STRUCTURE/FUNCTION CORRELATIONS IN *PSEUDOMONAS AERUGINOSA* DNA LIGASE LIGD

by

ASWIN NATARAJAN

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Biochemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

	Dr. Ranajeet Ghose
Date	Chair of Examining Committee
	Dr. Edward J. Kennelly
Date	Executive Officer

Dr. Ranajeet Ghose, City College

Dr. Marilyn Gunner, City College

Dr. Iban Ubarretxena-Belandia, Mount Sinai School of Medicine

Dr. Alexander Shekhtman, University at Albany, SUNY

Dr. David Cowburn, Albert Einstein School of Medicine

Dr. Patrick Loria, Yale University

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

STRUCTURE/FUNCTION CORRELATIONS IN *PSEUDOMONAS AERUGINOSA* DNA LIGASE LIGD

by

Aswin Natarajan

Advisor: Professor Ranajeet Ghose

The ATP-dependent DNA ligase D (LigD) performs a major role in the non-homologous end-joining (NHEJ) pathway. *Pseudomonas aeruginosa* LigD contains a N-terminal phosphoesterase domain (PE) domain followed by a ligase domain and a C-terminal polymerase domain. The PE domain (187 residues), belonging to a class of unique 3'-end-processing enzymes, possesses manganese dependent phosphodiesterase and phosphomonoesterase activities as it sequentially removes the 3'-ribonucleoside from the primer strand of the primer-template DNA duplex and hydrolyzes the 3'-PO_4 produced finally to a 3'-OH group. Extensive mutagenesis and biochemical studies have identified critical residues and important features required for 3'- ribonuclease and 3'– phosphatase activities. Lack of sequence homology to other known nucleases lead to the belief that this enzyme possesses some unique motifs. However, in the absence of atomic level structural information clear structure/function correlations were lacking. This thesis describes the procedures used to obtain a high-resolution structure of PE domain obtained using solution NMR methods and to ascertain its interaction with DNA substrates.

Acknowledgements

First and foremost, I would like to thank my mom, dad, sister and grandma for their love and support through all walks of my life. I thank them for allowing me to pursue my dream and I am sure they are proud of my efforts.

I thank Dr. Ronnie Ghose for taking me under his mentorship and guiding me in each and every step in my research. He showed tremendous confidence in me for which I will always be grateful.

I thank the members of Ghose's Laboratory - Dr. Andrea Piserchio, Dr. Ertan Eryilmaz, Deniz Temel, Zhen Ren, Dr. Venkatesh Ramakrishnan, Dr. Sabestian Alphonse and Benjamin Reed. It was an enjoyable experience working with them. A special thanks to my elder sister Deniz, for her unconditional love and support.

I thank Dr. Stewart Shuman (Memorial Sloan Kettering Institute) who came up with the idea about this project and for his support throughout the years. A special thanks to Dr. Kaushik Dutta (New York Structural Biology Center) who guided me in each and every step in this research work.

I thank my committee members Dr. Marilyn Gunner, Dr. Iban Ubarretxena-Belandia, Dr. Alexander Shekhtman, Dr. David Cowburn and Dr. Patrick Loria for their valuable guidance throughout the years.

A special thanks to my roomies, Dr. Vijaishankar Balachandran and Sudharshan Devaraju whose support has taken me to great heights. I thank my friends who migrated with me to US for pursuing our dreams and those back in India for their support.

Table of Contents

List of Tables

Lists of Figures

1. Introduction

1.1. Double strand breaks in cells

Maintaining genomic stability is crucial for cell survival. About 10 double strand breaks (DSBs) occur in an eukaryotic cell per day (Lieber 2010). Even a single DSB can cause genomic instability and lead to aneuploidy, chromosomal rearrangements, carcinogenesis and cell death (Pitcher et al, 2007). DSBs, caused by various factors, can be broadly categorized into multiple damaged sites and free DSB ends. DSBs with multiple damage sites are caused by ionizing radiation (IR), radiomimetic agents, drugs, enzymatic degradation and reactive oxygen species. Usually these DSBs possess overhangs that can weakly reassociate to start the end joining process. However, the ends may not carry the 5'-PO_4 or 3'-OH that are needed for direct ligation. So, some forum of repair machinery is required to facilitate the end joining.

Free-end DSBs are products of failed cellular processes like DNA replication and mitosis (Daley et al., 2005). Presence of single-strand breaks during replication can lead to formation of free end DSB due to polymerase stalling. Failed telomerase, topoisomerase or other nuclear enzyme action can produce free end DSBs which are exposed to further damaging factors. Sheer stress caused by pulling during mitosis can also lead to free end DSBs (Daley et al, 2005). Some kind of cellular machinery is needed for pre-processing and ligation of these free end DSBs as they are not good substrates for direct reassociation.

1.2. Non-homologous End Joining

Cells have evolved two independent processes to repair DSBs – homologous recombination (HR) and non-homologous end joining (NHEJ) (Pitcher et al, 2007). Homologous recombination occurs during the later stages of the cell cycle, namely the S and G2 phase. The presence of the identical sister chromatid, a product of DNA replication, is required to act as a template for the repair machinary. 5'to 3' exonucleases act on either one or both the DSBs to generate a 3'-OH overhang which then invades the sister chromatid. This invasion is facilitated by RecA protein in prokaryotes or Rad51 protein in eukaryotes (Shuman, Glickman 2007). The DNA polymerase enzyme uses the invading strand as a primer and one of the strands of the invaded sister chromatid as a template to generate the damaged strand. Once enough DNA is generated on the broken strand it is ligated back to the parent strand by the replicative DNA ligase. This type of repair mechanism usually does not lead to any loss of genetic information.

NHEJ occurs throughout the cell cycle but predominantly during the G0 and G1 phases of the cell cycle when there is no identical sister chromatid present (Pitcher et al, 2007). This phenomenon does not require extensive homology and results in direct joining of the broken ends. This process is sometimes mutagenic if the broken ends are processed by the nucleases that are part of the repair machinary. These mutations are not always detrimental as it provides much needed diversity for genes involved in immune responses (Shuman, Glickman 2007). NHEJ plays a crucial role in class switch recombination and variable (diversity) recombination.

NHEJ can be error-free if a DNA ligase can directly ligate broken ends which are fully complementary to each other. This simple religation cannot occur for damaged ends. In some cases, the broken DNA strands are aligned based on microhomology and the incompatible bases are subsequently resected, the gaps extended and ligated. Incompatible ends that arise from different DSBs can undergo this process. The most fascinating aspect of NHEJ is the variety of substrates that can be acted upon. This variation is provided by the robustness, flexibilty and independent functionality of the enzymes involved (Lieber 2010).

1.3. NHEJ in Eukaryotes

NHEJ system of DNA repair was first identified in eukaryotes (Lieber 2010). The pathway involves multiple protein system that includes a Ku 70/80 heterodimer, ligase LXX complex, DNA-PKCs, the MRX complex, polymerases, phosphatases and nucleases. Ku (Blier et al, 1993) represents the first response unit after a DSB occurs in a eukaryotic cell. It binds to the damaged ends with high affinity in a sequence independent manner to protect it and thereby forming the Ku:DNA complex (**Figure 1**). This complex at each damaged end acts as a recruiter for the polymerase, ligase and nuclease complexes (Lieber 2008). The Artemis:DNA-PKcs in higher eukaryotes (Ma et al, 2002) and MRX complex (Mre11:Rad50:Xrs2) (Daley et al, 2005) in yeast perform the necessary nuclease action to rectify the damaged DNA ends. Pols μ and λ (Pol4 in yeasts) (Lieber 2010) perform the necessary polymerase activities by adding bases in a template-independent or dependent manner. The XLF:XRCC4:DNA ligase IV complex

(Nej1:Lif1:Dnl4 in yeasts) can ligate simple nicks, across gaps, incompatible ends and complex structured ends (Gu et al, 2007).

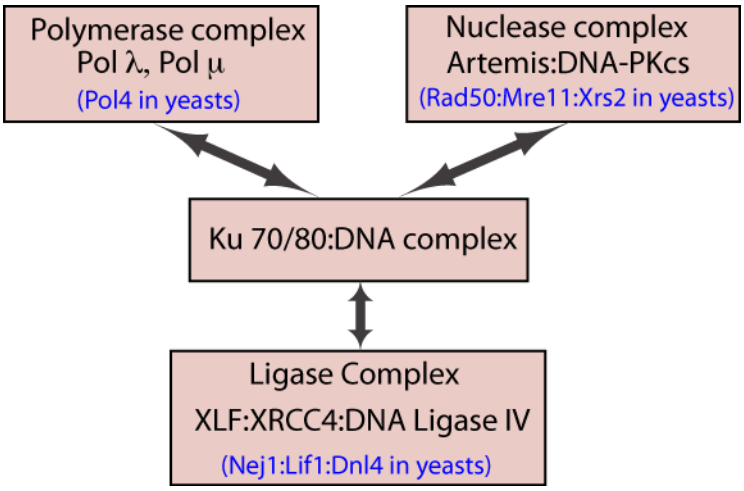

Figure 1: Illustration showing the various complexes involved in NHEJ in eukaryotes.

1.4. NHEJ in Prokaryotes

Irrefutable evidence about the presence of NHEJ pathway in bacteria was obtained when a transfected linear plasmid with blunt ends or overhangs were found to circularize in *Mycobacterium* (Gong et al, 2005). A two- protein NHEJ system, comprising of Ku and DNA ligase in bacteria was first identified based on phylogenetic analysis (Aravind, Koonin 2001). Putative Ku homologs (~30-40 kDa) coexpressed with putative DNA ligase genes were identified in several bacterial genomes (Weller, Doherty 2001). Interestingly, Ku-like proteins have been identified across the three domains of life (Pitcher et al, 2007) suggesting a role of evolution in creating a more complex NHEJ machinary centered on Ku for higher organisms. Bacterial Ku exists as a homodimer and binds to double stranded linear DNA (Weller et al. 2002). It lacks the von Willebrand

factor A and SAP domains seen in its eukaryotic counterpart (Pitcher et al, 2007). **Figure 2** compares the structure of bacterial Ku and its eukaryotic counterpart.

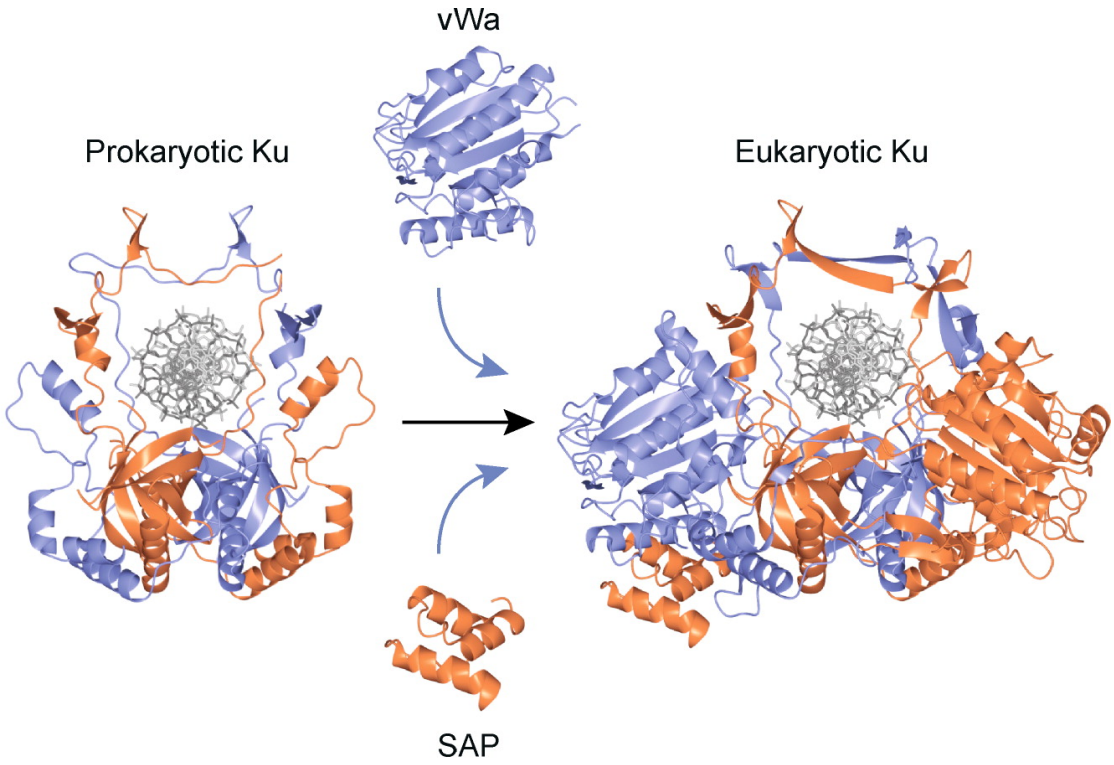

Figure 2: Comparison of Ku 70/80 and Prokaryotic Ku homodimers (modelled from Ku 70/80 structure). Figure modified from Pitcher et al, 2007.

The other indication of functional NHEJ in bacteria was the presence of DNA ligase genes. LigB, LigC and LigD are the 3 ATP-dependent ligases seen in bacteria and are dispensable for cell growth and survival (Weller et al, 2002). Of these, LigC and LigD have been shown to be involved in NHEJ (Gong et al, 2005). The presence of these 2 ligases differs between *Pseudomonas aeruginosa* (1 LigD, no LigC), *Mycobacterium tuberculosis* (1 LigD, 1 LigC), *Agrobacterium tumefaciens* (2 LigD, 3 LigC) (Shuman, Glickman 2007). The ping-pong mechanism of action of DNA ligase involves a 3 step process. It begins with the formation of a ligase-adenylate intermediate following the attack of α-phosphorus of ATP on Lys side chain of the ligase. AMP is transferred from

the Lys to the 5'-PO_4 of the DNA strand in the next step. The final step is the attack of 3'-OH of the nicked DNA on the DNA-adenylate to form the phosphodiester bond and the AMP moiety is released (Shuman, Glickman 2007).

1.5. DNA ligase LigD and its constituents

LigC is a single domain protein with minimal activity (Gong et al, 2004; Zhu, Shuman 2007) whereas LigD is a multidomain protein composed of polymerase, ATP-dependent ligase and phosphoesterase domains (Della et al, 2004; Zhu, Shuman 2005; Zhu, Shuman 2007). The arrangement of the domains varies between different bacterial species with some of them possessing just a ligase domain or ligase and polymerase domains. *P. aeruginosa* possesses a N-terminal phosphoesterase domain followed by a ligase domain and a C-terminal polymerase domain whereas *M. tuberculosis* has a N-terminal polymerase domain, a central phosphoesterase domain and a C-terminal ligase domain (**Figure 3**).

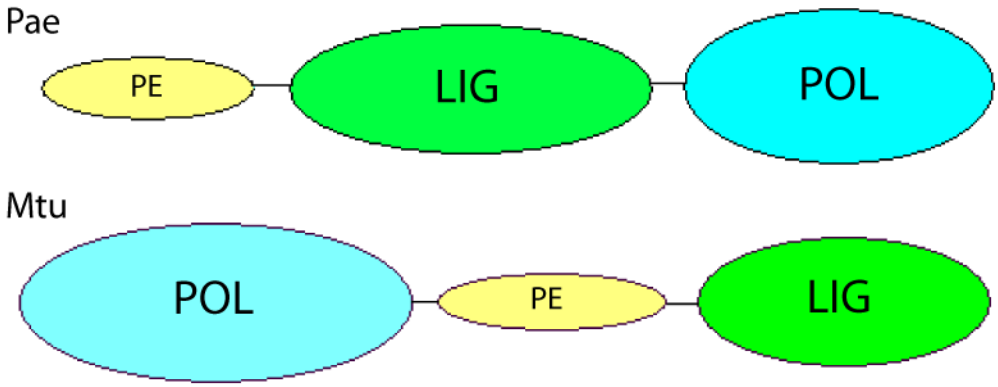

Figure 3: Domain organization of LigD in *Pseudomonas aeruginosa (Pae)* and *Mycobacterium tuberculosis (Mtu).*

Lig D polymerase (POL) possesses unique Mn^{2+}- dependent terminal transferase activities as it can extend ssDNA, blunt end dsDNA in a template independent manner and dsDNA with 5'- overhangs in a template dependent manner (Pitcher et al, 2007). The POL domain preferentially adds NTPs over dNTPs, which is understandable since NHEJ primarily occurs during stationary phase when the dNTP pool is limiting (Della et al, 2004). Up to four ribonucleotide bases can be added to the blunt end primer-template DNA duplex by the polymerase (Zhu, Shuman 2005). During template dependent synthesis, the primer is extended to the end of the template strand and then further extended with a single deoxyribonucleotide in a primer-template duplex (Zhu, Shuman 2005). Interestingly, template dependent synthesis in a primer-template duplex with a 5'-overhang in the presence of rNMPs only yielded primers that had just four added ribonucleotides. This activity illustrates a need for deoxyribonucleotides in the primer strand (Zhu, Shuman 2005). Moreover, LigD POL can fill in short single-stranded gaps in a template dependent manner, fill opposite modified bases and effectively bypass abasic sites. In all the forementioned cases, rNTPs are preferentially added instead of dNTPs (Pitcher et al, 2007). POL can also add incorrect bases to a primer-template duplex if the correct rNTP is missing (Shuman, Glickman 2007).

The POL domain contains unique catalytic motifs (I-III) seen in archaeo-eukaryotic primase (AEP) superfamily (Weller, Doherty 2001). Crystal structures of *Mtb*POL (Pitcher, Andrade et al, 2007) and *Pae*POL (Zhu et al, 2006) show conserved aspartate residues in motifs I and III that are involved in coordinating the metal binding site and conserved residues in motif II that are involved in nucleotide binding. It contains eleven β-strands arranged as two central sheets in the core and seven α-helices in the

periphery (**Figure 4**). The fold resembles the structure seen in archeal DNA primase (Augustin et al, 2001) family and is unlike eukaryotic PolX polymerase superfamily (Ramadan et al, 2004). The protein makes contact with the O2' and O3' of the ribose moiety of ATP in the crystal structure of *Pae*POL which indiates the preference of rNMP over dNMP (Zhu et al, 2006; Pitcher, Andrade et al, 2007). Ku recruits LigD to DSB site by interacting with the POL domain. POL has a strong tendency to bind 5'-PO_4 moiety at the broken ends and together with Ku brings the DSB sites together before the end-healing process (Pitcher et al, 2005).

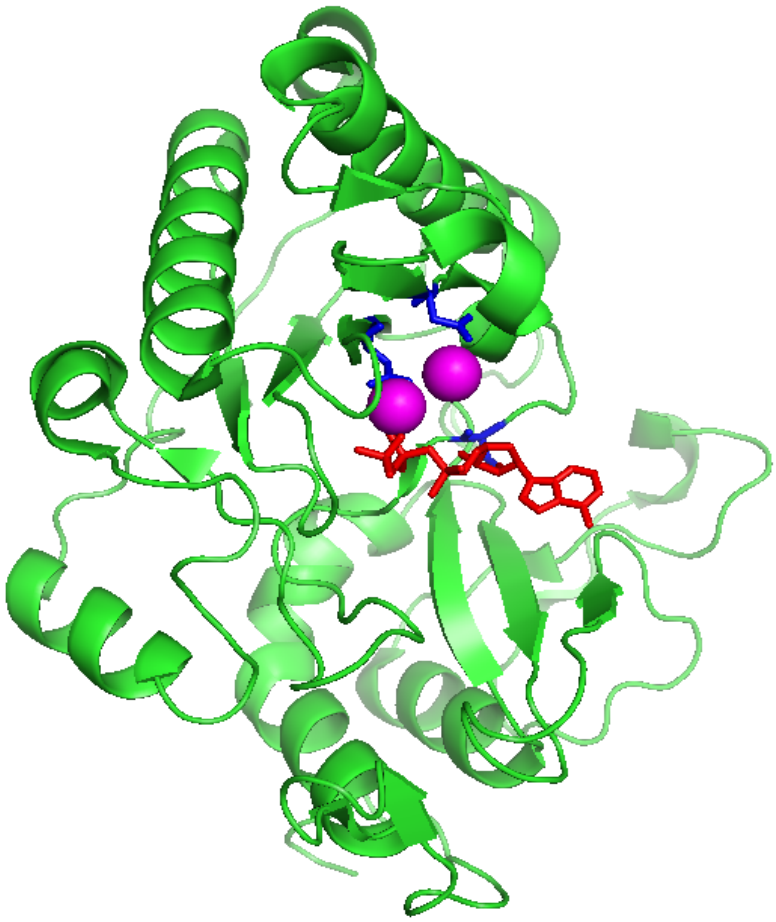

Figure 4: Crystal structure of *Pseudomonas aeroginosa* POL domain at 1.5 Å (2FAQ). Cartoon shows secondary structure elements in green, ATP in red and manganese ions in magenta spheres. The active site Asp669 and Asp759 coordinated by Mn ions and Ser768 interacting with ribose moiety of ATP are shown in blue.

