Structure, Vol. 12, 205–214, February, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2004.01.004

DNA Binding and Degradation by the HNH Protein ColE7

Kuo-Chiang Hsia,[1] Kin-Fu Chak,[3] Po-Huang Liang,[2] Yi-Sheng Cheng,[1] Wen-Yen Ku,[1] and Hanna S. Yuan[1,*]
[1]Institute of Molecular Biology
[2]Institute of Biological Chemistry
Academia Sinica
[3]Institute of Biochemistry
National Yang-Ming University
Taipei
Taiwan

Summary

The bacterial toxin ColE7 bears an HNH motif which has been identified in hundreds of prokaryotic and eukaryotic endonucleases, involved in DNA homing, restriction, repair, or chromosome degradation. The crystal structure of the nuclease domain of ColE7 in complex with a duplex DNA has been determined at 2.5 Å resolution. The HNH motif is bound at the minor groove primarily to DNA phosphate groups at and beyond the 3′ side of the scissile phosphate, with little interaction with ribose groups and bases. This result provides a structural basis for sugar- and sequence-independent DNA recognition and the inhibition mechanism by inhibitor Im7, which blocks the substrate binding site but not the active site. Structural comparison shows that two families of endonucleases bind and bend DNA in a similar way to that of the HNH ColE7, indicating that endonucleases containing a "ββα-metal" fold of active site possess a universal mode for protein-DNA interactions.

Introduction

Chromosomal DNA degradation is the hallmark of programmed cell death (apoptosis) (Zhang and Xu, 2002). Several nucleases have been identified that are involved in chromosome fragmentation during apoptosis, including DFF40/CAD (40 kDa DNA fragmentation factor/caspase-activated deoxyribonuclease) (Enari et al., 1998; Liu et al., 1997), mitochondrial endoG (Li et al., 2001), and CRN (cell death-related nucleases) (Parrish and Xue, 2003). The first identified apoptotic nuclease, DFF40/CAD, is normally inhibited by its inhibitor DFF45/ICAD in cytoplasm and is activated by caspase 3, which cleaves the inhibitor, resulting in DNA fragmentation in nuclei. In prokaryotes, there are also several nucleases capable of DNA degradation and cell killing, such as *Escherichia coli* nuclease-type colicins (James et al., 1996) ColE2, ColE7, ColE8, and ColE9, and *Pseudomonas aeruginosa* pyocins (Michel-Briand and Baysse, 2002) S1, S2, S3, and AP41. These colicins or pyocins are secretory endonucleases which are expressed and released to kill other closely related bacterial cells by degrading their chromosomal DNA when the host bacteria encounter environmental stresses. Colicins and pyocins are inhibited by their cognate inhibitors in the host cell and the DNase activity is released after removal of inhibitor proteins (Kleanthous and Walker, 2001). Remarkably, DFF40/CAD, colicins, and pyocins all contain an HNH motif in their endonuclease domains (see Figure 1). Therefore, colicins and DFF40/CAD seem to degrade DNA chromosomes during the programmed cell death in a comparable way in both prokaryotic and eukaryotic cells, respectively.

The HNH motif was first recognized by its sequence similarity in several phage and bacteria intron-encoded homing endonucleases about a decade ago (Gorbalenya, 1994; Shub et al., 1994). To date, more than 200 proteins containing the HNH motif have been identified in all phylogenic kingdoms, from bacteria phage up to human (Sui et al., 2002; Walker et al., 2002). Many HNH proteins are site-specific group I or II homing endonucleases, which initialize the process of transferring a mobile intervening sequence into a homologous allele that lacks the sequence (Chevalier and Stoddard, 2001; Lambowitz and Belfort, 1993). These include I-*Hmu*I (Goodrich-Blair et al., 1990), I-*Hmu*II (Goodrich-Blair and Shub, 1996), I-*Hmu*III (Goodrich-Blair, 1994), I-*Tev*III (Eddy and Gold, 1991), I-*Two*I (Landthaler et al., 2002), I-*Cmoe*I (Drouin et al., 2000), yosQ (Lazarevic et al., 1998), ORF253 (Foley et al., 2000), Avi (Ferat and Michel, 1993), Cpc1 (Ferat and Michel, 1993), I-*Sec*V (Zimmerly et al., 1995), and PetD (Kuck, 1989).

Homing endonucleases are classified into four subgroups, within which the HNH subfamily is the least structurally characterized. Except HNH, all other subfamilies of homing endonucleases have crystal structures reported to show how they recognize and cleave DNA, such as the protein/DNA complex structures of I-*Cre*I (Jurica et al., 1998) and PI-*Sce*I (Moure et al., 2002) in the LAGLIDADG subfamily, I-*Tev*I (Van Roey et al., 2002; Van Roey et al., 2001) in the GIY-YIG subfamily, and I-*Ppo*I (Flick et al., 1998) in the His-Cys box subfamily of homing endonucleases. A unique feature for the HNH subfamily of homing endonucleases is that they are most divergent in their endonuclease activity, making either one or two nicks in DNA and generating 5′ or 3′ overhangs in their recognition sequences (Chevalier and Stoddard, 2001). Moreover, the HNH proteins are not only involved in homing but also carry other biological functions, such as DNA degradation (Walker et al., 2002), repair (Eisen, 1998), and restriction (Hiom and Sedgwick, 1991; Piekarowicz et al., 1991).

The DNase-type colicins and pyocins constitute the second biggest subgroup of HNH proteins. These toxins contain three functional domains responsible respectively for receptor binding, membrane translocation, and cytotoxic endonuclease activity against target cells (James et al., 1996). The nuclease domains of colicins share high sequence identity (~65%), all containing the HNH motif at the C-terminal end with a consensus sequence of **EXH$\underline{H}$X$_{14}\underline{N}$X$_8\underline{H}$X$_3$H** (Sui et al., 2002). The crystal

*Correspondence: hanna@sinica.edu.tw

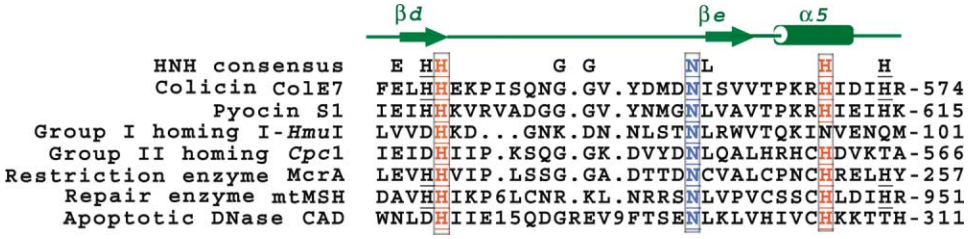

Figure 1. Sequence Alignment of the HNH Motif in Several Representative Protein Subfamilies

One example of each subgroup of HNH endonucleases are listed, including colicin ColE7 (Chang et al., 1992), pyocin S1 (Sano et al., 1993), group I homing endonuclease I-*Hmu*I (Goodrich-Blair et al., 1990), group II homing endonuclease Cpc1 (Ferat and Michel, 1993), restriction enzyme McrA (Hiom and Sedgwick, 1991), repair enzyme mtMSH (Eisen, 1998), and apoptotic endonuclease CAD/DEF40 (Enari et al., 1998; Liu et al., 1997). The secondary structure for the HNH motif in nuclease-ColE7 is shown on the top row of the sequences. The numbers in the sequences indicate the insertion amino acids.

structures of the nuclease domains of ColE7 (Cheng et al., 2002; Ko et al., 1999) and ColE9 (Kleanthous et al., 1999) in complex with their respective inhibitors Im7 and Im9 are the only structures reported thus far to reveal the structural fold of the HNH motif. This motif contains two antiparallel β strands connected to an α helix with a centrally located zinc ion bound to three conserved histidines. Biochemical assays demonstrate that the zinc ion in the HNH motif is involved in structural stabilization (Pommer et al., 1999) and catalytic hydrolysis (Ku et al., 2002). One of the conserved histidines in the motif (His545 in ColE7) likely functions as the general base (Ku et al., 2002) while cleaving at the 3′ side of the phosphodiester bond with a preference to make nicks at thymine bases in double-stranded DNA (Pommer et al., 2001).

