The EMBO Journal vol.14 no.16 pp.4108–4120, 1995

Novel DNA binding motifs in the DNA repair enzyme endonuclease III crystal structure

Maria M.Thayer, Holly Ahern[1], Dongxia Xing[1], Richard P.Cunningham[1,2] and John A.Tainer[2]

The Scripps Research Institute, Department of Molecular Biology–MB4, 10666 North Torrey Pines Road, La Jolla, CA 92037 and [1]Department of Biological Sciences, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222, USA

[2]Corresponding authors

The 1.85 Å crystal structure of endonuclease III, combined with mutational analysis, suggests the structural basis for the DNA binding and catalytic activity of the enzyme. Helix–hairpin–helix (HhH) and [4Fe–4S] cluster loop (FCL) motifs, which we have named for their secondary structure, bracket the cleft separating the two α-helical domains of the enzyme. These two novel DNA binding motifs and the solvent-filled pocket in the cleft between them all lie within a positively charged and sequence-conserved surface region. Lys120 and Asp138, both shown by mutagenesis to be catalytically important, lie at the mouth of this pocket, suggesting that this pocket is part of the active site. The positions of the HhH motif and protruding FCL motif, which contains the DNA binding residue Lys191, can accommodate B-form DNA, with a flipped-out base bound within the active site pocket. The identification of HhH and FCL sequence patterns in other DNA binding proteins suggests that these motifs may be a recurrent structural theme for DNA binding proteins.

Keywords: DNA binding motif/DNA recognition/DNA repair/helix–turn–helix motif/iron–sulfur cluster

Introduction

Endonuclease III from *Escherichia coli*, originally identified as a DNA nicking activity seen after heavy ultraviolet irradiation (Radman, 1976), removes cytosine hydrate from DNA by a DNA glycosylase activity (Boorstein *et al.*, 1989) and cleaves the phosphodiester backbone at an apurinic/apyrimidinic (AP) site via a β-elimination reaction (Bailly and Verly, 1987; Kim and Linn, 1988; Mazumder *et al.*, 1991). Endonuclease III has a broad specificity for DNA base excision repair in that it removes numerous forms of modified thymine and cytosine bases from DNA (Breimer and Lindahl, 1980, 1984; Katcher and Wallace, 1983; Boorstein *et al.*, 1989; Dizdaroglu *et al.*, 1993; Hatahet *et al.*, 1994). Enzymes with similar substrate specificities have been found in another prokaryote (Jorgensen *et al.*, 1987), in yeast (Gosset *et al.*, 1988) and in mammalian cells (Doetsch *et al.*, 1987; Higgins *et al.*, 1987; Ganguly *et al.*, 1990), suggesting that the function of endonuclease III is conserved from bacteria to man.

Endonuclease III is encoded by the *nth* gene, which has been cloned (Cunningham and Weiss, 1985) and sequenced (Asahara *et al.*, 1989), and has been overexpressed and purified in large quantities, allowing for its characterization as a [4Fe–4S] protein (Cunningham *et al.*, 1989) and crystallization (Kuo *et al.*, 1992a) for structural studies (Kuo *et al.*, 1992b). The [4Fe–4S] cluster in the enzyme is apparently not involved in redox chemistry (Cunningham *et al.*, 1989; Fu *et al.*, 1992), so structural studies may aid in understanding its function. Several genes which code for other predicted DNA glycosylases have been sequenced and the protein products have homology to endonuclease III, including the Cys-X_6-Cys-X_2-Cys-X_5-Cys sequence, which ligates the [4Fe–4S] cluster. The adenine-DNA glycosylase coded by the *E.coli mutY* gene (MutY protein) (Au *et al.*, 1989), which removes the adenine from an 8-oxoguanine•adenine mismatch in DNA (Michaels *et al.*, 1992), shows significant homology to endonuclease III (Michaels *et al.*, 1990). The activity of MutY is recovered after renaturation of the apoprotein only when ferrous iron and sulfide are present, suggesting that it also has an iron–sulfur cluster (Tsai-Wu *et al.*, 1992). A homologous MutY protein from *Salmonella typhimurium* has also been identified and sequenced (Desiraju *et al.*, 1993). An archael plasmid from a thermophile, *Methanobacterium thermoformicicum*, has an open reading frame (ORF) that apparently codes for a protein homologous to endonuclease III (Nölling *et al.*, 1992). As it is found next to the genes for a restriction–modification system that creates the modification 5-methylcytosine in DNA, this ORF apparently codes for a thymine-DNA glycosylase that would remove thymine from a G•T mismatched base pair arising from the thermal deamination of 5-methylcytosine. A recently identified ORF from *Bacillus subtilis* also appears to be an endonuclease III homolog (GenBank accession No. U11289). Besides their overall sequence similarities, these endonuclease III homologs show a high level of sequence conservation within a loop that we previously identified as the binding site for free thymine glycol (Kuo *et al.*, 1992b).

Here we report the refined 1.85 Å crystal structure of endonuclease III. In combination with mutational analysis, this structure identifies two catalytically important residues and the enzyme active site pocket. As the loop previously associated with thymine glycol binding is conserved in enzymes which recognize different substrates, we suggest that this loop is not responsible for the endonuclease III specificity. In addition, new results presented here suggest that the endonuclease III iron–sulfur cluster, which appears conserved in homologous enzymes but inactive in catalysis, plays a role in DNA binding. The combination of mutagenesis and structure-based sequence comparisons

Table I. Crystallographic data and refinement statistics

Quality of data	
Resolution (Å)	10.0–1.85
No. of reflections	24 309
No. of observations	259 089
Completeness of data (%)	99.8
$\langle I \rangle \geq 2\sigma$ (%)	85.0
R_{sym} (%)[a]	6.99
Quality of model	
R (%)	18.5
No. of residues	211
No. of atoms	1662
No. of atoms in alternate conformations	12
No. of water molecules	123
r.m.s. bond length deviation (Å)	0.015
r.m.s. bond angle deviation (degrees)	2.8
ϕ,ψ distribution (Laskowski *et al.*, 1993)	
Residues in most favored regions (%)	86.7
Residues in allowed regions (%)	13.3
Morris classification (Morris *et al.*, 1992)[b]	
ϕ,ψ distribution	1
χ_1 standard deviation	1
Hydrogen bond energy	2
Temperature factor distribution	Å^2 (esd[c])
Overall	23.7 (11.5)
Protein	22.6 (10.6)
Backbone	20.7 (8.7)
Water	39.2 (12.1)

[a]R_{sym} is the unweighted R value on intensity between symmetry mates.
[b]The Morris classification is on a scale from 1 to 4, with 1 representing the highest quality structures.
[c]esd stands for estimated standard deviation.

suggests that this DNA repair enzyme contains two novel DNA binding motifs consisting of a helix–hairpin–helix (HhH) and an [4Fe–4S] cluster loop (FCL). Based on the locations of the HhH and FCL motifs, we have built a model for DNA binding that is consistent with an extra-helical damaged base binding in the solvent-filled pocket, which lies between the catalytically important Lys120 and Asp138 side chains. The occurrence of the HhH and FCL motif sequence patterns among other DNA binding proteins suggests that these motifs could be of general importance for other protein–DNA interactions.

Results and discussion

Structural refinement and overview

We refined the three-dimensional crystal structure of endonuclease III to an R factor of 18.5% for diffraction data from 10–1.85 Å resolution (Table I). The 2 Å starting model (Kuo *et al.*, 1992b) was rebuilt into $2F_o - F_c$ and $F_o - F_c$ maps between cycles of least squares refinement in XPLOR (Brünger *et al.*, 1987) (Figure 1). Stereochemistry was monitored during refinement in XPLOR and finally checked using the PROCHECK program (Morris *et al.*, 1992; Laskowski *et al.*, 1993) (Table I). The root mean square (r.m.s.) deviations from ideal geometry are 0.015 Å for bond lengths and 2.8° for bond angles. The stereochemistry is excellent, with all the backbone dihedral angles falling into the most favored core region, 86.7%, or allowed region, 13.3%, of a ϕ,ψ plot (Morris *et al.*, 1992). The mean temperature factor is 22.6 (σ = 10.6) Å^2 for the protein and 20.7 (σ = 8.7) Å^2 for the peptide backbone. Temperature factors increase along the backbone from residue 209–211; however, all of these residues have ϕ,ψ values inside the allowed region (Morris *et al.*, 1992).

Endonuclease III has two domains and overall dimensions of 35×35×60 Å^3 (Figure 2). The two domains were named after their structural features, the [4Fe–4S] cluster and the 6-helix barrel (Kuo *et al.*, 1992b). The secondary structure is primarily α-helical and the helices are sequentially lettered αA–αJ. The αD helix is kinked with Gly68 in a non-helical conformation between the αD1 and αD2 portions of the helix. The 6-helix barrel domain consists of residues 22–133 (helices αB–αG); and the [4Fe–4S] cluster domain consists of residues 1–21 and 134–211 (helices αA, αI–αJ and the loop around the [4Fe–4S] cluster). The [4Fe–4S] cluster is nestled against helices αH and αJ and is protected from solvent on the other side by the C-terminal loop, which was substantially rebuilt in this refinement. Between the two domains there is a deep groove (Figure 3) with a solvent-filled pocket and a nearly continuous solvent channel along αG (Figure 4).

Active site pocket

Within the interdomain groove and near the thymine glycol binding loop there is a solvent-filled pocket (Figure 4), which is lined by polar side chains. This deep pocket is part of a solvent channel that has seven water molecules in the pocket, two water molecules in an internal cavity and three water molecules in a shallow depression on the surface opposite the interdomain groove. A 'gateway' formed by Val124, His176 and His177 separates the internal cavity from the deep water-filled pocket. Asn127 'bridges' the gap between the internal cavity and the solvent on the surface opposite the groove by forming one hydrogen bond to a water in the cavity (2.72 Å) and one to a water in the depression (2.80 Å).

