Pocketing the difference: structures of uracil excision repair proteins

The recognition of abnormal DNA structure by proteins is fundamentally different from sequence-specific DNA binding. Recent crystal structures of uracil-DNA glycosylases show a novel uracil-binding pocket that explains the enzyme's selectivity for uracil-containing DNA and provide a structural basis for exploring the catalytic mechanism of amino–glycosylic bond cleavage.

Chemistry & Biology June 1995, **2**:351–354

The conservation of genetic information in successive cell generations depends on accurate DNA synthesis and the equally precise repair of DNA damage caused by spontaneous hydrolytic decay of bases or exposure to environmental toxins [1]. Although DNA is replicated *in vivo* with remarkable speed and accuracy, occasional synthesis errors create mutations that may have disastrous consequences if left uncorrected. Furthermore, DNA is chemically reactive and subject to damage by oxidants, ultraviolet radiation, or alkylating agents including the endogenous methyl donor S-adenosylmethionine. The integrity of the genome is protected by two lines of defense. Many DNA polymerases have a 3'-to-5' exonuclease activity that removes misincorporated bases, providing for immediate correction of errors in DNA synthesis. Mismatched bases escaping this proofreading activity and chemically modified bases are subject to repair by lesion-specific pathways. Defects in DNA repair are associated with a number of human diseases, including ataxia telangiectasia, xeroderma pigmentosum, Cockayne syndrome, Fanconi's anemia and hereditary nonpolyposis colon cancer [2,3]. The significance of DNA repair to human health and recent discoveries of DNA repair proteins with roles in transcription [4–6] and cell cycle control [7] have heightened interest in enzymes catalyzing the repair of damaged DNA. Recent crystal structures of two uracil-DNA glycosylases complexed with substrate analogs [8,9] show an unprecedented strategy for the recognition of abnormal bases in DNA. The structures reveal features that are likely to be conserved in other base excision repair enzymes.

Uracil is a normal constituent of RNA, which is absent from DNA primarily because low cellular levels of deoxyuridine triphosphate (dUTP) limit its incorporation by most DNA polymerases. Cellular pools of dUTP are held in check by UTPase, which converts dUTP to the dUMP substrate for thymidylate synthase. Inhibition of either UTPase activity or thymidylate synthase activity increases the intracellular ratio of dUTP to dTTP, leading to greatly exaggerated uridine incorporation in DNA and chromosome fragmentation, the hallmarks of 'thymineless death' [10]. Deoxyuridine lacks only the C5 methyl group of thymidine and it is readily incorporated by most DNA polymerases, creating occasional U:A mismatches during normal cell growth. A second source of uracil in DNA is the deamination of deoxycytosine, a reaction that

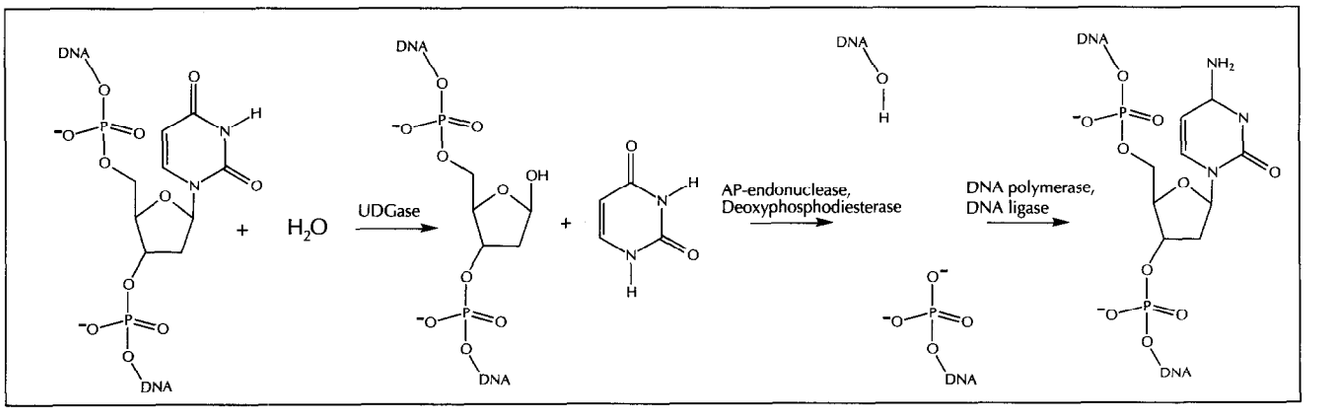

Fig. 1. Pathway of uracil-DNA repair. Uracil-DNA glycosylase (UDGase) specifically cleaves the glycosylic bond joining uracil to deoxyribose, creating an abasic site. The first step in repair is endonucleolytic cleavage of the DNA backbone. The repair is then completed by DNA polymerase and DNA ligase. AP, apurinic/apyrimidinic. Reprinted with permission from [8]