The ligase domain (LIG) is responsible for the ATP-dependent adenylation and sealing of DSBs. The crystal structure of *Mtb*LIG (Akey et al, 2006) bound with AMP shows an adenylation domain and an oligo binding domain similar to other ligases. The adenylation domain consists of core of β-strands surrounded by α-helices, and the oligonucleotide binding domain consists of a 5-stranded β-barrel and a single α-helix (**Figure 5**). Interestingly, the active site mutant (K481A) does not primarily abolish NHEJ *in-vivo* suggesting a backup role of LigC for ligation. An important feature was the requirement for a 3'-OH ribonucleotide at the break site for nick sealing by both LigD and LigC (Zhu, Shuman 2008). This again compliments the preferences shown by POL and phosphoesterase domains for ribonucleotides. Ku has been shown to enhance LIG activity although direct interaction between them is not shown (Akey et al, 2006).

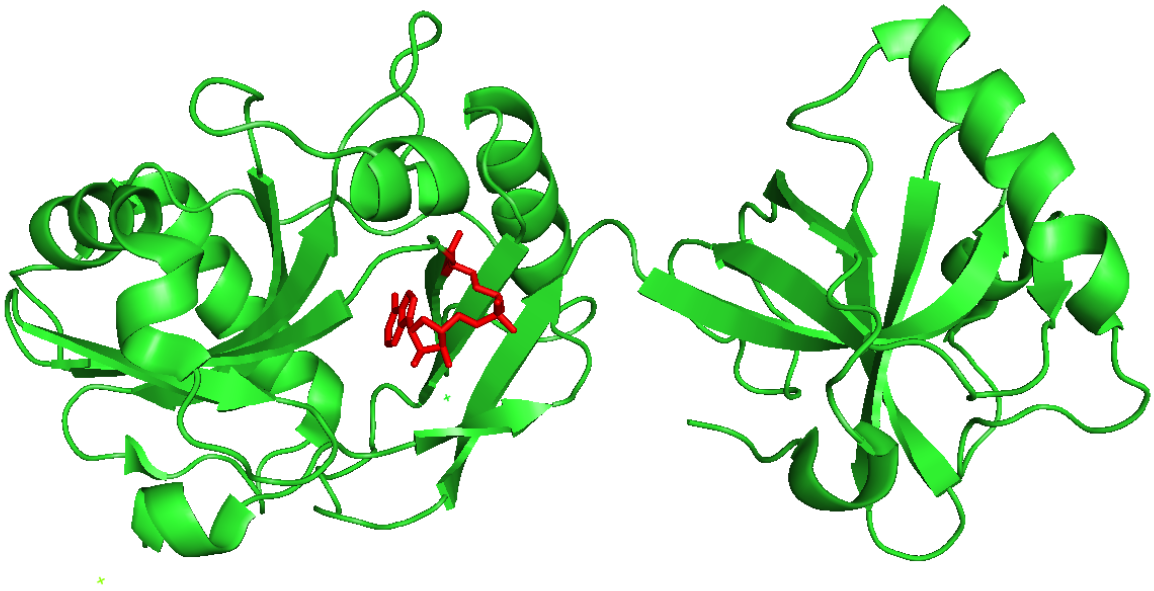

Figure 5: Crystal Structure of Mycobacterium tuberculosis LIG domain at 2.4 Å (1VS0). Cartoon shows secondary structure elements in green with adenylation domain in the right and oligo binding domain in right and AMP conjugated to Lys481 in red.

1.6. Phosphoesterase domain of LigD

The phosphoesterase (PE) component of *P. aeruginosa* LigD possess both phosphodiesterase (ribonuclease) and phosphomonoesterase (phosphatase) activities. The 3' phosphodiesterase and 3' phosphomonoesterase reactions of LigD PE are characteristically dependent on Mn^{+2} (Mg^{2+} is ineffective in promoting catalysis). As a 3' phosphodiesterase, it removes the short 3'-ribonucleotide tracts generated by POL at 5'-overhang DSBs, yielding a 3'-PO_4 monoribonucleotide. The 3'-phosphodiesterase seems to work sequentially by removing ribonucleotides from the primer-template duplex one by one instead of a dinucleotide. The inability to remove beyond the last ribonucleotide indicates the need for 3'-ribonucleotide as the penultimate base in the primer-template duplex. Modification of the 2'-OH of the penultimate base to 2'-F, 2-NH_2, or 2'-OCH_3 abolished the phosphodiesterase activity. Moreover, an increase in the length of the 5'-overhang had a positive effect on phosphodiesterase activity. Another interesting feature is the preference to work on a duplex rather than single strand DNA (Zhu, Shuman 2005; Zhu et al, 2005; Zhu, Shuman 2006).

PE also has a 3' phosphomonoesterase activity that converts DNAs with non-ligatable 3'-PO_4 ends to 3'-OH ends that can be extended by POL or sealed by LIG. This activity is highly crucial as breaks with 3'-PO_4 are often created by nucleolytic cleavage or ionizing radiation. This 3'-PO_4 is not the ideal substrate for either POL or LIG enzyme. But unlike phosphodiesterase, the phosphomonoesterase shows no preference for the presence of ribonucleotide as the penultimate base in the primer-template duplex. The preference for duplex DNA and lengthy 5'-overhang is also seen for 3'-phosphomonoesterase activity (Zhu, Shuman 2005; Zhu et al, 2005).

Mutational studies on the PE domain from *Pseudomonas aeroginosa* ligD have identified key residues that have an effect on each of the dual functions of PE (Zhu, Shuman 2006). His42, His48, Asp50, Arg52, His84 and Tyr88 (Group 1) contribute to both the monoesterase and disesterase activities; Arg14, Asp15, Glu21 and Glu82 (Group 2) contributed solely to the monoesterase activity; and Lys66 and Arg76 (Group 3) only to the diesterase activity.

The activity of PE domain in NHEJ and the overall mechanism of NHEJ in blunt end joining has been clearly illustrated in **Figure 6** and **Figure 7** respectively.

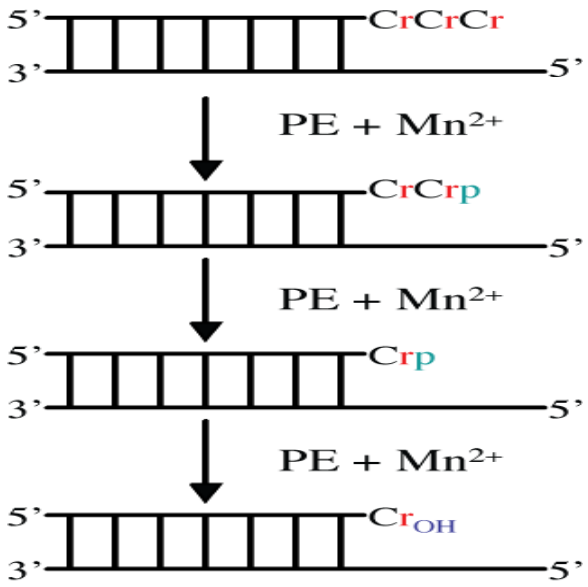

Figure 6 : PE resects ribonucleotide bases added to a primer-template duplex by the POL domain. The resection occurs in a sequential manner leaving a 3'-PO_4 at every step. The phospho-monoesterase activity takes over at the last base and the 3'-PO_4 is converted to 3'-OH.

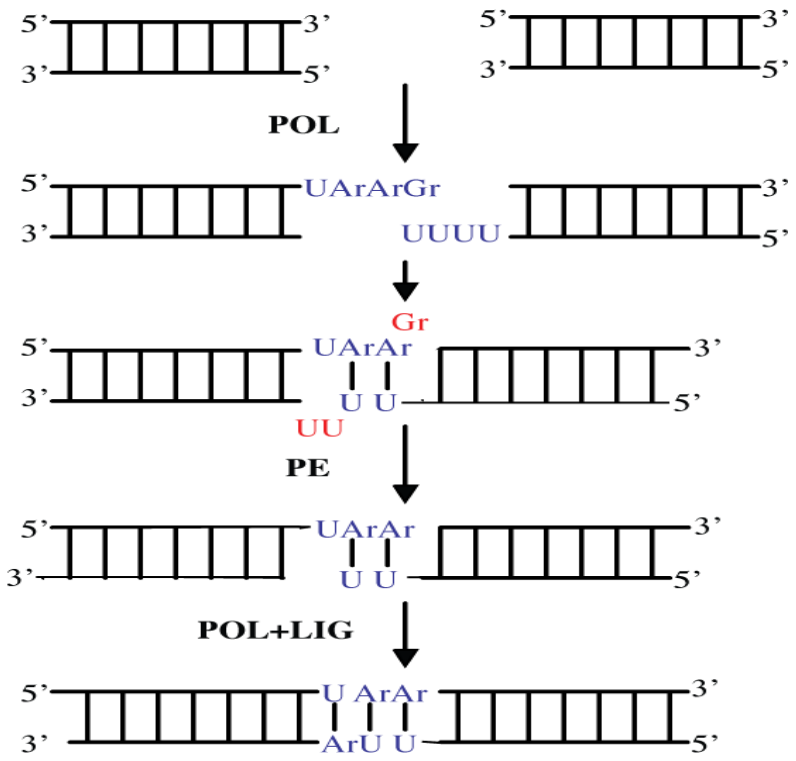

Figure 7: Mechanism of NHEJ in healing blunt end DSBs. The LigD POL adds few non-templated ribonucleotide bases to the blunt ends. The overhangs generated are base paired based on the microhomology between them. The non complimentary bases are resected by the PE domain to leave a 3'-OH that can be acted upon by the POLor LigD LIG domains. The gap is filled by the POL and LIG seals the nick to complete the repair.

Figure 8 depicts the overall sequence alignment of PE domains from various bacterial and archeal species. The alignment indicates the presence of highly conserved histidines and aspartates residues that form some kind of metal coordinating β-stranded structure (Aravind, Koonin 2001). Most of the conserved residues has been shown to be involved in the activity by mutagenesis (Zhu, Shuman 2005).

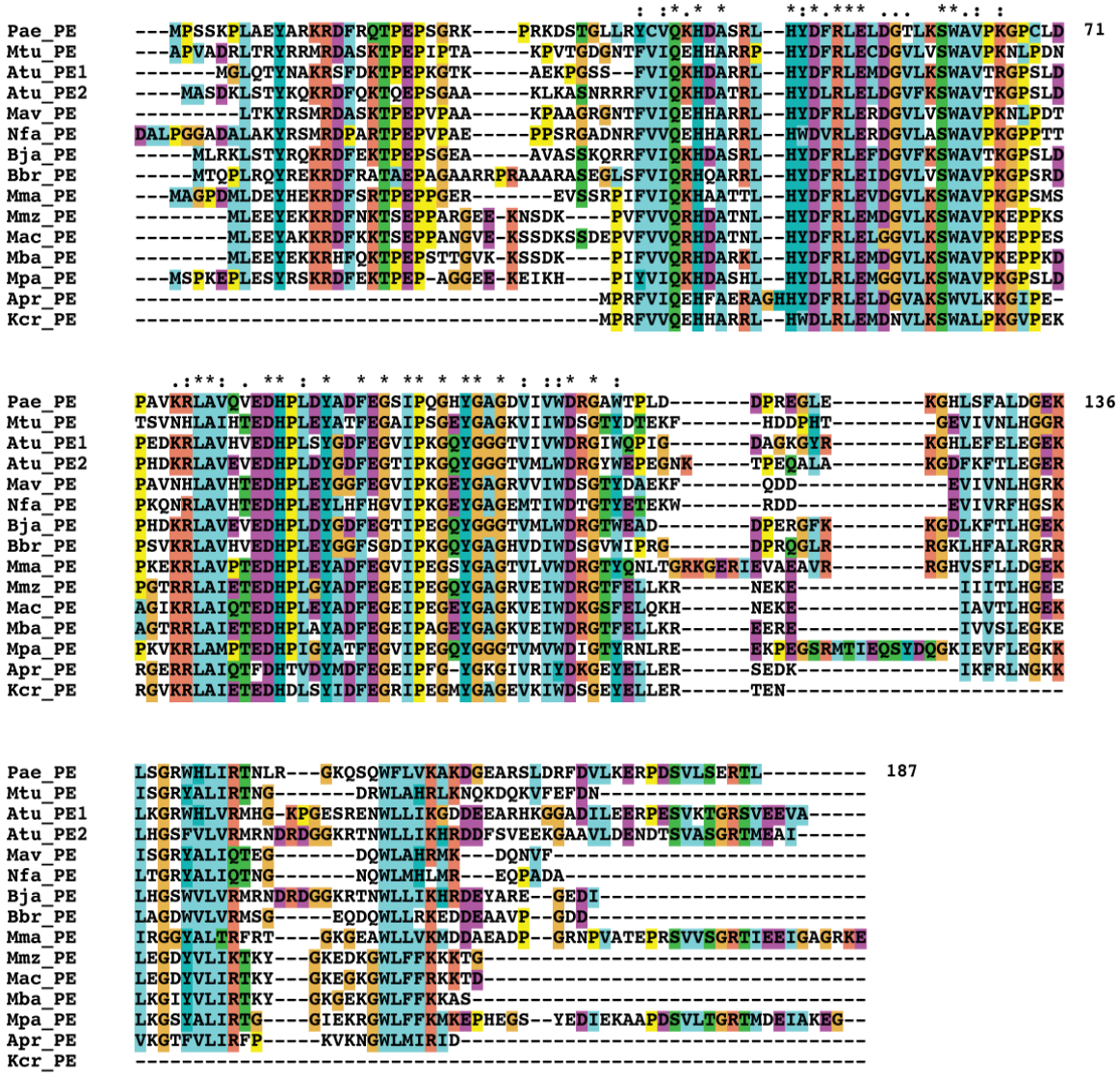

Figure 8: Sequence alignment of phosphoesterase domain of *Pseudomonas aeruginosa* and related domains from other bacteria (*Mycobacterium tuberculosis, Agrobacterium tumefaciens, Mycobacterium avium, Nocardia farcinica, Bradyrhizobium japonicum, Bordetella bronchiseptica*) and archeal species – (*Methanoculleus marisnigri, Methanosarcina mazei, Methanosarcina acetivorans, Methanosarcina barkeri, Methanocella paludicola, Archaeglobus profundus, Korarchaeum cryptofilum*). Conserved residues are indicated by asterix (*) and alignments gaps by dashes (-).

The PE component of LigD differs from other known ribonucleases in many notable ways suggesting that it is an unique end-healing enzyme. It differs from members of the RNase-A family due to the nature of its metal cation requirement for activity. Moreover, members of the RNase-A family do not possess 3'-monoesterase activity. PE

also differs from other metal dependent 3'-exonucleases of DEDD and RBN family (Zhu, Shuman 2005) because these enzymes directly yield 3'-OH. The phosphatases from T4 bacteriophage, mycobacteriophage and baculovirus differ from PE as they rely on Mg^{2+} as a cofactor and show no requirement for a template strand (Zhu, Shuman 2005).

The lack of sequence homology and mechanistic similarity to other phosphoesterase enzymes suggests that the PE module in LigD belongs to a unique class of nucleic acid end-processing enzymes. A significant gap in understanding fully all aspects of the functional regulation of LigD was the lack of structural information on the PE domain. I used solution NMR techniques to obtain the high-resolution structure of the *P. aeruginosa* PE domain and determine its mode of recognition of substrate DNA.

1.7. NMR spectroscopy

Nuclear Magnetic Resonance (NMR) is one of the principle techniques used in the field of structural biology (Greenbaum, Ghose 2010). Structure determination of proteins, RNA, DNA and other biologically important molecules has been done by X-ray crystallography, NMR and electron microscopy techniques. NMR as a powerful biophysical technique provides information about atomic-level structures, dynamic features as well as intermolecular interactions of biological molecules. One definite advantage of NMR is that the molecules can be studied at near-physiological solution conditions. The protein structure database (PDB – Protein Data Bank) has 8715 structures solved by NMR. NMR has come in handy for determining the structure of proteins that resist crystallization. NMR can be used to study membrane proteins that are dissolved in detergent micelles or reconstitued bilayers (Marassi, Opella 1998). The restriction on the

size of the proteins studied by NMR has been partially ameliorated in recent years due to advancement in the methodology and labeling techniques.

The behavior of an atom's nucleus in the presence of an external magnetic field forms the basis of NMR spectroscopy. NMR-active nuclei such as ^{1}H,^{13}C, ^{15}N and ^{31}P possess two distinct energy levels (spin S =1/2). Transitions between these discrete levels can be brought about by excitation with radiofrequency pulses in a magnetic field. These transitions generate magnetization which can be detected as well as transferred to nearby nuclei. The nearby nuclei can be covalently bonded to the excited nuclei or lie in spatial proximity (~ < 5 Å) to it. This ability to probe local conformation of molecules forms the basis of structure determination of proteins by NMR.

Different nuclei precess or spin at a different rate in different magnetic fields. Chemical shift is the measure of resonant frequency or precession rate of a nuclei scaled by the strength of the magnetic field. Each atom in a protein has its unique chemical shift because of the uniqueness of its local environment. Scalar coupling or *J*-coupling is the measure of interaction between nuclei that are separated by up to four covalent bonds (Cavanagh et al, 2007). Dipolar coupling is similar to scalar coupling except that the interaction is through-space instead of through-bond. In isotropic solutions, dipolar couplings, unlike scalar couplings, do not split resonant frequencies since they average out to zero. Nuclear Overhauser Effect (NOE) is the result of a dipolar interaction (**Equation 1**) between spatially close nuclei and the intensity of which is inversely correlated to the distance between them.

$$V_{ij} = \alpha(d_{ij})^{-6} \qquad (1)$$

Where V_{ij} and d_{ij} are the volume of the NOE peak and distance between i and j atoms respectively and α is the scaling factor which depends on the experimental and system conditions.

The measurement of scalar couplings and NOE between various atoms in a protein can lead to the determination of its atomic level structure.

1.8. Isotopic labeling of Proteins

^{1}H and ^{31}P are the only naturally high abundant S=1/2 nuclei found in biological systems. For efficient solution NMR studies, the S=1/2 nuclei, ^{15}N and ^{13}C, have to be enriched in a protein to perform NMR (for structure determination). *E.coli* cells, transformed with the appropriate plasmid encoding the protein to be expressed, are grown in minimal media containing ^{15}N-labeled ammomium chloride and ^{13}C-labeled glucose as the only nitrogen and carbon sources respectively. The protein expressed and purified in this medium contains only ^{15}N- and ^{13}C - labeled amino acids allowing efficient analysis by NMR technqiues. Additionally sometimes cells are grown in minimal media containing D_2O instead of H_2O. This is done to negate the deleterious effect of signal relaxation caused by ^{1}H nuclei leading to significant loss in signal and sensitivity. This is especially useful in larger systems.

1.9. Structure Determination by NMR

The first step in protein structure determination by NMR starts with the identification of the resonant frequencies of each of the atom present in the primary sequence of the protein. This identification is done with the help of several NMR

experiments that correlate different nuclei in the protein. The simplest of these correlation experiments is the ^{15}N-heteronuclear single quantum coherence (^{15}N-HSQC) which correlates the amide protons to the amide nitrogens of the backbone (Bodenhausen, Ruben 1980). A^{15}N-HSQC with well resolved cross peaks is an indication for a well folded protein. Usually, this spectrum serves as a handle to which all backbone and side chain nuclei are correlated in other experiments. ^{15}N-Transverse relaxation spectroscopy (^{15}N-TROSY) is the variation of this experiment wherein the slow relaxing component of the signal is collected. This is extremely useful with high molecular weight deuterated proteins (Pervushin et al, 1997).

Once the chemical shifts of all the atoms are identified, angular restraints can be generated from them. Further, distance restraints from nuclear overhauser effect spectroscopy (NOESY) and orientation restraints from residual dipolar couplings (Prestegard et al, 2004) can be complemented with the angular restraints to calculate atomic level structures. **Figure 9** illustrates the steps involved in structure determination in a schematic manner (Greenbaum, Ghose 2010).

