Crystal Structure of YdcE Protein From *Bacillus subtilis*

Arhonda Gogos,1 Haiyan Mu,2 Fabiana Bahna,2 Carlos A. Gomez,2 and Lawrence Shapiro1,2,3

1Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, New York
2Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, New York
3Naomi Berrie Diabetes Center, Columbia University College of Physicians and Surgeons, New York, New York

**Introduction.** Addiction modules, which consist of two genes encoding a toxin and an antitoxin, mediate plasmid maintenance in *Escherichia coli* by selectively killing plasmid-free cells (postsegregational killing).1 They are most often encoded on low–copy number plasmids. The operons controlling these modules are autoregulated at the transcription level by their corresponding toxin–antitoxin complex. Maintenance of this complex prevents the lethal effect of the toxin. Toxin activation in plasmid-free cells is a result of higher decay rates for the antitoxins, which are substrates for cellular proteases. Two structures of toxins involved in plasmid maintenance have recently been reported: Kid from *E. coli* plasmid R12 and CcdB from *E. coli* plasmid F.3 Despite low sequence identity, these toxins have similar three dimensional structures.

Pairs of genes homologous to addiction modules have also been found in the *E. coli* chromosome in two different loci, chpA/mazEF and chpB.4,5 It has been proposed that these modules are involved in programmed cell death6 that can be triggered by antibiotics,7 and that they may be part of a cellular response to nutritional stress by regulating the synthesis of macromolecules.8 Occurrence of similar systems in other bacteria is suggested by sequence analysis.8,9

Sequences of chromosomally encoded proteins homologous to the mazF toxin from *E. coli* are classified as members of a Cluster of Orthologous Groups of proteins (COG2337) as defined in the National Center for Biotechnology Information (NCBI) database.10,11 COG2337 includes representatives from *Bacillus subtilis*, Mycobacterium tuberculosis, Chlamydia pneumoniae, Xylella fastidiosa, Neisseria meningitidis, and Deinococcus radiodurans, that have not been functionally characterized but are annotated as “growth inhibitors.” Here, we present the three-dimensional (3D) structure of YdcE from *B. subtilis*, the first from COG2337.

**Methods.** We amplified the ydcE gene by polymerase chain reaction (PCR) from *B. subtilis* genomic DNA using primers (5′-GGCCGCCATTTGTTTGAACCGG-GCGATGT-3′) and (5′-GCCGCGAGCTTCTAACAATCTGAAGTGGC-3′). These primers incorporated 5′ BamHI and 3′ HindIII sites that were subsequently used to clone the PCR product into the pSMT3 expression vector,12 which encodes a 6His-Sumo N-terminal tag. Selenomethionine (Se-Met)-substituted protein was expressed in *E. coli* BL21(DE3) cells and purified to homogeneity by affinity chromatography and gel filtration. The tag was removed with protease Ulpl, leaving an N-terminal serine. The protein was concentrated to 8.9 mg/ml in 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 5 mM dithiothreitol (DTT). Crystals were obtained by vapor diffusion in 2 μL hanging drops that contained equal volumes of protein and 12% polyethylene glycol (PEG) 4000, 0.1 M sodium acetate, pH 4.6, 0.2 M ammonium acetate at 22°C.

The crystals belong to space group P61,22 with a = 56.63 Å, b = 56.63 Å, c = 138.257 Å, with 1 protein molecule per asymmetric unit. The crystals were flash-frozen at 100 K in the well solution supplemented with 30% glycerol. Data were collected at beamline X9A of the National Synchrotron Light Source (NSLS), and processed and merged with the HKL program suite.13 Se positions were located with SOLVE,14 and the first model was built with Resolve.15 Waters were added with Arp_waters (ARP/WARP version 5.0).16 and refinement was performed with Refmac 5.017 from the CCP4 program suite.18 Figures were made with the programs SETOR19 and GRASP.20 Coordinates have been deposited in the Protein Data Bank (PDB accession code: 1NE8).