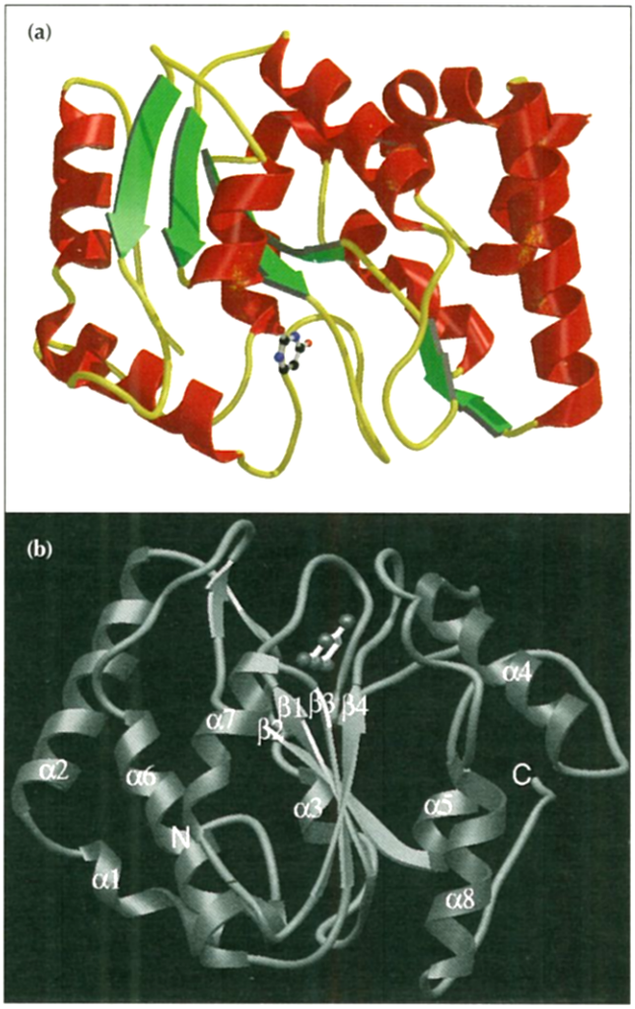

Fig. 2. Overall structures of **(a)** herpes simplex virus UDGase and **(b)** human UDGase. Both are mixed α–β proteins with a central parallel four-stranded β-sheet surrounded by α-helices. The active site, located at the carboxy-terminal end of the β-sheet, is characterized by a groove rich in basic amino acids and a pocket that accepts the uridine targeted for cleavage. This arrangement is consistent with an extrahelical orientation of the target uridine in DNA substrates. Reprinted with permission from [8] and [9], respectively.

is stimulated by exposure of cells to ionizing radiation or oxidants [1,11]. The resulting uridine is mispaired with guanine in a potentially mutagenic DNA lesion.

Removal of uracil in either DNA context is accomplished by a base excision repair pathway that is initiated by a uracil-specific DNA glycosylase (UDGase). UDGase cleaves the amino–glycosylic bond between uracil and deoxyribose, releasing the offending base from DNA (Fig. 1). Repair of the resulting abasic site is completed by the sequential actions of a site-specific endonuclease, a DNA polymerase and a DNA ligase. UDGase selectively excises uridine from DNA, yet it is inactive towards thymine in DNA or uridine in RNA. This exquisite substrate selectivity suggests that UDGase activity requires the unique pattern of exocyclic groups of the deoxyuridine target base. The crystal structures of uracil-DNA glycosylases from herpes simplex virus type-1 (HSV-1) [8] and human [9] provide a structural basis for this substrate specificity and offer clues that may lead to an understanding of the mechanism of enzyme-catalyzed glycosylic bond cleavage.

