PROTEINS: Structure, Function, and Bioinformatics 54:814–816 (2004)

STRUCTURE NOTE

Crystal Structure of *Human* p120 Homologue Protein PH1374 From *Pyrococcus horikoshii*

Idumi Ishikawa,[1] Naoki Sakai,[1] Tomohiro Tamura,[2] Min Yao,[1] Nobuhisa Watanabe,[1]* and Isao Tanaka[1]
[1]*Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo, Japan*
[2]*Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology (AIST), Sapporo, Japan*

Introduction. Human p120 is a nucleolar proliferation-associated antigen, and has been detected in the nucleoli of a variety of malignant tumors, including cancers of the breast, liver, gastrointestinal tract, genitourinary tract, blood, lymph system, lung, and brain.[1] It contains 855 amino acid residues, and its central hydrophobic domain (residues 380–583) is conserved among many organisms. Expression of p120 is increased 60- to 80-fold in several tumor cells as compared to normal resting cells and correlates with the tumor growth rate. Furthermore, p120 is expressed in the G1 phase of the cell cycle and its level peaks during S phase. In 1994, using computer methods for identification of amino acid motifs in sequence databases and multiple alignment, it was suggested that the central domain (380–583) of p120 and its homologues comprises a new family of S-adenosyl-L-methyonine (SAM)-utilizing methyltransferases (MTase).[2] A sequence homologue of the p120 central domain from *Saccharomyces cerevisiae*, Nop2p, was thought to play a direct role in pre-25S rRNA processing and 60S subunit production and yeast viability.[3] Another homologous protein from *Escherichia coli*, Fmu, was reported to catalyze transference of a methyl group to carbon 5 of a pyrimidine nucleotide at position 967 cytosine of *E. coli* 16S rRNA; this represents the first RNA m^5C MTase whose amino acid sequence is known.[4,5] In this context, this series of p120 homologue proteins is thought of as an RNA m^5C MTase family.

PH1374 from *Pyrococcus horikoshii* OT3 consists of 315 amino acids. This protein has 16% identity to all amino acids of *human* p120, but near the central domain of p120, it has 41% identity in 288 aa overlap (p120 residues 300–581, PH1374 residues 35–315). Though it is known that four genes are paralogues to PH1374 in *P. horikoshii* genomes by COGs database search, of the five proteins, PH1374 has the highest similarity to p120. Also, proteins that have sequence homology with human p120 characteristically contain two highly conserved cysteines. These cysteines are thought of as essential to the protein's function, but studies with Nop2p and Fmu seemed to be contradictory in terms of which residue acts as the active-site nucleophile in covalent catalysis.[6–8]

Materials and Methods. The PH1374 gene was amplified by polymerase chain reaction (PCR) from the genomic DNA of *P. horikoshii*, and cloned into a pET-22b(+) vector (Novagen). The $(His)_6$-tagged recombinant PH1374 was expressed in cells of the *E. coli* strain BL21-Star™(DE3)-RIL (Stratagene), and two additional amino acids (leucine and glutamic acid) derived from the vector were inserted between the last amino acid of the protein and the $(His)_6$-tag. The cells were grown at 37°C in 6 L Luria–Bertani (LB) medium containing 50 μg/mL ampicillin and 34 μg/mL chloramphenicol. Expression of recombinant protein was induced by adding 0.4 m*M* isopropyl-γ-D-galactoside (IPTG). After IPTG injection, *E. coli* was cultured for 15 h with shaking. Cells were harvested, then disrupted using a French press in A buffer (50 m*M* sodium phosphate pH 8.0, 300 m*M* NaCl). Cell debris was removed by centrifugation and the supernatant was mixed to Ni-NTA resin (Qiagen) and eluted using a linear gradient of 0.0–0.4 *M* imidazole. The fractions containing PH1374 were dialyzed against buffer B (10 m*M* sodium phosphate pH 8.0, 200 m*M* NaCl) and applied to a DE52 (Whatman) equilibrated with buffer B. PH1374 was passed through DE52, and contaminants were adsorbed. The purified protein was concentrated to 5 mg/mL. The selenomethionine (Se-Met) derivative was also expressed in BL21-Star™(DE3)-RIL and purified under almost the same conditions as the native protein.

The crystallization conditions for PH1374 consisted of 0.1 *M* citric acid pH5, 15% PEG6K, 1% 1-5 pentandiol, and 5% glycerol. Crystals were grown at 20°C to a size of up to 0.3 mm × 0.04 mm × 0.01 mm within a few months. X-ray diffraction data of Se-Met PH1374 were collected to 1.9 Å resolution at 100 K at the BL41XU in SPring-8 (Hyogo,

Grant sponsor: National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

*Correspondence to: Nobuhisa Watanabe, Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo, Japan. Email: nobuhisa@sci.hokudai.ac.jp.