To test the possibility that this pocket is the enzyme active site, mutant enzymes were made, expressed and characterized for two charged side chains at the mouth of the pocket: Lys120 and Asp138 (Figure 4). In addition, a Glu112→Gln, E112Q, mutant was expressed to test our previous suggestion that Glu112 could be catalytically important (Kuo *et al.*, 1992b). In the refined structure, Glu112 forms a salt bridge with Arg119 that links the αF and αG helices. Consistent with this structural role for Glu112, the E112Q mutant precipitated and could not be purified. Because Glu112 is not conserved and is not adjacent to the probable active site pocket, further attempts at mutagenesis of Glu112 were abandoned. Based upon computer graphics modeling, Lys120→Gln (K120Q) and Asp138→Asn (D138N) mutations were chosen to alter side chain chemistry while minimizing any potential structural effects. The K120Q mutation shows a 100 000-fold decrease in k_{cat}, indicating that the enzyme is catalytically impaired (Table II). The K_m is lower than that of the wild-type enzyme, which shows that substrate binding is not decreased. The D138N mutation shows a similar behavior: k_{cat} is reduced ~100-fold and the K_m is only 4-fold higher, again suggesting that substrate binding is relatively unimpaired while catalysis is compromised (Table II). The ellipticity of the D138N mutant, as measured by circular dichroism, was found to be within experimental error of the ellipticity of the native protein, indicating the structural integrity of this mutant. The

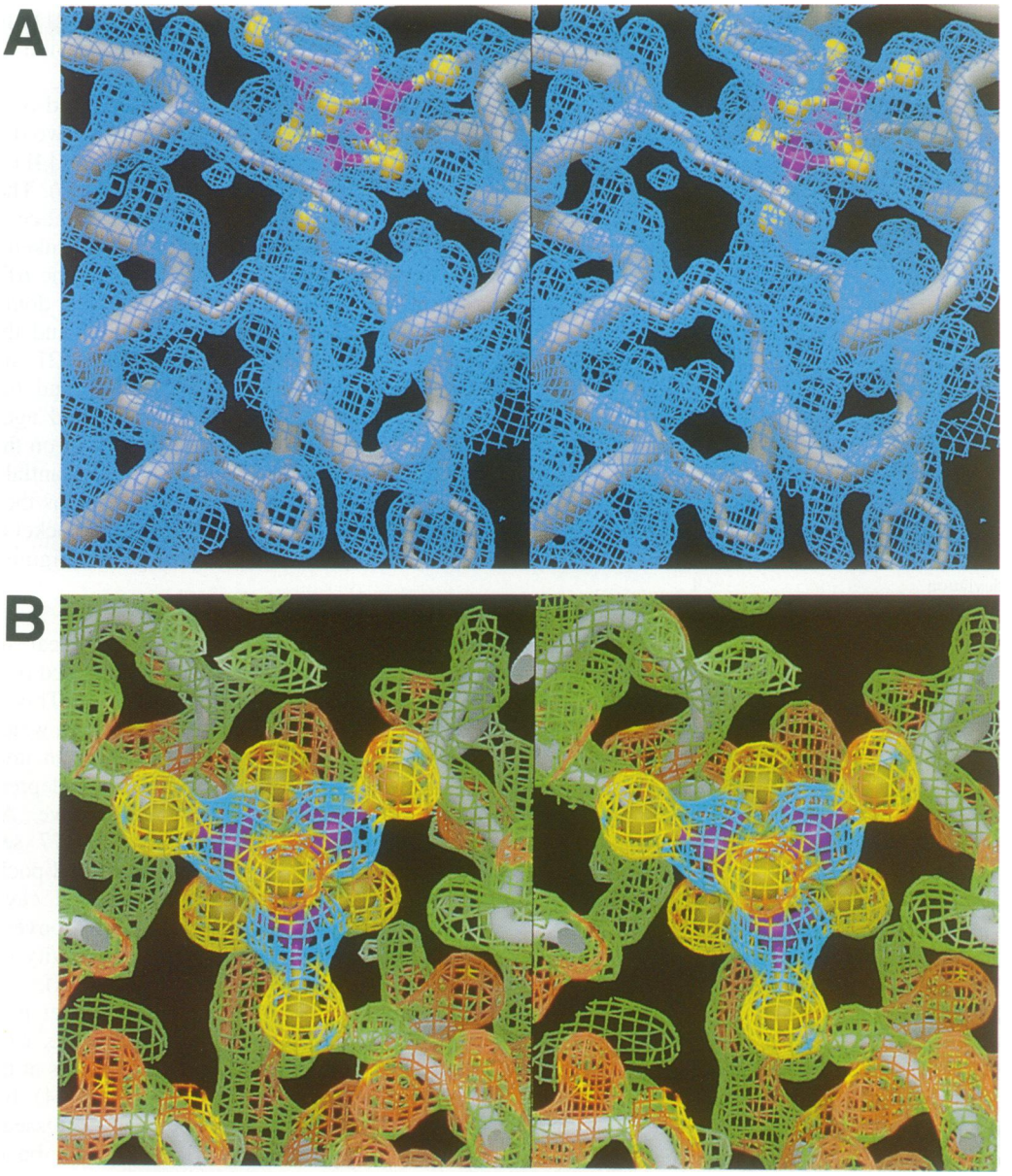

Fig. 1. Stereo electron density maps and the refined model. The $2F_o - F_c$ electron density of endonuclease III is contoured at 1.5σ and shown with the refined model Cα backbone (white tubes). (**A**) The αH helix and side chains (from bottom right to top left) of Tyr185, His140, Arg143, Arg147 and Tyr205 show the quality of the structure and the well-ordered side chains protecting the [4Fe–4S] cluster from solvent. The carbonyl oxygen electron density is also clearly visible (atomic model not shown). (**B**) Electron density colored by gradient in the region of the [4Fe–4S] cluster reveals that the gradient can be used to distinguish atom types (Purvis and Culberson, 1986): iron (blue contours and purple spheres) has the highest gradient, then sulfur (yellow) and oxygen (red) and carbon (green) has the lowest gradient.

binding of DNA to the mutant and wild-type proteins was found to be similar by gel shift assays, also indicating that the structure was not perturbed by the mutations. Based upon the identification of this potential active site pocket, we examined the surrounding surface for regions that might be associated with DNA binding.

HhH motif

Lys120 is positioned at the N-terminus of helix αG and over the mouth of the deep solvent-filled pocket, such that Lys120 could act in a transimination reaction (Kow and Wallace, 1987) on a damaged base bound in the pocket (Figure 4). The αG helix is part of a αF–αG helical hairpin (Figure 2) that has high sequence conservation among the homologous DNA repair enzymes (Figure 3). In addition, the αF–αG interhelical turn has been identified as the binding site for free thymine glycol (Kuo *et al.*, 1992b). Based on this evidence, we propose this helical hairpin as a DNA binding motif and named it the HhH motif.

Within the overall enzyme fold the HhH motif is located in the interdomain interface. The HhH motif (Figure 5), residues 108–127, consists of helices αF, Arg108–Ala113, and αG, Arg119–Phe130, with a five residue connection, Leu114–Gly118, which includes a type II β-turn, Leu114–Val117. The hydrophobic core between the two helices Leu111, Val125 and Leu126 allows helix αF to pack against the second turn of helix αG. This proposed HhH DNA binding region in the interhelical turn has greater local flexibility than other surrounding helical regions.

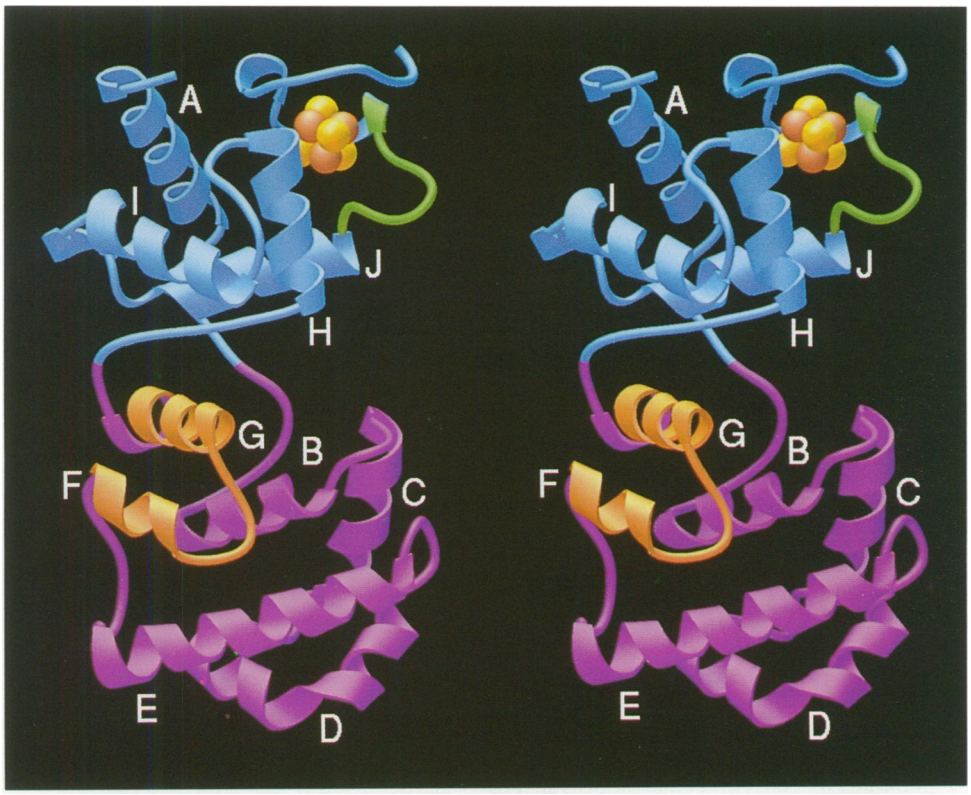

Fig. 2. Endonuclease III fold, domains and DNA binding motifs. The 6-helix barrel domain (pink) consists of antiparallel helices αB–αG connected in a +1 topology and contains the HhH motif (orange). Helices αA and αH–αJ and the 4[Fe–4S] cluster (red and yellow spheres) comprise the 4[Fe–4S] cluster domain (blue) containing the FCL motif (green). Ribbons representing helices and tubes representing loops were calculated using RIBBONS (Carson, 1991).

Only a single hydrogen bond joins the turn between these helices and adjacent protein regions: Leu114 N, the first atom of the turn, hydrogen bonds to Leu111 O (2.98 Å). The mobility suggested by these limited hydrogen bond and other stabilizing interactions is evident in the main chain temperature factors (30.6–38.6 Å^2). The β-turn is, however, stabilized by the hydrophobic packing of Leu114 and Val117.

