J. Mol. Biol. (1995) 249, 360–375

JMB

Crystal Structure of a Pyrimidine Dimer-specific Excision Repair Enzyme from Bacteriophage T4: Refinement at 1.45 Å and X-ray Analysis of the Three Active Site Mutants

Kosuke Morikawa[1]*, Mariko Ariyoshi[1], Dmitry G. Vassylyev[1] Osamu Matsumoto[1,3], Katsuo Katayanagi[1,4] and Eiko Ohtsuka[2]

[1]*Protein Engineering Research Institute, 6-2-3 Furuedai Suita, Osaka 565, Japan*

[2]*Faculty of Pharmaceutical Sciences, Hokkaido University Sapporo Hokkaiso 060, Japan*

[3]*Present address: Faculty of Pharmaceutical Sciences Kyoto University, Sakyo-ku Kyoto 606, Japan*

[4]*Present address: Central Research Laboratories of Mitsubishi Kasei Corporation 1000, Kamoshida, Midori-ku Yokohama, Kanagawa 227 Japan*

**Corresponding author*

Crystallographic study of bacteriophage T4 endonuclease V, which is involved in the initial step of the pyrimidine dimer-specific excision repair pathway, has been carried out with respect to the wild-type and three different mutant enzymes. This enzyme catalyzes the cleavage of the *N*-glycosyl bond at the 5'-side of the pyrimidine dimer, and subsequently incises the phosphodiester bond at the apyrimidinic site through a β-elimination reaction. The structure of the wild-type enzyme refined at 1.45 Å resolution reveals the detailed molecular architecture. The enzyme is composed of a single compact domain classified as an all-α structure. The molecule is stabilized mainly by three hydrophobic cores, two of which include many aromatic side-chain interactions. The structure has a unique folding motif, where the amino-terminal segment penetrates between two major α-helices and prevents their direct contact, and it is incompatible with the close-packing category of helices for protein folding. The concave surface, covered with many positive charges, implies an interface for DNA binding. The glycosylase catalytic center, which comprises Glu23 and the surrounding basic residues Arg3, Arg22 and Arg26, lie in this basic surface. The crystal structures of the three active-site mutants, in which Glu23 was replaced by Gln(E23Q) and Asp (E23D), respectively, and Arg3 by Gln (R3Q), have been determined at atomic resolution. The backbone structures of the E23Q and R3Q mutants were almost identical with that of the wild-type, while the E23D mutation induces a small, but significant, change in the backbone structure, such as an increase of the central kink of the H1 helix at Pro25. In the catalytic center of the glycosylase, however, these three mutations do not generate notable movements of protein atoms, except for significant shifts of some bound water molecules. Thus, the structural differences between the wild-type and each mutant are confined to the remarkably small region around their replaced chemical groups. Combined with the biochemical studies and the difference circular dichroism measurements, these results allow us to conclude that the negatively charged carboxyl group of Glu23 is essential for the cleavage of the *N*-glycosyl bond, and that the positively charged guanidino group of Arg3 is crucial to bind the substrate, a DNA duplex containing a pyrimidine dimer. The amino terminal α-amino group is located at a position approximately 4.4 Å away from the carboxyl group of Glu23. These structural features are generally consistent with the reaction scheme proposed by Dodson and co-workers.

Keywords: active site mutants; crystal structure; DNA repair; pyrimidine dimer; T4 endonuclease V

Abbreviations used: $R = [\Sigma Ii\text{-}\langle Ii\rangle/\Sigma Ii]\times 100$, where $\langle Ii\rangle$ is the average of Ii over all symmetrical equivalents; r.m.s.d., root-mean-square displacement; endoV, endonuclease V; 3D, three-dimensional; MIR, multiple isomorphous replacement.

0022–2836/95/220360–16 $08.00/0

Introduction

Ultraviolet (UV) light induces the formation of a pyrimidine dimer, which is the most typical lesion within a DNA duplex. Endonuclease V (endoV; endodeoxyribonuclease (pyrimidine dimer); deoxyribonuclease (pyrimidine dimer), EC 3.1.25.1), encoded by the bacteriophage T4 *denV* gene, is responsible for the first step of the catalytic reaction in the excision repair pathway specific for the *cis-syn* cyclobutane pyrimidine dimer (Yasuda & Sekiguchi, 1970; Friedberg, E. & King, 1971). This enzyme can scan a non-target sequence along a DNA duplex by non-specific electrostatic interaction, before the specific recognition of a pyrimidine dimer (Grunskin & Lloyd, 1986; Dodson & Lloyd, 1989). The most interesting functional aspect of this enzyme is that the two distinct catalytic activities reside in a relatively small protein consisting of 138 residues. The enzyme first hydrolyzes the *N*-glycosyl bond at the 5'-side of a pyrimidine dimer, and subsequently catalyzes the cleavage of the phosphodiester bond at the resulting abasic site (Dodson & Lloyd, 1989; Nakabeppu & Sekiguchi, 1981; Macmillan *et al.*, 1981). This second reaction proceeds through β-elimination other than the actual hydrolysis (Manoharan *et al.*, 1988; Kim & Linn 1988). Dodson *et al.* (1993) have found evidence for an intermediate in the course of the catalytic reaction, and proposed a reaction scheme for T4 endoV that is similar to that of AMP *N*-glycosylase. Thus, T4 endoV has unique catalytic functions, the molecular mechanisms of which can be most effectively elucidated in terms of the three-dimensional (3D) structure.

We have reported the crystal structure of T4 endoV determined at 1.6 Å resolution by X-ray crystallography (Morikawa *et al.*, 1992). That report communicated the first implications of the structure-function relationships deduced from the crystal structure. Furthermore, in combination with various results from extensive site-directed mutagenesis studies, the structural features of the enzyme, such as the biased charge distribution on the molecular surface, have allowed the identification of the glycosylase catalytic center that comprises Glu23 and the surrounding basic residues (Morikawa *et al.*, 1992; Doi *et al.*, 1992). The crystal structure of *Escherichia coli* endonuclease III (Kuo *et al.*, 1992), which is a DNA repair enzyme containing a [4Fe-4S] cluster, has implicated that this enzyme also may have a glycosylase catalytic center similar to that of T4 endoV.

The crystal structure of the wild-type enzyme has now been more precisely refined at 1.45 Å resolution, based on the intensity data collected using the synchrotron radiation source. We have determined the crystal structures of three different mutants, in which two residues constituting the putative active center for the glycosylase were replaced by others. Here, we first describe the details of the molecular architecture on the basis of the highly refined 3D structure of the wild-type enzyme. Secondly, the crystal structures of the three active-site mutants, all of which were refined at atomic resolution, have provided new insights into the catalytic reaction of glycosylase, and therefore we discuss the 3D structure-function relationships of the enzyme, in conjunction with recent biochemical results.

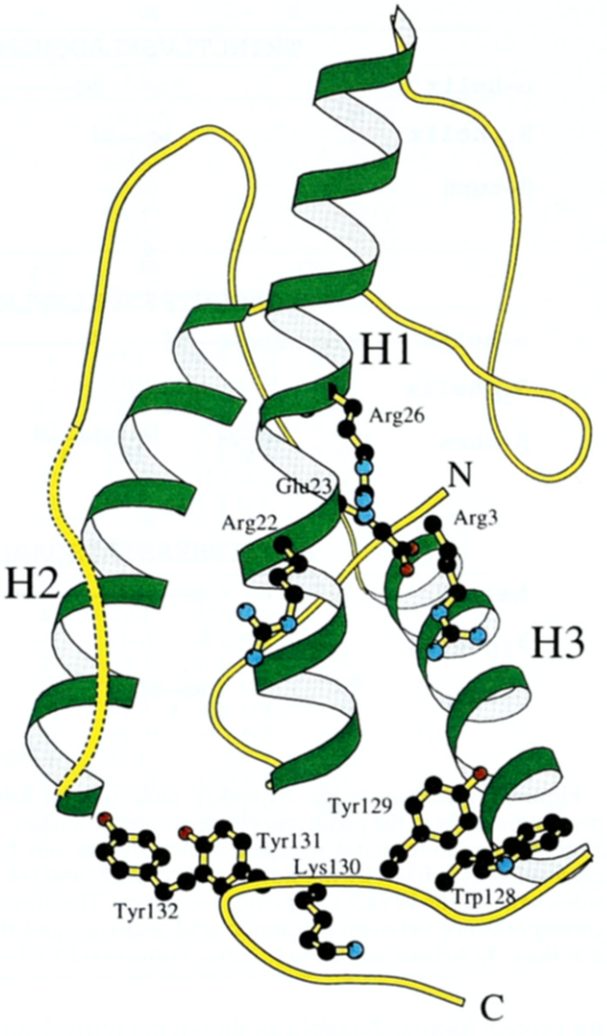


Figure 1(a) *(caption opposite)*

Results and Discussion

Main-chain folding

As reported previously (Morikawa *et al.*, 1992), the enzyme consists of three α-helices, seven reverse turns and six loops (Figure 1(a)). The ranges of α-helices and reverse turns are defined in Figure 1(b). Thus, the molecule, bearing the two distinct catalytic activities, forms a single compact domain classified into an all-α structure. The most striking feature in the main-chain folding is the amino-terminal segment, which penetrates between the two major α-helices, H1 and H3, and prevents their direct contact (Figure 1(a)). All the proteins classified into the all-α structure are known to comply with the close-packing category for assembling helices (Richardson, 1981). However, this folding pattern of T4 endoV is inconsistent with this central category

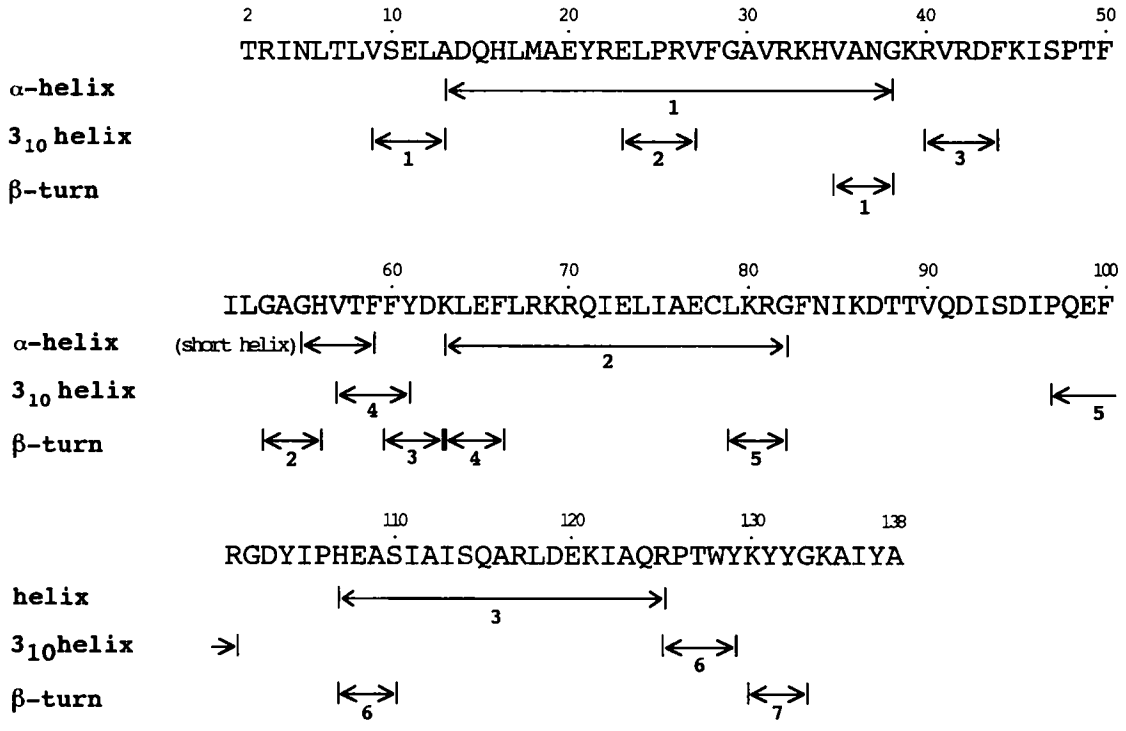


Figure 1(b)

