a "paradigm shift," but a gradual convergence of technique and need. The many approximations and compromises of these 2-D simulations will gradually be eliminated and what promises to emerge is a rich and detailed picture of one of the most dramatic natural events.

REFERENCES AND NOTES

1. A. Burrows, *Annu. Rev. Nucl. Part. Sci.* **40**, 181 (1990).
2. _____ and J. M. Lattimer, *Astrophys. J.* **299**, L19 (1985).
3. T. A. Weaver, S. E. Woosley, G. E. Fuller, in *Numerical Astrophysics*, J. M. Centrella, J. M. LeBlanc, R. L. Bowers, Eds. (Jones and Bartlett, Boston, 1985), p. 374.
4. W. D. Arnett, B. A. Fryxell, E. Müller, *Astrophys. J.* **341**, L63 (1989).
5. P. Woodward and P. Colella, *J. Comp. Phys.* **54**, 115 (1984).
6. S. E. Woosley and T. A. Weaver, *Annu. Rev. Astron. Astrophys.* **24**, 205 (1986).
7. A. Burrows, *Phys. Today* **40**, 28 (1987).
8. T. J. Mazurek, *Nature* **252**, 287 (1974); K. Sato, *Prog. Theor. Phys.* **54**, 1325 (1975); W. D. Arnett, *Astrophys. J.* **218**, 815 (1977); W. Hillebrandt and E. Muller, *Astron. Astrophys.* **103**, 147 (1981).
9. T. J. Mazurek, *Astrophys. J.* **259**, L43 (1982); S. W. Bruenn, *Astrophys. J. Suppl.* **58**, 771 (1985).
10. A. Burrows, *Astrophys. J.* **318**, L57 (1987).
11. _____ and T. J. Mazurek, *ibid.* **259**, 330 (1982).
12. J. R. Wilson, in *Numerical Astrophysics*, J. Centrella, J. LeBlanc, R. L. Bowers, Eds. (Jones and Bartlett, Boston, 1985), p. 422.
13. R. Mayle and J. R. Wilson, *Phys. Rep.* **163**, 63 (1988).
14. S. W. Bruenn, in preparation.
15. K. Nomoto, T. Shigeyama, M. Hashimoto, in *Proceedings of the IAU Colloquium 108 on Atmospheric Diagnostics of Stellar Evolution: Chemical Peculiarity, Mass Loss and Explosion*, K. Nomoto, Ed. (Springer-Verlag, Berlin, 1987), p. 319.
16. S. E. Woosley and E. Baron, *Astrophys. J.* **391**, 228 (1992).
17. J. P. Cox and R. T. Guili, *Principles of Stellar Structure* (Bordon and Breach), New York, 1968).
18. H. A. Bethe, *Rev. Mod. Phys.* **62**, 801 (1990).
19. M. Herant, W. Benz, S. A. Colgate, *Astrophys. J.* **395**, 642 (1992).
20. A. Burrows and J. M. Lattimer, *Phys. Rep.* **163**, 51 (1988); H. A. Bethe, G. E. Brown, J. Cooperstein, *Astrophys. J.* **322**, 201 (1987).
21. S. E. Woosley and R. Hoffman, *Astrophys. J.* **395**, 202 (1992); B. S. Meyer *et al.*, *ibid.*, in press.
22. A. Burrows and J. M. Lattimer, *ibid.* **307**, 178 (1986).
23. R. Duncan and C. Thompson, *ibid.*, in press; C. Thompson and R. Duncan, *ibid.*, in press.

24. We thank NSF and NASA for support under grants AST89-14346 and NAGW-2145, respectively, and acknowledge the hospitality of the Aspen Center for Physics, where some of this manuscript was generated. Conversations with D. Arnett, T. Weaver, S. Woosley, and W. Benz materially improved the paper. These calculations were performed at the National Center for Supercomputer Applications (NCSA) at Urbana, IL, and the graphics were created on a Silicon Graphics 4D/35 workstation.

24 July 1992; accepted 15 September 1992

Atomic Structure of the DNA Repair [4Fe-4S] Enzyme Endonuclease III

Che-Fu Kuo, Duncan E. McRee, Cindy L. Fisher, Suzanne F. O'Handley, Richard P. Cunningham, John A. Tainer*

The crystal structure of the DNA repair enzyme endonuclease III, which recognizes and cleaves DNA at damaged bases, has been solved to 2.0 angstrom resolution with an *R* factor of 0.185. This iron-sulfur [4Fe-4S] enzyme is elongated and bilobal with a deep cleft separating two similarly sized domains: a novel, sequence-continuous, six-helix domain (residues 22 to 132) and a Greek-key, four-helix domain formed by the amino-terminal and three carboxyl-terminal helices (residues 1 to 21 and 133 to 211) together with the [4Fe-4S] cluster. The cluster is bound entirely within the carboxyl-terminal loop with a ligation pattern (Cys-X_6-Cys-X_2-Cys-X_5-Cys) distinct from all other known [4Fe-4S] proteins. Sequence conservation and the positive electrostatic potential of conserved regions identify a surface suitable for binding duplex B-DNA across the long axis of the enzyme, matching a 46 angstrom length of protected DNA. The primary role of the [4Fe-4S] cluster appears to involve positioning conserved basic residues for interaction with the DNA phosphate backbone. The crystallographically identified inhibitor binding region, which recognizes the damaged base thymine glycol, is a seven-residue β-hairpin (residues 113 to 119). Location and side chain orientation at the base of the inhibitor binding site implicate Glu^{112} in the *N*-glycosylase mechanism and Lys^{120} in the β-elimination mechanism. Overall, the structure reveals an unusual fold and a new biological function for [4Fe-4S] clusters and provides a structural basis for studying recognition of damaged DNA and the *N*-glycosylase and apurinic/apyrimidinic–lyase mechanisms.

Damage inflicted on DNA by active oxygen species has been implicated in many degenerative processes, including cancer and aging (*1*). All cells have an arsenal of highly efficient DNA repair mechanisms to cope with the multitude of lesions formed. The base excision repair mechanism is initiated by the sequential action of ubiquitous DNA *N*-glycosylases and AP endonucleases, suggesting that all organisms make use of this mode of DNA repair (*2*). Endonuclease III of *Escherichia coli* acts both as a DNA *N*-glycosylase, removing oxidized pyrimidines from DNA, and an apurinic/apyrimidinic (AP) lyase, introducing a single-strand nick at the AP site from which the damaged base was removed (*3*). Investigation of the *E. coli* enzyme has the advantage that results from the extensively studied bacterial systems for protection and repair responses to oxidative stress are often applicable to more complex and less defined systems, including human cells (*4*). Bacteria (*5*, 6), yeast (*7*), and bovine and human cells (*8*) share a conserved class of enzymes with the same substrate specificity for damaged DNA cleavage, similar molecular weights, and no requirement for divalent cations. Because endonuclease III is the prototype for this conserved class of enzymes, studies of the protein have broad implications in understanding DNA repair and, potentially, carcinogenesis and aging in higher organisms. Endonuclease V, whose structure was recently reported (*9*), is a different enzyme of distinct structure but related function that is found in phages and is without known homologs in higher organisms.

