respectively.

Although the overall fold of the three UDGs is similar, significant structural differences are observed in regions with insertions, poor sequence homology, as well as the active site. The first difference occurs in the first and second α helices because of the insertion of Gly34 in *e*UDG. This insertion results in a half-turn and quarter-turn shift in these helices relative to those in *v*UDG and *h*UDG, respectively (Fig. 3b). This structural difference propagates several residues beyond Gly34, after which the three UDG structures start to converge. The second difference is

associated with the insertion of Ile175 in the loop preceding β-strand four of *e*UDG and *h*UDG structures. This insertion results in Cα shifts of up to 2.2 Å in this region as compared with the *v*UDG structure. Finally, backbone shifts as large as 6.4 Å are seen because of the insertion of residue Gly199 in the loop preceding helix 8 of the *e*UDG and *h*UDG structures as compared with that of *v*UDG.

In addition to the above structural changes associated with insertions, large differences in the backbone (>5 Å) are also found in nonconserved regions such as the loop containing residues 100–111 (Fig. 3b), and in helix 3 around residue Thr51 of vUDG (which is a cysteine in both *v*UDG and *h*UDG). In addition, the last six to seven C-terminal residues in *e*UDG go beyond the C-termini in *v*UDG and *h*UDG and reach out as an extended and unique tail. This tail is involved in crystal packing, and may contribute to the high-quality diffraction data obtained for *e*UDG using an in-house X-ray source.

Structural Comparisons at the Active Site

A uracil-recognition pocket is formed at the active site of *e*UDG by residues Gln63, Asp64, Tyr66, Phe77, Asn123, and His187 as shown in Figure 4. These residues are strictly conserved in all three UDG enzymes with C_α RMSD values of 0.44 Å and 0.35 Å when *e*UDG is compared with *v*UDG and *h*UDG, respectively. Although the differences in the C_α positions of these residues are small, significant conformational differences are observed between the side chains of the residues corresponding to Gln63, Asp64, Phe77, and His187 (Fig. 4). The side-chain conformation of Gln63 is similar in *e*UDG and *h*UDG but is rotated about 100° around the C_α-C_β bond in free *v*UDG. This conformation of Gln63 is not close to any of the preferred rotamers, because of a 2.5-Å H-bond between its side-chain carbonyl and the side-chain $N^\delta H$ of a His residue (His167 in *e*UDG). In the structures of *v*UDG with bound uracil or 5′-pdTdTdT-OH-3′, the conformation of Gln63 changes to become very similar to that observed in *e*UDG and *h*UDG. Thus, for *v*UDG, this conserved residue appears to change its side-chain conformation upon uracil or nucleotide binding.

For the putative active-site base in *e*UDG (Asp64), differences in both the side-chain and main-chain conformation are observed when compared with *v*UDG (Fig. 4). In *e*UDG *and h*UDG, Asp64 is turned away from the active site and is locked in position by a 2.5-Å hydrogen bond between its carboxyl oxygen and the backbone amide of His134. The other side-chain carboxyl oxygen of Asp64 is hydrogen bonded to a water molecule that in turn interacts with three other residues. Thus, the energetic penalty for disruption of this intramolecular hydrogen bond network must be paid for on productive DNA binding, allowing Asp64 to rotate ~120° into its catalytic position as seen in the *h*UDG -product complexes.[4,5]

Backbone Shifts in the Active-site Loop Containing His187 and Leu191

On the basis of crystal structures of hUDG-abasic DNA product complexes, damage site recognition by UDG has been suggested to proceed by a “”push-pull” uracil flipping mechanism.[4,5] In the push part of this mechanism, the loop containing a conserved leucine residue (Leu191 in *e*UDG) moves about 2 Å on productive binding to uracil-containing DNA. This movement allows the leucine to penetrate the DNA minor groove, and push the uracil base into the active-site pocket. In the pull component, this movement repositions the active-site histidine (His187 of *e*UDG) to directly interact with uracil O2.[4,5] This loop movement appears to be an important aspect of uracil-specific recognition because it is absent in *v*UDG-trinucleotide, *h*UDG 6-aminouracil, and *h*UDG-UDG inhibitor protein complexes. In contrast, this loop appears to be preorganized in Y19H *e*UDG-uracil complex, where the C_α of Leu191 is 1.2 Å and 1.3 Å closer to the active site, and the N ϵ of His187 is 1.1 Å closer to O2 of uracil, as compared with the *v*UDG and *h*UDG structures, respectively. Since the active-site residues in free *e*UDG and *v*UDG agree very well with their respective uracil complexes, the latter are used in Figure 5 to illustrate the more closed loop position in *e*UDG with respect to *v*UDG and *h*UDG. The absence of strong crystal-packing interactions involving this loop, and its relatively high B factors, suggests that these conformational differences are real, and that the loop is fairly rigid. Thus, the average conformation of this loop in *e*UDG may be more closed, and require a smaller conformational change to interact productively with the bound DNA and uracil. This structural result is consistent with the faster DNA association kinetics of *e*UDG as compared with the viral enzyme (see below).