Figure 1. Structure of the T4 endoV molecule. (a) Ribbon diagram (Kraulis, 1991) of the molecule viewed almost perpendicularly to the positively charged concave surface. The H1 helix is kinked at Pro25, which is located in the middle. Note the amino-terminal segment, which thrusts out between the H1 and H2 helices. The catalytic center for the glycosylase is marked by the side-chains of Glu23 and the surrounding three basic residues, Arg3, Arg22 and Arg23. The side-chains of the WYKYY sequence are shown. The broken lines indicate the mobile segment (residues 83 to 92). (b) The correspondence between the amino acid sequence and the secondary structure elements. The numbers, which denote α-helices, 3_{10} helices and reverse turns, correspond to those in Table 4.

for protein folding. Notably, as shown in Figure 2, all the main-chain carbonyl oxygen atoms and imino protons in this amino-terminal segment participate formation of hydrogen bonds with the surrounding atoms. This finding implies that the polar main-chain peptide atoms are allowed to lie in the interior of the protein molecule only by having polar interactions with the surrounding atoms.

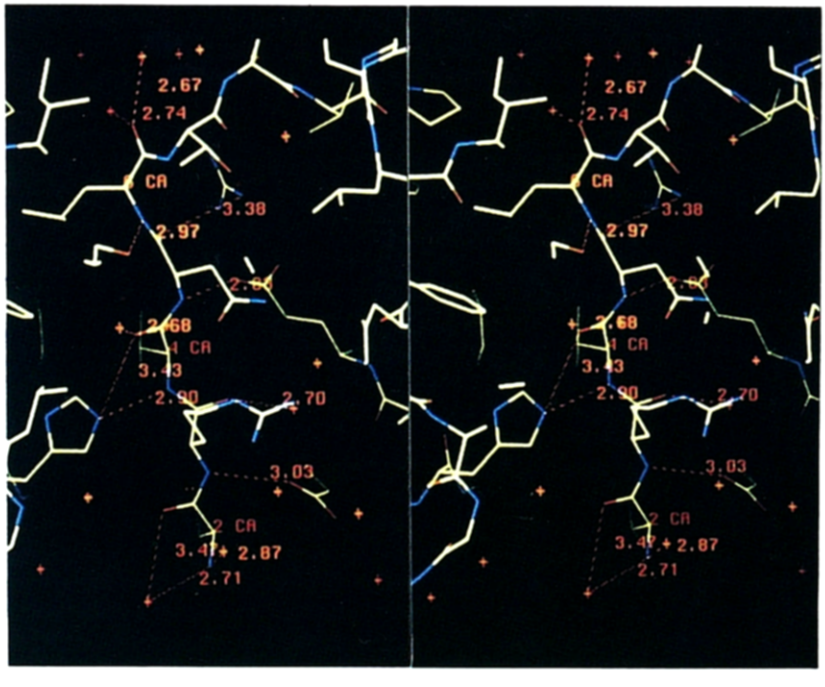


Figure 2. Stereo pair showing hydrogen bonds (red broken lines) formed between the main-chain atoms of the amino-terminal segment and the surrounding atoms. Oxygen and nitrogen atoms are indicated by red and blue, respectively. Crosses indicate water molecules. The NH group of Arg3 forms a hydrogen bond with the carboxyl side-chain of Glu23, the CO group of Arg3 with a water molecule (140WAT), the NH group of Ile4 with the imidazole side-chain of His56, the CO group of Ile4 with the same side-chain of His56, the NH group of Asn5 with the carboxyl group of Glu20, the CO group of Asn5 with the guanidino side-chain of Arg70, the NH group of Leu6 with the hydroxyl group of Ser114, and the CO group of Leu6 with two water molecules (200WAT and 145WAT).

Table 1

Parameters of α-helices, reverse turns and 3_{10} helices

α-Helix	Residues	⟨φ⟩ (deg.)[a]	⟨ψ⟩ (deg.)[a]
Helix H1	14–23	−66 (10)	−41 (4)
	27–35	−62 (3)	−43 (3)
Helix H2	66–79	−62 (5)	−43 (5)
Helix H3	110–124	−62 (4)	−42 (6)
Helix overall		−63 (6)	−43 (5)
Mean helix from data base[b]		−62 (7)	−41 (7)

	H-bond	N-O (Å)	φ2 (°)	ψ2 (°)	φ3 (°)	ψ3 (°)	Type[c]	Sequence	C^{α}(1–4) (Å)	Previous notation[d]
A. *Reverse turn*										
1.	35 O–38 N	3.03	−63	−27	−88	8	RT-I	V A N G	5.17	
2.	53 O–56 N	2.81	−59	119	97	−2	RT-II	G A G H	5.37	
3.	60 O–63 N	2.90	−63	−25	−112	19	NR-I	F Y D K	4.80	
4.	63 O–66 N	3.16	−65	−28	−67	−37	RT-III	K L E F	5.45	
5.	79 O–82 N	2.96	−55	−38	−82	−2	RT-I	L K R G	5.09	
6.	107 O–110 N	3.02	−60	−27	−70	−33	RT-III	H E A S	5.41	
7.	131 O–134 N	2.97	59	39	74	2	RT-I'	Y Y G K	5.10	
B. 3_{10} *helix*										
1.	8 O–11 N	3.01	−63	−30	−72	−11	RT-III	L V S E	5.48	G1
	9 O–12 N	3.24	−72	−11	−101	−3	NR-I	V S E L	5.43	
2.	23 O–26 N	3.03	−54	−44	−64	−21	RT-III	E L P R	5.51	
	24 O–27 N	3.13	−64	−21	−56	−28	NR-III	L P R V	6.05	
3.	40 O–43 N	2.91	−58	−29	−68	−10	RT-III	R V R D	5.55	G2
	41 O–44 N	3.11	−68	−10	−80	−9	NR-I	V R D F	5.76	
4.	57 O–60 N	2.87	−59	−32	−54	−27	RT-III	V T F F	5.67	G3
	58 O–61 N	2.87	−54	−27	−95	−6	RT-I	T F F Y	5.21	
5.	97 O–100 N	3.09	−55	−34	−60	−23	NR-III	P Q E F	6.08	G4
	98 O–101 N	2.92	−60	−23	−87	−17	RT-I	Q E F R	5.09	
6.	125 O–128 N	2.97	−63	−23	−69	−15	NR-III	R P T W	5.88	G5
	126 O–129 N	3.01	−69	−15	−80	−6	RT-I	P T W Y	5.51	

[a] Values in parentheses are standard deviations.
[b] Barlow & Thornton (1988).
[c] Classified according to Crawford *et al.* (1973). RT, reverse turn; NR, near reverse turn.
[d] Morikawa *et al.* (1992).

According to the notation used by Kabsch & Sander (1983), the previous paper (Morikawa *et al.*, 1992) defined the five reverse turns, all of which are located on the molecular surface. The reexamination based on the classification proposed by Crawford *et al.* (1973) are summarized in Table 1. All of the previous five reverse turns are now defined as 3_{10} helices. The near-reverse turn structure (residues 24 to 27), corresponding to the central kink of the H1 helix at Pro25, is noteworthy, since this proline residue is located on the opposite side of the H1 helix to Glu23, which is essential for the glycosylase reaction (Morikawa *et al.*, 1992; Doi *et al.*, 1992).

Intramolecular ionic interactions

The T4 endoV molecule contains relatively many intramolecular ionic interactions, which would contribute to maintaining the molecular architecture. The interactions with distances less than 3.4 Å are summarized in Table 2. The hydrogen bonds formed between the amino-terminal segment and the surrounding atoms are described above. Other notable hydrogen bonds, particularly around the carboxyl side-chain of Glu23 are discussed in a later section related to the catalytic center of the glycosylase. In addition to a number of hydrophobic interactions, several polar interactions, such as two hydrogen bonds between the Glu20 and Arg70 side-chains, a hydrogen bond between the Tyr21 carbonyl group and the Glu71 side-chain, and a hydrogen bond between the Asp14 and Arg81 side-chains contribute to the bundling of the H1 and H2 helices. Asp14, lying at the amino end of the H1 helix, forms several hydrogen bonds with Lys130, Tyr132 and Gly133. These polar interactions appear to contribute to the connection of the carboxyl-terminal loop with the remaining major region of the molecule, which includes the three α-helices.

Intramolecular hydrophobic interactions

The locations of the hydrophobic residues within the enzyme are shown in Figure 3. There are three hydrophobic cores within the molecule. Two of them contain a number of aromatic residues, suggesting that aromatic side-chain interactions stabilize the molecular architecture of this enzyme. The first hydrophobic core is surrounded by the amino-terminal half of the H1 helix, the amino terminus of the H2 helix, the irregular loop between these two helices and a part of the long loop connecting the H1 and H2 helices. In this core, the hydrophobic side-chains, such as Val27, Phe28, Val31, Val35, Leu64, Ile93 and

Ile96, are closely packed, while the aromatic side-chain of Tyr61 is stacked between the Phe44 and Val31 side-chains. The second core is formed by hydrophobic residues (Ile4 and Leu6) located in the amino-terminal segments, residues (Phe50, Leu52, Phe59 and Phe60) in the loop between the H1 and H2 helices, and hydrophobic residues (Phe66, Leu67, Tyr104, Phe106 and Ile113) near the H2 and H3 helices. In this core, the aromatic side-chains of Phe50, Phe59, Phe60, Phe66 and Tyr104 are clustered. The third is the smallest among the three hydrophobic cores. It contains Val9 and Leu12 located around the 3_{10} helix I, Tyr21 in the H1 helix, Leu74 in the middle of the H2 helix, and Phe83 and Ile85 near the carboxyl end of the H2 helix.