Cloning and overexpression of the *nth* gene has allowed large-scale purification and biochemical characterization of endonuclease III (*5*, *10*). Surprisingly, the protein, unlike any of the other well-characterized DNA repair enzymes, contains an iron-sulfur [4Fe-4S] cluster. Sequence similarity suggests that the recently sequenced *mutY* gene codes for an adenine glycosylase that also contains an Fe-S cluster (*11*) and that endonuclease III is a representative of a new class of Fe-S proteins. The three-dimensional (3-D) structure of endonuclease III provides detailed information to guide biochemical investigations into the enzyme's recognition of damaged DNA, the biological function of the [4Fe-4S] cluster, and the mechanisms for the *N*-glycosylase and AP lyase activities. The combined structural and biochemical results show that endonuclease III contains novel structural motifs for the recognition and repair of damaged DNA previously unidentified in a

C.-F. Kuo, D. E. McRee, C. L. Fisher, and J. A. Tainer are at the Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037. S. F. O'Handley and R. P. Cunningham are at the Center for Biochemistry and Biophysics, Department of Biological Sciences, State University of New York at Albany, Albany, NY 12222.

*To whom correspondence should be addressed.

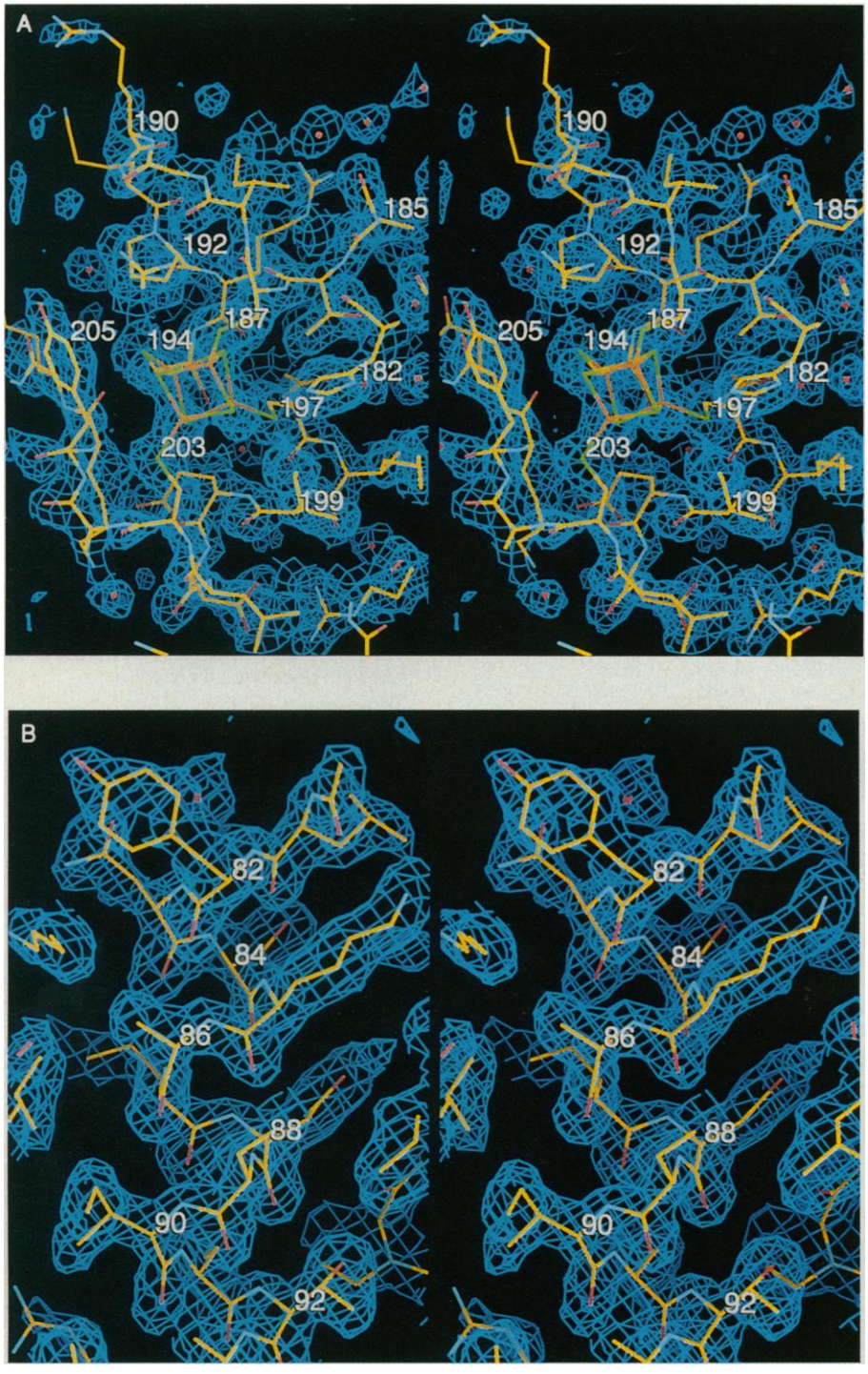

Fig. 1. Stereo pairs showing the refined model and electron density map. The map (blue contours at 1σ) was calculated at 2 Å resolution with ($|2F_o| - |F_c|$) amplitudes and model phases. The model is colored by atomic element (oxygens, red; nitrogens, blue; carbons, yellow; sulfurs, green; and irons, orange). Model and maps are displayed with the new program XtalView running on a Sun workstation (*45*). (**A**) The [4Fe-4S] cluster and its environment. The 26-residue carboxyl-terminal loop (residues 186 to 211) forms a right-handed spiral around the [4Fe-4S] cluster and contributes all four Cys ligands (residues 187, 194, 197, and 203), which are close together in sequence. This [4Fe-4S] cluster geometry was clearly defined even in the 2.7 Å–resolution MIR electron density map calculated with solvent-flattened combined phases. The [4Fe-4S] cluster within its carboxyl-terminal loop together with helices αA and αH-J form the [4Fe-4S] cluster domain. (**B**) A portion of the long helix αE (residues 82 to 99), which forms part of the 6-helix barrel domain. In the refined maps, the electron density shows the well-characterized side chain and carbonyl oxygen positions. Helix αE, which is the longest helix in the structure, appears to play a role in DNA recognition.

structural taxonomy of DNA-binding domains (*12*).

Structure determination. Crystals of endonuclease III were obtained by dialysis against low ionic strength buffer and improved by microseeding and macroseeding (*13*). The crystals belong to the space group $P2_12_12_1$ with $a = 48.5$ Å, $b = 65.8$ Å, and $c = 86.8$ Å and contain one 23-kD monomer per asymmetric unit and 55% solvent, with $V_m = 3.01$ Å^3 per dalton. Complete diffraction data to 1.98 Å resolution were collected from three crystals with 1.54 Å CuKα radiation on a Siemens area detector and processed with the Xengen data-processing software (version 1.3) (*14*). The structure was solved with the multiple isomorphous replacement (MIR) method by using four different derivatives and the native anomalous signal of the Fe-S cluster for phase determination (Table 1). A 2.7 Å–resolution MIR electron density map calculated with solvent-flattened (*15*) combined phases was used for initial tracing of the polypeptide chain with the program GRINCH (*16*). Main chain refitting and side chain assignment were accomplished with the graphics program FRODO (*17*). The model was refitted and then positionally refined with X-PLOR (version 2.1) (*18*) from an initial *R* factor of 0.48 to one of 0.286 to 2.5 Å resolution. One cycle of simulated annealing refinement with the slow-cooling protocol and a starting temperature of 1000 K reduced the *R* factor to 0.266. Positional and temperature factor refinements were repeated with manual refitting between refinement cycles. The refinement resolution was increased to 1.98 Å after the *R* factor reached 0.23. Each portion of the structure was checked for errors with fragment difference maps omitting the part in question to reduce phase bias. Well-ordered water molecules (118 total) were identified with a ($|F_o| - |F_c|$) difference map. The final 2 Å–resolution model (Fig. 1, A and B) containing 1778 atoms had an *R* factor of 0.185, with a bond length root-mean-square deviation of 0.016 Å, a bond angle deviation of 2.9°, an average temperature factor of 23.9 Å^2, and excellent main chain geometry (Fig. 2).