Structural comparison (Friedhoff et al., 1999; Miller et al., 1999) further reveals that the ββα-metal topology observed for the HNH motif is similar to those of the active sites in several families of endonucleases, including the His-Cys box homing endonuclease I-*Ppo*I (Flick et al., 1998), nonspecific *Serratia* nuclease (Miller et al., 1994), and the Holliday junction-specific phage T4 endonuclease VII (Raaijmakers et al., 1999). The crystal structure of Vvn (Li et al., 2003), which belongs to a new family of nonspecific extracellular nucleases, also shows an active site with a topology similar to that of HNH motif. Moreover, the apoptotic endonucleases, including DEF40/CAD which contains a region homologous to the HNH motif (Walker et al., 2002), and endoG (Li et al., 2001) which shares sequence homology in the active site to that of *Serratia* nuclease, might also bear a similar topology of active site to that of ColE7. Therefore, a structural study of how DNA is recognized by the HNH protein may address the mode of DNA interactions, not only for proteins in the HNH family but also for the ones in a variety of endonuclease families. Here we report the crystal structure of the endonuclease domain of ColE7 in complex with an 8 bp duplex DNA at a resolution of 2.5 Å. This structure shows for the first time how an HNH family protein interacts with duplex DNA and represents an example of nonspecific recognition between proteins and DNA.

Results

Substrate Characterization

The nuclease domain of ColE7 (referred to as nuclease-ColE7 hereafter) cleaves double-stranded DNA in the presence of divalent metal ions (Ku et al., 2002). The homologous ColE9 cleaves double- and single-stranded DNA of lengths longer than 10 bases with better activity, but not small substrates, such as dinucleotides or nucleotide analogs (Pommer et al., 2001). Here we used a 100 base single-stranded poly(dC) as the substrate to assay the minimum length of DNA required for cleavage by nuclease-ColE7. We found that a small oligonucleotide, of length less than 7 bases, was left uncleaved (see Figure 2A), indicating that a minimum size of DNA is indeed required for enzyme activity. Further analysis using 4–8 bases of DNA with different sequences as the substrate (see Figure 2B) showed that DNA with a length of at least 8 bases is required for nuclease-ColE7 cleavage.

One of the nuclease-type colicins, ColE2, was previously shown as being unable to degrade RNA prepared from phage R17 (Schaller and Nomura, 1976). However, ColE9 contains RNase activity when incubated with a synthetic, fluorescently labeled ssRNA with a length of 10 bases (Pommer et al., 2001). To clarify this issue with regards to ColE7, we used a zymogram method to measure the RNase activity of nuclease-ColE7. Yeast total RNA containing both single-stranded and double-stranded RNA were used as substrates which were suspended in an electrophoresis gel for enzyme cleavage. The RNA was digested by nuclease-ColE7 and RNase A (positive control) but not by DNase I (negative control) at their respective positions in the gel (see Figure 2C). Therefore, this experiment confirms that nuclease-ColE7 indeed contains RNase activity.

Structure Determination and Overall Structure

The nuclease domain of ColE7 containing residues 444–576 was cocrystallized with the duplex DNA with a palindromic sequence of 5′-GCGATCGC-3′. To avoid digestion of DNA, EDTA was added into the crystallization solution to remove the divalent metal ion required for enzyme activity (Ku et al., 2002). The apo-enzyme/DNA complex crystallized in monoclinic $P2_1$ space group, diffracting X-rays to a resolution of 2.5 Å using synchrotron radiations. The structure was solved by molecular replacement using the crystal structure of nuclease-ColE7 (PDB entry: 7CEI) in the heterodimer of nuclease-ColE7/Im7 (Ko et al., 1999) as the search model. The final model contained two protein molecules (chain A: residues 449–576 and chain B: residues 449–576), three

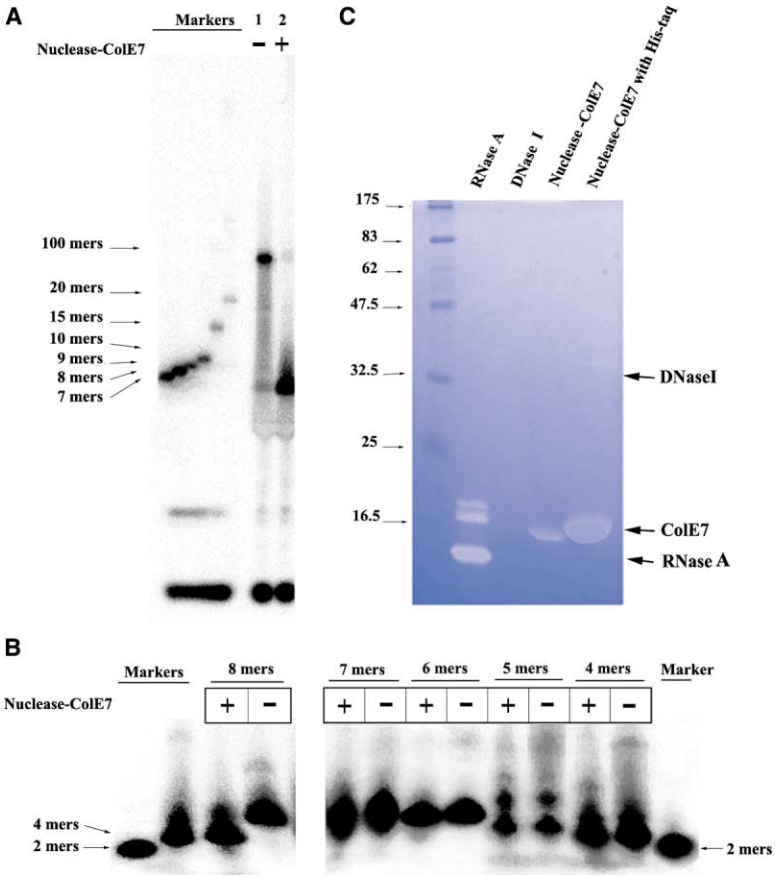

Figure 2. Characterization of the Endonuclease Activity of ColE7

(A) A ^{32}P-labeled single-stranded poly(dC) of 100 nucleotides was used as the substrate for the cleavage by nuclease-ColE7. Lanes 1 and 2 show the cleaved DNA in the absence or presence of nuclease-ColE7. Based on the size of the markers, small nucleotides <7 bases are left uncleaved.

(B) Various sizes of ssDNAs, 4-mer (GATC), 5-mer (CGATC), 6-mer (GATCG), 7-mer (CGAT CGC), and 8-mer (GCGATCGC), were labeled with ^{32}P at their 5′ ends and incubated with nuclease-ColE7 overnight. Only the 8-mer DNA was cleaved by nuclease-ColE7.

(C) The RNase activity of nuclease-ColE7 was assayed by a zymogram method using the total yeast RNA suspended in the gel as the substrate. The white bands indicate that RNA substrates were digested by nuclease-ColE7 (with or without His-taq) and RNase A (positive control), but not by DNase I (negative control).

duplex DNA molecules (chains C/D, E/F, and G/H), and 232 water molecules.