Temperature factors

The mean *B*-factor variations plotted *versus* residue numbers are shown in Figure 4 with respect to both the main-chain and side-chain atoms. In contrast to the mobility found in usual proteins, the amino-terminal segment of this protein has considerably less mobility, presumably because it is restricted by the internal hydrogen bonds of the main-chain atoms, as represented in Figure 2. Several residues in the vicinity of the carboxyl end of the H1 helix show greater mobility. These residues form a prominent region that is more exposed to the solvent. The most flexible is a segment from residue 84 to 91. In this region, the densities corresponding to main-chain atoms are not well defined even at 1.45 Å resolution, suggesting conformational disorder within the crystal. It should be noted that the sequence Lys-Arg-Gly-Phe-Asn-Ile-Lys-Asp-Thr-Thr-Val, which partly includes the mobile segment, has been reported to exhibit a significant similarity to a sequence (residues 151 to 161; Lys-Arg-Gly-Phe-Lys-Phe-Val-Gly-Thr-Thr-Ile) from 3-methyladenosine glycosylase from *Escherichia coli* (Steinum & Seeberg, 1986). Photoaffinity labeling of the enzyme with a substrate analog suggested that this segment is involved in DNA binding (Hori *et al.*, 1992a,b).

Water molecules

In the course of the refinement, 172 water molecules were identified (Figure 5). Among them, 74 molecules have temperature factors less than 30 Å^2. Fewer water molecules are observed on the

Table 2

Intramolecular interactions (≤3.4 Å) (for the side-chains)

Residue	Atom	Residue	Atom	Distance (Å)
Arg3	N^{ϵ}	Asn5	$O^{\delta 1}$	3.16
Arg3	N	Glu23	$O^{\epsilon 1}$	3.03
Ile4	N	His56	$N^{\epsilon 2}$	2.90
Asn5	N	Glu20	$O^{\epsilon 1}$	2.83
Asn5	$N^{\delta 2}$	Glu20	$O^{\epsilon 2}$	2.93
Asn5	O	Arg70	N^{ϵ}	3.04
Asn5	$O^{\delta 2}$	Arg70	$N^{\eta 2}$	3.38
Leu6	N	Ser114	O^{γ}	2.97
Thr7	O	Arg70	$N^{\eta 1}$	3.01
Thr7	O	Arg70	$N^{\eta 2}$	3.05
Val9	O	Arg81	$N^{\eta 2}$	3.26
Ser10	O^{γ}	Glu11	N	3.35
Ser10	O	Tyr131	O^{η}	2.74
Leu12	O	Arg81	$N^{\eta 1}$	2.80
Leu12	O	Arg81	$N^{\eta 2}$	2.85
Asp14	$O^{\delta 2}$	Arg81	$N^{\eta 1}$	2.83
Asp14	O	Lys130	O	2.72[a]
Asp14	$O^{\delta 1}$	Tyr132	N	3.15
Asp14	$O^{\delta 2}$	Tyr132	N	2.78
Asp14	$O^{\delta 1}$	Gly133	N	2.75
Gln15	$O^{\epsilon 1}$	Lys130	N	2.89
Glu20	$O^{\epsilon 1}$	Arg70	N^{ϵ}	2.73
Glu20	$O^{\epsilon 2}$	Arg70	$N^{\eta 2}$	2.90
Tyr21	O	Gln71	$N^{\epsilon 2}$	3.10
Tyr21	O^{η}	Asp87	N	3.31
Tyr21	O^{η}	Asp87	O	2.75
Arg32	$N^{\eta 1}$	Gln91	$O^{\epsilon 1}$	2.74
Arg32	$N^{\eta 2}$	Gln91	$O^{\epsilon 1}$	3.20
Arg32	$N^{\eta 1}$	Asp92	O	3.06
Arg32	N^{ϵ}	Asp95	$O^{\delta 2}$	2.76
Arg32	$N^{\eta 2}$	Gln95	$O^{\delta 1}$	3.04
His34	$N^{\epsilon 2}$	Tyr61	O^{η}	3.39
Arg40	N	Asp43	$O^{\delta 2}$	2.77
Arg40	$N^{\eta 1}$	Glu99	$O^{\epsilon 2}$	3.23
Arg40	$N^{\eta 2}$	Glu99	$O^{\epsilon 2}$	2.80
Val41	N	Gln99	$O^{\epsilon 1}$	2.79
Ser47	O^{γ}	Pro48	N	3.04

continued

Table 2 *(continued)*

Intramolecular interactions (≤3.4 Å) (for the side-chains)

Residue	Atom	Residue	Atom	Distance (Å)
Ser47	O^{γ}	Thr49	N	3.17
Ser47	O^{γ}	Thr49	O	2.72
Thr49	$O^{\gamma 1}$	Phe50	N	2.99
Ile51	O	His56	$N^{\delta 1}$	2.84
Phe59	O	Lys63	N^{ζ}	2.76
Asp62	$O^{\delta 1}$	Lys63	N^{ζ}	2.78
Asp62	O	Gly102	N	2.99[a]
Lys63	N	Phe100	O	3.31[a]
Leu64	N	Phe100	O	3.00[a]
Glu65	$O^{\epsilon 1}$	Arg68	N^{ϵ}	3.30
Glu65	$O^{\epsilon 1}$	Arg68	$N^{\eta 2}$	2.71
Glu65	$O^{\epsilon 2}$	Arg68	N^{ϵ}	2.73
Glu65	N	Gly102	O	2.95[a]
Lys69	N^{ζ}	Asp103	$O^{\delta 1}$	2.39
Leu74	O	Cys78	S^{γ}	3.38
Glu77	$O^{\epsilon 1}$	Arg81	N^{ϵ}	2.82
Glu77	$O^{\epsilon 1}$	Arg81	$N^{\eta 2}$	3.03
Asn84	$N^{\delta 2}$	Ile85	O	3.39
Asp87	$O^{\delta 1}$	Thr88	N	2.86
Asp87	$O^{\delta 1}$	Thr89	N	3.13
Asp87	$O^{\delta 1}$	Thr89	$O^{\gamma 1}$	2.89
Asp87	$O^{\delta 2}$	Thr89	$O^{\gamma 1}$	3.16
Asp92	$O^{\delta 1}$	Ile93	N	2.99
Asp92	$O^{\delta 1}$	Ser94	N	2.89
Asp92	$O^{\delta 1}$	Ser94	O^{γ}	2.74
Ser94	O^{γ}	Asp95	N	3.22
Ile96	O	Arg101	N^{ϵ}	2.86
Ile96	O	Arg101	$N^{\eta 2}$	3.27
Asp103	$O^{\delta 2}$	Tyr104	N	2.75
His107	$N^{\delta 1}$	Glu108	N	3.04
His107	$N^{\delta 1}$	Ala109	N	2.98
Gln115	$N^{\epsilon 2}$	Asp119	$O^{\delta 2}$	2.87
Arg117	N^{ϵ}	Glu120	$O^{\epsilon 1}$	3.24
Lys121	N^{ζ}	Gln124	$O^{\epsilon 1}$	3.04
Arg125	N^{ϵ}	Thr127	$O^{\gamma 1}$	2.68
Tyr132	O	Lys134	N^{ζ}	3.14

[a] Main-chain interaction that does not affect secondary structure.

Table 3

Intermolecular interactions

Residue	Atom	Residue	Atom	Distance (Å)	Symmetry operation
A. *Hydrogen bonds (≤3.4 Å)*					
Arg3	$N^{\eta 1}$	Glu108	$O^{\varepsilon 2}$	3.25	−X,−1/2 + Y,−Z
Arg3	$N^{\eta 2}$	Glu108	$O^{\varepsilon 2}$	3.22	−X,−1/2 + Y,−Z
Glu11	$O^{\varepsilon 1}$	Arg117	$N^{\eta 1}$	3.03	−X,1/2 + Y,1−Z
Glu11	$O^{\varepsilon 1}$	Arg117	$N^{\eta 2}$	3.11	−X,1/2 + Y,1−Z
His16	$N^{\varepsilon 2}$	Glu108	$O^{\varepsilon 1}$	2.81	−X,−1/2 + Y,1−Z
Arg32	$N^{\eta 1}$	Lys80	O	2.82	−X,−1/2 + Y,−Z
Lys33	N^{ζ}	Asn84	$O^{\delta 1}$	2.82	−X,−1/2 + Y,−Z
Ala36	O	Arg101	$N^{\eta 1}$	2.85	−1−X,−1/2 + Y,−Z
Ala36	O	Arg101	$N^{\eta 2}$	3.29	−1−X,−1/2 + Y,−Z
Ile46	O	Arg125	$N^{\eta 1}$	3.37	−1−X,1/2 + Y,−Z
Ile46	O	Arg125	$N^{\eta 2}$	3.27	−1−X,1/2 + Y,−Z
Ile46	O	Thr127	$N^{\eta 1}$	3.33	1−X,1/2 + Y,−Z
Asp62	$O^{\delta 2}$	Lys130	N^{ζ}	2.90	−1−X,1/2 + Y,−Z
Glu73	$O^{\varepsilon 1}$	Lys86	N^{ζ}	2.52	−X,1/2 + Y,−Z
Glu73	$O^{\varepsilon 2}$	Lys86	$N\zeta$	3.22	−X,1/2 + Y,−Z
Glu108	$O^{\varepsilon 1}$	Tyr129	$O\eta$	2.53	−X,1/2 + Y,1−Z
B. *Non-bonded contacts (≤4.0 Å)*					
Glu11	$O^{\varepsilon 1}$	Gly53	N	3.46	−X,1/2 + Y,1−Z
His16	$N^{\varepsilon 2}$	Glu108	$O^{\varepsilon 2}$	3.77	−X,−1/2 + Y,1−Z
Arg32	N^{η}	Lys80	O	3.89	−X,−1/2 + Y,−Z
Arg42	O	Trp128	$N^{\varepsilon 1}$	3.56	−1−X,1/2 + Y,−Z
Ile46	O	Arg125	$N^{\eta 1}$	3.48	−1−X,1/2 + Y,−Z

basic concave surface of the molecule, which includes the glycosylase reaction center, presumably because this surface makes extensive contact with a neighboring molecule in the crystal (Figure 6).

Crystal packing

The packing of the T4 endoV molecule in the monoclinic $P2_1$ lattice is represented in Figure 6, which is viewed along the Z-axis. Consistent with the low water content of 35%, the molecules are rather closely packed. Intermolecular polar interactions and non-bonded contacts with neighboring molecules are summarized in Table 3. The side-chains of Arg3 and His16 near the glycosylase catalytic center form ionic bonds with the carboxyl side-chain of Glu108 belonging to the (−X, −1/2 + Y, −Z) molecule. The two basic side-chains of Arg32 and Lys33 make polar interactions with the Lys80 main-chain and the Asn84 side-chain in the same molecule. The Glu11 side-chain, lying at the rear of the molecule, makes an ionic interaction with Arg117 in the basic concave surface of the (−X, 1/2 + Y, 1 − Z) molecule. The main-chain atoms of Ala36 and Ile46, near the carboxyl end of the H1 helix, are involved in the contact with the neighboring two molecules.

