Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural basis for the DNA-binding activity of the bacterial β-propeller protein YncE

Wataru Kagawa,[a] Tomohiko Sagawa,[a] Hironori Niki[b] and Hitoshi Kurumizaka[a]*

[a]Laboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan, and [b]Microbial Genetics Laboratory, Genetic Strains Research Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan

Correspondence e-mail: kurumizaka@waseda.jp

Received 2 September 2011
Accepted 27 October 2011

PDB References: $YncE_{31-353}$, 3vgz; $YncE_{31-353}$–DNA complex, 3vh0.

β-Propellers are widely utilized in nature as recognition modules. The well conserved β-propeller fold exhibits a high degree of functional diversity, which is probably accomplished through variations in the surface properties of the proteins. Little is known about the interactions between β-propeller proteins and nucleic acids. In the present study, it has been found that the bacterial β-propeller protein YncE binds to DNA. Crystal structures of YncE in the free form and complexed with DNA revealed that the surface region of YncE corresponding to the 'canonical' substrate-binding site forms essential contacts with DNA. A single DNA base within a single-stranded DNA region is trapped in the hydrophobic pocket located within the central channel of the β-propeller protein. These data provide physical evidence for the DNA-binding ability of the previously uncharacterized YncE and also suggest that the 'canonical' substrate-binding site may be commonly adapted to facilitate nucleic acid binding in a subset of β-propeller proteins.

1. Introduction

β-Propeller proteins are present in all domains of life and form several large families that are based on specific sequence patterns. Their diverse functions provide platforms for interactions with a wide range of biomolecules, including proteins, ions, sugars, low-molecular-weight compounds and nucleic acids (Stirnimann *et al.*, 2010). Despite the diverse functions displayed by β-propeller proteins, their structures are highly similar, consisting of four or more four-stranded β-sheets arranged in a circle (Fülöp & Jones, 1999; Stirnimann *et al.*, 2010). The surface properties of β-propeller proteins are considered to have diversified to perform a variety of functions. Consequently, surface analysis has been used to identify new β-propeller proteins with a particular function and to classify β-propeller proteins into groups of related function that otherwise would not be predicted as being functionally homologous by sequence analysis (Fülöp & Jones, 1999; Smith *et al.*, 1999).

Previous structural studies of β-propeller proteins bound to their substrates revealed a 'canonical' substrate-binding site, which is the narrow side of the funnel-shaped protein. A wealth of structural information is available for low-molecular-weight and peptide substrates bound to this region. In contrast, only a few studies have described the detailed interactions of β-propeller proteins with their nucleic acid

substrates. We wondered whether β-propeller proteins that interact with nucleic acids share similar surface properties. If so, then comparative structural studies of several β-propeller protein–DNA complexes may help to illuminate the common structural features in the protein–DNA interactions and may also identify new nucleic acid-binding β-propeller proteins. We unexpectedly identified a β-propeller protein called YncE as a protein from an *Escherichia coli* cell extract that binds to a DNA column (H. Niki, unpublished data). YncE is a highly basic protein (pI 9.2) and shares sequence similarity with the YVTN family. It has an N-terminal signal sequence for potential periplasmic localization (Baars *et al.*, 2006). A previous genetic study indicated that YncE participates in iron metabolism (McHugh *et al.*, 2003) and a structural study was initiated (Baba-Dikwa *et al.*, 2008). However, the precise function of YncE is unknown.

In the present study, we found that YncE has DNA-binding activity. Crystal structures of YncE in the free form and in a complex with DNA were solved. The structures revealed that the surface region of YncE corresponding to the 'canonical' substrate-binding site is highly positively charged and binds to single-stranded regions of DNA by accommodating a single DNA base in its hydrophobic pocket, which is located within the central channel. The DNA-recognition features of YncE are similar to those of the DNA-repair factor DDB2, suggesting that the 'canonical' substrate-binding site may be commonly utilized among nucleic acid-binding β-propeller proteins that interact with unpaired DNA bases.

2. Experimental procedures

2.1. Protein expression and purification

For structural studies, the *E. coli* $YncE_{31–353}$ gene was cloned into the pET-21a vector (Novagen). A stop codon was inserted at the 3′ end of the gene, allowing $YncE_{31–353}$ to be expressed without a C-terminal hexahistidine tag. $YncE_{31–353}$ was expressed in *E. coli* strain BL21 (DE3). All procedures after cell harvesting were performed at 277 K. For purification, 2 l LB culture was incubated at 310 K and protein expression was induced at an optical density (A_{600}) of 0.6 with 0.1 m*M* isopropyl β-D-1-thiogalactopyranoside (final concentration). After overnight induction at 303 K, the cells were harvested, resuspended in buffer *A* (50 m*M* MOPS–NaOH pH 7.0, 1 m*M* EDTA) containing 500 m*M* NaCl and 1 m*M* PMSF and lysed by sonication. The cell lysate was cleared of insoluble material by centrifugation at 15 000 rev min^{-1} for 25 min (Beckman Type JA-20 rotor). Nucleic acids in the cell lysate were removed by adding polyethyleneimine [5%(*v*/*v*), pH 7.5] to a final concentration of 0.1%, followed by gentle mixing at 277 K for 30 min and centrifugation at 10 000 rev min^{-1} for 10 min (Beckman Type JA-20 rotor). The resulting supernatant was subjected to two rounds of ammonium sulfate precipitation, the first with 55% saturation and the second with 85% saturation. The precipitant from the second fractionation, which contained $YncE_{31–353}$, was washed with 95% saturated ammonium sulfate solution to remove the residual polyethyleneimine and was dissolved in buffer *A* containing 50 m*M* NaCl. The $YncE_{31–353}$-containing fraction was then applied onto a Sephadex G-25M gel-filtration column (GE Healthcare) using the same buffer. The elution fractions containing $YncE_{31–353}$ were combined and loaded onto a 5 ml HiTrap SP HP column (GE Healthcare). After washing to remove the unbound proteins, $YncE_{31–353}$ was eluted using a linear gradient of 50–500 m*M* NaCl in buffer *A*. The peak fractions were collected, combined and applied onto a HiLoad Superdex 200 pg (26/60) column (GE Healthcare) equilibrated with buffer *B* (20 m*M* MOPS–NaOH pH 7.0, 0.1 m*M* EDTA). The $YncE_{31–353}$ fractions were then loaded onto a MonoS 5/50 GL column (GE Healthcare) and eluted using a stepwise gradient of 0, 0.2 and 1 *M* NaCl in buffer *B*. Peak fractions were collected and dialyzed against either buffer *B* containing 1 m*M* DTT and 200 m*M* NaCl for crystallization without DNA or buffer *C* (10 m*M* HEPES–NaOH pH 7.5, 50 m*M* NaCl) for cocrystallization with DNA. Purified proteins were stored at 277 K. Selenomethionine-substituted $YncE_{31–353}$ was expressed in *E. coli* B834 (DE3) cells cultured in M9 medium. Protein expression was induced at an optical density (A_{600}) of 0.6 with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (final concentration). After overnight induction at 303 K, the derivatized protein was purified using the same protocol as used for the native protein.