To determine whether this endonuclease III motif might occur within other proteins, sequences similar to the HhH consensus sequence, LXALPGVGRXTAXALL, were identified using the BLAST program (Altschul *et al.*, 1990; Henikoff and Henikoff, 1992) and compared (Figure 6). Of the 16 highest scoring sequences, 13 were in enzymes that act on DNA and one was in a protein of unknown function. Only these 14 sequences with possible DNA binding activities were included in the analysis. In order to determine if the HhH motif was part of a structural conservation of the 6-helix barrel domain, sequence alignments (Lipman *et al.*, 1989) and profile searches (Bowie *et al.*, 1991) of the fragment of each protein starting 80 residues prior to and ending four residues after its HhH motif yielded significant similarity for only the endonuclease III homologs. This implies that the HhH motif does not include other portions of the 6-helix barrel domain. For our analysis the amino acid residues in the motif were numbered from HhH1 to HhH20 (Figure 6). Seven features resulting from this analysis appear significant: (i) the interhelical packing residues at HhH4, HhH18 and HhH19 (Leu111, Val125 and Leu126 in endonuclease III) are large hydrophobic residues, with the exception of one Ala at position HhH19; (ii) large hydrophobic residues are conserved for HhH7 and HhH10 (Leu114 and Val117); (iii) HhH8 and HhH9 (Pro115 and Gly116), the central residues of the β-turn, are conserved; (iv) the Gly at HhH11 (Gly118) is conserved and may play a role in the flexibility of the interhelical loop; (v) the Thr at HhH14 (Thr121), which has oxygen Oγ1 exposed to solvent and methyl group Cγ2 interacting with hydrophobic residues, is conserved in all but one case, where it is an isosteric residue, Val; (vi) there is a strong preference for a Lys or an Arg at HhH12; (vii) the Pro–Gly sequence of the 5′→3′ exonuclease domain of the bacterial DNA polymerases at HhH5 and HhH6, the last two residues of the αF helix, may distort the end of the helix and the interhelical loop to allow for the appropriate DNA interaction for these enzymes. Although proline is not classified as a helix forming residue, Pro does occur as the penultimate residue of helices (Richardson and Richardson, 1988), such as in helix αC in this endonuclease III structure, and is compatible at HhH5.

Comparison of the HhH and HtH motifs

The proposed HhH motif is similar to, but distinct from, the previously described HtH motif (Figure 7). Both of these motifs have two helices connected by a short turn. In the HtH motif the second or recognition helix binds to DNA with the helix in the major groove of the DNA (Ohlendorf *et al.*, 1982). Interhelical packing in the HtH motif and specific contacts between the DNA and various residues throughout the protein mediate recognition of the DNA (Brennan and Matthews, 1989). The HhH motif is

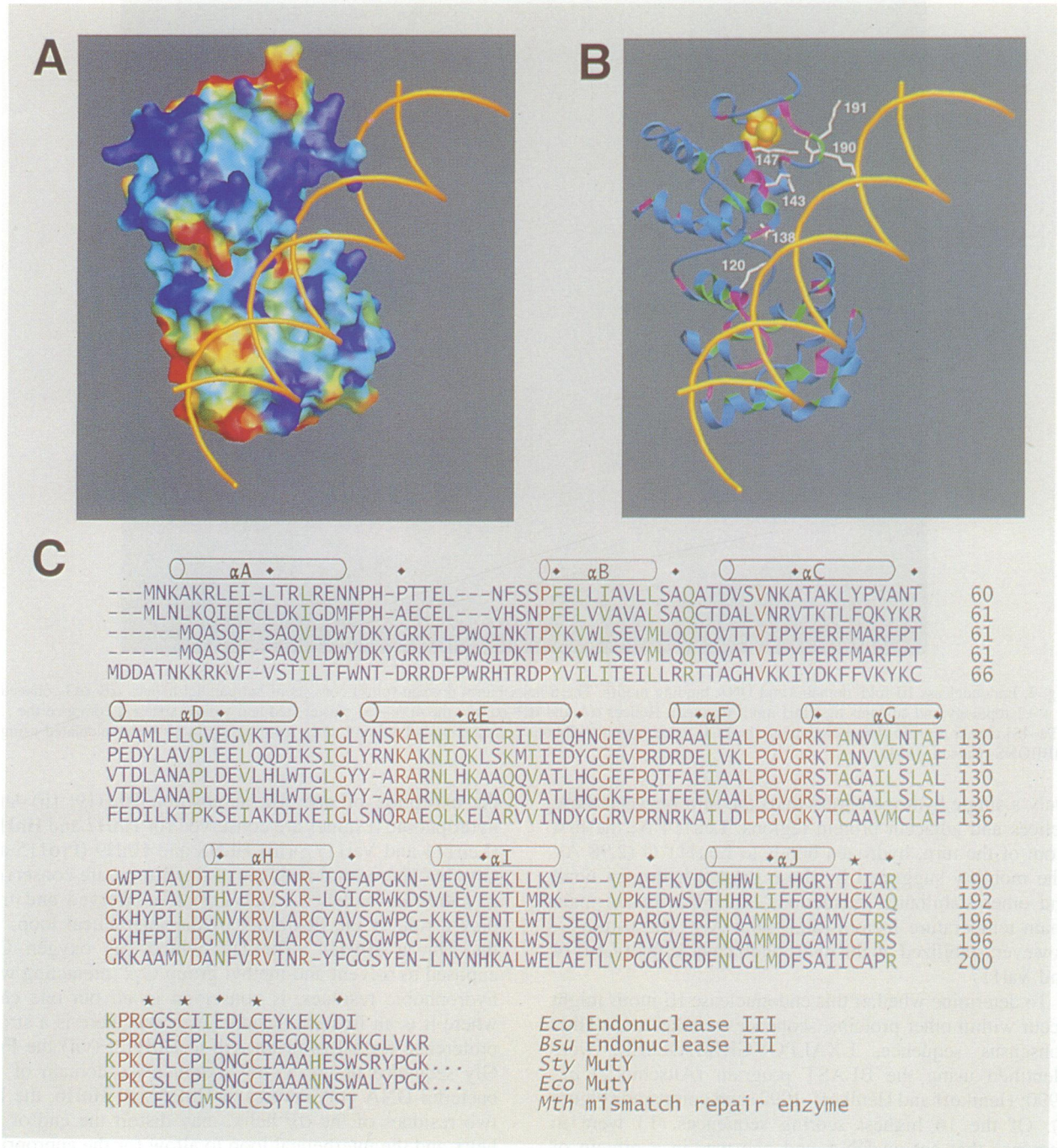

Fig. 3. Endonuclease III interdomain groove, electrostatic surface, sequence conservation and DNA binding model. The endonuclease III molecular surface (**A**) is colored by the Coulombic electrostatic potential with a gradient from dark blue for $+1\sigma$ to red for -1σ and the sequence homology mapped onto a ribbons diagram of the structure (**B**) is colored as indicated by the sequence alignment (**C**), with the homologous residues in red, similar residues in green and unconserved residues in blue. (A) and (B) have the same orientation so as to show the relative positions of the interdomain groove and conserved residues. In two places in the sequence alignment similar residues with an offset of one residue were counted as similar. B-DNA (phosphate backbone in yellow) is modeled against the HhH and the FCL motifs such that the backbone is in the interdomain groove near the solvent-filled pocket, which is next to Lys120.

distinct from the HtH motif because it interacts with DNA differently, it has a different sequence pattern and it is found in a different class of proteins, primarily those involved in enzymatic activity on DNA, rather than those with a regulatory role.

There are major differences between the HtH and the new HhH motifs in the turn between the helices, in the surface exposure of the second helix and in the packing of the two helices. First, there is a type II β-turn in the HhH motif that is not dependent on either helix for hydrogen bonding. In contrast, the turn between the helices in the HtH motif is dependent on the left-handed helical conformation of the Gly at HtH9, which is not part of a β-turn. Next, although the second helix in the HhH motif is exposed to solvent along its entire length, the helix does not protrude from the surface of the protein and therefore cannot lie in the major groove of the DNA in a similar manner to the recognition helix of the HtH motif.

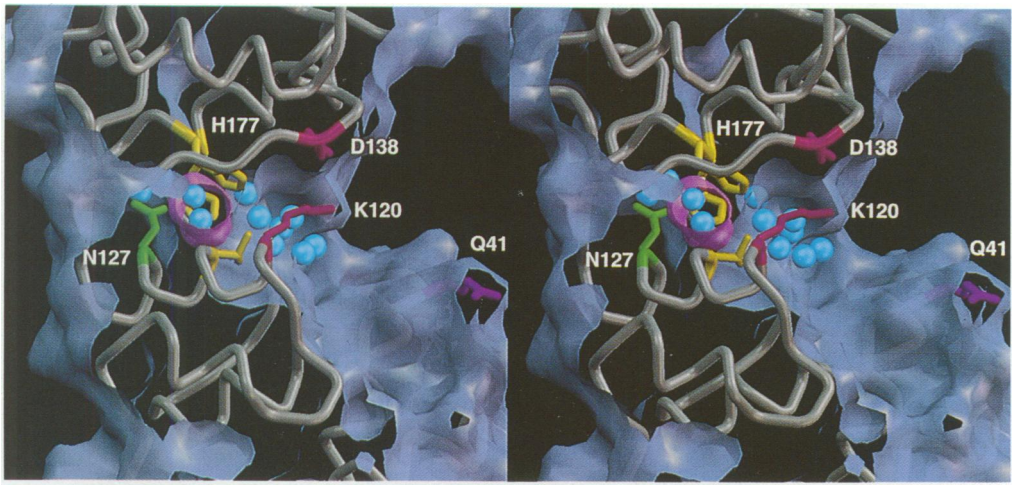

Fig. 4. Active site residues at the mouth of the solvent-filled pocket and solvent channel. The solvent-accessible molecular surface (blue-gray) near the HhH motif, the internal water cavity (purple surface) and solvent-filled pocket are shown with ordered water molecules (blue) from the endonuclease III structure. Sequence conserved Lys120 and Asp138 (red), at the mouth of the pocket, are important for the DNA lyase activity. Gln41 (magenta), conserved as Gln or Arg, is also at the mouth of the pocket and may be important for DNA binding. Val124, His176 and His177 (yellow) form a gateway between the solvent pocket and the internal cavity. Asn127 (green) forms a hydrogen bond bridge between the internal water molecules and the water molecules on the opposite side of the protein from the water pocket.