The overall structure of the two nuclease-ColE7 molecules and three duplex DNA in one asymmetric unit is shown in Figure 3. The protein molecule A is bound to C/D chains of DNA, and molecule B is bound to E/F chains of DNA, while the G/H chains of DNA, located in between the two protein/DNA complexes, are unbound. The three duplex DNA molecules appear to pack as a continuous DNA, but the twist angles between the interfaced base pairs are $-32°$ and $-29°$, indicating that they are not stacked as a pseudo-continuous DNA. The two protein/DNA complexes, A/CD and B/EF, match well with each other, with an average rms difference of 1.36 Å for all atoms when the two complexes are superimposed. When only the protein molecules are superimposed, they differ by an average rms difference of 0.6 Å (for all Cα atoms), but the two bound DNA duplexes differ by 1.9 Å (for all P atoms). Therefore, the two protein molecules are almost identical to each other, but the two DNA duplexes have slightly different conformations when binding to the protein. For clarity, we have used only the complex of A/CD for structural analysis and comparison in most cases in the following sections.

The structure of the nuclease-ColE7 has a mixed α/β fold, containing a central three-stranded antiparallel β sheet (βb-βe-βd) packed against four α helices (α1–α4) with a two-stranded β sheet (βa–βc) capping on the side and a small α helix (α5) flanking at the other side. The two strands, βd and βe, and the helix α5 in the HNH motif located in the C terminus of nuclease-ColE7 are displayed in red in Figure 3A. Nuclease-ColE7 binds to DNA as a monomer with the HNH motif facing the minor groove of DNA. The N terminus of the α2 helix contacts the DNA at the neighboring major groove so that one of the DNA strands is clamped in between the HNH motif and α2 helix. Previous biochemistry data had suggested that nuclease-ColE7 may bind DNA as a homodimer (Cheng et al., 2002); however, the DNA used for crystallization here is only an octamer which might be too short to promote dimer formation. The nuclease-ColE7 in the protein/DNA complex fits well with the free form of nuclease-ColE7 (Cheng et al., 2002), shifting with an average rms difference of only 1.03 Å for Cα atoms, indicating that nuclease-ColE7 does not change its overall conformation upon DNA binding.

Endonuclease Active Site and Protein-DNA Interactions

The endonuclease active site in nuclease-ColE7 is located around the Zn^{2+} binding site where the metal ion binding is required for the DNase activity (Ku et al., 2002). Although the zinc ion in the nuclease-ColE7/DNA complex was removed by EDTA during crystallization, the electron density surrounding the metal ion binding site revealed that the three histidine side chains, His544, His569, and His575, still retain a well-defined conformation in the absence of direct coordination to Zn^{2+} (see Figure 3B). The superimposition of the apo-nuclease-ColE7/DNA complex (A/CD complex) with the zinc-bound nuclease-ColE7/HPO_4^{-2} complex (Sui et al., 2002) further shows that the histidine side chains, His544 and

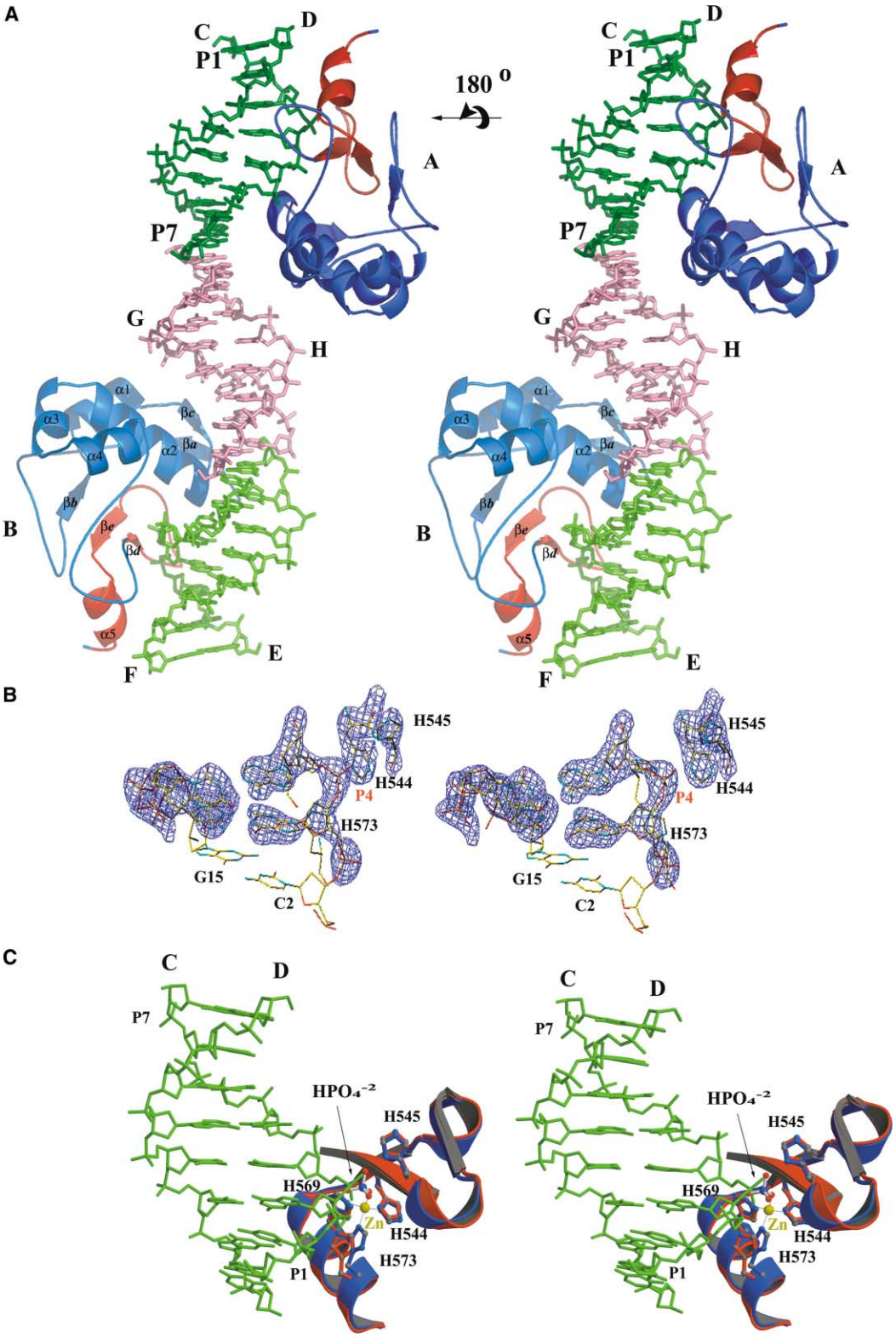

Figure 3. Crystal Structure of the Nuclease Domain of ColE7 in Complex with DNA

(A) Stereo view of the two nuclease-ColE7 (A and B chains) and three DNA duplexes (C/D, E/F, and G/H chains) in one asymmetric unit of the $P2_1$ monoclinic unit cell. The HNH motif (βd, βe, and α5) facing the DNA minor groove is displayed in red. The α2 helix contacts DNA at the neighboring major groove.