In order to locate the active site of endonuclease III, diffraction data were collected from a crystal soaked in a solution containing thymine glycol, an inhibitor for the *N*-glycosylase activity of endonuclease III and a major DNA lesion produced by oxidative stress. A difference Fourier map was used to identify the binding site (Table 1). Refinement of these data from the complex was then used to examine possible structural changes associated with binding.

Overall fold and dimensions. Endonuclease III is an elongated molecule (60 by 35 by 30 Å) with primarily α-helical sec-

Table 1. Crystallographic data statistics. Crystals were derivatized by soaking at 15°C in mother liquor (0.1 M Hepes buffer, pH 7.0, 5% glycerol) containing specified heavy atom reagents. Each heavy atom derivative data set was collected from a single crystal by using a 0.4° oscillation step size to facilitate data collection and to circumvent the rapid decay of derivatized crystals. The positions of heavy atoms were located by isomorphous difference Patterson maps. The single isomorphous phases obtained from the single-site platinum derivative were used to calculate the crossed-phase difference Fourier maps for location of additional heavy atom binding sites and orientation of different heavy atom site coordinates on the same origin. The heavy atom coordinates were refined with the program HEAVY (*43*). The [4Fe-4S] cluster model of *Peptococcus aerogenes* ferredoxin obtained from the Brookhaven Protein Data Bank (*44*) was fit into the [4Fe-4S] cluster density clearly recognizable in the original MIR map, with the four Fe atoms positioned into four protrusions of the cluster density. The coordinates of the Fe atoms were then used for including the anomalous signals in the phase determinations. The combined phases were improved by the solvent-flattening protocol implemented in the program PHASES with a 40% solvent cutoff (W. Furey, Biocrystallography Laboratory, Veterans Affairs Medical Center, Pittsburgh, Pennsylvania). In order to obtain data for the inhibitor-bound endonuclease III crystal, the crystal was soaked at 15°C in mother liquor containing 2 mM thymine glycol for 12 hours. Data were collected with 0.2° oscillation steps and processed as described in the text.

Parameters	Native	$K_2Pt(NO_2)_4$	$OsCl_3$	Thimerosal	$KAu(CN)_2$	Thymine glycol inhibitor
Soaking concentration (mM)		10	10	10	10	2
Soaking time (hours)		15	16	16	48	12
Data resolution (Å)	1.98	2.7	2.8	2.9	2.7	2.4
Data completeness (percent)	100	85	93	97	87	94
I/σ at resolution limit	3.5	3.3	3.0	5.8	3.6	4.8
Observations (no.)	141,196	22,733	19,479	22,785	16,365	33,182
Unique reflections	19,833	7,573	6,703	6,461	7,169	11,405
R_{sym}* (percent)	6.4	6.6	7.4	4.8	5.7	5.6
R_{Cullis}†	0.51	0.58	0.60	0.60	0.64	
R_{merge}‡ (percent)		13.1	18.4	17.7	18.7	9.7
Sites (no.)	4	1	3	1	3	1
Phasing power§	1.51	1.91	1.96	1.53	1.38	

*R_{sym} is the unweighted R factor on F between symmetry mates. †For native, R was calculated for the largest (upper 25 percent) of the Bijvoet differences; $R_{Cullis} = \Sigma[|F_h| - (|F_{ph}| - |F_p|)]/\Sigma|F_h|$. ‡$R_{merge} = \Sigma(|F_{der}| - |F_{nat}|)/\Sigma|F_{nat}|$. §Phasing power = $(F_h)/$(closure error). The solvent-flattened combined phases have a figure of merit of 0.67 to 2.7 Å resolution.

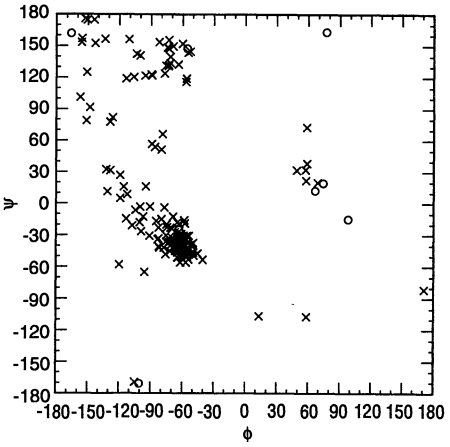

Fig. 2. Ramachandran plot showing the main chain dihedral angles for the refined endonuclease III model. The circles denote glycines, and crosses denote nonglycine residues. The two nonglycine residues outside the stereochemically favorable region of the plot are residues Lys^{206} and Val^{209}. Beyond residue 205, the conformations of the carboxyl-terminal residues were not well defined in the MIR map; hence residue conformation beyond 205 is uncertain. The polypeptide chain direction, however, appears to be correct.

ondary structure (Fig. 3). Ten α helices, which contain a total of 121 residues and range in length from 7 to 18 residues, are named in sequence order αA through αJ. The loop and turn connections range from 4 to 12 residues in length and are named for the helices they connect. The 211 residues are partitioned between two similarly sized globular domains (111 and 100 residues, respectively) separated by a deep cleft. Residues 22 to 132 fold into one domain, descriptively termed the 6-helix barrel domain for its distinctive structure. The 21 amino-terminal residues (residues 1 to 21) and 79 carboxyl-terminal residues (residues 133 to 211) form the [4Fe-4S] cluster domain. Both domains have extensive interior hydrophobic cores and polar charged surfaces. The absence of extensive crystal packing interactions and of any hydrophobic patches on the molecular surface suggests that the enzyme is unlikely to form dimers or larger multimers.

6-Helix barrel domain. This novel, sequence-continuous domain is formed by six antiparallel helices αB through αG, all with nearest-neighbor connections. The packing of αB through αE resembles that of a standard four-helix bundle (*19, 20*). The helix αD contains two seven-residue parts (αD1 and αD2) kinked by a 60° angle at Gly^{68}. Helices αF and αG form an additional layer, placing αB roughly in the center of the other five helices. As seen in four-helix bundles, αC, αE, and αG are packed against two antiparallel neighbors. The most extensive interactions are formed by αB, which packs against αC, αE, and αG and has polar residues only at its ends. The most loosely packed helix is αF, which only forms tight interactions with αG. Although essentially antiparallel, most of the helix pairs cross at greater angles than those seen in four-helix bundles: only three helix pairs (αB-αE, αC-αD2, and αF-αG) have shallow helix packing angles. Helices αB through αF all have standard N-cap residues (*21*) that form side chain to main chain hydrogen bonds to initiate the helices. Terminal C-caps occur on most of the helices, with water molecules forming C-caps on αC and αE. Helix αG is exceptional and has no N- or C-cap interactions.