This and the positive charge at HhH12 (Arg119 in endonuclease III) suggests that DNA recognition is at the N-terminal end of this helix and in the loop between the two helices. Finally, the packing of the two helices is also affected by the interhelical turn and the positions of the interhelical hydrophobic contacts. In the HtH motif the first helix packs partially along the loop between the helices and the core of the hydrophobic interhelical packing is approximately one turn down the recognition helix. In the HhH motif the first helix packs approximately two turns down the second helix and the conserved hydrophobic residues involved in this packing are also further down the second helix. These differences in the two motif structures lead to a difference in the interhelical packing angles for the HtH motif (107.5°, s = 6.5°) and the endonuclease III HhH structure (125.6°) (Table III).

The differences in the tertiary structure of the HtH and HhH motifs are reflected in differences in their primary structure. The method of Dodd and Egan (1987) was used to analyze the relatedness of the HhH sequences retrieved from the homology searches (Figure 6) to the HtH motif. The highest scoring of the HhH sequences fell in the lowest scoring of the evaluated groups of proteins. This group appeared to have a background proportion of proteins with similarity to the HtH motif (≤7%). These differences in primary structure can be explained by the differences in the packing features described above (Figure 5). The key interhelical packing residues in both motifs are conserved hydrophobic residues, but they are further apart in the HhH motif (HhH4 and HhH18–19) than in the HtH motif (HtH4 and HtH15) by one turn of the second helix. In addition, hydrophobic residues in the interhelical loop are conserved in both motifs, but are one residue further apart in the HhH motif (HhH7 and HhH10) than in the HtH motif (HtH8 and HtH10).

A structural feature that has similarities to both the HhH and the HtH motifs occurs in the c-*myb* proto-oncogene product Myb. The structure of Myb from both NMR (Ogata *et al.*, 1992) (coordinates not available) and model building analysis (Frampton *et al.*, 1991) has been

Table II. Wild-type and mutant kinetic data

Enzyme	k_{cat} (s^{-1})	K_m (M)
Wild-type	1.1	6.1×10^{-8}
K120Q	8.1×10^{-6}	1.4×10^{-8}
D138N	8.5×10^{-3}	2.5×10^{-7}
K191E	7.0×10^{-1}	8.3×10^{-6}

reported. The Myb structure has two helices that are connected by a β-hairpin turn that is like the turn in the HhH motif and unlike the turn in the HtH motif. An additional difference between the Myb structure and the HtH motif is a difference in the relative orientations of the two helices (Ogata *et al.*, 1992). However, the environment of the recognition helix in the Myb structure is more closely related to the HtH motif than the HhH motif in two respects: (i) this helix is a protruding surface helix that can bind along the major groove of the DNA; (ii) the hydrophobic interhelical core residue, Ile40 (Protein Data Bank), corresponds to the conserved hydrophobic interhelical packing residue at HtH15. Additional residues that may affect interhelical packing are the positioning of Ile28 at motif position HtH3, rather than HtH4, and the interactions of two Trp residues at motif positions HtH –1 and HtH19. Finally, the hairpin turn in the Myb structure is not packed against a domain and is accessible to solvent from both sides, rather than from only one side as in the endonuclease III structure.

Structure and environment of the [4Fe–4S] cluster

The endonuclease III iron–sulfur cluster, which lies at the other end of the interdomain groove from the HhH motif, appears to be conserved in homologous enzymes but inactive in catalysis, so its structural characterization may provide clues to its function. The [4Fe–4S] cluster is ligated by four Cys residues, Cys187, Cys194, Cys197 and Cys203, in a pattern unique to the homologs of endonuclease III, Cys-X_6-Cys-X_2-Cys-X_5-Cys (Kuo *et al.*, 1992b). The cluster ligands, the interligand loops, the C-

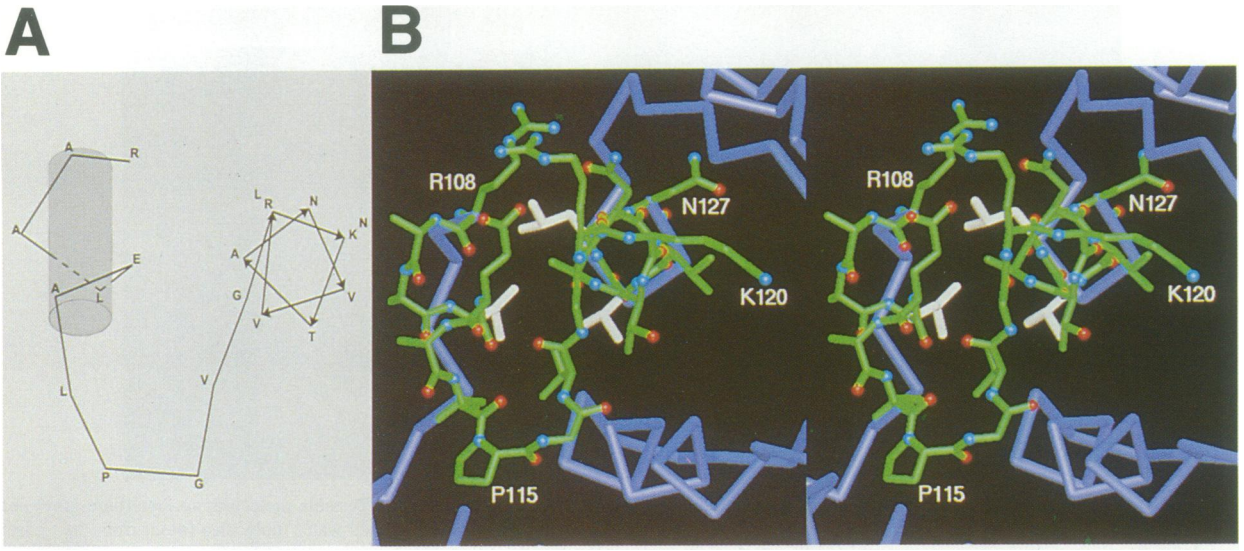

Fig. 5. HhH motif. (**A**) A schematic diagram of the HhH motif with helix αF (shaded cylinder), the β-hairpin turn and helix αG (helical wheel). One letter amino acid codes show the residue types. (**B**) The stereo pair of the HhH motif is labeled at residues Arg108 (first motif residue), Asn127 (last motif residue), Pro115 (Pro–Gly sequence in the β-hairpin) and at Lys120 (important for DNA lyase activity). The HhH motif atomic structure (green bonds with blue nitrogen and red oxygen spheres) is shown in the context of the surrounding endonuclease fold. Note that the second helix of the HhH motif is not protruding.

protein	code	res no.	sequences	score
motif numbering			4· 89 15 19	
consensus HhH motif			LxALPGVGRxTAxALL	
Eco endonuclease III	J02857	108	raa LeALPGVGRkTAnvVL n	51
Mth mismatch repair enzyme	X68366	114	rka IldLPGVGKyTcaAVM c	40
Eco MutY	X52391	108	fee VaALPGVGRsTAgAIL s	55
Sty MutY	M86634	108	fae IaALPGVGRsTAgAIL s	56
Bsu endonuclease III	U11289	109	rde LvkLPGVGRkTAnvVV s	43
Eco 3-MeA glycosylase II	K02498	206	mkt LqtTPGIGRwTAnyfa l	37
Eco RuvA	P08576	107	vga LvkLPGIGKkTAerLI v	42
Eco DNA polymerase I	J01663	187	sdn IpgVPGVGekTAqALL q	43
Bca DNA polymerase	D12982	187	sdn IpgVPGIGekTAvkLL r	37
Hsa replication factor C	L23320	646	aal LsgpPGVGKtTtaSLV c	37
Mmu replication factor C	X72711	630	aal LsgpPGVGKtTtaSLV c	37
Taq hypothetical DNA bp	M33159	191	aev LmALPGVGpqvAaAVL a	44
Mla pilin gene invert. protein	M34367	194	asl LatIPGIGKkTlphLL v	37
Sco hypothetical protein	P19780	273	spv LtSMPGVGvrTAavLL v	43
San Oxidoreductase	M96551	46	vqe LaAagGaGRvTAeALd l	37
Rca beta-ryanodine bp	D21071	1646	esk rhgLPGVGRsTclksd l	37
			F G	
Hsa Myb		167	aei akeLPGrtdnaiknhw n	
434 Cro - HtH		19	qte LAtkAGVkqqsIqlie a	

Fig. 6. HhH motif sequences in other proteins. The protein name, an NCBI code, starting residue number, the sequence with residue matches in upper case letters (exact homology in bold, β-turn highlighted) and the BLOSUM62 score for the BLAST search results (central column of sequence block) are shown. Helices αF and αG are shown as cylinders below the sequences retrieved from the BLAST search. The motif was numbered with the left-handed Gly at position HhH9, which puts the hydrophobic core residues at positions HhH4, HhH18 and HhH19. The proteins are listed in order from top to bottom: the three proteins used to derive the consensus sequence, the other two homologous DNA repair enzymes, other enzymes with a probable DNA binding function and the two proteins with no probable DNA binding function. Sequences from two proteins that have structural characteristics of the HtH motif (listed below the helix cylinders) are Myb, a DNA binding protein with a β-turn between two helices, and Cro, a representative protein containing the HtH motif (listed with structurally important residues for the HtH motif capitalized). HhH, helix–hairpin–helix; 3-MeA, 3-methyladenine; bp, binding protein; invert., inverting; *Eco*, *Escherichia coli*; *Mth*, *Methanobacterium thermoformicicum*; *Sty*, *Salmonella typhimurium*; *Bsu*, *Bacillus subtilis*; *Bca*, *Bacillus caldotenax*; *Hsa*, *Homo sapiens*; *Mmu*, *Mus musculus*; *Taq*, *Thermus aquaticus*; *Sco*, *Streptomyces coelicolor*; *San*, *Streptomyces antibioticus*; *Rca*, *Rana catesbeiana*.

terminal residues (203–211), helix αH and the C-terminus of helix αJ form a hydrophobic pocket that protects the [4Fe–4S] cluster from water. There are five hydrogen bonds to the [4Fe–4S] cluster and the Cys Sγ ligands: one to the cluster and one to each of the cluster ligands (Table IV).