1.10. Experiments for backbone and side chain assignments

Triple resonance experiments like HNCACB/CBCA(CO)NH, HNCA/HN(CO)CA and HNCO/HN(CA)CO correlate the backbone carbon atoms to the backbone amide atoms. One of the experiments in the pair correlate carbon atoms of the previous as well as its own while the other experiment correlates the carbon atom on the previous amino acid in the sequence. For example, by comparing the signals from an HNCA (two peaks representing self and previous C_α for each amide) with HN(CO)CA (a peak representing

previous C_α for each amide), the individual resonances of each carbon atom can be identified through the sequence (Cavanagh et al, 2007).

Side chain experiments like C(CO)NH and H(CCO)NH are extended versions of CBCA(CO)NH wherein the side chain carbon and protons are correlated to the backbone amides (Grzesiek et al, 1993). HCCH-TOCSY (Total Correlation Spectroscopy) (Cavanagh et al, 2007) is another useful experiment that correlates side chain protons to carbons.

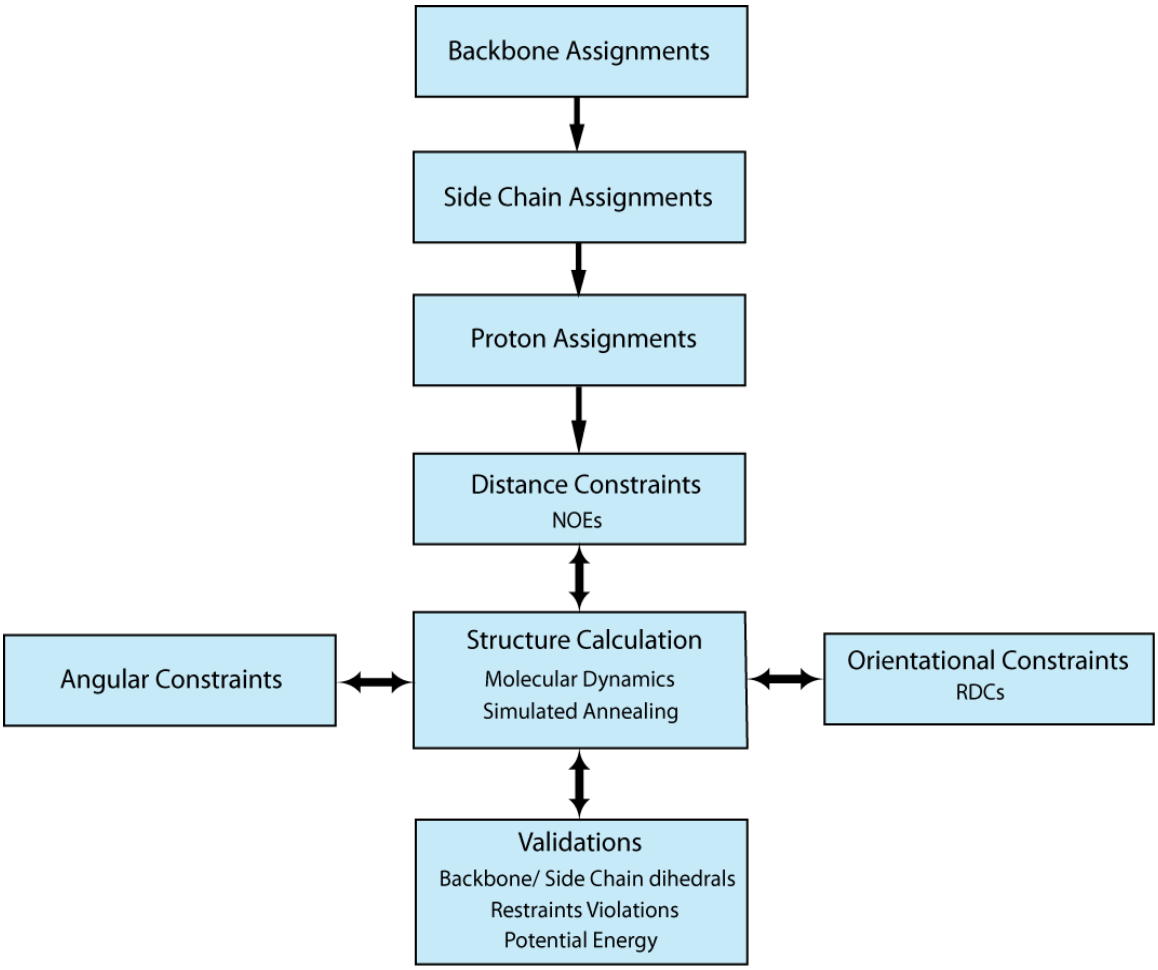

Figure 9: The flow diagram lists the sequential steps involved in obtaining NMR struture from experimental data.

1.11. Restraints to Structure

Distance restraints are obtained from NOESY experiments where the peak intensity is proportional to distance (r^{-6}) between the nuclei. ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC uses protons attached to carbons and backbone amide protons respectively as the handle and correlates protons that are closer in space to it (Sattler et al, 1999). Nuclei that are up to ~5 Å away can be detected by the NOESY experiments. NOE data are collected either in H_2O or in D_2O based on the experimental requirements distinguishing exchangeable and non-exchangeable protons.

C_α and C_β chemical shifts observed can be used to identify the existence of secondary structural elements in the protein. A residue in a β-strand conformation exhibits negative deviation for C_α and positive deviation for C_β from the random coil chemical shifts values for that particular residue type. The converse is true for residues in α-helix (Wishart, Sykes 1994). Further, backbone φ and ψ dihedral angles can be identified from chemical shifts by using databases (Shen et al, 2009).

Orientational restraints are obtained by measuring residual dipolar couplings (RDCs) (**Equation 2**) in partially orienting media like polyacrylamide gels, nonionic liquid crystalline media and viruses (Prestegard et al, 2004). The RDCs appear as an alteration in scalar coupling normally seen in HSQC-type experiments.

$$D_{ij} = - \mu_0\gamma_i\gamma_j h(2\pi r)^{-3}[(3\cos^2\theta - 1)/2] \qquad (2)$$

where D_{ij} is the dipolar coupling between the atoms in Hz, μ_0 is the permittivity of space, γ is the magnetogyric ratio of the atoms, h is the Planck's constant, r is the distance between the nuclei and θ is the angle between the internuclear vector and the magnetic field.

Structure calculations are performed as a molecular dynamics simulation with a suitable force field with the distance, angular and orientational restraints as the input (Greenbaum, Ghose 2010). Initially, unambiguous distance restraints along with dihedral restraints are used to produce initial structures with proper secondary structural elements. Then, a process of iterative refinement is carried out where more NOEs are added and subsequently checked for violations (Greenbaum, Ghose 2010). Programs like ARIA (Habeck et al, 2004) can perform these calculations in a fairly automated way. Programs like PROCHECK (Laskowski et al, 1996) generates the Ramachandran plot (Ramachandran et al, 1963) for the structures calculated which can be used to assess the quality of the strutures. The final ensemble of structures, containing > 10-15 structures, consistent with the all available experimental data with optimal structural statistics is taken to constitute the final structural ensemble.

1.12. Study of interaction by NMR

Since the resonant frequency of an atom is highly influenced by its local environment, NMR is an ideal technique to study protein-protein, protein-ligand and protein-nucleic acid interaction (Zuiderweg 2002). Once the NMR resonances (backbone and/or sidechain) corresponding to the protein have been assigned, ligands, nucleic acids or unlabeled protein partner can be slowly titrated with the protein and chemical shifts can be monitored. The residues that are involved in the interaction experience a change in the local environment, due to direct interaction or a conformational change which results in an alteration in the NMR signal (chemical shift changes or intensity changes). The kinetics of the interaction plays a major role in chemical shift changes seen during a

titration. Fast chemical exchange (suggesting weak binding) between associated-dissociated state results in a single set of peaks corresponding to average chemical shift of both bound and unbound states. Slow exchange between states (indicating strong binding) leads to two sets of peaks corresponding to bound and free states. Disappearance of resonances during titration can be due to intermediate exchange when the kinetics of interaction is poorly defined (Zuiderweg 2002). These changes can be mapped onto the protein structure to identify the binding pocket. Moreover, a good estimation of dissociation constants, stoichiometry and kinetics of binding can be obtained.

2. Materials and Methods

2.1. Expression and purification of *Pseudomonas aeruginosa* PE (187 aa) domain

The plasmid pET16b(+) cloned with PE gene was a gift from Dr. Stewart Shuman at Memorial Sloan Kettering Institute, New York. The construct had a N-terminal His_{10} tag and the protein with the tag had a molecular weight of 23,839.7 Da. The plasmid was transformed into *E. coli* BL21-(DE3) cells and plated in ampicillin containing agar plates that produced good isolated colonies the next day. Cells from individual colonies were inoculated into 20 mL of M9 medium prepared in 100% H_2O. After overnight growth, cells were subsequently transferred step-wise into 20 mL M9 media containing successively increasing proportions of D_2O (25%, 50%, 75% and finally 100%). Cells grown overnight in M9-(100% D_2O) were inoculated into 1 L of M9-(100% D_2O) containing 1 g of ^{15}N-labeled ammonium chloride and 2 g of ^{13}C,^{2}H-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 *Pae*PE 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 (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 14,500 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 (pH 7.5), 0.5 mM NaCl, 500 mM

Imidazole, 10% sucrose. The eluate was 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 (NMR buffer). The peak fraction containing pure *Pae*PE uniformly-labeled with ^{15}N,^{13}C,^{2}H was concentrated to 300 μM protein in 300 μL for use in NMR experiments.

A similar procedure was followed to obtain ^{15}N,^{13}C-labeled *Pae*PE except that the pre-adaptation D_2O steps were eliminated and the M9 containing 2 g of ^{13}C,^{1}H-glucose (as the only carbon source) was prepared in 100% H_2O. Specific ^{15}N glycine or alanine labeled samples was prepared by using 50 mg ^{15}N-labeled glycine or alanine in 500 mL LB media (Englander et al, 2006). The expression and purification is already explained above. 100 μM ^{15}N,^{13}C-labeled protein was resuspended in 5% PEG/n-octanol media (molar ratio, r = 0.87) (Rueckert, Otting 2000) and used to make residual dipolar coupling (RDC) measurements.

2.2. Production and purification of *Pae*PE177 (C-terminal truncation)

A pET28b-based plasmid encoding the N-terminal 177 residues of *P. aeruginosa* LigD PE (molecular weight - 20220.8 Da) fused to an N-terminal His_{10}-SUMO tag (Mossessova, Lima 2000) was transformed into *E. coli* BL21 (DE3)-RIPL cells. The growth and initial TALON purification procedure were identical to the purification of full length *Pae*PE. 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 1L 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 *Pae*PE177 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 NMR buffer. The peak fraction containing pure *Pae*PE177 uniformly-labeled with ^{15}N,^{13}C,^{2}H was concentrated to 300 μM in 300 μL protein for use in NMR experiments.

2.3. Size identification of *Pae*PE and *Pae*PE177 by gel-filtration

Superdex 200 (Column volume – 24 mL) was equilibrated with five column volumes of 50 mM $NaPO_4$ (pH 7.0), 150 mM NaCl. 100 μl of 1 mg/mL Blue Dextran (mol wt. 2,000,000 Da) was passed through the column to identify the void volume. 100 μL of sample containing 3 mg/mL Conalbumin (mol wt. 75,000 Da), 4 mg/mL Ovalbumin (mol wt. 43,000 Da), 3 mg/mL Carbonic Anhydrase (mol wt. 29,000 Da), 3 mg/mL Ribonuclease A (mol wt. 13,700 Da) and 3 mg/mL Aprotinin (mol wt. 6,500 Da) was passed through the column to make a calibration profile. 2 mg/mL of pure *Pae*PE and *Pae*PE177 was passed through the column (equilibrated with NMR buffer) individually and their elution volume was noted. The K_{av} of the eluted proteins is given by,

$$K_{av} = (V_e - V_o)/ (V_c - V_o) \qquad (3)$$

Where V_e is the elution volume, V_c is the column volume and V_o is the void volume. K_{av} is plotted against the log (mol.wt) of the calibration proteins to obtain a straight line. The observed molecular weights of the PE proteins were calculated from the slope and the intercept.

2.4. NMR spectroscopy of *Pae*PE

NMR experiments were performed at 25° C using either Bruker Avance (600 MHz, 700 MHz or 800 MHz) or Varian Inova (600 MHz) spectrometers equipped with cryogenic probes capable of applying pulsed-field gradients along the z-axis. Sequential backbone assignments were obtained from TROSY-based HNCA, HN(CO)CA experiments (600 MHz; 512, 32 and 64 complex points and sweep-widths of 15 ppm, 39.5 ppm and 32 ppm in the ^{1}H, ^{15}N and ^{13}C dimensions respectively), HNCO (800 MHz; 512, 32 and 64 complex points and sweep-widths of 13 ppm, 32 ppm and 12 ppm in the ^{1}H, ^{15}N and ^{13}C dimensions respectively), HN(CA)CO (600 MHz; 512, 32 and 32 complex points and sweep-widths of 15 ppm, 39.5 ppm and 14 ppm in the ^{1}H, ^{15}N and ^{13}C dimensions respectively), HNCACB and HN(CO)CACB (800 MHz; 512, 32 and 40 complex points and sweep-widths of 13 ppm, 32 ppm and 65 ppm in the ^{1}H, ^{15}N and ^{13}C dimensions respectively) (Salzmann et al, 1998) with a relaxation delay of 1.5 seconds on a uniformly ^{13}C,^{15}N,^{2}H-labeled protein. The assignment procedure was facilitated by residue type identification using (H)C(CO)NH-TOCSY (600 MHz; 512, 32 and 64 complex points and sweep-widths of 15 ppm, 36.2 ppm and 80 ppm in the ^{1}H, ^{15}N and ^{13}C dimensions respectively; mixing time = 12 ms) (Cavanagh et al, 2007) on a ^{13}C,^{15}N,^{1}H-labeled protein. Simple ^{15}N-HSQC (600 MHz, 512 and 128 complex points and sweep-widths of 15 and 36.2 ppm in the ^{1}H and ^{15}N dimensions respectively) (Bodenhausen, Ruben 1980) were run for specific amino acid labeled samples. The data were processed with NMRPipe (Delaglio et al, 1995) and analyzed using NMRView (Johnson 2004). ^{15}N-^{1}H steady state NOE experiment (Ferrage et al, 2010) was done on a

$^{13}C,^{15}N,^{2}H$-labeled sample with 6 s ^{1}H saturation (600 MHz; 512 and 128 complex points and sweep-widths of 15 ppm and 37.8 ppm in the ^{1}H and ^{15}N dimensions respectively).

IPAP (Ottiger et al, 1998) experiment was performed for measuring the ^{15}N-^{1}H RDC on an isotropic (700 MHz, 512 and 128 complex points and sweep-widths of 14 ppm and 32 ppm in the ^{1}H and ^{15}N dimensions respectively) and PEG/n-octanol aligned sample (700 MHz, 512 and 100 complex points and sweep-widths of 14 ppm and 32 ppm in the ^{1}H and ^{15}N dimensions respectively).

2.5. NMR spectroscopy of *Pae*PE177

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 ^{1}H, ^{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 ^{1}H, ^{15}N and ^{13}C dimensions respectively) experiments were used for backbone-directed sequential assignments utilizing uniformly $^{13}C,^{15}N,^{2}H$-labeled protein. A pre-scan delay of 1.5 s was used for all experiments. HCCH-TOCSY (800 MHz; 512, 35 and 35 complex points and sweep-widths of 14 ppm, 64 ppm and 64 ppm in the ^{1}H, 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 ^{1}H, ^{15}N and ^{13}C dimensions respectively; mixing time = 12 ms) experiments on uniformly $^{13}C,^{15}N$-labeled *Pae*PE177 were also performed. Main-chain and side-chain ^{1}H assignments were obtained using a H(CCO)NH experiment (600 MHz; 512, 30 and 43 complex points and sweep-widths of 13 ppm, 32 ppm and 7.5 ppm in the direct ^{1}H, ^{15}N and indirect ^{1}H

dimensions respectively; mixing time = 18 ms). Additional ^{1}H 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 ^{1}H, ^{15}N and indirect ^{1}H 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 ^{1}H, ^{13}C and ^{1}H dimensions respectively; mixing time = 150 ms). ^{15}N-^{1}H steady state NOE experiment was done on a ^{13}C,^{15}N-labeled sample with 3 s ^{1}H saturation (600 MHz; 512 and 128 complex points and sweep-widths of 14 ppm and 32 ppm in the ^{1}H and ^{15}N dimensions respectively).

2.6. Distance and Angular Constraints for *Pae*PE177

Backbone $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and ^{15}N chemical shifts were converted into dihedral ϕ and ψ angular restraints using the TALOS+ software suite (Shen et al, 2009). The NOE-derived distance constraints were obtained using a 3D NOESY-TROSY experiment (150 ms mixing time), and 3D ^{13}C-edited NOESY experiment in D_2O (150 ms mixing time, separately optimized for aliphatic and aromatic regions of the spectrum) (Sattler et al, 1999). NOESY crosspeak volumes/intensities were obtained using NMRView and converted into distance restraints using the ADR (ambiguous distance restraints) protocol using the ARIA (Habeck et al, 2004) software suite. Manual NOE assignments were supplemented with those assigned automatically using the ARIA scheme as described previously.

2.7. *Pae*PE177 NMR structure calculation

Manually-assigned NOEs were combined with dihedral and hydrogen-bonding constraints into a standard assignment/structure-calculation protocol using the ARIA 2.3 (Habeck et al, 2004) software suite. Additional assignments that were obtained automatically during a run were inspected for correctness and incorporated into subsequent runs. Structure calculations were performed using a Cartesian dynamics simulated annealing protocol that consisted of: (i) a high temperature dynamics at 2000 K (10,000 steps), (ii) 4000 steps of refinement, (iii) a Cartesian dynamics cooling stage from 2000 to 1000K (6000 steps) and (iv) a second Cartesian dynamics cooling stage from 1000K to 50K (4000 steps). Force constants of 5, 25 and 200 kcal mol^{-1} $Å^{-2}$ for the dihedral constraints and 10, 10 and 50 kcal mol^{-1} $Å^{-2}$ for the distance constraints (ambiguous, unambiguous and hydrogen-bond) were used during the three temperature stages of the Cartesian dynamics protocol. Prochiral valine and leucine methyl groups were treated using the floating chirality method (Habeck et al, 2004). For the final refinement step, the lowest energy structures were solvated in a 7.5 Å box of water and refined with the full Lenard-Jones non-bonded potential and electrostatic interactions from the OPLS force-field using a protocol described in detail previously (Linge et al, 2003). The resulting NMR structures were evaluated with PROCHECK (Laskowski et al, 1996). In the final run, 1500 structures were calculated and the 100 lowest energy structures were used in the water-refinement protocol. 20 of the lowest energy water-refined structures with no violations of experimental constraints (no violations > 0.5 Å for distance restraints and > 5.0 ° for dihedral restraints) were chosen to represent the final structural ensemble of *Pae*PE177.

2.8. Preparation of DNA Oligos

HPLC purified DNA oligo D18 (5'- TTT TTT GGG ACC GGT CCC –3'), D18p ('5'- TTT TTT GGG ACC GGT CCC –3'PO_4) and D17R1 (5'- TTT TTT GGG ACC GGT CC_fC_r –3' – the penultimate base is a 2-fluoro cytidine and the last base has a ribose moiety) were purchased from IDT DNA. DNA stock solutions containing 1158.2 nmol of D18 in 100 μL of water, 300.1 nmol of D18p in 50 μL of water and 924.9 nmol of D17R1 in 100 μL of water were annealed in an Eppendorf thermocycler at 95° C for 3 min, at 60° C for 2 min, at 40° C for 10 min and finally at 25° C for 10 min. The concentrations of the annealed DNA oligos were estimated from absorption measurements at wavelengths of 260 and 280 nm.

2.9. *Pae*PE177 interaction with DNA oligo by Gel-Filtration

Superdex 75 (GE Life Sciences, Column Vol - 24 mL) was equilibrated with four column volumes of NMR buffer. 100 μM sample of *Pae*PE177 and 50 μM sample of DNA oligo D18 were run individually through the column as a control. Then, *Pae*PE177: D18 samples of the following concentrations 50:50, 50:25 and 50:200 μM were run through the column. The *Pae*PE177: D18 mixtures eluted as a single peak indicating a complex formation whose elution volumes were noted.