(B) Stereo view of the omit difference (Fo-Fc) map contoured at 2.5 σ shows that the DNA and proteins all have a well-defined density. The map was calculated after omitting DNA base pairs Gua3-Cyt14 and Ade4-Thy13, protein residues His545 and His544, and all the atoms within

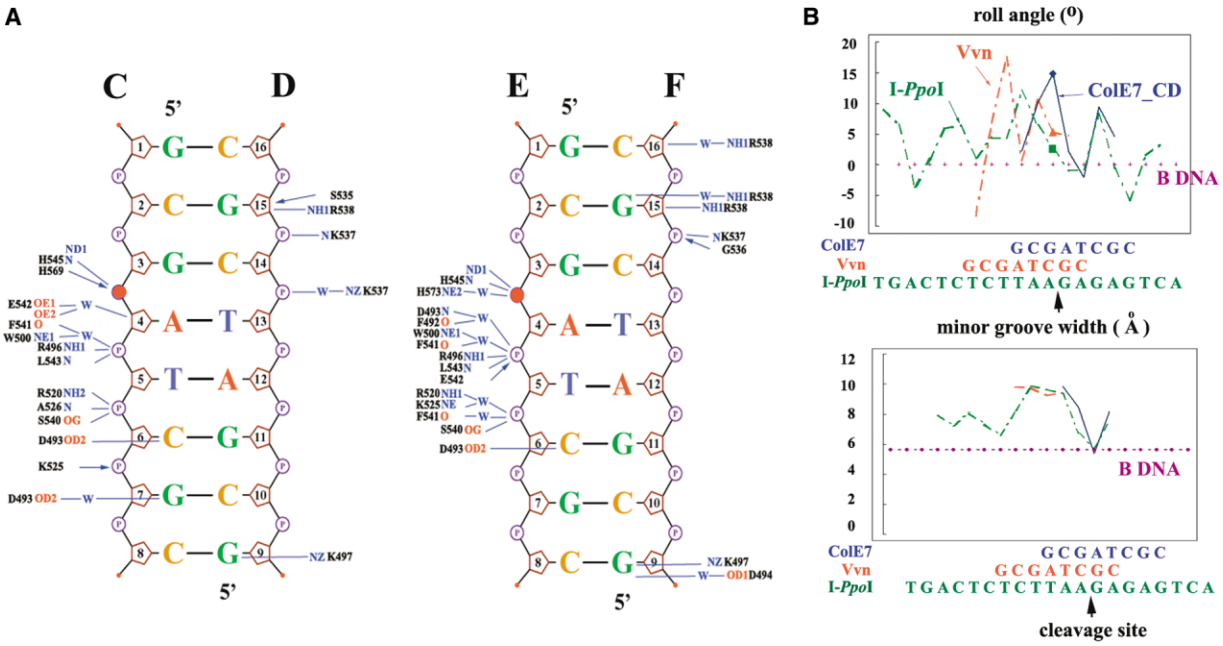

Figure 4. The Interactions between the Nuclease-ColE7 and DNA

(A) The schematic diagrams display the contacts between A/CD and B/EF of nuclease-ColE7/DNA complexes. The solid lines indicate hydrogen bonds (<3.5 Å), and arrows indicate van der Waals contacts (<3.35 Å) between protein and DNA molecules. The scissile phosphates (P4) are circled in red.

(B) The DNA conformation parameters, roll angles, and minor groove widths are plotted for each base step of the DNA bound to nuclease-ColE7 (this study, in blue), Vvn (pdb entry: 1OUP, in red), and I-PpoI (pdb entry: 1A74, in green). DNA cleavage sites are marked by a solid square, circle, or triangle in the figure, and marked by an arrow in the DNA sequences which are aligned at the cleavage sites and listed below the figure. The roll angles are positive values of ~5°–15° around the scissile phosphates and the minor groove widths are widened to ~9 Å at the region bound to the ββα-metal motif in all three protein-DNA complexes.

His569, fit well with each other, but those of His573 deviate slightly. The HPO_4^{-2} in the holo-enzyme almost matches with the P4 phosphate in the octa-nucleotide (see Figure 3C). This result indicates that the P4 group is likely a scissile phosphate which might bind to the zinc ion in the HNH motif during hydrolysis.

Most of the minor-groove binding proteins greatly bend DNA and widen the minor groove (Bewley et al., 1998); however, the DNA in the nuclease-ColE7/DNA complex seems relatively straight. The DNA octamers are only slightly bent with a bending angle of ~7° (for A/CD complex, calculated by Curve [Lavery and Sklenar, 1988]). Upon closer examination of DNA base stacking parameters, large positive roll angles are observed between the base pair steps C2-G3 and G3-A4, consistent with a widened minor groove width of ~9 Å in this region of DNA (see Figure 4). The minor groove width in other regions of the DNA is close to the value of 5.6 Å of an ideal straight B-form DNA. Therefore, the binding of the HNH motif at the minor groove indeed induces DNA conformational change. A more drastic DNA bending is not observed because the HNH motif is bound at the 5′ end of the duplex DNA, but not at the center of this short duplex DNA, as is the case with Vvn.

There are altogether 14 hydrogen bonds and three van der Waals contacts in the buried protein surface of 682 Å^2 for the A/CD complex. These numbers are lower than the average of 20–30 contacts in the buried surface of ~1600 Å^2 found in most of protein/DNA complexes (Jones et al., 1999; Nadassy et al., 1999). This is likely because a short octamer DNA is used in this study. More contacts and buried surface might be revealed with a longer DNA. In the complex, the protein residues involved in the interactions are clustered in two regions, the βd strand in HNH motif binding at the minor groove and α2 helix binding at the major groove. The DNA phosphate groups are dominantly responsible for the protein contact, especially on the side 3′ to the scissile phosphate, including P4, P5, and P6 phosphate groups, which make extensively interactions. There are only two direct hydrogen bonds between DNA base and protein side chain atoms (O6 of GUA9 bound to Nz of K497, N4

a 3 Å spherical shell, followed by a full round of positional and simulated annealing refinement at 2.5 Å resolution. This view is rotated 180° horizontally to that of the A/CD molecule in (A).

(C) Stereo view of the superimposition of the HNH motif in the nuclease-ColE7/DNA complex (in red) with that of the zinc-bound nuclease-ColE7/HPO_4^{-2} (in blue, PDB entry: 1MZ8). This view is rotated 180° horizontally to that of (A). The zinc ion is displayed as a yellow sphere, which is bound to the three histidine residues (His545, His569, and His573) and the HPO_4^{-2} in the holo-enzyme. The HPO_4^{-2} is located closely to the P4 phosphate group in the DNA octamers. Therefore, in the presence of zinc ion, the scissile phosphate, presumably the P4 phosphate, likely binds to the zinc ion in the HNH motif.

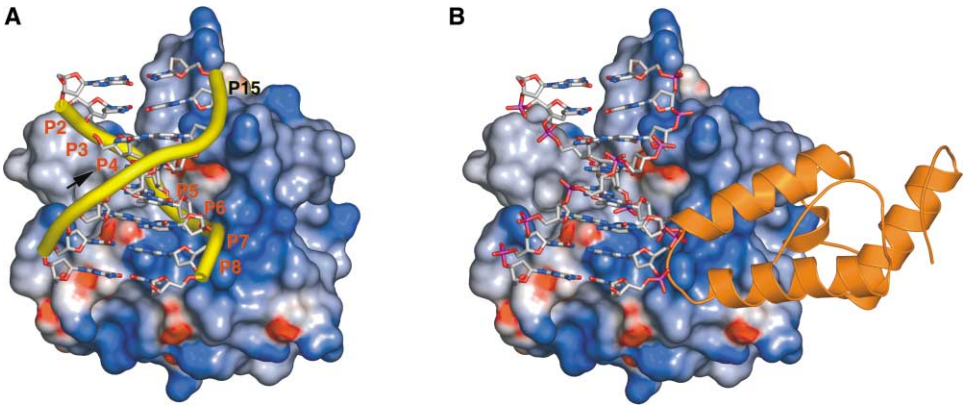

Figure 5. The Molecular Surface of Nuclease-ColE7, Color Coded to Represent Electrostatic Potential

(A) One of the DNA strands is bound at the central cleft region and DNA is cleaved at the P4 phosphate group, indicated by an arrow. The most basic surface (displayed in blue) of the protein contacts P5, P6, and P7 phosphate groups located at the 3′ side of the scissile phosphate. (B) The inhibitor Im7 protein is bound to this basic region, based on the structure of nuclease-ColE7/Im7 complex (PDB entry: 7CEI) (Ko et al., 1999). Therefore, Im7 blocks the major substrate binding site but not the endonuclease active site.

of CYT6 bound to OD2 of D493). Compared to other site-specific or nonspecific protein-DNA complexes with an average of six contacts at the bases (Luscombe and Thornton, 2002), the structure of nuclease-ColE7/DNA demonstrates a rare example of little contact at DNA bases. This result may explain the fact that nuclease-ColE7 is a sugar and sequence nonspecific endonuclease.