For biochemical studies, *Strep*-tag II (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) was fused to the N-termini of $YncE_{31–353}$ and its alanine point mutants to allow rapid purification. Similar to the production of $YncE_{31–353}$ used in the structural studies, the *Strep*-tag II-fused proteins were expressed using the pET-21a vector in *E. coli* strain BL21 (DE3) and lacked a C-terminal hexahistidine tag. All procedures after cell harvesting were performed at 277 K. For purification, a 100 ml LB culture was incubated at 310 K and protein expression was induced at an optical density (A_{600}) of 0.6 with 1 m*M* isopropyl β-D-1-thiogalactopyranoside (final concentration). After overnight induction at 303 K, the cells were harvested, lysed and cleared of insoluble material and nucleic acids following the same procedures as described for the $YncE_{31–353}$ used for structural studies. The recovered supernatant was applied onto a 0.5 ml *Strep*-Tactin Sepharose column and the unbound proteins were removed by a wash step with 5 ml buffer *A* containing 1 m*M* DTT and 500 m*M* NaCl. The *Strep*-tag II-fused proteins were eluted using buffer *A* containing 1 m*M* DTT, 500 m*M* NaCl and 2.5 m*M* D-desthiobiotin. The peak fractions were collected, dialyzed against buffer *D* (20 m*M* HEPES–NaOH pH 7.5, 0.1 m*M* EDTA, 1 m*M* DTT, 200 m*M* NaCl) and stored at 277 K. All protein concentrations were determined using a Bradford protein assay kit with bovine serum albumin (Nacalai Tesque) as the standard.

2.2. Crystallization and data collection

Crystals of selenomethionine-substituted $YncE_{31–353}$ were grown at 293 K using the hanging-drop vapour-diffusion method by mixing 1 µl concentrated protein (25 mg ml^{-1}) and 1 µl reservoir solution consisting of 0.1 *M* sodium acetate pH

4.4, 0.2 *M* ammonium sulfate and 25% PEG 4000. Crystals typically appeared after one week and grew to dimensions of 0.1 × 0.3 × 0.7 mm. For harvesting, a cryoprotectant solution consisting of 0.1 *M* sodium acetate pH 4.4, 0.2 *M* ammonium sulfate, 25% PEG 4000 and 20% ethylene glycol was directly added to the drop. The crystals were immediately captured in a loop, flash-cooled in a stream of N_2 gas (∼100 K) and stored in liquid nitrogen. Multiwavelength anomalous dispersion (MAD) data were collected at three wavelengths (0.97902, 0.97944 and 1.00000 Å) on the SPring-8 BL26B2 beamline. To reduce the overlap of diffraction spots, the oscillation range was set to 0.3°.

The $YncE_{31–353}$–DNA complex was prepared by mixing equal volumes of $YncE_{31–353}$ solution (0.3 m*M*) and an 11-nucleotide single-stranded DNA (5′-CGGGTACTCAG-3′) solution (0.6 m*M*) and incubating the mixture at 310 K for 10 min. Crystals of the complex were grown at 293 K using the hanging-drop vapour-diffusion method by mixing 1 µl of the protein–DNA mixture and 1 µl of a reservoir solution consisting of 0.1 *M* trisodium citrate pH 5.6, 2% Tacsimate pH 5.0 and 16% PEG 3350. The crystals were harvested by adding approximately 60 µl of a cryoprotectant solution consisting of 0.1 *M* trisodium citrate pH 5.6, 2% Tacsimate pH 5.0, 16% PEG 3350 and 20% ethylene glycol to the drop. The crystals were immediately captured in a loop, flash-cooled in a stream of N_2 gas (∼100 K) and stored in liquid nitrogen. Diffraction data were collected on the BL41XU beamline of SPring-8.

Table 1
Statistics for crystallographic structure determination.

Values in parentheses are for the highest resolution shell.

	YncE			
Data set	SeMet (peak)	SeMet (edge)	SeMet (remote)	YncE–DNA
Data collection				
Space group	$C222_1$			$I4_1$
Unit-cell parameters (Å, °)	a = 119.17, b = 139.31, c = 173.69, $\alpha = \beta = \gamma = 90$			$a = b$ = 171.17, c = 177.22, $\alpha = \beta = \gamma = 90$
Wavelength (Å)	0.97902	0.97944	1.00000	1.00000
Resolution (Å)	50–1.7	50–1.7	50–1.7	50–2.9
Reflections (measured/unique)	3324589/157493	3411711/157707	3487701/157624	828530/56581
R_{merge}† (%)	7.2 (32.7)	7.0 (45.4)	7.2 (60.7)	6.4 (37.7)
Mean $I/\sigma(I)$	12.6 (5.9)	11.1 (4.2)	10.1 (2.7)	13.3 (6.3)
Completeness (%)	99.9 (100)	99.9 (100)	99.9 (99.5)	100 (100)
Multiplicity	7.1 (7.1)	7.1 (7.0)	7.1 (6.4)	7.8 (7.8)
Refinement				
Resolution (Å)	50–1.7			50–2.9
R factor/free R factor‡ (%)	22.0/24.4			21.0/23.4
B factors (Å^2)				
Protein	26.3			62.3
DNA	—			159
Water	27.6			41.6
R.m.s. deviations				
Bond lengths (Å)	0.005			0.007
Bond angles (°)	1.41			1.40
Ramachandran plot, residues in (%)				
Most favourable regions	85.1			80.3
Additionally allowed regions	14.4			19.0
Generously allowed regions	0.5			0.7

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl)\rangle$ is the mean value for that reflection. ‡ R factor = $\sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. The free R factor was calculated with 5% of the data excluded from the refinement.

Data sets were processed and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997). The crystals of $YncE_{31–353}$ and the $YncE_{31–353}$–DNA complex belonged to space groups $C222_1$ and $I4_1$, respectively (Table 1).