Received 2 September 2003; Accepted 5 September 2003

TABLE I. X-Ray Data and Refinement Statistics

	Remote	Se-MAD edge	Peak
Data collection			
Beamline		BL41XU, SPring-8	
Wavelength (Å)	0.9000	0.9805	0.9800
Crystallographic data			
Space group	$P2_1$		
Cell parameters (Å)	$a = 63.55$ $b = 41.51$ $c = 71.07$ $\beta = 114.18$		
Data reduction			
Resolution	23.77–1.9 (2.0–1.9)	25.57–1.9 (2.0–1.9)	25.54–2.0 (2.1–2.0)
Number of observed reflections	152004	131357	130678
Unique reflections	26745	23045	22968
Completeness (%)	99(98.6)	99.1(99.1)	99.1(99.1)
Averaged redundancy	5.7(5.7)	5.7(5.8)	5.7(5.7)
Averaged I/s(I)	6.8(2.6)	6.0(2.6)	5.9(3.3)
R_{meas} (%)	9.7(32.9)	8.7(27.1)	10.2(26.9)
R_λ (%)		6.9(12.5)	5.9(10.8)

$R_{meas} = \Sigma_h[m/(m-1)]^{1/2}\Sigma_j|\langle I\rangle_h - I_{h,j}|/\Sigma_h\Sigma_j I_{h,j}$, where $<I>_h$ is the mean intensity of symmetry-equivalent reflections and m is redundancy.
$R_\lambda = \Sigma_h|F_{h\lambda i} - F_{h\lambda o}|/\Sigma_h|F_{h\lambda o}|$, where λ_o is the remote wavelength.

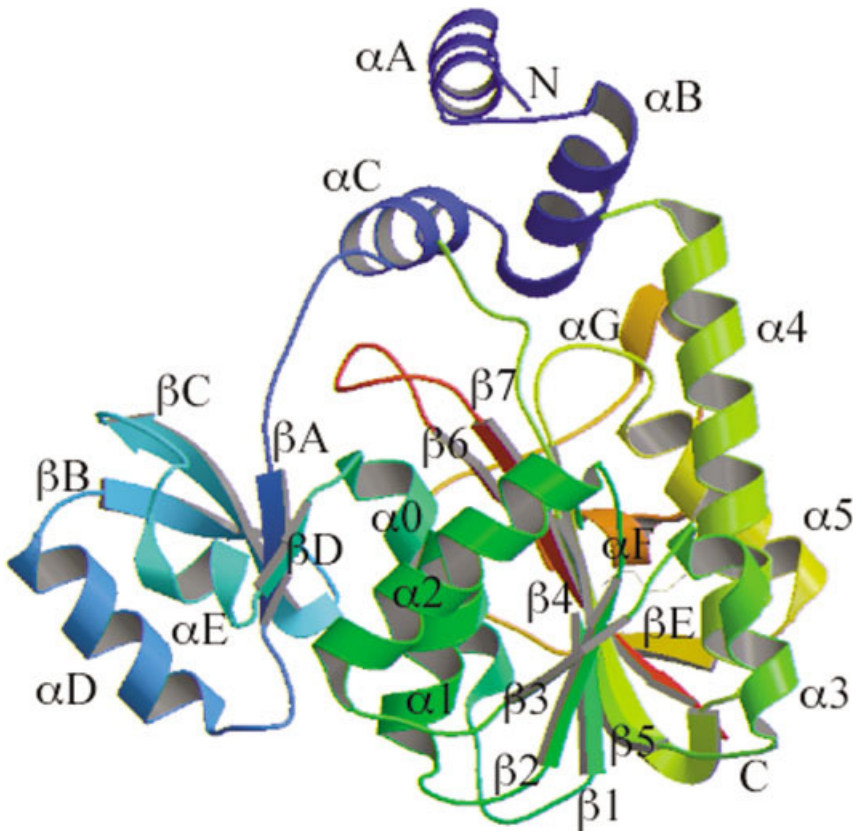

Fig. 1. A ribbon diagram of PH1374 structure. The secondary structural elements common among MTase domains are labeled β1–β7 and αA–αE, consistent with standard nomenclature.

Japan). The crystal was transferred to cryoprotectant solution containing 25% glycerol in the reservoir. The diffraction images were integrated with MOSFLM[9] and processed with the CCP4 suite.[10] Statistics for this data are shown in Table I.

The structure of PH1374 was determined by the multiple-

wavelength anomalous dispersion (MAD) method. The positions of 7 selenium atoms were determined by the program SOLVE.[11] Heavy-atom parameters were refined and phase calculation was carried out using the program SHARP.[12] Initial phases were improved further with SOLOMON[13] using the procedure in SHARP. The model of PH1374 was built automatically using ARP/wARP,[14] except two N-terminal residues, Met1 and Lys2, and 8 residues from Lys205 to Asn212. The model was refined, and water molecules were located using the program CNS.[15] The manual fitting of the modal was performed using O.[16] The final model consisted of 305 amino acids and 151 water molecules. The refinement statistics are also presented in Table I. Atomic coordinates have been deposited at the Protein Data Bank (PDB) with access code 1IXK.