Correlation of Structure and Substrate Specificity

UDG shows exquisite specificity for cleavage of the *N*-glycosidic bond of deoxyuridine in DNA. Indeed, recent biochemical and kinetic studies on the *E. coli* UDG[6,7] show that the enzyme has a specificity ratio $(k_{cat}/K_m)^{sp}/(k_{cat}/K_m)^{nonsp} \geq 10^6$. By using noncleavable 2′-fluorodeoxyuridine substituted substrate analogs in combination with 2-aminopurine fluorescent reporter groups, it has been shown that site-specific recognition occurs in two steps.[7] The first step involves diffusion-controlled formation of a weak nonspecific encounter complex with the DNA (k_{on} = 300 μmol^{-1} L s^{-1}, K_D = 1.2 μmol/L). The second step involves a fast concerted conformational change in the DNA and enzyme (k_{conf} = 1,300 s^{-1}) that leads to a stable extrahelical uracil base. The conformational change in UDG is absolutely specific for deoxyuridine-containing (or 2′-fluorodeoxyuridine-containing) DNA, and contributes at least 10^4-fold to the overall specificity of the UDG reaction. These kinetic data provide direct support for enzyme-assisted uracil flipping, and exclude a mechanism involving enzymatic capture of an extrahelical uracil base in the free substrate.[7] Recent structural and kinetic studies on *h*UDG also support this conclusion.[5] However, these conclusions for the bacterial and human enzymes differ considerably from recent studies of the substrate recognition mechanism of the viral UDG.[8] In this work it was concluded that the viral enzyme binds the uracil in DNA directly in a simple one-step binding reaction with associa-