Viewed down the long axis, the six-helix bundle in endonuclease III has an unusual, overall right-handed twist, unlike the left-handed twist of most helix bundles (*22*). Because of this right-handed twist and the large angle between most helix pairs, this domain resembles a helical version of a β-barrel, which, to our knowledge, has not been described previously for soluble protein domains (*19, 20*). Except for the seven-helix bundles in bacteriorhodopsin (*23*) and the pore-forming δ-endotoxin (*24*), other known antiparallel clusters with five or more helices have at least one distant-neighbor connection and all have an overall left-handed twist (*19*).

[4Fe-4S] cluster domain. The amino-terminal helix αA packs with the three carboxyl-terminal helices αH, αI, and αJ at roughly 90° angles to form the four-helix domain containing the [4Fe-4S] cluster. The three carboxyl-terminal helices are joined by short loops. Asp and Asn side chains form N-caps on αA, αH, and αI. The N-cap for αJ is formed by two water molecules that also participate in a network of hydrogen bonds to side chains in αG, the G-H loop, αJ, and the I-J loop. C-caps occur on αH, αI, and αJ, but not on αA. These four helices interact with three sets of helix lap joints (*25, 26*), a distinctive type of slightly offset and overlapped right-angle corners that occur in the E-F hands of reversible calcium-binding (*26*) and DNA-binding (*12, 27, 28*) proteins. One difference between the helix-turn-helix pattern of endonuclease III and that of other DNA-binding proteins is that the longer loop between αH and αI of endonuclease III allows the lapping interaction to be at the amino-terminal ends of both αH and αI. Lap joints of the amino-terminal end of αH

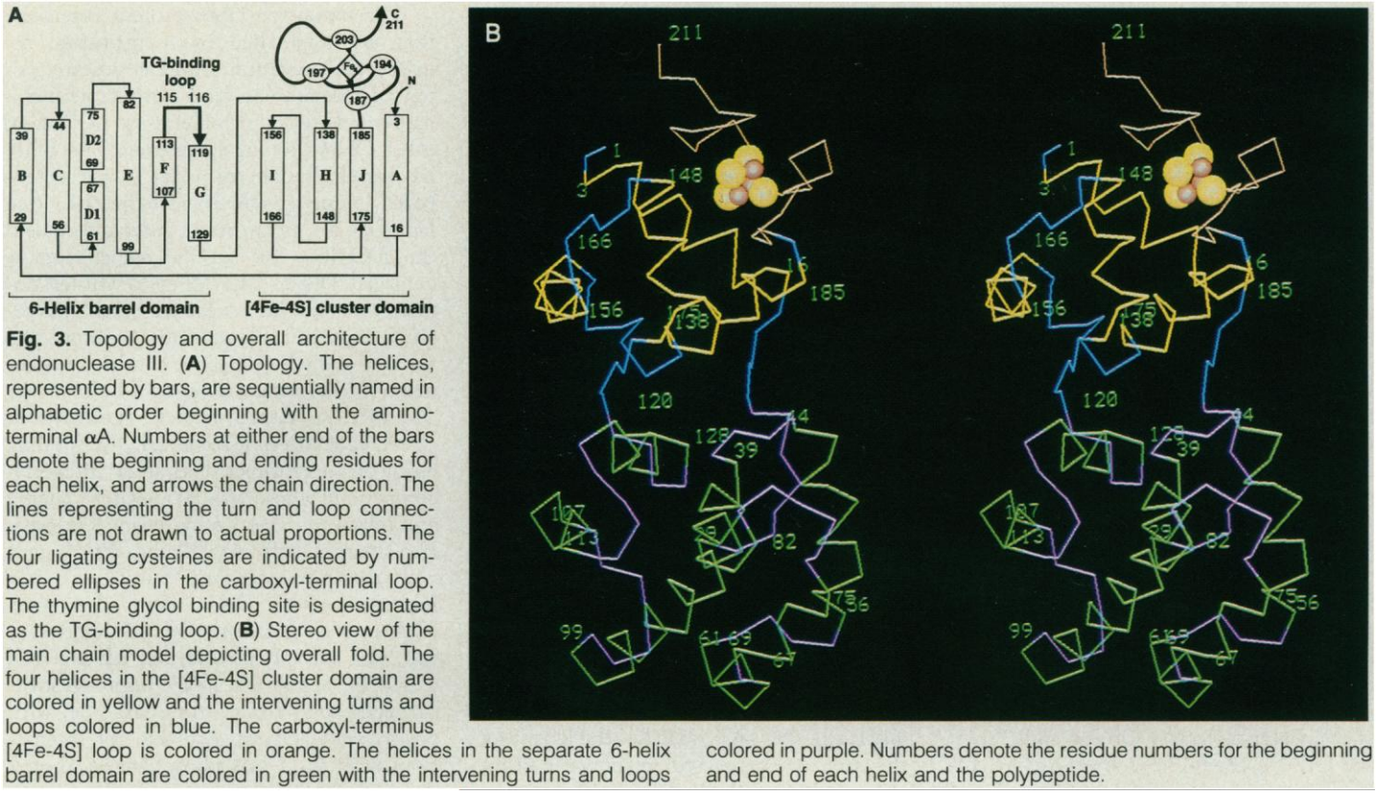

Fig. 3. Topology and overall architecture of endonuclease III. (**A**) Topology. The helices, represented by bars, are sequentially named in alphabetic order beginning with the amino-terminal αA. Numbers at either end of the bars denote the beginning and ending residues for each helix, and arrows the chain direction. The lines representing the turn and loop connections are not drawn to actual proportions. The four ligating cysteines are indicated by numbered ellipses in the carboxyl-terminal loop. The thymine glycol binding site is designated as the TG-binding loop. (**B**) Stereo view of the main chain model depicting overall fold. The four helices in the [4Fe-4S] cluster domain are colored in yellow and the intervening turns and loops colored in blue. The carboxyl-terminus [4Fe-4S] loop is colored in orange. The helices in the separate 6-helix barrel domain are colored in green with the intervening turns and loops colored in purple. Numbers denote the residue numbers for the beginning and end of each helix and the polypeptide.

with the carboxyl-terminal end of αJ and amino-terminal end of αJ with the carboxyl-terminal end of αA occur between noncontiguous helices. Based on its topology (Fig. 3A), which has a +2 connection between both helix pairs αA-αH and αI-αJ, this domain is a Greek-key helix bundle (*19*). As a result of this connectivity, αJ at the carboxyl terminus packs between αA at the amino terminus and αH, which starts the domain.