2.3. Structure determination and refinement

The structure of selenomethionine-derivatized $YncE_{31–353}$ was determined using MAD data collected from a single crystal. The selenium sites were found with *SHELXD* (Sheldrick, 2008) using all three data sets. The phase data calculated using *SHELXD* were converted to mtz format with the *CCP*4 program suite (Winn *et al.*, 2011), followed by phase extension, density modification and automated model building using *RESOLVE* (Terwilliger, 2003). Although the resulting model was incomplete, the β-propeller structure was readily apparent. In the crystal, four $YncE_{31–353}$ molecules are present in the asymmetric unit. The partial model was subjected to iterative rounds of rigid-body, energy-minimization and *B*-factor refinement using *CNS* (Brünger *et al.*, 1998) and model building using *Coot* (Emsley *et al.*, 2010). The structure of the $YncE_{31–353}$–DNA complex was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007), using the coordinates of the $YncE_{31–353}$ structure as a search model. In the crystal, four YncE and two DNA molecules were present in the asymmetric unit cell. Prime-and-switch density modification was performed using *RESOLVE*, starting with the phases calculated by *Phaser*. Rigid-body, energy-minimization and *B*-factor refinement of the molecular-replacement solution using *CNS* yielded clearly interpretable electron densities for the DNA. Iterative rounds of model building using *Coot* and energy-minimization and *B*-factor refinement using *CNS* were performed. We note that the lower final R factors of the complex compared with those of the DNA-unbound structure of $YncE_{31–353}$ suggest potential model bias, although care was taken to minimize the bias (prime-and-switch density modification and resetting the B factors to 30.0 Å^2 prior to rigid-body refinement).

All structure figures were created using *PyMOL* (http://www.pymol.org). The atomic coordinates of $YncE_{31–353}$ and the $YncE_{31–353}$–DNA complex have been deposited in the Protein Data Bank under accession codes 3vgz and 3vh0, respectively.

2.4. Oligonucleotides used in DNA-binding assays

The following 55-mer oligonucleotides were used in the competitive DNA-binding assay. Oligo-1 (5′-GGC GCT TTA ACT TCC AAG ATA TCG TTC CTT TTA GAC AGG TCT TCG CCA TAT CTT C-3′) was used as the ssDNA substrate. The dsDNA was prepared by annealing Oligo-2 (5′-GGC GCT TTA ACT TCC ATT TTT GCG GGT ACT CAG CAA AAT TCT TCG CCA TAT CTT C-3′) and Oligo-3 (5′-GAA GAT ATG GCG AAG AAT TTT GCT GAG TAC CCG CAA AAA TGG AAG TTA AAG CGC C-3′), and the cruciform DNA was prepared by annealing Oligo-1 and Oligo-4 (5′-GAA GAT ATG GCG AAG TTA AAA CGA CTC ATG GGC GTT TTT AGG AAG TTA AAG CGC C-3′). The resulting cruciform DNA contains 15-base-pair and ten-base-pair homologous regions. For hairpin DNA-binding assays, hairpin DNA substrates containing a 20-base-pair duplex region were used. To prepare the hairpin substrates, Oligo-5 (3-mer stem loop, 5′-TTG GAT CCT GGA ACA GGA TCC AA-3′), Oligo-6 (4-mer stem loop, 5′-TTG GAT CCT GTA CTC AGG ATC CAA-3′), Oligo-7 (5-mer stem loop, 5′-TTG GAT CCT GGT ACT CAG GAT CCA A-3′), Oligo-8 (7-mer stem loop, 5′-TTG GAT CCT GCG TAC TCC AGG ATC CAA-3′) and Oligo-9 (9-mer stem loop, 5′-TTG GAT CCT GGC GTA CTC ACA GGA TCC AA-3′) were separately heated at 373 K for 10 min and the DNA was then quickly chilled in ice water.

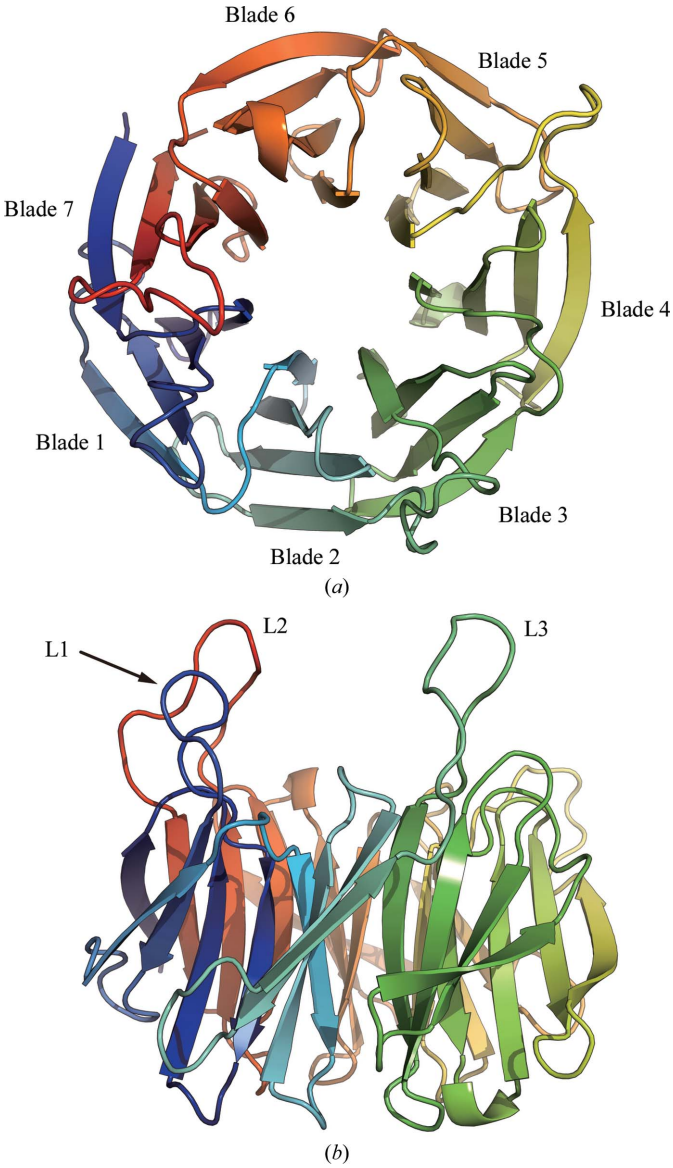

Figure 1
Crystal structure of *E. coli* $YncE_{31–353}$. (*a*) $YncE_{31–353}$ viewed down the central channel. The protein adopts a seven-bladed β-propeller structure. (*b*) $YncE_{31–353}$ viewed from the side. The face of the β-propeller ring commonly used for substrate binding corresponds to the top surface of the $YncE_{31–353}$ structure. It includes three characteristic loops (L1, L2 and L3).