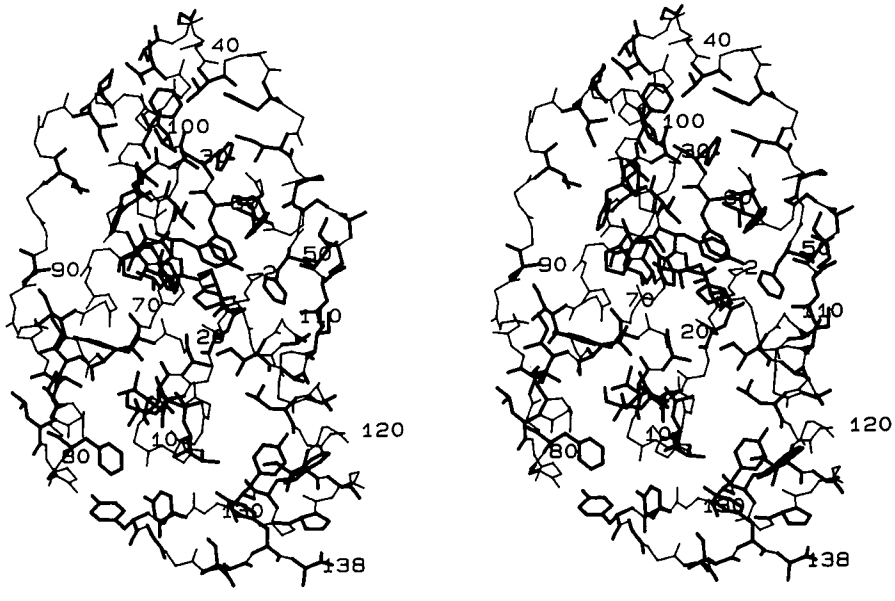


Figure 3. Locations of hydrophobic residues within the T4 endoV molecule, which are indicated by bold lines.

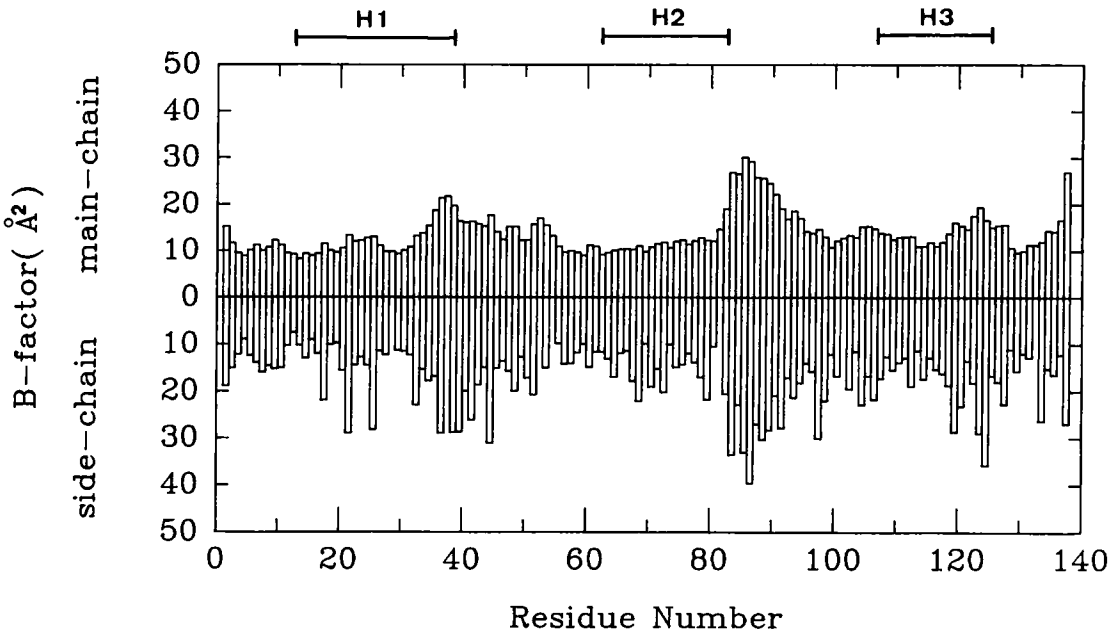


Figure 4. *B*-factor variations *versus* residue numbers. *B*-factors are average for all atoms belonging to each residue. The main three α-helices are indicated by bars.

Basic surface and the catalytic center of the glycosylase

Electrostatic potentials on the accessible surface is shown in Figure 7(a), together with the corresponding backbone structure (Figure 7(b)). It reveals the remarkably biased charge distribution of the enzyme. In fact, many basic residues, such as Arg3, Arg22, Arg26, Lys33, Lys86, Arg117, Lys121, Arg125 and Lys130, are concentrated in the concave surface of the comma-shaped molecule. This feature of the molecular surface is somewhat similar to that of the TFIID protein, the crystal structure of which was determined both in the free form and in complex with a DNA duplex (Y. Kim *et al.*, 1993; J. L. Kim *et al.*, 1993; Burley, 1994). This positively charged concave surface of T4 endoV appears to be essentially consistent with the DNA binding scheme, in which the enzyme scans the non-target sequence by electrostatic interaction with a DNA duplex, prior to specific binding to a pyrimidine dimer (Grunskin & Lloyd, 1986; Dodson & Lloyd, 1989). Further examination of the electrostatic potential revealed a particular region where only a single acidic residue, Glu23, is surrounded by many basic residues. In combination with these structural features of the enzyme, extensive site-directed mutagenesis allowed the identification of the catalytic center of the glycosylase (Morikawa *et al.*, 1992; Doi *et al.*, 1992). The catalytic center comprises Glu23 and the

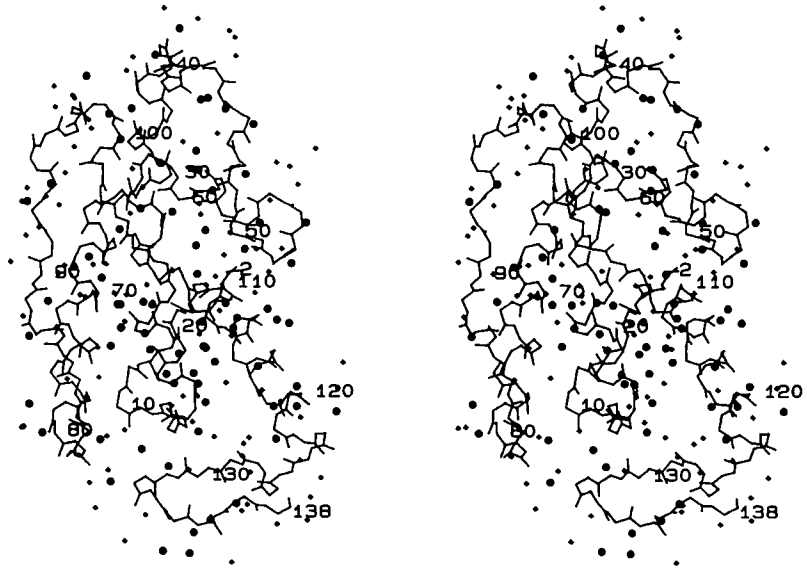


Figure 5. Distribution of water molecules. Filled circles and crosses indicate molecules with temperature factors below and above 30 Å², respectively.

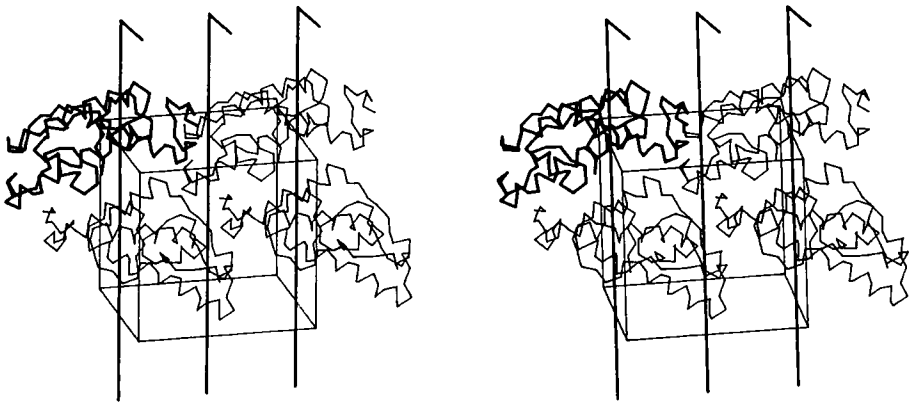


Figure 6. Molecular arrangements within the monoclinic $P2_1$ lattice. The bold line indicates a molecule at (X, Y, Z).

surrounding basic residues, such as Arg3, Arg22 and Arg26. All the mutations at these residues abolish the glycosylase activity (Doi *et al.*, 1992). The R3K mutation, which substitutes similar basic side-chains, increases the K_m value by more than 2000-fold, indicating the almost complete loss of substrate binding. The mutations, such as R22Q, R22K, R26Q and R26K, also exhibit substantial increases of the K_m value. The guanidino groups of Arg22 and Arg26, both of which protrude from the basic concave surface, are directed towards the solvent, while the side-chains of Glu23 and Arg3 lie