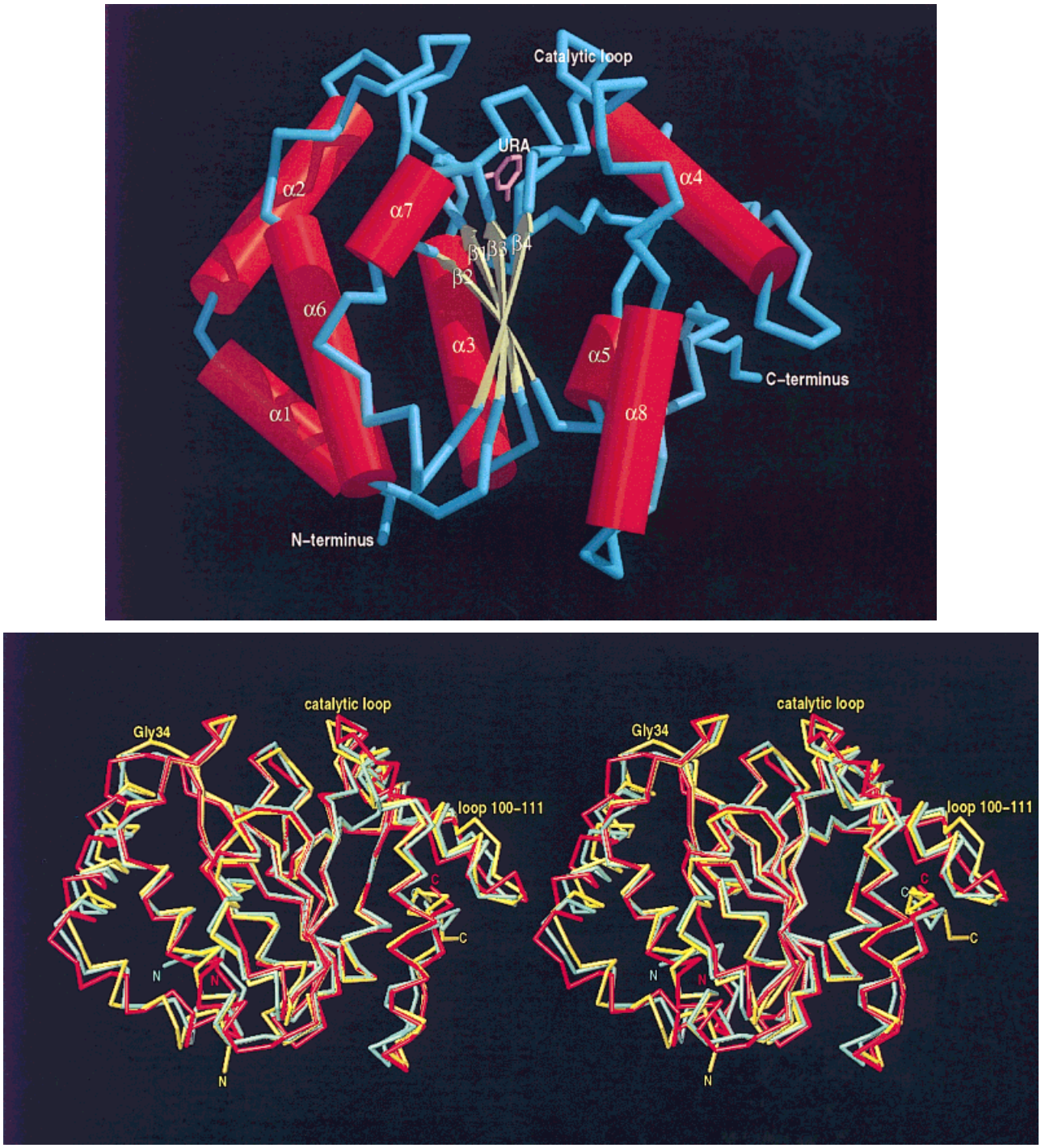

Fig. 3. **a:** Raster three-dimensional image[22] for wild-type *e*UDG with secondary structures labeled. **b:** Stereoview of the C_α trace of *e*UDG (in yellow) aligned with those of *v*UDG (1UDG in red, RMSD 1.3 A, 206 target pairs) and *h*UDG (1AKZ in green, RMSD. 0.9 A, 213 target pairs). The figure was prepared using TURBO–FRODO.[35]

tion rate constants 100–10,000-fold less than those measured for the *E. coli* enzyme.[7] These slower association rates could in part reflect the larger conformational changes in the viral enzyme that are required to adjust the side-chain positions of Gln63 and Asp64 (Fig. 4), and the larger loop movement that is required to place His187 and Leu191 in catalytically active positions (Fig. 5).

The structure of the Y19H *e*UDG-uracil binary complex reveals the specific interactions that stabilize the flipped-out uracil base (Fig. 4). The side-chain amide group of

