2007

A STRUCTURE-FUNCTION STUDY OF THE HUMAN DNA POLYMERASE LAMBDA (λ)

Francis Mwai Kariuki
Western University

A STRUCTURE-FUNCTION STUDY OF THE HUMAN DNA POLYMERASE LAMBDA (λ)

(Thesis format: Monograph)

by

Francis Mwai Kariuki

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF GRADUATE STUDIES

CERTIFICATE OF EXAMINATION

Supervisor	Examiners
Dr. Hong Ling	Dr. Harvey A. Goldberg
Supervisory Committee	Dr. Megan J. Davey
Dr. David B. Haniford	Dr. Thomas G. Linn
Dr. James Choy	

The thesis by

Kariuki, Francis Mwai

entitled:

A Structure-Function Study of the Human DNA Polymerase Lambda (λ)

is accepted in partial fulfillment of the
requirements for the degree of
Master of Science

Date_____

_____
Chair of the Thesis Examination Board

ABSTRACT

The human DNA polymerase lambda (Pol λ) is a recently discovered member of the X-family DNA polymerases that has been shown to be involved in DNA repair and meiotic recombination, where it plays a gap filling role in these two processes. The protein's primary sequence is organized into an N-terminal BRCA1-C terminal (BRCT) domain, a proline-rich domain and a C-terminus polymerase beta-like (Pol β-like) domain. Studies have shown that the polymerase's activity is localized at the C-terminus of the protein and that this activity is partially suppressed by the proline-rich domain. Primer extension assays confirmed the inhibitory effects of the proline-rich domain, however, it also showed that the BRCT domain appeared to subdue or compensate for this effect. This assay also revealed that the proline-rich domain by itself did not significantly increase the protein's fidelity but when combined with the BRCT domain, there was approximately a 10 fold increase in fidelity. DNA binding assays also showed that the proline-rich domain inhibited the C-terminus domain's ability to bind DNA; however, it was also observed that this negative effect was suppressed or compensated for by the BRCT domain. NMR and CD data analysis confirmed secondary structure predictions that indicated the proline-rich domain to be essentially a disordered region. The NMR experiment further presented substantiation that there was a weak domain-domain interaction between the proline-rich domain and the C-terminus domain, which may play a crucial role in the protein function.

Keywords: human DNA polymerase lambda, X-family DNA polymerases, proline-rich domain, BRCT domain, core domain, protein activity, protein fidelity, DNA binding, NMR spectroscopy

ACKNOWLEDGEMENTS

First I would like to express gratitude to my supervisor, Dr. Hong Ling who gave me the opportunity to work on this project. She has been my supervisor for the past 4 years while I worked in her laboratory as a 4^{th} year undergraduate student, a summer student, a research assistant and finally as a masters student. I would like to thank her for all her support, patience, encouragement and valuable knowledge and experience she has provided over the years.

Next I would like to express appreciation to my advisory committee: Dr. David Haniford and Dr. James Choy, for their very useful suggestions and guidance throughout my project. When the project seemed to hit a road block, they provided their support and expertise to help me overcome the challenges. I would also like to say thanks again to Dr. Choy and members of his lab in helping me carry out the NMR experiments.

I would also like to give a special thanks to all the past and present laboratory members who have made the past 4 year in the Ling lab a fun and memorable experience. I would also like to say thanks to Lee-Ann Briere for her technical expertise in running the CD experiments. In addition, I would like to give special thanks the entire University Of Western Ontario Biochemistry department for their hard work, friendship and assistance in making my past 4 years in the department a remarkable experience. I thank everyone for the great job they are doing.

Much gratitude also goes to my parents, my older brother Stephen and younger sister Martha for their constant support and encouragement throughout the university years. They have been there for me through the good times and the tough times and could not ask for a better bunch of people. I love them all dearly and I am forever grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

Amino Acids

Ala (A) alanine
Arg (R) arginine
Asn (N)asparagine
Asp (D)aspartic Acid
Cys (C) cysteine
Gln (Q) glutamine
Glu (E) glutamic acid
Gly (G) glycine
His (H) histidine
Ile (I) isoleucine
Leu (L) leucine
Lys (K) lysine
Met (M) methionine
Phe (F) phenylalanine
Pro (P) proline
Ser (S) serine
Thr (T) threonine
Trp (W) tryptophan
Tyr (Y) tyrosine
Val (V) valine

BER	Base Excision Repair
BSA	Bovine serum albumin
CD	Circular Dichroism
DNA	Deoxyribonucleic acid
D_2O	Deuterium oxide
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
dNTP	deoxyribonucleotide triphosphate
dRPase	deoxyribose 5` phosphate lyase

DSS	2,2-Dimethyl-2-silapentane-5-sulfonate
E.coli	Escherichia coli
EDTA	Ethylene-diamine-tetra-acetic acid
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kiloDaltons
K_2HPO_4	Potassium phosphate dibasic anhydrous
LB	Luria-Bertani
$MgCl_2$	Magnesium Chloride
MS	Mass Spectrometry
MW	Molecular weight
NHEJ	Non-Homologous End Joining repair
N_2HPO_4	Sodium phosphate dibasic anhydrous
NaCl	Sodium Chloride
NMR	Nuclear magnetic resonance spectroscopy
PAGE	Polyacrylamide Gel Electrophoresis
PMSF	Phenyl-methyl-sulfonyl-fluoride
RPC	Reverse-Phase Chromatography
SDS	Sodium Dodecyl Sulfate
TdT	Terminal deoxynucleotidyl transferase
Tris	Tris (Hydroxymethyl) Aminomethane
UV	Ultra Violet

Protein Constructs:

fpol λ	full length DNA polymerase lambda (residues 1-575). (65.6 kDa).
dpol λ	truncated DNA polymerase lambda (residues 132-575). (50.1 kDa)
tdpol λ	truncated DNA polymerase lambda (residues 154-575). (46.7 kDa)
tpol λ	DNA polymerase lambda core domain (residues 245-575). (38.2 kDa)
Proline-rich domain	Proline-rich domain of DNA polymerase lambda (residues 132-244). (12.3kDa)

CHAPTER 1: INTRODUCTION

1.1 DNA damage and repair

DNA based organisms are faced with a great challenge of maintaining genomic stability over many generations. This preservation of genomic stability is complicated by the frequent occurrence of DNA damage that arises spontaneously (e.g. deamination) or is caused by environmental conditions (such as UV and ionizing radiation) as well as many chemical agents [2]. Without any form of repair, the probability of cell survival and the ability to maintain a stable genome would decrease significantly.

A variety of DNA repair machineries have evolved to deal with multiple kinds of damage and genomic alterations which may occur within the DNA as a result of these unfavorable conditions. These systems not only replicate the genomes accurately but also remove damaged nucleotides and replace them with undamaged nucleotides by means of DNA synthesis [3]. Some of the common repair mechanisms utilized by the replication machinery are base excision repair (BER); which is a major pathway of repairing modified bases in DNA (Figure 1A) and non homologous end joining (NHEJ); which is a common way of repairing double strand breaks in DNA (Figure 1B) [2,3,5]. The BER mechanism involves multiple proteins such as DNA glycosylases that remove damaged/modified bases creating an apurinic/apyrimidine (AP) site, AP endonucleases that nick the damaged DNA strand upstream of the AP sites thus creating a 3` hydroxyl terminus adjacent the AP site, DNA polymerases that extend the 3`-OH terminus and lastly, DNA ligases that seal the DNA backbone. The NHEJ repair mechanism also involves multiple proteins such as KU factors (Ku70/Ku80) and DNA dependent protein kinases that recognize breaks in the DNA, recruit repair factors and align the broken DNA, DNA polymerases that process DNA ends and fill the gap in the DNA, and finally

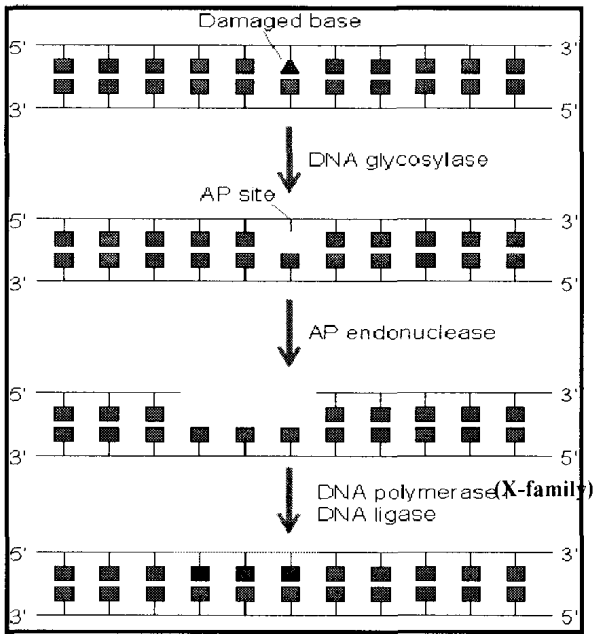
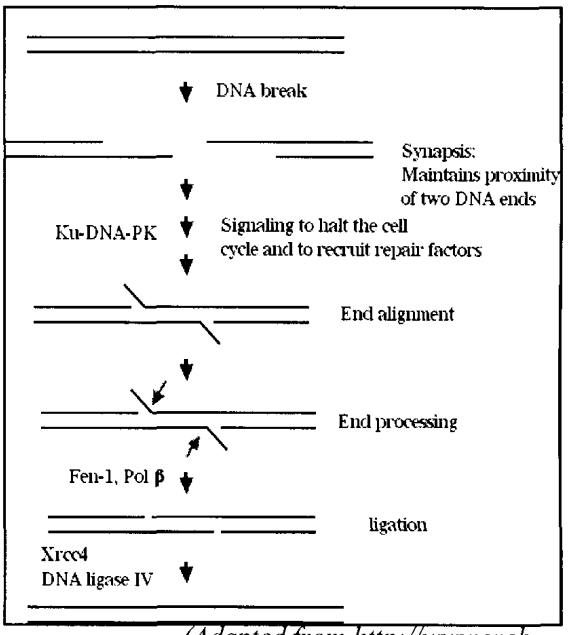


Figure 1A. Schematic representation of Base excision repair. The first step involves removal of the incorrect base by an appropriate DNA N-glycosylase to create an AP (apurinic/apyrimidine) site followed by the nicking of the damaged DNA strand by AP endonuclease upstream of the AP site, thus creating a 3'-OH terminus adjacent to the AP site and then finally extending the 3'-OH terminus by a DNA polymerase, accompanied by excision of the AP site.



(Adapted from http://www.web-books.com/MoBio/Free/Ch7G.htm)

Figure 1B. Schematic representation of the NHEJ pathway of DNA repair. The various steps and proteins involved are listed, beginning with the recognition of a double strand break and ending with the sealing of the DNA backbone.

DNA ligases that seal the DNA backbone. Both of these processes involve multiple proteins to carry out various steps of the repair pathway, however, the handover or switching mechanism between the different players remains unclear. Crucial players involved in both of these DNA repair processes are a group of proteins belonging to the X-family of DNA polymerases, which are involved in the gap filling step in these systems.

1.2 X-family DNA polymerases

The X-family polymerases are generally small enzymes that belong to a subdivision of a larger superfamily of nucleotidyl transferases, and can be found in various organisms ranging from viruses to higher eukaryotes [12]. Members of this family include enzymes like: terminal deoxynucleotidyltransferase (TdT) which is thought to be mainly restricted in the lymphoid tissues and catalyzes template-independent nucleotide addition at V(D)J junctions [11], DNA polymerase beta (Pol β) which is highly expressed in the gametes and removes the 5'-deoxyribose phosphate moiety and catalyzes gap-filling synthesis during BER, DNA polymerase mu (Pol μ) which is preferentially expressed in secondary lymphoid tissues and is suggested to play a role in V(D)J recombination thereby complementing the biological functions of TdT [14], African swine fever virus DNA polymerase X (ASFV PolX) that plays a role in BER analogous to the function of Pol β [17], yeast DNA polymerase IV (Pol IV) that possibly functions in both NHEJ of double strand breaks and BER [18] and yeast DNA polymerase sigma (Pol σ) that couples DNA replication to the establishment of sister chromatid cohesion [15,16].

Although members of this family have been shown to possess a high sequence similarity with each other, they differ considerably in their biochemical properties. Over the past few years, a number of studies on four members of this family: Pol β, Pol λ, Pol μ and TdT, has provided a significant amount of structural and biochemical information [2, 3, 7]. Most studies on the X-family polymerases have been done on Pol β, which has been crystallized in its native form and as different binary and ternary complexes. This Pol β structures have provided a wealth of information into the catalytic cycle of these X-family polymerases, suggesting a conformational change upon substrate binding that correctly positions three carboxylate residues that coordinate two divalent metal ions that participate in catalysis, in a manner similar to other DNA polymerases [3, 10, 11].

1.3 Human DNA polymerase Lambda (Pol λ)

The recently discovered DNA polymerase Lambda (Pol λ) has been identified as a member of the X family of polymerases based on the high degree of sequence similarities with other members: Pol β (54%), Pol μ (47%) and TdT (44%) (Figure 2) [4, 6-8]. The Pol λ primary sequence is organized into 3 main domains consisting of a Breast Cancer Susceptibility Gene 1 C-terminus (BRCT) domain and a proline-rich domain, which make up the N-terminus of the protein, and a polymerase beta like (Pol β like) domain, which makes up the C-terminus of the protein. Due to the high degree of homology, Pol λ is also predicted to possess similar enzymatic properties to that of other X-family polymerases, such as, dRPase (deoxyribose 5` phosphate lyase) activity that involves the removal of a 5`deoxyribose phosphate moiety, TdT activity which is a template independent extension of DNA, and gap-filling polymerase activity [1,3,6, 8,10].