2.10. *Pae*PE177-DNA interactions by fluorescence measurements

1 μM sample of *Pae*PE177 in a buffer containing 50 mM Bis-Tris, 200 mM NaCl, 5 mM DTT, pH 6.5 was used for the fluorescence measurements using a 3 mL cuvette

that allowed continuous mixing. All measurements were performed on a Spex Fluorolog-3 (Horiba Jobin Yvon) at 25° C. Tryptophan emission was monitored following excitation at 280 nm. Spectra were recorded from 300 to 450 nm in 1 nm increments using a 0.5 s integration time and averaged over four scans. For the annealed D17R1 oligo, the following *Pae*PE177:DNA ratios 1:0, 1:0.05, 1:0.1, 1:0.2, 1:0.3, 1:0.5, 1:0.75, 1:1, 1:1.5, 1:2, 1:3, 1:5, 1:8, 1:12, 1:18, 1:25 and 1:40 were used (ratios refer to dsDNA) in the absense of any metal ions, in the presence of 100 μM Mg^{2+} and in the presence of 100 μM Mn^{2+} ions. K_d values were obtained by fitting the change in fluorescence intensity at the emission maximum in the absence of DNA (331 nm) with increasing DNA concentration to **Equation 4** using in-house software that utilized the ODRPACK libraries (Boggs et al, 1989).

$$I_0 - I = \frac{K_d + [D] + [P]_0 - \sqrt{\left(K_d + [D] + [P]_0\right)^2 - 4[D][P]_0}}{2[P]_0} \qquad (4)$$

Where I is the intensity at 331 nm at a dsDNA concentration of [D], I_0 is the corresponding intensity in the absence of DNA and $[P]_0$ is the protein concentration (1 μM).

$$\Theta = [D]^n / (K_d + [D]^n) \qquad (5)$$

is the Hill equation used to identify co-operativity where θ is the fraction of occupied sites and n is the Hill coefficient.

Additionally, fluorescence titrations of 1 μM *Pae*PE177 with metal ions (Mg^{2+} and Mn^{2+}) in the absence of dsDNA were done. The following *Pae*PE177:metal ions ratios 1:0, 1:0.05, 1:0.1, 1:0.2, 1;0.5, 1:1, 1:2, 1:5, 1:10, 1:20, 1:40 and 1:80 were used.

2.11. *Pae*PE177-DNA interactions by NMR titration analysis

100 μM samples of uniformly-^{2}H,^{13}C,^{15}N-labeled *Pae*PE177 in a buffer containing 50 mM Bis-Tris, 200 mM NaCl, 5 mM DTT, pH 6.5 was used for NMR titrations with both D18 and D18p. ^{15}N,^{1}H TROSY spectra were acquired in the presence of increasing amounts of DNA to a final concentration of 1, 3, 5, 8, 12, 20, 30, 50, 100, 200, 300 and 470 μM for D18; 5, 10, 20, 30, 50, 100, 200 and 300 μM for D18p. All spectra were acquired at 600 MHz and 25 °C using sweep-widths of 15 ppm (512 complex points) and 36.2 ppm (128 complex points) in the direct and indirect dimensions respectively using a recycle delay of 1.5 s and 48 scans per t_1 point.

Similarly, 100 μM sample of uniformly-^{2}H,^{15}N-labeled *Pae*PE177 in the above mentioned buffer was titrated with D17R1 with the following concentrations: 5, 10, 20, 30, 50, 100, 200 and 300 μM. ^{15}N,^{1}H TROSY spectra were acquired at 800 MHz spectrometer equipped with a microprobe at 25 °C using sweep-widths of 14 ppm (512 complex points) and 32 ppm (128 complex points) in the direct and indirect dimensions respectively using a recycle delay of 1.5 s and 64 scans per t_1 point.

2.12. ^{15}N-edited NOESY-TROSY experiments on the *Pae*PE177-DNA Complex

180:180 μM ^{2}H,^{15}N *Pae*PE177: D18 (93 nmol in 10 μl water stock was made and annealed by previously explained procedure) was used to perform a ^{15}N-edited NOESY-HSQC experiment at 25 °C on a Bruker Avance spectrometer operating at a ^{1}H frequency of 800 MHz (512, 40 and 128 complex points and sweep-widths of 14 ppm, 32 ppm and 12 ppm in the direct ^{1}H, ^{15}N and indirect ^{1}H dimensions respectively; mixing time = 150 ms).

2.13. Mn^{2+} titration with *Pae*PE177:D17R1 complex by NMR

To a 100:50 μM $^{2}H,^{15}N$-*Pae*PE177:D17R1 sample with 2 mM phosphate, the following concentrations of Mn^{2+} from $MnCl_2$ stock solution were added: 20 and 50 μM. $^{15}N,^{1}H$ TROSY spectra were acquired at 600 MHz spectrometer equipped with a microprobe at 25 °C using sweep-widths of 14 ppm (512 complex points) and 36.2 ppm (128 complex points) in the direct and indirect dimensions respectively using a recycle delay of 1.5 s and 64 scans per t_1 point.

3. Results and Discussion

3.1. Gel-Filtration of *Pae*PE and *Pae*PE177

Both *Pae*PE and *Pae*PE177 eluted as a single peak from the Superdex 200 size exclusion column. The calculated size from the K_{av} value for *Pae*PE and *Pae*PE177 was 32,344.4 Da (Actual weight 23,839.7 Da) and 27,353.3 Da (Actual weight 20,220.8 Da) respectively. **Figure 10c** is the overlay of the chromatogram from size exclusion column for *Pae*FL and the calibration standards.

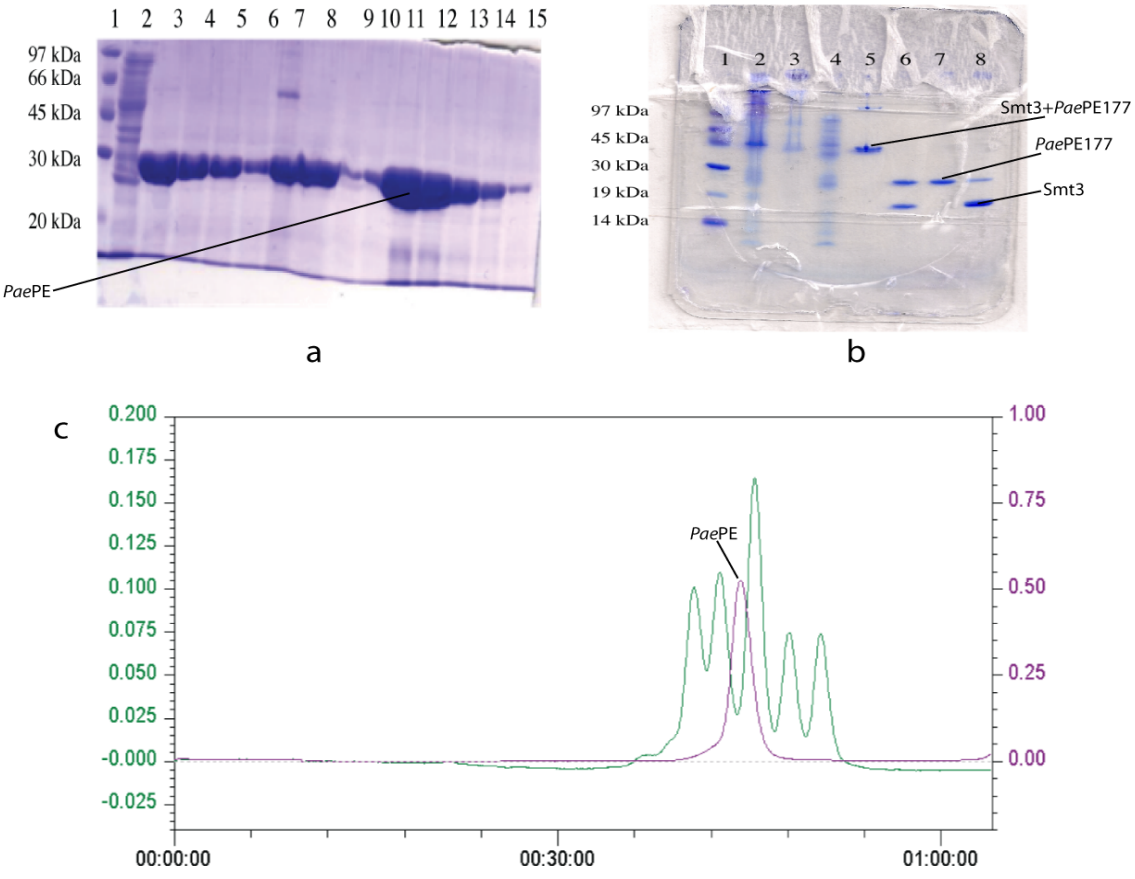

Figure 10: a) SDS-PAGE gel for *Pae*FL purification. Lane 1 – Marker, 2- TALON flow through, 3-6 - TALON elution, 7 – TALON beads, 8 – GF sample, 9-15 - GF elutions. b) SDS-PAGE for gel *Pae*PE177 purification. Lane 1 – Marker, 2 – cell lysate, 3 – cell debris, 4 – TALON flow through, 5 – TALON elution, 6 – After Ulp1 digestion, 7 – TALON flowthrough after Ulp1 (pure *Pae*PE177 seen), 8 – TALON elution after Ulp1. c) Gel-filtration overlay of *Pae*PE (magenta) and standards (green).

3.2. *Pae*PE backbone assignments and analysis

The backbone assignments of *Pae*PE were obtained by performing previously described triple resonance experiments and using the runabout automatic module in NMRViewJ analysis software. 136 out of 174 (78.2 %) expected backbone N, H_N resonances (excluding the N-terminal 21 residue tag and 13 prolines) were assigned. Further, 153 of 187 (81.8 %) C_α resonances, 121 of 172 (70.3 %) C_β resonances and 156 of 187 (83.4 %) C' resonances of the backbone were assigned.

There were three stretches of amino acid sequences 42 to 49, 105 to 109 and 167 to 177 that were not assigned. Residues His42, His48 and Tyr88 which are critical for both diesterase and monoesterase activites (Zhu et al, 2005) were not assigned. Incomplete back-exchange of amide protons was unlikely to be the cause of the missing assignments, because identical ^{15}N-^{1}H 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 **Figure 11**) and could not be sufficiently resolved to allow unambiguous assignments. Even HNCO peaks were not well resolved in that region. Moreover, the presence of the N-terminal 21 residue tag, which likely is completely disordered, could account for some of the peaks in the crowded region. Only six of the 21 residues of the tag were unambiguously assigned. The plasmid pET-16b(+) possess a Factor X_a cleavage site after the His_{10} tag. Efforts were undertaken to remove the tag after purification by using the enzyme. The results were unfruitful as non-specifically cleaved products were seen.

Moreover, the H(CCO)NH experiments performed to obtain the side chain ^{1}H assignments failed due to insufficient transfer from far off protons to the backbone

amides. These assignments are much needed for any protein structure calculation. I was not able to proceed any further with this construct.

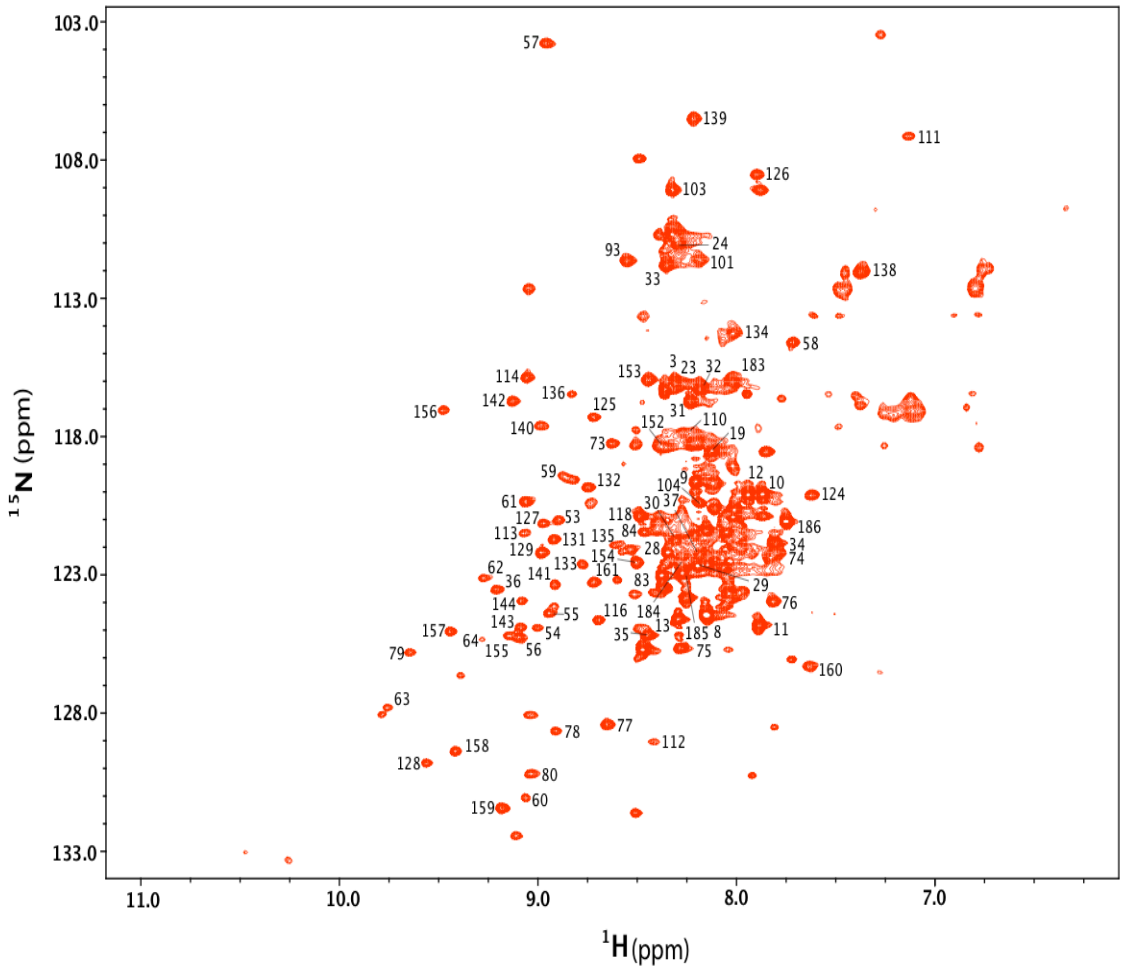

Figure 11: ^{15}N,^{1}H-TROSY spectrum of *Pae*PE. Data collected in Bruker spectrometer with ^{1}H frequency of 800 MHz equipped with cryoprobe with sweep-width - 14 ppm (512 complex points) for direct dimension and sweep-width - 32 ppm (128 complex points) for indirect dimension with recycle delay of 1.5 s. Sample concentration – 300 µM. NMR buffer – 50 mM Bis-Tris, 200 mM NaCl, 5 mM DTT, pH 6.5

Just by looking at the ^{15}N-^{1}H TROSY spectrum of *Pae*PE, the presence of both ordered (well resolved peaks) and disordered regions (crowded intense peaks in the middle of the spectra) were seen. I was curious to find the regions of the protein that were

ordered and disordered. The steady state ^{15}N-^{1}H NOE and residual dipolar coupling measurements gave a pretty good idea about the regions of the protein that were ordered.

Figure 12 shows the plot between NOE and RDCs versus residue number. Two regions were identified that were clearly disordered. One was the N-terminus of the protein from 1- 30 and the other residues belonging to the C-terminus starting from 160, even though the second region had many missing assignments.

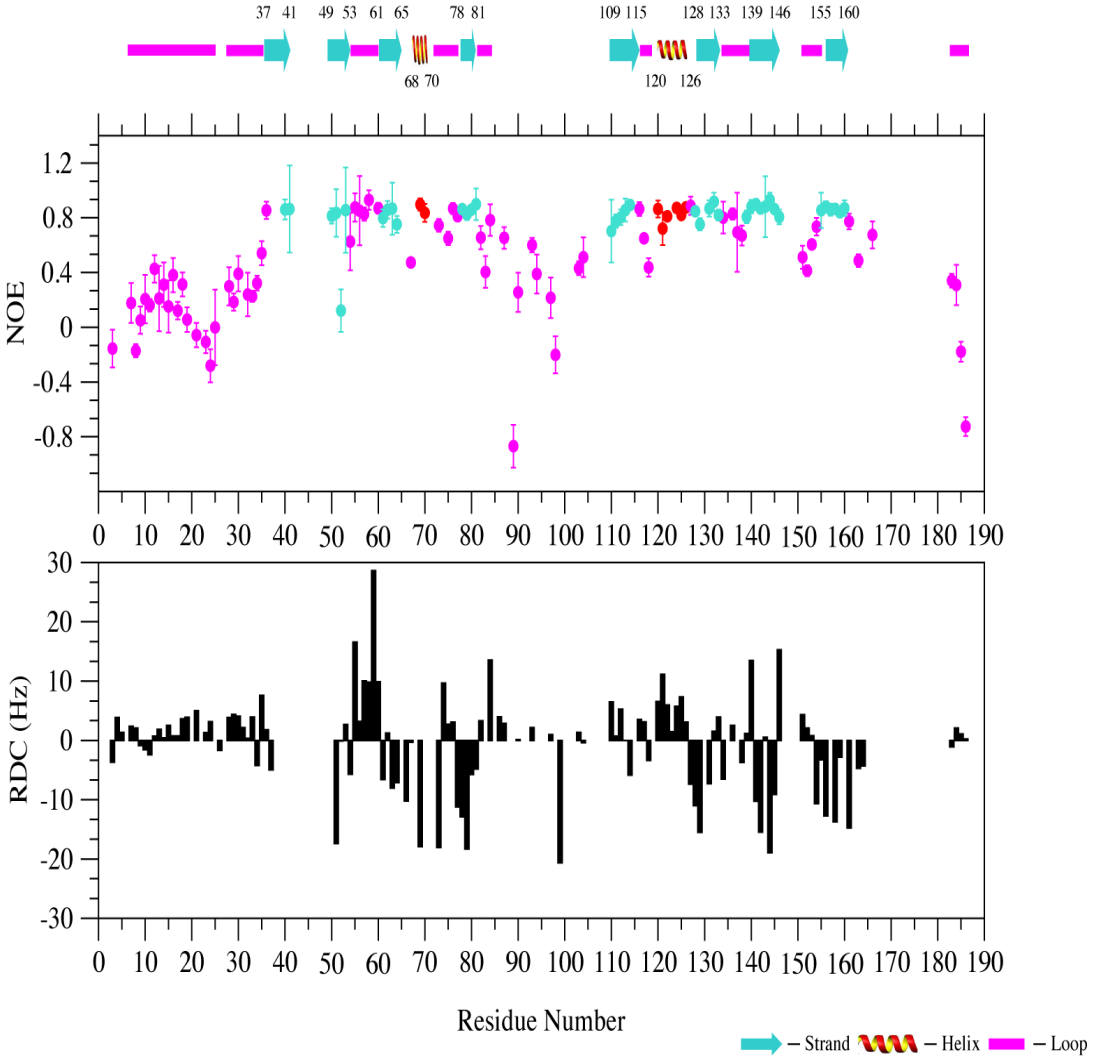

Figure 12: Illustrates NOEs and RDCs versus residue number. Secondary structural elements with residue numbers are indicated on top identified from chemical shift index. Steady state NOEs were measured at 600 Mhz Varian spectrometer equipped with cryoprobe. Sample in 50 mM Bis-Tris, 200 mM NaCl, 5 mM DTT, pH 6.5 at 25°C, Conc. = 300 µM. RDCs were measured with sample in PEG/Octanol media at 700 MHz Bruker spectrometer equipped with cryoprobe, Conc. = 100 µM.

3.3. *Pae*PE truncation analysis

Due to the disordered nature of the C-terminus, we decided to investigate the role of the C-terminal tail in the activity of *Pae*PE. Our collaborators at MSKCC performed a series of truncations on *Pae*PE and measured the activity of the protein. The assay studies the ability of the protein to resect a single ribonucleotide in a labeled primer-template duplex. The end products of both phosphodiesterase and phosphomonoesterase activities were identified by electrophoresis as resected products run at different positions in the gel (**Figure 13**). *Pae*PE, *Pae*PE 1-180, *Pae*PE 1-177, *Pae*PE 1-172, *Pae*PE 1-167 and *Pae*PE 1-162 all showed equivalent activities. Out of these, *Pae*PE 1-180 and *Pae*PE 1-177 expresses well, were soluble and could be purified in high concentrations for NMR processes. I decided to pursue further studies with *Pae*PE 1-177 (hereby referred as *Pae*PE177) construct since it was expected that the spectral quality corresponding to this construct would be significantly better.

