PROTEINS: Structure, Function, and Genetics 53:320–322 (2003)

STRUCTURE NOTE

Crystal Structure of YdcE Protein From *Bacillus subtilis*

Arhonda Gogos,[1] Haiyan Mu,[2] Fabiana Bahna,[2] Carlos A. Gomez,[2] and Lawrence Shapiro[1,2,3]
[1]*Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, New York*
[2]*Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, New York*
[3]*Naomi 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 R1[2] 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 death[6] 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′-GGCCGGGGATCCTTGATTGTGAAACGCG-GCGATGT-3′) and (5′-GCCGCGAAGCTTCTACTA-AAAATCAATGAGTGCCAAACT-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 m*M* Tris-HCl, pH 8.0, 150 m*M* NaCl, and 5 m*M* dithiothneitol (DTT). Crystals were obtained by vapor diffusion in 2 μL hanging drops that contained equal volumes of protein and 12% polyethelene 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 was $P6_522$ 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 5.0[17] from the CCP4 program suite.[18] Figures were made with the programs SETOR[19] and GRASP.[20] Coordinates have been deposited in the Protein Data Bank (PDB accession code: 1NE8).

Results and Discussion. We determined the structure of YcdE 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

*Correspondence to: Lawrence Shapiro, Department of Biochemistry and Molecular Biophysics, Columbia University, 630 West 168th Street, New York, NY 10032. E-mail: shapiro@convex.hhmi.columbia.edu

Grant sponsor: New York Structural Genomics Research Consortium; Grant number: NIH S/G IP50 GM62529.

Received 28 February 2003; Accepted 6 March 2003

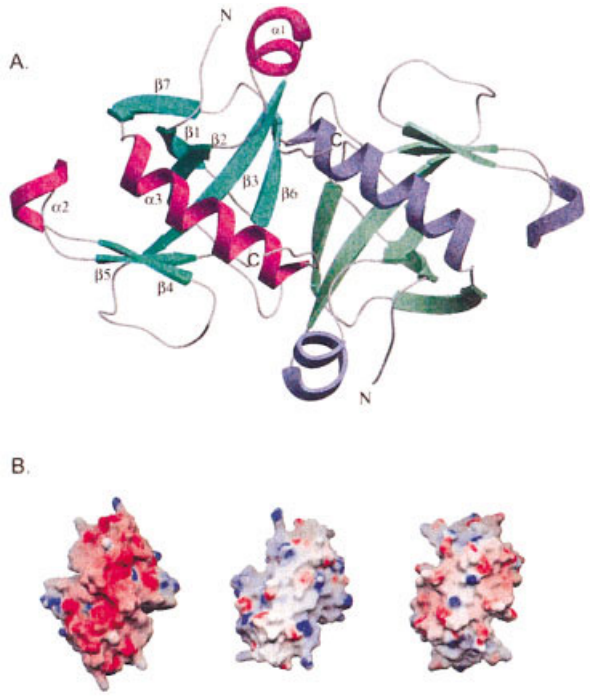

Fig. 1. (**A**) Ribbon diagram of *B. subtilis* YdcE protein dimer with color-coded secondary structure elements: α-helices are magenta and purple for the two monomers, and β-strands dark- and light-green, respectively. The NH_2- and COOH-termini are indicated. (**B**) Comparison of the electrostatic potential on the C-terminal helix–containing surface of YdcE, Kid, and CcdB dimers (left to right) calculated with the program GRASP.[20] The charge distribution is color-coded with blue for positive (≥15 kT/electron) and red for negative (≤15 kT/electron).

TABLE I. Statistics From the Crystallographic Analysis

Diffraction Data Statistics			
Data set	Peak	Edge	Remote
Wavelength (Å)	0.9790	0.97938	0.97163
Resolution (Å)	30.0–2.1	30.0–2.1	30.0–2.1
Measured reflections	80,833	44,102	54,751
Unique reflections	8244	8231	8246
Completeness[a]	99.5% (98.5%)	99.2% (97.3%)	99.4% (99.1%)
R_{sym}[b]	0.054 (0.198)	0.046 (0.217)	0.091 (0.165)
$\langle I \rangle / \langle \sigma I \rangle$	19.6	18.9	19.6
Refinement Statistics			
Resolution range		20.86–2.1	
Number of reflections (observed)		8194	
Number of reflections (R_{free})		380	
R-factor/R_{free}		0.159/0.21	
RMSD bond lengths		0.02	
RMSD bond angles		1.64	