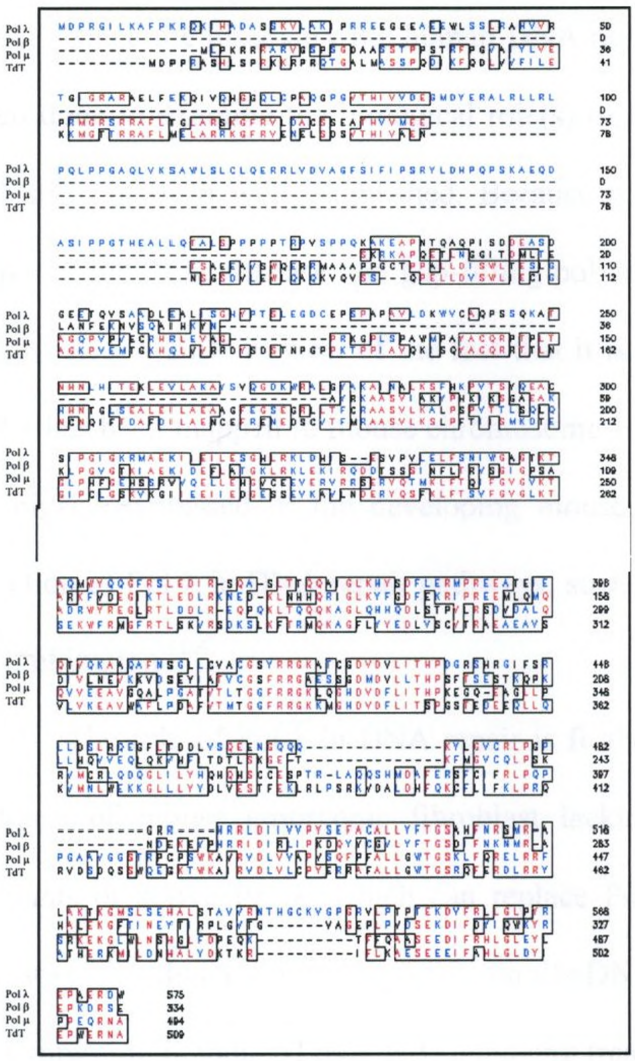
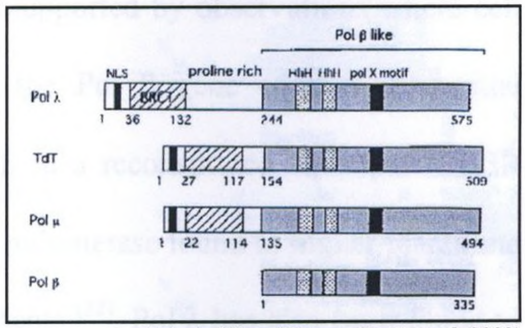


Figure 2A. Multiple sequence alignment of X-family DNA polymerases (Pol λ, Pol β, Pol μ and TdT). Red color indicates identical residues, blue color indicates similar residues and black indicates other residues



(Adapted from Shimazaki, N. et. al. 2002)

Figure 2B. Schematic representation of the X-family DNA polymerases. The figure shows the shared domains between the well documented members of the X-family polymerases and Pol λ. Also shown is the domain arrangement of Pol λ (the BRCT, proline-rich and Pol β-like domains) and other regions of the protein thought to play important roles in the protein activity (nuclear localization signal [NLS], helix-hairpin-helix motif [HhH] and pol X motifs [active center])

1.3.1 Functional role of human DNA polymerase λ in DNA replication and repair

Since the discovery the human DNA polymerase lambda, extensive studies have been done to examine the biological role(s) of the protein; however, the exact function has still not been well established. Because it possesses the two enzymatic activities required for BER; dRPase and gap-filling polymerase activities, it is believed that Pol λ contributes to BER in view of the fact that it is related to Pol β [20]. The gene encoding Pol λ has been mapped to mouse chromosome 19 and like Pol β, Pol λ has been found to be highly expressed in the developing mouse testes, therefore suggesting a possible function of Pol λ in DNA repair pathways, such as BER that are associated with meiotic recombination [19].

The role of Pol λ in DNA repair is further supported by observations where cell extracts of mouse embryonic fibroblast lacking the Pol β gene contain substantial amounts of active Pol λ, which can replace Pol β in a reconstituted short-patch BER pathway [21]. Also, Pol λ is the only X-family DNA polymerase found in higher plants and its expression is induced by DNA-damaging treatments [22]. Pol λ has also been found to protect mouse fibroblasts against oxidative DNA damage and is recruited to oxidative DNA damage sites. Therefore, these studies suggest that Pol λ may complement or support the function of Pol β in BER *in vivo* [23].

Based on recent biochemical data, it has also been proposed that Pol λ plays a biological role in the repair of double-stranded breaks (DSBs) through NHEJ pathways [24,25]. This hypothesis is supported by the analysis of immunodepletion studies that suggest that Pol λ rather than other X-family polymerases, is responsible for the gap-filling synthesis associated with NHEJ in human nuclear extracts [24]. Finally, the ability

of Pol λ to bypass abasic sites in the presence of Mn^{2+} *in vitro*, has led to the suggestion that it may also play a role in bypassing DNA lesions *in vivo* [26, 27].

Although the role of Pol λ *in vivo* has not yet been well characterized, studies have shown that Pol $\lambda^{-/-}$ mice display hydrocephalus, situs inversus, chronic sinusitis and male infertility [9]. And although these Pol λ knock-out experiments are often complicated by the presence of other DNA polymerases, in particular Pol β that could fill in and compensate for the loss of functions of Pol λ [23], the human DNA polymerase lambda has been clearly shown to be an essential protein in maintaining genomic stability and plays a crucial role in DNA repair [20].

1.3.2 Structure of human DNA polymerase λ

Like other family X polymerases, Pol λ is a single-subunit enzyme that lacks $3' \rightarrow 5'$ exonuclease activity. It is comprised of a 575 amino acid polypeptide chain that is organized into 3 major domains [21]: an N-terminal BRCT domain, that is believed to mediate protein–protein interactions based on observations within other proteins that have a similar domain [8], a Pol β-like domain (catalytic core) that is composed of an N-terminal 8 kDa domain unique to X-family polymerases plus a polymerase domain that includes the fingers, palm (that contains the catalytic carboxylates) as well as the thumb subdomains common to all polymerases [30] and finally a proline-rich domain that connects the BRCT domain and the catalytic core (Fig. 3). This proline-rich domain is also present in the yeast homologue Pol IV, and has been suggested to be a target for posttranslational modifications [21] due to the significant number (10%) of serine residues. However, its functional role still remains unclear.

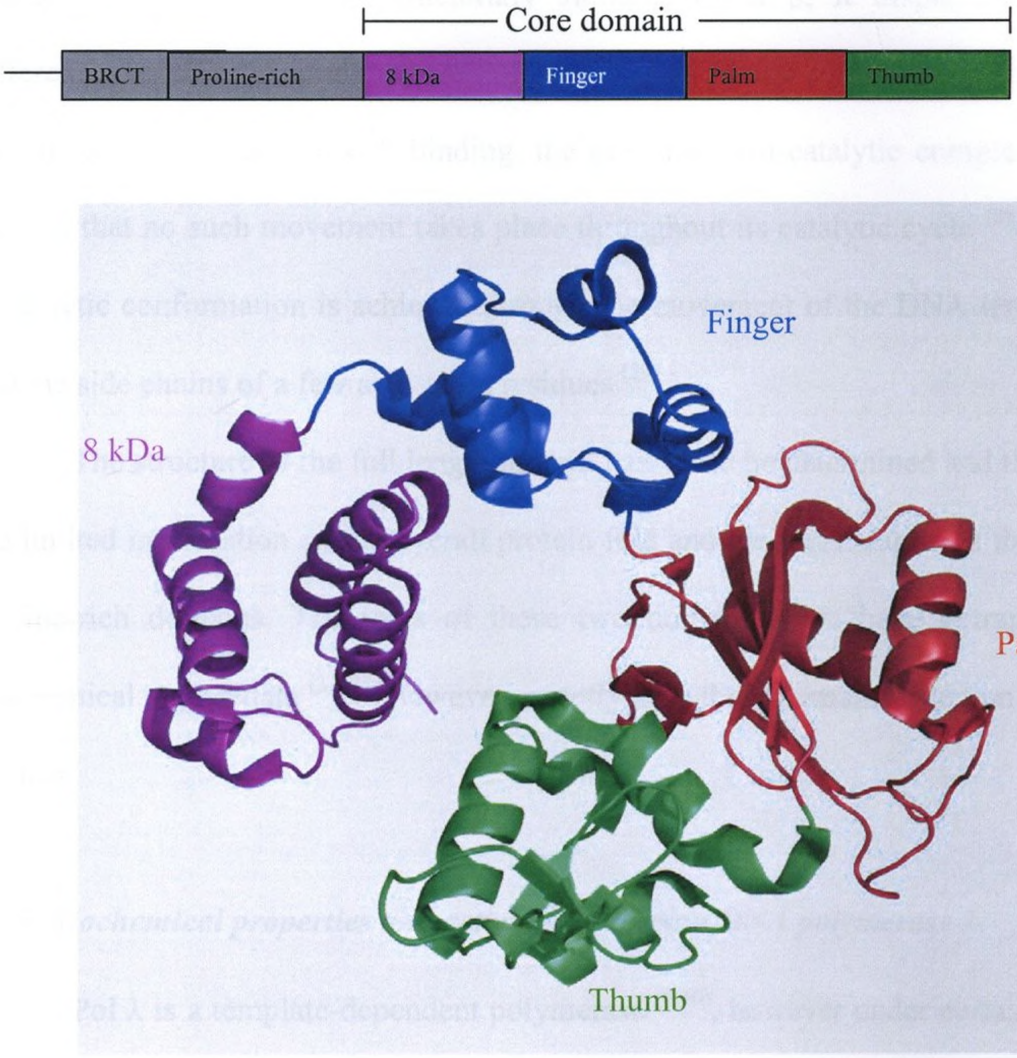


Figure 3. Domain and subdomain organization of Pol λ.
(A) Linear illustration showing the different domains that compose the full-length Pol λ protein: the BRCT domain, Proline-rich domain, 8 kDa domain and the polymerase catalytic domain, composed of finger, palm, and thumb subdomains.
(B) Ribbon representation of the catalytic core of Pol λ generated using MacPymol v0.99 (Delano Scientific LLC). The subdomains of the catalytic core are shown in: purple (8 kDa domain unique to the X-family polymerases), blue (Finger subdomain), red (Palm subdomain) and green (Thumb subdomain).
(PDB 1RZT)

The structure of the catalytic core of human Pol λ has been extensively characterized and although structurally similar to Pol β, it displays some unique differences [7, 8]. For example, where Pol β undergoes a large conformational change in its thumb subdomain upon dNTP binding, the pre- and post-catalytic complexes of Pol λ suggest that no such movement takes place throughout its catalytic cycle [29] and instead, a catalytic conformation is achieved through the movement of the DNA template strand and the side chains of a few active site residues [29].

The structure of the full length protein has yet to be determined and this has led to the limited information on the overall protein fold and the organization of the BRCT and proline-rich domains. The roles of these two domains have been extrapolated from biochemical assays/data [1,4,21] however, exactly how these domains function still remains unclear.

1.3.3 Biochemical properties and activities of human DNA polymerase λ

Pol λ is a template-dependent polymerase [6,10], however under certain conditions, it has been shown to carry out template independent incorporation, although with very low efficiency [31,32]. Studies have also shown that the preferred substrate for Pol λ is a short gap with a phosphate on the 5' end, which is a characteristic shared by other X-family polymerases [8]. Whereas Pol λ is distributive in nature on a primed single-strand template, its polymerase activity is stimulated during synthesis of a short gap containing a 5' phosphate at the end, while at the same time its strand displacement activity is limited [6].

DNA polymerase λ has also been shown to have a high affinity for dNTPs that is attributed in part to the presence of an uncharged side chain (Ala) at residue 510, positioned against the base of the incoming dNTP in the growing base pair binding pocket. This position is replaced by an aspartate residue (Asp ^{276}) in Pol β, which makes Van der Waals interactions with the incoming dNTP and also restricts dNTP binding [6-8]. This high affinity of dNTPs has led to the suggestions that Pol λ may conduct DNA synthesis when the concentration of precursors (eg. dNTPs) in the cell is low; for instance, outside S-phase in cycling cells or in inactive cells [8].

It has also been determined that Pol λ has an unusually low fidelity and unique error specificity. It is the only human DNA polymerase studied thus far whose average single base deletion error rate exceeds its average single base substitution error rate [8]. It deletes single nucleotides at a rate much higher than that of Pol β and even higher than the notoriously inaccurate polymerases in the Y family [32]. In contrast, the single base substitution errors rates of Pol λ are only slightly higher than that of Pol β [8, 32].

Similar to Pol β, Pol λ also possesses a dRP lyase activity that can remove a 5' terminal sugar–phosphate (dRP) group and is linked to the N-terminal 8 kDa domain of the catalytic core [33]. This process of eliminating the abasic dRP group does not involve a divalent metal ion and is thought to proceed either by a hydrolysis reaction, as it occurs for most nucleases, or by means of an elimination reaction that proceeds via formation of a Schiff's base intermediate [33].