The amino-acid sequences of human and HSV-1 UDGases are 39 % identical, and both proteins have a mixed α–β structure consisting of a parallel, four-stranded β-sheet flanked by eight (human) or ten (HSV-1) α-helices (Fig. 2). Loops extending from the carboxyl termini of two β-strands define a uracil-binding pocket that was experimentally identified by soaking the crystals with 6-aminouracil or uracil (Fig. 3). Residues surrounding the uracil-binding pocket are highly conserved among eukaryotic and prokaryotic UDGases, and many of these are required for catalytic activity [9]. The uracil-binding pocket lies in a groove rich in basic amino acids that form a DNA-docking surface. Deoxythymidine trinucleotide binds in the groove of the HSV-1 UDGase with the middle thymine draping across the uridine-binding pocket [8]. UDGases are more active towards single-stranded DNA substrates than towards duplex DNA. This may be because of the narrow width of the DNA-binding groove, which suggests that duplex DNA contacts the enzyme active site primarily with one strand.

The geometry of the UDGase active site requires the target DNA base to adopt an extrahelical orientation for binding in the substrate pocket. This orientation of the substrate gives the enzyme complete access to the functional groups of uracil in any DNA sequence context and exposes the glycosylic bond for cleavage. A precedent for the extrusion of the target base by a DNA modification enzyme is found in the structure of *Hha*I methyltransferase bound to a DNA-based inhibitor [12]. Uracil is distinguished from other naturally occurring pyrimidines by the presence of exocyclic oxygens O4 and O2, and the absence of a substituent at C5. UDGase recognizes these features through an extensive set of protein–base interactions within the uracil-binding pocket (Fig. 3). A conserved Phe residue lying against one face of the pyrimidine ring forms the floor of the uracil-binding pocket. A nearby, conserved Tyr stacks against C5 on the edge of the pyrimidine ring, effectively blocking access to thymine or other 5-substituted pyrimidines. Cytosine binding is disfavored by electrostatic repulsion between its N4 and nearby polar residues in the uridine-binding pocket. Purines are similarly excluded from the pocket because of their bulky imidazole ring. The orientation of the conserved Tyr is fixed by van der Waals contacts with neighboring conserved residues and by a hydrogen bond to a nearby peptide carbonyl oxygen. Uracil's exocyclic oxygens participate in hydrogen-bonding interactions with residues of the active site; O4 accepts hydrogen bonds from a conserved Asn and from a nearby main-chain amide, while O2 bonds with a main-chain amide. Thus, UDGase's substrate range is uniquely determined by the small size of