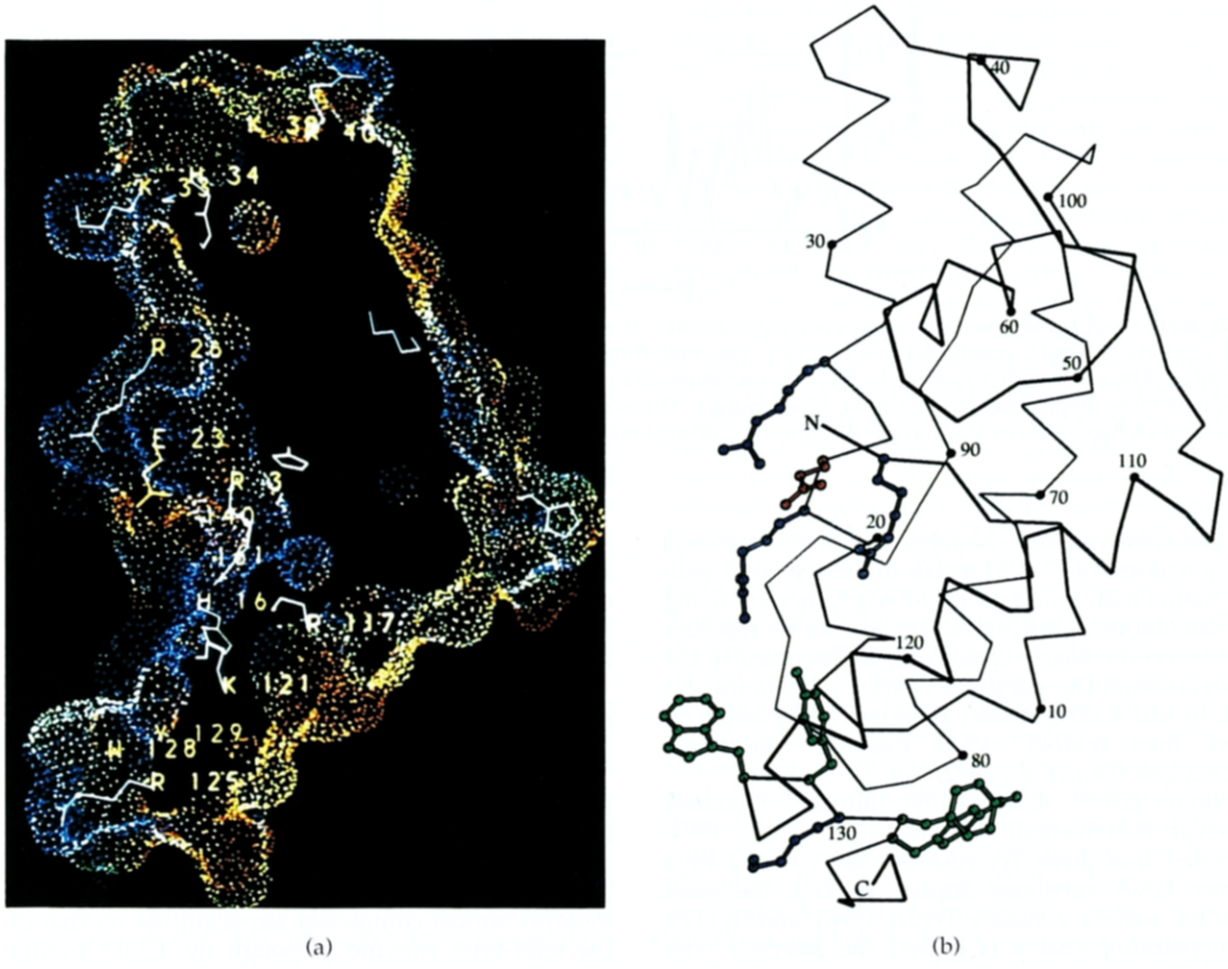


Figure 7. Basic concave surface and the glycosylase catalytic center. (a) Distribution of the electrostatic potential on the solvent-accessible surface. Red denotes electrostatic potential values below –0.1 V, and blue denotes values above 0.1 V; intermediate values follow the spectrum from blue to red. Key residues are indicated by their side-chains. (b) The backbone structure viewed in the same direction. The active center for glycosylase is indicated by the side-chains of Glu23 and the surrounding basic residues, Arg3, Arg22 and Arg26. The WYKYY sequence is shown by the side-chains of the constituents.

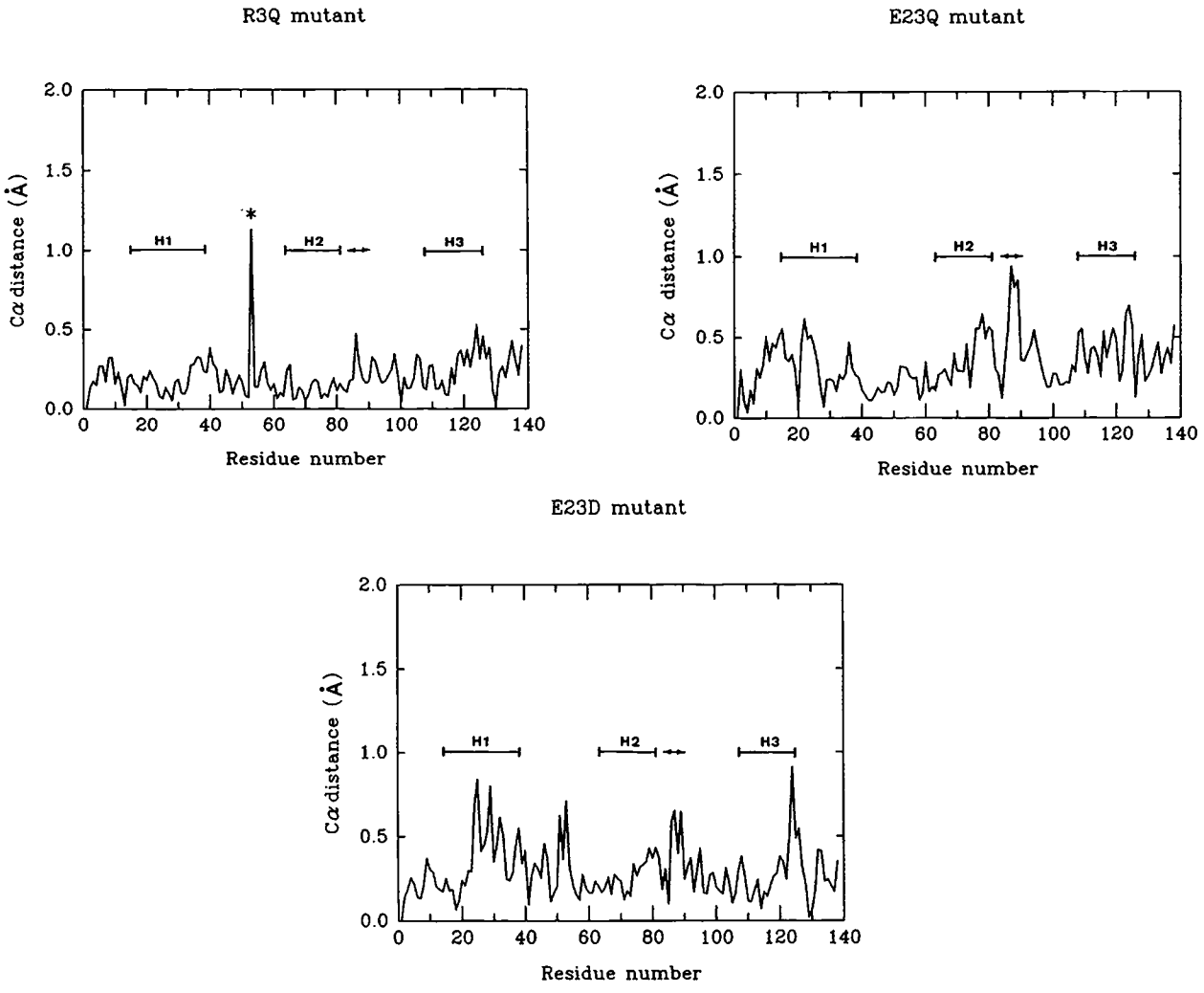


Figure 8. Displacement variations between equivalent C^{α} atoms of each mutant and the wild-type. The glycine residue with a poorly defined C^{α} position is marked by *. The main three α-helices are indicated by bars. The disordered segment is indicated by a bar with arrowheads. Note that residues 23 to 35, belonging to the carboxyl half of the H1 helix, exhibit strikingly large displacements only in the E23D mutant. This corresponds to the increased helical kink (see the text). This structural change induces significant shifts in the carboxyl end of the H3 helix (residues 122 to 123).

on this concave plane. Notably, one of the carboxyl oxygen atoms of Glu23 forms a hydrogen bond with the main-chain imino group of Arg3, thereby fixing the orientation of the carboxyl group. Other residues that lie on the basic surface may be involved in the recognition of the substrate DNA. For instance, the R117Q and K121Q mutations raise the K_m value by more than tenfold, while the K_{cat} values are unchanged (Doi *et al.*, 1992).

Photo-CIDNP (Chemically Induced Dynamic Nuclear Polarization) analyses (Lee *et al.*, 1994) revealed that three Tyr residues are covered by a 10mer DNA substrate duplex. In fact, the basic concave surface contains Tyr21, Tyr61 and Tyr129, demonstrating that it is indeed the interface with DNA. The size of the basic surface implicates that the enzyme would require a considerable length of a substrate DNA duplex for recognition. Actually, the photo-dimer of a dinucleotide, dTpT, cannot be cleaved by the enzyme. In agreement with this finding, NMR analyses (Lee *et al.*, 1994), using the E23Q mutant, demonstrated that the enzyme cannot recognize this dTpT photo-dimer itself, whereas a 10mer DNA duplex containing the single photo-dimer is strongly bound.

Overall structures of active-site mutants

The crystal structures of the three active-site mutants, R3Q, E23Q and E23D, were determined by the molecular replacement method (see Tables 4 and 5). They are very similar to that of the wild-type. In particular, the C^{α} backbones of the R3Q and E23Q mutants almost completely superimpose on that of the wild-type enzyme, although the E23D mutant exhibits some significant differences. The variations of the r.m.s.d. values, which are calculated for the C^{α} atoms between the wild-type and each mutant, are plotted *versus* the residue numbers in Figure 8. The overall r.m.s.d. values for the C^{α} atoms between the

wild-type and each mutant are 0.24 Å for R3Q, 0.27 Å for E23Q and 0.34 Å for E23D. When including the side-chain atoms in the calculation, their values increase to 0.63 Å for R3Q, 0.76 Å for E23Q and 1.10 Å for E23D. The segment (residues 84 to 91) in all the mutants remains disordered. The E23D mutation produced the largest difference in the backbone structure among the three mutants. The striking structural change in this mutant is a substantial increase in the kink of the H1 helix, which is presumably coupled with a newly formed internal hydrogen bond between the replaced carboxyl group of Asp23 and its own main-chain imino group. The E23Q crystal, produced at the same pH as that for the E23D, does not yield such alteration of the backbone conformation. Furthermore, the wild-type enzyme crystal exhibits very constant unit cell parameters between pH 4.5 and 8.0, suggesting that the conformation is invariant in this pH range. Therefore, it is very unlikely that the structural change of the E23D mutant is induced by pH difference in crystal growth.

Mutational effects on the reaction center for the glycosylase

Even including the side-chain conformations, the structural differences in the R3Q and E23Q mutants are confined to a small region around the replaced residue (Figure 9(a) and (b)). The notable features in the R3Q mutant are that the replacement of the guanidino group of Arg3 by the amido group does not induce a significant structural change in the active site, except for some water molecules and the side-chain of Arg22, which points into the solvent and hence presumably would be sensitive to crystal packing. Even the replacement of the negatively charged carboxyl side-chain of Glu23 by the neutral amide group hardly induces a significant structural change, although two water molecules close to the replaced group are slightly moved. The hydrogen bond between the carboxyl side-chain of Glu23 and the main-chain imino group of Arg3 is essentially maintained in the two mutants, although the hydrogen bond length is different in the E23Q mutant.

In parallel with the X-ray structural analysis, difference circular dichroism (CD) measurements were carried out to observe the DNA-binding ability of the enzyme, separately from the glycosyl bond cleavage activity (Doi *et al.*, 1992). The results have demonstrated that the E23Q mutant retains the full DNA binding ability, whereas its cleavage activity is almost completely lost. On the other hand, the R3Q mutant loses the DNA-binding ability. Together with these findings, the structural features of the E23Q and R3Q mutants suggest that the negatively charged carboxyl side-chain of Glu23 is essential for the catalytic activity of the glycosylase and the positively charged guadinino group of Arg3 plays a crucial role in correct substrate binding.