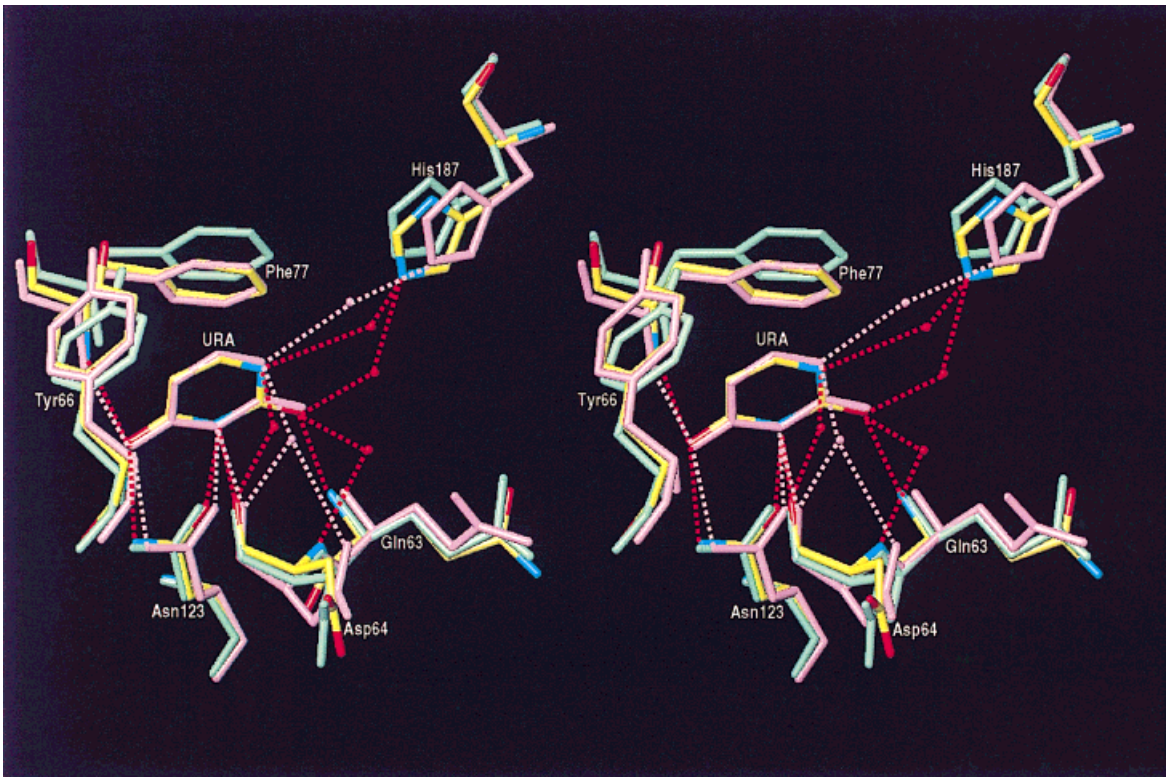

Fig. 4. Stereoview of the active-site residues in the Y19H *e*UDG–uracil complex (colored by atom type) aligned with the *v*UDG-uracil complex (pink) and free *h*UDG (green). The figure was prepared using TURBO–FRODO.[35]