The [4Fe–4S] cluster with its four Cys ligands has a net negative charge of –2 that sequesters positively charged residues. The seven positively charged residues (143, 147, 190, 191, 193, 206 and 208) near the cluster and five negatively charged residues (200, 201, 204, 207 and 210) have a net +2 charge that effectively counterbalances the cluster charge. Of the negatively charged residues, only Glu207 and Asp205 are conserved in two out of the five homologs, but the positively charged residues are highly conserved. Two positive residues, Arg143 and Arg147, are absolutely conserved. These residues donate a hydrogen bond to a cluster ligand, Arg147 Nη2–Cys187

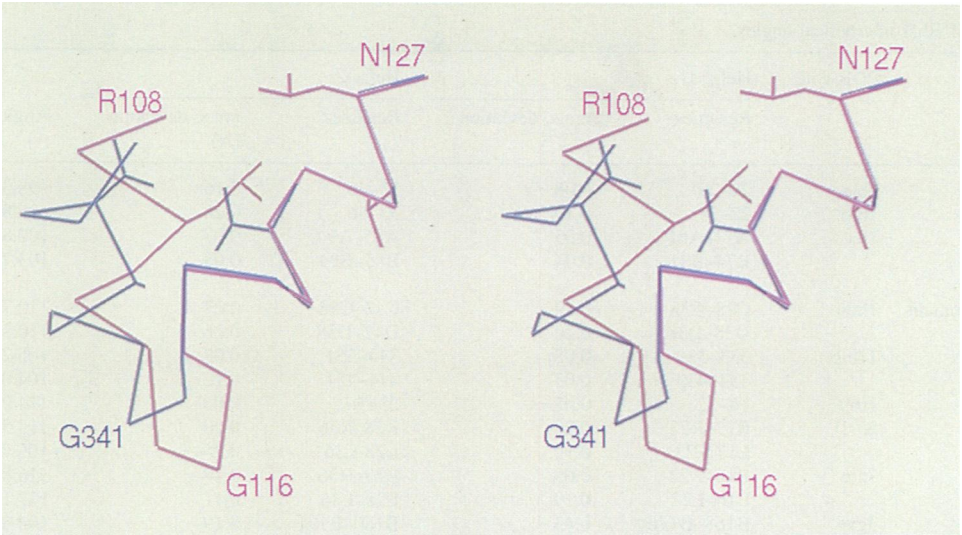

Fig. 7. Comparison of the new HhH motif and the well-characterized HtH motif. The HtH motif of the λ repressor/operator complex (blue) and the HhH motif of endonuclease III (violet) have been superimposed using the backbone of the second helix of the motifs to reveal the 20° difference in the interhelical angles. The conserved glycines at position 9 of the respective motifs are labeled to clarify the differences in the interhelical turn. The HhH motif contains a true type II β-hairpin turn between the helices, whereas the interhelical turn in the HtH motif is dependent on the left-handed conformation at Gly HtH9. Similarities and differences in the interhelical packing residues suggest that these motifs are distinct, but related.

Sγ (3.4 Å) and help stabilize the loop between the first two cluster ligands, Arg143 Nε–Cys187 O (2.9 Å), Arg147 Nη1–Ala189 O (3.0 Å) and Arg147 Nη2–Ile188 O (2.8 Å). Two other positive charges, Lys191 and Arg193, are conserved charges in four or five of the five structures (Figure 3). The remaining positive charges are less conserved. Lys206 is unconserved and Arg190 is conserved in only two of five cases. Lys208 is conserved in three of five proteins. However, the side chain of Lys208 crosses the C-terminal loop and forms a hydrogen bond to the main chain, Lys208 Nζ–Glu200 (2.9 Å), so the His residues at the corresponding position in the two MutY sequences may have a similar function.

The position of the two conserved positive charges Lys191 and Arg193, along with Arg190, in a surface exposed loop (Figure 3) implicates these side chains and this [4Fe–4S] loop region in potential DNA binding. A possible DNA binding function for a hydrogen bonding side chain at endonuclease III position 190 in the homologous proteins is conserved. To test the possible involvement of the FCL, Cys187–Cys194, in DNA binding, the mutation Lys191→Glu (K191E) was chosen as least disruptive to local structure and expressed and characterized (Table II). The K191E mutant has a >100-fold increase in K_m, showing that substrate binding is severely affected by this change. The catalytic efficiency of this mutant enzyme is relatively unaffected, since k_{cat} remains essentially unchanged. These are the results expected of a change which diminishes DNA binding without affecting active site catalysis.

The FCL motif

The combination of the data from the K191E mutant and the existence of solvent-exposed and sequence conserved positive charges in the protruding loop between the first two cluster ligands (Figure 3) suggests that this may be a new DNA binding motif, termed the FCL motif (Figure 8). In the endonuclease III structure this motif consists of residues Thr186–Cys194. Thr186 packs in the core of the protein with Thr186 Oγ1 exposed to solvent away from the loop. Ile188 and Ala189 are stabilized by hydrogen bonds to Arg143 and Arg147, as mentioned above. These hydrophobic residues serve to space the main chain of the positively charged residues away from the core of the protein and to place Arg190 and Lys191 away from the surface of the protein. The conserved Pro192 acts as a spacer between the charged residues, so that Arg193 could interact with a different portion of the DNA than Arg190 and Lys191.

The consensus sequence derived from the homology between *E.coli* endonuclease III, *E.coli* MutY and the *M.thermoformicicum* G•T mismatch repair enzyme, ICXXRKPKC, provided a basis for identifying sequences homologous with the FCL motif (Figure 9A). This single intercysteine loop was chosen because of the limited homology in the other intercysteine loops and to avoid fixing an exact pattern for the cysteine ligands. Trial searches, using the FastA algorithm (Pearson and Lipman, 1988) as implemented in the GCG package (Genetics Computer Group Inc.) and using all four Cys ligands matched more than three of the Cys ligands in only the endonuclease III homologs and yielded no more specificity than the searches using the BLAST algorithm. Of the nine highest scoring proteins, four are the homologous DNA repair enzymes, three are DNA binding proteins from fungi that contain zinc binuclear clusters, one is an ORF transcribed during sexual development of *Saccharomyces pombe* and one is a phospholipase A2. There are three important features in the retrieved set of proteins. First, *B.subtilis* endonuclease III was not retrieved. This is due to a difference in the positions of the positively charged residues from the other DNA repair enzymes. This protein has positively charged residues at FCL3 and FCL8, rather than at FCL5, FCL6 and FCL8 as in endonuclease III from *E.coli*. The *B.subtilis* enzyme has a Gln at FCL5 that is a good candidate for DNA binding interactions in the major groove. Second, in the yeast Gal4 sequence the loop between the third and fourth zinc ligands was

Table III. HtH and HhH interhelical angles

Protein	PDB code	Helix 1		Helix 2			
		Residues	r.m.s. deviation (Å)	Residues	r.m.s. deviation (Å)	Angle (°)	Resolution (Å)
E.coli BirA	1bia	22–29	0.08	33–46	0.16	98.3	2.3
E.coli BirA	1bib	22–29	0.19	33–46	0.20	96.8	2.8
E.coli FIS	1fia	A74–A81	0.05	A85–A94	0.07	102.8	2.0
		B74–B81	0.04	B85–B94	0.08	103.7	
D.melanogaster *engrailed* homeodomain	1hdd	C28–C38	0.19	C42–C58	0.27	110.7	2.8
		D28–D38	0.20	D42–D58	0.26	110.8	
λ repressor	1lmb	333–339	0.08	344–351	0.08	105.2	1.8
		433–439	0.03	444–451	0.12	104.0	
434 repressor	1r69	16–22	0.05	28–36	0.06	114.0	2.0
434 repressor	2or1	R17–R23	0.20	R28–R36	0.20	115.5	2.5
		L17-L23	0.12	L28-L36	0.25	109.7	
434 cro	3cro	R16–R22	0.09	R28–R36	0.18	116.8	2.5
		L16–L22	0.19	L28–L36	0.11	107.1	
E.coli CAP	3gap	B168–B176	0.43	B180–B191	0.19	104.9	2.5
		A168–A176	0.23	A180–A191	0.44	102.2	
E.coli Trp repressor	3wrp	68–74	0.15	79–91	0.13	118.1	1.8
Mean						107.5	
Standard deviation						6.4	
Endonuclease III		108–113	0.09	119–130	0.15	125.6	1.85

retrieved, rather than the more conserved loop between the second and third zinc ligands (Pfeifer *et al.*, 1989; Bai and Kohlhaw, 1991). These two loops have a similar distribution of positive charges. However, there is a Pro to match the Pro at FCL7 in the loop between the third and fourth zinc ligands. Third, the relative positions of the Cys residues in the two proteins without a DNA binding function does not match the pattern of Cys ligands, implicating either a zinc binuclear cluster or an endonuclease III-like [4Fe–4S] cluster. Therefore, both these motifs are likely to be dependent on the metal cluster for organizing the structure of the DNA binding loop.

A second search using a consensus without the Pro, ICXXRKXKC, was performed (Figure 9B) to match the loop between the second and third zinc ligands in Gal4. Of the nine proteins retrieved, one was a DNA repair enzyme, seven contained zinc binuclear clusters and one was a catalase from upland cotton. In this search the DNA binding loop between the second and third zinc ligands for Gal4 was found, rather than the very similar loop found in the previous search. As with the previous search, the non-DNA binding protein does not have a Cys pattern consistent with either a zinc binuclear cluster or a [4Fe–4S] cluster.

These sequence similarities suggest structural similarities between these DNA binding motifs. The three-dimensional X-ray structure of Gal4 with bound DNA has been solved (Marmorstein *et al.*, 1992). In this structure the positively charged residues from the second intercysteine loop, residues 15–20, and one from the third intercysteine loop, residues 22–26, wrap around the DNA such that Arg15 forms a hydrogen bond to a phosphate oxygen, Lys18 forms a hydrogen bond in the major groove, Lys20 is in Van der Waals contact with the sugar as it crosses the backbone and Lys23 forms a hydrogen bond to an oxygen on the next phosphate in the 3′ direction. The FCL motif in the endonuclease structure is not superimposable on the corresponding DNA binding loop in the Gal4 structure, but the relative positions of residues Arg190, Lys191 and Arg193 are such that endonuclease III could bind DNA in a similar manner to Gal4.