I did not tamper with N-terminus of the protein since there was biochemical and mutagenesis data that suggested the need for several N-terminal residues for 3'-phosphomonoesterase activity (Zhu et al, 2005).

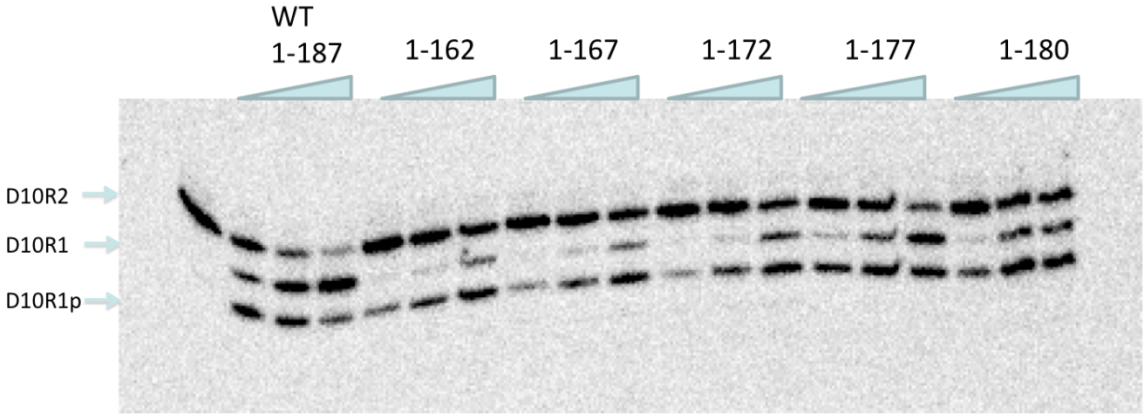

Figure 13: Truncation assays were done in 50 mM Tris (pH 7.5), 5 mM DTT, 0.5 mM $MnCl_2$, 0.5 pmol 5'-^{32}P-D10R2 primer-template duplex with incubation at 37 ˚C for 30 min and quenched with EDTA/formamide.

3.4. *Pae*PE177 assignments and analysis

146 of 164 (excluding the Met residue at the extreme N-terminus) main-chain amide ^{15}N, $^{1}H_N$ (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 (see **Figure 14**).

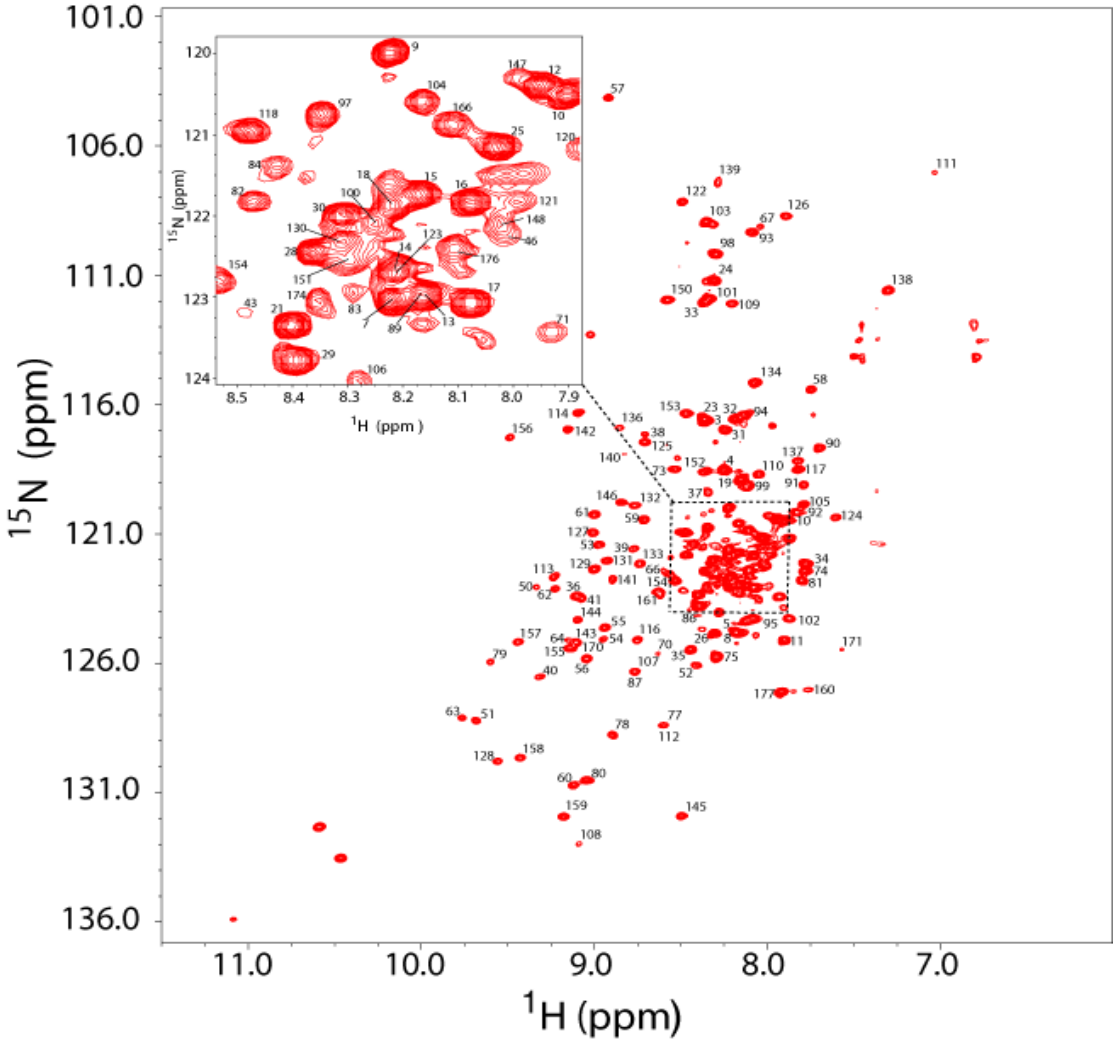

Figure 14: Assigned ^{15}N, ^{1}H TROSY spectrum of uniformly ^{13}C,^{15}N,^{2}H-labeled *Pae*PE177 in 50 mM Bis-Tris propane (pH 6.5), 200 mM NaCl, 5 mM DTT. Data was acquired at a ^{1}H frequency of 600 MHz at 25° C using sweep-widths of 15 ppm (512 complex points) in ^{1}H dimension and 36.2 ppm (128 complex points) in ^{15}N dimension with a recycle delay of 1.5 s. The central crowded region is shown in the inset.

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 side-chains). 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.

Of the 17 residues for which ^{15}N, ^1H_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. Of the residues important for catalytic activity (Zhu et al, 2005), assignments for main-chain ^{15}N, ^1H_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, 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, 2007) software utilizing $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$ and ^{15}N chemical shifts after correction of $^{13}C_\alpha$ and $^{13}C_\beta$ 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).

We noted another interesting feature upon careful analysis of the ^{15}N, ^1H_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 (**Figure 15**). Inspection of **Figure 15** 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 (**Figure 15**). This indicates conformational heterogeneity around the edges of one face (β_5-β_1-β_2-β_3) of the β-barrel.

The N-terminal 32-aa peptide of *Pae*PE, which is important for the phosphomonoesterase activity, was disordered on the ps-ns timescale with average ^{15}N-^{1}H NOE values of 0.03 ± 0.28, determined using an improved pulse sequence (Ferrage et al, 2009), 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 (0.45 ± 0.02) for this region predicted from the Random Coil Index (RCI) (Berjanskii, Wishart 2005) (**Figure 15**) 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}H NOE values for this segment were also marginally higher (0.16 ± 0.04) than the rest of the disordered N-terminus.

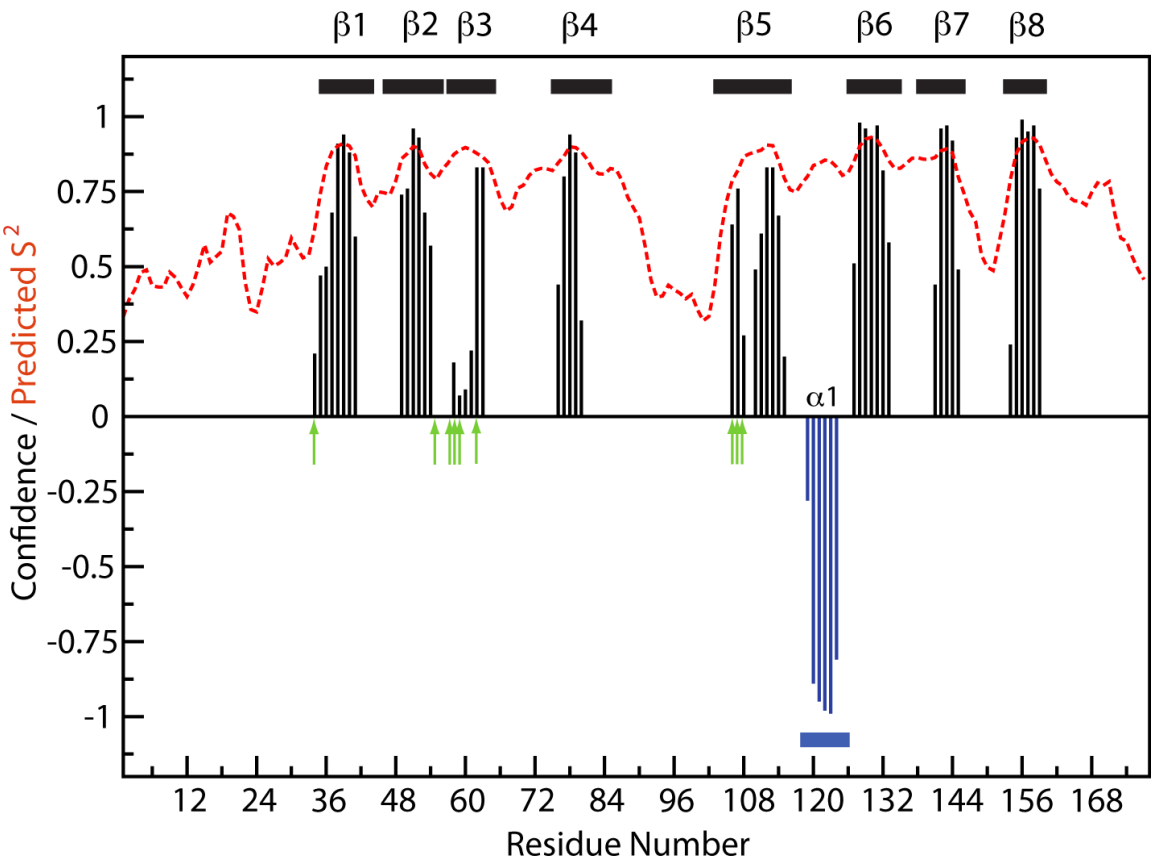

Figure 15: Secondary structure prediction using chemical shifts using TALOS+ (Shen et al, 2009). The confidence values for regions predicted to be β-strand (black) or α-helix (blue) are plotted against residue numbers. Also shown are the predicted order parameters (S^2) based on chemical shifts. Residues that show peak doubling are indicated by green arrows. Secondary structure elements identified by chemcal shift indexing are indicated with β-strands as black rectangles and α-helices as blue rectangles.

3.5. Comparison of *Pae*PE177 and *Pae*PE spectra

^{15}N-^{1}H TROSY spectra of *Pae*PE and *Pae*PE177 had much dissimilarity. The effect of the 21-residue N-terminus tag was discounted for by using a *Pae*PE187 construct that did not possess the N-terminus tag. The purification of *Pae*PE187 was exactly similar to *Pae*PE177. The ^{15}N-^{1}H HSQC of *Pae*PE and *Pae*PE187 was exactly identical except for a few extra peaks pertaining to the tag in the crowded region in the spectrum of *Pae*PE. The scaled chemical shift differences were particularly high (δppm > Std.deviation) for residues (His84, Asp90, Gln97, His99, Gly103, Arg110 and Trp113) in

the loop between strands β_4-β_5 and β_5. This indicates that the tail residues are involved with some interactions with the core. Some of these residues showed marked differences when steady state NOE of *Pae*PE and *Pae*PE177 were compared (**Figure 16**).

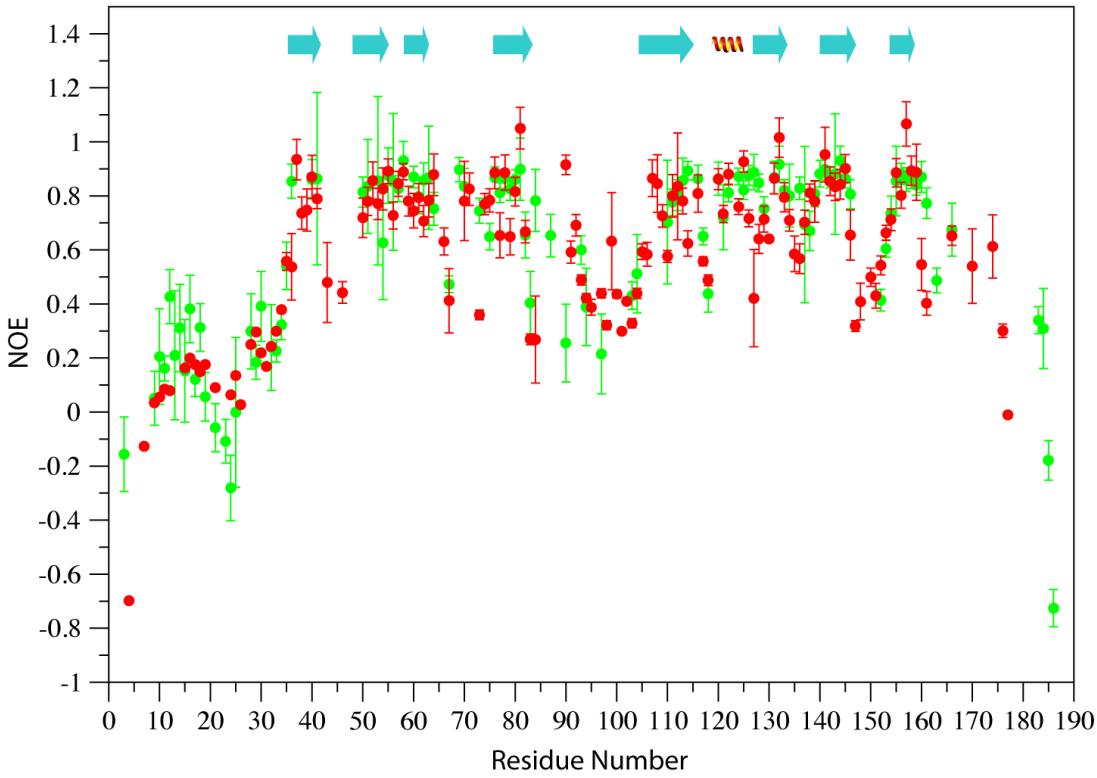

Figure 16: Comparison between *Pae*PE (green) and *Pae*PE177 (red) ^{15}N-^{1}H NOEs. Residue numbers in x-axis and NOE values in y-axis with positions of β-strands in cyan and position of α-helix in red (predicted by chemical shift indexing).

3.6. *Pae*PE177 solution structure

The structure of *Pae*PE177 is the first instance in which a NMR structure has been solved in the class of bacterial NHEJ phosphoesterase domains. The core structure of *Pae*PE177 is a β-barrel comprising eight antiparallel β-strands surrounded by a α-helix and a 3_{10} helix. The residues forming the eight β-strands are 36-42 (β_1), 48-55 (β_2), 58-63(β_3), 76-84 (β_4), 105-115 (β_5), 127-134 (β_6), 141-147 (β_7) and 154-159 (β_8). A short α-

helix spanning residues 120-125 and a 3_{10} helix spanning 96-99 are also present. An analysis with PROMOTIF (Hutchinson, Thornton 1996) indicated the presence of 14 β-turns and 3 γ-turns. About 45% of the residues were in loop conformation suggesting a highly dynamical protein (**Table 1**).

Table 1: Promotif Documentation

Secondary structure summary

Strand	α -helix	3_{10} - helix	Other	Total Residues
62 (47.3%)	6 (4.6%)	4 (3.1%)	59(45%)	131

1 β-sheet

Sheet	No. strands	Type	Barrel	Topology
A	8	Antiparallel	Y	-1-1-1-4X 1 1 1

6 β-hairpins

Strand 1			Strand 2			Hairpin
Start	End	Length	Start	End	Length	Class
Arg36	His42	7	His48	Leu55	8	5:5
His48	Leu55	8	Thr58	Ala63	6	2:2 IP
Thr58	Ala63	6	Arg76	His84	9	15:15
Val105	Pro115	11	His127	Gly134	8	12:14
His127	Gly134	8	Trp141	Asn147	7	10:10
Trp141	Asn147	7	Gln154	Lys159	6	6:6

3 β-bulges

Bulge type	ResX	Res 1	Res 2	Res 3	Res 4
Antiparallel classic	Val39A	Trp108A	Asp109A		
Antiparallel classic	Lys60A	Glu82A	Asp83A		
Antiparallel classic	Gln154A	Thr146A	Asn147A		

14 β-turns

Turn	Sequence	Turn type	H-bond
His42-Ser45	HDAS	IV	
Leu55-Thr58	LDGT	I’	Yes

Ala63-Lys66	AVPK	IV	
Leu70-Ala73	LDPA	VIb	
Pro85-Tyr88	PLDY	IV	Yes
Tyr88-Phe91	YADF	I	
Ala89-Glu92	ADFE	IV	
Phe91-Ser94	FEGS	IV	
Gly103-Ile106	GDVI	IV	
Pro115-Asp118	PLDD	IV	
Gly134-Leu137	GEKL	IV	
Glu135-Ser138	EKLS	VIII	
Thr146-Arg149	TNLR	IV	
Lys151-Gln154	KQSQ	IV	

3 γ-turns

Start	End	Sequence	Turn type
Pro72	Val74	PAV	INVERSE
Gly93	Ile95	GSI	INVERSE
Lys159	Lys161	KAK	INVERSE

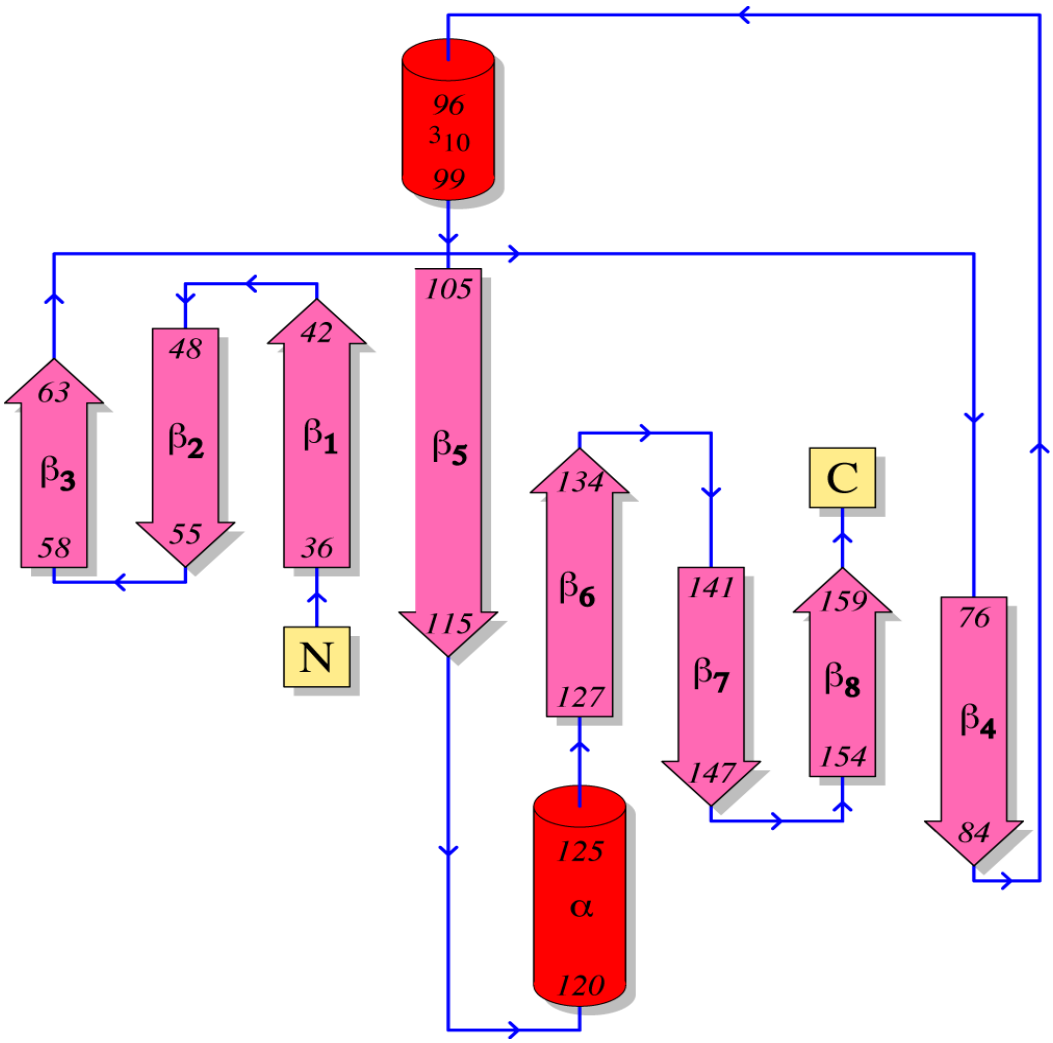

Figure 17: Topology of *Pae*PE177 generated by PROMOTIF with β-strands in magenta and helices in red. The naming pattern is indicated in the middle of the secondary structural element.