While the function of the proline-rich domain in Pol λ is not clear, it has been reported to suppress the polymerase activity of the enzyme *in vitro* [10].

1.4 The Scope of Thesis

Most structural studies performed thus far have been done on the C-terminus of DNA polymerase λ, which has been crystallized in its native form as well as different binary and ternary complexes. This has allowed the structure-function relationship between various subdomains of the catalytic core to be well documented. However, little is known about the structures of the BRCT domain and the proline-rich domain whose specific role remains vague.

The goal of this study is to obtain structural and functional information about the proline-rich domain and determine if there is any interaction between it and the core domain as well as establish what functional role this interaction might have. We believe that understanding the role of this domain may help us better realize the biological functioning of DNA polymerase lambda and also provide a firm basis for future studies on the domain-domain interactions between the protein's core domain and BRCT domain, as well as between the BRCT domain and the proline-rich domain.

1.5 Hypothesis

Based on the inhibitory effects of the human DNA polymerase λ proline-rich domain on the overall protein activity [10], we hypothesize that this region makes specific contacts with important residues within or near the active site of the protein and that these interactions play a greater role in regulating the protein activity during DNA repair. Furthermore, we also propose that the proline-rich domain suppresses the protein activity by one or two ways: either by limiting the protein's ability to bind DNA and/or limiting the proteins ability to incorporate deoxyribonucleotide triphosphates (dNTPs).

CHAPTER 2: MATERIALS AND METHODS

2.1.1 Bacterial strains and plasmids

The pET28b-fpolλ, pET24b-dpol λ and pET24b-tpolλ plasmids for the full length Polλ (fpolλ), truncated Polλ (dpolλ) and the core domain of Polλ (tpolλ) respectively, were provided by Dr. Z. Suo (Ohio State University) [4]. A list of all expressed proteins with the bacterial strains and the plasmids has been summarized in Table 1.

2.1.2 List of buffers and Media

A summary of the media used in protein expression and the buffers used in protein purification as well as protein preparation in this study have been listed in Table 2. The reaction buffer "L" contains 40 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 10 mM DTT, 0.25 mg/ml BSA and 2.5 % (v/v) glycerol. All the DNA extension and binding experiments reported here were carried out in buffer L.

2.2 Cloning of HIS_6-tdPol λ and His_6-proline-rich domain.

The pET28b-fpolλ, pET24b-dpolλ and pET24b-tpolλ plasmids contain the fragment carrying the full-length Pol λ (fpol λ), Pol λ lacking the BRCT domain (dpol λ) and the core domain of Pol λ (tpolλ) (Figure 4) genes respectively. The human genes encoding tdpol λ (dpol λ construct lacking the first 22 N-terminal residues of the proline-rich domain (Figure 4) and the proline-rich domain were PCR amplified using the pET28b-fpolλ plasmid and separately inserted into the *NcoI/XhoI* sites of the pHIS.parallel 1 vector [36] to construct pHis-tdpolλ and pHis-proline-rich plasmids. The forward and reverse primers are shown in Table 1 with the mismatched bases used to generate the insertion/restriction enzyme cleavage site underlined.

Table 1. Summary of the plasmids along with bacterial strains for the expressed and purified proteins used in this study

Protein	**Parental plasmid**	***E.coli* strain**	**Oligonucleotide primers (forward and reverse)**
fpolλ	pET28b	BL21(DE3)pLysS	
dpolλ	pET24b	BL21(DE3)pLysS	
tdpolλ	pHIS.parallel 1	BL21(DE3)pLysS	5'-GTGCCCTCGAGTCACCAGTCCCGCTCAGCAGGTT-3' 5'-CGCTTCCATGGTTCCTCCTG GCACCCATGAGG-3'
tpolλ	pET24b	BL21(DE3)pLysS	
Proline-rich domain	pHIS.parallel 1	BL21(DE3)	5'-GCACCGTTCCATGGGCATCTTCATCCCCAGTAGGT-3' 5'-GCGCTTCTCGAGTGGGCTGTGCACAGACCCACTT-3'

The underlined based represent the mismatched residues used to generate the restriction enzyme (*Nco*I and *Xho*I) cut sites.

Table 2. List of Media and Buffers used for this study

Medium / Buffer	Recipe
LB-Agar Medium	10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl, 15 g/L Agar, pH 7.5
LB Medium	10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl, pH 7.5
M9 Minimal Medium	6.8 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 10 μg/ml Thiamine, 10 μg/ml Biotin, 1 g/L NH_4Cl, 3 g/L Glucose, pH 7.4
His Binding Buffer	20 mM Tris-HCl, pH 7.5, 10% glycerol (v/v), 800mM NaCl, 5 mM imidazole and 0.1% β-mercaptoethanol.
His Wash Buffer	20 mM Tris-HCl, pH 7.5, 10% glycerol (v/v), 800 NaCl, 40 mM imidazole and 0.1% β-mercaptoethanol.
His Elution Buffer	20 mM Tris-HCl, pH 7.5, 10% glycerol (v/v), 800 NaCl, 300 mM imidazole and 0.1% β-mercaptoethanol.
Buffer A	20 mM Tris-HCl, pH 7.5, 10% glycerol (v/v), 1mM EDTA and 1 mM DTT.
Buffer B	20 mM Tris-HCl, pH 7.5, 1 M NaCl, 10% glycerol (v/v), 1mM EDTA and 1 mM DTT
Buffer C	20 mM Hepes, pH 7.5, 0.1 mM EDTA, 10% glycerol (v/v) and 1 mM DTT
Buffer D	20 mM Hepes, pH 7.5, 1M NaCl, 10% glycerol (v/v), 0.1 mM EDTA and 1 mM DTT
Protein sizing buffer	20 mM Tris-HCl, pH 7.5, 10% glycerol (v/v), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA and 1 mM DTT.
Circular Dichroism (CD) sizing buffer	10 mM Na_2HPO_4 pH 7.5, 150 mM NaF, 1 mM EDTA and 0.1% β-mercaptoethanol.
Nuclear Magnetic Resonance (NMR) sizing Buffer	50 mM K_2HPO_4 pH 7.0, 100mM NaCl, 0.1 mM EDTA and 1 mM DTT

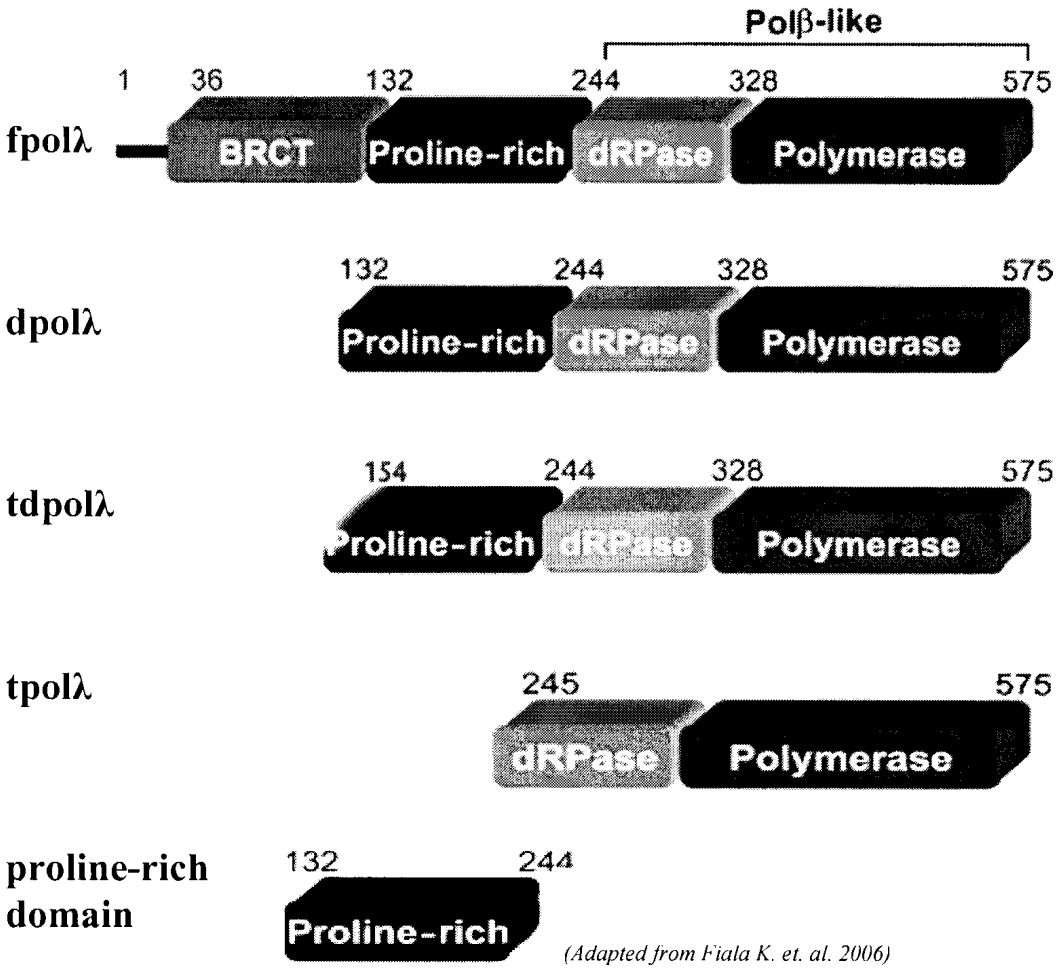


Figure 4. Schematic representation of the gene fragments carried by the various plasmids and the proteins of interest they express. pET28b-fpolλ plasmid carries the full-length Pol λ gene [residues 1-575] that expresses an N- and C-terminal His_6-tagged fpol λ protein. pET24b-dpolλ plasmid carries the gene of Pol λ lacking the BRCT domain (dpol λ) [residues 132-575] and expresses a C-terminal His_6-tagged dpol λ protein. pHis-tdpolλ plasmid carries the gene of the dpol λ construct lacking the first 22 N-terminal residues of the proline-rich domain (tdpolλ) [residues 154-575] and expresses an N-terminal His_6-tagged tdpolλ protein. pET24b-tpolλ plasmid contains the gene of the core domain of Pol λ (tpolλ) [residues 245-575] and expresses a C-terminal His_6-tagged tpolλ protein. The pHis-proline-rich plasmid carries the gene fragment of the Pol λ proline-rich domain [residues 132-244] and expresses an N-terminal His_6-tagged proline-rich domain protein fragment.

The cloning of the tdpolλ and proline-rich domain genes was performed by two-step site directed PCR mutagenesis as previously described with minor modifications [35]. In brief, the amplifications were carried out with 50 ng templates, 125 ng primer (Sigma), 200 μM dNTPs and 2U of *Pfu* Turbo DNA polymerase (Fermentas). An initial extension reaction for the forward and reverse primers was performed individually in the first step of the reaction with the extension products and reactants being combined for subsequent reactions. After completion of the PCR cycles, the reaction was allowed to cool down to room temperature after which, the resulting gene fragment was purified from a 1% agarose gel. The purified fragments were then digested using the restriction enzymes *Nco*I and *Xho*I, and then ligated into *Nco*I and *Xho*I sites of the pHIS.parallel 1 plasmid and transformed into DH5α competent cells and grown over night at 37°C in LB agar plate with the corresponding antibiotic (Ampicillin). The plasmids from the resulting colonies were harvested and confirmed by DNA sequencing. The plasmid containing the correct DNA sequence was subsequently transformed into *E. coli* cells for protein expression.

2.3 Expression and Purification of fPol λ, dPol λ, tdPol λ, tpol λ and the proline-rich domain.

2.3.1 Protein Expression

All plasmids were transformed into *E. coli* cells using standard methods [37]. The general protocol for the expression, harvesting and lysis of all proteins was similar unless stated otherwise.

Freshly transformed single colonies were grown in 50 ml LB supplemented with the appropriate antibiotic (Table 3), overnight at 37° C. 5 ml of the overnight starter

culture was then used to inoculate 1L LB supplemented with the corresponding antibiotic and grown at 37° C until the OD_{600} reached 0.7. The cell culture was subsequently induced with 0.4 mM IPTG to express the protein of interest, and incubated for 12 hours at 22°C.

The cells were harvested by centrifugation at 6000 rpm in a Beckman JLA-9.1000 rotor for 15 minutes and re-suspended in His Binding buffer (Table 2). Cells were then transferred to a 50 ml Falcon conical tube and centrifuged at 7000 rpm (4 500 g) in a Tomy MX301 AR500-04 rotor for 10 minutes after which the excess buffer was disposed and only the pellet was flash frozen and stored at -80°C for later purification.

2.3.2 Protein Purification

Since all proteins were His_6-tagged, they were first purified by affinity chromatography using a HiTrap Nickel column (Amersham Bioscience). Frozen cell pellets were thawed on ice and re-suspended in His Binding buffer containing protease inhibitors (2 mM PMSF, 1 mM benzamidine and 1 mM pepstatin A). The re-suspended cells were lysed by passing through a French pressure cell at 20,000 psi three times and the resulting cell lysate was cleared by ultracentrifugation at 38 000 rpm (100 000 g) in a Beckman Ti-45 rotor for 1 hour at 4 °C. The protein supernatant was then loaded onto a HiTrap nickel column pre-equilibrated with 5 column volumes of His Binding buffer after which the loaded column was washed extensively with His Wash buffer to remove any non-specifically bound proteins and contaminants. The His_6-tagged recombinant proteins were eluted using the His Elution buffer after which they were dialyzed against buffer A containing 10 mM NaCl.