Table IV. [4Fe–4S] cluster and cluster ligand hydrogen bonds

Donor		Acceptor		Distance (Å)
Residue	Atom	Residue	Atom	
His182	Nε2	[4Fe–4S]	S1	3.5
Arg147	Nη2	Cys187	Sγ	3.4
Val209	N	Cys194	Sγ	3.5
Ile199	N	Cys197	Sγ	3.5
Tyr205	N	Cys203	Sγ	3.5

Endonuclease III–DNA interactions

The combination of the refined structure, mutagenesis and DNA binding motif results provides a basis for assessing the structural implications for DNA recognition. To synthesize this information into a structural model, the binding of an ideal 25mer of B-form DNA was modeled against the crystallographic structure (Figure 3). Several pieces of evidence guided the model building, including the HhH motif, the FCL motif, the homology of the FCL motif with the Gal4 structure, kinetic evidence implicating Lys120 as a key residue for lyase activity (Table II), the positively charged surface of the molecule and the sequence homology of endonuclease III with MutY and the G•T mismatch repair enzyme. The enzyme–DNA complex model is more detailed than the previously reported model (Kuo *et al.*, 1992b), which was simply an orientation of the endonuclease III positively charged face to the negatively charged DNA.

DNA could be positioned against the positively charged electrostatic face of the protein and close to the positively charged residues in both motifs, including the active site residue Lys120. Based on the homology with Gal4 and the other DNA repair enzymes, the DNA was aligned such that the side chain of Arg190 could contact the major

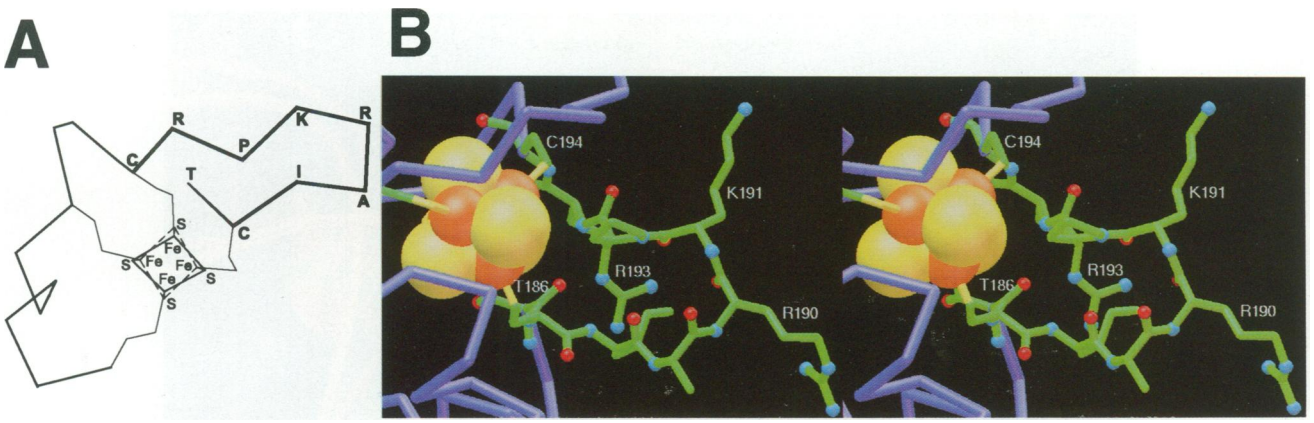

Fig. 8. FCL motif. (**A**) The schematic of the FCL motif (dark lines and one letter amino acid codes) also shows the [4Fe–4S] cluster (labeled with atomic symbols), the Cys ligands (thin lines) and the α carbon trace between the ligands (medium lines). (**B**) The stereo pair shows the positioning of positively charged residues Arg190 and Lys191 (nitrogens, blue spheres and oxygens, red spheres) away from the rest of the protein (trace in blue) ready to interact with DNA and Arg193 positioned to interact with a different piece of the DNA.

groove of the DNA and Lys191 and Arg193 could contact the backbone of the DNA. The resulting model is a feasible binding mode with no unresolvable collisions between the protein and the DNA. The backbone of the DNA fits into the interdomain groove of the protein near the active site residue Lys120 and the solvent-filled pocket. All of the steric collisions occur in flexible loops or side chains of the protein and could be accommodated by minor conformational changes in either the DNA or the protein in the actual reaction. In addition, the current model of DNA binding is consistent with the DNA footprinting results in the damaged, AP-containing, strand using (methidiumpropyl-EDTA)Fe(II) or hydroxyl radicals as nicking reagents (O'Handley *et al.*, 1995). This endonuclease III–DNA complex thus provides an experimentally testable model that can guide mutant design to probe residue function in DNA recognition and binding.

The endonuclease III–DNA complex model has similarities with the crystallographic structure of the trapped intermediate of HhaI DNA cytosine-5-methyltransferase with a 13mer oligonucleotide that has 5-fluorocytosine substituted for the target cytosine (Klimasauskas *et al.*, 1994). In this structure the DNA sits in a groove on the protein surface and the analog for the substrate base is extrahelical. Although the mechanism for extrusion of the cytosine base from the helix is unclear, it is apparently associated with binding of the DNA to the protein. The endonuclease III–DNA model also has the DNA bound in a groove on the surface of the protein such that the solvent pocket could accommodate an extrahelical base (Figure 10). Consistent with this possibility, intriguing recent evidence suggests that the thymine glycol substrate in DNA is extrahelical in solution prior to binding to endonuclease III (Kao *et al.*, 1993).

Functional interactions and specificity

The atomic structure of the base excision repair enzyme endonuclease III combined with mutational analysis suggests four specific features that relate to recognition of the DNA substrate by the protein. The electrostatic recognition surface spans the deep cleft between the domains. Within the interdomain cleft, two residues, Lys120 and Asp138, which are important for activity

A

protein	code	res. no.	sequences	score
motif numbering			1 2 3 4 5 6 7 8 9	
consensus FCL motif			I C x x R K P K C	
Eco endonuclease III	J02857	186	t C i a R K P R C	36
Mth mismatch repair enzyme	X68366	196	I C a p R K P K C	42
Eco MutY	X52391	191	I C t r s K P K C	37
Sty MutY	M86634	111	V C t r s K P K C	36
Sce Gal4	K01486	20	k C s k e K P K C	34
Sce Nim2	D23663	836	I C r l R K k K C	34
Sce Cyp1	X13793	66	I C r k R K v K C	33
Spo sexual development ORF1	D14060	75	s C e k K K P K C	35
Mmu Enhancing Factor PLA2	X74266	138	f C k g K K P K C	35
Bsu endonuclease III	U11289	186	h C k a q s P R C	
consensus zinc binuclear cluster			x C R x *K K* x K C	

B

protein	code	res. no.	sequences	score
motif numbering			1 2 3 4 5 6 7 8 9	
consensus FCL motif-no Pro			I C x x R K x K C	
Mth mismatch repair enzyme	X68366	196	I C a p R K P K C	33
Sce Gal4	K01486	13	I C r l K K l K C	31
Sce Nim2	D23663	836	I C r l R K k K C	34
Sce Cyp1	X13793	66	I C r k R K v K C	34
Sce YK44	S75352	15	n C k k R K s K C	31
Eni NirA	M68900	44	a C r r R K s K C	31
Ncr Nit4	M80368	56	a C r r R K s K C	31
Cal Suc1	S75352	15	s C s f R K v K C	31
Ghi catalase	X52135	412	v C t g R R e K C	31

Fig. 9. FCL motif in other protein sequences. For each sequence a three-letter code for the organism, the protein name, an NCBI code, the starting residue number, the sequence with residue matches in upper case letters (exact matches in bold face) and the BLOSUM62 score are given. (**A**) The proteins retrieved from the search with the consensus sequence, ICXXRKPKC, are listed in order: the three DNA repair enzymes used to derive the consensus, the other homologous DNA repair enzymes retrieved, the three zinc binuclear cluster proteins and the two proteins that probably do not have a DNA binding function. Below are listed the unretrieved DNA repair homolog endonuclease III and the consensus for the zinc binuclear cluster DNA binding loop (Pfeifer *et al.*, 1989). In this consensus sequence residues conserved as positive charges are italicized. (**B**) The proteins retrieved from the BLAST search without the Pro. These proteins include, in order: one of the DNA repair homologs used to develop the consensus, the DNA binding loop in Gal4 and six other zinc binuclear cluster homologs and one protein without a probable DNA binding function. Abbreviations not in the legend to Figure 6: *Sce*, *Saccharomyces cerevisiae*; *Spo*, *Schizosaccharomyces pombe*; *Eni*, *Emericella nidulans*; *Ncr*, *Neurospora crassa*; *Cal*, *Candida albicans*; *Ghi*, *Gossypium hirsutum*.

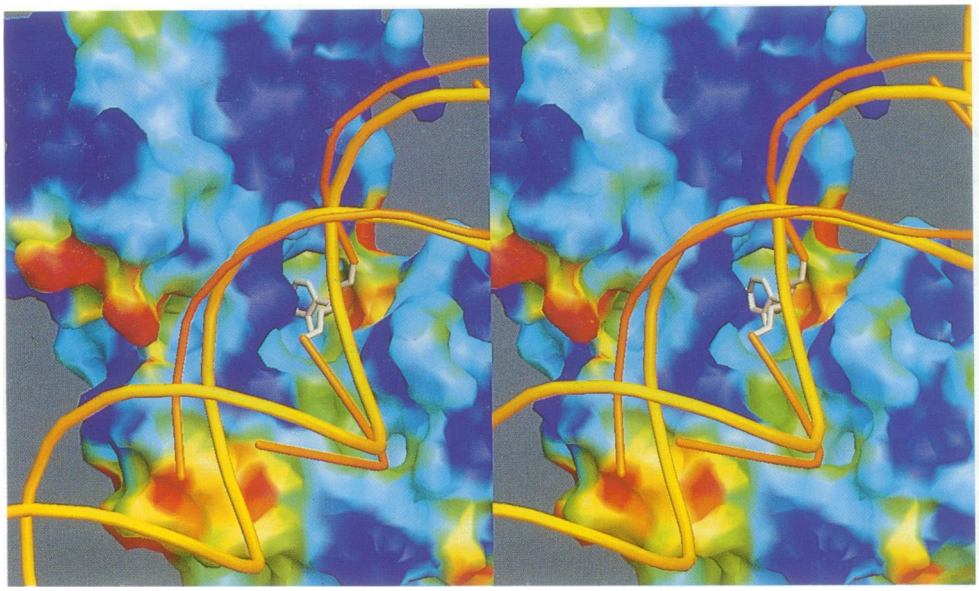

Fig. 10. Accommodation of a flipped-out base within the endonuclease III active site pocket. Alignment of the backbone of DNA from the HhaI methyltransferase–DNA complex structure (thin orange tubes) with the backbone of the DNA for our proposed endonuclease III–DNA complex model (thick yellow tubes) reveals an excellent fit for a flipped-out pyrimidine (white) within the active site pocket adjacent to Lys120 (surface colored by electrostatic potential).

bracket a deep solvent-filled pocket that could accommodate an extrahelical base. Structure-based sequence comparisons identify two novel DNA binding motifs, the HhH motif and the FCL motif. A model of an endonuclease III–DNA complex, which is consistent with the above features, has similarities with the structure of the HhaI methyltransferase–DNA crystal structure, suggesting that endonuclease III also binds an extrahelical base. Thus our results suggest a structural mechanism underlying the broad specificity of the enzyme, which may result from the variable replacement of bound water molecules within the active site pocket by the binding of extrahelical modified thymine and cytosine bases.

