FEBS Letters 589 (2015) 2675–2682

journal homepage: www.FEBSLetters.org

Crystal structure of family 4 uracil–DNA glycosylase from *Sulfolobus tokodaii* and a function of tyrosine 170 in DNA binding

Akito Kawai [a,*], Shigesada Higuchi [a], Masaru Tsunoda [c], Kazuo T. Nakamura [b], Yuriko Yamagata [d], Shuichi Miyamoto [a]

[a] Faculty of Pharmaceutical Sciences, Sojo University, 4-22-1 Ikeda, Nishi-ku, Kumamoto 860-0082, Japan
[b] School of Pharmacy, Showa University, 1-5-8 Hatanodai, Shinagawa-ku, Tokyo 142-8555, Japan
[c] Faculty of Pharmacy, Iwaki Meisei University, 5-5-1 Chuodai-iino, Iwaki 970-8551, Japan
[d] Graduate School of Pharmaceutical Sciences, Kumamoto University, 5-1 Oe-honmachi, Chuo-ku, Kumamoto 862-0973, Japan

ARTICLE INFO

Article history:
Received 20 May 2015
Revised 23 July 2015
Accepted 14 August 2015
Available online 28 August 2015

Edited by Christian Griesinger

Keywords:
Uracil–DNA glycosylase
Crystal structure
Leucine-intercalation loop
Sulfolobus tokodaii

ABSTRACT

Uracil–DNA glycosylases (UDGs) excise uracil from DNA by catalyzing the *N*-glycosidic bond hydrolysis. Here we report the first crystal structures of an archaeal UDG (*sto*UDG). Compared with other UDGs, *sto*UDG has a different structure of the leucine-intercalation loop, which is important for DNA binding. The *sto*UDG–DNA complex model indicated that Leu169, Tyr170, and Asn171 in the loop are involved in DNA intercalation. Mutational analysis showed that Tyr170 is critical for substrate DNA recognition. These results indicate that Tyr170 occupies the intercalation site formed after the structural change of the leucine-intercalation loop required for the catalysis.

1. Introduction

Deamination is a common base modification in DNA. When cytosine and adenine are deaminated, they are converted into uracil and hypoxanthine, respectively. Deamination of guanine produces xanthine or oxanine. These deaminated bases are recognized as risk sites for transition mutations in thermophilic archaea. The archaeal family-B DNA polymerase possesses a "read-ahead" scanning mechanism: a running polymerase recognizes the base at position +4 of the DNA template strand and stalls replication if uracil or hypoxanthine is found [1]. Archaeal family-D polymerase are also inhibited by the presence of uracil in the DNA template strand [2]. Therefore, we believe that the study of the repair system in the presence of the deaminated bases is important for further understanding of the DNA replication mechanism in archaea.

Uracil–DNA glycosylase (UDG) is a monofunctional DNA glycosylase that initiates the base excision repair pathway. UDGs are widely identified in archaea, eukaryotes, bacteria, and large DNA viruses and are well-studied examples of the removal of deaminated bases from DNA. UDG-family enzymes are classified into six families on the basis of their substrate specificity, conserved motifs, and structural similarities [3,4]. Archaea commonly carry the genes for UDGs from family 4, 5, and 6 [4–6]. Biochemical studies of these UDGs have demonstrated that family 4 and 5 UDGs possess four conserved cysteine residues required to coordinate the [4Fe–4S] iron–sulfur cluster, and the substrate specificities of family 4, 5 and 6 UDGs are different as follows. Family 4 UDGs remove uracil from both double- and single-stranded DNA [7]. Family 5 UDGs have a broad substrate specificity for uracil, hypoxanthine, and xanthine in double-stranded DNA [8,9]. Family 6 UDGs exhibit a hypoxanthine-DNA glycosylase activity but do not have UDG activity [4]. Moreover, it has been reported that archaeal family 4 UDGs interact with proliferating cell nuclear antigen (PCNA), which is a processivity factor for replicative DNA polymerase [10,11]. These findings suggest a PCNA-mediated

Abbreviations: BSA, bovine serum albumin; HEX, hexachloro-fluorescein; MES, 2-(*N*-morpholino)ethanesulfonic acid; MR, molecular replacement; PCNA, proliferating cell nuclear antigen; PDB, protein data bank; TLS, translation, liberation and screw; UDG, uracil–DNA glycosylase

Author contributions: A.K., S.H. and S.M. designed the project. A.K. performed all experiments. M.T. and K.T.N. contributed the enzyme preparations. Y.Y. contributed the structure determinations and the data interpretations. A.K. and S.M. wrote the manuscript. All authors read and approved the final manuscript.

* Corresponding author. Fax: +81 96 326 5048.
E-mail address: akawai@ph.sojo-u.ac.jp (A. Kawai).

http://dx.doi.org/10.1016/j.febslet.2015.08.019
0014-5793/

repair system whereby archaeal family 4 UDG is recruited to PCNA and then removes the uracil base from DNA.