Table 3. List of the cell lines (and plasmids the contain) used to express the proteins of interest used in this study

Cell line (plasmid)	**antibiotic resistance marker (s)**	**Concentration of antibiotic used**
BL21(DE3)pLysS (pET28b-fpolλ)	Kanamycin	40 μg/ml
	Chloramphenicol	50 μg/ml
BL21(DE3)pLysS (pET24b-dpolλ)	Kanamycin	40 μg/ml
	Chloramphenicol	50 μg/ml
BL21(DE3)pLysS (pHis-tdpolλ)	Ampicillin	100 μg/ml
	Chloramphenicol	50 μg/ml
BL21(DE3)pLysS (pET24b-tpolλ)	Kanamycin	40 μg/ml
	Chloramphenicol	50 μg/ml
BL21(DE3) (pHis-proline-rich domain)	Ampicillin	100 μg/ml

The dialyzed protein solutions of the fpol λ, dpolλ, tdpolλ and tpolλ were then subjected to another step of affinity chromatography using a HiTrap Heparin HP column (GE Healthcare). Each of the proteins was separately applied to a HiTrap Heparin HP column equilibrated with buffer A supplemented with 5 % buffer B and eluted with a linear salt gradient of 0.05-1 M NaCl using buffer B. The fractions containing the proteins of interest were pooled and subjected to a final step of purification that involved ion exchange chromatography. The fpol λ and tpol λ protein solutions were diluted with buffer C to a salt concentration of 50 mM after which they were applied to a HiTrap SP cation exchange column (GE Healthcare) equilibrated with buffer C supplemented with 5% buffer D, and eluted with a linear salt gradient of 0.05-1M NaCl using buffer D. The dialyzed proline-rich domain (from the nickel column), dpol λ and tdpol λ (from the heparin column) protein solutions were diluted with buffer A to a salt concentration of 50mM after which they were applied to a HiTrap Q anion exchange column (GE Healthcare) pre-equilibrated with buffer A supplemented with 5% buffer B, and eluted with a linear salt gradient of 0.05-1 M NaCl using buffer B. The final protein solutions were individually pooled and concentrated using centrifugal concentrators (Vivaspin) and stored at 4 °C or -20°C (in 50% glycerol) for later use.

Each step of the purification and the protein purity was monitored by Coomassie-Brilliant Blue stained SDS-PAGE.

2.4 Synthetic Oligonucleotides.

The DNA substrates listed in Table 3 were purchased from either Sigma or W. M. Keck Oligoncleotide Synthesis Facility (Yale University) and were purified by reverse phase chromatography (RPC) or polyacrylamide gel electrophoresis (PAGE).

Table 4. List of the oligonucleotides used in this study to carry out various reactions and the methods used to purify them.

Template (41mer) 5'– GGACGGCATTGGATCGACGATGAGTTGGTTGGACGGCTGCG – 3' Upstream primer (21mer) 5' – CGCAGCCGTCCAACCAACTCA - 3' Downstream primer(19mer) 5' – $^{PO4-}$CGTCGATCCAATGCCGTCC – 3' Annealed product (41/ 21-19) 3' - GCGTCGGCAGGTTGGTTGAGTAGCAGCTAGGTTACGGCAGG – 5' 5' – CGCAGCCGTCCAACCAACTCA- CGT CGATCCAATGCCGTCC – 3'	DNA extension and binding substrates	PAGE purified
Template (11mer) 5'-CGG CAA CGC AC-3' Upstream primer (5mer) 5'-GTGC G -3' Downstream primer (4mer) 5'-$^{Po4-}$GCC G-3' Annealed product (11/4-5) 5'-CGG CAACGC AC-3' 3'-GCC G GCG TG -5'	Crystallography substrate	RPC purified
Template (16mer) 5'-CCG AGC CGC GAT CAG C-3' Upstream primer (9mer) 5'-GCT GAT CGC-3' Downstream primer (5mer) 5'- $^{PO4-}$CTC GG-3' Annealed product 5'- CCG AGCC GCG ATC AGC-3' 3'- GGC T C CGC TAG TCG-5'	Crystallography substrate	RPC purified

In order to do DNA extension and DNA binding assays as well as grow ternary crystals, we had to design DNA substrates for these experiments. Based on previous studies, and the role of Pol λ in base excision repair, the DNA had to be gapped and possess a 5'-phosphate downstream from the gap, which is the preferred substrate for Pol λ [8].

The DNA substrate used in the DNA extension and binding studies was composed of a 41 nucleotide long template duplexed to a downstream 19 nucleotide long primer that was 5'-phosphorylated and an upstream 21 nucleotide long primer that was 5'-^{32}P radiolabeled. The upstream primer strand 21-mer was 5'-^{32}P radiolabeled by incubation with T4 polynucleotide kinase (NEBL) and [γ-^{32}P] ATP (ICN) for 1 h at 37 °C. The unreacted [γ-^{32}P] ATP was subsequently removed by centrifugation via a Bio-Spin-6 column (Bio-Rad). The 5'-^{32}P-labeled 21-mer primer was then annealed with the corresponding nonradiolabeled downstream primer 19-mer to the 41-mer template at a molar ratio of 1.0:1.15:1.25, respectively, to form the 21-19/41-mer single-nucleotide (adenine) gapped substrate. The DNA mixture to be annealed was first denatured at 95 °C for 8 min and then cooled slowly to room temperature over several hours.

The DNA oligonucleotides for crystallization of the ternary complex (composed of protein, DNA and incoming nucleotide) were first purified using reverse phase chromatography and then the corresponding template-primer mixtures (Table 3) were annealed in a 1:1:1 ratio to form the 4-5/11-mer and 5-9/16-mer substrates with a two-nucleotide gap. Mixtures to be annealed were first denatured at 60 °C for 10 min and then cooled slowly to 4 °C over several hours and stored at -20°C. Due to the low melting

temperatures of these oligonucleotides, all the reactions where they were involved were carried out on ice.

2.5 DNA Extension and Binding Assays.

To assay the ability of the different protein constructs (fpol λ, dpol λ, tdpol λ and tpol λ) to incorporate different nucleotides, and their ability to bind DNA, we carried out DNA extension and binding studies. All the assays described here were carried out in buffer L (previously mentioned), and used the 21-19/41-mer duplex DNA as a substrate. All concentrations reported in these experiments refer to the concentrations of the components after mixing. The amount of reacted substrate in the DNA extension assays (protein activity and protein fidelity assays) was quantified and plotted as a ratio of total substrate. This ratio was calculated using the following equation:

$$\textit{Fraction of reacted substrate} = \frac{[S]_T - [S]}{[S]_T} \qquad \textit{Equation 1}$$

Where $[S]_T$ is the total substrate (reacted substrate + unreacted substrate, or substrate at the zero time point [t=0]) and [S] is the amount of unreacted substrate measured after a given amount of time when the reaction is stopped (Time=t). The time course of product formation was fitted to a single exponential equation using GraphPad Prism for Windows (GraphPad Software version 5.00, San Diego California USA, www.graphpad.com). All the reactions were carried out in triplicates so as to adjust for any errors in substrate measurements resulting from pipetting and gel loading inaccuracies.

2.5.1 Protein Activity assay.

To examine the activity of the different protein constructs, we performed DNA replication experiments. The replication complexes were formed by pre-incubating the

21-19/41-mer single-nucleotide (dA) gapped DNA substrate (10 nM) with 1 nM of protein for 10 minutes at room temperature. The replication reaction was initiated by adding 100 μM of the correct dNTP (dTTP) at 37°C, after which the reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of stop solution containing 98% formamide, 20 mM EDTA, 0.3% bromophenol blue, and 30% cyanol blue. The reaction products were resolved on a 10% polyacrylamide gel containing 7 M urea and visualized using a Molecular Dynamics Storm PhosphoImager and quantified using ImageQuant software (ImageQuant Version 5.2, Molecular Dynamics, www.mdyn.com).

2.5.2 Protein Fidelity assay.

To analyze how the different protein constructs dealt with mismatch insertions, we carried out replication assays (previously mentioned) with a few adjustments. In these experiments, we used 25 nM of protein and initiated replication by adding 100 μM of the wrong dNTP (dATP or dCTP or dGTP) individually. All other procedures remained the same.

2.5.3 Trans-replication assay.

To investigate if there was any interaction between the proline-rich domain and tpol λ and whether this interaction had any effect on the protein activity, we performed a trans-replication assay. This experiment was similar to the replication assay (section 2.5.1) with a few adjustments, where we used a 1:1 molar ratio (2 nM + 2 nM) of proline-rich domain to core domain to a final concentration of 1 nM. All other procedures remained the same. A replication (activity) assay of the core domain alone and the 1:1 mixture was then carried out and compared for any differences/similarities.

2.5.4 DNA binding assay.

One way that the proline-rich domain might influence the protein activity may be by regulating or manipulating DNA binding, and to examine if this is the case, we carried out DNA binding (Gel shift) assays. The DNA binding ability of the different protein constructs was assayed by incubating varying concentrations of protein (1, 5 and 10 μm) with excess DNA substrate (50 μM) for 30 minutes at room temperature. The experimental products were then resolved on a 10% native gel and visualized using a Molecular Dynamics Storm PhosphoImager.

2.6 Circular Dichroism (CD) Spectropolarimetry.

CD experiments were performed to characterize the structural properties of the proline-rich domain and also to investigate if there was any interaction with the core domain. The CD experiment was performed on a Jasco J-810 spectrophotometer (Easton, MD) equipped with a PTC-423S Peltier temperature control unit, using a cuvette with a path length of 0.1 mm. All the spectra were acquired at 24 °C as the average of three scans in the far-UV region from 180 to 260 nm at a scan speed of 20 nm/min and a response time of 1 second. The spectrum of the sample buffer was subtracted from the raw data of protein samples using Jasco supplied software to get the raw unnormalized ellipticity (θ) in millidegrees. All the proteins used were run through a Superdex S200 10/300 GL (GE Healthcare) gel-filtration column using the Circular Dichroism (CD) sizing buffer prior to performing the CD experiments to ensure that all the samples had identical buffer composition/components. The CD sizing buffer was also used as the sample buffer (blank).

To characterize the structure of purified proline-rich domain, CD spectra were collected using 50 nM of individual Proline-rich domain in CD sizing buffer.

To examine if there was any interaction between the Proline-rich domain and the core domain, CD spectra of the individual core domain (50 nM), individual proline-rich domain (50 nM), individual dpol λ protein (50 nM) and a 1:1 molar ratio of proline-rich to core domain (50 nM final concentration) in CD sizing buffer was acquired. A hypothetical spectrum simulating a case where the proline-rich and core domains did not interact was also generated by summing up the individual proline-rich and core domain spectra. All the spectra were then plotted together to identify if there were any changes.

2.7 His Pull-down assays

To further determine if there was interactions between the core domain and the proline-rich domain, we carried out his pull-down assays using nickel charged beads (Sigma). The buffers used had salt concentrations of 150 mM NaCl, similar to the sizing buffer. The core domain was used as the His_6-tagged protein while the proline-rich domain was used as the binding partner. Prior to carrying out the pull-down assay, the proline-rich domain was digested using Tobacco etch virus (TEV) protease [78] to remove the His_6-tag using previously described methods [78-80].

The experiment was carried out by incubating the core domain and 3 fold excess proline-rich domain together for approximately 1 hr at room temperature, followed by incubating the mixture with the nickel charged beads that had been equilibrated with the His Binding buffer (containing only 150 mM NaCl) for another 1 hr while rocking the whole mixture gently. After the incubation, the beads were washed thoroughly with His

Wash buffer (containing only 150 mM NaCl) to remove any unbound/excess proteins. Finally, the nickel bound His_6-fusion protein with the interaction partner was eluted using the His Elution buffer and the proteins were detected using SDS gel analysis along with Coomassie Brilliant Blue staining.

2.8 Nuclear Magnetic Resonance (NMR) Spectroscopy

2.8.1 Overexpression of Human ^{15}N-labelled Proline-rich domain.

A single colony of cells containing the pHis-proline plasmid (previously described) was used to inoculate 1 ml of LB media containing 100 μg/ml Ampicillin. The culture was allowed to grow at 37 °C for 8 hours, and then transferred into 200 ml of M9 minimal media supplemented with 5 ml LB media and 100 μg/ml Ampicillin. After an overnight incubation at 37 °C, the bacterial cells were pelleted by centrifugation (3,000 × g, 15 min, Tomy MX301 AR500-04 rotor) and were resuspended into 1 L of M9 minimal media containing 50 μg/ml Ampicillin to obtain a starting OD_{600} of 0.10 – 0.15. For the expression of ^{15}N-labelled proline-rich domain protein, 1 g of $^{15}NH_4Cl$ (Cambridge Isotope Laboratories) was added to the 1-litre M9 media as the only source of nitrogen. The cell culture was incubated at 37 °C until the OD_{600} value reached 0.7 – 0.8 at which point, protein overexpression was induced with 0.4 mM isopropyl-beta-D-thiogalactopyranoside (IPTG, BioShop) at 22 °C and incubated overnight (12 hrs). The cells were harvested by centrifugation at 6000 rpm for 15 min in a Beckman JLA-9.1000 rotor. The cell pellets were washed with His Binding buffer and collected by centrifugation at 5000 rpm for 15 min in a Tomy AR500-04 rotor, followed by storage at -80 °C until needed.

2.8.2 Purification of Human ^{15}N-labelled Proline-rich domain.

The purification of the ^{15}N-labelled Proline-rich domain was similar to that of the unlabelled protein (section 2.3.2). The His_6-tag was cleaved off as previously mentioned (section 2.7)

2.8.3 Two-Dimensional ^{1}H-^{15}N HSQC Experiments

The Heteronuclear Single Quantum Coherence (HSQC) NMR experiment described below was performed at 25 °C on a Varian INOVA 600 MHz spectrometer equipped with a xyz-gradient triple resonance probe. Each NMR sample contained 10% (v/v) D_2O (Cambridge Isotope Laboratories) and 1 mM sodium 2,2-Dimethyl-2-silapentane-5-sulfonate (DSS, Sigma) was added as an internal standard for chemical shift referencing [39]. All data sets were processed using NMRPipe [39] and the spectra were analyzed with NMRView [40].

To study the interaction of the core domain of DNA polymerase λ (tpol λ) with the ^{15}N-labeled proline-rich domain, we carried out an NMR addition experiment and observed the changes in ^{1}H and ^{15}N chemical shifts of the proline-rich domain. The addition experiment involved collecting a series of ^{1}H-^{15}N HSQC spectra of 70 nM ^{15}N-labelled proline-rich domain protein (in NMR sizing buffer – see Table 2) in the presence of different concentrations of core domain (tpol λ) protein (0, 70 and 140 nM) with the following molar ratios of proline-rich domain to tpol λ: 1:0, 1:1 and 1:2. The spectra were then overlaid to identify regions of significant change in chemical shift.

2.9 Protein Crystallization screening

Using x-ray crystallography, we hoped to obtain a 3-dimensional representation of the DNA polymerase lambda with the BRCT and proline-rich domains, which would help us better understand how the proline-rich domain influences the protein activity. Using freshly purified protein that had been subjected to size exclusion chromatography, we set up the reaction mixture that was composed of 6 to 10 mg/ml of protein (fpol λ, dpol λ, tdpol λ or tpol λ), 1:1.2 molar ratio of protein to DNA, 1 mM of dideoxynucleotide triphosphate (ddNTP) and 10 mM of magnesium chloride ($MgCl_2$). Due to the short length and low melting temperature of the template and primer oligonucleotides (Table 3), we carried out the experiment at a low temperature (4°C) to ensure that the system remained homogenous and that the duplex DNA remained annealed.

CHAPTER 3: RESULTS

3.1 Expression and purification of proteins

3.1.1 Expression and purification of fpol λ, dpol λ, tdpo lλ and tpol λ.

The different constructs of the human DNA polymerase lambda (fpol λ, dpol λ, and tpol λ) were provided by Dr. Suo, expressed in *E.coli* (Table 1) and purified to homogeneity (Figure 5) through a series of affinity and ion exchange chromatography steps as described in the methods section. However, the dpol λ construct appeared to co-purify with a second species of protein that we were unable to separate (Figure 5, lane 3)

Mass spectrometry analysis combined with trypsin digestion was performed on the purified dpol λ sample to identify the extra species that co-purified with it. The results revealed that the second species was a similar protein to dpol λ but differed from the original construct by a mass of approximately 2121 Da (Figure 6), which corresponds to the loss of the first 18 N-terminal residues.

Based on the mass spectrometry results, we created a deletion mutant that lacked the first 22 N-terminal residues of the original dpol λ protein to deal with the problem of having a non-homogeneous sample. After the PCR reaction, the mutant plasmid (pHis-tdpolλ) was sequenced and the deleted residues confirmed. The protein construct was referred to as tdpol λ and was overexpressed in *E. coli* cells (Table 1) and purified as a single species (Figure 5, lane 4) and in a soluble form by a series of affinity and ion exchange chromatography steps, similar to the dpol λ purification procedure.

All the protein samples were purified from 1 liter cultures and the final yield in each case was: 10 mg (fpol λ), 8 mg (dpol λ), 15 mg (tdpol λ), 15 mg (tpol λ) and 10 mg (proline-rich domain)

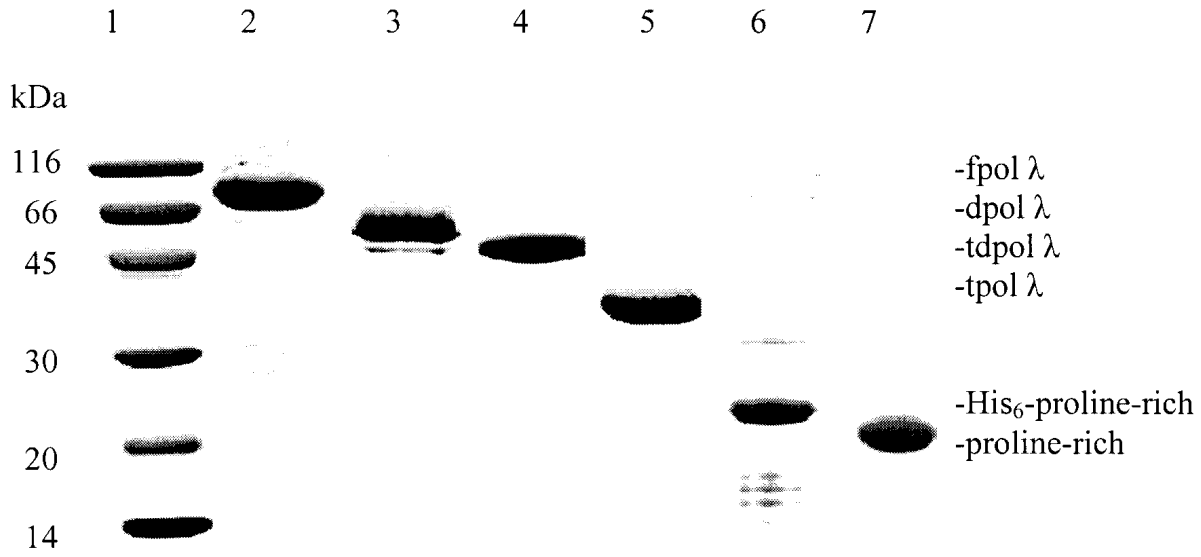


Figure 5. SDS-PAGE of the purified proteins of interest.
The protein samples from the final step of purification were analyzed by SDS-PAGE using a 20% polyacrylamide gel and stained with Coomassie Brilliant Blue R-250 dye. Lane 1 represents the low molecular weight marker with the corresponding sizes listed on the left. Lane 2 contains fpol λ sample (65.6 kDa) from the HiTrap SP column. Lane 3 contains dpol λ sample (50.1 kDa) from the HiTrap Q column. Lane 4 contains tdpol λ sample (46.7 kDa) from the HiTrap Q column. Lane 5 contains tpol λ sample (38.2 kDa) from thc HiTrap SP column. Lane 6 contains the proline-rich domain with the His_6-tag (15.2 kDa) from the HiTrap Q column. Lane 7 contains the proline-rich domain without the His_6-tag (12.3 kDa) from the HiTrap Q column.

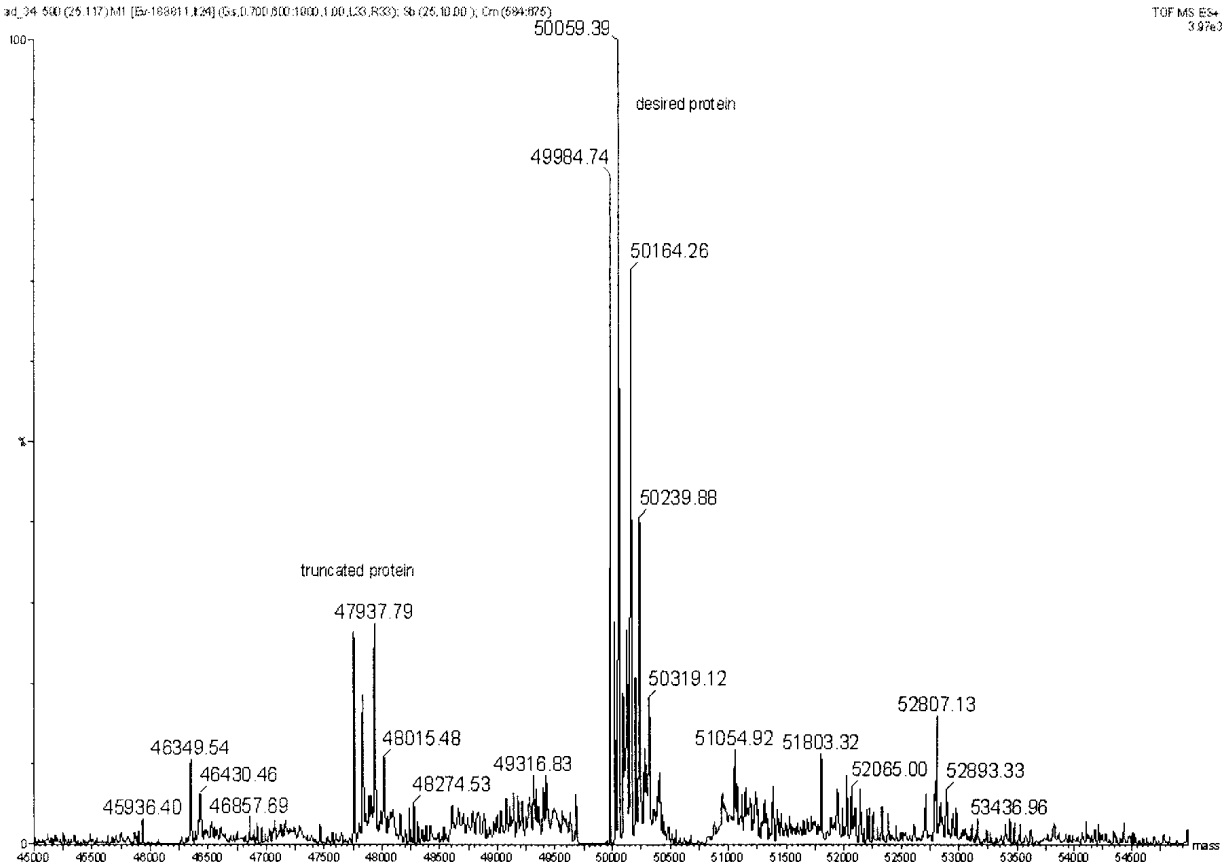


Figure 6. Mass spectrometry analysis of the purified dpol λ protein.
The final purified sample of the dpol λ protein (theoretical protein mass is 50.11 kDa) was analyzed by SDS-PAGE using a 10% polyacrylamide gel and stained with Coomassie Brilliant Blue R-250 dye to identify the different species in the sample. The different bands on the gel were then cutout and subjected to trypsin digestion, after which the digests were analyzed by mass spectrometry. The analysis revealed a second species with a molecular weight of 47.93 kDa; a difference of approximately 2121 Da between the two major proteins/species identified. This difference represents the loss of first 18 N-terminal residues from the original construct (dpol λ) and using this information, the tdpol λ protein construct that lacks the first 22 N-terminal amino acids of dpol λ (previously described) was generated.

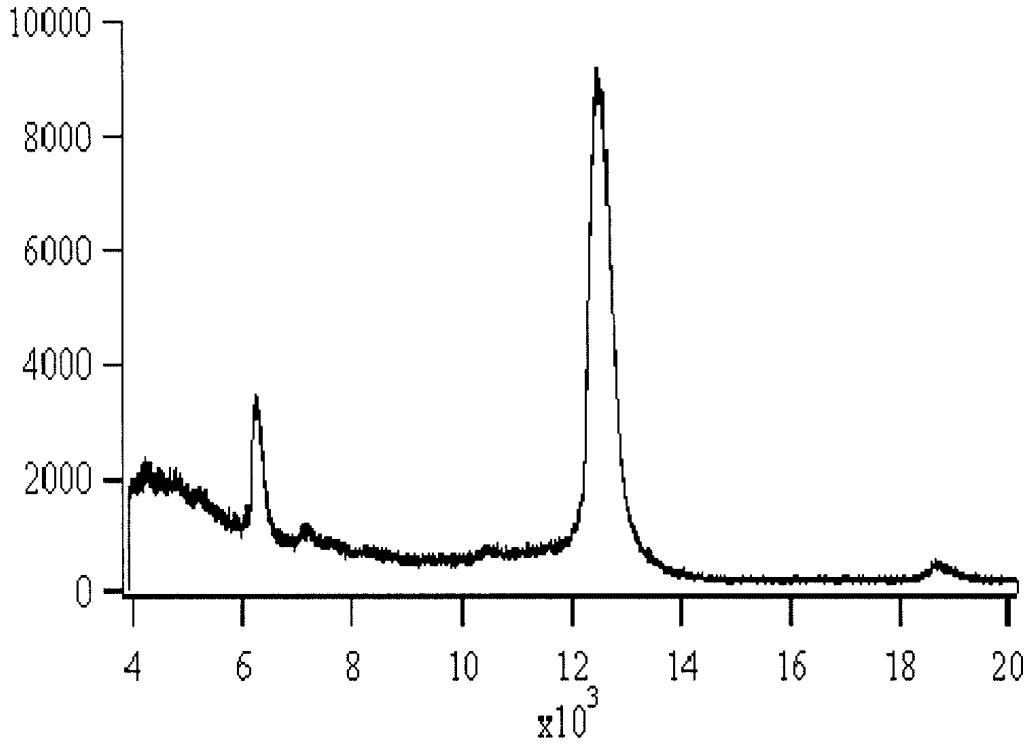


Figure 7. Mass spectrometry analysis (MALDI) of the purified proline-rich domain. The proline-rich domain was purified using a series of affinity and ion (anion) exchange chromatographies. The first step of purification involved using a HiTrap Nickel affinity column to purify the His_6-fusion proline-rich domain, after which it was further purified using a HiTrap Q anion exchange column. The His_6-tag of the fusion protein was then cleaved off using TEV protease, and using a HiTrap Nickel affinity column we were able to separate the protease and his_6-tag from the actual protein of interest. The final protein sample was subjected to mass spectrometry, which confirmed the theoretical mass (12.3 kDa) of the protein and purity of the sample.