2.5. DNA-binding assays

For competitive DNA-binding assays, the reaction mixture contained final concentrations of 10 m*M* HEPES–NaOH pH 7.5, 150 m*M* NaCl, 11 µ*M* of each DNA substrate and 0.5–8 µ*M* $YncE_{31–353}$. The mixture was incubated at 310 K for 10 min. After adding sixfold loading dye (15% Ficoll, 0.1% bromophenol blue, 0.1% xylene cyanol), the reaction mixtures were subjected to 10% polyacrylamide gel electrophoresis in 0.5× TB (45 m*M* Tris, 45 m*M* boric acid) buffer. For hairpin DNA-binding assays, the reaction mixture contained final concentrations of 20 m*M* sodium citrate pH 5.6, 50 m*M* NaCl, 0.5 µ*M* hairpin DNA and 0.5–4 µ*M* $YncE_{31–353}$. The mixture was incubated at 310 K for 10 min. After adding sixfold loading dye (15% Ficoll, 0.1% bromophenol blue, 0.1% xylene cyanol), the reaction mixtures were subjected to 3.5% polyacrylamide gel electrophoresis in 0.5× TBE buffer. For DNA-binding assays using $YncE_{31–353}$ alanine mutants, the reaction mixture contained final concentrations of 10 m*M* HEPES–NaOH pH 7.5, 150 m*M* NaCl, 0.2 µ*M* of the cruciform DNA used in the competitive DNA-binding assays and 0.5–4 µ*M* $YncE_{31–353}$ or its alanine mutants. The mixture was incubated at 310 K for 10 min. After adding sixfold loading dye (15% Ficoll, 0.1% bromophenol blue, 0.1% xylene cyanol), the reaction mixtures were subjected to 6% polyacrylamide gel electrophoresis in 0.5× TB buffer. Products for all assays were visualized by SYBR Gold (Invitrogen) staining of the gels.

3. Results

3.1. Crystal structure of YncE

The *E. coli* YncE protein contains an N-terminal signal sequence for periplasmic export. Full-length YncE prepared as a recombinant protein was susceptible to proteolytic degradation, and mass spectrometry and amino-acid sequencing analyses of the proteolytic fragments revealed that the degradation primarily occurred within the signal sequence (data not shown). We subsequently prepared a YncE construct lacking the signal sequence ($YncE_{31–353}$). This construct was resistant to degradation and crystallized in space group $C222_1$ with four chains in the asymmetric unit. The crystal structure of $YncE_{31–353}$ was solved by the multiwavelength

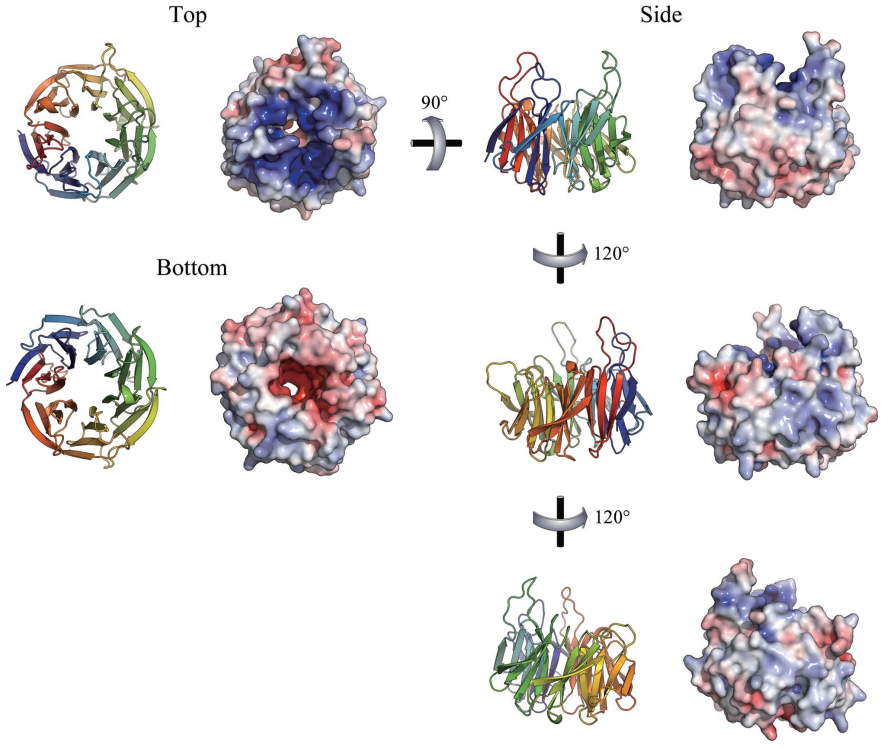

Figure 2
Solvent-accessible surface of $YncE_{31–353}$ coloured according to the electrostatic potential from $-6k_BTe^{-1}$ (red) to $+6k_BTe^{-1}$ (blue), where k_B is Boltzmann's constant, T is temperature (K) and e is the charge of an electron.