The [4Fe-4S] cluster distinguishes endonuclease III as the prototype for Fe-S cluster–containing DNA repair enzymes. The [4Fe-4S] cluster is ligated by cysteines 187, 194, 197, and 203 in an unusually short (17 residues), sequence-continuous region that folds into a right-handed spiral around the [4Fe-4S] cluster and packs against the four-helix domain core (see Fig. 1A). The ligand spacing (Cys-X_6-Cys-X_2-Cys-X_5-Cys) is unlike that found in any other [4Fe-4S] proteins of known structure: ferredoxin (distal Cys + Cys-X_2-Cys-X_2-Cys); high-potential iron protein (Cys-X_2-Cys-X_{16}-Cys-X_{13}-Cys); and trimethylamine dehydrogenase (Cys-X_2-Cys-X_2-Cys-X_{12}-Cys) (*29*). Sequence conservation of the endonuclease III pattern of Cys ligands in the DNA repair protein MutY (*11*) suggests that these proteins belong to a new class of [4Fe-4S] enzymes. In the endonuclease III model, the Fe–Fe (2.72 to 2.81 Å) and Fe–S (2.29 to 2.31 Å) bond lengths vary by less than 0.1 Å; thus all four irons are equivalent within the expected error at this resolution (~0.2 Å). This result is consistent with Mössbauer spectroscopy, which indicated that the four irons in the cluster are equivalent and reside in similar environments (*10*). In both packing and hydrogen-bonding interactions, the buried [4Fe-4S] cluster interacts primarily with the 26-residue carboxyl-terminal loop through two hydrogen bonds to Cys sulfur ligands ($Cys^{194}N\cdots Cys^{194}SG$ and $Tyr^{205}N\cdots Cys^{203}$-SG) and packing of the aromatic ring of Tyr^{205} with an inorganic sulfur (S4) of the cluster (see Fig. 1A). Interactions of the cluster outside the carboxyl-terminal loop include: the hydrogen bond with His^{182} ($S1\cdots His^{182}NE2$) that links the cluster to the carboxyl-terminal end of the preceding αJ; the αH side chains (Arg^{143} and Arg^{147}) that shield the cluster from solvent; and the packing of two turns of αA against the [4Fe-4S] loop. Given side chain flexibility, the greatest potential solvent accessibility of the cluster lies between three side chains contributed by αH and the carboxyl-terminal loop (Arg^{143}, Arg^{147}, and Tyr^{205}) on one side and three side chains contributed by the carboxyl-terminal loop (Arg^{190}, Lys^{191}, and Pro^{192}) on the other, with 8 to 11 Å separating the Cβ atoms of these two sets of side chains.

Interdomain interactions. The 6-helix and [4Fe-4S] domains are covalently linked by a 12-residue extended loop (residues 17 to 28) between αA and αB and an 8-residue extended loop (residues 130 to 137) between αG and αH. Prolines 18, 20, and 29 in the first linker and 133 in the second contribute to the bends observed for these connections. A solvent-filled tunnel, about one water molecule in width, goes entirely through the domain junction, which also has a single internal cavity about the size of three water molecules that is surrounded by αA and αJ, the H-I loop, and the carboxyl-terminal loop (residues Ile^9, Leu^{13}, Phe^{150}, Leu^{179}, His^{182}, and Ile^{199}). There are no aromatic and few hydrophobic packing interactions between the two domains. The principal polar side-chain interactions are charged hydrogen bonds and salt bridges along one side of the domain interface cleft (Glu^{23}-Arg^{184}, Asp^{44}-Arg^{184}, and Glu^{23}-Tyr^{185}). Taken together, the deep cleft, solvent tunnel, minimal hydrophobic packing interactions, and polar side chain connections suggest that the two domains have segmental flexibility so that the interdomain region acts as a hinge.

Thymine glycol recognition. Difference Fourier maps of the inhibitor-soaked crystals (Fig. 4) located the thymine glycol binding site, which is associated with recognition of damaged DNA, as adjacent to the main chain of the F-G loop, a seven-residue β-hairpin (residues 113 to 119; Ala-Leu-Pro-Gly-Val-Gly-Arg). Comparisons of the free and thymine glycol–complexed endonuclease III structures suggest that thymine glycol binding displaces at least two bound water molecules but does not induce

significant conformational changes (Fig. 4). Two main-chain atoms ($Glu^{112}O$ and $Gly^{116}N$) apparently form hydrogen bonds to thymine glycol, although the current electron density map does not allow unambiguous orientation of the inhibitor. The sequence and conformation of this F-G β-hairpin loop region cause an unusual exposure of the main chain of residues 112 to 118, which would allow close contacts with DNA. At lower contour levels, the thymine glycol electron density appears to extend across the amino-terminal end of αG, the only helix without a shielding, hydrogen-bonding N-cap. This helical dipole, enhanced by three nearby positive side chains (Arg^{108} in αF and Arg^{119} and Lys^{120} in αG), could provide favorable interactions with the DNA phosphate backbone.

When mapped onto the endonuclease III structure (Fig. 5), the sequence of MutY (*11*) shows complete conservation in the F-G β-hairpin (Fig. 4). This flexible, highly solvent exposed, yet sequence-conserved, loop packs loosely (at Val^{117}) with helices αB (Leu^{33}), αF (Leu^{111}), and αG (Val^{125}) and forms an internal cavity about the size of one water molecule. On the end (Pro^{115} and Gly^{116}), the F-G β-hairpin packs tightly against the more rigid αE (Asn^{88} and Lys^{91}). Thus the F-G β-hairpin appears to be involved in substrate recognition and is positioned to provide a flexible reading head to accommodate different types of damaged DNA substrates. Differential enzyme specificity between endonuclease III and MutY, which excises the adenine of a G-A mispair, may arise from variability in the 3-D structure and sequence surrounding the loop ($Ala^{40}Gln$, $Ser^{84}Ala$, $Ile^{95}Gln$, $Gln^{99}Leu$, $Ala^{109}Glu$, $Ala^{110}Glu$, $Glu^{112}Ala$, and $Lys^{120}Ser$).

Structural and biochemical implications for endonuclease III–DNA interactions. Although uncertainties in the exact nature of the DNA-enzyme interaction arise from possible local and global conformational changes in endonuclease III (in the apparently flexible thymine glycol binding loop and between domains) and in DNA (including the damaged or abasic site and overall bending), the wealth of detailed information from the 3-D structure of endonuclease III suggests mechanisms for binding and catalysis. Based on our structural analysis, we propose here a potential mode of DNA binding to endonuclease III in order to suggest target areas for biochemical and mutagenic studies. Alternative modes of binding are possible; further biochemical and crystallographic results should provide definitive evidence.

The electrostatic potential around endonuclease III (net charge of +1e) was calculated to investigate possible electrostatic interactions with DNA (*30*). Sequence-conserved residues (Fig. 5A) form a surface-exposed patch of about 25 by 45 Å, with Lys^{120} near the center surrounded by an area of positive electrostatic potential (Fig. 5B). The most highly positive potential surrounds the shallow 24 Å–diameter surface groove and sequence-conserved patch above the [4Fe-4S] cluster, implicating the Fe-S cluster region in DNA phosphate backbone recognition. The opposite face of the enzyme is surrounded by a volume of negative potential, making it an unlikely DNA-binding surface. By analogy with the *trp* repressor-operator complex (*31*), we suggest that the protein's positive charge potential follows the contours of B-DNA.

The proposed combination of nonspecific electrostatic binding to DNA and specific higher affinity binding to damaged DNA by the flexible β-hairpin region that is suggested by the endonuclease III structure is also supported by biochemical binding studies. Abasic DNA, which structurally (*32*) shows little distortion from duplex B-DNA, is protected from deoxyribonuclease I (DNase I) cleavage by endonuclease III. The footprint (*33*), mapped on B-DNA containing an abasic site (Fig. 5C), matches the length for the sequence-conserved region of positive potential (Fig. 5B). Binding constants of the enzyme to DNA oligonucleotides (*34*) indicate nonspecific binding to undamaged DNA with affinities ranging from 3×10^3 for 13-bp molecules to 1×10^5 to 5×10^5 for 39-bp molecules (*33*). This is consistent with nonspecific electrostatic interactions involving the minor groove phosphate backbone. A 300-fold increase in affinity for 13-bp oligonucleotides containing an abasic site is consistent with specific recognition of damaged residues in the context of extremely diverse sequences.

