PROTEINS: Structure, Function, and Bioinformatics 54:371–374 (2004)

STRUCTURE NOTE

Crystal Structure of the *Escherichia Coli* YjiA Protein Suggests a GTP-Dependent Regulatory Function

Pavel P. Khil,[1] Galina Obmolova,[2] Alexey Teplyakov,[2] Andrew J. Howard,[3] Gary L. Gilliland,[2] and R. Daniel Camerini-Otero[1]

[1]*Genetics and Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, Maryland*

[2]*Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute and the National Institute of Standards and Technology, Rockville, Maryland*

[3]*Center for Synchrotron Radiation Research and Instrumentation, Biological, Chemical and Physical Sciences Department, Illinois Institute of Technology, Chicago, Illinois*

Introduction. The *YjiA* gene of *Escherichia coli* encodes an uncharacterized protein of 36-kDa molecular weight. Its homologs are present in most bacteria and eukaryotes but not in archaea. The amino acid sequence contains several conserved consensus motifs that identify YjiA as a P-loop guanosine triphosphatase (GTPase).[1] These motifs include the Walker A and B sequences involved in the binding of the phosphates of NTP and also the TKXD sequence, which provides specificity for guanine. More specifically, as a member of the G3E subfamily of the SIMIBI class[1] YjiA has a third conserved glycine residue (Gly14) in the Walker A motif $G_{11}XXGXGKT_{18}$ and a conserved aspartate residue at the beginning of the Walker B β-strand (Asp92). A unique feature of YjiA among the GTPases is the presence of a glutamate residue (Glu97) in the Walker B motif instead of a more typical aspartate, and also a substitution of Asp for Glu at the end of the second β-strand (Glu39). The functional role of these substitutions is not clear, nor was the function of the protein itself. *YjiA* was among the genes significantly induced in response to the DNA damage caused by mitomycin C.[2] Therefore, although *yjiA* is not an essential gene, its function seems to be important for viability of the cell.

The crystal structure determination of YjiA was undertaken as part of a structural genomics effort[3] to assist with the functional assignment of the protein. The YjiA protein from *E. coli* was cloned, expressed, and the crystal structure determined at 2.4-Å resolution.

Materials and Methods. *Cloning, expression, and purification*. The *yjiA* gene was PCR amplified from *E. coli* MG1655 genomic DNA and subcloned into a pDONR201 plasmid using the Gateway technology (Invitrogen). For expression, the coding sequence was transferred into a pDEST14 plasmid using site-specific recombination (Invitrogen). The protein was produced in *E. coli* strain BL21 Star (DE3) (Invitrogen) that was transformed with pDEST14. Cells were grown on a minimal LB media containing 50 mg/l seleno-L-methionine and 100 μg/μl ampicillin at 37°C to an A_{600} of 0.6 and induced with 1 mM isopropyl β-D-thiogalactoside for 3 h. The protein was purified by column chromatography in two steps using Source 30Q (Pharmacia) and Butyl-560M (Toyopearl).

Crystallization and structure determination. YjiA crystals were grown by vapor diffusion in hanging drops at room temperature from 0.1 M HEPES, pH 7.5, and 1 M lithium sulfate. The crystals belong to the space group C2 with unit cell parameters: a = 83.8 Å, b = 66.5 Å, c = 56.1 Å, and β = 115.0°. There is one protein molecule in the asymmetrical unit. Crystals were frozen in the mother liquor supplemented with 1.5 M sodium formate.[4] X-ray data were collected at 100 K at the IMCA-CAT beamline 17-ID at the APS (Argonne, IL) equipped with a ADSC Quantum-4 CCD detector and processed with HKL2000.[5]

The structure was solved by the multiwavelength anomalous diffraction method using one SeMet protein crystal. X-ray data to 2.4-Å resolution were measured at three wavelengths (Table I). Six selenium sites were located by the Shake-and-Bake method[6] and were used for phasing

Abbreviations: NTP, nucleotide triphosphate; TNP-GTP, 2'-(or 3')-*O*-(trinitrophenyl)guanosine 5'-triphosphate; PCR, polymerase chain reaction; HEPES, 2-hydroxyethylpiperazineethanesulphonic acid; IMCA-CAT, Industrial Macromolecular Crystallography Collaborative Access Team; APS, Advanced Photon Source; CCD, charge-coupled device; RMSD, root mean square deviation.

Certain commercial materials, instruments, and equipment are identified in this article to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the materials, instruments, or equipment identified is necessarily the best available for the purpose. The accepted SI units of concentration, mol/l, and of unified atomic mass unit, u, have been represented by the symbol M and the symbol Da, respectively, to conform to the conventions of this journal.