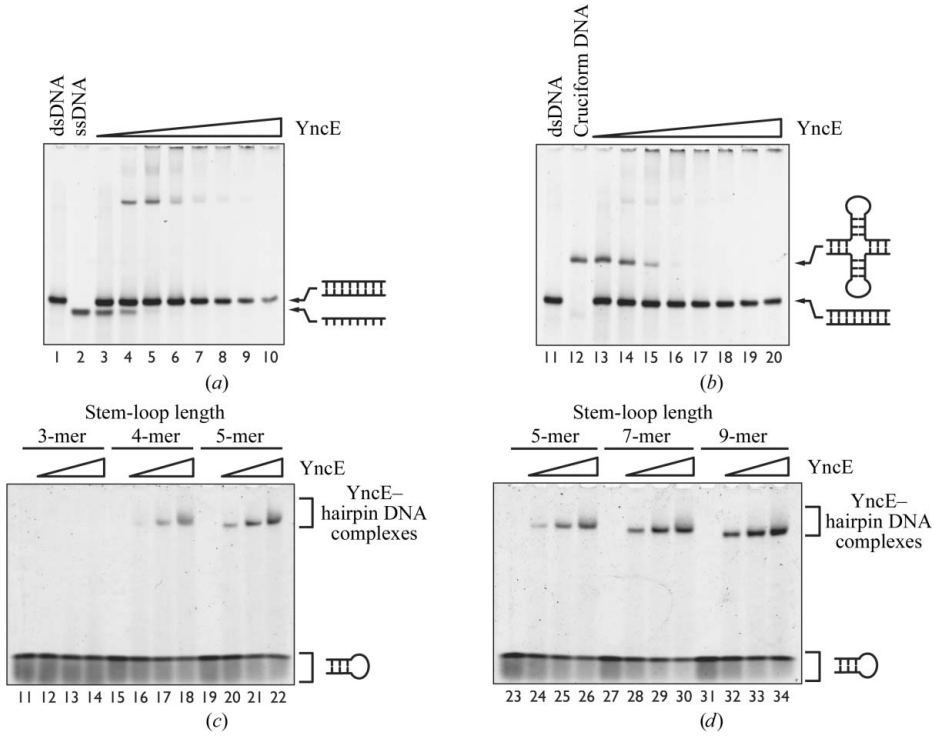

Figure 3
DNA-binding activity of $YncE_{31–353}$. (*a*) Preferential binding of $YncE_{31–353}$ to single-stranded DNA. Increasing amounts of $YncE_{31–353}$ (0, 0.5, 1, 2, 3, 4, 6 and 8 μ*M*) were incubated with a mixture containing equimolar amounts of ssDNA and dsDNA for 10 min at 310 K and then fractionated through a 10% polyacrylamide gel (lanes 3–10). ssDNA and dsDNA were separately fractionated in lanes 1 and 2, respectively. (*b*) Preferential binding of $YncE_{31–353}$ to cruciform DNA. Increasing amounts of $YncE_{31–353}$ (0, 0.5, 1, 2, 3, 4, 6 and 8 μ*M*) were incubated with a mixture containing equimolar amounts of cruciform DNA and dsDNA for 10 min at 310 K and then fractionated through a 10% polyacrylamide gel (lanes 13–20). Cruciform DNA and dsDNA were separately fractionated in lanes 11 and 12, respectively. (*c*, *d*) $YncE_{31–353}$ binds to the stem-loop region of hairpin DNA. Increasing amounts of $YncE_{31–353}$ (0.5, 1, 2 and 4 μ*M*) were incubated with hairpin DNAs with varying lengths (three, four, five, seven or nine nucleotides) of the stem-loop region for 10 min at 310 K. Products were fractionated through a 3.5% polyacrylamide gel.

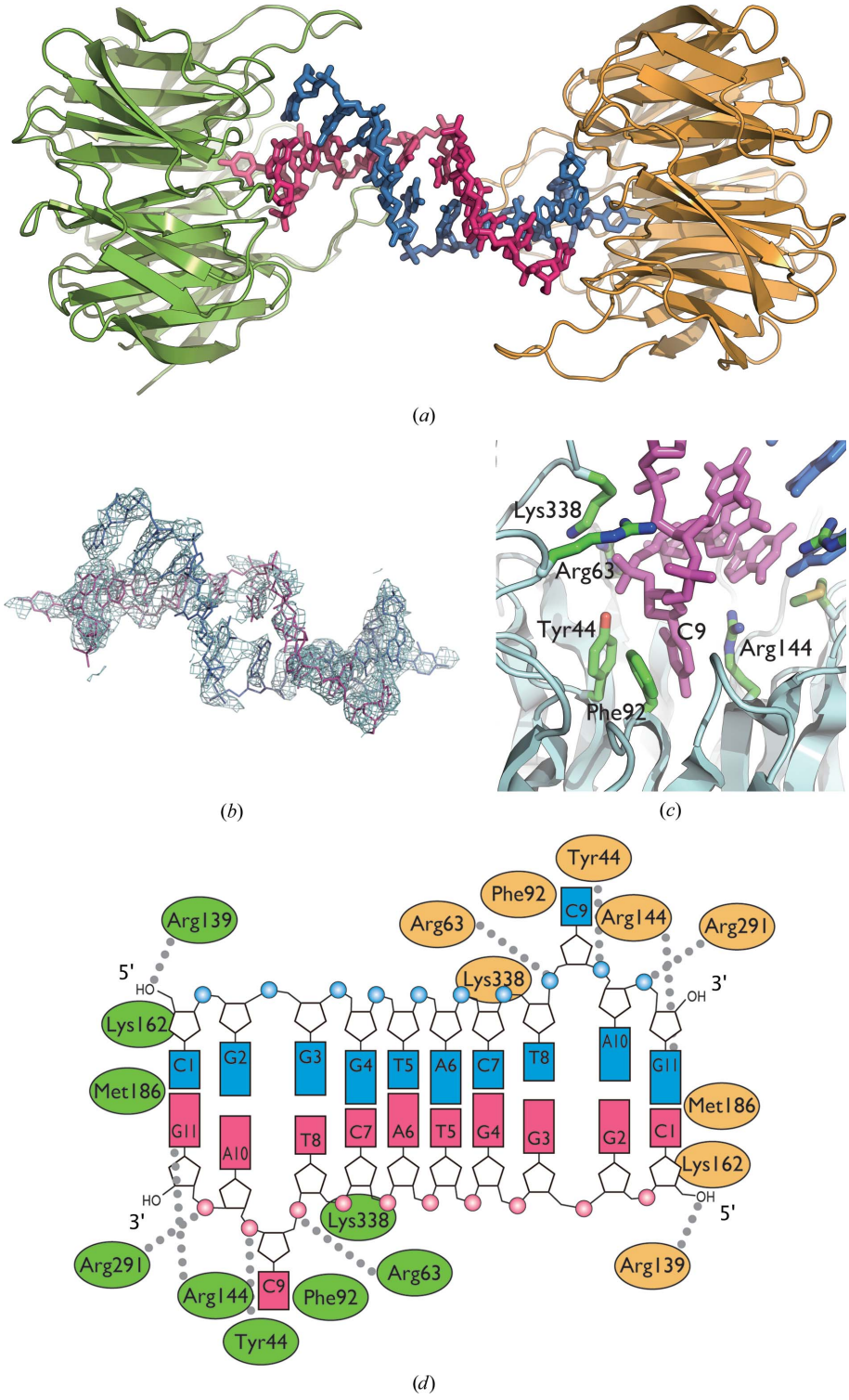

Figure 4
Crystal structure of the $YncE_{31–353}$–ssDNA complex. (*a*) Structure of the complex in the asymmetric unit. Two $YncE_{31–353}$ protein molecules (coloured green and orange) are bound to two ssDNA strands, shown in stick representation (coloured magenta and blue). The ssDNA strands form a partial duplex near the central region of the sequence. (*b*) $|F_o| - |F_c|$ map of the ssDNA strands shown in (*a*), which was calculated prior to model building. The map is superimposed with the final ssDNA model, shown in stick representation (coloured magenta and blue). (*c*) Close-up view of the flipped base (C9) and the $YncE_{31–353}$ residues that directly interact with the ssDNA. The ssDNA is coloured magenta and the interacting residues are coloured green. (*d*) Schematic interactions between $YncE_{31–353}$ and ssDNA. Colours correspond to those shown in (*a*). The grey dotted lines indicate hydrogen bonds or polar contacts.