The extensive DNA-protein interfaces that are implicated by footprinting, sequence conservation, and electrostatic potential calculations cross both protein domains and are comparable in size to those of other DNA-protein complexes. The 6-helix barrel domain provides a surface-exposed, sequence-conserved helix (αE) (see Fig. 1B) for electrostatic interactions and the F-G loop for specific recognition of damaged DNA. The [4Fe-4S] cluster domain also provides sequence-conserved regions of positive electrostatic potential and the αH and αI arrangement, which resembles a DNA-binding helix-turn-helix motif. The [4Fe-4S] cluster may act indirectly in DNA recognition by stabilizing this binding motif or, more directly, as an architectural component whose overall negative charge and relatively large size positions four positive charges (αH Arg^{143} and Arg^{147} and cluster loop residues Arg^{190} and Lys^{191}) for interaction with DNA. With the DNA aligned at an angle of about 20° relative to the long axis of the enzyme, αE and αH would lie across the DNA minor groove, one turn apart and on the same side, with positive side chains available for interactions with the DNA phosphate backbone. A similar mode of nonsequence-specific binding to DNA involving positive charges on helices interacting with the minor groove has been proposed for the binding of DNase I, the β-subunit of DNA polymerase, and the HU class of proteins (*35*). This proposed mode of binding would position the F-G β-hairpin in the major groove. The DNA would span most of the sequence-conserved regions, including αE, the αG dipole, αH, αI, the F-G thymine glycol binding loop, and the [4Fe-4S] cluster loop between Cys^{187} and Cys^{194}. Sequence-con-

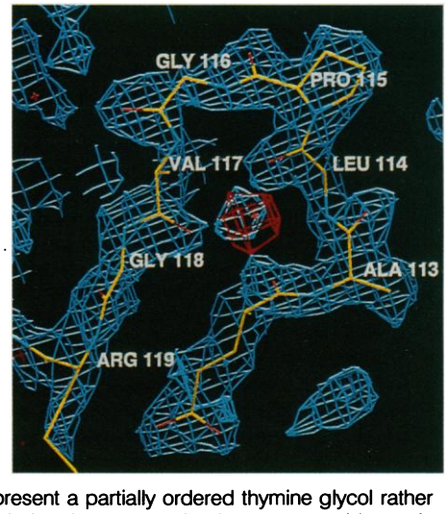

Fig. 4. Thymine glycol inhibitor binding site. The electron density for the native enzyme, calculated with ($|2F_o| - |F_c|$) amplitudes, is displayed with blue contours at 1σ. The protein model without thymine glycol, shown colored by atom type (oxygens, red; nitrogens, blue; and carbons, yellow), was refined against the diffraction amplitudes ($|F_{TG}|$) from the thymine glycol–soaked crystal (Table 1). Thymine glycol binding is shown by a difference ($|F_{TG}| - |F_{nat}|$) electron density map (red contours at 5σ) calculated with phases from the refined model. The highest contour level for the thymine glycol peak is at 7σ as compared to the next highest peak at 3σ in the difference map. The small size of the thymine glycol peak suggests that bound thymine glycol is disordered when separated from DNA. The refined temperature factors for a water molecule placed into this density are lower than those for almost all of the main chain within the β-hairpin loop, suggesting that this peak does represent a partially ordered thymine glycol rather than a change in the existing bound water. Although the electron density does not unambiguously define a bound orientation, the thymine glycol inhibitor appears to displace two bound water molecules to bind across the β-hairpin loop and extend above the amino terminus of helix αG.

served Lys and Arg residues 85, 91, 119, 143, 147, 162, 191, and 193 form areas of positive electrostatic potential. Besides these positively charged residues, five negatively charged residues (Glu^{112}, Asp^{138}, Glu^{157}, Glu^{160}, and Glu^{161}) and five polar side chains (Gln^{41}, Ser^{84}, Asn^{88}, His^{100}, and His^{140}) would fall within the interface region and could be probed by site-directed mutagenesis for roles in DNA building. Although the proposed DNA–endonuclease III interactions would not require large DNA or protein distortions, some movements, including domain-domain adjustments, could be involved in the recognition and catalysis. Conformational flexibility established for abasic sites in DNA (*36*) and DNA bending toward protein found for repressor-DNA structures (*28*) could enhance protein-DNA interactions.

Implications for catalysis from correlated structural and biochemical data. The roles for Fe-S clusters as electron carriers and Lewis acid catalysts in proteins are well documented (*37*). Alternative noncatalytic roles have been discovered recently; for example, the [4Fe-4S] cluster in *Bacillus subtilis* glutamine phosphoribosylpyrophosphate amidotransferase, which is resistant to redox chemistry, maintains the protein's structural integrity rather than participating in its catalytic function (*38*). The endonuclease III [4Fe-4S] cluster is similarly resistant to oxidation and reduction and probably not essential for DNA *N*-glycosylase or AP lyase activity, since the activity-related T4 endonuclease V lacks an Fe-S cluster (*9*). The proposed absence of a direct catalytic role for the endonuclease III [4Fe-4S] cluster is consistent with the absence of spectroscopic changes on base or oligonucleotide binding as shown by Mössbauer results with the inhibitor thymine glycol (*10*) and Raman results with both thymine glycol and oligonucleotide-containing abasic sites (*39*). Overall, these results suggest a structural analogy between the [4Fe-4S] cluster in this class of DNA repair enzymes and the zinc in the zinc finger DNA-binding domains (*12, 27, 28*), with both types of metal sites acting to form an appropriate surface for DNA binding.

Although endonuclease III cleaves the DNA phosphate backbone at AP sites, it does so by a β-elimination rather than the classic phosphodiesterase mechanism (*40*). A strategically located basic residue has been proposed to be involved in this elimination, as evidenced by the ability of certain basic tripeptides (such as Lys-Trp-Lys and Lys-Tyr-Lys) to cleave the DNA backbone at AP sites (*41*), although the enzymatic mechanism is still unknown.

Two charged side chains appear poised for interaction with the DNA backbone and damaged base: Glu^{112} at the carboxyl terminus of αF and Lys^{120} at the amino terminus of αG. Single-bond rotations can bring the charged ends of each of these residues within 3.5 Å of the thymine glycol electron density without collisions or other side chain rearrangements. Based on its chemistry and central location in the proposed DNA binding site, Lys^{120} is the most likely candidate for the formation of the Schiff base associated with AP lyase activity. Glu^{112} is surrounded by several basic side chains (Lys^{91}, Arg^{94}, Arg^{108}, and Arg^{119}), as is the glutamic acid of endonuclease V implicated by mutagenesis to be responsible for *N*-glycosylase activity (*9*). Glu^{112} underlies the implicated thymine glycol binding site, making it the most likely nucleophile in the *N*-glycosylase reaction. There are at least four other negatively charged side chains within the proposed DNA contact region: αI Glu^{157}; Glu^{160}; Glu^{161}; and αH Asp^{138}. The nearest of these cannot be brought closer than 12 Å to the thymine glycol binding site with single-bond rotations and would thus require considerable conformational change to participate directly in the mechanism.

