Published in final edited form as:

Biomol NMR Assign. 2011 October ; 5(2): 151–155. doi:10.1007/s12104-010-9289-7.

Sequence-specific ^1H , ^{13}C and ^{15}N assignments of the phosphoesterase (PE) domain of *Pseudomonas aeruginosa* DNA ligase D (LigD)

Kaushik Dutta^{1,§}, Aswin Natarajan^{2,3,§}, Pravin A. Nair⁴, Stewart Shuman⁴, and Ranajeet Ghose^{2,3}

¹The New York Structural Biology Center, 89 Convent Avenue, New York, NY 10027

²Department of Chemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031

³The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016

⁴Molecular Biology Program, Sloan-Kettering Institute, New York, New York 10021, USA

Abstract

DNA ligase D (LigD), consisting of polymerase, ligase and phosphoesterase domains, is the essential catalyst of the bacterial non-homologous end-joining pathway of DNA double-strand break repair. The phosphoesterase (PE) module performs manganese-dependent 3'-phosphomonoesterase and 3'-ribonucleoside resection reactions that heal broken ends in preparation for sealing. LigD PE exemplifies a structurally and mechanistically unique class of DNA end-processing enzymes. Here, we present the resonance assignments of the PE domain of *Pseudomonas aeruginosa* LigD comprising the N-terminal 177 residues.

Keywords

DNA repair; non-homologous end-joining; DNA ligase

Biological context

Bacteria repair double-stranded DNA breaks (DSBs) by either of two distinct pathways – RecA-dependent homologous recombination (HR) or Ku-dependent nonhomologous end joining (NHEJ) (Shuman and Glickman, 2007). Bacterial NHEJ is catalyzed by a multifunctional DNA ligase D (LigD). *Pseudomonas aeruginosa* LigD consists of an N-terminal phosphoesterase (PE) domain (residues 1-187), a central ATPdependent ligase (LIG) domain (residues 188-527) and a C-terminal polymerase (POL) domain (residues 533-840). POL catalyzes the incorporation of deoxyribonucleoside or ribonucleoside monophosphates at the 3'-ends of DSBs prior to their sealing by LIG. Atomic structures have been determined for the POL and LIG domains of bacterial LigD proteins bound to

Address correspondence to: Ranajeet Ghose, Department of Chemistry, The City College of New York, 160 Convent Avenue, New York, NY 10031. Phone: (212) 650-6049; Fax: (212) 650-6107; rghose@sci.ccny.cuny.edu.

[§]KD and AN contributed equally to this work.

substrates at various functional states along their respective reaction pathways (Shuman and Glickman, 2007). On the other hand, only recently has the PE domain been characterized structurally (Nair et al., 2010).

The PE component of LigD is a 3' phosphodiesterase that removes the short 3'-ribonucleotide tracts generated by POL at 5'-overhang DSBs, yielding a 3'-monoribonucleotide end that is the preferred substrate for sealing by NHEJ ligases (Zhu et al., 2005). PE also has a 3' phosphomonoesterase activity that converts DNAs with non-ligatable 3'-PO₄ ends to 3'-OH ends that can be extended by POL or sealed by LIG. The 3' phosphodiesterase and 3' phosphomonoesterase reactions of LigD PE are characteristically dependent on Mn²⁺ (Mg²⁺ is ineffective in promoting catalysis).

We are applying the complementary methods of X-ray crystallography and solution NMR spectroscopy to obtain atomic structures of LigD PE and thereby illuminate its catalytic mechanism, substrate specificity, and conformational dynamics. We recently reported the 1.9 Å crystal structure of the active phosphodiesterase core of *Pseudomonas aeruginosa* PE (*PaePE*), spanning residues 30-187 (Nair et al., 2010). *PaePE* has a novel fold in which an 8 stranded β-barrel with a hydrophobic interior supports a crescent-shaped hydrophilic active site on its outer surface. Six essential side chains (His42, His48, Asp50, Arg52, His84, Tyr88) were found to coordinate a manganese ion and a sulfate mimetic of the scissile phosphate. The crystal structure accounted for the manganese requirement (insofar as two of the metal ligands were histidines) and showed that PE active site is unlike that of any known DNA/RNA endhealing enzymes or any known phosphotransferases (Nair et al., 2010).