anomalous diffraction method using selenomethionine-substituted protein. The final model was refined to 1.7 Å resolution with good stereochemistry and no Ramachandran outliers. Data-collection, phasing and structure-refinement statistics are presented in Table 1.

The structure revealed that $YncE_{31–353}$ adopts a commonly observed β-propeller structure, with seven four-stranded β-sheets arranged in a closed circle (Fig. 1*a*). A structural homology search using the *DALI* server (Holm & Rosenström, 2010) identified nearly 200 nonredundant PDB structures with essentially identical β-propeller architectures. The highest structural similarities were found with the human WDR5 protein (PDB entry 2h6n, *Z* score = 34.5; Ruthenburg *et al.*, 2006), the archaeal surface-layer protein (SLP; PDB entry 1l0q, *Z* score = 33.4; Jing *et al.*, 2002) and the bacterial virginiamycin B lyase (Vgb; PDB entry 2z2o, *Z* score = 32.8; Korczynska *et al.*, 2007). Despite the high level of similarity amongst the structural homologues of YncE identified by the *DALI* server, the loop regions connecting the β-strands vary in both their lengths and conformations. YncE has three relatively long loops: L1 ($YncE_{60–68}$) in blade 1, L2 ($YncE_{334–345}$) in blade 7 and L3 ($YncE_{129–141}$), which connects blades 2 and 3 (Fig. 1*b*). All three loops protrude from the same face of the β-propeller structure, which corresponds to the face commonly used for substrate binding by many β-propeller proteins. Thus, it is highly likely that the characteristic loops L1, L2 and L3 play an important role in YncE function.

3.2. YncE has DNA-binding activity

To determine the functional role of YncE, we inspected the crystal structure for surface regions that might perform its function. The electrostatic surface of YncE revealed a highly positively charged region near the central channel of the protein that is surrounded by the characteristic loops (Fig. 2, top left). The other regions of YncE are mostly neutral, suggesting that YncE may interact with a negatively charged

molecule through its positively charged surface. Given that protein surfaces involved in DNA binding are usually positively charged, we investigated whether YncE interacts with DNA. Using single-stranded and double-stranded oligonucleotides, we performed electrophoretic mobility-shift assays (EMSAs) to assess whether YncE binds to DNA. YncE was mixed with equimolar amounts of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA). Increasing the YncE concentration resulted in a decrease in the unbound ssDNA and dsDNA. A predominant complex was formed at low protein:DNA ratios and higher order complexes that were not resolved by the gel were also generated (Fig. 3*a*). These observations indicate that YncE binds to both ssDNA and dsDNA but has higher affinity towards ssDNA.

To further characterize the DNA-binding activity of YncE, we examined whether YncE interacts with unpaired DNA bases, which may be the molecular basis for the higher affinity towards ssDNA than dsDNA. We predicted that if YncE exhibits high affinity towards unpaired DNA bases then it would bind to the loop region of a cruciform DNA and would show a binding preference for cruciform DNA over dsDNA in a competitive EMSA using equimolar amounts of these substrates. Both DNA substrates were formed by annealing two 55-mer oligonucleotides, and the cruciform DNA contained a five-base loop region at the middle of the top and bottom strands (Fig. 3*b*). Similar to the EMSA performed using ssDNA and dsDNA, increasing the YncE concentration resulted in a decrease in the unbound cruciform DNA and dsDNA and the formation of higher order complexes that were not resolved by the gel (Fig. 3*a*). We found that YncE had a higher affinity for the cruciform DNA. To gain further evidence that YncE interacts with the loop region of the cruciform DNA, EMSAs using hairpin DNA substrates with various loop-region lengths were performed. Shortening the loop region of the hairpin DNA clearly decreased the level of YncE binding (Fig. 3*c*), whereas elongating the loop region increased the binding (Fig. 3*d*). Thus, these observations suggest that YncE recognizes single-stranded regions in DNA by interacting with bases that are unpaired and accessible. Moreover, YncE formed specific complexes with hairpin DNA, suggesting that hairpin DNA may be a more physiologically relevant substrate than either ssDNA or cruciform DNA.

3.3. Structure of the YncE–DNA complex

To understand the molecular details of the YncE–DNA interactions, we crystallized $YncE_{31–353}$ bound to ssDNA. ssDNAs with various lengths and sequences were screened to obtain crystals of the complex. We found that an 11-mer ssDNA with the sequence 5′-CGGGTACTCAG-3′ yielded cocrystals that diffracted X-rays to 2.9 Å resolution. The structure was solved by the molecular-replacement method using the substrate-free $YncE_{31–353}$ structure as a search model (Table 1). The asymmetric unit of the crystal contained four $YncE_{31–353}$ proteins, with two ssDNA strands bound to two of the $YncE_{31–353}$ proteins (Fig. 4*a*). The electron density of the ssDNA was clearly visible in the $|F_o| - |F_c|$ map that was