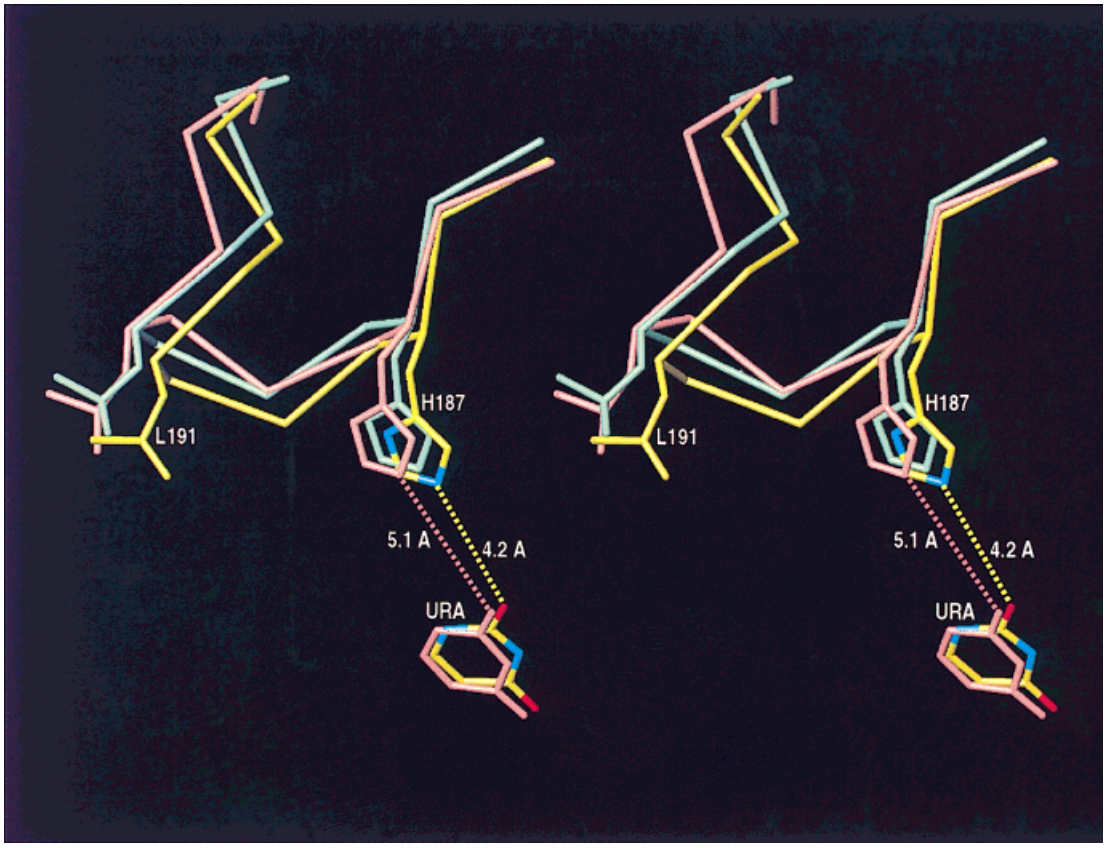

Fig. 5. Stereoview of the shifts in the loop carrying H187 and Leu191 in the Y19H *e*UDG–uracil complex (colored by atom type) as compared with free *v*UDG–uracil (in pink) and *h*UDG (in green). The figure was prepared using TURBO–FRODO.[35]

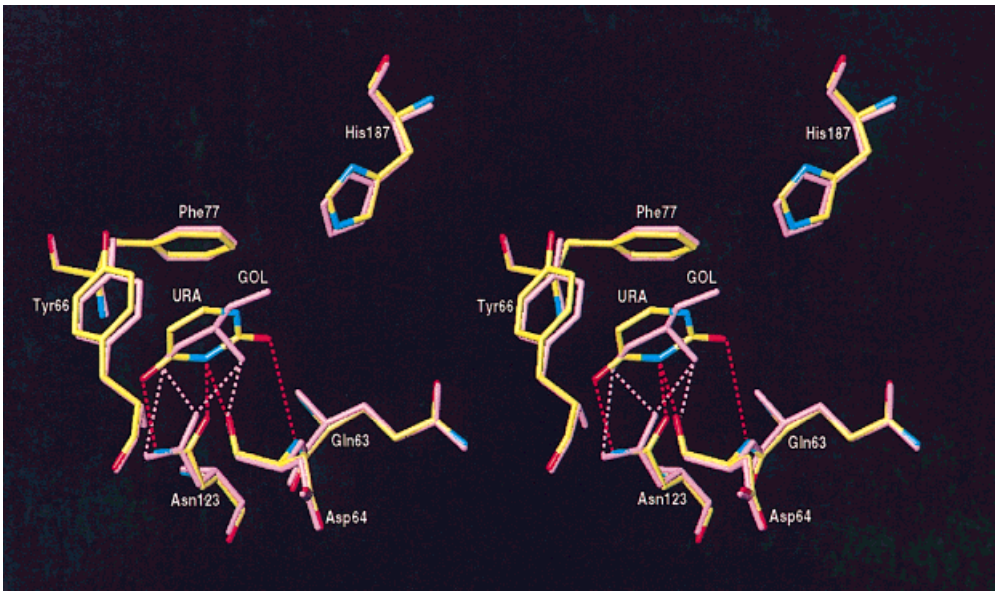

Fig. 6. Stereoview of the interactions of glycerol with the active-site residues in the Y19H *e*UDG–glycerol complex (colored by atom type) as compared with those of uracil (pink). The figure was prepared using TURBO–FRODO.[35]

Asn123, the backbone amide proton of Gln63, and the backbone carbonyl of Asp64 all show hydrogen bonding interactions with uracil O4 and N3. In addition, the imidazole side-chain of the conserved active-site residue His187 forms a hydrogen bond to uracil O2 through a bridging water molecule (Fig. 4). Three active-site waters are displaced from the active site of the free enzyme upon uracil binding, suggesting that solvent displacement provides a favorable entropic component to the free energy of binding. As discussed previously, the bound uracil base is stacked against a conserved phenylalanine residue (Phe77) and binds orthogonally to the ring plane of a conserved tyrosine (Tyr66). The proximity of Tyr-66 to the C5 position of uracil contributes to specificity by sterically excluding the 5-methyl group of thymidine from the binding site.[2–5] These aromatic interactions with the uracil base are also seen in *v*UDG-uracil complex and *h*UDG complex ternary product and may contribute to catalysis by exclusion of bulk solvent from the active site. In addition, a similar arrangement of aromatic residues is seen in the active sites of enzymes with related functions such as the phosphoribosyl transferases[24,25] and inosine-uridine nucleoside *N*-ribohydrolase.[26] Thus, this mode of base binding has been conserved throughout the evolution of many enzymes that carry out similar reactions.