Grant sponsor: National Institutes of Health; Grant number: P01-GM57890.

*Correspondence to: Pavel P. Khil, Genetics and Biochemistry Branch, NIDDK, National Institutes of Health, 5 Memorial Drive, Bethesda, MD 20892. E-mail: pk94j@nih.gov

Received 18 January 2003; Accepted 28 January 2003

TABLE 1. X-Ray Data and Refinement Statistics

Data set	Remote	Peak	Inflection
Wavelength (Å)	0.9649	0.9794	0.9796
Resolution (Å)	2.4	2.4	2.4
Number of unique reflections[a]	21,164	21,109	20,942
Completeness (%)	98.5	98.3	97.5
Redundancy	2.3	2.3	2.3
R_{sym} ($\Sigma\|I - \langle I\rangle\|/\Sigma I$)	0.034	0.034	0.034
$\langle I/\sigma\rangle$ (outer shell)	5.0	4.5	4.0
R_{cryst} ($\Sigma\|\|F_o\| - \|F_c\|\|/\Sigma\|F_o\|$)	0.198		
R_{free} (3% data)	0.232		
Number of protein atoms	2,498		
Number of water molecules	136		
RMSD in bonds (Å)	0.019		
RMSD in angles (°)	2.1		
Mean B factor from the model (Å^2)	45.6		
Mean B factor from Wilson plot (Å^2)	46.8		

[a]Anomalous pairs not merged and all reflections were used.

with MLPHARE/DM.[7] The atomic model was built using O[8] and refined with REFMAC.[9] The final model contains all 318 residues except the N-terminal methionine; 86% of the residues are in the most favored conformation. One residue, Gln277, has unusual main-chain torsion angles: $\phi = 51.9°$, $\psi = -120.1°$. The electron density for this residue is unambiguous. The atomic coordinates of the YjiA protein were deposited in the Protein Data Bank (PDB) under accession code 1LA9.

Fluorescence measurements. Binding of trinitrophenyl-labeled GTP (TNP-GTP) (Molecular Probes) to YjiA was measured by its fluorescence at 16°C with an excitation wavelength of 408 nm and a maximum of the emission spectrum at 540 nm using a Fluoromax-2 spectrofluorometer (Jobin Yvon). The reaction mixture contained 0–2.5 μM protein, 2 mM $MgCl_2$, and 50 mM Tris-HCl, pH 7.5. Titration curves were obtained by adding TNP-GTP to the reaction mixture at a fixed protein concentration. In the competition experiment, GTP was added to the preincubated solution that contained 2.0 μM protein, 30 μM TNP-GTP, 2 mM $MgCl_2$, and 50 mM Tris-HCl, pH 7.5.

Results and Discussion. The YjiA protein is folded into two distinct domains with a 22-residue linker between them (Fig. 1). The N-terminal domain includes residues 1–202 and is dominated by a 9-stranded mostly parallel β-sheet. The fold is characteristic of P-loop NTP-binding proteins.[10] The P-loop with the consensus sequence $G_{11}XXXXGKT_{18}$ follows the N-terminal β-strand located in the middle of the β-sheet.

The distribution of conserved residues reveals a cluster at the NTP-binding site that includes the four canonical motifs of G-proteins,[11] namely: (1) the phosphate-binding Walker A motif (P-loop), (2) the Walker B motif with the invariant residue Glu97, (3) the switch I region with the invariant residues Asn38 and Glu39 that bind the γ-phosphate of NTP, and (4) the base recognition motif $T_{158}KTD_{161}$ (usually NKXD in the G-proteins) with the invariant residues Lys159 and Asp161. The presence of the latter motif indicates the specificity of YjiA toward guanine.

Specific binding of GTP by YjiA was established in a series of experiments using a nucleotide analog with the

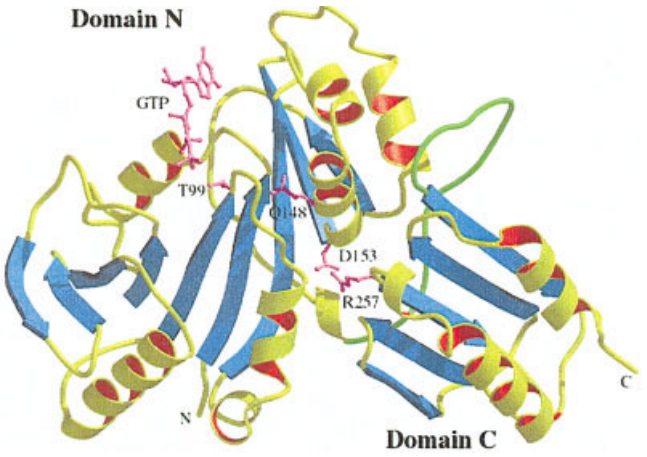

Fig. 1. Ribbon presentation of the polypeptide fold of the YjiA protein. The linker between domains is shown in green. The GTP molecule was modeled by superposition on the Ras-GTP structure (PDB code 1CTQ). Residues implicated in the signal transfer between the GTP site and the C-domain are shown as ball-and-stick models.