Discussion

Nuclease Inhibition and DNA Substrate Length Requirement

The crystal structure of the nuclease-ColE7/Im7 complex (Ko et al., 1999) reveals that Im7 does not block the HNH endonuclease active site, instead it binds at a region distantly away from the active site. The structural comparison between the free form of nuclease-ColE7 and the nuclease-ColE7/Im7 complex further suggests that Im7 is not an allosteric inhibitor that induces backbone conformational changes in nuclease-ColE7 (Cheng et al., 2002). Thus, an intriguing question is how Im7 inhibits the DNase activity of ColE7, if it does not block the active site or distort the enzyme? The electrostatic potential surface of nuclease-ColE7 (Figure 5A) shows that the protein contacts the DNA primarily in a basic region near the P5, P6, and P7 phosphate groups, but far from the active site (at P4). Interestingly, this region is also the binding site for the acidic Im7 protein (see Figure 5B), indicating that Im7 inhibits the endonuclease activity by blocking the major substrate binding site. Biochemistry data indeed show that DNA cannot bind to nuclease-ColE7 in the presence of Im7 (Ku et al., 2002). It has been suggested that an inhibitor blocking the substrate binding site instead of the DNase active site may have an advantage in differentiating colicins with conserved active sites but more varied surfaces around their DNA binding sites (Kleanthous and Walker, 2001).

Another intriguing question raised in this study is why ColE7 does not cleave small DNA substrates of less than 8 bases. Most of the nonspecific endonucleases cleave mono- or dinucleotides, such as *Serratia* nuclease and DNase I (Friedhoff et al., 1996). The crystal structure of nuclease-ColE7/DNA complex also offers a possible explanation. Since the major DNA binding site is located in a region far from the active site, shorter DNA might bind preferably to the basic region in nuclease-ColE7 around the P5, P6, and P7 phosphate binding sites. Therefore, a short DNA might not bind DNA with sufficient affinity or may not be long enough to extend to the active site. Hence, for this DNase, a longer DNA ($>$7 bp) is required for binding and cleavage.

DNA Degradation Mechanisms

ColE7 forms a complex with Im7 after it is expressed; therefore, its cytotoxic endonuclease activity is inhibited in host cells. The crystal structures of the enzyme (nuclease-ColE7 [Cheng et al., 2002]), the enzyme inhibitor Im7 (Chak et al., 1996), the enzyme/inhibitor complex (Ko et al., 1999), and enzyme/substrate complex (this study) have all been resolved. These structures together demonstrate how ColE7 and Im7 form a tight and specific complex and how Im7 blocks the DNA binding site in nuclease-ColE7 to inhibit the enzyme from binding to substrates. The ColE7/Im7 heterodimer is released from the host cell to search for other bacteria with recognizable outer membrane cobalamin transporter, BtuB. After entering the periplasm of the target cell, ColE7 is cleaved to produce an enzyme containing only the cytotoxic endonuclease domain of ColE7, which is then traversed across the inner membrane and enters the cytosol of target cells (Liao et al., 2001).

To make nuclease-ColE7 an efficient enzyme for killing a cell, it degrades all nucleic acids that it encounters, including RNA and DNA without sequence specificity. The crystal structure of nuclease-ColE7/DNA complex supports the earlier suggestion (Sui et al., 2002) that His545 functions as the general base. First, His545 is located at the same position as the general base residue in the endonucleases which also contain a $\beta\beta\alpha$-metal active site, such as His98 in I-*Ppo*I (Galburt et al., 1999)

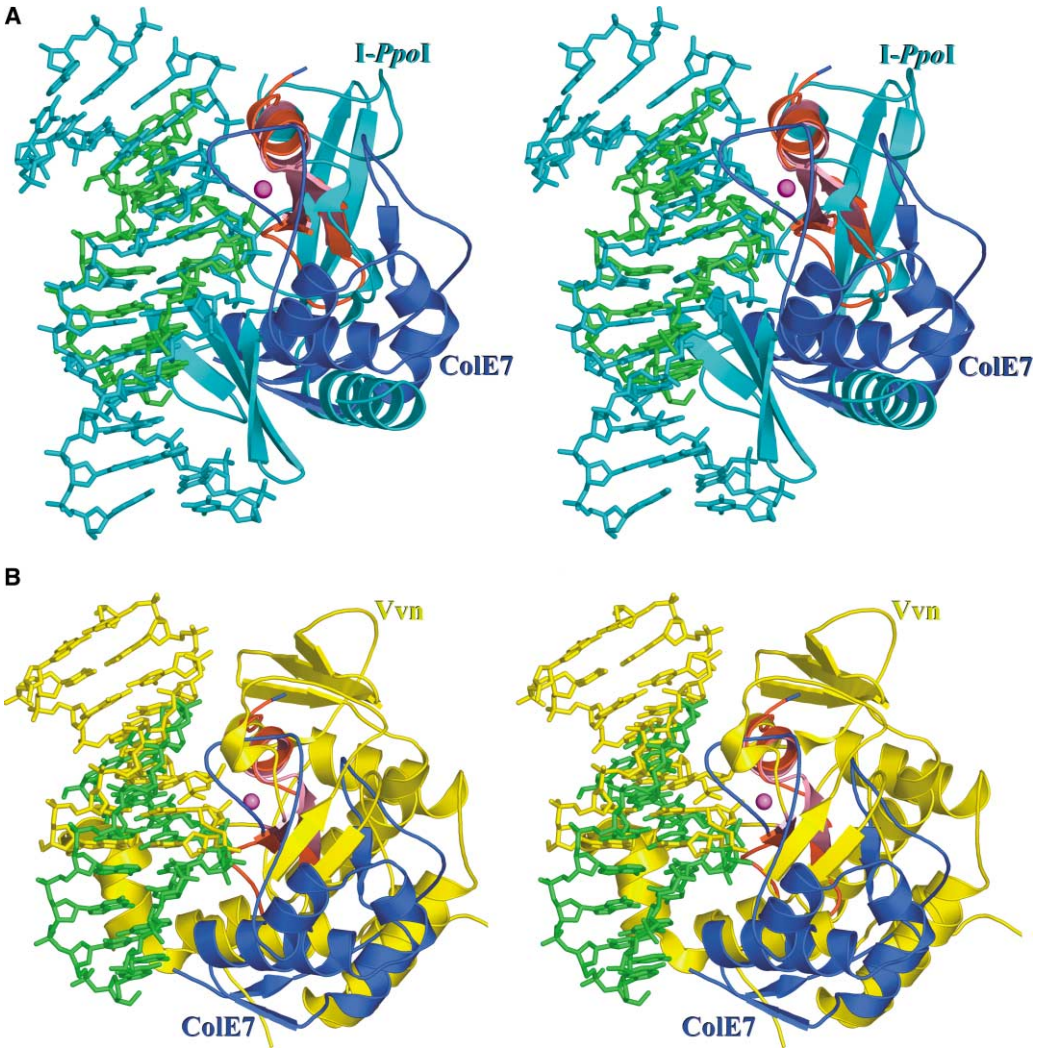

Figure 6. The Stereo View of the Superimposition of the HNH Motif of Nuclease-ColE7/DNA with the Active Site of I-*Ppo*I/DNA and Vvn/DNA, Respectively

(A) Only the HNH motif (in red) in nuclease-ColE7 and the ββα-metal motif (in pink) in I-*Ppo*I (pdb entry: 1A74) are used for superimposition. The average rms difference between the two active sites is 1.2 Å, as calculated over 17 Cα atoms. For clarity, for I-*Ppo*I/DNA complex, only one monomer in the I-*Ppo*I dimer and 15 (out of the 20) base pairs of the DNA duplex are shown in blue. The magnesium ion in the active site of I-*Ppo*I is shown as a pink ball. The relative orientations between the DNA duplex and the "ββα-metal" fold of active site in the two complexes are similar. The nuclease-ColE7 is in navy blue and its DNA is in green.