Figure 17 illustrates the topology of the secondary structural elements and naming pattern in *Pae*PE177. A search on the DALI server (Holm, Rosenström 2010) for structural homology indicates structure of *Pae*PE177 as a novel fold. The best hit was human Spindlin1 protein with a Z-score of 3.0. Spindlin1 consists of three repeats of four or five antiparallel β-strands. Functionally, it is unrelated to phosphoesterase as it possesses no nuclease or phosphate activity.

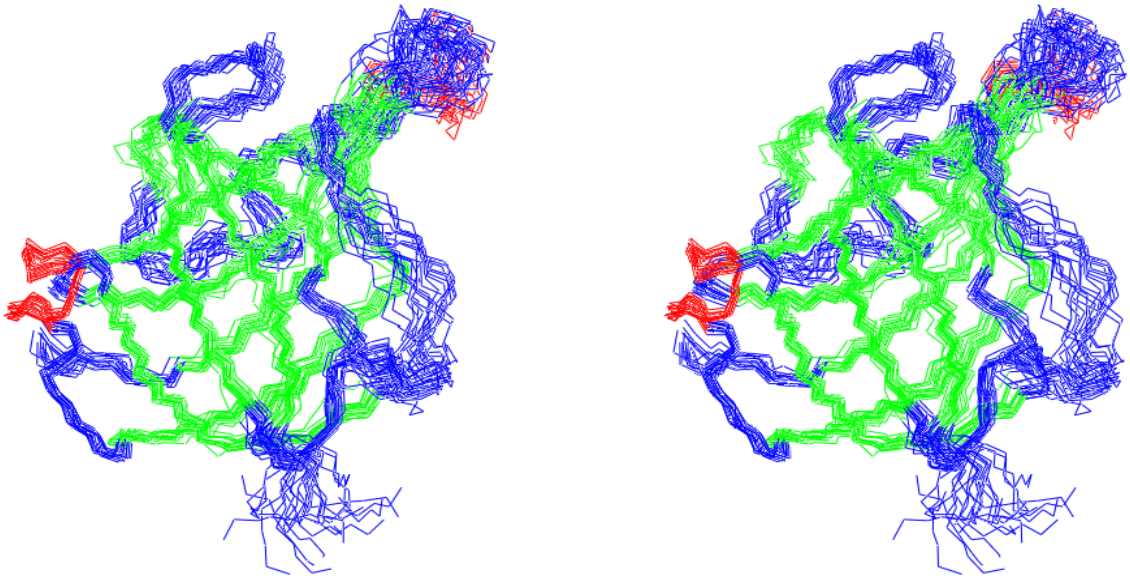

Figure 18: 20 best solution structure generated by ARIA after round 21. Just the backbone of the 20 structures lacking 32 residues from the N-terminus has been shown for clarity. β-strands are in green and helices in red.

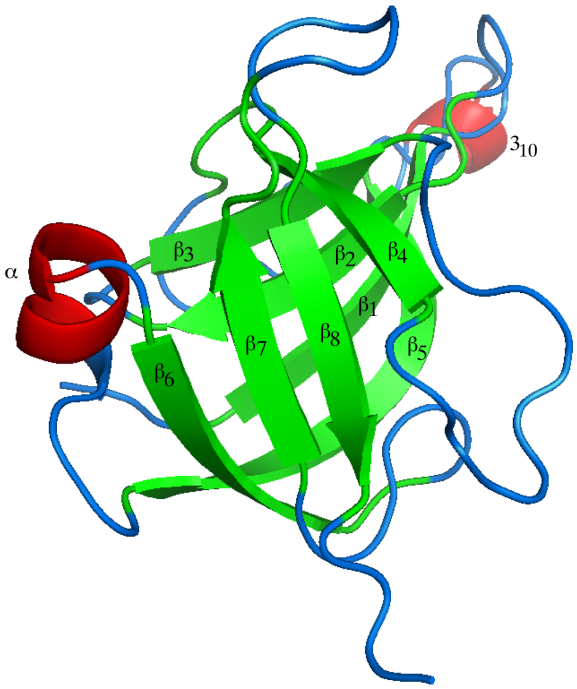

Figure 19: Cartoon representation of *Pae*PE177. β-strands are in green and helices in red.

For clarity purposes, the N-terminus of the protein spanning residues 1-32 has not been included in the structure calculation at this stage because of the lack of significant

number of NOEs to the protein core. It will be added at the end of the calculation to the final structure. **Figure 18** represents the 20-best structures and **Figure 19** is the cartoon represention of *Pae*PE177 generated by ARIA at this point in the calculation. **Table 2** lists the different type of restraints used, RMSD statistics and Ramachandran numbers for the structures.

Table 2. NMR restraints and structural statistics for the best 20 structures.

Restraints and statistics	
Total number of restraints	2189
NOE restraints	1881
Unambiguous	*1434*
Intraresidue	*461*
Sequential	*380*
Short-range	*115*
Medium-range	*46*
Long-range	*432*
Ambiguous	*447*
Dihedral angle restraints	224
Hydrogen bond restraints[a]	84
Structure Statistics[b]	
NOE violations > 0.5 Å	<1%
Dihedral violations > 5°	<0.1%
RMSD from average structure[c, d]	
Backbone (N, Cα, C) (Å)	0.53 ± 0.11 (0.91 ± 0.10)
Heavy atoms (Å)	1.33 ± 0.16 (1.59 ± 0.12)
Ramachandran Statistics[e]	
Most favored region (%)	79.2(62.1)
Additionally allowed (%)	16.8(28.5)
Generously allowed (%)	3.6(6.3)
Disallowed (%)	0.3(3.1)

[a] Hydrogen bond restraints were H^N-O distance of 1.8-2.3 Å and an N-O distance of 2.8-3.3 Å.
[b] Structural characteristics for the final ensemble of 20 water-refined structures.
[c] RMSD of the mean structure from individual structures in the ensemble.
[d] RMSD for residues is structured region 34-42, 47-55, 58-63, 79-84, 111-134, 138-144, 156-159 shown. The numbers in the parenthesis indicate the RMSD for residues 33-163.
[e] Ramachandran plot data shown for residues 34-42, 47-55, 58-63, 79-84, 111-134, 138-144, 156-159. The numbers in the parenthesis indicates the statistics for residues 33-163.

Structurally, the core of the β-barrel is composed of several conserved hydrophobic residues (see **Figure 8**- Introduction). Further, the β-barrel is stabilized by extensive hydrogen bonding between various strands. Hydrogen bonding is seen between atoms

present in the backbone (e.g., Trp108 N-Val39 O-Trp108 H_N), backbone-side chain (e.g., Trp113 $N_{\varepsilon 1}$-Glu54 O-Trp113 $H_{\varepsilon 1}$) and side chain (e.g., Trp155 $N_{\varepsilon 1}$-Gln80 O $_{\varepsilon 1}$-Trp155 $H_{\varepsilon 1}$).

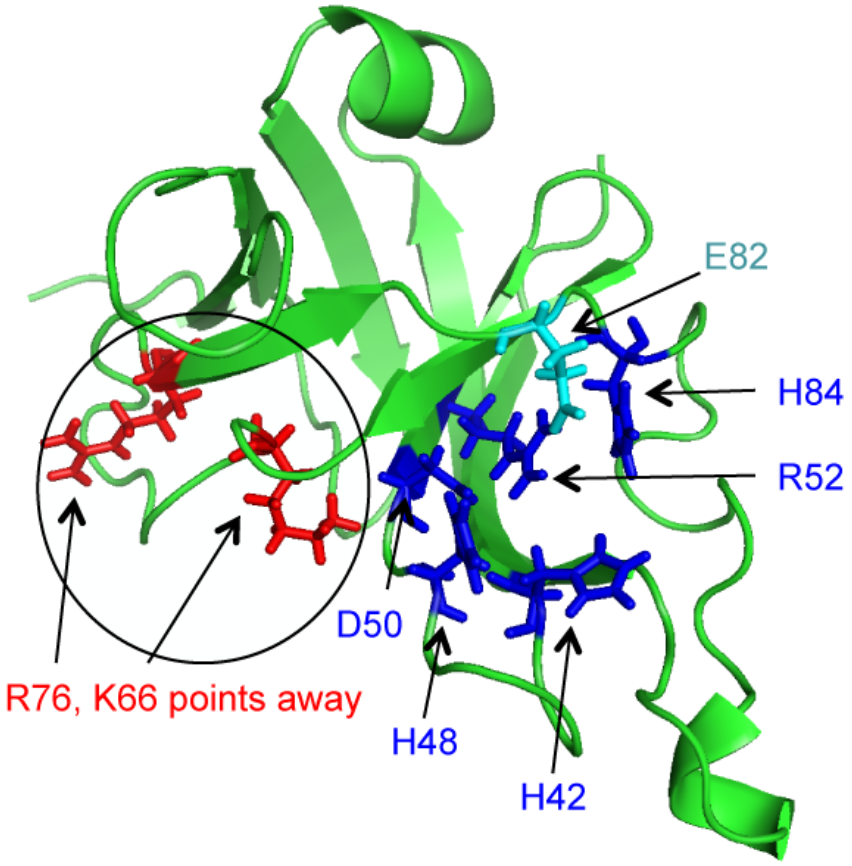

Figure 20: *Pae*PE177 shown in green cartoon with the side chain residues of active site (His42, His48, Asp50, Arg52 and His84) shown in blue, Glu82 (cyan) and Lys66, Arg76 (in red).

The active site involving group 1, group 2 and group 3 residues (see Introduction) is present on the outer surface of the β-barrel rather than the core. The side chains of group 1 residues His42 (β_1), His48, Asp50, Arg52 (all 3 in β_2), His84 (β_4), Tyr88 and group 2 residue Glu82 (β_4) are in close proximity to each other in the active site. Interestingly, the side chain of Lys66 and Arg76 (Group 3 residues) are farther away from the active site. This suggests that there might be some conformation change

involving these residues in the presence of the DNA substrate and Mn^{2+} ion. **Figure 20** shows the protein with the active site side chains highlighted.

3.7. Comparison of solution structure of *Pae*PE177 and crystal structure of *Pae*PE

When the NMR structure was being solved, our collaborators belonging to Prof. Stewart Shuman's group at Sloan Kettering institute solved the crystal structure of *Pae*PE spanning residues 17-187 (Nair et al, 2010). This construct did not possess the phosphomonoesterase activity because residues Arg14 and Asp15 were missing. Overall, the fold between the two structures was similar with the secondary structures at the same places except for a few key differences. The RMSD of the solution and the crystal structure for the structured regions was 1.1 Å. **Figure 21** shows the overlay of the two structures with the secondary structural elements.

One of the key contrasting features between 2 structures was the difference in the proline backbone conformation. Val64-Pro65 and Asp71-Pro72 (**Figure 22a**) are fixed to be in *cis* conformation in the NMR structure whereas they are in the *trans* conformation in the crystal structure. This fixing was done after detailed ^{13}C chemical shift analysis was performed. ^{13}C chemical shifts are key indicators of the conformation of the peptide backbone about an Xaa-Pro bond. An exhaustive statistical analysis of the differences between $^{13}C_\beta$ and $^{13}C_\gamma$ chemical shifts ($\Delta_{\beta\gamma}$) 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 $\Delta_{\beta\gamma}$ values were both 8.4 ppm after correcting for ^{2}H isotope shifts (Venters et al, 1996). Using the program Promega (Shen, Bax 2009) the absolute probabilities of the

Val64-Pro65 and Asp71-Pro72 bonds being in the *cis* conformation were found to be 91% and 86%, respectively.

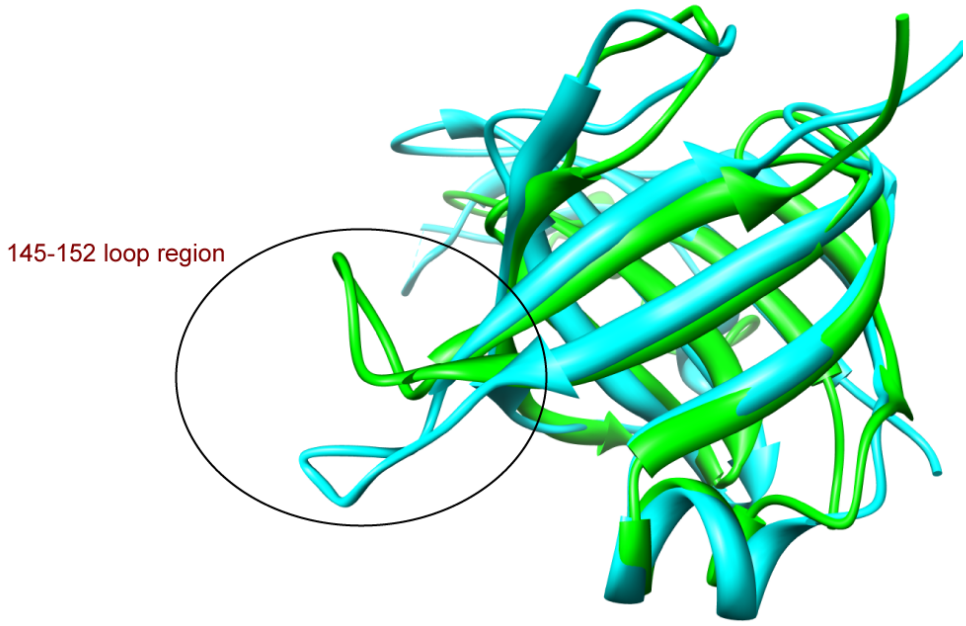

Figure 21: Overlay of NMR structure (green) and crystal struture (cyan) showing the secondary structural elements in cartoon. The loop 145-152 is indicated.

The loop region from 145-152 seems to have a different orientation in the solution structure in comparison with the crystal structure. In the crystal structure, the loop seems to be further extended away from the core. But in the solution structures the loop seems to fold back slightly towards the core. Some of the β-strands (β_2, β_5 and β_7) (**Figure 22b**) are kinked when compared to its crystal counterparts.

Another important feature to be noted in the crystal structure is the presence of four yttrium ions bridging a *Pae*PE protomer with two other *Pae*PE protomers (Nair et al, 2010). Asp104, Asp117, Glu124, Glu135, Asp179 and Glu184 are chelated by yttrium ions that aided in crystal stabilization. Yttrium-chelated Asp104 which is spacially close

to β_5 could be the cause of the difference in this β-strand between NMR and crystal structures.

The most important aspect of the crystal structure is the presence of Mn^{2+} and sulfate (considered as a mimic for phosphate) ions in the active site. Mn^{2+} is seen coordinating $His42N_{\delta}$, $His48N_{\varepsilon}$, $Asp50O_{\varepsilon}$, sulfate oxygen and two water molecules. This explains the requirement of Mn^{2+} over Mg^{2+}, which prefers coordination with side chain nitrogens instead of oxygens (Nair et al, 2010). The side chains of Arg52, His84 and Tyr88 hydrogen bonds with the sulfate ion either directly or through water.

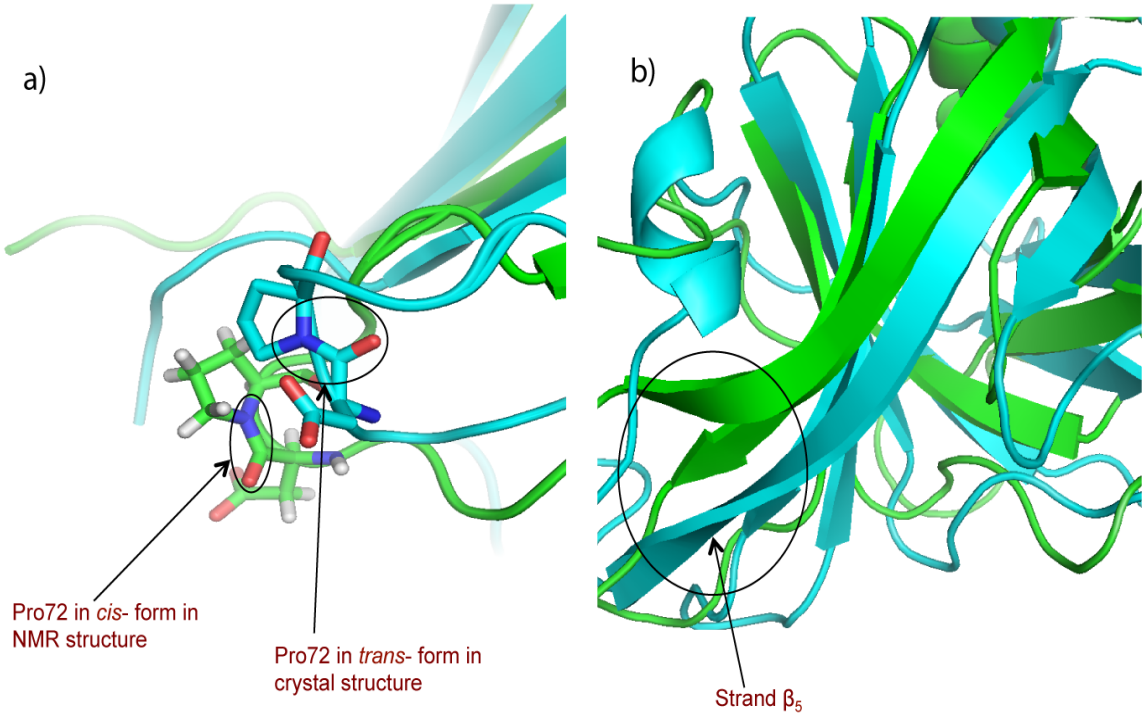

Figure 22: a) shows the overlay in the Pro72 region between NMR (green) and crystal (cyan) structures. b) shows the overlay in the β_5 regions (NMR structure – green, crystal structure – cyan)

3.8. *Pae*PE177-DNA interaction by Gel-Filtration

100 µM sample of *Pae*PE177 eluted at the volume of 12.9 ml (226.4 mAU) from the Superdex 75 column as a monomer (calculated molecular weight 20,137.25 Da)

(**Figure 23**). 50 μM of DNA oligo D18 eluted slightly earlier at 11.95 ml with higher absorbance (989.3 mAU). 50:50 μM *Pae*PE177:D18 eluted as a single peak at 11.6 ml (863.7 mAU) indicating complex formation. The slight decrease in the absorbance is further evidence for complex formation due to quenching. 50:25 μM *Pae*PE177:D18 eluted at 11.6 ml (414.9 mAU) as well but a small shoulder corresponding to free *Pae*PE177 was seen. An intense peak (3323 mAU) was seen for 50:200 μM *Pae*PE177:D18 at 11.95 ml corresponding to free DNA oligo but a shoulder at 11.6 ml was seen corresponding to a complex.

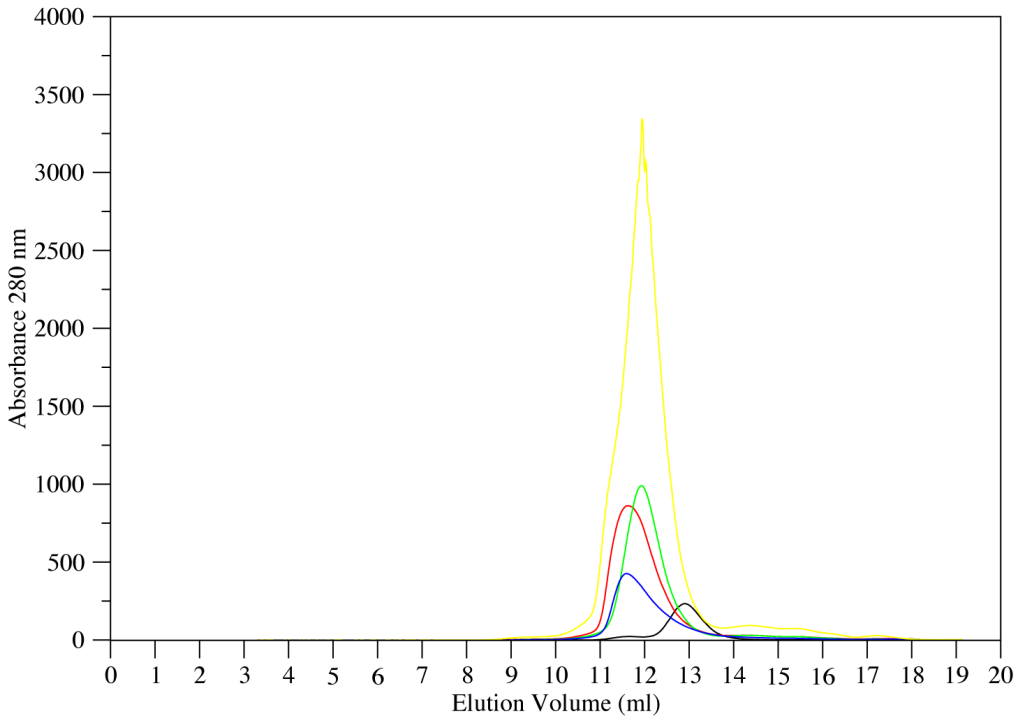

Figure 23: Overlay of gel filtration chromatograms of *Pae*PE177 only (black), D18 only (green), *Pae*PE177:D18-1:1 (red), *Pae*PE177:D18 – 1:0.5 (blue) and *Pae*PE177:D18 – 1:4 (yellow).

3.9. *Pae*PE177-DNA interaction by NMR

There are several positively charged residues distributed over the surface of *Pae*PE177. Any or all of these residues could potentially acts as loci for DNA binding.

However, a binding event that is likely to be catalytically productive should involve the same face as the catalytic site thus allowing the bound DNA to access it with minimal distortions of the sugar-phosphate backbone. Thus, it is expected that DNA binding in *Pae*PE177 could involve both productive and non-productive binding modes.

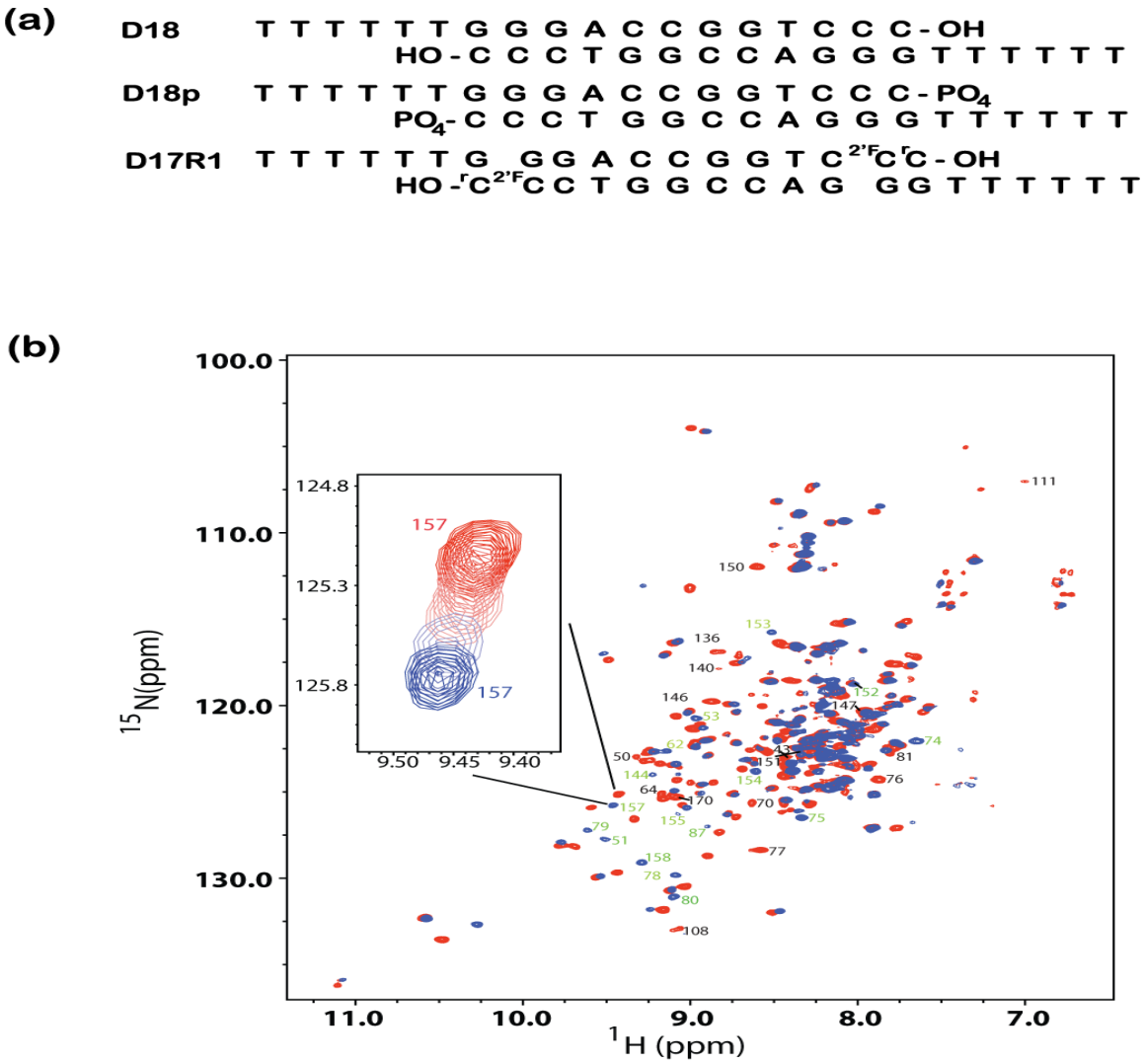

Figure 24: (a) DNA primer-template duplex used in titrations. (b) Overlay of $^{15}N,^{1}H$-TROSY of *Pae*PE177 (apo) and *Pae*PE177:D18 – 1:1. The residues attenuated are labeled in black and residues that are shifting are labeled in green. The inlet shows shifts for the entire course of titration for Leu157.

In order to determine the DNA-binding surface of *Pae*PE177, the spectral perturbations induced in $^{15}N,^{1}H$ TROSY spectra of uniformly $^{15}N,^{2}H$-labeled protein by increasing amounts of the 18-mer primer-template duplex D18 (see **Figure 24a**) containing 5'-overhangs were analyzed. The presence and length of a 5'-tail affects the efficiency of primer-template processing by PE (Zhu Shuman, 2006). A large number of spectral perturbations were seen in the presence of D18 (see **Figure 24b**) reflecting the alteration of the chemical environment of the main chain nuclei due to a combination of direct binding and conformational changes. A large number of residues were broadened beyond the threshold of detection early in the titration series, some at a protein: DNA ratio as low as 1:0.05 (see **Table 3**). These largest perturbations were seen for: (1) several residues near the active site, (2) on the β_3- β_4 loop, (3) on strand β_4, (4) on strand β_5, and (5) on the β_7- β_8 loop (**Figure 25**). In addition several positively charged residues including His127, Lys136 and Arg140 were also completely broadened out early during the titration. This extensive line-broadening indicates exchange on the intermediate timescale (Zuiderweg 2002). Of these residues several could presumably involve a productive binding event since their involvement would enable the DNA duplex to access the catalytic residues with minimal distortions. On the other hand, β_5 lies on the opposite face of the protein and while it could certainly comprise a binding event, a substantial distortion would be required of DNA molecule bound to this patch to access the catalytic site. A conformational change in this region in response to DNA binding at the opposite face, while also possible, is difficult to rationalize without further evidence. In addition to these extensive line-broadening effects, large chemical shift changes were also seen for several residues in and around the strands β_4 and β_8. These residues too lie on the same

face as the catalytic site. It is therefore likely that the spectral perturbations in the presence of D18 seen result from both productive and non-productive binding events.

Given the demonstrated ability of *Pae*PE177 to catalyze the hydrolysis of the 3'-PO_4 for an all-DNA primer-template (Zhu Shuman, 2005), the presence of a phosphate group at the 3'-end (D18p, see **Figure 24a**) was tested for any alteration in the spectral perturbations. The overall distribution of the perturbations was quite similar, as were the nature of chemical shift changes (there were some small differences < 0.05 ppm seen near the active site). However, the residues on β_5, the putative non-productive binding site, seemed to be broadened at a higher protein:DNA ratio compared to D18 (see **Table 3**). Thus, the presence of a 3'-PO_4 group did not seem to have a substantial effect on the nature of spectral perturbations at least in the absence of the catalytic Mn^{+2} ions.

The preferred substrate for sealing by LigD LIG domain consists of a ribonucleotide at the 3'-end, the end product of PE activity. Given this scenario, the influence of primer-template containing a ribonucleotide 3'-end on the spectral perturbations in *Pae*PE177 was checked. The primer-template construct (D17R1, **Figure 24a**) that is also 2'-fluoro modified at the penultimate base was used in the NMR titrations. This construct cannot be cleaved by PE and provides possible access to the catalytic conformations in the presence of Mn^{+2} ions. In the presence of D17R1, dramatic changes in the nature of chemical shift perturbations were seen compared to D18 and D18p. The perturbations for residues comprising β_5 were no longer seen (see **Table 3, Figure 26**) and all line-broadening effects were limited to residues near the active site (Tyr37, Gln40 and Lys41), the β_3-β_4 loop (Val64, Leu70), strand β_4 (Arg76, Leu77, Ala78, Val79, Val81, Glu82 and His84), strand β_7 (Thr146 and Asn147), the β7-β8 loop

(Gly150). In addition, large chemical shift changes (> 0.11 ppm at an equimolar ratio with D17R1) were seen near the catalytic site (Glu50: 0.12 ppm), the β_3-β_4 loop (Val74: 0.15 ppm; Lys75:0.25 ppm), β_4 (Gln80: 0.11 ppm), β_7 (Ile144: 0.16 ppm), β_7-β_8 loop (Gln152: 0.22 ppm, Ser153: 0.12 ppm), β_8 (Gln154: 0.33 ppm, Trp155: 0.37 ppm, Phe156: 0.13 ppm, Leu157: 0.2 ppm and Lys159: 0.15 ppm).

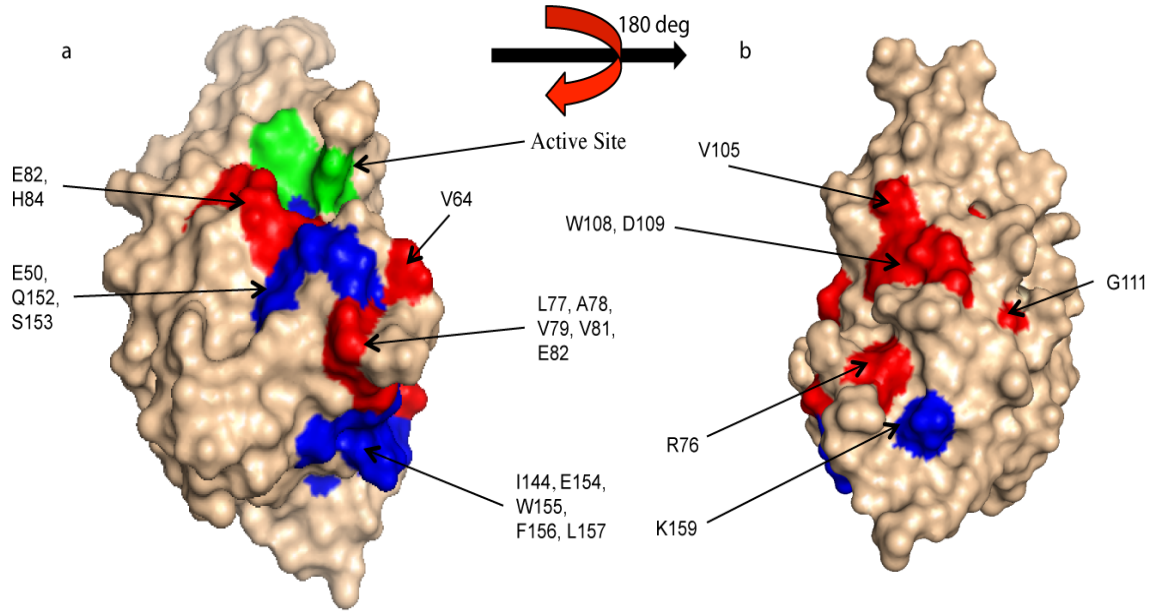

Figure 25: Surface map of *Pae*PE177 in the presence of D18/D18p is shown in wheat with the active site in green. The residues that are attenuated are in red and those showing chemical shift perturbations are in blue.

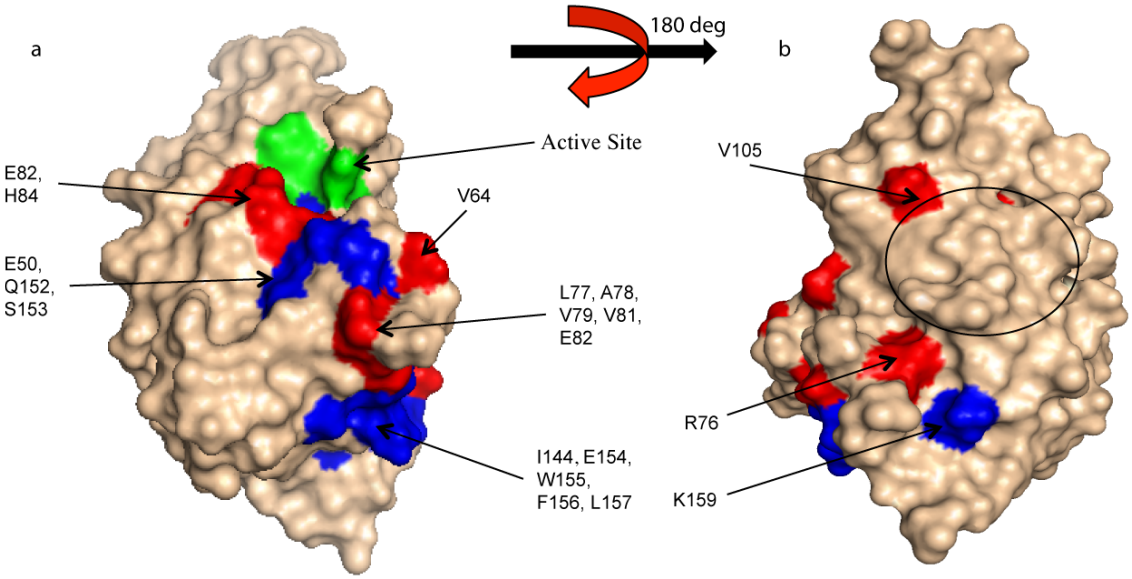

Figure 26: Surface map of *Pae*PE177 in the presence of D17R1 is shown in wheat with the active site in green. The residues that are attenuated are in red and those showing chemical shift perturbations are in blue.

Table 3. Residues on *Pae*PE177 that are broadened out during NMR titration with the primer-template duplexes[a]

Residue	Secondary Structure	D18	D18p	D17R1
Tyr37	β1	-	-	0
Gln40	β1	5	-	-
Lys41	β1	-	-	3
Asp43		-	4	2
Asp50	β2	2	4	-
Phe51	β2	2[c]	2[c]	-
Trp62	β3	-	-	0[b]
Val64		1	1	1
Leu70		1	2	1
Arg76	β4	1	1	3
Leu77	β4	1	1	1
Ala78	β4	1[c]	2[c]	3
Val79	β4	1[b]	4	2
Val81	β4	2	2	0
Glu82	β4	-	-	3
His84	β4	-	-	3
Asp87		3	2	-
Val105	β5	-	-	1
Trp108	β5	1	4	-
Asp109	β5	1	3	-
Gly111	β5	4	3	-
His127	β6	2	2	-
Lys136		2	4	-
Arg140		2	2	-
Ile144	β7	3[c]	-	-
Thr146	β7	1	1	1
Asn147	β7	1	2	2
Leu148		4	4	-
Gly150		1	1	3
Lys151		0	2	-
Trp155	β8	2[b]	-	-
Trp158	β8	-	-	2[b]
Ala160		3	4	-
Leu174		-	-	2
Glu176		-	-	5

[a]R = (concentration of dsDNA)/(concentration of *Pae*PE177). 0: Not seen in the presence of DNA; 1: not seen at R > 0.05, 2: not seen at R > 0.1; 3: not seen at R > 0.2; 4: not seen at R > 0.3 and 5: not seen at R > 0.5. Protein concentration was 100 mM in all the titrations and R is the ratio of the concentration protein to the concentration of DNA. The "-" indicates that the resonance corresponding to the residue was visible throughout the titration series up to a R value of 1.
[b]Reappears at an R value of 0.5.
[c]Reappears at an R value of 1.0.

Based on these observations, we may conclude that the binding site for DNA involves the sheet composed of β_4, β_7 and β_8. In the absence of the ribose base, DNA binding occurs to additional non-productive sites at the opposite face of the protein and the extensive line-broadening effects likely corresponding to exchange between the productive and non-productive sites. The presence of the ribose base enables the selection of the productive sites. However, some degree of exchange with the non-productive sites even for D17R1 cannot be completely ruled out given the line-broadening effects seen for the productive site residues. It is possible that these effects can be largely eliminated under low salt conditions that are difficult to achieve for *Pae*PE177 due to sample stability issues. No significant changes in the chemical shifts were seen for the N-terminal segment important for the monoesterase activity for any of the primer-templates used in this study. This suggests that this segment does not seem to play any role in protein-DNA interactions.

Analysis of the charged residues present on the binding surface based on the sequence analysis of putative PE domains in the prokaryal and archaeal taxa indicated that Lys75, Arg76, Arg145 and Lys159 are highly conserved. This further highlights the importance of these residues in DNA binding.

^{15}N-NOESY-TROSY experiment on a 1:1 ^{15}N,^{2}H-*Pae*PE177:D18 complex showed direct interaction between DNA protons and the protein backbone amide protons. NOE cross peak corresponding to one of the DNA ribose moiety proton (4.1 ppm) and one of the DNA nucleotide base proton (7.43 ppm) were seen for Lys75 and Leu157 respectively (**Figure 27**).

PDBsum (Laskowski 2009) identified the presence of two clefts with volume 725.20 $Å^3$ (near the active site) and 254.39 $Å^3$ in *Pae*PE177 solution structure. Interestingly, these clefts match well with the binding site identified by NMR studies (see **Figure 28**).

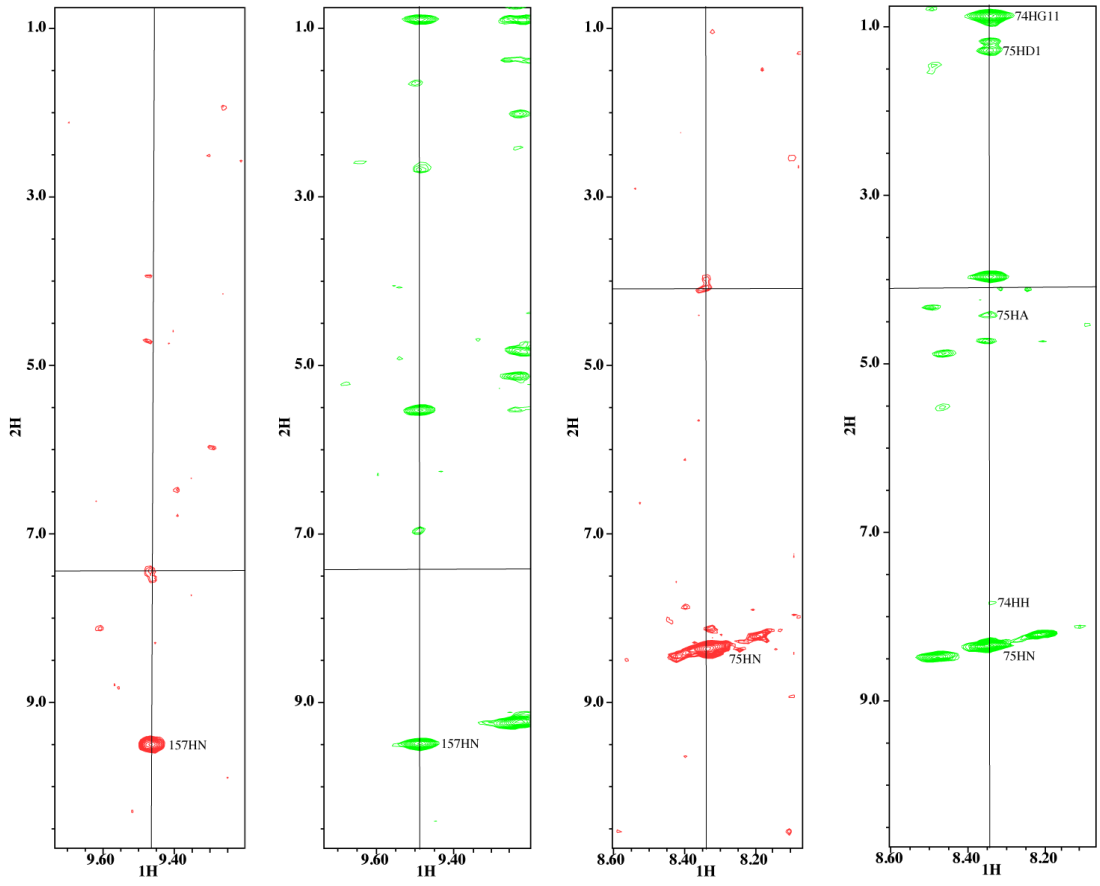

Figure 27: ^{15}N-NOESY-TROSY spectrum strips for ^{15}N,^{2}H-*Pae*PE177:D18 (1:1) in red and ^{15}N,^{13}C,^{1}H-*Pae*PE177 in green. The DNA NOE cross peak is indicated in the crosswire.

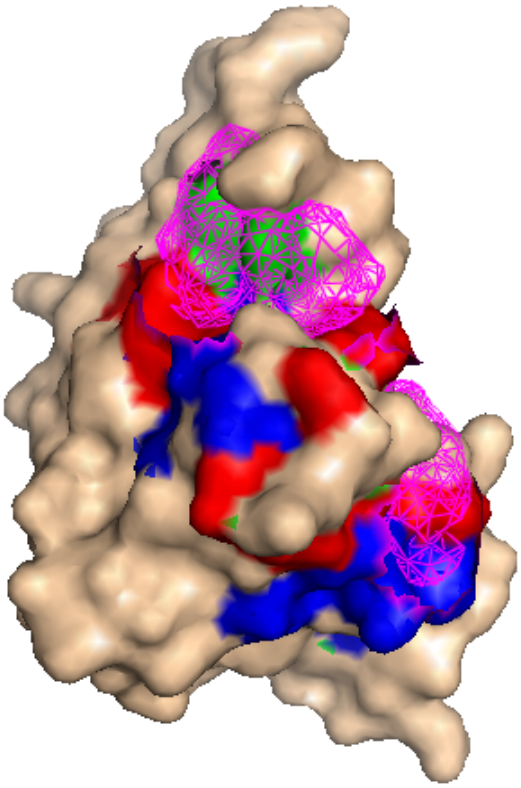

Figure 28: Surface map of *Pae*PE177 is shown in wheat with the active site in green. The residues that are attenuated are in red and those showing chemical shift perturbations are in blue. The clefts are shown in magenta sticks.

Addition of Mn^{2+} to $^{15}N,^{2}H$-*Pae*PE177:D17R1 (1:0.5) sample in the presence of phosphate ions did not produce dramatic change in the $^{15}N,^{1}H$-TROSY spectrum except that a few residues near the active site were attenuated. The N-terminal 32 residues, which are required for the monoesterase activity did not shift or change in intensity indicating no involvement by them. This might be because of low amounts of manganese used in this experiment. Using high amounts of manganese ions in this experiment lead to quenching and non-meaningful results. The role of the N-terminus has to be probed by other experimental methods.

3.10. *Pae*PE177-DNA interaction by fluorescence spectroscopy

In addition to the NMR titration experiments discussed above, the native Trp fluorescence for *Pae*PE177 was used to probe its interations with duplex DNA. The protein contains five Trp residues (62, 108, 113, 141 and 155). All but one (108) of them is buried as seen in the NMR structure. The fluorescence emission maximum was around 331 nm following excitation at 280 nm. A large change in fluorescence emission was seen in the presence of D17R1. The binding isotherm was fit to a simple one-site binding model (**Equation 4**) yielding an apparent K_d value of 5.6 ± 0.6 μM (**Figure 29**). However, improvement in the overall quality of the fit was obtained using a Hill equation (**Equation 5**) yielding a value of 1.4 ± 0.1 for the Hill co-efficient suggesting positive cooperativity. Note that the change in fluorescence emission is not due to direct quenching by DNA, rather due to a conformational change in the protein. Improved fits to a quadratic isotherm were obtained in the presence of either Mg^{+2} or Mn^{+2} with concurrent decreases in the Hill co-efficient to around 1. Neither of these metal ions resulted in any significant change in fluorescence emission in the absence of DNA. These results are summarized in **Table 4**. As stated before, PE relies exclusively on Mn^{+2} ions for catalytic activity. The fluorescence results suggest that Mn^{+2} binding do not cause any significant conformational changes in the protein core. The similar effects of Mg^{+2} and Mn^{+2} on the seeming loss of cooperativity, is more difficult to rationalize. It could simply be the result of an increased ionic strength in the buffer. Interestingly, I was also unable to fit the chemical shift changes in the NMR titrations described above to a simple quadratic binding isotherm (**Equation 4**). Improved fits were obtained when using the

corresponding Hill equation with all residues yielding n values > 1, indicating positive co-operativity in DNA binding in the absence of metal ions.

Table 4. Fluorescence analysis of the interaction of *Pae*PE177 with D17R1

State	K_d (quadratic)	K_d (Hill)	*n*	Var_Q	Var_H
D17R1	5.62 ± 0.62 μM	8.04 ± 0.58 μM	1.36 ± 0.07	1.06 x 10^9	3.04 x 10^8
D17R1 + Mn^{+2}	3.00 ± 0.14 μM	3.85 ± 0.16 μM	1.13 ± 0.05	1.67 x 10^8	1.67 x 10^8
D17R1 + Mg^{+2}	2.75 ± 0.13 μM	3.51 ± 0.15 μM	1.06 ± 0.05	2.11 x 10^8	2.64 x 10^8

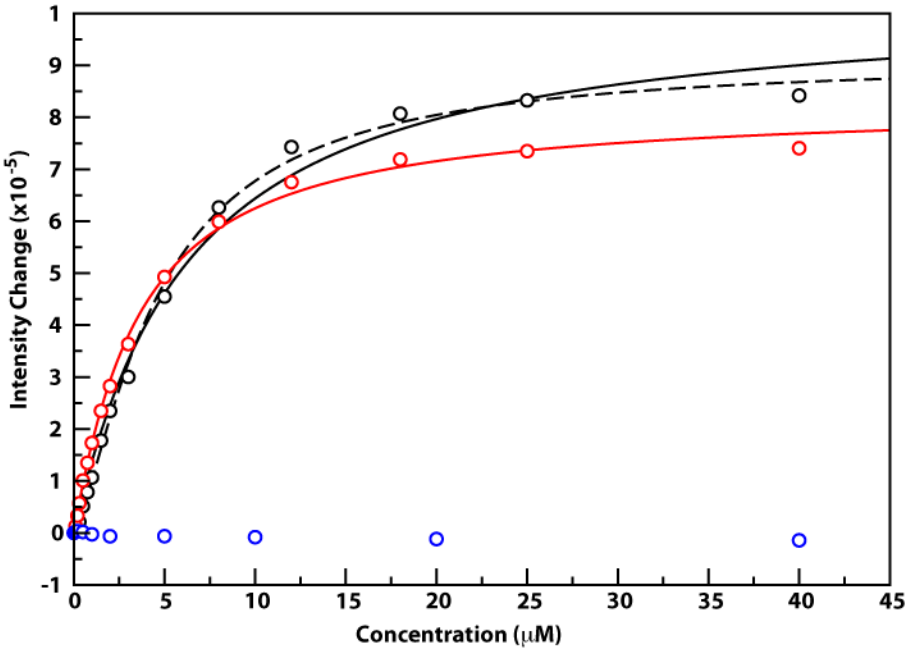

Figure 29: Fluorescence binding curves: Blue circles – *Pae*PE177 with Mn^{2+} (absence of D17R1), Black solid - *Pae*PE177 with D17R1 (absence of Mn^{2+}) – Quadratic fit, Black dotted - *Pae*PE177 with D17R1 (absence of Mn^{2+}) – Hill fit, Red circles - *Pae*PE177 plus Mn^{2+} with D17R1.

4. Conclusion

Most of the cells, other than those in laboratory cultures, tend to stay in the stationary phase to survive the harsh environmental conditions. NHEJ has been shown to be critical for cells in stationary phase. NHEJ is also crucial for survival against IR and desiccation. *Mycobacterium tuberculosis*, needs NHEJ to survive the human immune system by overcoming the oxidative stress it experiences in macrophages. NHEJ has been shown to be important to spore forming bacteria like *Bacillus subtilis* (Shuman Glickman, 2007). Bacterial NHEJ is not yet completely understood. There are still regulators and additional components that could play an role in NHEJ.

This study has identifed a class of DNA repair proteins that has a unique fold. Further, its interactions with DNA substrate and the mode of binding has been identified. The void that was caused by the lack of atomic structure for PE domain in LigD has been filled. The study is also an indicator of the power of NMR to study a highly dynamical protein.

Once the final NMR structure is obtained with the addition of the N-terminal tail to the structure, the next logical step would be to calculate the *Pae*PE177:DNA complex structure through NMR. Complex structure determination requires the complete assignments of the DNA oligo. Experiments like HNCCCH, HCCNH-TOCSY and HCCH-TOCSY can be used for DNA residue identification (Greenbaum, Ghose 2010). NOESY experiments using intra and internucleotide correlations can be used for sequential assignments as well as generating structural restraints for structure calculations. Once the DNA oligo structure is calculated, intermolecular NOE using

isotope filtered experiments (Sattler et al, 1999) on labeled DNA/unlabeled protein and unlabeled DNA/labeled protein can be used to generate restraints for complex structure calculations. These types of experiments has already been shown feasible in the previous section (**Figure 27**). Moreover, Paramagnetic relaxation enhancement (PRE) measurements of ^{1}H using conjugated dT-EDTA chelated with Fe^{2+} (paramagnetic centers) at specific DNA sites can be used to generate long-range distance restraints (Iwahara et al, 2003).

Further NHEJ is a tightly regulated process and there is communication between different LigD domains and bacterial Ku. PE's interaction with POL and LIG domains can be extensively studied by NMR. Labeled PE can be titrated with unlabeled POL and LIG domains and the chemcial shifts of PE can be monitored. The deleterious effects of relaxation can be overcome by using POL and LIG domains that are purified from cultures grown in D_2O with unlabeled ammonium chloride and glucose. This study can map the binding pocket of POL and LIG domains on PE.

5. References

Akey D, Martins A, Aniukwu J, Glickman MS, Shuman S, Berger JM, Crystal structure and nonhomologous end-joining function of the ligase component of Mycobacterium DNA ligase D, J Biol Chem. 2006, 281: 13412-13423.

Aravind L, Koonin EV, Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system, Genome Res. 2001, 11:1365-1374.

Augustin MA, Huber R, Kaiser JT, Crystal structure of a DNA-dependent RNA polymerase (DNA primase), Nature Struc. Biol., 2001, 8: 57-61.

Berjanskii MV, Wishart DS, A simple method to predict protein flexibility using secondary chemical shifts, J Am Chem Soc, 2005, 127: 14970-14971.

Blier PR, Griffith AJ, craft J, Hardin JA, Binding of Ku protein to DNA, J.Biol. Chem., 1993, 268: 7594-7601.

Bodenhausen G, Ruben DJ, Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, Chemical Physics Letters, 1980, 69: 185-189.

Boggs PT, Donaldson JR, Byrd RH, Schnabel RB, Algorithm676 ODRPACK: Software for weighted orthogonal distance regression, ACM Trans. Math. Software, 1989, 15: 348-364

Cavanagh J, Fairbrother WJ, Palmer AG, Rance M, Skelton NJ, Protein NMR spectroscopy: Principles and practice, 2007, Elsevier academic press.

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.

Daley JM, Palmbos PL, Wu D, Wilson TE, Nonhomologous end joining in yeast, Annu Rev Genetics. 2005, 39: 431-451.

Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, Topper LM, Pitcher RS, Tomkinson AE, Wilson TE, Doherty AJ, Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine, Science, 2004, 306: 683-685.

Englander J, Cohen L, Arshava B, Estephan R, Becker JM, Naider F, Selective labeling of a membrane peptide with ^{15}N-amino acids using cells grown in rich medium, Biopolymers, 2006, 84: 508-518.

Ferrage F, Cowburn D and Ghose R, Accurate sampling of high-frequency motions in proteins by steady-state ^{15}N-{^{1}H} nuclear Overhauser effect measurements in the presence of cross-correlated relaxation, J Am Chem Soc, 2009, 131: 6048-6049.

Ferrage F, Riechel A, Bhattacharya S, Cowburn D, Ghose R, On the measurement of ^{15}N-{^{1}H} nuclear Overhauser effects 2. Effects of the saturation scheme and water signal suppression.*J. Magn. Reson.2010,* 207, 294-303.

Gong C, Martins A, Bongiorno P, Glickman M, Shuman S, Biochemical and genetic analysis of the four DNA ligases of mycobacteria, J. Biol. Chem., 2004, 279: 20594-20606.

Gong C, Bongiorno P, Martins A, Stephanou NC, Zhu H, Shuman S, Glickman MS, Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C, Nat Struct Mol Biol. 2005, 12:304-312.

Greenbaum NL, Ghose R, Nuclear magnetic resonance (NMR) spectroscopy:Structure determination of proteins and nucleic acids. *In Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd.:Chichester, 2010.

Grzesiek S, Anglister J, Bax, A., Correlation of backbone amide and aliphatic side-chain resonances in $^{13}C/^{15}N$-enriched proteins by isotropic mixing of ^{13}C magnetization, 1993, J. Magn. Reson., B 101: 114-119.

Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR, XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps, EMBO J., 2007, 26: 1010-1023.

Habeck M., Rieping W, Linge, JP, Nilges M, NOE assignment with ARIA 2.0: the nuts and bolts, Meth. Mol. Biol., 2004, 278: 379-402.

Holm L, Rosenström P, Dali server: conservation mapping in 3D. Nucl. Acids Res., 2010, 38, W545-549.

Hutchinson EG, Thornton JM, PROMOTIF - a program to identify and analyze structural motifs in proteins, Protein Sci., 1996, 5: 212-20.

Iwahara J, Anderson E, Murphy EC, Clore MG, EDTA-derivatized deoxythymine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement, J. Amer. Chem. Soc., 2003, 125: 6634-6635.

Johnson BA, Using NMRView to visualize and analyze the NMR spectra of macromolecules. Methods Mol Biol, 2004, 278: 313-52.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, J. Biomol. NMR, 1996, 8: 477-486.

Laskowski RA, PDBsum new things, Nucleic Acids Res., 2009, 37, D355-D359.

Lieber MR, The mechanism of human nonhomologous DNA end joining, J.Biol.Chem., 2008, 283:1-5.

Lieber MR, The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway, Annu. Rev. Biochem., 2010, 79:181-211.

Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M, Refinement of protein structures in explicit solvent, Proteins: Struc. Funct. Genet., 2003, 50: 496-506.

Ma Y, Pannicke USchwarz K, lieber MR, Hairpin opening and overhang processing by an Artemis:DNA-PKcs complex in V(D)J recombination and in nonhomologous end joining, Cell, 2002, 108:781-794.

Marassi FM, Opella SJ, NMR structural studies of membrane proteins, Curr Opin Struc Biol., 1998, 8:640-648.

Marsh JA, Singh VK, Jia Z, Forman-Kay, JD, Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, Protein Sci, 2006, 15: 2795-804.

Mossessova E, Lima CD, Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast, Mol Cell., 2000, 5:865-76.

Nair PA, Smith P, Shuman S, Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily, Proc Natl Acad Sci USA, 2010, 107: 12822-12827.

Ottiger M, Delaglio F, Bax A, Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, J.Magn.reson., 1998, 1311:373-378.

Pervushin, K, Riek R, Wider G, Wuthrich K, Attenuated T_2 relaxation by mutual cancellation of dipole-dipole oupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proc. Natl. Acad. Sci. U.S.A, 1997, 94:12366-12371.

Pitcher RS, Tonkin LM, Green AJ, Doherty AJ, Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis, J Mol Biol. 2005, 351:531-544.

Pitcher RS, Brissett NC, Picher AJ, Andrade P, Juarez R, Thompson D, Fox GC, Blanco L, Doherty AJ, Structure and function of a mycobacterial NHEJ DNA repair polymerase, J Mol Biol. 2007, 366:391-405.

Pitcher RS, Brissett NC, Doherty AJ, Nonhomologous End-joining in Bacteria: A microbial perspective, Annu. Rev. Microbiol. 2007, 61: 259-282.

Prestegard JH, Bougault CM, Kishore AI, Residual dipolar couplings in structure determination of biomolecules, Chem. Rev., 2004, 104:3519-3540.

Ramachandran GN, Ramakrishnan C, Sasisekharan V, Stereochemistry of polypeptide chain configurations, J.Mol.Biol., 1963, 7: 95-99.

Ramadan K, Shevelev I, Huebscher U, The DNA-polymerase-X family: controllers of DNA quality? Nature Rev. Mol. Cell. Biol., 2004, 5: 1038-1043.

Ruekert M, Otting G, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, J. Am. Chem.Soc, 2000, 122:7793-7797.

Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich, K, TROSY in triple-resonance experiments: New prospectives for sequential NMR assignment of large proteins, Proc. Natl. Acad. Sci. U.S.A., 1998, 95: 13585-13590.

Sattler M, Schleucher J, Griesinger C, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog. Nucl. Magn. Reson. Spec., 1999, 34: 93-158.

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.

Shen Y, Delaglio F, Cornilescu G, Bax A, Talos+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J. Biomol. NMR, 2009, 44: 213-223.

Shen Y, Bax A, Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts, J Biomol NMR, 2009, 46: 199-204.

Shuman S, Glickman MS, Bacterial DNA repair by non-homologous end joining, Nat Rev Microbiol., 2007, 11: 852-861.

Venters RA, Farmer BT, Fierke CA, Spicer LD, Characterizing the use of perdeuteration in NMR studies of large proteins: ^{13}C, ^{15}N and ^{1}H assignments of human carbonic anhydrase II, J Mol Biol, 1996, 264: 1101-16.

Weller GR, Doherty AJ, A family of DNA repair ligases in bacteria?, FEBS Lett. 2001, 505:340-342.

Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, Della M, Devine SK, Day JP, Wilkinson A, d'Adda di Fagagna F, Devine KM, Bowater RP, Jeggo PA, Jackson SP, Doherty AJ, Identification of a DNA nonhomologous end-joining complex in bacteria, Science. 2002, 297:1686-1689.

Wishart DS, Sykes BD, The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data, J Biomol NMR.,1994, 4:171-80.

Zhu H, Wang LK, Shuman S, Essential constituents of the 3'-phosphoesterase domain of bacterial DNA ligase D, a nonhomologous end-joining enzyme, J Biol Chem. 2005, 280:33707-33715.

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, 280:25973-25981.

Zhu H, Shuman S, A primer-dependent polymerase function of pseudomonas aeruginosa ATP-dependent DNA ligase (LigD), J Biol Chem. 2005, 280:418-427.

Zhu H, Nandakumar J, Aniukwu J, Wang LK, Glickman MS, Lima CD, Shuman S, Atomic structure and nonhomologous end-joining function of the polymerase component of bacterial DNA ligase D, Proc Natl Acad Sci U S A. 2006, 103:1711-1716.

Zhu H, Shuman S, Substrate specificity and structure-function analysis of the 3'-phosphoesterase component of the bacterial NHEJ protein, DNA ligase D, J Biol Chem. 2006, 281:13873-13881.

Zhu H, Shuman S, Characterization of *Agrobacterium tumefaciens* DNA ligases C and D, Nucleic Acids Res., 2007, 35: 3631-3645.

Zhu H, Shuman S, Bacterial nonhomologous end joining ligases preferentially seal breaks with a 3'-OH monoribonucleotide, J Biol Chem. 2008, 283:8331-8339.

Zuiderweg ER, Mapping protein-protein interactions in solution by NMR spectroscopy, Biochemistry, 2002, 41:1-7.