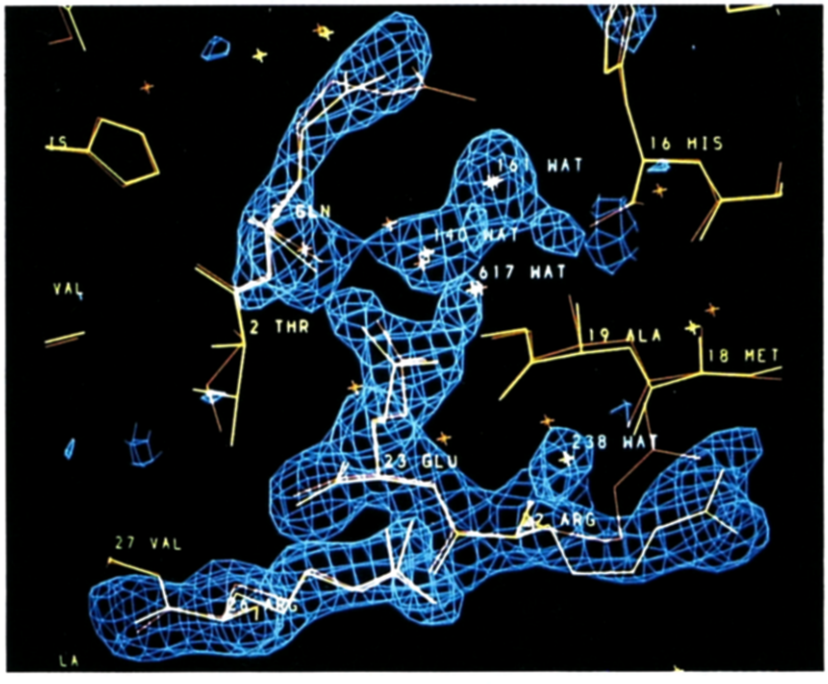
In spite of the considerable change in the backbone structure of the E23D mutant, the reaction center for the glycosylase has minimal structural differences, except for the local change around the substituted side-chain, such as the loss of the hydrogen bond between the carboxyl group of Glu23 and the imino group of Arg3, and the movements of some bound water molecules (Figure 9(c)). The difference CD measurements of this mutant have demonstrated that it almost completely retains DNA-binding ability (N. Hori, personal communication), although the glycosylase activity is lost. This finding implies that the one methylene bond shift of the carboxyl group perhaps induces the loss of the glycosylase activity. Thus, very strict positioning of the functional groups in the active-site pocket appears to be required for the catalytic activity.

Implications of catalytic mechanism

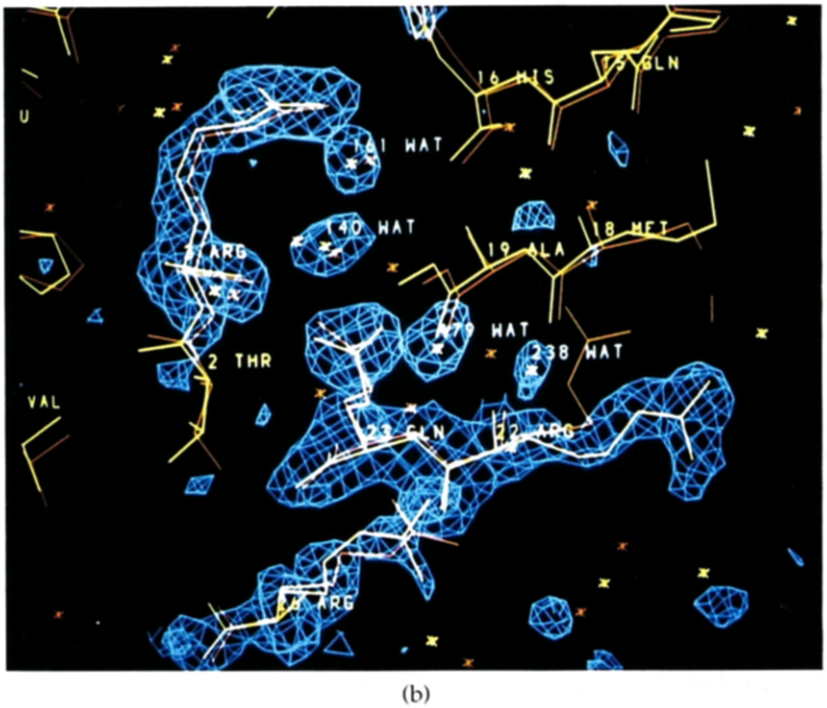
Chemical modification and site-directed mutagenesis studies implicated that the N-terminal α-amino group of the enzyme is involved in the catalytic mechanism (Schrock & Lloyd, 1991, 1993). NMR studies combined with reductive methylation predicted that an imino covalent enzyme-DNA substrate intermediate is formed between the protein N-terminal α-amino group and C-1' of the 5'-deoxyribose moiety of the pyrimidine dimer substrate subsequent to (or concomitantly with) the glycosylase step (Dodson *et al.*, 1993). Furthermore, inhibition experiments by cyanide and the identification of a stable crosslinked substrate-enzyme complex verified that the catalytic mechanism through this intermediate is correct (Dodson *et al.*, 1993). This terminal α-amino group, which is the attacking nucleophile, lies at a distance of 4.4 Å from the free oxygen $O^{\epsilon 1}$ of the Glu23 carboxyl group, and these chemical groups both point into the solvent in a similar direction from the basic concave surface. Combined with functional properties and CD measurements, the structural features of the three active-site mutants imply that the negatively charged carboxyl side-chain of Glu23 is essential for the glycosylase activity, while the side-chain of Arg3 highlights its importance for the substrate binding. Dodson *et al.* (1993) suggested that a role of Glu23 is to stabilize a protonated intermediate with a positive charge in the catalytic pathway. This acidic residue may play an additional role in the abstraction of 2'-hydrogen as a general base (Hori *et al.*, 1992*a*,*b*), since recent kinetic analyses, using a DNA substrate analog with a phosphorothiolate linkage, demonstrated that the β-elimination is catalyzed by an amino acid residue but not by the inter-nucleotide phosphate (Iwai *et al.*, 1995). The cluster of Arg3, Arg22 and Arg26, in particular the contiguity of the latter two residues, would lower the pK_a value of the guanidino side-chains. We assume that one of these basic residues may donate a proton to the 5' side thymine base of the pyrimidine dimer. Thus, the structural features of the wild-type and three active-site mutants are generally consistent with the

reaction scheme proposed by Dodson *et al.* (1993). Furthermore, the global view of the active-site geometry suggests that the substrate binding would not produce a large conformational change, except for slight movements of the Arg22 and Arg26 side-chains, which point into the solvent from the DNA-binding concave surface.

It is interesting to compare the crystal structure of T4 endoV with that of *E. coli* endonuclease III (Kuo *et al.*, 1992), which similarly has a DNA glycosylase activity and an apurinic/apyrimidinic (AP) lyase activity. In particular, the latter reaction in both enzymes proceeds through β-elimination rather than hydrolysis. The endonuclease III consists of the six-helix barrel domain and the [4Fe-4S] domain. This enzyme appears to belong to a new class of [4Fe-4S] DNA repair enzymes (Cunningham *et al.*, 1994). The [4Fe-4S] cluster, which is missing in T4 endoV, is covered by positive charges and hence, in endonuclease III, appears to serve as a structural element for DNA binding (Kuo *et al.*, 1992). Thus, endonuclease III is dissimilar to T4 endoV in terms



(a)



(b)

Figure 9(a)–(b) *(caption on page 12)*

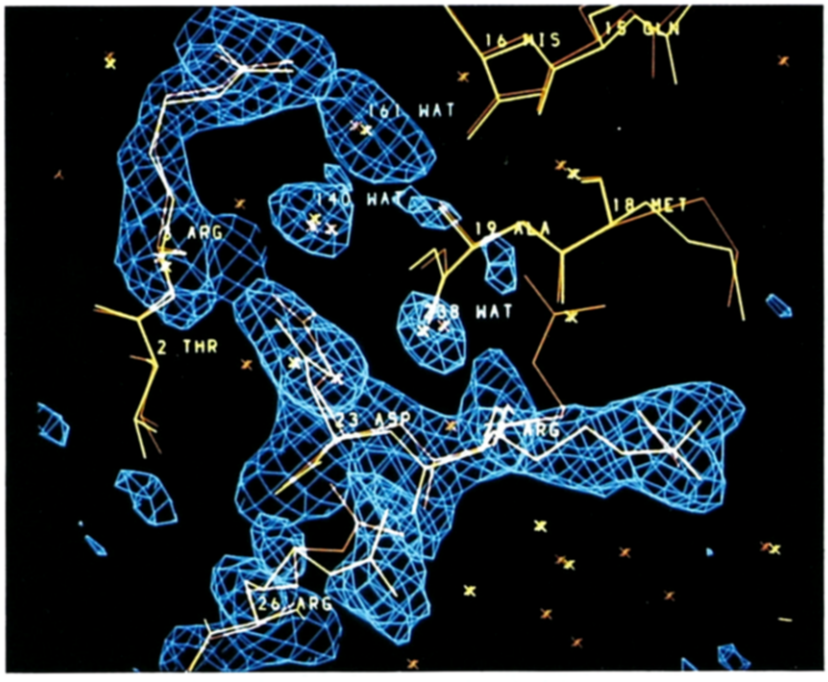


Figure 9(c)

Figure 9. ($F_o - F_c$) electron density maps (contoured at 2.3σ) of the mutant enzymes around the active center. The four residues in the catalytic site (Glu23, Arg3, Arg22 and Arg26) and its nearby water molecules are eliminated from the calculation (omit map). For comparison, the wire models of each mutant (yellow) and the wild-type (red) are superimposed. Crosses indicate the positions of water molecules. (a) The R3Q mutant at 2.2 Å resolution. Except for shifts of the side-chain of Arg22 and two water molecules (238WAT and 617WAT), no substantial movement is observed. (b) The E23Q mutant at 1.8Å resolution. The structural difference is confined to two water molecules (238WAT and 479WAT) and the replaced side-chain of Glu23. (c) The E23D mutant at 2.0 Å resolution. The situation is similar to those in the former two mutants.

of the domain architecture and the backbone structure. There is no significant similarity in the sequences between the two enzymes. Endonuclease III exhibits a very broad spectrum of substrate specificity in the glycosylase reaction and acts upon various damages, such as ring-saturated, ring-rearranged and ring-contracted pyrimidines, whereas T4 endoV strictly recognizes the cyclobutane type of pyrimidine dimer. A putative active site of endonuclease III was located from the position of thymine glycol, which is an inhibitor of the enzyme. The inhibitor was bound to the β-hairpin between the F and G helices (Kuo *et al.*, 1992). It is of interest that this inhibitor interacts with Glu112, which is surrounded by several basic residues. This structural feature is similar to the environments around the glycosylase catalytic center of T4 endoV, where Glu23 is surrounded by several basic residues. In endonuclease III, Lys120 lies in the vicinity of Glu112. This basic residue is the most likely candidate at which a DNA-endonuclease III covalent intermediate might form, whereas the corresponding intermediate in the T4 endoV reaction is formed between the terminal α-amino group and C-1' of the 5'-deoxyribose moiety of the pyrimidine dimer. Thus, the catalytic schemes of these two enzymes appear to be essentially similar, although the amino groups for the covalent bond formation with DNA are different. It remains uncertain whether this scheme may be expanded to other glycosylases involved in DNA repair. However, it seems worthwhile to mention that Dodson *et al.* (1993) pointed out a similarity between T4 endoV and AMP nucleosidase, of which the 3D structure is unknown.