We found that when a cryoprotectant containing glycerol was used for obtaining crystals of *e*UDG, a molecule of glycerol was bound in the uracil binding pocket (Fig. 2b). Accordingly, kinetic inhibition studies using the AU**P**AA substrate show that 200 mM glycerol inhibits the reaction rate by ~50% (for [S] = K_m). Remarkably, many of the interactions seen in the *e*UDG-glycerol complex are also seen in the Y19H *e*UDG-uracil complex (Fig. 6). The three hydroxyl groups of glycerol are seen to mimic the interactions of uracil O2, O4, and N3 with the enzyme described above, and replace the three water molecules seen in the free enzyme structure. The mechanism of glycerol binding even includes the same through-water hydrogen bond to the imidazole side-chain of the conserved His187.

Mechanism of Glycosidic Bond Cleavage

The detailed mechanism by which the *N*-glycosidic bond in DNA is cleaved by enzymes is largely unknown. This is especially true for the hydrolysis of the glycosidic bond in *pyrimidine* nucleotides, because much of the extant mechanistic work has focused on the hydrolysis of *purine* ribonucleosides. It can be estimated that UDG enhances the spontaneous rate by a factor of ~10^{12}-fold and has a catalytic proficiency of 10^{18} mol $^{-1}$ L.[7] These large factors make UDG one of the most powerful *N*-glycosylases yet investigated.

A key question is the nature of the chemical interactions in the UDG active site that promote this tremendous catalytic power. Many enzymatic glycosyl transfer reactions are thought to proceed through oxycarbenium-ion transition-states (Fig. 7).[28,29] This transition-state is often stabilized by protonation of the leaving group base,[30] or by geometric and electronic distortion of sugar ring that favors the oxycarbenium-ion transition state.[31] In addition, deprotonation of the incoming water nucleophile by an active-site carboxylate is a common feature of these reactions.[28,31] Thus, it is of interest to ask whether the chemistry and architecture of the UDG active site is consistent with this type of transition state.

Although Asp64 is pointed away from the active-site pocket in the free enzyme and the uracil complex, it could easily serve as the basic group for deprotonation of water by simply rotating its sidechain carboxylate towards the

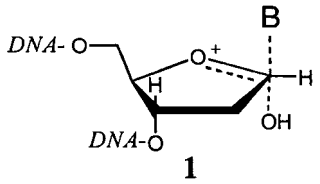

Fig. 7. Hypothetical oxycarbenium-ion transition-state for glycosidic bond hydrolysis in DNA.

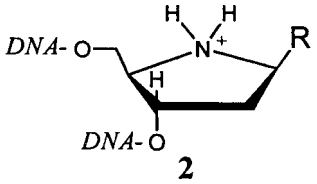

Fig. 8. Pyrrolidine-based transition-state analog for DNA glycosylases.

active-site pocket on substrate binding, as observed in the human UDG-DNA complex.[4,5] Likewise, with minimal movement on substrate binding, His187 could directly donate a proton or hydrogen bond to uracil O2, thereby assisting leaving group departure, as observed in the human UDG-DNA complex.[4,5] Alternatively, this interaction may occur through a bridging water molecule as indicated in the uracil-*e*UDG complex (Fig. 4). High-resolution crystal structures of *e*UDG bound to nonreactive deoxyuridine-containing substrate analogs will be required to definitively answer these questions.

The essential pK_a ~ 6.4 in the ES complex for the wild-type enzyme, and its disappearance in the D64N mutant, supports the structural evidence implicating Asp64 as the general base catalyst. In contrast, these pH studies provide no evidence for general acid catalysis, implying that the p*K*a for His187 may be outside the range studied, that the protonation equilibrium for this group is frozen while the substrate and products are present, or that general-acid catalysis is not important.

The unusually high pK_a value for Asp64, and the absence of a second pK_a in the ES complex suggests that Asp64 and His187 are in unique environments. In the free enzyme, Asp64 is located in a hydrophobic niche with its side-chain carboxylate turned away from solvent and hydrogen bonded to the backbone amide proton of His134 (2.5 Å). Thus, this short hydrogen bond to His134 indicates that it may be difficult for solvent protons to access the side chain carboxylate of Asp64 in the free enzyme. From the structure of the human UDG-DNA complex, Asp64 rotates by 120° from its position in the free enzyme. Nevertheless, it is still in a hydrophobic environment produced by the productively bound DNA, and a "lid" created by residues 125–141 in the loop connecting β strand 2 and helix 6. In addition, the close proximity of the negatively charged phosphodiester groups of the deoxyuridine could also play a role in increasing the pK_a value for Asp64.