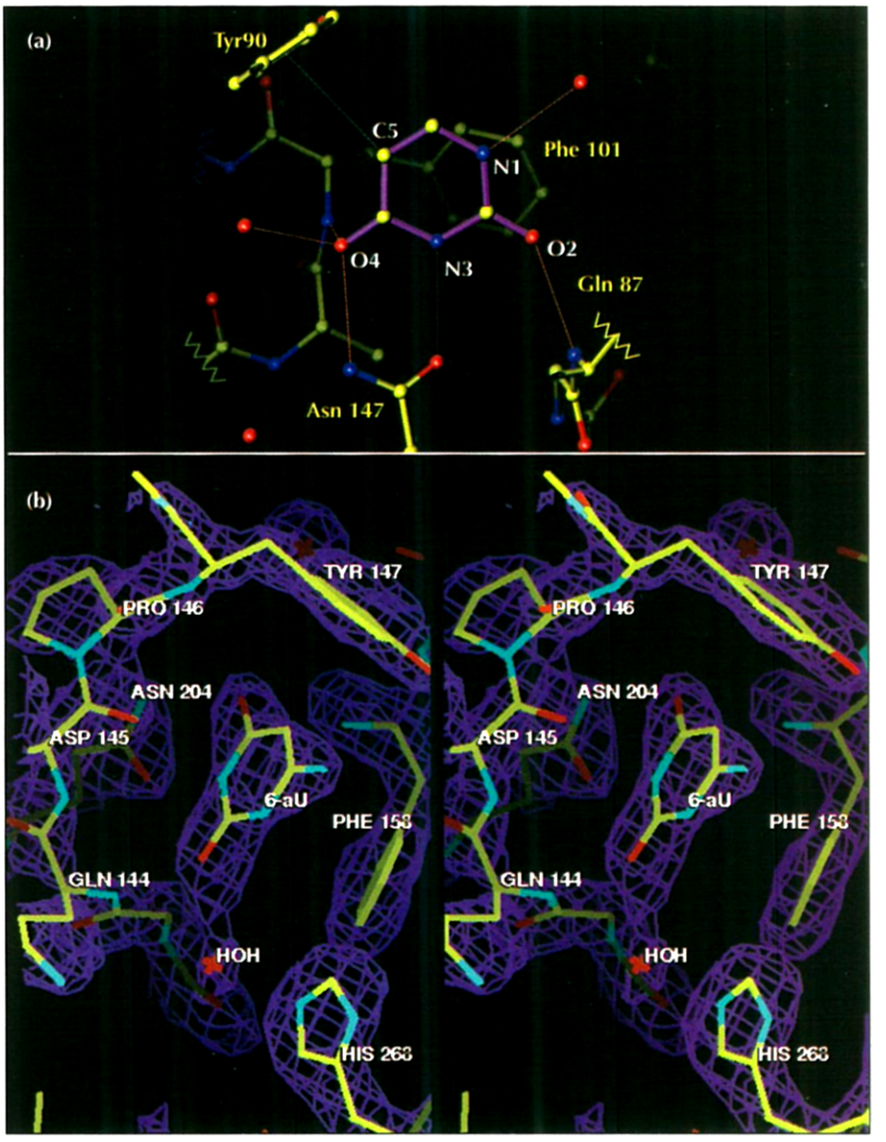

Fig. 3. Structure of the uracil-binding pocket. **(a)** Interactions between HSV-1 UDGase and bound uracil include van der Waals contacts with conserved Phe101 and Tyr90 and electrostatic interactions involving the polar atoms of the pyrimidine ring and neighboring residues. **(b)** 6-Aminouracil (6-aU) binds to human UDGase in a similar orientation with many of the same conserved residues interacting with O2, N3, O4 and C5. These extensive interactions between the protein and the target base contribute to the UDGase's high substrate selectivity. Reprinted with permission from [8] and [9], respectively.

the uracil-binding pocket and the distribution of polar and nonpolar groups lining the pocket. The structural basis for the selective cleavage of uridine from DNA, and not from RNA, is less certain. A model of the HSV-1 UDGase with bound DNA, based on structures of the enzyme complexed with uracil or thymine trinucleotide, suggests that the 2'-OH of an RNA substrate would clash with the conserved phenylalanine stacking against C5 of uracil [8]. In addition, the 2'-OH of a bound RNA substrate would block the access of a catalytically important His to the catalytic site [9]. Confirmation of these models awaits additional structural data.

The human and HSV-1 UDGase structures suggest mechanisms for enzyme-assisted cleavage of the deoxyuridine amino–glycosylic bond (Fig. 4). This cleavage reaction proceeds via nucleophilic attack of deoxyribose C1', with polarization of the scissile bond causing a buildup of positive charge on the sugar and negative charge on the base in the transition state proposed for this reaction. Residues rimming the uridine-binding pocket of both UDGase structures are appropriately situated to assist in nucleophilic attack of the anomeric carbon and protonation of the pyrimidine leaving group. Savva *et al.* [8] propose that aspartic acid 88 of HSV-1 UDGase is a general base that activates water for attack on C1'. The uridine leaving group is protonated by nearby His210, and its orientation is stabilized by the hydrogen bonds inferred from the structure of the enzyme complex with uracil. Mol *et al.* [9] suggest two alternative mechanisms for human UDGase that are based on the structure of the human enzyme and on analyses of its substrate binding and catalytic activities by site-directed mutagenesis. In their reaction schemes, His268 either directly attacks C1' or serves as a general base forming an OH^- nucleophile that attacks C1'. Uridine's orientation is fixed by the same conserved Asn described above for the HSV-1 enzyme, and additional hydrogen bonds form between main-chain atoms and uracil's O2, O4 and N3 atoms. These contacts with the protein orient uracil for cleavage and provide favorable interactions in the postulated transition-state structure.