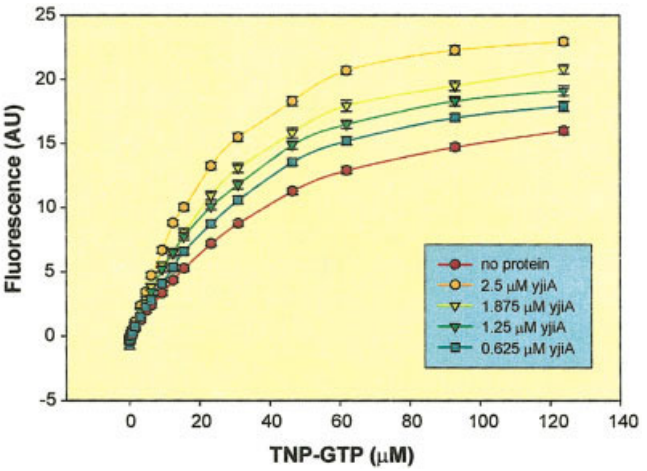

Fig. 2. Effect of YjiA on the fluorescence of TNP-GTP at different protein concentrations. The reaction mixture contained 2 mM $MgCl_2$ and 50 mM Tris-HCl, pH 7.5.

fluorescent label trinitrophenyl (TNP) attached to the ribose of GTP. Figure 2 shows titration curves observed as an increase in fluorescence upon increasing the concentration of TNP-GTP for several protein concentrations as well as for the protein-free solution. The relative protein-induced increase in fluorescence reaches 50%, which indicates the binding of TNP-GTP to the protein. Figure 3 shows the quenching of fluorescence of TNP-GTP bound to the protein upon addition of GTP. YjiA at a concentration of 2.0 μM was preincubated with 30 μM TNP-GTP in the presence of 2 mM $MgCl_2$, and then GTP was added up to the concentration of 1 mM. The sigmoidal curve with an inflection at about 50 μM GTP indicates the replacement of TNP-GTP by GTP implying their competition for the same binding site.

The three-stranded β-meander (β3-β4-β5 at one edge of the β-sheet is an interesting topological feature of YjiA. This fragment is likely to play an important functional role

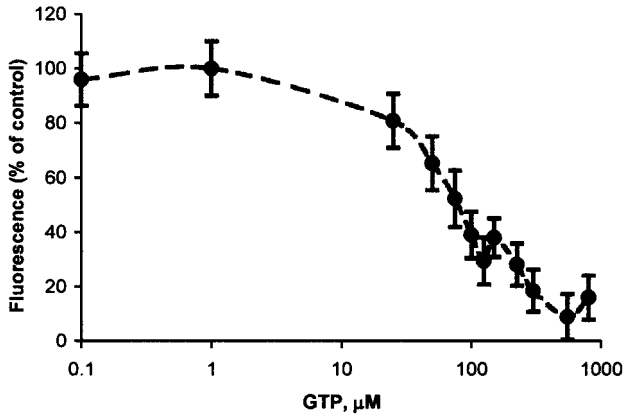

Fig. 3. Quenching of fluorescence of TNP-GTP bound to YjiA upon addition of GTP. The reaction mixture contained 2.0 μM YjiA, 30 μM TNP-GTP, 2 mM $MgCl_2$, and 50 mM Tris-HCl, pH 7.5.

due to the presence of the invariant Asp46 and the pentapeptide $G_{63}CXCC_{67}$ with four of five residues being conserved among the YjiA homologs. One possibility is the chelating of a metal ion by these residues; another is the interaction with a target protein. A combined function of binding a metal ion and presenting it to the receptor also cannot be ruled out. It is worth noting that the β-meander is directly attached to the switch I region of the protein, and therefore its conformation may be sensitive to GTP/GDP exchange.

The linker between the N and C domains includes residues 203–224. It may allow the relative movement of the domains. The area of direct contact between domains is only about 600 $Å^2$, that is, less than 10% of the accessible surface of the C domain and 6% of the N domain surface. Given the small interdomain interface and a sufficiently long linker, one can expect significant flexibility of the two-domain structure. In some species, the linker is much longer (up to 100 residues), and has a histidine-rich sequence, which may be related to Ni binding by the proteins. The sequence of the linker in the *E. coli* protein does not suggest such a possibility.