Results. The crystal structure of p120 homologue PH1374 was determined at 1.9 Å resolution (Table I). PH1374 was shown to be composed of 13α-helices and 12 β-strands divided into 3 distinct domains of different sizes (Fig. 1). The N-terminal small domain is composed of 3 α-helices (αA–αC). The second small domain is composed of 4 β-strands (βA–βD) and 2 α-helices (αD, αE). The C-terminal core domain has a mixed α/β structure with alternating α-helices and β-strands. The focal point of this core domain is a 7-stranded β-sheet (β1–β7), flanked by 3 α-helices on one side (α0–α2) and 3 α-helices on the other (α3–α5). The loop between β4 and α4, which contains 8 residues, Lys205–Asn212, was completely disordered. Also βE, αF, and αG, located between α5 and β6, protrude and form an upright loop to the 7-stranded β-sheet.

As predicted, a structural homology search using DALI[17] revealed that PH1374 has significant homology with other members of the SAM-dependent MTase superfamily. Especially, the C-terminal core domain of PH1374 has a consensus topology called SAM-dependent MTase fold.[18,19] In contrast to this conserved catalytic domain, the 2 N-terminal domains of PH1374 are novel in the known SAM-dependent MTase superfamily. According to the comparison with known structures, which were solved as SAM or S-adenosyl-L-homocysteine binding complexes, the environment around the expected SAM binding site of PH1374 seems not suitable for binding. No extra electron density peak for binding SAM was observed there. Especially, the candidate site of PH1374 for ribose binding is buried by Pro196, which is located in the loop between β4 and α4. These suggest that some structural change might be necessary for SAM binding to PH1374. It has been observed that two functionally significant cysteines are conserved in the RNA m^5C MTase family.[6–8] Cys197 (PC-Cys) and Cys247 (TC-Cys) of PH1374 correspond to them. However, they made a disulfide bridge in the crystal structure. A search for the biological substrate of PH1374 is currently under way.

Acknowledgments. We thank M. Kawamoto and H. Sakai for their kind help during data collection on the beamline BL41XU of SPring-8, Japan.

REFERENCES

1. Freeman JW, Busch RK, Gyorkey F, Gyorkey P, Ross BE, Busch H. Identification and characterization of a human proliferation-associated nucleolar antigen with a molecular weight of 120,000 expressed in early G1 phase. Cancer Res 1988;48:1244–1251.
2. Koonin EV. Prediction of an rRNA methyltransferase domain in human tumor-specific nucleolar protein P120. Nucleic Acids Res 1994;22:2476–2478.
3. Kressler D, Linder P, de La Cruz J. Protein trans-acting factors involved in ribosome biogenesis in *Saccharomyces cerevisiae*. Mol Cell Biol 1999;19:7897–7912.
4. Tscherne JS, Nurse K, Popienick P, Michel H, Sochacki M, Ofengand J. Purification, cloning, and characterization of the 16S RNA m5C967 methyltransferase from *Escherichia coli*. Biochemistry 1999;38:1884–1892.
5. Gu XR, Gustafsson C, Ku J, Yu M, Santi DV. Identification of the 16S rRNA m5C967 methyltransferase from *Escherichia coli*. Biochemistry 1999;38:4053–4057.
6. King M, Ton D, Redman KL. A conserved motif in the yeast nucleolar protein Nop2p contains an essential cysteine residue. Biochem J 1999;337:29–35.
7. Liu Y, Santi DV. m5C RNA and m5C DNA methyl transferases use different cysteine residues as catalysts. Proc Natl Acad Sci USA 2000;97:8263–8265.
8. King MY, Redman KL. RNA methyltransferases utilize two cysteine residues in the formation of 5-methylcytosine. Biochemistry 2002;41:11218–11225.
9. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 1992;26.
10. Collaborative Computational Project, No. 4. The CCP4 Suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50:760–763.
11. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 1999;55:849–861.
12. La Fortelle E de, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. Methods Enzymol 1997;276:472–494.
13. Abrahams JP, Leslie AGW. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr 1996;52:30–42.
14. Lamzin VS, Wilson KS. Automated refinement of protein models. Acta Crystallogr D Biol Crystallogr 1993;49:129–147.
15. Brunger AT, Adams PD, Clore GM, Delano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
16. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
17. Holm L, Sandler C. Mapping the protein universe. Science 1996; 273:595–602.
18. Schluckebier G, O'Gara M, Saenger W, Cheng X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. J Mol Biol 1995;247:16–20.
19. Cheng X, Roberts RJ. AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res 2001;29:3784–3795.