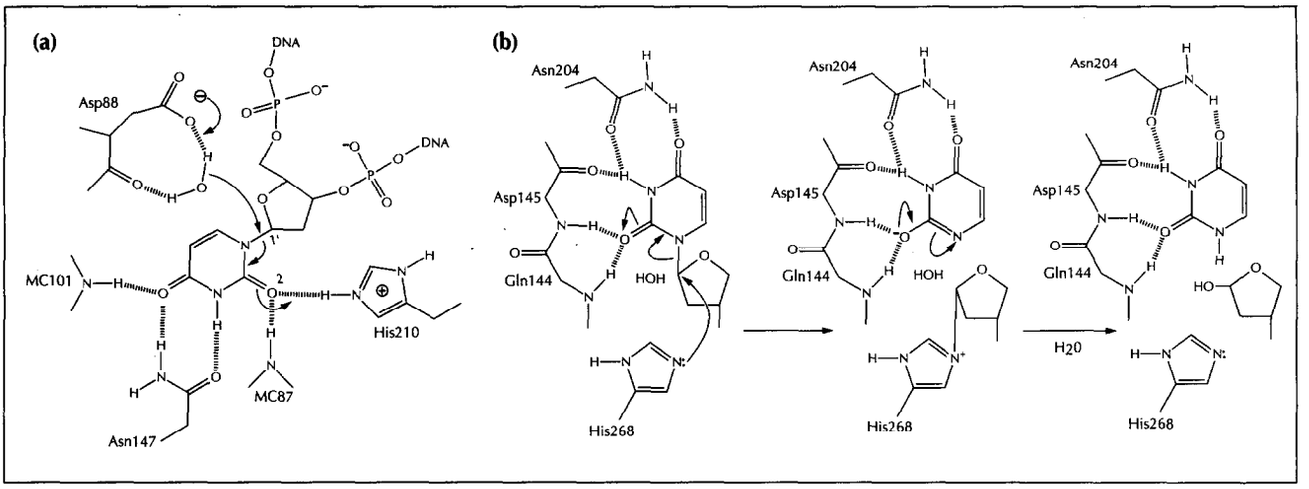

Fig. 4. Proposed mechanisms for base excision by UDGases. **(a)** The proposed cleavage mechanism for HSV-1 UDGase involves solvent-mediated attack on C1', activated by the general base Asp88 (*Savva et al.*, [8]). His210 protonates uracil O2 and neighboring residues help to stabilize the iminol tautomer. Product release is accompanied by protonation of His210 and the binding of solvent to D88. MC, main-chain atoms **(b)** Mol *et al.* [9] propose that His268 of human UDGase mediates attack on C1', either directly with the formation of an enzyme–uracil intermediate or indirectly by activation of solvent. Neighboring amides interacting with uracil O2 serve to polarize the glycosylic bond and stabilize the iminol intermediate in the cleavage reaction. Reprinted with permission from [8] and [9], respectively.

These structures of the human and HSV-1 UDGases are the first high resolution views of monofunctional DNA glycosylases, and they provide general insights into the process of lesion-specific recognition of DNA in any sequence context. Features such as the extrusion of the target base are likely to be used by other DNA glycosylases, whereas strategies for the selective recognition of damaged bases and the mechanism of glycosylic bond cleavage are no doubt tailored to the substrate ranges of individual enzymes in this family of DNA-repair proteins.

References

1. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.
2. Barnes, D.E., Lindahl, T. & Sedgwick, B. (1993). DNA repair. *Curr. Opin. Struct. Biol.* **5**, 424–433.
3. Cleaver, J.E. (1994). It was a very good year for DNA repair. *Cell* **76**, 1–4.
4. Selby, C.P. & Sancar, A. (1993). Molecular mechanism of transcription–repair coupling. *Science* **260**, 53–58.
5. Schaeffer, L., *et al.*, & Egly, J.-M. (1993). DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260**, 58–63.
6. Feaver, W.J., *et al.*, & Kornberg, R.D. (1993). Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. *Cell* **75**, 1379–1387.
7. Smith, M.L., *et al.*, & Fornace, A.J., Jr (1994). Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science* **266**, 1376–1380.
8. Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature* **373**, 487–493.
9. Mol, C.D., *et al.*, & Tainer, J.A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell* **80**, 869–878.
10. Kornberg, A. & Baker, T.A. (1992). *DNA Replication.* (2nd edn), W.H. Freeman, New York.
11. Demple, B. & Harrison, L. (1994). Repair of oxidative damage to DNA: enzymology and biology. *Annu. Rev. Biochem.* **63**, 915–948.
12. Klimasuaskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). *Hhal* methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369.

Tom Ellenberger, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA.