FEBS Letters 585 (2011) 452–458

journal homepage: www.FEBSLetters.org

Novel structure of an N-terminal domain that is crucial for the dimeric assembly and DNA-binding of an archaeal DNA polymerase D large subunit from *Pyrococcus horikoshii*

Ikuo Matsui [a,*], Yuji Urushibata [a,1], Yulong Shen [a,b], Eriko Matsui [a], Hideshi Yokoyama [c]

[a] Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba 305-8566, Japan
[b] State Key Laboratory of Microbiology Technology, Shandong University, Jinan, Shandong 250100, PR China
[c] School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka 422-8526, Japan

ARTICLE INFO

Article history:
Received 26 July 2010
Revised 30 November 2010
Accepted 22 December 2010
Available online 28 December 2010

Edited by Christian Griesinger

Keywords:
D-family DNA polymerase
DNA replication
Binding domain
Molecular structure
Hyperthermophilic archaea
Pyrococcus

ABSTRACT

Archaea-specific D-family DNA polymerase forms a heterotetramer consisting of two large polymerase subunits and two small exonuclease subunits. The N-terminal (1–300) domain structure of the large subunit was determined by X-ray crystallography, although ~50 N-terminal residues were disordered. The determined structure consists of nine alpha helices and three beta strands. We also identified the DNA-binding ability of the domain by SPR measurement. The N-terminal (1–100) region plays crucial roles in the folding of the large subunit dimer by connecting the ~50 N-terminal residues with their own catalytic region (792–1163).

Structured summary:
DP2 binds to **DP2** by molecular sieving (View interaction)
DP2 binds to **DP2** by fluorescence technology (View interaction)
DP2 binds to **DP2** by circular dichroism (View interaction)

1. Introduction

D-family DNA polymerases (PolD), which were originally discovered in Euryarchaeota [1], have also been identified in Korarchaeota, which may have diverged early from the major archaeal phyla Crenarchaeota and Euryarchaeota [2]. This suggests that PolD are responsible for the replication of the ancestral genome of archaea [3]. From recent analysis of the evolution of DNA replication apparatus, it is likely that the last common ancestor of archaea had two DNA polymerases from the B-family and one from the D-family [3].

PolD from *Pyrococcus horikoshii* (PhoPolD) was proposed to be a heterotetramer (molecular weight: 420 kDa) consisting of two small subunits (DP1s) (PH0123, NCBI Accession No. NP_ 142131; 622 amino acids), and two large subunits (DP2s) (PH0121, NCBI Accession No. NP_ 142130; 1434 amino acids) [4]. DP2 is the catalytic subunit of DNA polymerase [4], while DP1 is the catalytic subunit of Mre11-like 3′–5′ exonuclease, which shows low but significant homology to the non-catalytic second subunit found in eukaryotic B-family DNA polymerases (Pols α, δ, and ε) [5]. Interestingly, it was reported that PolD demonstrated strong DNA polymerase and 3′–5′ exonuclease activities were acquired when the two subunits were mixed or co-expressed, even though each individual subunit demonstrated weak activity [5,6]. The domain containing the 300 N-terminal residues of DP2 [abbreviated as DP2(1–300); similar descriptions for other fragments will be used in the present manuscript] was reported to be essential for the folding of PolD and is probably the oligomerization domain [7]. Since the molecular mechanisms of the protein folding and biochemical function of the DP2(1–300) domain are unknown, we investigated the crystal structure of the domain and its key roles in the dimeric assembly and the self-cyclization of the DP2 subunit.

2. Materials and methods

2.1. Sample preparation for X-ray analysis

To prepare a selenomethionine (SeMet) derivative of DP2 (1–300), *Escherichia coli* BL21 (DE3) Codon-Plus RIL-X (Stratagene) was transformed with the previously reported co-expression plasmid pET15b/SL(1–300) [7]. The strain was grown in LeMaster's

* Corresponding author. Fax: +81 29 8616151.
E-mail address: ik-matsui@aist.go.jp (I. Matsui).
[1] Present address: NB Health Laboratory Co. Ltd., A1-551 12–18, Kamiaoki 3, Kawaguchi City, Saitama 333-0844, Japan.

0014-5793/$36.00
doi:10.1016/j.febslet.2010.12.040

medium [8] without methionine and supplemented with 50 mg/l L-SeMet (Sigma) and 0.8% (w/v) lactose. Crystallization drops were prepared by mixing equal volumes of protein solution (6 mg/ml) and reservoir solutions containing 22.5% PEG10000 and 0.1 M Hepes–NaOH (pH 7.5), and 40 mM guanidine–HCl was added to the droplets. Then, crystals were grown at 20 °C according to the hanging-drop vapor-diffusion method.

2.2. Data collection, structure determination and refinement of DP2(1–300)

A crystal was cryoprotected with 15% (v/v) glycerol containing the reservoir solution, and flash-frozen at 100 K. The X-ray diffraction data of the SeMet DP2(1–300) were collected at beamline BL6A of the Photon Factory in KEK (Tsukuba, Japan) with a Quantum-4R CCD detector, and were processed and scaled with HKL2000 [9]. The structure was determined using the multiwavelength anomalous dispersion (MAD) method. Four of eight possible selenium sites were found by the direct method with SHELXS [10] and refined with SHARP [11] along with the CCP4 suite [12]. The remaining four SeMet residues were in the disordered region. The initial model was built by automated chain tracing with ARP/wARP [13]. Crystallographic refinement was carried out against *remote* data (Table 1). Iterative cycles of model building and refinement were performed with TURBO-FRODO and CNS [14]. The last stage of refinement was performed with REFMAC5 in the CCP4 suite [12]. Data collection and refinement statistics are summarized in Table 1. Fig. 2A–D was produced with PyMOL (http://pymol.sourceforge.net/). Fig. 2E was produced with CueMol (http://cuemol.sourceforge.jp/en/). The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB) with the Accession Code 3O59.