(B) The ββα-metal motif (in pink) in Vvn/DNA complex (in yellow, PDB entry: 1OUP) is superimposed with the HNH motif (in red) in ColE7 (in navy blue), resulting in an average rms difference of 0.9 Å, as calculated over 17 Cα atoms. The two DNA molecules fit with each other, showing again that the ββα-metal motif in these endonucleases interact with DNA in a similar mode.

and His80 in Vvn (Li et al., 2003). The distance between His545 (Nδ1) to P4 atom is 4.5 Å in nuclease-ColE7/DNA complex, close to the distances of 4.6 Å between the general base (His98) and scissile phosphate in I-*Ppo*I/DNA. Second, His545 is located in an ideal position for functioning as a general base, at the opposite site of the leaving 3′ oxygen atom. The nucleophilic water molecule is not observed in the complex structure, likely because an apo-enzyme without a metal ion cofactor but not a holo-enzyme is used for the structural study here. The complex structure also supports the suggestion that the zinc ion may bind to the scissile phosphate and stabilize the phosphoanion transition state during hydrolysis (Sui et al., 2002).

Comparison to the Structure of I-*Ppo*I/DNA and Vvn/DNA Complexes

The ββα-metal fold has been observed in several different families of endonucleases, within which only the complex crystal structures of I-*Ppo*I/DNA (Flick et al., 1998) and Vvn/DNA (Li et al., 2003) have been reported. I-PpoI binds DNA site specifically as a homodimer, whereas Vvn and nuclease-ColE7 bind DNA in a sequence-independent manner as a monomer. This leads to the question whether the ββα-metal motif in these proteins binds DNA in a similar mode, considering that they share no sequence identity but only structural homology in a small region in the active site. Superimposition of only the HNH motif in nuclease-ColE7/DNA, with

Table 1. Data Collection and Refinement Statistics

Data Collection and Processing	
Space group	$P2_1$
Cell dimensions	a = 59.4 Å, b = 46.5 Å, c =78.5 Å, α = 90.0°, β = 90.4°, γ = 90.0°
Resolution (Å)	2.5
Observed reflections	61,443
Unique reflections	15,097
Completeness—all data (%)	96.7
Completeness—last shell (%)	83.2
(resolution range, Å)	(2.58–2.49)
R_{sym}[a]—all data (%)	4.0
R_{sym}[a]—last shell (%)	23.6
$I/\sigma(I)$—all data (50.0–2.49 Å)	36.4
$I/\sigma(I)$—last shell (2.58–2.49 Å)	7.8
Refinement	
Resolution range (Å)	50.0–2.5
Reflections	14,389
Nonhydrogen atoms	
Protein	2,064
DNA	965
Solvent molecules	232
R factor (%)[b]	21.6
R_{free} (%)	28.9
Model Quality	
Rms deviations in	
Bond lengths (Å)	0.011
Bond angles (°)	1.3
Average B factor (Å^2)	
Protein atoms	44.8
DNA atoms	44.4
Solvent atoms	40.6
Ramachandran plot (%)	
Most favored	86.6
Additionally allowed	11.6
Generously allowed	1.4
Disallowed	0.5

[a] $R_{sym} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>|/I_i(hkl)$.
[b] R factor = $\Sigma_{hkl}||Fo(hkl)|-|Fc(hkl)||/\Sigma_{hkl}|Fo(hkl)|$.

the respective ββα-metal motif in I-*Ppo*I/DNA and Vvn/DNA gives an rms difference of 1.2 Å, as calculated over 17 Cα atoms in I-*Ppo*I, and 0.9 Å, as calculated over 17 Cα atoms in Vvn. This result shows that these small ββα regions indeed share high structural homology, even though the overall folds of the three proteins are very different. Remarkably, the DNA duplexes in these complexes all match well after superimposition, as shown in Figure 6. Although these endonucleases use different ways to interact with DNA at the consecutive major groove, that I-*Ppo*I uses β strands and nuclease-ColE7 and Vvn use α helices for interactions. However, it is noteworthy that not only do the DNA molecules bind in a similar orientation to the ββα-metal motif, but these DNA duplexes also deform in a comparable manner. By lining up the cleavage sites in the three protein-DNA complexes (see Figure 4B), the roll angles for the base steps at and one step before the scissile phosphate are all positive values of ~10° and the minor groove widths, immediately before the cleavage site, are all widened to ~9 Å in the three protein-DNA complexes. This amazing resemblance in DNA conformation indicates that all the ββα-metal motifs in these endonucleases not only bind DNA at minor grooves in a similar manner but also bend DNA to a similar extent as that of HNH ColE7.

Concluding Remarks

Here we show the recognition between a HNH protein and DNA. The HNH ColE7 binds DNA at the minor groove and contacts DNA primarily at phosphate groups. The structure provides a rare opportunity for the understanding of sequence-independent protein recognition of DNA at an atomic level. Remarkably the structure of the complex also explains the substrate length requirement and inhibition mechanism of the specific inhibitor protein. In summary, this structure provides a solid foundation for understanding how an endonuclease degrades DNA randomly in cells during cell death and a general DNA recognition mode for all endonucleases carrying a ββα-metal motif.

Experimental Procedures

Protein Overexpression and Purification

The expression vector pQE-70 (Qiagen, Germany) containing a 6 histidine affinity tag at the C terminus of the cloning site was used for the overexpression of nuclease-ColE7 and Im7 proteins in *E. coli* M15 (Sui et al., 2002). Cells were incubated in LB medium at 37°C and induced by IPTG for 4 hr once the $A_{600\ nm}$ reached 0.6 O.D. Crude cell extracts were loaded onto a Ni-NTA resin affinity column (Qiagen, Germany), and the bound protein was then eluted by an imidazole gradient solution. The eluent was dialyzed against 20 mM glycine-HCl buffer (pH 3.0) overnight to denature the protein complex. The resulting protein solution was loaded onto a Sepharose-SP column (HiTrap SP, Pharmacia) equilibrated with 20 mM glycine-HCl buffer (pH 3.0). The nuclease-ColE7 was eluted by a NaCl gradient (0–2.0 M) at pH 3.0 and Im7 was eluted afterward by a 20 mM sodium phosphate buffer at pH 7.0. The eluent containing nuclease-ColE7 was dialyzed against 20 mM sodium phosphate buffer (pH 7.0) and then applied to a heparin column (HiTrap SP, Pharmacia). The free nuclease-ColE7 (containing residues 444–576 without a 6 histidine tag) was then eluted by a NaCl gradient (0–1.0 M) and concentrated to 10 mg/ml in 50 mM Tris-HCl (pH 7.5).

Characterization of the Minimum Length of DNA Substrates

Various sizes of single-stranded DNAs were 5′ end labeled by T4 polynucleotide kinase (Promega) and [γ-^{32}P]ATP at 37°C for 20 min, and then the kinase acitivities were inactivated by heating the mixture at 75°C for 30 min. Free [γ-^{32}P]ATP was removed by MicroSpin G-25 columns (Amersham). The radiolabeled ssDNAs were cleaved by nuclease-ColE7 in 50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, and 5 mM DTT at 37°C overnight and the reactions stopped with a solution containing 95% (v/v) formamide, 10 mM EDTA, 0.1% (w/v) bromophenol blue, and 0.05% (w/v) xylene cyanol. These DNA samples were then analyzed by electrophoresis, using a 20% (19:1 [w/w] acrylamide to bisacrylamide) acrylamide gel.

Characterization of the ColE7 Nuclease Activity by a Zymogram Method

The RNase activity of nuclease-ColE7 was assayed using Baker's yeast (Worthington Biochemical) as the RNA substrate (12 μg/μl), which was first added to the separating gel and heated at 50°C for 5 min Before being loaded onto the gel, nuclease-ColE7 with or without His-taq, RNase A (positive control; Sigma), and DNase I (negative control; Worthington Biochemical) were denatured by a solution containing 125 mM Tris-HCl (pH 6.8), 2% SDS, 2.5% glycerol, and then incubated at 55°C for 30 min. After electrophoresis using a Bio-Rad Mini-Protean II Dual Slab Gel model apparatus, the SDS was extracted from the gel by soaking the gel in 25% (v/v) isopropanol and 10 mM Tris-HCl (pH 7.0). The isopropanol was then removed by soaking the gel in 10 mM Tris-HCl (pH 7.0) buffer, and the gel was further incubated in Tris-HCl buffers, containing 3 μM Zn^{2+} and 3 μM Mg^{2+} at 37°C for 60 min to recover the enzyme

activity. The reaction was stopped with 200 μl of stop solution (7% phosphoric acid and 0.1% uranyl acetate dihydrate) at 4°C and the free RNA products in the gel were removed by soaking the gel in 10 mM Tris-HCl buffer at pH 7. Finally, the gel was incubated in stain solution, containing 0.2% (w/v) toluidine blue O and 10 mM Tris-HCl, (pH7.0).

Crystallization and X-Ray Data Collection

Crystals of nuclease-ColE7/DNA were obtained by the hanging drop vapor-diffusion method. A drop containing 10 mg/ml of nuclease-ColE7 with equal molar ratio of the octa-nucleotide 5′-GCGATCGC-3′, 2.5 mM EDTA, 12.5 mM Tris-HCl (pH 7.5), 0.1 M ammonium formate, and 10% PEG 3350 was set up against a reservoir of 0.2 M ammonium formate and 20% PEG 3350. Plates of crystal appeared within a week at room temperature. Crystals were soaked in the reservoir solution containing an extra 20% glycerol, prior to being flash cooled for data collection by a Quantum 4 CCD detector at Taiwan beam line BL12B2 in SPring-8, Japan. The nuclease-ColE7/DNA crystallized in a monoclinic $P2_1$ space group with two proteins and three DNA duplex per asymmetric unit. Cell parameters and diffraction statistics are listed in Table 1.

Structure Determination and Refinement

The crystal structure of the nuclease-ColE7/DNA complex was determined by molecular replacement using the structure of nuclease-ColE7 (PDB accession number 7CEI) as the search model by the program CNS (Brunger et al., 1998). Two solutions were identified and structural refinement was carried out by the program CNS, with 10% of data selected randomly and set aside for the calculation of R_{free} values. After positional and B factor refinement of the two protein molecules, the R factor decreased to 39.6%. At this stage, the Fourier maps gave a clear continuous electron density for the three DNA octamers and the DNA structural models were built accordingly. The final complex model has an R factor of 21.6% and an R_{free} of 28.8% for 14,389 reflections in the resolution range of 40.0–2.5 Å. The final refinement statistics are listed in Table 1.

Acknowledgments

We thank Mau-Tsu Tang and Yi-Shan Huang for their help during data collection at BL12B2 at SPring-8. This work was supported by research grants from the Academia Sinica and National Science Council (NSC92-2321-B001-012) of the Republic of China to H.S.Y.

Received: October 8, 2003
Revised: October 21, 2003
Accepted: October 21, 2003
Published: February 10, 2004

References

Bewley, C.A., Gronenborn, A.M., and Clore, G.M. (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. Annu. Rev. Biophys. Biomol. Struct. *27*, 105–131.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N.S., et al. (1998). Crystallography and NMR System (CNS): a new software system for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

Chak, K.-F., Safo, M.K., Ku, W.-Y., Hsieh, S.-Y., and Yuan, H.S. (1996). The crystal structure of the ImmE7 protein suggests a possible colicin-interacting surface. Proc. Natl. Acad. Sci. USA *93*, 6437–6442.

Chang, M.C., Chang, S.Y., Chen, S.L., and Chuang, S.M. (1992). Cloning and expression in *Escherichia coli* of the gene encoding an extracellular deoxyribonuclease (DNase) from *Aeromonas hydrophila*. Gene *122*, 175–180.

Cheng, Y.-S., Hsia, K.-C., Doudeva, L.G., Chak, K.-F., and Yuan, H.S. (2002). The crystal structure of the nuclease domain of ColE7 suggests a mechanism for binding to double-stranded DNA by the H-N-H endonucleases. J. Mol. Biol. *324*, 227–236.

Chevalier, B.S., and Stoddard, B.L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. *29*, 3757–3774.

Drouin, M., Lucas, P., Otis, C., Lemieux, C., and Turmel, M. (2000). Biochemical characterization of I-*Cmo*eI reveals that this H-N-H homing endonuclease shares functional similarities with H-N-H colicins. Nucleic Acids Res. *28*, 4566–4572.

Eddy, S.R., and Gold, L. (1991). The phage T4 *nrd*B intron: a deletion mutant of a version found in the wild. Genes Dev. *5*, 1032–1041.

Eisen, J.A. (1998). A phylogenomic study of the MutS family of proteins. Nucleic Acids Res. *26*, 4291–4300.

Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A., and Nagata, S. (1998). A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature *391*, 42–50.

Ferat, J.L., and Michel, F. (1993). Group II self-splicing introns in bacteria. Nature *364*, 358–361.

Flick, K.E., Jurica, M.S., Monnat, R.J., and Stoddard, B.L. (1998). DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. Nature *394*, 96–101.

Foley, S., Bruttin, A., and Brussow, H. (2000). Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilius* bacteriophages. J. Virol. *74*, 611–618.

Friedhoff, P., Franke, I., Meiss, G., Wende, W., Krause, K.L., and Pingoud, A. (1999). A similar active site for non-specific and specific endonucleases. Nat. Struct. Biol. *6*, 112–113.

Friedhoff, P., Meiss, G., Kolmes, B., Pieper, U., Gimadutdinow, O., Urbanke, C., and Pingoud, A. (1996). Kinetic analysis of the cleavage of natural and synthetic substrates by the *Serratia* nuclease. Eur. J. Biochem. *241*, 572–580.

Galburt, E.A., Chevalier, B., Tang, W., Jurica, M.S., Flick, K.E., Monnat, R.J., and Stoddard, B.L. (1999). A novel endonuclease mechanism directly visualized for I-*Ppo*I. Nat. Struct. Biol. *6*, 1096–1099.

Goodrich-Blair, H. (1994). The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames. Nucleic Acids Res. *22*, 3715–3721.

Goodrich-Blair, H., and Shub, D.A. (1996). Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. Cell *84*, 211–221.

Goodrich-Blair, H., Scarlato, V., Gott, J.M., Xu, M.-Q., and Shub, D.A. (1990). A self-splicing group I intro in the DNA polymerase gene of *Bacillus subtilis* bacteriophage SPO1. Cell *63*, 417–424.

