PROTEINS: Structure, Function, and Genetics 52:633–635 (2003)

STRUCTURE NOTE

Crystal Structure of a Flavin-Binding Protein From *Thermotoga maritima*

Weiru Wang,[1] Rosalind Kim,[1] Jaru Jancarik,[2] Hisao Yokota,[1] and Sung-Hou Kim[1,2]*
[1]*Physical Biosciences Division of the Lawrence Berkeley National Laboratory, Berkeley, California*
[2]*Department of Chemistry, University of California, Berkeley, California*

Introduction. The Berkeley Structural Genomics Center (BSGC) pursues an integrated structural genomics program designed to obtain a near-complete structural representation of two minimal genomes: *Mycoplasma genitalium* (MG) and *Mycoplasma pneumoniae* (MP), two related human pathogens. The MP gene gi:674379 (MP379) encodes an enzyme with no sequence homologue in the current PDB structure database.[1] The protein structure determined in this work, gi:4981319 from *Thermotoge maritima* (TM379), was selected as a target based on sequence homology to MP379 (log ($E_{\text{PSI-BLAST}}{}^{2}$) = −57, coverage = 98.88%). Both MP379 and TM379 gene were annotated as FAD synthetase,[3] a bifunctional enzyme that catalyzes two reactions: riboflavin kinase (EC 2.7.1.26) and FMN adenylyltransferase (EC 2.7.7.2). This sequence annotation is awaiting experimental verification. FAD synthetase is present among all kingdoms of life. Mammals use separate enzymes for FMN and FAD formation,[4,5] whereas lower level organisms depend on the bifunctional enzyme. To date, there are no structural data available for riboflavin kinase, FMN adenylyltransferase, or the bifunctional enzyme from any organism.

Results. TM379 is present as a dimer in the crystallographic asymmetric unit. Each monomer is in an elongated shape with dimensions of 32 × 32 × 62 Å^3. The TM379 structure is composed of two structural domains with a single loop linkage in between. The N-terminal domain (residues 2–134) adopts a typical nucleotide-binding fold (Rossmann fold). The C-terminal domain (residues 135–288) contains a six-stranded antiparallel β-barrel architecture. The C-terminal domain also includes three α-helices. Two of them are on the N-terminal end and one long helix is on the C-terminal end. The two N-terminal helices are located at the bottom of the β-barrel shown in Figure 1. The C-terminal helix runs across the side of the β-barrel and is oriented approximately perpendicular to the β-barrel's axis.

Structural homology search using DALI[6] algorithm revealed that TM379 N-terminal domain has significant homology with other nucleotidyltransferases including glycerol-3-phosphate cytidylyltransferase (1COZ), nicotinamide mononucleotide adenylyltransferase (1F9A), and phosphopantetheine adenylyltransferase (1B6T). Such structural homology with nucleotidyltransferases may imply that the N-terminal domain of TM379 is involved in the FMN adenylyltransferase, the second reaction catalyzed by FAD synthetase. The DALI search also showed that the TM379 C-terminal domain is homologous to several flavin-binding proteins including the FAD-binding domain of benzoate dioxygenase reductase (1KRH), the FMN-binding domain of phthalate dioxygenase reductase (2PIA), the FAD-binding domain of flavohemoglobin (1CQX), and the flavin-binding domain of flavin oxidoreductase (1QFJ). All of these flavin-binding domains contain a six-stranded antiparallel barrel and helices on both N- and C-termini with a similar arrangement as in the C-terminal domain of TM379. The structural homology strongly suggests that the C-terminal domain of TM379 contributes to binding of flavin and probably is involved in both reactions. Although both the N- and C-terminal domains have homology to known structures, the TM379 structure exhibits a novel combination of a Rossmann fold and a flavin-binding fold.

Methods. Cloning. Primers (Invitrogen Corp., Carlsbad, CA) for PCR amplification from genomic DNA contained an NdeI restriction site in the forward primer (5′-CATATGGTTGTCAGCATCGGAGTTTTC-3′) and BamHI site in the reverse primer (5′-GGATCCTTACCCCTCTTTTTCGAACTTCGAATT-3′). PCR was performed by using Deep Vent Polymerase (New England Biolabs, Inc., Beverly, MA) and *Thermotoga maritima* genomic DNA. The PCR product was cloned into pCR-BluntII-TOPO vector (Invitrogen) and the TM379 gene insert was confirmed by DNA sequencing. The amplified TOPO vector was restricted with NdeI and BamHI, and the gene insert was purified by agarose gel electrophoresis extraction. This insert was ligated into pET21a (Novagen, Inc., Madison, WI) digested with NdeI and BamHI and transformed into DH5. A plasmid containing the gene insert was confirmed and then transformed into BL21(DE3)pSJS1244 for protein expression.[7]

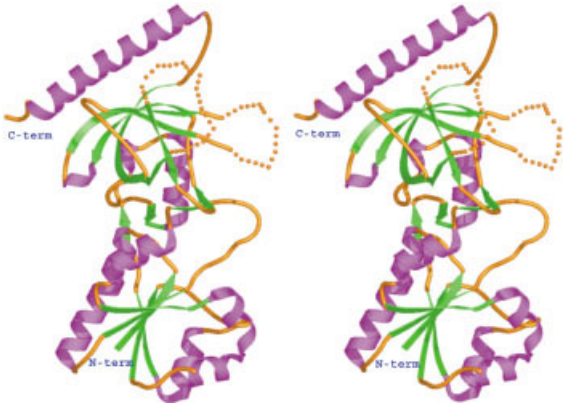

Fig. 1. A ribbons diagram of TM379 structure. The color coding is based on secondary structure. The dotted loops represent the disordered regions.

Protein expression and purification. pET21a.TM379 protein was overexpressed in *E. coli* strain B834(DE3)/pSJS1244[6] cells upon induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside. Selenomethionyl (Se-Met) protein was prepared according to the method of Doublie.[8] Bacteria were lysed by sonication in buffer A (50 mM Tris, pH 8, 0.5 mM PMSF, 1 μg/mL antipain, 1 μg/mL chymostatin, 0.5 μg/mL leupeptin, and 0.7 μg/mL pepstatin A), and cell debris were pelleted by centrifugation at 39,000 × *g* for 20 min in a Sorvall centrifuge. Because this protein is heat stable, the cleared lysate was incubated at 70°C for 20 min to denature most of the *E. coli* proteins. The lysate was then spun in a Beckman ultracentrifuge Ti45 rotor at 60,000 × *g* for 20 min at 4°C. The eluted sample was dialyzed against 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, bound onto a Mono Q 10/10 column (Pharmacia Biotech, Uppsala, Sweden) and eluted with a 15-column volume (CV) linear gradient from 0 to 1.0 M NaCl in the same buffer. The target protein was found in the flow through; because its predicted pI is 8.8, fractions were pooled and dialyzed against 20 mM Bis-Tris, pH 6.5, 1 mM EDTA. The pooled fractions were bound onto a 5-mL HiTrap S column (Pharmacia Biotech, Uppsala, Sweden) and eluted in 10 CVs from 0 to 1.0 M NaCl. The protein eluted in the gradient at 150 mM NaCl.

The purity of the expressed protein was determined by sodium dodecyl sulfate (SDS) gel electrophoresis and electrospray mass spectrometry. The protein was concentrated in an Ultrafree 10 K unit (Millipore, Bedford, MA) in buffer containing 20 mM Tris-HCl, pH7.5, 1 mM EDTA, and 10% glycerol to 6.5 mg/mL.

Crystallization. The native protein was further concentrated in 20 mM Tris, pH 7.5, 1 mM EDTA, 10% glycerol, to a final concentration of 21.5 mg/mL. Crystallization conditions were tested with a sparse matrix sampling method[9] using hanging drop vapor diffusion method at room temperature with commercially prepared reagents (Hampton Research, Laguna Niguel, CA). The optimum conditions were found in 0.2 mM sodium citrate, 10% polyethylene glycol 4000 (PEG 4K), 5% glycerol, and 10% propanol. The PEG4K concentration was raised to 30% in the reservoir 1 day before data collection. Crystals reached a maximum size of 0.05 × 0.1 × 0.2 mm^3. Se-Met-labeled protein was crystallized in the same condition.

TABLE I. X-ray Data and Refinement Statistics

Data set	SAD	Native
Space group	C2	$P2_1$
Wavelength (Å)	0.97911	0.91963
Resolution (Å)	30.0–2.5	30.0–1.9
Redundancy	6.3 (4.6)	3.8 (3.5)
Unique reflections	30829 (1357)	51627 (2190)
Completeness (%)	88.3 (99.2)	96.9 (82.2)
I/σ	21.8 (2.9)	27.4 (3.0)
Rsym[b] (%)	7.1 (48.9)	3.8 (29.9)
Space group	$P2_1$	
Total number residues	536	
Total non-H atoms	4816	
No. of water molecules	442	
Average temperature factors ($Å^2$)	21.8	
R factor (%)	21.6	
Free R factor (%)	23.3	
Stereochemical ideality		
Bond (Å)	0.007	
Angle (°)	1.12	
Improper (°)	2.51	

[a]Numbers in parentheses are related to the highest resolution shell, which is 2.56–2.50 Å for the SAD data and 1.93–1.90 Å for the native data set.
[b]$R_{sym} = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I \rangle_{hkl}| / \Sigma |\langle I \rangle_{hkl}|$.

Structure determination and refinement. The X-ray diffraction data were collected at the advanced light source (ALS) beam line 5.0.2 by using an Area Detector System Co. (ADSC) Quantum 4 CCD detector. A single wavelength anomalous data set was collected with a Se-Met crystal, and a monochromatic data set was collected with a native crystal. The diffraction data were processed by using the programs DENZO and SCALEPACK.[10]

X-ray data statistics are shown in Table I. The Se atom positions were located by using program SOLVE.[11] Twelve Se sites were found in an asymmetric unit with a figure of merit (FOM) of 0.35 using the SAD data up to 2.5 Å. SOLVE-calculated phases were subject to 10 cycles of refinement, and solvent flattening[12] followed by noncrystallographic symmetry (NCS) averaging using program DM.[13] A twofold NCS matrix was determined on the basis of the Se positions. The FOM after NCS averaging was 0.881 overall and 0.776 in the 2.56–2.50 Å resolution shell.

The model building was performed by using O.[14] The initial model was traced in the 2.5 Å NCS averaged experimental map for a monomer. A model containing 258 residues was derived from progressive improvement of the electron density map using rounds of phase combination and manual building. This model was then subjected to preliminary refinement using program CNS.[15] The native TM397 crystal diffracted to a much higher resolution, 1.9Å (Table I). However, it was crystallized in $P2_1$ space group, different from the Se-Met crystal. Therefore, we performed a molecular replacement using the refined Se-Met struc-

ture in C2 as the search model. The molecular replacement calculation was performed with program AmoRe.[16]

The molecular replacement solution was then refined against the $P2_1$ native data set to 1.9 Å using CNS.[15] At this point, the NCS constraints and restraints were completely released throughout the refinement. Ten percent of the data were randomly picked for free R factor cross-validation. The refinement statistics are shown in Table I. Atomic coordinates have been deposited at the protein data bank with access code 1MRZ.

Acknowledgments. We thank Dr. David King for mass spectrometric analysis of the protein. We are also grateful to Dr. Thomas Ernest and Dr. Gerry McDermott for help during the data collection at ALS beam line 5.0.2. We are grateful to Barbara Gold and Andrew Waight for cloning of the gene of TM379.

Grant sponsor: National Institutes of Health; Grant number: GM 62412.

W. Wang's present address is Plexxikon, Inc., Berkeley, CA 94710.
*Correspondence to: Sung-Hou Kim, Department of Chemistry, University of California, Berkeley, CA. E-mail: SHKim@cchem.berkeley.edu

Received 5 November 2002; Accepted 7 November 2002

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Protein data bank. Nucleic Acids Res 2000;28:235–242.
2. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
3. Manstein DJ, Pai EF. Purification and characterization of FAD synthetase from *Brevibacterium ammoniagenes*. J Biol Chem 1986;261:16169–16173.
4. Merrill AH, McCormick DB. Affinity chromatographic purification and properties of flavokinase (ATP:riboflavin 5′-phosphotransferase) from rat liver. J Biol Chem 1980;255:1335–1338.
5. Oka M, McCormick DB. Complete purification and general characterization of FAD synthetase from rat liver. J Biol Chem 1987;262:7418–7422.
6. Holm L, Sander C. Mapping the protein universe. Science 1996;273:595–602.
7. Kim R, Sandler SJ, Goldman S, Yokota H, Clark AJ, Kim S-H. Overexpression of archaeal proteins in *Escherichia coli*. Biotech Lett 1998; 20:207–210.
8. Doublie S. Preparation of selenomethionyl proteins for phase determination. Methods Enzymol 1997; 276:523–529.
9. Jancarik J, Kim S-H. Sparse matrix sampling: a screening method for crystallization of proteins. J Appl Crystallogr 1991;24:409–411.
10. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1996;276:307–326.
11. Terwilliger TC, Kim S-H, Eisenberg D. Generalized method of determining heavy-atom positions using the difference Patterson function. Acta Crystallogr 1987;43:1–5.
12. Wang B-C. Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol 1985;115:90–112.
13. Cowtan K. Joint CCP4 and ESF-EACBM newsletter on protein crystallography 1994;31:34–38.
14. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr 1991;47:110–119.
15. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jlang JS, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr 1998;54:905–921.
16. Navaza J. AM. Re: An automated package for molecular replacement. Acta Crystallogr 1994;A50:157–163.