2.3. SPR measurements

The interaction of the DP2(1–300) domain with 3′-recess DNA, single-strand DNA (ssDNA), and double-strand DNA (dsDNA) was quantitatively analyzed on a BIAcore X apparatus (Biacore) at 25 °C. To generate biotinylated 3′-recess DNA, 500 pmol of 5′ biotinylated A strand (5′-bio-GAGCTAGATGTCGGACTCTGCCTCAAGAC GGTAGTCAACGTGCACTCGAGGTCA-3′, 54mer) was boiled with 500 pmol of ssDNA (5′-TGACCTCGAGTGCACGTTGACTACCGT-3′, 27mer) in 20 μl of binding buffer (50 mM Tris–HCl buffer (pH 8.0) containing 10 mM NaCl, 5 mM EDTA, and 0.005% Tween 20) for 5 min and cooled down to room temperature for 30 min. To generate 5′ biotinylated dsDNA, 500 pmol of the biotinylated A strand was annealed with 500 pmol of the complementary ssDNA (54mer) in 20 μl of the binding buffer. The biotinylated A strand was also used as a ssDNA ligand. Each template DNA was immobilized as a ligand on a streptavidin–dextran layer on the surface of the Sensor Chip SA (Biacore) at 660–970 resonance units with the binding buffer. The flow cell was routinely equilibrated with running buffer (50 mM Tris–HCl buffer (pH 8.0) containing 50 mM NaCl, 1% glycerol, and 0.005% Tween 20). The concentrations of DP2(1–300) as analytes varied from 0.25 to 80 μM. The obtained sensorgrams were analyzed by evaluation software (Biacore) on the basis of a simple 1:1 binding model.

Table 1
Data collection and refinement statistics.

Data collection			
Space group	$P4_32_12$		
Cell dimensions			
a, *c* (Å)	54.0, 173.3		
	Remote	Edge	Peak
Wavelength (Å)	0.9900	0.9793	0.9788
Resolution range (Å)	50–2.2 (2.28–2.20)[a]	50–2.2 (2.28–2.20)	50–2.2 (2.28–2.20)
$R_{merge}(I)$[b]	0.058 (0.233)	0.085 (0.381)	0.108 (0.512)
Average $I/\sigma I$	77.9 (14.4)	63.2 (11.2)	48.6 (7.6)
Unique reflections	13924	13907	13949
Redundancy	27.3	27.5	27.5
Completeness (%)	99.0 (92.4)	99.6 (97.9)	99.6 (98.3)
Refinement			
Resolution (Å)	42.33–2.2		
Reflections used	12379		
R_{work}[c]/R_{free}[d]	0.206/0.272		
No. atoms			
Protein	1909		
Water	107		
B-factors (average) (Å^2)	31.1		
R.m.s. deviations			
Bond length (Å)	0.020		
Bond angle (°)	1.774		

[a] Values in parentheses are for the highest-resolution shell.
[b] $R_{merge}(I) = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed diffraction intensity.
[c] $R = \Sigma|F_o - F_c|/\Sigma F_o$, where F_o and F_c are the observed and calculated structural amplitudes, respectively.
[d] R_{free} is the R value for 10% of the reflections chosen randomly and omitted from refinement.

2.4. Effective refolding of the catalytic domain DP2(792–1163) by reconstitution with the N-terminal domain of DP2, and the isolation of a stable complex by gel filtration

DP2 contains the 166 residue proteinous intron (mini-intein) between Asn954 and Cys1121, as shown in Fig. 1A. After genetically removing the mini-intein, the resultant 206-aa polypeptide of DP2(792–1163) was fused with an N-terminal histidine-tag (20 residues) and overexpressed in *E. coli* cells. The fused catalytic domain DP2(792–1163) was expressed as insoluble inclusion bodies and purified by Ni-affinity chromatography in 6 M urea, as described previously [15]. The catalytic domain refolding trial was started in 3 M urea by mixing the catalytic domain with an equimolar amount of the N-terminal domain of DP2 (i.e., the protein concentration of each domain was 44 μM in 3 M urea), followed by successive stepwise dialysis with 2 M urea and 50 mM Tris–HCl buffer (pH 8.0) containing 100 mM NaCl for 5 h and then with the same buffer containing 100 mM NaCl without urea overnight. The dialyzate was centrifuged at 15000×*g* for 15 min to remove the precipitate and concentrated with a Centricon YM-10 (Amicon). The resultant soluble complexes were purified with a Superdex 200 (10/300 GL) gel filtration column (GE Healthcare) equilibrated with 50 mM Tris–HCl buffer (pH 8.0) containing 100 mM NaCl using the FPLC system.

2.5. Determination of the molar ratio of DP2(1–100) to DP2(792–1163) in the complex

The intensity of the stained protein bands on SDS–PAGE gel was measured with a ChemiDoc XRS scanning imager (Bio-Rad) equipped with the Quantity One ver. 4.4. software (Bio-Rad). The molecular ratio of DP2(1–100) to DP2(792–1163) in the complex was determined by calculations involving the scanning intensity and the molecular mass of each band.

3. Results and discussion

3.1. Crystal structure of DP2(48–291)

The crystal structure of DP2(1–300) was determined at 2.2 A resolution according to the MAD method using a selenomethionine (SeMet) derivative (Table 1). The refined model contains the residues (48–291) and 107 water molecules; i.e., the N-terminal and C-terminal ends of DP2(1–300) are absent. Hereafter, the crystal