In the present study, we determined the crystal structures of a family 4 UDG isolated from the thermoacidophilic crenarchaea *Sulfolobus tokodaii* (*sto*UDG), in the free form and in the complex form with uracil (*sto*UDG–uracil complex). To date, approximately eighty crystal structures of UDG-family enzymes isolated from eukaryotes, bacteria, and viruses have been deposited in PDB. These structural studies revealed that the UDG-family enzymes has a common α/β/α sandwich fold that a four-strands parallel β-sheet is located at the center of the molecule and bordered by the α-helices despite low protein sequence similarities among them [3,12,13]. The crystal structures of family 4 UDGs isolated from the hyperthermophilic bacteria *Thermus thermophilus* (*tth*UDGa) [14] and *Thermotoga maritima* (*tma*UDG) have been reported and showed that the overall structures and the active site arrangement are similar to those of family 1 UDGs [14]. However, the crystal structure isolated from archaea had never been reported. Our present study provides the first report regarding to the archaeal UDG structures. To characterize the *sto*UDG structure, we compared it with the crystal structures of the bacterial family 4 UDGs. Our results indicated that the *sto*UDG structure involved in substrate DNA recognition differs from the bacterial family 4 UDG structures. Thus, to identify the residue important for the substrate DNA recognition, we created model structures of family 4 UDG complexed with DNA (UDG–DNA complex model) and performed mutational analyses of *sto*UDG.

2. Materials and methods

2.1. Enzyme and DNA preparation

The recombinant wild-type *sto*UDG for the mutational analyses and the *sto*UDG mutant truncated the C-terminal region (Tyr195–Lys220) for the crystallizations were overexpressed in *Escherichia coli* and purified as previously described [15]. The *sto*UDG mutants for the mutational analyses were constructed by PCR-mediated mutagenesis using KOD-Plus-Ver.2 DNA polymerase (TOYOBO) and DpnI nuclease (NEW ENGLAND BioLabs). Overexpression and purification of *sto*UDG mutants were performed with the same procedure as used for wild-type *sto*UDG. All of the *sto*UDG enzymes did not contained the additional amino acids such as affinity-tags in their protein sequences. The oligonucleotide sequences used in the UDG assay and the fluorescence anisotropy-based DNA binding assay are summarized in Supplementary Table S1.

2.2. Crystallization, data collection and structure determination

Crystallizations and data collections for *sto*UDG crystals in the free form and the *sto*UDG–uracil complex were performed as pre-

Table 1
Data-collection and refinement statistics. Values in parentheses are for the highest resolution shell.

Data-collection	Uracil complex form	Free form	
PDB code	4zby	4zbx	4zbz
Source	SPring-8 BL44XU	SPring-8 BL44XU	MicroMAX 007
Wavelength (Å)	0.9000	0.9000	1.5418
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell	a = 52.06,	a = 52.19,	a = 52.03,
Parameters (Å)	b = 52.22,	b = 52.25,	b = 52.35,
	c = 74.06	c = 74.45	c = 74.69
Resolution range (Å)	50.0–1.70 (1.76–1.70)	50.0–1.70 (1.76–1.70)	50.0–1.90 (1.95–1.90)
No. of observed reflections	123,660	168,247	133,938
No. of unique reflections	23,005 (2,248)	23,097 (2,275)	16,590 (1,188)
Multiplicity	5.4 (5.2)	7.3 (7.1)	8.1 (4.6)
Completeness (%)	99.9 (99.6)	100 (100)	99.7 (97.5)
R_{merge} (%)[a]	5.6 (39.7)	5.0 (33.9)	5.9 (32.5)
$\langle I/\sigma(I)\rangle$	37.6 (4.5)	50.8 (6.4)	25.3 (5.2)
Refinement			
Resolution (Å)	42.7–1.70 (1.78–1.70)	36.9–1.70 (1.78–1.70)	26.2–1.90 (2.02–1.90)
Reflection used	22,831 (2,807)	23,045 (2,831)	16,566 (2,689)
R_{work} (%)[b]	16.5 (20.1)	16.9 (19.1)	16.0 (17.1)
R_{free} (%)[c]	20.7 (26.2)	21.6 (24.0)	20.6 (24.5)
Completeness (%)	99.9 (99.8)	99.9 (99.4)	99.6 (98.7)
No. of non-hydrogen atoms	1,885	1,818	1,854
Protein	1,657	1,648	1,674
Ligands	28	20	26
Water	200	150	154
r.m.s deviation from ideality			
Bond length (Å)	0.010	0.008	0.009
Bond angle (°)	1.195	1.059	1.126
Avg. *B*-factor	27.5	29.4	25.7
Protein	26.3	28.2	24.3
Ligands	29.8	40.3	31.1
Waters	37.2	40.1	40.4
Ramachandran plot			
Favored region (%)	98.0	98.0	98.5
Allowed region (%)	2.0	2.0	1.5
Outlier region (%)	0.0	0.0	0.0
Clashscore	3.5	2.7	3.2

[a] $R_{merge} = 100 \times \sum_{hkl}\sum_i|I_i(hkl) - \langle I(h\ k\ l)\rangle|/\sum_{hkl}\sum_i I_i(h\ k\ l)$, where $\langle I(hkl)\rangle$ is the mean value of $I(hkl)$.
[b] $R_{work} = 100 \times \sum_{hkl}||F_o| - |F_c||/\sum_{hkl}|F_o|$, where F_o and F_c the observed and calculated structure factors, respectively.
[c] R_{free} is calculated as for R_{work}, but for the test set comprising 5% reflections not used in refinement.