The structure of *E. coli* endonuclease III,

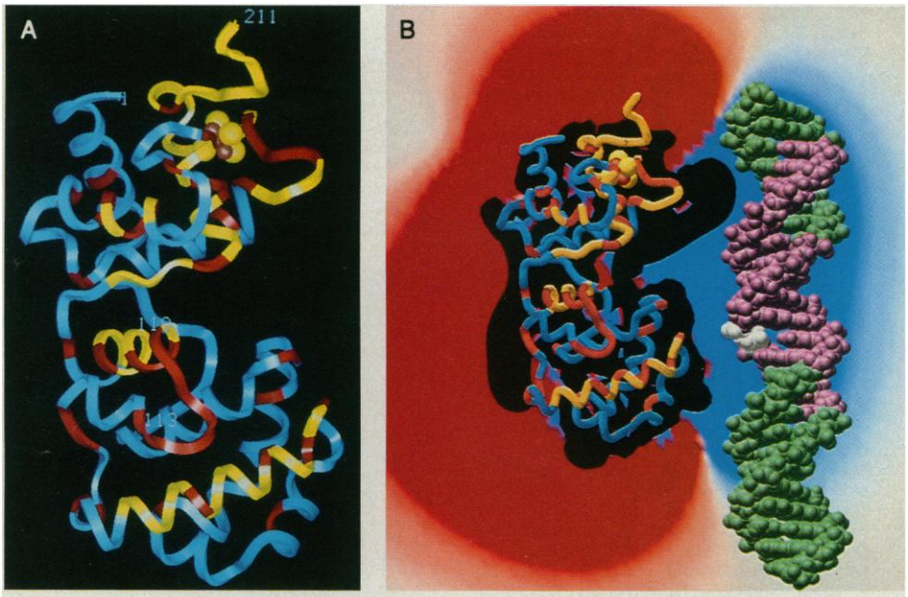

Fig. 5. DNA binding surface implicated by sequence conservation, electrostatic potential, and DNA protection. (**A**) Sequence-conserved regions. A ribbon representation of the endonuclease III model is colored to highlight the sequence conservation between endonuclease III and the homologous adenine glycosylase MutY. Red denotes identical residues in both proteins (three Arg-Lys substitutions are treated as identical residues). Yellow denotes homologous regions as judged by the high degree of sequence conservation within a secondary structural element, indicating potentially conserved secondary structures and critical residues that may be involved in the interaction with substrate DNA. Nonhomologous residues are blue, and the [4Fe-4S] cluster is shown by brown (iron) and yellow (sulfur) spheres. The red F-G loop between residues 113 and 119 is the thymine glycol binding site as defined by difference maps. The sequence of this conserved seven-residue loop is Ala-Leu-Pro-Gly-Val-Gly-Arg. (**B**) Association of the sequence-conserved face with a long stripe of positive electrostatic potential. Electrostatic potential, calculated with a Poisson-Boltzmann approximation (*30*), is color-coded (0.75 to 0.1 kcal/mol is blue for positive and −0.75 to −0.1 kcal/mol is red for negative) to illustrate the dimensions of positive and negative potential. Positive electrostatic potential surrounds Lys^{120} and is contributed to by amino-terminal Lys^{3}, Lys^{5}, and Arg^{6}; αE Lys^{85}, Lys^{91}, and Arg^{94}; αF Arg^{108}; αG Arg^{119} and Lys^{120}; αH His^{140}, Arg^{143}, and Arg^{147}; H-I loop Lys^{154}; αI Lys^{162} and Lys^{165}; [4Fe-4S] loop Arg^{190}, Lys^{191}, and Arg^{193}; and carboxyl-terminal loop Lys^{208}. The positive and negative electrostatic potential surfaces are shown for a slice running through the sequence-conserved face and the [4Fe-4S] cluster (AVS software on a Stardent workstation). The endonuclease III chain tracing (tubes) and [4Fe-4S] cluster are colored as noted above, and the protein is in the same orientation as in (A). The DNase I footprint of the endonuclease III binding site (purple) is shown within the context of the double-stranded 25-bp oligonucleotide (green) used experimentally (see C) and modeled as B-DNA (right). The DNA abasic site is white and placed near the F-G β-hairpin (red tubes). If we assume that the DNase footprint gives a 3-bp overestimate at the 5′ end (*33*), the offset in the protected region on the two strands produces an overall protected extent of 15 bp (~46 Å), which closely matches the length of area defined by the highly positive potential (>0.75 kcal/mol). (**C**) Schematic of DNase I footprint for binding of endonuclease III to a synthetic duplex containing a reduced AP (rAP) site (*33*). Brackets denote the protected base sequences on both strands.

which represents a common class of enzymes in bacteria, yeast, bovine, and human cells, may be considered a prototype for the *N*-glycosylase and AP lyase activities in bacteria and higher organisms. The suggested *N*-glycosylase active site residue Glu^{112} is consistent with the Glu^{23} proposed from mutational analysis in the recently determined structure of the T4-phage enzyme endonuclease V (*9*). A separate study indicates that the amino terminus of endonuclease V, which is near Glu^{23}, may be involved in the AP lyase activity (*42*), as proposed for endonuclease III Lys^{120}. Thus these completely different enzymes may have evolved similar mechanisms of activity. The detailed structure of endonuclease III, including the association of the β-hairpin with recognition of thymine glycol and of the novel [4Fe-4S] cluster with the orientation of basic side chains for potential DNA binding provides the structural basis to guide future biochemical and mutagenic studies addressing the detailed molecular mechanisms for damaged DNA recognition and for the DNA *N*-glycosylase and AP lyase activities.