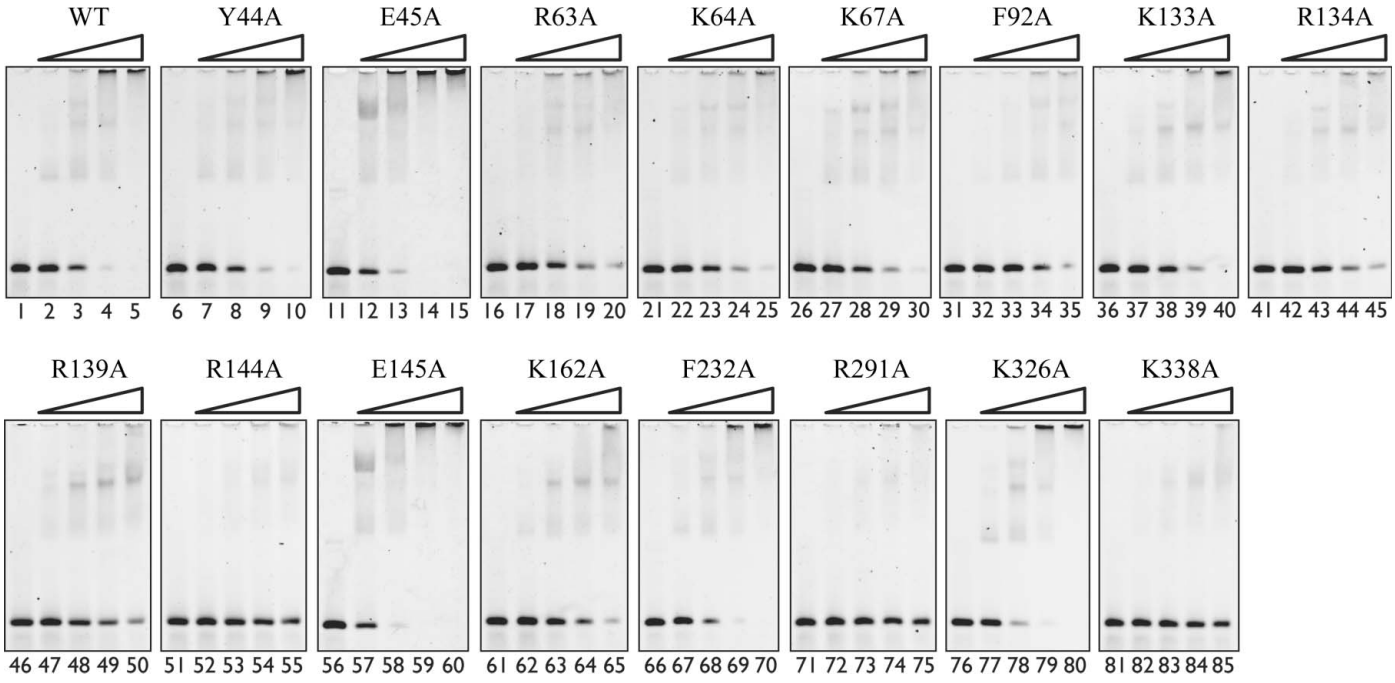

Figure 5
DNA binding by $YncE_{31–353}$ and its alanine point mutants. Arg, Lys, Tyr, Phe and Glu residues located on the positively charged surface of $YncE_{31–353}$ were mutated to alanine and the abilities of the mutants to bind to cruciform DNA were examined by EMSA. Increasing amounts of $YncE_{31–353}$ and its alanine point mutants (0.5, 1, 2 and 4 μ*M*) were incubated with the cruciform DNA used in the competitive assay for 10 min at 310 K (lanes 2–5, 7–10, 12–15, 17–20, 22–25, 27–30, 32–35, 37–40, 42–45, 47–50, 52–55, 57–60, 62–65, 67–70, 72–75, 77–80 and 82–85). Cruciform DNA without protein was fractionated in lanes 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 66, 71, 76 and 81.

generated after rigid-body, energy-minimization and *B*-factor refinement of the molecular-replacement solution in the absence of an ssDNA model (Fig. 4*b*). Interestingly, model building of the ssDNA revealed that the ssDNA annealed into a pseudo-duplex along almost the entire length of the DNA (Figs. 4*a* and 4*b*). The pseudo-duplex is anchored to YncE at either end by a bona fide Watson–Crick C1·G11 base pair, with internal mismatches formed by the G2·A10 and G3·T8 pairs (Fig. 4*d*).