A shortcoming of the X-ray structure was that it omitted the protease-sensitive N-terminal peptide of *PaePE* that includes several residues important for the 3' phosphomonoesterase activity (Zhu et al., 2005). Our aim in the present study was to obtain an NMR structure of an active version of *PaePE* that includes the N-terminal peptide and the core PE fold. After testing incremental truncations from the C-terminus of the PE module, we found that *PaePE*-(1-177) retained 3' phosphodiesterase and monoesterase activities comparable to that of *PaePE*-(1-187). We exploited *PaePE*-(1-177) for solution NMR studies, after finding that this construct provided spectra of higher quality than *PaePE*-(1-187), especially in the experiments required for the assignment of side chain ¹H resonances, a crucial step toward obtaining its solution structure. Here we report the main chain and side chain ¹⁵N, ¹H and ¹³C resonance assignments for *PaePE*-(1-177).

Methods and experiments

Production and purification of *PaePE*

A pET28b-based plasmid encoding the N-terminal 177 residues of *P. aeruginosa* LigD fused to an N-terminal His₁₀-SUMO tag, was transformed into *E. coli* BL21 (DE3)-RIPL cells. Cells from individual colonies were inoculated into 20 mL of M9 medium prepared in 100% H₂O. After overnight growth, cells were subsequently transferred stepwise into 20 mL M9 media containing successively increasing proportions of D₂O (25%, 50%, 75% and finally 100%). Cells grown overnight in M9-(100% D₂O) were inoculated into 1 L of M9-(100% D₂O) containing ¹⁵N-labeled ammonium chloride and ¹³C,²H-labeled glucose as the

only nitrogen and carbon sources, respectively. The cultures were incubated at 37° C with shaking at 250 rpm until the A_{600} reached 0.6, then 1 mM IPTG was added to induce *PaePE* production. The induced culture was incubated at 16° C for 16 h. The cells were then harvested by centrifugation and resuspended in 40 mL of lysis buffer that contained 50 mM Tris-HCl (pH 7.5), 0.5 mM NaCl, 15 mM Imidazole, 10% sucrose, 0.1% Triton X-100, and one crushed protease inhibitor tablet (Roche). All subsequent procedures were performed at 4° C. The cell suspension was lysed by sonication, followed by removal of the cell debris by centrifugation at 14500 rpm for 30 min. The soluble lysate was then passed through a TALON affinity column (Clontech) pre-equilibrated with lysis buffer. The protein was eluted with 40 mL of 50 mM Tris-HCl (pH 7.5), 0.5 mM NaCl, 500 mM Imidazole, 10% sucrose. SUMO-protease Ulp1 (20 units; a kind gift from Dr. Christopher Lima, SKI) was added to the imidazole eluate and the mixture was then dialyzed overnight in 1 L of a buffer containing 50 mM Tris-HCl (pH 8.0), 400 mM NaCl, 5 mM β -mercaptoethanol. The Ulp1-treated dialysate was applied to a second column of TALON beads, and the tag-free *PaePE*-(1-177) protein was recovered in the flow-through fraction. The flow-through fractions were pooled, concentrated and purified further by gel-filtration through a Superdex 200 column (GE Life Sciences) equilibrated with 50 mM Bis-tris (pH 6.5), 200 mM NaCl, 5 mM DTT. The peak fraction containing pure *PaePE*-(1-177) uniformly-labeled with ^{15}N , ^{13}C , ^2H was concentrated to 300 μM protein for use in NMR experiments. A similar procedure was followed to obtain ^{15}N , ^{13}C -labeled *PaePE*-(1-177) except that the pre-adaptation D_2O steps were eliminated and the M9 containing ^{13}C , ^1H -glucose (as the only carbon source) was prepared in 100% H_2O .

We also confirmed the absence of any bivalent metals ions that could have been acquired through the expression and purification process by noting that the ^{15}N , ^1H TROSY spectra acquired in the absence and presence of a 10-fold excess of EDTA in the buffer were identical. Thus all further experiments were carried out in the absence of chelating agents.

NMR spectroscopy

NMR experiments were performed at 25° C using either Bruker Avance (600 MHz or 800 MHz) or Varian Inova (600 MHz) spectrometers equipped with cryogenic probes capable of applying pulsed-field gradients along the z-axis. TROSY-based HNCACB, HN(CO)CACB experiments (600 MHz; 1024, 32 and 64 complex points and sweep-widths of 14 ppm, 32 ppm and 65 ppm in the ^1H , ^{15}N and ^{13}C dimensions respectively), HNCO and HN(CA)CO (600 MHz; 512, 32 and 32 complex points and sweep-widths of 15 ppm, 36.2 ppm and 14 ppm in the ^1H , ^{15}N and ^{13}C dimensions respectively) experiments were used for backbone-directed sequential assignments utilizing uniformly ^{13}C , ^{15}N , ^2H -labeled protein. A pre-scan delay of 1.5 s was used for all experiments. The assignment procedure was facilitated by residue type identification using HCCH-TOCSY (800 MHz; 512, 35 and 35 complex points and sweep-widths of 14 ppm, 64 ppm and 64 ppm in the ^1H , and the two ^{13}C dimensions respectively; mixing time 10.8 ms) and (H)C(CO)NH-TOCSY (600 MHz; 512, 32 and 64 complex points and sweep-widths of 15 ppm, 39.5 ppm and 76 ppm in the ^1H , ^{15}N and ^{13}C dimensions respectively; mixing time = 12 ms) experiments on uniformly ^{13}C , ^{15}N -labeled *PaePE*-(1-177). The data were processed with NMRPipe (Delaglio et al., 1995) and analyzed using NMRViewJ (Johnson, 2004). Main-chain and side-chain ^1H assignments

were obtained using a H(CCO)NH experiment (600 MHz; 512, 30 and 43 complex points and sweepwidths of 13 ppm, 32 ppm and 7.5 ppm in the direct ^1H , ^{15}N and indirect ^1H dimensions respectively; mixing time = 18 ms). Additional ^1H assignments were obtained using a ^{15}N -edited NOESY-HSQC experiment in H_2O (800 MHz; 512, 40 and 128 complex points and sweep-widths of 14 ppm, 32 ppm and 12 ppm in the direct ^1H , ^{15}N and indirect ^1H dimensions respectively; mixing time = 150 ms) and a ^{13}C -edited NOESY-HSQC experiment in D_2O (800 MHz; 512, 38 and 99 complex points and sweep-widths of 14 ppm, 64 ppm and 10.5 ppm in the ^1H , ^{13}C and ^1H dimensions respectively; mixing time = 150 ms).

Resonance assignments and data deposition

146 of 164 (excluding the Met residue at the extreme N-terminus) main-chain amide ^{15}N , $^1\text{H}^{\text{N}}$ (89 %), 153 of 165 C_α (93 %), 139 of 150 C_β (93 %) and 146 of 165 C' (89 %) resonances could be unambiguously assigned for the non-proline residues. For the 12 prolines, 11 of the 12 C^α , C^β and C' resonances could be assigned. Further, 60 of 100 (60 %) C^γ , 36 of 66 (55 %) C^δ and 5 of 14 C^ϵ (36 %) were assigned (including proline but excluding aromatic side-chains). For ^1H resonances, 89 % of H_α , 86 % of H_β , 73 % of H_γ , 76 % of H_δ and 7 % of H_ϵ were assigned (including prolines but excluding aromatic sidechains). The side-chain N_δ , H_δ of the only Asn, 1 of 6 Gln N_ϵ , H_ϵ and 2 of 5 Trp side-chain N_ϵ , H_ϵ pairs were assigned. 34 % of aromatic protons were assigned. The completeness of resonance assignments was highly satisfactory given the quality of spectra that could be obtained even for this optimal construct (Fig. 1).

Of the 17 residues for which ^{15}N , $^1\text{H}^{\text{N}}$ chemical shifts could not be assigned, 10 (Asp162, Gly163, Glu164, Ala165, Ser167, Leu168, Asp169, Asp172, Val173 and Lys175) belonged to the C-terminal region that is dispensable for the enzymatic activity (Nair and Shuman, unpublished results). Incomplete back-exchange of amide protons was unlikely to be the cause of the missing assignments, because identical ^{15}N - ^1H HSQC spectra were recorded for samples prepared in H_2O or in D_2O minimal media. However, it is possible that some of these missing resonances were present at the crowded region of the spectra (see Fig. 1) and could not be sufficiently resolved to allow unambiguous assignment. Of the residues important for catalytic activity (Zhu et al., 2005), assignments for main-chain ^{15}N , $^1\text{H}^{\text{N}}$ resonances for His42, His48, and Tyr88 could not be obtained. It is likely that these residues are in conformational exchange in the absence of the catalytic Mn^{+2} ion that is essential (Zhu and Shuman, 2005) for both the phosphomonoesterase and phosphodiesterase activities. Mn^{+2} was not utilized in the present studies because of its paramagnetic nature and resultant line-broadening effects.

Secondary structural elements were identified using the TALOS+ (Shen et al., 2009) software utilizing $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}'$ and ^{15}N chemical shifts after correction of $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts for ^2H -isotope effects (Venters et al., 1996). Thus, eight β strands (β_1 :34-41, β_2 :49-54, β_3 :58-63, β_4 :76-80, β_5 :106-115, β_6 :127-133, β_7 :141-145 and β_8 :154-159) were identified in addition to single α -helix (120-124). These regions of secondary structure were consistent with those seen in the crystal structure (Nair et al., 2010) as shown in Fig. 2.

We noted another interesting feature upon careful analysis of the ^{15}N , $^1\text{H}^{\text{N}}$ TROSY spectra - two sets of peaks (a major and a minor peak) corresponding to residues Leu34, Leu55, Gly57, Thr58, Leu59, Trp62, Ile106, Val107 and Trp108 were seen in a freshly prepared sample. Over time, the minor peaks disappeared and only the major peak could be detected, indicating a conformational transition that was very slow on the chemical shift timescale. These residues were located near the N-termini of β_1 and β_5 , and the region between β_2 and β_3 (Fig. 2). Inspection of Fig. 2 reveals the confidence in strand prediction by TALOS+ to be quite low at the N-termini of β_1 and β_3 . This is consistent with conformational dynamics on the slow to very slow timescale. Additionally, the same analysis seems to suggest a break in β_5 around residue 109 (Fig. 2). This indicates conformational heterogeneity around the edges of one face (β_5 - β_1 - β_2 - β_3) of the β -barrel (Nair et al., 2010).

^{13}C chemical shifts are key indicators of the conformation of the peptide backbone about a Xaa-Pro bond. An exhaustive statistical analysis of the differences between $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\gamma}$ chemical shifts (β_{γ}) for *trans* and *cis* conformations of the Xaa-Pro peptide bond were 4.5 ± 1.2 ppm and 9.6 ± 1.3 ppm, respectively (Schubert et al., 2002), suggesting a significantly larger difference for a *cis* conformation. For Pro65 and Pro72, the β_{γ} values were both 8.4 ppm after correcting for ^2H isotope shifts (Venters et al., 1996). Using the program Promega (Shen and Bax, 2009), we found that the absolute probabilities of the Val64-Pro65 and Asp71-Pro72 bonds being in the *cis* conformation were 91% and 86%, respectively. Notably, both these bonds are in the *trans* conformation in the crystal structure (Nair et al., 2010). All other prolines for which $^{13}\text{C}^{\beta}$ and $^{13}\text{C}^{\gamma}$ assignments were available (this excludes Pro68) were all predicted to be in the *trans* conformation.

The N-terminal 32-aa peptide of *PaePE*, which is important for the phosphomonoesterase activity, was disordered on the ps-ns timescale with average ^{15}N - $\{^1\text{H}\}$ NOE values of 0.03 ± 0.28 . However, an analysis of the secondary structure propensities using SSP scores (Marsh et al., 2006) revealed significant helical tendencies ($>28\%$) for residues 6-11 (data not shown). S^2 values for this region (0.45 ± 0.02) predicted from the Random Coil Index (RCI) (Fig. 2) indicated some degree of order, though a higher degree of order (0.58 ± 0.07) was predicted for residues 14-21, a segment that includes three side chains (Arg14, Asp15 and Glu21) shown to be important for the 3'-phosphomonoesterase activity but not for the 3'-phosphodiesterase activity (Zhu et al., 2005). The ^{15}N - $\{^1\text{H}\}$ NOE values for this segment were also marginally higher (0.16 ± 0.04) than the rest of the disordered N-terminus.