[a]Completeness for the highest resolution shell in parentheses.
[b]R_{sym} for the highest resolution shell in parentheses.
$R_{sym} = I - \Sigma|I - \langle I \rangle|/\Sigma I$, where I is observed intensity and $\langle I \rangle$ is average intensity. $R_{cryst} = 100 \times \Sigma\|F_{obs}\| - \|F_{calc}\|/\Sigma|F_{obs}|$, where F_{obs} is the observed structure factors and F_{calc} is the calculated structure factors. The crystallographic *R* factor, R_{cryst}, is based on 95% of the data used in refinement, and the free *R* factor, R_{free}, is based on 5% of the data withheld for the cross-validation test. RMSD, root-mean-square deviation. Over 90% of the main chain dihedrals fall within the "most favored regions" of the Ramachandran plot.[24]

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 Nε of Arg5. Kid toxin and CcdB toxin have a fold that is similar to that of YdcE,[2,3] 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.[21] The ccdB gene encodes a protein that acts on the A subunits of gyrase (GyrA) and inhibits its activity.[22,23] 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.[20] 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.[3] 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 Radhakannan at beamline X9A of the National Synchrotron Light Source (NSLS) for assistance with data collection.

REFERENCES

1. Jensen RB, Gerdes K. Programmed cell death in bacteria: Proteic plasmid stabilization systems. Mol Microbiol 1995;17:205–210.
2. Hargreaves D. Santos-Sierra S, Giraldo R, Sabariegos-Jareno R, de la Cueva-Mendez G, Boelens R, Diaz-Orejas R, Rafferty J. Structural and functional analysis of the kid toxin protein from *E. coli* plasmid R1. Structure 2002;10:1425–1433.
3. Loris R, Dao-Thi MH, Bahassi EM, Van Melderen L, Poortmans F, Liddington R, Couturier M, Wyns L. Crystal structure of CcdB, a topoisomerase poison from *E. coli.* J Mol Biol 1999;285:1667–1677.
4. Aizenman E, Engelberg-Kulka H, Glaser G. An *Escherichia coli* chromosomal "addiction module" regulated by guanosine 3′,5′-bispyrophosphate: a model for programmed bacterial cell death. Proc Natl Acad Sci U S A 1996;93:6059–6063.
5. Masuda Y, Miyakawa K, Nishimura Y, Ohtsubo E. chpA and chpB, *Escherichia coli* chromosomal homologs of the pem locus responsible for stable maintenance of plasmid R100. J Bacteriol 1993;175:6850–6856.
6. Hazan R, Sat B, Reches M, Engelberg-Kulka H. Postsegregational killing mediated by the P1 phage "addiction module" phd-doc requires the *Escherichia coli* programmed cell death system mazEF. J Bacteriol 2001;183:2046–2050.
7. Sat B, Hazan R, Fisher T, Khaner H, Glaser G, Engelberg-Kulka H. Programmed cell death in *Escherichia coli:* Some antibiotics can trigger mazEF lethality. J Bacteriol 2001;183:2041–2045.
8. Gerdes K. Toxin–antitoxin modules may regulate synthesis of macromolecules during nutritional stress. J Bacteriol 2000;182: 561–572.
9. Mittenhuber G. Occurence of MazEF-like antitoxin/toxin systems in bacteria. J Mol Microbiol Biotechnol 1999;1:295–302.
10. Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science 1997;278:631–637.
11. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 2001;29:22–28.
12. Mossessova E, Lima CD, Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell 2000;5:865–876.
13. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
14. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr 1999;D55:849–861.
15. Terwilliger TC. Maximum likelihood density modification. Acta Crystallogr 2000;D56:965–972.
16. Perrakis A, Sixma TK, Wilson KS, Lamzin VS. wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. Acta Crystallogr 1997;D53:448–455.
17. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr 1997;D53:240–255.
18. Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. Acta Crystallogr D 1994;50: 760–763.
19. Evans SV. SETOR: Hardware-lighted three-dimensional solid model representations of macromolecules. J Mol Graph 1993;11: 127–138.
20. Nicholls A, Sharp KA, Honig B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
21. Ruiz-Echevarria MJ, Gimenez-Gallego G, Sabariegos-Jareno R, Diaz-Orejas R. Kid, a small protein of the parD stability system of plasmid R1, is an inhibitor of DNA replication acting at the initiation of DNA synthesis. J Mol Biol 1995;247:568–577.
22. Bernard P, Kezdy KE, Van Melderen L, Steyaert J, Wyns L, Pato ML, Higgins PN, Couturier M. The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase J Mol Biol 1993;234:534–541.
23. Maki S, Takiguchi S, Miki T, Horiuchi T. Modulation of DNA supercoiling activity of *Escherichia coli* DNA gyrase by F plasmid proteins: Antagonistic actions of LetA (CcdA) and LetD (CcdB) proteins. J Biol Chem 1992;267:12244–12251.
24. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK—A program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.