3.1.2 Expression and purification of the proline-rich domain protein fragment.

To determine the interactions and effects of the proline-rich domain, we cloned the individual domain from the full length (fpol λ) protein as previously described. The plasmid/PCR product (pHis-proline-rich) sequence was confirmed by DNA sequencing and was found to be correct. The protein was overexpressed in *E. coli* cells (Table 1) and purified (Figure 5, lane 6) in a soluble form by a series of affinity and ion exchange chromatography steps. The His_6-tag of the protein was cleaved off using TEV protease and the mass of the final product confirmed by mass spectrometry (MALDI) (Figure 7) before concentrating the protein and storing it.

3.2 DNA Extension Assays.

3.2.1 Protein Activity assay

Once all the 4 different constructs (fpol λ, dpol λ, tdpol λ and tpol λ) had been purified to homogeneity, we carried out replication assays to test and determine their activity (section 2.5.1) using the DNA substrate 21-19/41 mer and correct incoming nucleotide, which was a 2'-deoxythymidine 5'-triphosphate (dTTP)

The reaction products were resolved using a 10% polyacrylamide gels containing 7M urea and visualized using a Molecular Dynamics Storm PhosphoImager. The series of gels (Figure 8) illustrates the separation of the substrate (21-mer) and product (22-mer) after nucleotide incorporation of the different constructs at various time points. Based on qualitative analysis, we observed that all our constructs were active; as indicated by the presence of a 22 nucleotide long product on the sequencing gel.

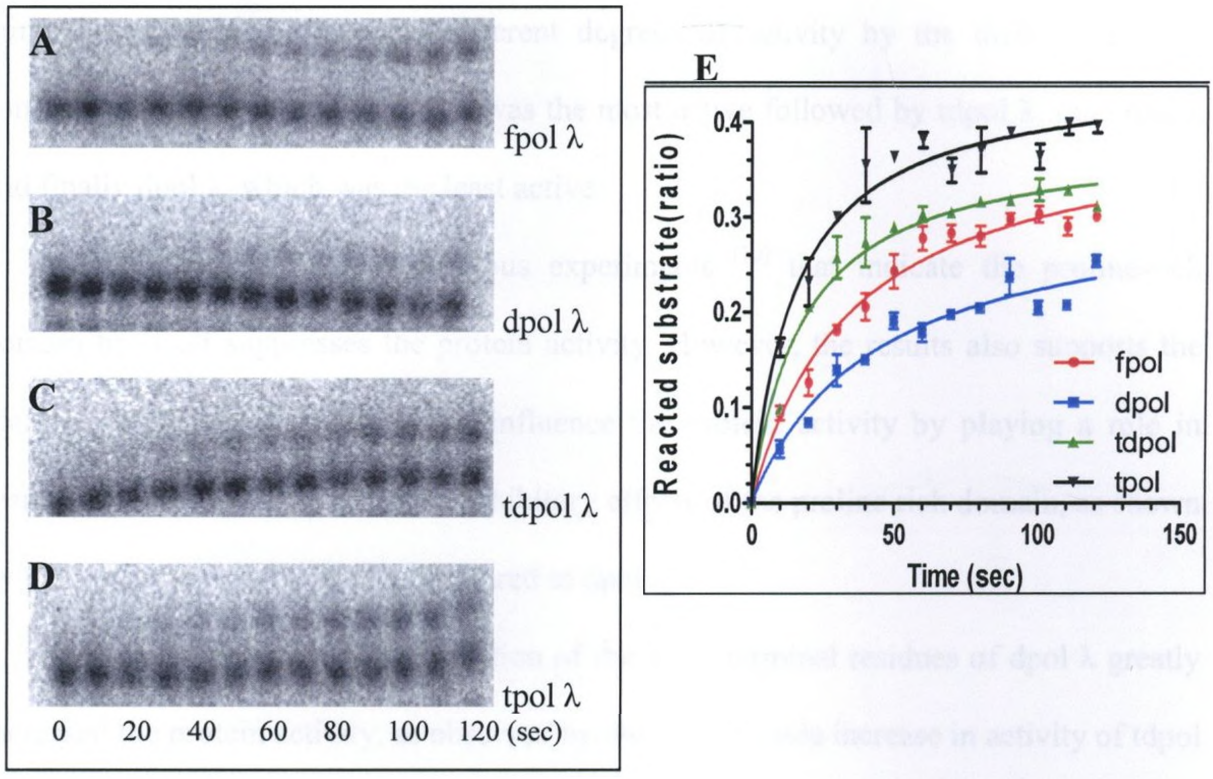


Figure 8. Protein activity assay.
Series of gels illustrating the progression of product formation as a function of time under conditions of 0.1 mM dTTP, 1 nM of protein and 10 nM of 21-19/41-mer single-nucleotide (adenine) gapped substrate. The replication reactions were carried out by first pre-incubating the protein and DNA substrate for 10 minutes at room temperature and starting the reaction by adding dTTP and placing the whole mixture in 37°C, thus creating an Adenine to Thymine (A:T) match. The reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of Stop solution, after which the products were resolved on a 10% polyacrylamide gel containing 7M urea and visualized as well as quantified using a Molecular Dynamics Storm PhosphoImager.
The unreacted substrate (21-mer) is shown at the bottom of each gel picture with the extended product (22-mer) located sequentially above the corresponding unreacted substrate. The reaction time intervals (seconds) are shown at the bottom of the panel. **(A)** Incorporation reaction by fpol λ, **(B)** Incorporation reaction by dpol λ, **(C)** Incorporation reaction by tdpol λ, **(D)** Incorporation reaction by tpol λ. **(E)** The fraction of reacted substrate in each reaction was quantified using a Molecular Dynamics Storm PhosphoImager and calculated using *equation 1* and plotted as a function of time. fpol λ (___), dpol λ (___), tdpol λ (___), tpol λ (___).

Further quantification (Figure 8E) of the fraction of substrate reacted within the 2 minute reaction time, revealed different degrees of activity by the different protein constructs. We observed that tpol λ was the most active followed by tdpol λ, then fpol λ and finally dpol λ, which was the least active.

These results support previous experiments [10] that indicate the proline-rich domain by itself suppresses the protein activity. However, the results also supports the idea that the BRCT domain could influence the protein activity by playing a role in disrupting or compensating for the inhibitory effects of the proline rich domain, as shown by the higher activity of fpol λ compared to dpol λ.

We also noticed that the deletion of the 22 N-terminal residues of dpol λ greatly increased the protein activity, as observed by the considerable increase in activity of tdpol λ compared to the dpol λ. This would suggest that a number of residues located in this region do play some significant role in regulating the proline-rich domain's ability to suppress protein activity.

3.2.2 Protein Fidelity assay.

To analyze the ability of the human DNA polymerase Lambda to discriminate between correct and incorrect incorporation of dNTPs, I carried out a fidelity test (previously described) to examine how the different constructs replicated past the gapped DNA substrate when supplied with the 3 wrong nucleotides (dATP, dCTP or dGTP) individually. The reaction products were resolved and visualized in a manner similar to the activity assay. The series of gels (Figure 9-11) illustrates the separation of the substrate and product after nucleotide incorporation of the different constructs at various time points for the different nucleotide incorporations.

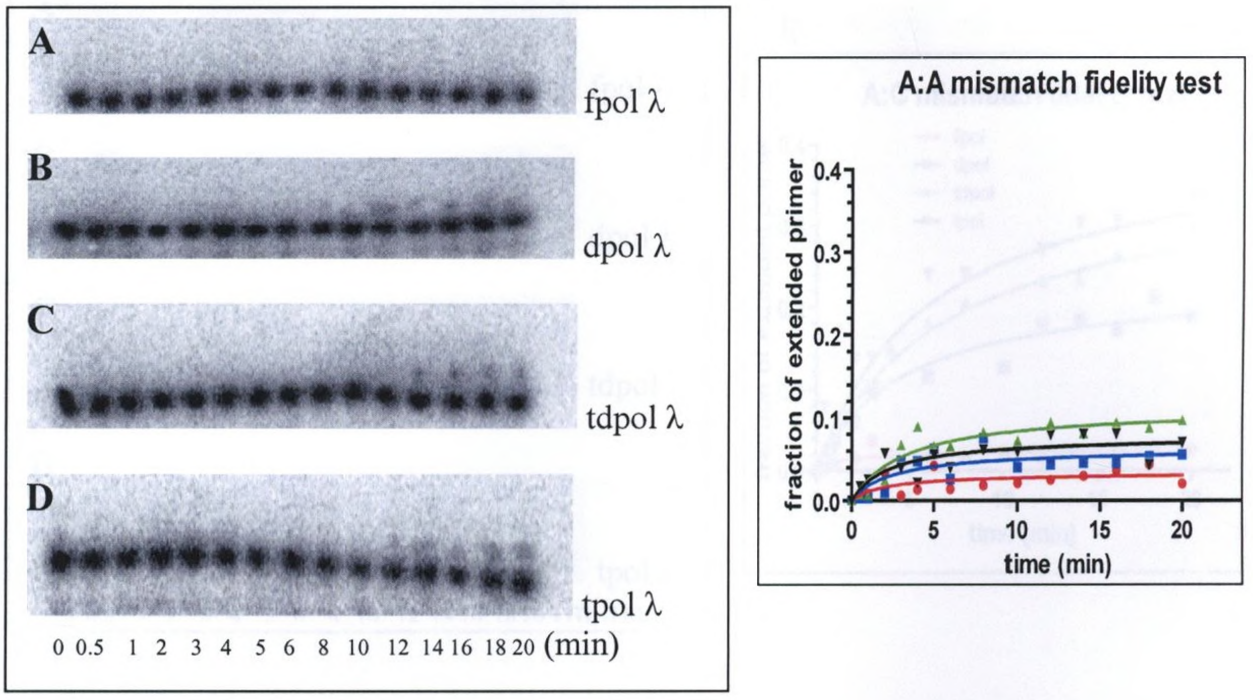


Figure 9. Fidelity assay

Series of gels showing the progression of product formation as a function of time under conditions of 0.1 mM dATP, 25 nM of protein and 10 nM of 21-19/41-mer single-nucleotide (adenine) gapped substrate. The replication reactions (fidelity assay) were carried out by first pre-incubating the protein and DNA substrate for 10 minutes at room temperature and starting the reaction by adding dATP as the incoming nucleotide and placing the whole mixture in 37°C, thus creating an Adenine to Adenine (A:A) mismatch. The reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of Stop solution, after which the products were resolved on a 10% polyacrylamide gel containing 7M urea and visualized as well as quantified using a Molecular Dynamics Storm PhosphoImager.

The unreacted substrate (21-mer) is shown at the bottom of each gel picture with the extended product (22-mer) located sequentially above the corresponding unreacted substrate. The reaction time intervals (minutes) are shown at the bottom of the panel. **(A)** Incorporation reaction by fpol λ, **(B)** Incorporation reaction by dpol λ, **(C)** Incorporation reaction by tdpol λ, **(D)** Incorporation reaction by tpol λ. **(E)** The fraction of reacted substrate in each reaction was quantified using a Molecular Dynamics Storm PhosphoImager and calculated using *equation 1* and plotted as a function of time. fpol λ (___), dpol λ (___), tdpol λ (___), tpol λ (___).