His187 is also located in a pocket lined with hydrophobic residues, with no anionic groups or helix dipoles in the near vicinity to stabilize its cationic form, suggesting that the pK_a for H187 in the free enzyme is much lower than a solvent-exposed histidine (pK_a ~6.5). Consistent with a neutral ionization state for His187 in the free enzyme, the N^{δ} nitrogen of His187 is hydrogen bonded to the backbone amide proton of Ser189 in the crystal at pH 8.5. Because k_{cat} for the UDG reaction is pH *independent* in the range 7–10 (Table I), then the pK_a of His187 must *increase* considerably on productive substrate binding if it is to act as a cationic general acid. To rigorously address these issues, the ionization states for Asp64 and His187 are currently being investigated by the direct method of heteronuclear NMR spectroscopy.

Does UDG stabilize an oxycarbenium ion transition state? In this transition state, a substantial positive charge is accumulated at O4′ and C1′. Accordingly, positively charged pyrrolidine-based transition-state analogues have been found to bind tightly to a number of DNA glycosylases, and related *N*-ribohydrolases (Fig. 8, R = base, H).[31,32] In general, an electrostatic interaction or hydrogen bond from the enzyme to the imino nitrogen (or O4 in the true transition state) is thought to be an important component for stabilization. Indeed, a conserved aspartic acid is believed to play this role in the helix-hairpin-helix superfamily DNA glycosylases,[33] and the crystal structure of inosine-uridine nucleoside hydrolase bound to a pyrrolidine-based transition-state inhibitor indicates that the 5′-hydroxyl of the inhibitor serves this function.[31] However, recent studies have shown that *e*UDG does *not* bind tightly to a pyrrolidine abasic site analog,[34] and there is no candidate electron-donating group in the active site of the bacterial enzyme studied here, or the human enzyme bound to the abasic DNA product,[4] that supports this mechanism of transition-state stabilization. The transition-state structure for the UDG reaction is currently being addressed directly using isotope effect measurements.

CONCLUSION

The structures of prokaryotic, viral, and human UDG enzymes are remarkably well conserved. However, significant differences in the side-chain conformations of key active residues Asp64 and Gln63, and the loop region containing His187 and Leu191, may reflect significant differences in the uracil recognition mechanisms used by the viral and bacterial enzymes. In general, the *E. coli* enzyme appears to have a preorganized active site, requiring smaller conformational changes to place active-site groups in position for catalysis, which is consistent with its rapid association kinetics. Structural and mutagenesis studies, as well as the pH dependence of k_{cat}, suggest that the catalytic mechanism used by UDG involves deprotonation of water by a high pK_a Asp64. In contrast, the pH dependence of k_{cat} provides *no* evidence for a general-acid catalyst over the pH range 5.5–9.5. This result suggests that general-acid catalysis is not occurring, or that the pK_a for the putative general-acid (His187) is outside of this pH range. Assuming additivity of free energies, the damaging

effects of the Asp64 and His187 mutations account for *at most* $10^{7.4}$-fold of the estimated 10^{12}-fold rate enhancement of UDG. Thus, much remains to be understood about the catalytic power of this impressive enzyme.

ACKNOWLEDGMENTS

This work was supported by the National Institute for Standards and Technology and National Institutes of Health grant GM56834 (J.T.S.)