REFERENCES AND NOTES

1. B. N. Ames, *Science* **221**, 1256 (1983); P. A. Cerutti, *ibid.* **227**, 375 (1985); C. G. Fraga, M. K. Shigenaga, J.-W. Park, P. Degan, B. N. Ames, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4533 (1990); D. Harman, *ibid.* **78**, 7124 (1981); R. Adelman, R. L. Saul, B. N. Ames, *ibid.* **85**, 2706 (1988).
2. T. Lindahl, *Annu. Rev. Biochem.* **51**, 61 (1982); *Prog. Nucleic Acids Res.* **22**, 135 (1979); *Mutat. Res.* **238**, 305 (1990).
3. H. L. Katcher and S. S. Wallace, *Biochemistry* **22**, 4071 (1983); P. W. Doetsch and R. P. Cunningham, *Mutat. Res.* **236**, 173 (1990).
4. E. C. Friedberg, *DNA Repair* (Freeman, New York, 1985); H. Ahern, *ASM News* **57**, 627 (1991); G. Storz, L. A. Tartaglia, S. B. Farr, B. N. Ames, *Trends Genet.* **6**, 363 (1990).
5. H. Asahara, P. M. Wistort, J. F. Bank, R. H. Bakerian, R. P. Cunningham, *Biochemistry* **28**, 4444 (1989).
6. T. J. Jorgenson, Y. W. Kow, S. S. Wallace, W. D. Henner, *ibid.* **26**, 6436 (1987).
7. J. Gossett, K. Lee, R. P. Cunningham, P. W. Doetsch, *ibid.* **27**, 2629 (1988).
8. D. E. Helland, P. W. Doetsch, W. A. Haseltine, *Mol. Cell. Biol.* **6**, 1983 (1986); P. W. Doetsch, W. D. Henner, R. P. Cunningham, J. H. Toney, D. E. Helland, *ibid.* **7**, 26 (1987).
9. K. Morikawa *et al.*, *Science* **256**, 523 (1992).
10. R. P. Cunningham *et al.*, *Biochemistry* **28**, 4450 (1989).
11. M. L. Michaels, L. Pham, Y. Nghiem, C. Cruz, J. H. Miller, *Nucleic Acids Res.* **18**, 3841 (1990).
12. S. C. Harrison, *Nature* **353**, 715 (1991).
13. C.-F. Kuo, D. E. McRee, R. P. Cunningham, J. A. Tainer, *J. Mol. Biol.* **227**, 347 (1992).
14. A. J. Howard *et al.*, *J. Appl. Crystallogr.* **20**, 383 (1987).
15. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
16. T. Williams, thesis, University of North Carolina, Chapel Hill (1982).
17. T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
18. A. T. Brunger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
19. J. S. Richardson and D. C. Richardson, in *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, Ed. (Plenum, New York, 1990), pp. 2–98.
20. S. Pascarella and P. Argos, *Protein Eng.* **5**, 121 (1992); C. Chothia and A. V. Finkelstein, *Annu. Rev. Biochem.* **59**, 1007 (1990).
21. J. S. Richardson and D. C. Richardson, *Science* **240**, 1648 (1988).
22. J. S. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981).
23. R. Henderson *et al.*, *J. Mol. Biol.* **213**, 899 (1990).
24. J. Li, J. Carroll, D. J. Ellar, *Nature* **353**, 815 (1991).
25. J. S. Richardson and D. C. Richardson, *Proteins* **4**, 229 (1988).
26. C. W. Heizmann and W. Hunziker, *Trends Biochem. Sci.* **16**, 98 (1991).
27. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 933 (1990); R. G. Brennan and B. W. Matthews, *J. Biol. Chem.* **264**, 1903 (1989).
28. T. A. Steitz, *Q. Rev. Biophys.* **23**, 205 (1990).
29. H. Beinert, *FASEB J.* **4**, 2483 (1990).
30. The electrostatic potential surface around endonuclease III was calculated from the structure with united atom partial charges rather than formally charged residues. AMBER partial charges for the standard amino acid residues [S. J. Weiner *et al.*, *J. Am. Chem. Soc.* **106**, 765 (1984)] were assigned to the atomic coordinates, including polar hydrogen atoms added by InsightII (BIOSYM Technologies, Inc.). Local density functions [W. Ravenek, in *Scientific Computing on Supercomputers*, J. T. Devreese and P. E. Van Camp, Eds. (Plenum, New York, 1989), pp. 201–218; F. M. Bickelhaupt, E. J. Baerends, W. Ravenek, *Inorg. Chem.* **29**, 350 (1990); P. M. Boerrigter, G. Te Velde, E. J. Baerends, *Int. J. Quantum Chem.* **33**, 87 (1988)] and CHELPG [a program that calculates atomic charges from electrostatic potentials determined in ab initio molecular orbital computations; C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.* **11**, 361 (1990); L. E. Chirlian and M. M. Francl, *ibid.* **8**, 894 (1987)] results (J.-L. Chen, L. Noodleman, J.-M. Mouesca, D. A. Case, D. Bashford, unpublished results) provided charges for the [4Fe-4S] cluster and the four Cys ligands. Assuming all lysines, arginines, and the amino terminus to have an overall +1e charge, all aspartates, glutamates, and the carboxyl terminus to have an overall −1e charge, and all histidines to be neutral, the net charge calculated for endonuclease III was +1e. The electrostatic potential around the protein was calculated from a linearized Poisson-Boltzmann equation with the UHBD program [developed by J. D. Madura and M. E. Davis at the University of Houston working in collaboration with J. A. McCammon; M. E. Davis and J. A. McCammon, *Chem. Rev.* **90**, 509 (1990)] with the use of a solvent ionic strength of 150 μM and a dielectric constant of 2 within and 78 outside the protein. The potential at the grid boundary was set to the value calculated when treating the molecule as a single Debye-Hückel sphere of radius 30.0 Å with dielectric boundary smoothing. B-DNA model was built by using the Biopolymer module of InsightII.
31. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
32. M. W. Kalnik, C.-N. Chang, F. Johnson, A. P. Grollman, D. J. Patel, *Biochemistry* 28, 3373 (1989); J. M. Withka, J. A. Wilde, P. H. Bolton, A. Mazumder, J. A. Gerlt, *ibid.* **30**, 9931 (1991).
33. A 25-bp oligonucleotide with a single uracil in the center was synthesized and the uracil removed with uracil-DNA glycosylase. The resulting AP site was reduced with sodium borohydride. The modified strand was annealed to its complement to yield a noncleavable substrate (unpublished results). For the footprinting experiments, DNA concentration was 10 nM and that of endonuclease III was 20 μM. The cleavage reactions were carried out with varying amounts of DNase I for 1 min and quenched with 20 mM EDTA. Samples were run on a 20% denaturing polyacrylamide gel. DNase I footprinting may overestimate the recognition site due to steric effects, particularly at the 5′ end [T. D. Tullius, *Annu. Rev. Biophys. Biophys. Chem.* **18**, 213 (1989)]. Nuclear magnetic resonance studies indicate that there are primarily local changes in B-DNA associated with forming an abasic site (*32*). The binding constant for reduced AP DNA was determined by quantitative footprint experiments. Binding constants for undamaged B-DNA were determined by competition assay in a gel shift experiment.
34. S. O'Handley and R. P. Cunningham, unpublished results.
35. M. E. A. Churchill and A. A. Travers, *Trends Biochem. Sci.* **16**, 92 (1991); X.-P. Kong, R. Onrust, M. O'Donnell, J. Kuriyan, *Cell* **69**, 425 (1992).
36. A. Schwartz, L. Marrot, M. Leng, *J. Mol. Biol.* **207**, 445 (1989); G. S. Manning, *Biopolymers* **22**, 689 (1983).
37. R. Cammack, *Adv. Inorg. Chem.* **38**, 1 (1991).
38. R. L. Switzer, *Biofactors* **2**, 77 (1989).
39. W. Fu, S. O'Handley, R. P. Cunningham, M. K. Johnson, *J. Biol. Chem.* **267**, 16135 (1992).
40. A. Mazumder *et al.*, *Biochemistry* **30**, 1119 (1991); V. Bailly and W. G. Verly, *Biochem. J.* **242**, 565 (1987).
41. T. Behmoaras, J.-J. Toulmé, C. Hélène, *Nature* **292**, 858 (1981).
42. M. L. Augustine, R. W. Hamilton, M. L. Dodson, R. S. Lloyd, *Biochemistry* **30**, 8052 (1991).
43. T. C. Terwilliger, S. H. Kim, D. Eisenberg, *Acta Crystallogr.* **A43**, 1 (1987).
44. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
45. D. E. McRee, *J. Mol. Graphics* **10**, 44 (1992).

46. Supported by NIH grant GM 46312 and NIH training grant HL07695 to C.L.F. We thank H. Parge, E. Getzoff, D. Goodin, A. Karplus, J. Li, D. Stout, and P. Wright for their intellectual contributions; L. Noodleman for [4Fe-4S] cluster charges; and M. Pique for computer graphics using the AVS graphics software. Atomic coordinates will be deposited in the Brookhaven Protein Data Bank upon publication.

3 June 1992; accepted 25 August 1992