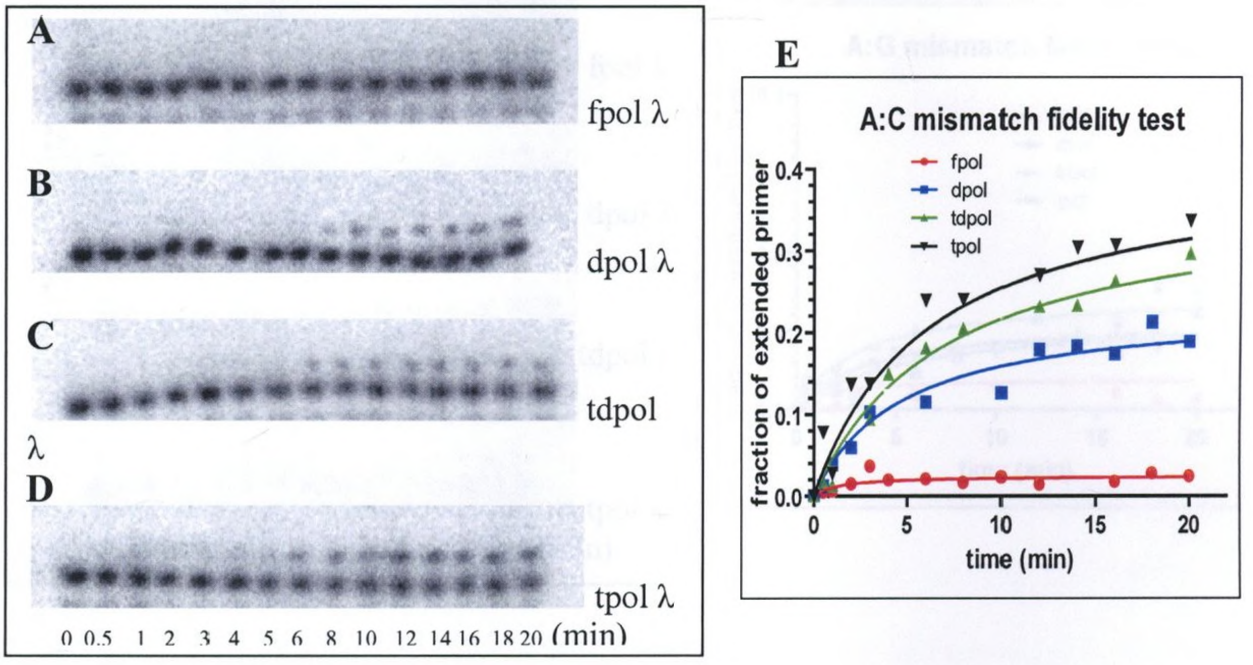


Figure 10. Fidelity assay
Series of gels showing the progression of product formation as a function of time under conditions of 0.1 mM dCTP, 25 nM of protein and 10 nM of 21-19/41-mer single-nucleotide (adenine) gapped substrate. The replication reactions (fidelity assay) were carried out by first pre-incubating the protein and DNA substrate for 10 minutes at room temperature and starting the reaction by adding dCTP as the incoming nucleotide and placing the whole mixture in 37°C, thus creating an Adenine to Cytosine (A:C) mismatch. The reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of Stop solution, after which the products were resolved on a 10% polyacrylamide gel containing 7M urea and visualized as well as quantified using a Molecular Dynamics Storm PhosphoImager.
The unreacted substrate (21-mer) is shown at the bottom of each gel picture with the extended product (22-mer) located sequentially above the corresponding unreacted substrate. The reaction time intervals (minutes) are shown at the bottom of the panel. **(A)** Incorporation reaction by fpol λ, **(B)** Incorporation reaction by dpol λ, **(C)** Incorporation reaction by tdpol λ, **(D)** Incorporation reaction by tpol λ. **(E)** The fraction of reacted substrate in each reaction was quantified using a Molecular Dynamics Storm PhosphoImager and calculated using *equation 1* and plotted as a function of time. fpol λ (___), dpol λ (___), tdpol λ (___), tpol λ (___).

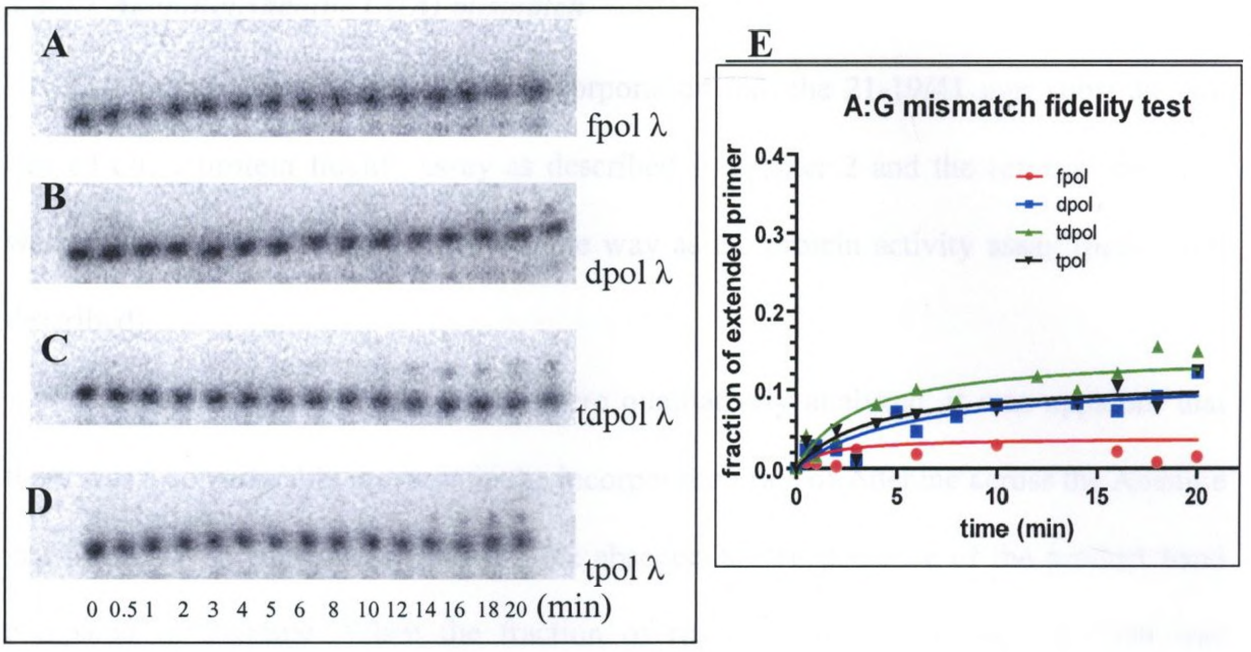


Figure 11. Fidelity assay

Series of gels showing the progression of product formation as a function of time under conditions of 0.1 mM dGTP, 25 nM of protein and 10 nM of 21-19/41-mer single-nucleotide (adenine) gapped substrate. The replication reactions (fidelity assay) were carried out by first pre-incubating the protein and DNA substrate for 10 minutes at room temperature and starting the reaction by adding dGTP as the incoming nucleotide and placing the whole mixture in 37°C, thus creating an Adenine to Guanine (A:G) mismatch. The reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of Stop solution, after which the products were resolved on a 10% polyacrylamide gel containing 7M urea and visualized as well as quantified using a Molecular Dynamics Storm PhosphoImager.

The unreacted substrate (21-mer) is shown at the bottom of each gel picture with the extended product (22-mer) located sequentially above the corresponding unreacted substrate. The reaction time intervals (minutes) are shown at the bottom of the panel. **(A)** Incorporation reaction by fpol λ, **(B)** Incorporation reaction by dpol λ, **(C)** Incorporation reaction by tdpol λ, **(D)** Incorporation reaction by tpol λ. **(E)** The fraction of reacted substrate in each reaction was quantified using a Molecular Dynamics Storm PhosphoImager and calculated using *equation 1* and plotted as a function of time. fpol λ ( ), dpol λ ( ), tdpol λ ( ), tpol λ ( ).

3.2.2.1 Adenine:Adenine (A:A) mismatch

To study the incorrect dATP incorporation into the 21-19/41 mer substrate, we carried out a protein fidelity assay as described in chapter 2 and the reaction products were visualized and analyzed in the same way as the protein activity assay (previously described).

When the gels (Figure 9A-D) were qualitatively analyzed, it was apparent that there was a considerable decrease in the incorporation rate of Adenine across the Adenine gap in the DNA substrate as seen by the absence/limited presence of the product band compared to thymine. When the fraction of reacted substrate in each reaction was quantitated using a Molecular Dynamics Storm PhosphoImager and plotted as a function of time (Figure 9E), it was shown that generally less than 10% of the substrate had reacted.

The 4 different protein constructs (fpol λ, dpol λ, tdpol λ and tpol λ) did not demonstrate significant differences in dATP incorporation, although tdpol λ did show a slightly higher misincorporation than other protein constructs.

3.2.2.2 Adenine:Cytosine (A:C) mismatch

To study the incorrect dCTP incorporation into 21-19/41 mer substrate, we carried out a protein fidelity assay as described in section 2.5.2 and the reaction products were visualized and analyzed in the same way as the protein activity assay.

Unlike the A:A mismatch, the qualitative analysis of the gels (Figure 10B-D) showed a substantial increase in the incorporation rate of cytosine across the Adenine gap in the substrate DNA as observed by the presence to the 22-mer product, except for the fpol λ reaction (Figure 10A), which showed a much lower misincorporation.

Other than the overall reduced incorporation rate (Min. vs. Sec. time scale), the gels (Figure 10B-D) appeared almost similar to the thymine incorporation gels (Figure 8B-D). Furthermore, when the fraction of reacted substrate in each reaction was quantified and plotted as a function of time (Figure 10E), it was shown that approximately between 20% and 30% of the substrate had reacted for dpol λ, tdpol λ and tpol λ, while fpol λ had less than 5% reaction.

The 4 different protein constructs (fpol λ, dpol λ, tdpol λ and tpol λ) demonstrated much more differences in dCTP incorporation compared to dATP incorporation. The fpol λ protein showed the lowest misincorporation where less than 5% of the substrate had reacted. The tpol λ construct showed the highest misincorporation where it reacted about 30% of the substrate followed by tdpol λ that reacted about 27% of the substrate and then dpol λ, which reacted slightly less than 20% of the substrate.

3.2.2.3 Adenine:Guanine (A:G) mismatch

To investigate the incorrect dGTP incorporation into 21-19/41 mer substrate, we carried out another protein fidelity assay, with the reaction products being visualized and analyzed in the same way as the protein activity assay.

By visualizing the reaction gels (Figure 11A-D), it was noticeable that there was a considerable decrease in the incorporation rate of guanine that was quite similar to the A:A mismatch reactions. Quantification of the fraction of reacted substrate, that was plotted as a function of time (Figure 11E), revealed that less than 10% of the substrate had reacted in most cases.

The 4 different protein constructs (fpol λ, dpol λ, tdpol λ and tpol λ) did not demonstrate major differences in dGTP incorporation, although tdpol λ did show the highest misincorporation with slightly higher that 10% of substrate reacted while fpol λ showed the lowest misincorporation with less than 5% of reacted substrate.

When it comes to incorporating the wrong nucleotide across a gapped DNA substrate, DNA polymerase lambda has been shown to have a 10^3-10^4 slower rate of misincorporation compared to the rate of correct nucleotide incorporation [4]. This reduced nucleotide misincorporation rate was also observed in this study, where we observed slower incorporation of the adenine, cytosine and guanine across the adenine gapped DNA substrate, compared to the thymine incorporation. It took more than 10 times longer (20 min. vs. 2 min.) and 25 times (25 nM vs. 1 nM) more enzyme for the fidelity assay to be carried out compared to the protein activity assay. From the series of figures shown (Figure 9-11), there was a higher preference for cytosine misincorporation, followed by guanine and finally Adenine which was the least preferred nucleotide. For all the protein constructs (fpol λ, dpol λ, tdpol λ, tpol λ), the nucleotide incorporation efficiency for adenine and guanine observed, was much lower than that of the Cytosine. This would indicate a preference for similar nucleotide (purine or pyrimidine) substitutions when it comes to mismatch incorporations (A↔G or C↔T substitutions). In this fidelity experiment, the fpol λ protein demonstrated the lowest misincorporation (highest fidelity), followed by dpol λ, and then tpol λ (although it had the lowest fidelity when incorporating cytosine) and finally tdpol λ which had the highest misincorporations. Earlier studies [4] have shown fpol λ to have 10-100 fold higher

fidelity than tpol λ and only 2 fold higher than dpol λ. Our experimental results indicated that the proline-rich domain alone did not significantly increase the protein fidelity; as seen by the high misincorporation of dpol λ and tdpol λ. However, the presence of proline-rich domain together with the BRCT domain showed a much greater increase in the protein fidelity; as observed by the low nucleotide misincorporation of fpol λ. This would imply that the significant increase in fidelity from the tpol λ construct to the fpol λ protein is due to the presence of both the BRCT domain and the proline-rich domain. The deletion of the first 22 N-terminal amino acid of dpol λ was observed to have a negative effect when it came to the proteins ability to carry out misincorporation. We saw a marked increase in the incorporation of the wrong nucleotide across the 21-19/41 mer gapped DNA substrate in tdpol λ, compared to the original dpol λ protein. This pointed towards the truncated region possessing some residues that are responsible in increasing the tpol λ protein fidelity. This deleted region of the proline-rich domain has not been studied before, however, it appears to be involved in fidelity regulation of DNA polymerase lambda.

3.2.3 Trans-replication assay.

The effects of the proline-rich domain could be as a result of domain-domain interaction with the core domain. To confirm if the proline-rich domain interacted with the core domain, and whether the interaction was responsible for any changes in protein activity, we carried out a trans-replication assay using the individual domains (as described in the methods section). The replication reaction of the core domain alone and the reaction of a 1:1 mixture of proline-rich domain to core domain, were run separately, analyzed and compared for any differences (Figure 12). Changes in the activity of tpol λ in the 1:1 mixture would be indicative of interaction with the proline-rich domain.