Materials and Methods

Crystallization

The wild-type enzyme was prepared as described (Inaoka *et al.*, 1989). The mutant proteins were purified from an *E. coli* overproducing strain, as reported (Morikawa *et al.*, 1988).

Crystallization of all the T4 endoV proteins was carried out at 4°C by the hanging drop mode of the vapor diffusion technique, as reported (Morikawa *et al.*, 1988). The crystals of the enzyme, including the wild-type and the three mutants, all appeared in a broad pH range from 4.5 to 8.0 without alteration of their apparent shapes, although the optimum pH values for producing large crystals varied among the wild-type and the three mutants. The best crystal of the wild-type enzyme for data collection was grown at pH 4.7, while the R3Q mutant crystal was produced at pH 5.5. On the other hand, both of the E23Q and E23D mutants were crystallized at pH 7.5 to 8.0. It was a serious difficulty that T4 endoV crystals have a strong tendency of forming a cluster, and too few single crystals suitable for X-ray diffraction could be grown. Therefore, a crystal for data collection was often prepared by cracking off a single crystal piece from the remaining major cluster. Particularly, for the mutant proteins, successive seeding was required to grow a crystal of a size suitable for data collection. All the crystals of the mutant proteins belong to the same space group, $P2_1$, as that of the wild-type, whereas they showed significant changes in the unit cell parameters. On the other hand, the unit cell parameters of the wild-type remained invariant regardless of pH between 4.5 and 8.0. Crystal data are summarized in Table 4.

Data collection

For the initial structural analysis by the multiple isomorphous replacement method (MIR), intensity data to

1.6 Å for the native crystal and to 2.5 Å for the derivatives were collected at room temperature, using a CAD4 four-circle diffractometer (Enraf-Nonius) on a generator with a sealed Cu tube operated at 50 kV and 40 mA. The intensities of Bragg reflections were measured in the ω-scan mode, using a scan width of 0.6 to 0.8° and a scan speed of 1 to 3° per minute. The radiation damage was monitored by measuring several standard reflections every 1.5 hours, and X-ray exposure to a crystal was allowed, until a 20% loss of the initially registered intensities was observed. The intensity data were corrected for the absorption effect by monitoring the intensity variations of five near-axial reflections, according to the method of North *et al.* (1968). All the intensity data were scaled and merged using the PROTEIN program package (Steigemann, 1974).

For crystallographic refinements, intensity data up to 1.45 Å resolution were collected using a Weissenberg-type imaging-plate diffractometer for macromolecules (Sakabe, 1991) installed on beam line BL6A at the Photon Factory (Tsukuba, Japan). The wavelength was set at 1.00 Å. The diffraction path was filled with helium gas to avoid air scattering. The diffraction spots were digitized using a BA100 reader (Fuji Film). The data were processed by the use of the program WEIS (Higashi, 1989).

Intensity data from the three mutant crystals were collected using the CAD4 four-circle diffractometer or the Weissenberg-type imaging plate diffractometer for the synchrotron radiation X-ray source. The conditions for data collections were essentially the same as those for the wild-type crystal (Morikawa *et al.*, 1992). The X-ray facilities and data collections for all the crystals of the wild-type and three mutants are summarized in Table 4.

Structural analysis

The crystal structure of the wild-type enzyme was solved by the MIR method and refined at 1.6 Å resolution with the restrained least-squares program PROLSQ (Hendrickson, 1985) as described (Morikawa *et al.*, 1992). A subsequent refinement for the higher-resolution data set to 1.45 Å was carried out, using the same refinement program PROLSQ. The final R-factor for 17,350 reflections ($F_o > 1\sigma(F_o)$) was 0.161 for all data from 6.0 to 1.45 Å. The final refinement parameters at 1.45 Å resolution are summarized in Table 5. A representative portion of the $(2F_o - F_c)$ electron density map after the refinement is shown in Figure 10.

All the crystal structures of the mutant enzymes were solved using the molecular replacement method. The search model was the crystal structure (Morikawa *et al.*, 1992) of the wild-type enzyme refined at 1.6 Å resolution (protein data bank code 1END). The water molecules were excluded from the calculations for the rotation function. The overall *B*-factors of the initial models were fixed at 15 Å^2. Structure factors were calculated on the basis of the model placed in a *P*1 unit cell, with dimensions of $a = 80$ Å, $b = 80$ Å, $c = 80$ Å, $\alpha = \beta = \gamma = 90°$, using the PROTEIN program package (Steigemann, 1974). The rigid body refinement, using the program TRAREF (Huber & Schneider 1985), was applied to the respective mutants. For the R3Q mutant, the model was required to rotate by (0.3, −0.17, 0.01) in the angles (ψ, θ, ϕ) defined by Huber and to shift by −0.05 Å and 0.07 Å along the *x* and *z* axis, respectively. Similarly, the refinement for the E23Q mutant resulted in rotations of (0.24°, −0.33°, 0.69°) in the angles (ψ, θ, ϕ) and in shifts by −0.03 Å and −0.07 Å along the *x* and *z* axis, respectively. The E23D mutant showed larger rotations of (0.45°, −0.98°, 0.43°) and larger shifts of −0.10 Å and 0.27 Å along the *x* and *z* axis, respectively. At this stage, the *R*-factor for the R3Q, E23Q and E23D mutants were 28.1%, 29.5% and 32.0%, respectively. The $(2F_o - F_c)$ electron density maps calculated from the refined models for the three mutants showed clear continuous densities over their entire molecules. Inspection of the $(2F_o - F_c)$ electron density maps allowed the appropriate side-chain to be replaced at each mutation site by the corresponding residue, using the model building program FRODO (Jones, 1978) run on an Evans and Sutherland PS390 graphics system.

The models of the R3Q and E23Q mutants were subsequently refined using the restrained least-squares refinement program PROLSQ (Hendrickson, 1985). For the E23D mutant, the starting model was refined using the program XPLOR (Brunger, 1993) for energy restraints and optional dynamics at high temperature. Forty cycles of positional refinement were carried out before an annealing procedure. Simulated annealing was carried out using the protocol of SLOWCOOL, setting an initial temperature at 1000 K and applying gradual cooling in increments of 25 K to a final temperature of 300 K. After the application of 50 cycles of positional refinements following the annealing simulation, the *R*-factor decreased to 20.9% at 2.8 Å. Subsequent refinement was performed using the program PROLSQ. For all three mutants, the intermediate and resulting models were inspected on the basis of the $(F_o - F_c)$ and $(2F_o - F_c)$ electron density maps and their

Table 4

Crystal data and data collections of the wild-type and the mutant proteins

	Wild-type	R3Q	E23Q	E23D
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell parameters	$a = 41.4$	$a = 41.4$	$a = 41.4$	$a = 41.7$
(a, b, c Å; β deg.)	$b = 40.1$	$b = 40.7$	$b = 40.1$	$b = 40.2$
	$c = 37.6$	$c = 37.4$	$c = 37.4$	$c = 37.1$
	$\beta = 90.01$	$\beta = 90.1$	$\beta = 90.4$	$\beta = 92.0$
Data collection instrument	Weissenberg imaging plate	CAD4	CAD4	Weissenberg imaging plate
X-ray source	Synchrotron radiation	Sealed Cu tube	Sealed Cu tube	Synchrotron radiation
Resolution limit (Å)	1.45	2.2	1.8	2.0
Overall completeness (%)	80.2	84.6	68.9	93.0
R_{merge}[a]	0.048	—	—	0.084

[a] $R_{merge} = \Sigma\Sigma|I_{hi} - \langle I_h \rangle| / \Sigma\Sigma\langle I_h \rangle$, where I_{hi} = measured diffraction intensity, $\langle I_h \rangle$ = mean value of all intensity measurements of (h, k, l) reflections.

Table 5

Structure determination statistics

	Wild-type	R3Q	E23Q	E23D
Resolution (Å)	6–1.45	10–2.2	10–1.8	10–2.0
No. of reflections	17,350	5446	7922	7868
No. of protein atoms	1129	1127	1129	1128
No. of water atoms	172	111	98	101
R-factor	0.161	0.190	0.221	0.207
r.m.s. deviation from ideality				
Bond distance (Å)	0.021	0.015	0.017	0.016
Angle distance (Å)	0.039	0.036	0.036	0.040
Planar 1–4 distance (Å)	0.055	0.047	0.044	0.053
Planar groups (Å)	0.019	0.007	0.004	0.004
Chiral volumes (Å^3)	0.100	0.041	0.021	0.022
Non-bonded contact				
Single-torsion contact	0.173	0.190	0.203	0.187
Multiple-torsion contact	0.200	0.212	0.220	0.200
Possible ($x - y$)	0.181	0.256	0.253	0.256
Torsion angles				
Peptide plane (ω)	3.1	1.3	1.4	1.6
Staggered ($\pm 60°$ or $\pm 180°$)	11.8	21.9	24.1	25.0
Orthonomal ($\pm 90°$)	26.9	35.8	31.8	38.4
Isotropic thermal factors				
Main-chain bond (Å^2)	1.074	0.585	0.789	0.660
Main-chain angle (Å^2)	1.628	0.987	1.331	1.103
Side-chain bond (Å^2)	2.242	0.639	0.783	0.770
Side-chain angle (Å^2)	3.543	1.090	1.200	1.213

models were manually improved using the program FRODO. Finally, the crystal structures of the R3Q, E23Q and E23D mutants were refined for intensity data of 10 to 2.2 Å, 10 to 1.8 Å and 10 to 2.0 Å, respectively. Final refinement statistics of each mutant are summarized in Table 5, together with those of the wild-type.

The refined model coordinates of the wild-type and three active-site mutants have been deposited in the Protein Data Bank (Brookhaven National Laboratory Chemistry Department, Upton, NY, 11973, USA) under accession numbers, 2END (wild-type), 1ENI (R3Q), 1ENJ (E23Q) and 1ENK (E23D).