Materials and methods

Expression, purification and crystallization

Endonuclease III was purified from the cloned *nth* gene of *E.coli* as previously described (Asahara *et al.*, 1989). Preparations with a 410/280 nm absorbance ratio between 0.38 and 0.40, indicating that the [4Fe–4S] cluster is intact (Cunningham *et al.*, 1989), were used for crystallization. Crystals grown in 5 mM HEPES, pH 7.0, 150 mM NaCl, 5% glycerol as previously described have space group $P2_12_12_1$, with one monomer per asymmetric unit and unit cell constants a = 48.5 Å, b = 65.8 Å, c = 86.8 Å (Kuo *et al.*, 1992a).

Site-directed mutations in the *nth* gene carried on plasmid pHIT5 were produced using a unique site elimination procedure (Deng and Nickoloff, 1992). The mutant proteins were expressed in *E.coli* UC6444Δ*nth*, which carries a deletion of the endonuclease III gene, and purified as described previously (Asahara *et al.*, 1989). The kinetics of the enzymes carrying mutations in the two motifs were determined using an oligonucleotide containing a unique AP site as a substrate (Xing *et al.*, 1995). A duplex 19mer with a ^{32}P label at the 5′-end of the strand containing the AP site was treated with endonuclease III at 37°C in 50 mM HEPES, pH 7.5, 100 mM KCl buffer (O'Handley *et al.*, 1995). The reaction was stopped by addition of urea to a final concentration of 6 M and the cleaved and uncleaved strands were separated on a 20% denaturing polyacrylamide gel. The gel was dried and radioactivity in cut and uncut strands was quantitated using a Betagen betascope. The fraction of DNA cleaved was taken as the ratio of the counts in the band corresponding to the cleaved DNA to the counts in both bands, corresponding to cleaved and uncleaved DNA. Conditions were chosen such that the reaction went to ~10% completion. The cleavage reaction was analyzed using Michaelis–Menten kinetics. The structural integrity of mutants was measured by CD spectrometry using an AVIV 60DS spectrometer and the K_d of mutant and wild-type proteins was determined by gel shift assays (O'Handley *et al.*, 1995).

X-ray data collection

Data were collected at 18°C on a Siemens multiwire area detector using a Rigaku rotating anode X-ray source and a four circle goniostat. The XENGEN program package was used for data reduction (Howard *et al.*, 1987). Native data consisted of 272 851 observations of 27 787 reflections with an R_{sym} of 7.17%. These data were 99.8% complete to a resolution of 1.85 Å and 99.6% complete in the 1.97–1.85 Å shell. Data with $\langle I \rangle \geq 2\sigma$ were used for refinement. These data were 85% complete to a resolution of 1.85 Å and 56% complete in the 1.97–1.85 Å shell.

Crystallographic fitting and refinement

Refinement of the previously reported structure (Kuo *et al.*, 1992b) to a resolution of 1.85 Å was accomplished using Powell minimization and temperature factor refinement in Xplor 3.0 (Powell, 1977; Brünger *et al.*, 1987) alternated with cycles of model building using the XtalView package (McRee, 1992, 1993). Several types of maps were used for checking and rebuilding the model of endonuclease III. F_o maps with combined model and MIR phases were calculated from 10 to 2.5 Å resolution using the PHASES program package (Furey and Swaminathan, 1990). Difference maps ($2F_o - F_c$ and $F_o - F_c$) with model phases were calculated from 10 to 1.85 Å resolution using XtalView. Coloring the $2F_o - F_c$ maps by the magnitude of the electron density gradient has also been investigated (Purvis and Culberson, 1986).

During each rebuilding cycle the electron density for each water molecule was inspected and those with poor density were removed. Additional water molecules were added by hand or by a semi-automatic method (Borgstahl *et al.*, 1994). The solvent model consists of 126 water molecules with mean temperature factor 39.2 (σ = 12.1) Å^2.

Several side chains were omitted from the model for one or more cycles of refinement. These side chains were rebuilt into the observed density with slightly different conformations. Two of the side chains, Lys165 and Arg190, showed two possible conformations when the side chains were omitted entirely. When only one of these conformations was modeled, the other appeared as a positive peak in the resulting $F_o - F_c$ map. Each of these side chains has been modeled with two conformations at 50% occupancy. Residues 206–211 were omitted and then rebuilt into a different conformation during the course of several refinement cycles.

Structure analysis

The consensus sequences for the two motifs were derived from a sequence analysis of the three proteins: *E.coli* endonuclease III (Asahara *et al.*, 1989), *E.coli* MutY (Michaels *et al.*, 1990) and *M.thermoformicicum* putative G•T mismatch repair enzyme (Nölling *et al.*, 1992). Identification of sequences similar to the two consensus sequences were carried out using the BLAST (Altschul *et al.*, 1990) network service at the National Center for Biotechnology Information (NCBI). Contiguous peptide sequences in the non-redundant protein database were scored against the inquiry sequences using the BLOSUM62 homology matrix (Henikoff and Henikoff, 1992). The analysis of the motifs was limited to proteins with scores greater than or equal to a cut-off value that was determined by the score where a large majority of the proteins had a probable DNA binding function. This cut-off was always in the tail of the histogram of matches.

Structural homology of the sequences retrieved from the BLAST search on the HhH consensus sequence and the 6-helix barrel domain of endonuclease III was checked using multiple sequence alignments (Lipman *et al.*, 1989) and three-dimensional structure profile searches (Bowie *et al.*, 1991). In each protein the sequence corresponding to the 6-helix barrel domain was judged by the position of the HhH motif. The homology of the primary structure of each of these '6-helix barrel domain' sequences was tested against the endonuclease III homologs using a multiple sequence alignment program (Lipman *et al.*, 1989). The three-dimensional structure profile of endonuclease III was constructed and a score for the similarity between the tertiary structure of each of the '6-helix barrel domain' sequences and endonuclease III was obtained (Bowie *et al.*, 1991).

The comparison of the primary structure of the HhH motif with the HtH motif was accomplished by summing the likelihood of the amino acid at each position over a specified length of protein and then finding the probability that a given score is an HtH motif in a table (Dodd and Egan, 1987). Setting the invariant Gly in the β-turn of the HhH motif to position 9 facilitated the comparison of the HhH sequences with the scores for the HtH sequences.

One measure of tertiary structural differences between the HtH and HhH motifs is the differences in the helical angles. The end-points for a helical vector for each set of four adjacent α carbons over the length of the helix was calculated (Kahn, 1989a). The helical axis was calculated from these vectors using the simple least squares algorithm that has an r.m.s. deviation based on the displacement of the points from the best fit vector (Kahn, 1989b). The angle between the two helices was derived from the dot product or the dihedral angle at the point of closest approach of the vectors representing the two helical axes.

The implications for DNA recognition by the HhH and FCL motifs was examined by modeling their interactions with a 25mer of B-DNA built using the Biopolymer module of Insight II (Biosym Technologies Inc). Both the DNA and the protein were treated as rigid bodies. In order to facilitate modeling of endonuclease III interactions with DNA, a molecular surface with a 1.4 Å probe radius colored by a Coulombic electrostatic potential was calculated with a linear distance-dependent treatment of the dielectric constant (Getzoff *et al.*, 1992) and imported into Insight II. Additional guides for the modeling of DNA interaction with endonuclease III were the previously mentioned sequence conservation, homology between the FCL motif and the zinc binuclear cluster loop in Gal4 and positioning of the putative active site residue Lys120 near the AP site or damaged base. A detailed docking of endonuclease III to DNA was not attempted, due to possible conformational changes in both endonuclease III and the DNA upon complex formation.

Acknowledgements

We thank Elizabeth Getzoff, Brian Crane and Arthur Pardi for critical discussions; Cindy Fisher and Michel Sanner for help with electrostatic surfaces; Michael Hickey for the circular dichroism measurements and Peter Kahn for providing helix axis programs. Michael Pique's help with the computer graphics, particularly the gradient electron density maps, was invaluable. This work was supported by National Institute of Health grants GM46312 (R.P.C. and J.A.T.) and AG00080 (M.M.T.).

References

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Asahara,H., Wistort,P.M., Bank,J.F., Bakerian,R.H. and Cunningham,R.P. (1989) Purification and characterization of *Escherichia coli* endonuclease III from the cloned *nth* gene. *Biochemistry*, **28**, 4444–4449.

Au,K.G., Clark,S., Miller,J.H. and Modrich,P. (1989) *Escherichia coli mut*Y gene encodes an adenine glycosylase active on G-A mispairs. *Proc. Natl Acad. Sci. USA*, **86**, 8877–8881.

Bai,Y.L. and Kohlhaw,G.B. (1991) Manipulation of the 'zinc cluster' region of transcriptional activator LEU3 by site-directed mutagenesis. *Nucleic Acids Res.*, **19**, 5991–5997.

Bailly,V. and Verly,W.G. (1987) *Escherichia coli* endonuclease III is not an endonuclease but a β-elimination catalyst. *Biochem. J.*, **242**, 565–572.

Boorstein,R.J., Hilbert,T.P., Cadet,J., Cunningham,R.P. and Teebor,G.W. (1989) UV-induced pyrimidine hydrates in DNA are repaired by bacterial and mammalian DNA glycosylase activities. *Biochemistry*, **28**, 6164–6170.