The smaller, C-terminal domain includes residues 225–318 that form an antiparallel β-sheet flanked by two α-helices on one side. Topologically, the C domain resembles the ferredoxin fold, although it has an additional β-hairpin on one side of the β-sheet, and lacks a β-strand on the other side of the β-sheet. The ferredoxin type of α + β sandwich is common for many functionally unrelated and nonhomologous proteins, such as metallochaperones, protease propeptides, and RNA-binding domains of various enzymes.[12]

There are only three conserved residues, Arg257, Lys259, and Gly260, in the C domain. They are located on the surface facing the N domain, that is, inaccessible to solvent in the present structure. The buried location of the charged residues suggests two alternative hypotheses of their supposedly important role: One predicts the dissociation of the domains in the functional cycle, so that these residues become exposed and may interact with another molecule. According to the other hypothesis, these residues form a chain of communication between the two domains that also includes invariant residues Gln148 and Asp153. This idea is in agreement with the function of YjiA as a molecular switch that is typical for G-proteins. Thus, the GTP/GDP exchange and related conformational changes in the N domain may be transferred to the periphery of the molecule, namely, to the C domain.

While the crystal structure suggests the molecular function of YjiA as a GTP-dependent regulator, the biologic role of the protein in the cell remains unclear. Analysis of the gene expression levels in *E. coli* under the conditions of genotoxic stress[2] indicates that YjiA may regulate the pathway that is essential for cell viability. The *cobW* gene, which is orthologous to *E. coli yjiA*, was identified as part of the cob cluster of genes in *Pseudomonas denitrificans* and therefore may be somehow involved in cobalamin biosynthesis.[13] However, the position of the *yjiA* gene in cob cluster is not conserved in other species. Given the conserved motif GCXCC in the β-meander, a metallochaperone function is a distinct possibility. It is worth noting that two other GTPases closely related to YjiA, UreG and HypB, have been implicated in the assembly of the Ni^{2+} centers in metalloenzymes.[14,15] But, these proteins contain histidine-rich nickel binding motif absent in YjiA. Further biochemical and biophysical studies will shed more light on the molecular and cellular functions of the YjiA protein. These studies will be facilitated by the 3D structure of YjiA presented here.

Acknowledgments. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract W-31-109-Eng-38. This work was also supported in part by an award from the W.M. Keck Foundation.

REFERENCES

1. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 2002;317:41–72.
2. Khil PP, Camerini-Otero RD. Over 1000 genes are involved in the DNA damage response of *Escherichia coli*. Mol Microbiol 2002;44: 89–105.
3. Eisenstein E, Gilliland GL, Herzberg O, Moult J, Orban J, Poljak RJ, Banerjei L, Richardson D, Howard AJ. Biological function made crystal clear—annotation of hypothetical proteins via structural genomics. Curr Opin Biotech 2000;11:25–30.
4. Rubinson KA, Ladner JE, Tordova M, Gilliland GL. Cryosalts: suppression of ice formation in macromolecular crystallography. Acta Crystallogr D 2000;56:996–1001.
5. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Meth Enzymol 1997;276:307–326.
6. Weeks CM, Miller R. Optimizing Shake-and-Bake for proteins. Acta Crystallogr D 1999;55:492–500.
7. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D 1994;50: 760–763.
8. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
9. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by maximum-likelihood method. Acta Crystallogr D 1997;53:240–255.
10. Saraste M, Sibbald PR, Wittinghofer A. The P-loop—a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 1990;15:430–434.
11. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily:

conserved structure and molecular mechanism. Nature 1991;349: 117–127.
12. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
13. Crouzet J, Levy-Schil S, Cameron B, Cauchois L, Rigault S, Rouyez MC, Blanche F, Debussche L, Thibaut D. Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair *Pseudomonas denitrificans* DNA fragment containing five cob genes and identification of structural genes encoding Cob(I)alamin adenosyltransferase, cobyric acid synthase, and bifunctional cobinamide kinase–cobinamide phosphate guanylyltransferase. J Bacteriol 1991;173:6074–6087.
14. Lee MH, Mulrooney SB, Renner MJ, Markowicz Y, Hausinger RP. *Klebsiella aerogenes* urease gene cluster: sequence of ureD and demonstration that four accessory genes (ureD, ureE, ureF, and ureG) are involved in nickel metallocenter biosynthesis. J Bacteriol 1992;174:4324–4330.
15. Maier T, Jacobi A, Sauter M, Bock A. The product of the hypB gene, which is required for nickel incorporation into hydrogenases, is a novel guanine nucleotide-binding protein. J Bacteriol 1993;175:630–635.