Gorbalenya, A.E. (1994). Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. Protein Sci. *3*, 1117–1120.

Hiom, K., and Sedgwick, S.G. (1991). Cloning and structural characterization of the *mcr*A locus of *Escherichia coli*. J. Bacteriol. *173*, 7368–7373.

James, R., Kleanthous, C., and Moore, G.R. (1996). The biology of E colicins: paradigms and paradoxes. Microbiol. *142*, 1569–1580.

Jones, S., Heyningen, P.V., Berman, H.M., and Thornton, J.M. (1999). Protein-DNA interactions: a structural analysis. J. Mol. Biol. *287*, 877–896.

Jurica, M.S., Monnat, R.J., and Stoddard, B.L. (1998). DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-*Cre*I. Mol. Cell *2*, 469–476.

Kleanthous, C., and Walker, D. (2001). Immunity proteins: enzyme inhibitors that avoid the active site. Trends Biochem. Sci. *26*, 624–631.

Kleanthous, C., Kuhlmann, U.C., Pommer, A.J., Ferguson, N., Radford, S.E., Moore, G.R., James, R., and Hemmings, A.M. (1999). Structural and mechanistic basis of immunity toward endonuclease colicins. Nat. Struct. Biol. *6*, 243–252.

Ko, T.-P., Liao, C.-C., Ku, W.-Y., Chak, K.-F., and Yuan, H.S. (1999). The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure *7*, 91–102.

Ku, W.-Y., Liu, Y.-W., Hsu, Y.-C., Liao, C.-C., Liang, P.-H., Yuan, H.S., and Chak, K.-F. (2002). The zinc ion in the HNH motif of the endonuclease domain of colicin E7 is not required for DNA binding but is essential for DNA hydrolysis. Nucleic Acids Res. *30*, 1670–1678.

Kuck, U. (1989). The intron of a plasmid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. Mol. Gen. Genet. *218*, 257–265.

Lambowitz, A.M., and Belfort, M. (1993). Introns as mobile genetic elements. Annu. Rev. Biochem. *62*, 587–622.

Landthaler, M., Begley, U., Lau, N.C., and Shub, D.A. (2002). Two self-splicing group I introns in the ribonucleotide reductase large gene of *Staphylococcus aureus* phage Twort. Nucleic Acids Res. *30*, 1935–1943.

Lavery, R., and Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dyn. *6*, 63–91.

Lazarevic, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C., and Karamata, D. (1998). Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus substitls* temperate bacteriophage SPβ. Proc. Natl. Acad. Sci. USA *95*, 1692–1697.

Li, L.Y., Luo, X., and Wang, X. (2001). Endonuclease G is an apoptotic DNase when released from mitochondria. Nature *412*, 95–99.

Li, C.L., Hor, L.-I., Chang, Z.-F., Tsai, L.-C., Yang, W.-Z., and Yuan, H.S. (2003). DNA binding and cleavage by the periplasmic nuclease Vvn: a novel structure with a known active site. EMBO J. *22*, 4014–4025.

Liao, C.-C., Hsia, K.-C., Liu, Y.-W., Liang, P.H., Yuan, H.S., and Chak, K.-F. (2001). Processing of DNase domain during translocation of Colicin E7 across the membrane of *Escherichia coli*. Biochem. Biophys. Res. Commun. *284*, 556–562.

Liu, X., Zou, H., Slaoghter, C., and Wang, X. (1997). DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. Cell *89*, 175–184.

Luscombe, N.M., and Thornton, J.M. (2002). Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J. Mol. Biol. *320*, 991–1009.

Michel-Briand, Y., and Baysse, C. (2002). The pyocins of *Pesudomonas aeruginosa*. Biochemie *84*, 499–510.

Miller, M.D., Tanner, J., Alpaugh, M., Benedik, M.J., and Krause, K.L. (1994). 2.1 Angstrom structure of *Serratia* endonuclease suggests a mechanism for binding to double-stranded DNA. Nat. Struct. Biol. *1*, 461–468.

Miller, M.D., Cai, J., and Krause, K.L. (1999). The active site of *Serratia* endonuclease contains a conserved magnesium-water cluster. J. Mol. Biol. *288*, 975–987.

Moure, C.M., Gimble, F.S., and Quiocho, F.A. (2002). Crystal structure of the intein homing endonuclease PI-*Sce*I bound to its recognition sequence. Nat. Struct. Biol. *9*, 764–770.

Nadassy, K., Wodak, S.J., and Janin, J. (1999). Structural features of protein-nucleic acid recognition sites. Biochemistry *38*, 1999–2017.

Parrish, J.Z., and Xue, D. (2003). Functional genomic analysis of apoptotic DNA degradation in *C. elegans*. Mol. Cell *11*, 987–996.

Piekarowicz, A., Yuan, R., and Stein, D.C. (1991). Isolation of temperature-sensitive McrA and McrB mutations and complementation analysis of the McrBC region of *Escherichia coli* K-12. J. Bacteriol. *173*, 150–155.

Pommer, A.J., Kuhlmann, U.C., Cooper, A., Hemmings, A.M., Moore, G.R., James, R., and Kleanthous, C. (1999). Homing in on the role of transition metals in the HNH motif of colicin endonuclease. J. Biol. Chem. *274*, 27153–27160.

Pommer, A.J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C., Cooper, A., Connolly, B.A., Hemmings, A.M., Moore, G.R., et al. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. J. Mol. Biol. *314*, 735–749.

Raaijmakers, H., Vix, O., Toro, I., Golz, S., Kemper, B., and Suck, D. (1999). X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. EMBO J. *18*, 1447–1458.

Sano, Y., Matsui, H., Kobayashi, M., and Kageyama, M. (1993). Molecular structures and functions of Pyocins S1 and S2 in *Pseudomonas aeruginosa*. J. Bacteriol. *175*, 2907–2916.

Schaller, K., and Nomura, M. (1976). Colicin E2 is a DNA endonuclease. Proc. Natl. Acad. Sci. USA *68*, 3989–3993.

Shub, D.A., Goodrich-Blair, H., and Eddy, S.R. (1994). Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem. Sci. *19*, 402–404.

Sui, M.-J., Tsai, L.-C., Hsia, K.-C., Doudeva, L.-G., Chak, K.-F., and Yuan, H.S. (2002). Metal ions and phosphate binding in the H-N-H motif: crystal structures of the nuclease domain of ColE7/Im7 in complex with a phosphate ion and different divalent metal ions. Protein Sci. *11*, 2947–2957.

Van Roey, P., Waddling, C.A., Fox, K.M., Belfort, M., and Derbyshire, V. (2001). Interwined structure of the DNA-binding domain of intron endonuclese I-*Tev*I with its substrate. EMBO J. *20*, 3631–3637.

Van Roey, P., Meehan, L., Kowalski, J., Belfort, M., and Derbyshire, V. (2002). Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-*Tev*I. Nat. Struc. Biol. *9*, 806–811.

Walker, D.C., Georgiou, T., Pommer, A.J., Walker, D., Moore, G.R., Kleanthous, C., and James, R. (2002). Mutagenic scan of the H-N-H motif of colicin E9: implications for the mechanistic enzymology of colicins, homing enzymes and apoptotic endonucleases. Nucleic Acids Res. *30*, 3225–3234.

Zhang, J., and Xu, M. (2002). Apaptotic DNA fragmentation and tissue homeostasis. Trends Cell Biol. *12*, 84–89.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., and Lambowitz, A.M. (1995). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell *83*, 529–538.

Accession Numbers

Structural coordinates have been deposited in the Protein Data Bank (accession code 1PT3).