REFERENCES

1. Mosbaugh DW, Bennett SE. Uracil-excision DNA repair. Prog Nucleic Acid Res Mol Biol 1994;48:315–371.
2. Mol CD, Arvai AS, Slupphaug G, et al. Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. Cell 1995;80:869–878.
3. Savva R, McAuley-Hecht K, Brown T, Pearl L. The structural basis of specific base-excision repair by uracil-DNA glycosylase. Nature 1995;373:487–493.
4. Slupphaug G, Mol CD, Kavli B, Arvai AS, Krokan HE, Tainer JA. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature 1996;384:87–92.
5. Parikh SS, Mol CD, Slupphaug G, Bharti S, Krokan HE, Tainer JA. Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. EMBO J 1998;17:5214–5226.
6. Stivers JT. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. Nucleic Acids Res 1998;26:3837–3844.
7. Stivers JT, Pankiewicz KW, Watanabe KA. Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. Biochemistry 1999;38:in press.
8. Panayotou G, Brown T, Barlow T, Pearl LH, Savva R. Direct measurement of the substrate preference of uracil-DNA glycosylase. J Biol Chem 1998;273:45–50.
9. Varshney U, Hutcheon T, van de Sande JH. Sequence analysis, expression, and conservation of Escherichia coli uracil DNA glycosylase and its gene (ung). J Biol Chem 1988;263:7776–7784.
10. Lindahl T, Ljungquist S, Siegert W, Nyberg B, Sperens B. DNA N-glycosidases. Properties of uracil-DNA glycosidase from Escherichia coli. J Biol Chem 1977;252:3286–3294.
11. Leatherbarrow RJ. GraFit Version 3.0. Staines, U.K.: Erithacus Software Ltd.; 1992.
12. Segel IH. Enzyme kinetics. New York: John Wiley & Sons Inc.; 1975. p 898–902.
13. Kuzmic P. Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. Anal Biochem 1996;237:260–273.
14. Howard AJ, Gilliland GL, Finzel BC, Poulos TL, Ohlendorf DH, Salemme FR. Use of an imaging proportional counter in macromolecular crystallography. J Appl Crystallog 1987;20:383–387.
15. Navaza J. AmoRe: automated package for molecular replacement. Acta Crystallogr Sect A 1994;50:157–163.
16. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and location of errors in these models. Acta Crystallog Sect A 1991;47:110–119.
17. Furey W, Swaminathan S. PHASES-95: a program package for the processing and analysis of diffraction data from macromolecules. In: Carter C, Sweet R, editors. Macromolecular crystallography (volume 277 of Methods in enzymology). Orlando: Academic Press; 1995. p 590–620.
18. Brünger AT. X-PLOR version 3.1—a system for x-ray crystallography and NMR. New Haven: Yale University Press; 1992.
19. Brünger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992;355:472–475.
20. Sheldrick GM. SHELX-97, program for the solution of crystal structures. University of Gottingen, Germany; 1997.
21. Bernstein FC, Koetzle TF, Williams GJB, et al. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
22. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. Biopolymers 1983;22:2577–2637.
23. Bacon D, Anderson WA. A fast algorithm for rendering space-filling molecule pictures. J Mol Graphics 1988;6:219–220.
24. Scapin G, Grubmeyer C, Sacchettini JC. Crystal structure of orotate phosphoribosyltransferase. Biochemistry 1994;33:1287–1294.
25. Eads JC, Scapin G, Xu Y, Grubmeyer C, Sacchetini JC. The crystal structure of human hypoxantine-guanine phosphoribosyltransferase with bound GMP. Cell 1994;78:325–334.
26. Degano M, Gopaul DN, Scapin G, Schramm VL, Sacchettini JC. Three-dimensional structure of the inosine-uridine nucleoside N-ribohydrolase from Crithidia fasciculata. Biochemistry 1996;35:5971–5981.
27. Varshney U, van de Sande JH. Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by Escherichia coli uracil DNA glycosylase. Biochemistry 1991;30:4055–4061.
28. Sinnott ML. Catalytic mechanisms of enzymic glycosyl transfer. Chem Rev 1990;90:1171–1202.
29. Schramm VL, Horenstein BA, Kline PC. Transition state analysis and inhibitor design for enzymatic reactions. J Biol Chem 1994;28:18259–18262.
30. Mazzella LJ, Parkin DW, Tyler PC, Furneaux RH, Schramm VL. Mechanistic diagnoses of N-ribohydrolases and purine nucleoside phosphorylase. J Am Chem Soc 1996;118:2111–2112.
31. Degamo M, Almo SC, Sacchettini JC, Schramm VL. Trypanosomal nucleoside hydrolase. A novel mechanism from the structure with a transition-state inhibitor. Biochemistry 1998;37:6277–6285.
32. Deng L, Schärer OD, Verdine GL. Unusually strong binding of a designed transition-state analog to a base-excision DNA repair protein. J Am Chem Soc 1997;119:7865–7866.
33. Nash HM, Bruner SD, Schärer OD, et al. Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA-repair protein superfamily. Curr Biol 1996;6:968–980.
34. Schärer OD, Nash HW, Jiricny J, Laval J, Verdine GL. Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases. J Biol Chem 1998;272:8592–8597.
35. Roussel A, Inisan A-G, Cambillau C. TURBO-FRODO. Marseille, France: AFMB and Bio Graphics; 1996.