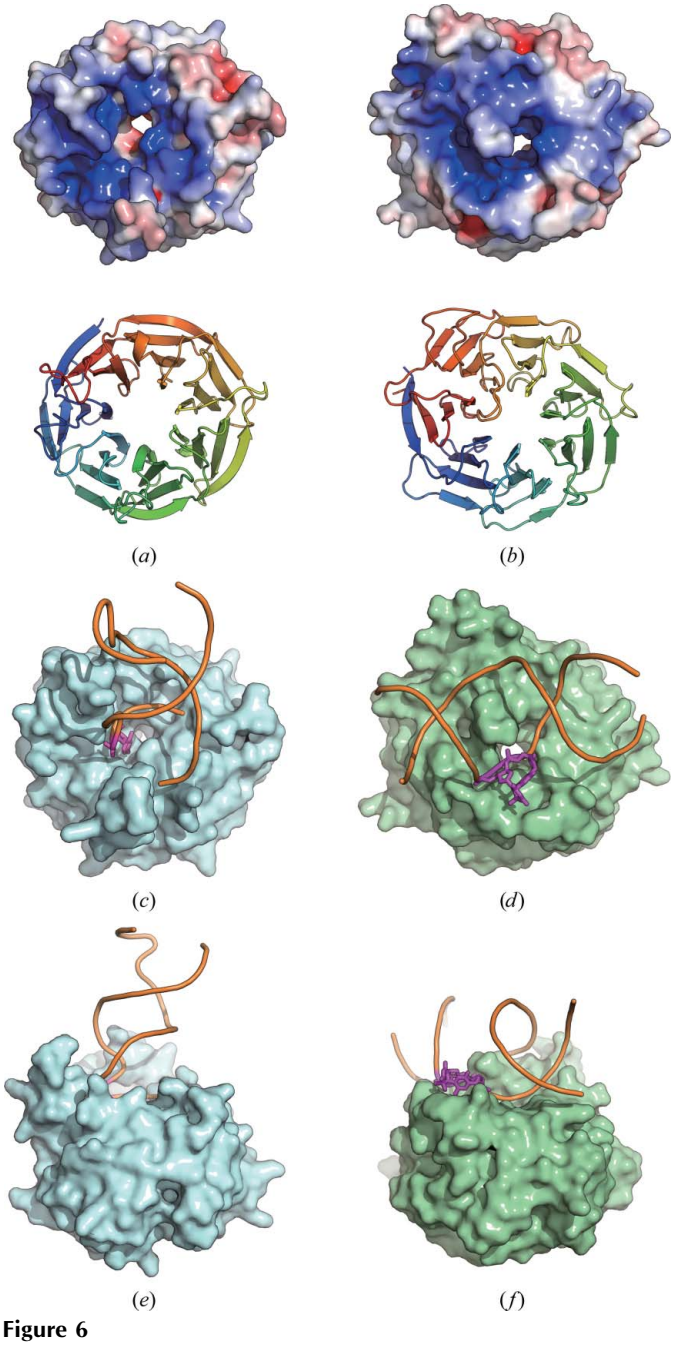

Figure 6
Comparison of the DNA-binding modes between YncE and DDB2. Solvent-accessible surfaces of $YncE_{31–353}$ (*a*) and $DDB2_{136–455}$ (PDB entry 3ei1) (*b*) coloured according to the electrostatic potential from $-6k_BTe^{-1}$ (red) to $+6k_BTe^{-1}$ (blue), where k_B is Boltzmann's constant, T is temperature (K) and e is the charge of an electron. Both proteins are viewed down the central channel. Cartoon representations of the proteins are shown below for clarity. (*c*, *d*) The flipped DNA bases (coloured magenta) lie on the 'canonical' substrate-binding surfaces of $YncE_{31–353}$ (*c*) and $DDB2_{136–455}$ (*d*). (*e*) The helical axis of the pseudo-duplex is nearly orthogonal to the DNA-binding surface of YncE. (*f*) The helical axis of the DNA bound to DDB2 runs parallel to the DNA-binding surface.

More importantly, the structure revealed that YncE interacts with an unpaired DNA base (C9), which is consistent with the results of the DNA-binding assays (Figs. 4*c* and 4*d*). The base of C9 is flipped out of the partial duplex DNA structure and is bound to the central channel of $YncE_{31–353}$ (Figs. 4*a* and 4*c*). Hydrophobic interactions occur between the base of C9 and the Arg144, Tyr44 and Phe92 residues of $YncE_{31–353}$. The shape and conformation of the nucleic acid-binding pocket formed by these residues indicate that YncE readily accepts and binds to DNA duplexes that contain an unpaired base. Furthermore, Arg144 also forms hydrogen bonds to the terminal guanine base (G11), which is the only side chain–base interaction in the structure (Fig. 4*d*). YncE also makes specific stacking interactions with the terminal base pair (C1·G11) *via* Met186 and Lys162 (Fig. 4*d*). These interactions suggest that YncE may possess a DNA end-binding activity.

To confirm that the $YncE_{31–353}$–ssDNA interactions in the crystal structure are not unique to the oligonucleotide used for crystallization, alanine-scanning mutagenesis was performed on the amino-acid residues located on the positively charged surface of $YncE_{31–353}$ and the ability of the alanine mutants to bind to a cruciform DNA was observed by EMSA (Fig. 5). Mutations in Arg144, Arg291 and Lys338 nearly abolished DNA binding, as observed from the lack of shifting of the unbound DNA substrate. Furthermore, mutations in Arg63 (in L1), Phe92, Arg139 (in L3) and Lys162 mildly affected DNA binding. All of these residues directly interact with the DNA in the crystal structure and therefore the results of the mutagenesis experiments are consistent with the crystal structure.

4. Discussion

A comparison of the DNA-binding modes between YncE and DDB2, a nucleotide-excision repair factor that recognizes ultraviolet-induced 6–4 pyrimidine–pyrimidone photoproducts in double-stranded DNA (Scrima *et al.*, 2008), reveals interesting similarities and differences. Both proteins utilize the 'canonical' substrate-binding site, which is the narrow side of the funnel-shaped β-propeller structure. This region is positively charged in both YncE and DDB2 (Figs. 6*a* and 6*b*) and is responsible for interactions with unpaired DNA bases (Figs. 6*c* and 6*d*). In the case of YncE a single base is trapped within its hydrophobic pocket, which is located in the central channel. In the case of DDB2 the 6–4 pyrimidine–pyrimidone is flipped out from the DNA duplex by the insertion of a hairpin-loop region into the minor-groove side of the lesion. DDB2 makes specific interactions with the flipped-out 6–4 pyrimidine–pyrimidone near the central channel. On the other hand, YncE and DDB2 exhibit differences in their DNA-binding modes. The major difference is the orientation of the DNA (Figs. 6*e* and 6*f*). In the case of DDB2 the DNA axis is

orthogonal to the central channel of the β-propeller structure. In contrast, the helical axis of the pseudo-duplex bound to YncE is nearly parallel to the central channel of YncE. This edge-on binding to YncE is novel and thus the YncE–DNA complex structure revealed in this study may provide an additional mechanism to be considered for future studies of β-propeller–nucleic acid interactions.

We thank Kengo Saito for critical reading of the manuscript. We also thank Sam-Yong Park (Yokohama City University, Japan) for advice on structure determination of $YncE_{31–353}$, Isao Sakane and Yoko Someya for technical assistance and the beamline scientists of the BL41XU and BL26B2 beamlines at SPring-8 for their assistance with data collection. This work was supported in part by Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology, Japan. HK is a Research Fellow of the Waseda Research Institute of Science and Engineering.

References

Baars, L., Ytterberg, A. J., Drew, D., Wagner, S., Thilo, C., van Wijk, K. J. & de Gier, J. W. (2006). *J. Biol. Chem.* **281**, 10024–10034.
Baba-Dikwa, A., Thompson, D., Spencer, N. J., Andrews, S. C. & Watson, K. A. (2008). *Acta Cryst.* F**64**, 966–969.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fülöp, V. & Jones, D. T. (1999). *Curr. Opin. Struct. Biol.* **9**, 715–721.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Jing, H., Takagi, J., Liu, L., Lindgren, S., Zhang, R., Joachimiak, A., Wang, J. & Springer, T. A. (2002). *Structure*, **10**, 1453–1464.
Korczynska, M., Mukhtar, T. A., Wright, G. D. & Berghuis, A. M. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 10388–10393.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
McHugh, J. P., Rodríguez-Quinoñes, F., Abdul-Tehrani, H., Svistunenko, D. A., Poole, R. K., Cooper, C. E. & Andrews, S. C. (2003). *J. Biol. Chem.* **278**, 29478–29486.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Ruthenburg, A. J., Wang, W., Graybosch, D. M., Li, H., Allis, C. D., Patel, D. J. & Verdine, G. L. (2006). *Nature Struct. Mol. Biol.* **13**, 704–712.
Scrima, A., Konícková, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N. P. & Thomä, N. H. (2008). *Cell*, **135**, 1213–1223.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. (1999). *Trends Biochem. Sci.* **24**, 181–185.
Stirnimann, C. U., Petsalaki, E., Russell, R. B. & Müller, C. W. (2010). *Trends Biochem. Sci.* **35**, 565–574.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.