Accuracy of the model

According to the method proposed by Luzzati (1952), the final *R*-values at 1.45 Å resolution for the wild-type crystal were plotted *versus* resolutions to estimate the

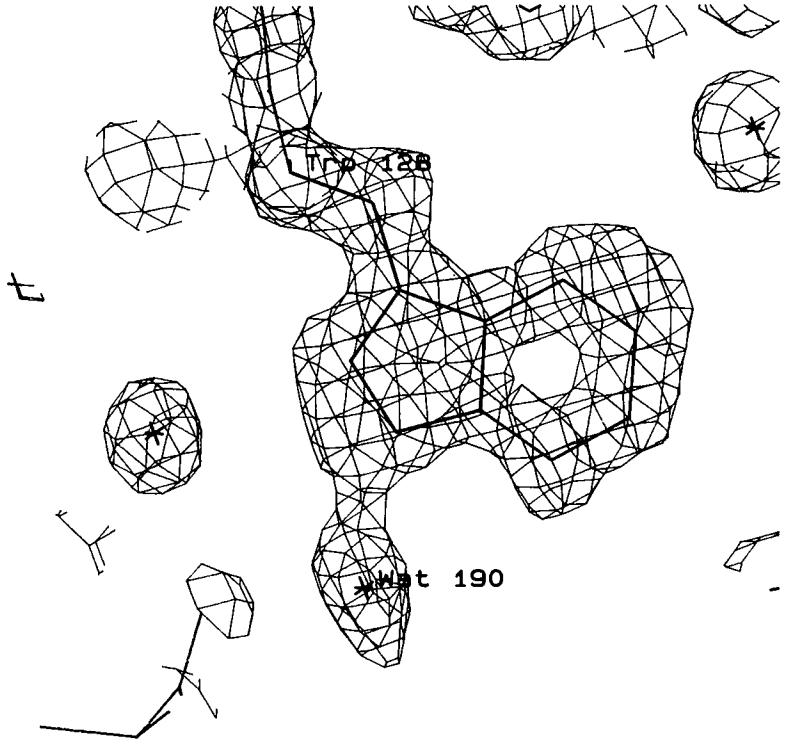


Figure 10. ($2F_o - F_c$) electron density map of the wild-type enzyme at 1.45 Å. The density is contoured at 1.0σ.

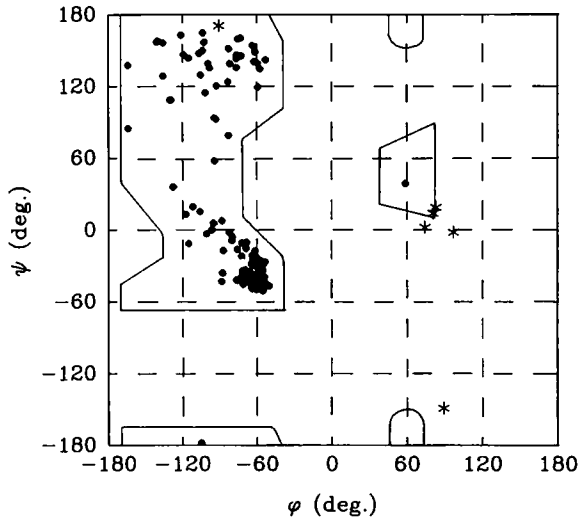


Figure 11. Ramachandran plot for the main-chain torsion angles (ϕ, ψ) of the wild-type enzyme. Filled circles and asterisks indicate non-glycine and glycine residues, respectively.

overall coordinate error. The resulting value is about 0.11 Å. A Ramachandran plot for the main-chain torsion angles (Ramachandran *et al.*, 1963) is displayed in Figure 11. Most of the torsion angles lie within the allowed regions, except for glycine residues, which are often observed in energetically unfavorable regions. The electron densities corresponding to the disordered segment (residues 84 to 91) were considerably improved as compared with the previous analysis at 1.6 Å (Morikawa *et al.*, 1992), but the corresponding residues were still not well defined even in the final map at 1.45 Å resolution. This region appears to have conformational disorder in the crystal.

Likewise, the overall coordinate errors of the three mutant crystals were estimated to be 0.17 Å for R3Q, 0.18 Å for E23Q and 0.18 Å for E23D. Ramachandran plots for all the three mutants show that most of the torsion angles are located in reasonable regions, except for glycine residues. The conformations in the disordered segment remained undefined in the three mutant crystals as well. In both of the E23Q and E23D mutants, the torsion angles of Asp87 in this segment lie within energetically unfavorable regions.

Acknowledgements

We are indebted to Drs N. Sakabe and A. Nakagawa, and to Mr T. Shimizu for their help in data collection. Drs M. Ikehara, H. Nakamura, T. Doi and A. Pähler are acknowledged for their helpful discussions and comments. This work was partly supported by a grant for international collaboration groups from the Human Frontier Science Program.

References

Barlow, D. J. & Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**, 631–640.

Brunger, A. T. (1993). *X-PLOR Version 3.1*. Yale University, New Haven, CT.

Burley, S. K. (1994). DNA-binding motifs from eukaryotic transcription. *Curr. Opin. Struct. Biol.* **3**, 3–10.

Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973). The reverse turn as a polypeptide conformation in globular proteins. *Proc. Natl Acad. Sci. USA*, **70**, 538–542.

Cunningham, R. P., Ahern, H., Xing, D., Thayerm, M. M. & Tainer, J. A. (1994). Structure and function of *Escherichia coli* endonuclease III. In *DNA Damage* (Wallace, S., Houten, B. V. & Kow, Y. W., eds), pp. 215–222, The New York Academy of Sciences, New York, NY.

Dodson, M. L. & Lloyd, R. S. (1989). Structure-function studies of the T4 endonuclease V repair enzyme. *Mutat. Res.* **218**, 49–65.

Dodson, M. L., Schrock, R. D., III & Lloyd, R. S. (1993). Evidence for an imino intermediate in the T4 endonuclease V reaction. *Biochemistry*, **32**, 8284–8290.

Doi, T., Recktenwald, A., Karaki, Y., Kikuchi, M., Morikawa, K., Ikehara, M., Inaoka, T., Hori, N. & Ohtsuka, E. (1992). Role of the basic amino acid cluster and Glu-23 in pyrimidine dimer glycosylase activity of T4 endonuclease V. *Proc. Natl Acad. Sci. USA*, **89**, 9420–9424.

Friedberg, E. & King, J. J. (1971). Dark repair of ultraviolet-irradiate deoxyribonucleic acid by bacteriophage T4: purification and characterization of a dimer-specific-phage-induced endonuclease. *J. Bacteriol.* **106**, 500–507.

Grunskin, E. A. & Lloyd, R. S. (1986). The DNA scanning of T4 endonuclease V. Effect of NaCl concentration on processive nicking activity. *J. Biol. Chem.* **2**, 2696–2710.

Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–271.

Higashi, (1989). The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography, *J. Appl. Crystallog.* **22**, 9–18.

Hori, N., Doi, T., Karaki, Y., Kikuchi, M., Ikehara, M. & Ohtsuka, E. (1992*a*). Participation of glutamic acid 23 of T4 endonuclease V in the β-elimination of an abasic site in a synthetic duplex DNA. *Nucl. Acids Res.* **20**, 4761–4764.

Hori, N., Iwai, S., Inoue, H. & Ohtsuka E. (1992*b*). Photoaffinity labeling of T4 endonuclease V with a substrate containing a phenyl-diazirine derivative. *J. Biol. Chem.* **267**, 15591–15594.

Huber, R. & Schneider, M. (1985). A group refinement proceure in protein crystallography using Fourier transforms. *J. Appl. Crystallogr.* **18**, 165–169.

Inaoka, T., Ishida, M. & Ohtsuka, E. (1989). Affinity of single- or double-stranded oligodeoxyribonucleotides containing a thymine photodimer for T4 endonuclease V. *J. Biol. Chem.* **264**, 2609–2614.

Iwai, S., Maeda, M., Shirai, M., Shimada, Y., Osafune, T., Murata, T. & Ohtsuka, E. (1995). Reaction mechanism of T4 endonuclease V determined by analysis using modified oligonucleotide duplexes. *Biochemistry*, in the press.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kim, J. & Linn, S. (1988). The mechanism of action of *E. coli* endonuclease III and T4 UV endonuclease (endonuclease V) at APsites. *Nucl. Acids Res.* **16**, 1135–1141.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature*, **365**, 520–527.

Kim, Y. Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature*, **365**, 512–520.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallog.* **21**, 572–576.

Kuo, C.-F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P. & Tainer, J. A. (1992). Atomic structure of the DNA repair [4Fe-4S] enzyme endonuclease III. *Science*, **258**, 434–440.

Lee, B. J., Sakashita, H., Ohkubo, T., Ikehara, M., Doi, T., Morikawa, K., Kyogoku, Y., Osafune, T., Iwai, S. & Ohtsuka, E. (1994). Nuclear magnetic resonance study of the interaction of T4 endonuclease V with DNA. *Biochemistry*, **33**, 57–64.

Luzzati, V. (1952). Traitement statisque des erreurs dans la determination des structure cristallines. *Acta Crystallog.* **5**, 802–810.

Macmillan, S., Edenberg, H. J., Radany, E. H., Friedberg, R. C. & Friedberg, E. C. (1981). *denV* gene of bacteriophage T4 codes for both pyrimidine dimer-DNA glycosylase and apyrimidinic endonuclease activities. *J. Virol.* **40**, 211–223.

Manoharan, M., Mazumder, A., Ranson, S. C., Gerlt, J. A. & Bolton, P. H. (1988). Mechanism of UV endonuclease V cleavage of abasic sites in DN determined by ^{13}C labeling. *J. Am. Chem. Soc.* **11**, 2690–2691.

Morikawa, K., Tsujimoto, M., Ikehara, M., Inaoka, T. & Ohtsuka, E. (1988). Preliminary crystallographic study of pyrimidine dimer-specific excision repair enzyme from bacteriophage T4. *J. Mol. Biol.* **202**, 683–684.

Morikawa, K., Matsumoto, O., Tsujimoto, M., Katayanagi, K., Ariyoshi, M., Doi, T., Ikehara, M., Inaoka, T. & Ohtsuka, E. (1992). X-ray structure of T4 endonuclease V: an excision repair enzyme specific for a pyrimidine dimer. *Science*, **256**, 523–526.

Nakabeppu, Y. & Sekiguchi, M. (1981). Physical association of pyrimidine dimer DNA glycosylase and apurinic/apyrimidinic DNA glycosylase essential for repair of ultraviolet-damage DNA. *Proc. Natl Acad. Sci. USA*, **78**, 2742–2746.

North, A. C. T., Phillips, D. C. & Matthews, F. S. (1968). A semi-empirical method of absorption correction. *Acta Crystallog. sect. A*, **24**, 351–359.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configuration. *J. Mol. Biol.* **7**, 95–99.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 168–364.

Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods Phys. Res. sect. A*, **303**, 448–463.

Schrock, R. D., III & Lloyd, R. S. (1991) Reductive methylation of the amino terminus of endonuclease V eradicates catalytic activities. *J. Biol. Chem.* **266**, 17631–17639.

Schrock, R. D., III & Lloyd, R. S. (1993). Site-directed mutagenesis of the NH_2 terminus of T4 endonuclease V. *J. Biol. Chem.* **268**, 880–886.

Steigemann, W. (1974). Die Entwicklung und Anwendendung von Rechenverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen am Beispiel des Trypsin-Trypsininhibitor Komplexes, des freien Inhibitors und der L-Asparginase. Ph.D. thesis, Technical University of Munich.

Steinum, A.-L. & Seeberg, E. (1986). Nucleotide sequence of the *tag* gene from *Escherichia coli*. *Nucl. Acids Res.* **14**, 3763–3772.

Yasuda, S. & Sekiguchi, M. (1970). T4 endonuclease involve in repair of DNA. *Proc. Natl Acad. Sci. USA*, **67**, 1839–1845.

Edited by R. Huber

(Received 12 December 1994; accepted 3 March 1995)