Borgstahl,G.E.O., Rogers,P.H. and Arnone,A. (1994) The 1.8 Å structure of carbonmonoxy-$\beta_4$hemoglobin: analysis of a homotetramer with the R quaternary structure of liganded α2-β2 hemoglobin, *J. Mol. Biol.*, **236**, 817–830.

Bowie,J.U., Lüthy,R. and Eisenberg,D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**, 164–170.

Breimer,L. and Lindahl,T. (1980) A DNA glycosylase from *Escherichia coli* that releases free urea from a polydeoxyribonucleotide containing fragments of base residues. *Nucleic Acids Res.*, **8**, 6199–6211.

Breimer,L.H. and Lindahl,T. (1984) DNA glycosylase activities for thymine residues damaged by ring saturation, fragmentation, or ring contraction are functions of endonuclease III in *Escherichia coli*. *J. Biol. Chem.*, **259**, 5543–5548.

Brennan,R.G. and Matthews,B.W. (1989) The helix–turn–helix DNA binding motif. *J. Biol. Chem.*, **264**, 1903–1906.

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

Carson,M. (1991) RIBBONS 2.0. *J. Appl. Crystallogr.*, **24**, 958–961.

Cunningham,R.P. and Weiss,B. (1985) Endonuclease III (*nth*) mutants of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **82**, 474–478.

Cunningham,R.P. *et al.* (1989) Endonuclease III is an iron–sulfur protein. *Biochemistry*, **28**, 4450–4455.

Deng,W.P. and Nickoloff,J.A. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem.*, **200**, 81–88.

Desiraju,V., Shanabruch,W. and Lu,A. (1993) Nucleotide sequence of the *Salmonella typhimurium mutB* gene, the homolog of *Escherichia coli mutY*. *J. Bacteriol.*, **175**, 541–543.

Dizdaroglu,M., Laval,J. and Boiteux,S. (1993) Substrate specificity of the *Escherichia coli* endonuclease III: excision of thymine- and cytosine-derived lesions in DNA produced by radiation-generated free radicals. *Biochemistry*, **32**, 12105–12111.

Dodd,I.B. and Egan,J.B. (1987) Systematic method for the detection of potential lambda cro-like DNA binding regions in proteins. *J. Mol. Biol.*, **194**, 557–564.

Doetsch,P.W., Henner,W.D., Cunningham,R.P., Toney,J.H. and Helland,D.E. (1987) A highly conserved endonuclease activity present in *Escherichia coli*, bovine and human cells recognizes oxidative DNA damage at sites of pyimidines. *Mol. Cell. Biol.*, **7**, 26–32.

Frampton,J., Gibson,T.J., Ness,S.A., Döderlein,G. and Graf,T. (1991) Proposed structure for the DNA-binding domain of the Myb oncoprotein based on model building and mutational analysis. *Protein Engng*, **4**, 891–901.

Fu,W., O'Handley,S., Cunningham,R.P. and Johnson,M.K. (1992) The role of the iron–sulfur cluster in *Escherichia coli* endonuclease III: a resonance Raman study. *J. Biol. Chem.*, **267**, 16135–16137.

Furey,W. and Swaminathan,S. (1990) PHASES: a program package for the processing and analysis of diffraction data from macromolecules. In *ACA Meeting Abstracts*, Vol. 73. American Crystallographic Association, Washington, DC.

Ganguly,T., Weems,K.M. and Duker,N.J. (1990) Ultraviolet-induced thymine hydrates in DNA are excised by bacterial and human DNA glycosylase activities. *Biochemistry*, **29**, 7222–7228.

Getzoff,E.D., Cabelli,D.E., Fisher,C.L., Parge,H.E., Viezzoli,M.S., Banci,L. and Hallewell,R.A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidence, *Nature*, **358**, 347

Gosset,J., Lee,K., Cunningham,R.P. and Doetsch,P.W. (1988) Yeast redoxyendonuclease, a DNA repair enzyme similar to *Escherichia coli* endonuclease III. *Biochemistry*, **27**, 2629–2634.

Hatahet,Z., Kow,Y.W., Purmal,A.A., Cunningham,R.P. and Wallace,S.S. (1994) New substrates for old enzymes. 5-Hydroxy-2′-deoxycytidine

and 5-hydroxy-2′-deoxyuridine are substrates for *Escherichia coli* endonuclease III and formamidopyrimidine DNA *N*-glycosylase, while 5-hydroxy-2′-deoxyuridine is a substrate for uracil DNA *N*-glycosylase. *J. Biol. Chem.*, **269**, 18814–18820.

Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.

Higgins,S.A., Frenkel,K., Cummings,A. and Teebor,G.W. (1987) Definitive characterization of human thymine glycol *N*-glycosylase activity. *Biochemistry*, **26**, 1683–1688.

Howard,A.J., Gilliland,G.L., Finzel,B.C., Poulos,T.L., Ohlendorf,D.H. and Salemme,F.R. (1987) The use of an imaging proportional counter in macromolecular crystallography. *J. Appl. Crystallogr.*, **20**, 383–387.

Jorgensen,T.J., Kow,Y.-W., Wallace,S.S. and Henner,W.D. (1987) Mechanism of action of *Micrococcus luteus* γ-endonuclease. *Biochemistry*, **26**, 6436–6443.

Kahn,P.C. (1989a) Defining the axis of a helix. *Comput. Chem.*, **13**, 185–189.

Kahn,P.C. (1989b) Simple methods for computing the least squares line in three dimensions. *Comput. Chem.*, **13**, 191–195.

Kao,J.Y., Goljer,I., Phan,T. and Bolton,P.H. (1993) Characterization of the effects of a thymine glycol residue on the structure, dynamics and stability of duplex DNA by NMR. *J. Biol. Chem.*, **268**, 17787–17793.

Katcher,H.L. and Wallace,S.S. (1983) Characterization of the *Escherichia coli* X-ray endonuclease, endonuclease III. *Biochem*istry, **22**, 4071–4081.

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

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

Kow,Y.W. and Wallace,S.S. (1987) Mechanism of action of *Escherichia coli* endonuclease III. *Biochemistry*, **26**, 8200–8206.

Kuo,C.-F., McRee,D.E., Cunningham,R.P. and Tainer,J.A. (1992a) Crystallization and crystallographic characterization of the iron–sulfur-containing enzyme endonuclease III from *Escherichia coli*. *J. Mol. Biol.*, **227**, 347–351.

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

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Lipman,D.J., Altschul,S.F. and Kececioglu,J.D. (1989) A tool for multiple sequence aligment. *Proc. Natl Acad. Sci. USA.*, **86**, 4412–4415.

Marmorstein,R., Carey,M., Ptashne,M. and Harrison,S.C. (1992) DNA recognition by Gal4: structure of a protein–DNA complex. *Nature (Lond.)*, **356**, 408–414.

Mazumder,A., Gerlt,J.A., Absalon,M.J., Stubbe,J., Cunningham,R.P., Withka,J. and Bolton,P.H. (1991) Stereochemical studies of the β-elimination reactions at aldehydic abasic sites in DNA: endonuclease III from *Escherichia coli*, sodium hydroxide and Lys-Trp-Lys. *Biochemistry*, **30**, 1119–1126.

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

McRee,D.E. (1993) *Practical Protein Crystallography.* Academic Press, San Diego, CA.

Michaels,M.L., Pham,L., Nghiem,Y., Cruz,C. and Miller,J.H. (1990) MutY, an adenine glycosylase active on G-A mispairs, has homology to endonuclease III. *Nucleic Acids Res.*, **18**, 3841–3845.

Michaels,M.L., Tchou,J., Grollman,A.P. and Miller,J.H. (1992) A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry*, **31**, 10964–10968.

Morris,A.L., MacArthur,M.W., Hutchinson,E.G. and Thornton,J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins Structure Function Genet.*, **12**, 345–364.

Nölling,J., van Eeden,F.J.M., Eggen,R.I.L. and de Vos,W.M. (1992) Modular organization of related Archael plasmids encoding different restriction–modification systems in *Methanobacterium thermoformicicum*. *Nucleic Acids Res.*, **20**, 6501–6507.

O'Handley,S., Scholes,C.P. and Cunningham,R.P. (1995) Endonuclease III interactions with DNA substrates. 1. Binding and footprinting studies with oligonucleotides containing a reduced apyrimidinic site. *Biochemistry*, **34**, 2528–2536.

Ogata,K., Hojo,H., Aimoto,S., Nakai,T., Nakamura,H., Sarai,A., Ishii,S. and Nishimura,Y. (1992) Solution structure of a DNA-binding unit of Myb: a helix–turn–helix-related motif with conserved tryptophans forming a hydrophobic core. *Proc. Natl Acad. Sci. USA.*, **89**, 6428–6432.

Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y. and Matthews, B.W. (1982) The molecular basis of DNA–protein recognition inferred from the structure of cro repressor. *Nature (Lond.)*, **298**, 718–723.

Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA.*, **85**, 2444–2448.

Pfeifer,K., Kim,K.S., Kogan,S. and Guarente,L. (1989) Functional dissection and sequence of yeast HAP1 activator. *Cell*, **56**, 291–301.

Powell,M.J.D. (1977) Restart procedures for the conjugate gradient method. *Math. Programming*, **12**, 241–254.

Purvis,G.D. and Culberson,C. (1986) On the graphical display of molecular electrostatic force-fields and gradients of the electron density. *J. Mol. Graphics*, **4**, 88–92.

Radman,M. (1976) An endonuclease from *Escherichia coli* that introduces single polynucleotide chain scissions in ultraviolet-irradiated DNA. *J. Biol. Chem.*, **251**, 1438–1445.

Richardson,J.S. and Richardson,D.C. (1988) Amino acid preferences for specific locations at the end of α-helices. *Science*, **240**, 1648–1652.

Tsai-Wu,J.-J., Liu,H.-F. and Lu,A.-L. (1992) *Escherichia coli* MutY protein has both N-glycosylase and apurinic/apyrimidinic endonuclease activities on A•C and A•G mispairs. *Proc. Natl Acad. Sci. USA*, **89**, 8779–8783.

Xing,D., Dorr,R., Cunningham,R.P. and Scholes,C.P. (1995) Endonuclease III interactions with DNA substrates. 2. The DNA repair enzyme endonuclease III binds differently to intact DNA and to apyrimidinic/apurinic DNA substrates as shown by tryptophan fluorescence quenching. *Biochemistry*, **34**, 2537–2544.

Received on February 10, 1995; revised on May 1, 1995