**Results and Discussion.** We determined the structure of YdcE from *B. subtilis* using phases derived from a three-wavelength multiwavelength anomalous diffraction (MAD) experiment performed on a single crystal of Se-MET–substituted protein. The final model, refined to 2.1 Å with an R factor of 15.9%, contains all 117 amino acids. It is a compact single-domain α/β protein consisting of 3
α-helices and seven β-strands [Fig. 1(A)]. Five of these strands (β1, β2, β3, β6, β7) form an antiparallel β-sheet, whereas strands β4, β5, and the C-terminus of β3 form another, smaller sheet. The structure of YdcE reveals a substantial dimer interface between monomers related by a crystallographic two-fold axis. Dynamic light scattering and gel filtration studies suggest that YdcE is also a dimer in solution (data not shown).

The dimer has a convex surface that is capped by the loops between strands β1 and β2, and a flat surface that includes helix α3, with protruding C-terminal tails. The extensive hydrophobic interface between the two monomers includes residues Ile30, Ile43, Ile111, Leu107, Ile80, and Leu114. Strands β6 from each monomer pair with each other through hydrogen bonds between the amide of Thr82 and the carbonyl oxygen of Ile80. Dimer interactions on the convex side include hydrogen bonds between the amides of Ser19 to the sidechains of Asp84 from each monomer and salt bridges between Glu20 and Arg87. Between these salt bridges, Arg81 from each monomer are buried in the dimer interface and stabilized by water-mediated hydrogen bonds. Other dimer interactions include hydrogen bonds between the carbonyl oxygen of Ser110 and the amide of Asn32, the carbonyl oxygen of Ala112 and Ne of Arg5. Kid toxin and CcdB toxin have a fold that is similar to that of YdcE, which includes a five-stranded antiparallel β-sheet, a smaller, three-stranded β-sheet, and a C-terminal α-helix. YdcE shares 27% sequence identity with Kid and 7% with CcdB. The dimer interface is very hydrophobic in all these structures and includes many interactions that are conserved between Kid and YdcE. In contrast to YdcE, both of these proteins have been subjected to biologic characterization. Kid toxin has been shown to form a complex with Kis antitoxin at a 1:1 ratio, and the replicative helicase DnaB has been implicated as its target. The ccdB gene encodes a protein that acts on the A subunits of gyrase (GyrA) and inhibits its activity. In the presence of CcdA antitoxin the action of CcdB is inhibited by the formation of a tight complex. If there exists a cognate antitoxin for YdcE, it remains to be identified.

We compared the electrostatic surface potentials of the three proteins using GRASP. The flat (C-terminal helix–containing) face of YdcE has a surface potential that is significantly more negative than the corresponding faces of Kid and CcdB [Fig. 1(B)]. This is mainly due to the differences in the amino acid content of the C-terminal helix. YdcE has six charged amino acids: Asp96, Asp97, Glu98, Asp101, Asp104, and Glu105. Based on our structural alignment, the corresponding residues in Kid include one charged amino acid (Pro94, Glu95, Thr96, Asn99, Leu102, and Gly103, respectively) and two charged amino acids in CcdB (Glu87, Asn88, Asp89, Asn92, Asn95, and Leu96, respectively). YdcE has one more charged amino acid on its C-terminal tail (Asp115). This tail is not present in Kid and is a few amino acids shorter in CcdB. Genetic analysis and model interactions between CcdB and GyrA indicate that this protein surface, and especially the C-terminus, is involved in the toxin’s interaction with its target. Sequence alignments of chromosomal and plasmid-
encoded homologs indicate that the C-terminal region that corresponds to the α-helix in the solved structures is highly variable. This variability might reflect differences in substrate specificity within this protein family.

ACKNOWLEDGMENTS.

We gratefully acknowledge Thirumuruhan Radhakan-nan at beamline X9A of the National Synchrotron Light Source (NSLS) for assistance with data collection.

REFERENCES