A detailed analysis of the structural and dynamic features of *PaePE*-(1-177) in solution and its interactions with substrate DNA will be discussed at length elsewhere (Natarajan et. al., in preparation).

The main-chain ^1H , $^{13}\text{C}^{\alpha}$, $^{13}\text{C}'$ and ^{15}N and side-chain ^{13}C and ^1H chemical shifts of *PaePE*-(1-177) have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) under accession number 17283. Corrections to account for the TROSY and ^2H -isotope shifts were not applied to the deposited shifts.

Acknowledgments

This work was supported by the following grants: NSF MCB 083141 (RG); NIH GM63611 (SS) and NIH 5G12 RR03060 (partial support of the core facilities at CCNY). SS is an American Cancer Society Research Professor. RG is a member of the New York Structural Biology Center, a STAR center supported by the New York State Office for Science, Technology and Academic Research.

References

- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995; 6:277–93. [PubMed: 8520220]
- Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Meth Mol Biol*. 2004; 278:313–52.
- Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between alpha- and gammasynuclein: implications for fibrillation. *Protein Sci*. 2006; 15:2795–804. [PubMed: 17088319]
- Nair PA, Smith P, Shuman S. Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily. *Proc Natl Acad Sci USA*. 2010; 107:12822–7. [PubMed: 20616014]
- Schubert M, Labudde D, Oschkinat H, Schmieder P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ^{13}C chemical shift statistics. *J Biomol NMR*. 2002; 24:149–54. [PubMed: 12495031]
- Shen Y, Bax A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR*. 2009; 46:199–204. [PubMed: 20041279]
- Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS $_{\alpha}$: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009; 44:213–23. [PubMed: 19548092]
- Shuman S, Glickman MS. Bacterial DNA repair by non-homologous end joining. *Nature Rev Microbiol*. 2007; 5:852–61. [PubMed: 17938628]
- Venters RA, Farmer BT 2nd, Fierke CA, Spicer LD. Characterizing the use of perdeuteration in NMR studies of large proteins: ^{13}C , ^{15}N and ^1H assignments of human carbonic anhydrase II. *J Mol Biol*. 1996; 264:1101–16. [PubMed: 9000633]
- Zhu H, Shuman S. Novel 3'-ribonuclease and 3'-phosphatase activities of the bacterial non-homologous end-joining protein, DNA ligase D. *J Biol Chem*. 2005; 280:25973–81. [PubMed: 15897197]
- Zhu H, Wang LK, Shuman S. Essential constituents of the 3'-phosphoesterase domain of bacterial DNA ligase D, a nonhomologous endjoining enzyme. *J Biol Chem*. 2005; 280:33707–15. [PubMed: 16046407]

Fig. 1.

Assigned ^{15}N , ^1H TROSY spectrum of uniformly ^{13}C , ^{15}N , ^2H -labeled *PaePE*-(1-177) in 50 mM Bis-Tris propane (pH 6.5), 200 mM NaCl, 5 mM DTT. Data were acquired at a ^1H frequency of 600 MHz at 25° C using sweep-widths of 15 ppm (512 complex points) in ^1H dimension and 36.2 ppm (128 complex points) in ^{15}N dimension with a recycle delay of 1.5 s. The crowded central region indicated by the dashed rectangle, is shown in the inset. Assigned backbone ^{15}N , ^1H resonances have been labeled. Sidechain NH groups have not been labeled.

Fig. 2.

Secondary structure prediction using TALOS+ (Shen et al., 2009). The confidence values for regions predicted to be β -strand (black) or α -helix (blue) are plotted against residue numbers. Also shown are the predicted order parameters (S^2) based on chemical shifts. Residues that show peak doubling (see main text) are indicated by green arrows. Secondary structure elements seen in the crystal structure of *PaePE*-(30-187) (PDB: 3N9B) are shown (β -strands as black rectangles and α -helices as blue rectangles) and labeled.