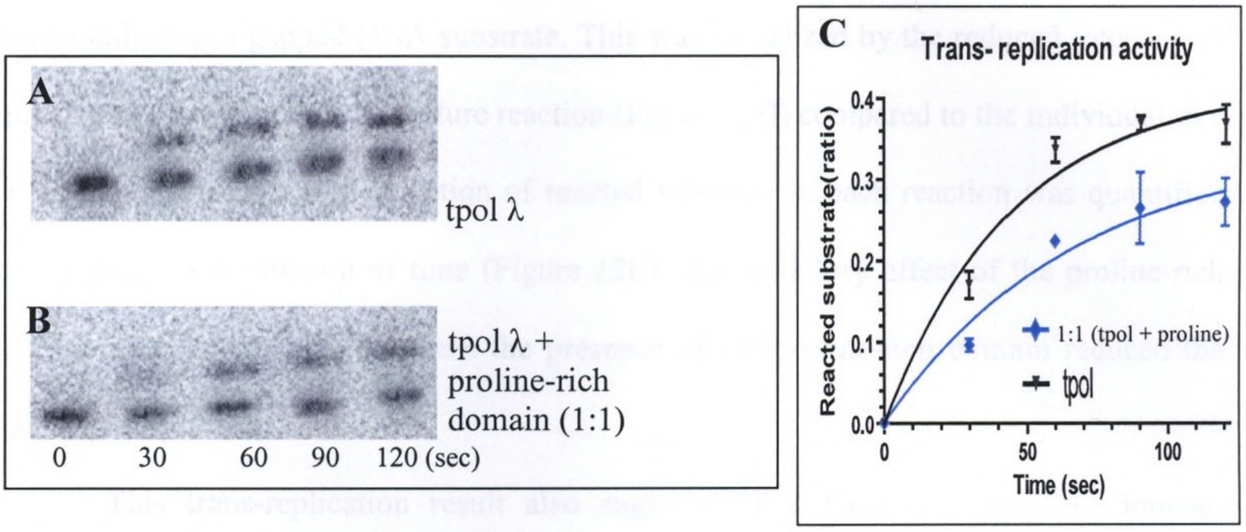


Figure 12. Trans-replication assay
Series of gels illustrating the progression of product formation as a function of time under conditions of 0.1 mM dTTP, 1 nM of protein (tpol λ or tpol λ + proline-rich domain in a 1:1 ratio) and 10 nM of 21-19/41-mer single-nucleotide (adenine) gapped substrate. The replication reactions were carried out by first pre-incubating the protein and DNA substrate for 10 minutes at room temperature (for the 1:1 mixture, the complex was first formed by incubating 2 nM of tpol λ and 2 nM of proline-rich domain at room temperature for 30 minutes) and starting the reaction by adding dTTP and placing the whole mixture in 37°C, thus creating an Adenine to Thymine (A:T) match. The reaction was quenched at various time points by mixing a sample of the reaction mixture with an equal volume of Stop solution, after which the products were resolved on a 10% polyacrylamide gel containing 7M urea and visualized as well as quantified using a Molecular Dynamics Storm PhosphoImager.
The unreacted substrate (21-mer) is shown at the bottom of each gel picture with the extended product (22-mer) located sequentially above the corresponding unreacted substrate. The reaction time intervals (seconds) are shown at the bottom of the panel. **(A)** Incorporation reaction by tpol λ, **(B)** Incorporation reaction by tpol λ + proline-rich domain (1:1). **(C)** The fraction of reacted substrate in each reaction was quantified using a Molecular Dynamics Storm PhosphoImager and calculated using *equation 1* and plotted as a function of time. tpol λ (___), tpol λ + proline-rich domain mixture (___)

The trans-replication experiment indicated that the presence of the proline-rich domain did reduce the activity of the core domain when incorporating the correct nucleotide past a gapped DNA substrate. This was visualized by the reduced intensity of the product band in the 1:1 mixture reaction (Figure 12B) compared to the individual tpol λ (Figure 12A). When the fraction of reacted substrate in each reaction was quantified and plotted as a function of time (Figure 12C), the inhibitory effect of the proline-rich domain was more evident; where the presence of the proline-rich domain reduced the activity of tpol λ by about 10%.

This trans-replication result also suggested that there was possible domain-domain interaction between the proline-rich domain and the core domain that was responsible for modulating the protein activity. The proline-rich domain by itself was tested for nucleotide incorporation activity (data not shown) and was found to be deficient. This implied that the change in the tpol λ activity was due to the presence and possibly the interaction with the proline-rich domain.

3.3 DNA Binding Assays.

To study the DNA binding ability of the human DNA polymerase Lambda (λ) and examine the effects of the various domains on the ability to bind DNA, we performed some DNA binding experiments (described in section 2.5.4) to test the degree to which the different protein constructs bound to DNA. One possible way that the proline-rich domain regulates the core domain's activity may be by altering how it binds to DNA during replication initiation, and this experiment would provide some insight if this was the case.

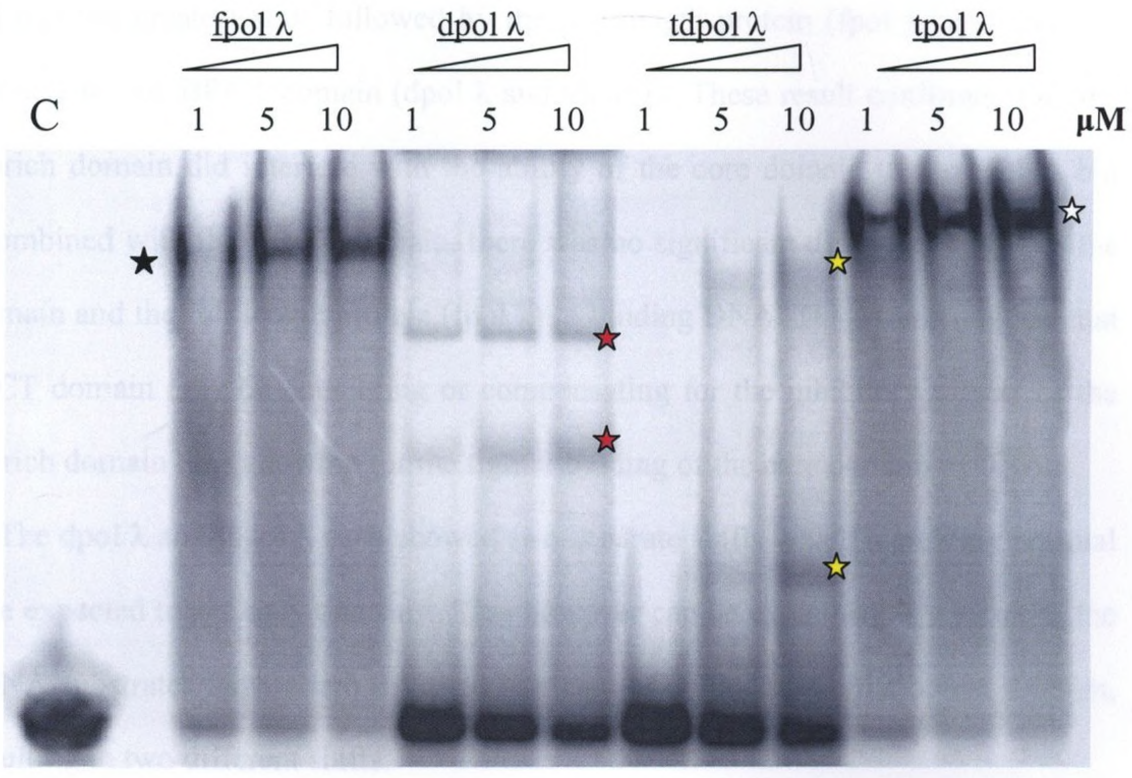


Figure 13. DNA Binding assay for the fpol λ, dpol λ, tdpol λ and tpol λ protein constructs.

Reactions were carried out by incubating different protein concentrations (1, 5, 10 μM) with excess 21-19/41-mer single-nucleotide (adenine) gapped DNA substrate (50 μM) for 30 minutes, after which the reaction products were resolved using a 10% native gel and analyzed/visualized using a Molecular Dynamics Storm PhosphoImager.

The native gel shows the position of the unbound 5'-^{32}P labeled 21 nucleotide long DNA substrate (21-mer) at the bottom of the gel and the product(s) after protein binding (indicated by star figures). The different protein constructs and concentrations used are listed at the top of the gel. Lane 'C' indicates the control lane that contains only the DNA substrate. (★) Indicates the DNA shift as a result of fpol λ binding the substrate. (★) Indicates the DNA shift as a result of dpol λ binding the substrate. (★) Indicates the DNA shift as a result of tdpolλ binding the substrate. (☆) Indicates the DNA shift as a result of tpol λ binding the substrate.

Analysis of the gel shift (DNA Binding) assay indicated that the core domain (tpol λ) had the greatest shift followed by the full length protein (fpol λ) and then the mutants lacking the BRCT domain (dpol λ and tdpol λ). These result confirmed that the proline-rich domain did interfere with the ability of the core domain to bind DNA but when combined with the BRCT domain, there was no significant difference between the core domain and the full length protein (fpol λ) in binding DNA. This would suggest that the BRCT domain may be countering or compensating for the inhibitory effects of the proline-rich domain thus allowing for the tighter binding of the core domain to DNA.

The dpol λ and tdpol λ each showed two separate shifts which were very unusual since we expected to see only one shift. This however can be explained as a result of the same DNA substrates having two different numbers of polymerase units bound to them, thus resulting in two different shifts.

3.4 Circular Dichroism (CD) Spectropolarimetry.

To characterize the structural properties of the proline-rich domain and investigate if there was any interaction with the core domain, we carried out CD experiments using individual proteins as well as a 1:1 equimolar ratio of both the core and proline-rich domains (section 2.6). The rationale behind this experiment is that, if there is interaction that causes any conformational changes, this will result in a 1:1 spectrum which will differ from the sum of the individual components [41,42].

Spectra of dpol λ and individual (proline-rich and core) domains were obtained separately and also a 1:1 mixture of the two domains. The individual domain spectra were then added up to generate a hypothetical spectrum that would represent a situation where the two domains did not interact. The hypothetical spectrum was then compared to the spectrum of the 1:1 mixture and dpol λ spectrum for any similarities/differences.

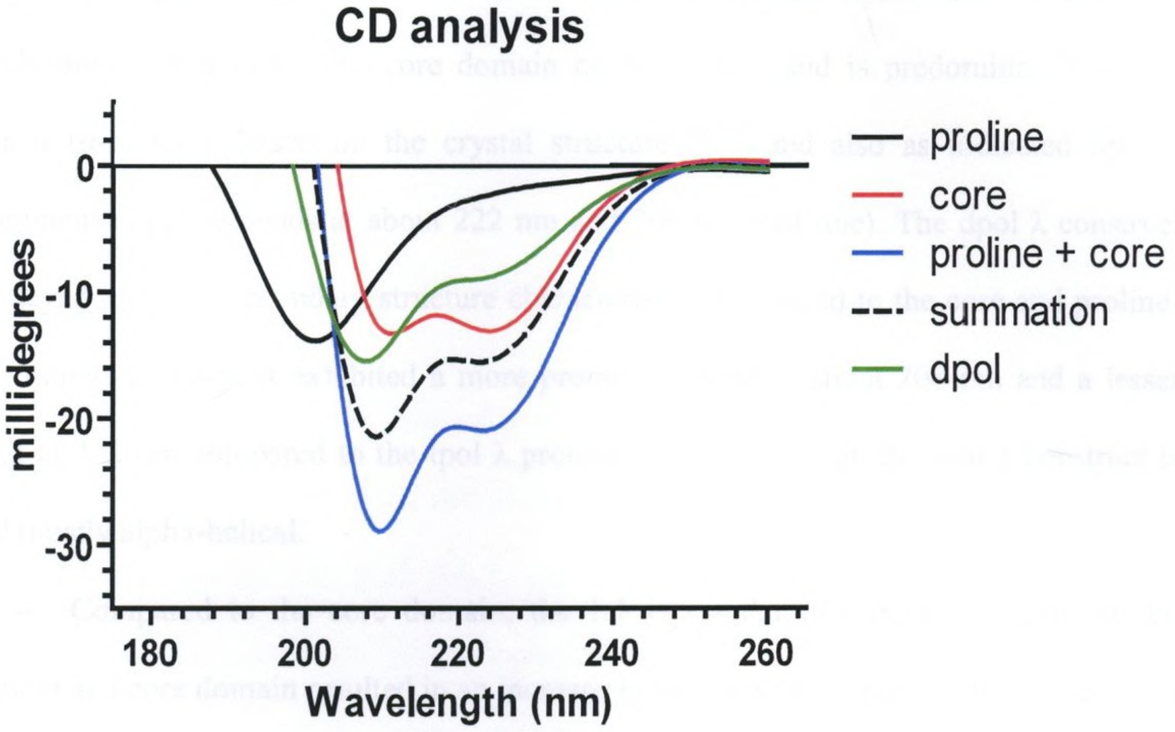


Figure 14. Circular dichroism spectra of core domain's interaction with the proline-rich domain.
The protein sample was analyzed in CD sizing buffer (Table 2) at a final protein concentration of 50 nM in each case. CD spectra of the individual tpol λ protein (____), the individual proline-rich domain (____) , the individual dpol λ protein (____) and a 1:1 equimolar mixture of the two (tpol λ + proline-rich) domains (____) were recorded as the average of three individual spectral scans at 24°C in the far-UV region from 180 to 260 nm, using a cuvette with 0.1 mm of path length. The hypothetical spectrum (----) was generated by summing up the individual tpol λ spectrum and individual proline-rich domain spectrum.

The prominent negative band (solid black line) centered about 198 nm clearly reflects that the Proline-rich domain is mostly disordered under the experimental conditions studied here. The core domain on the other hand is predominantly alpha-helical (α-helical), based on the crystal structure [6,7,8] and also as indicated by the prominent negative bands at about 222 nm and 208 nm (red line). The dpol λ construct displayed different secondary structure characteristics compared to the core and proline-rich domains, where it exhibited a more prominent band at about 208 nm and a lesser band at 222 nm compared to the tpol λ protein. This would indicate dpol λ construct is still mostly alpha-helical.

Compared to the core domain, the 1:1 equimolar mixture of the proline-rich domain and core domain resulted in an increase in intensity about the 222 nm and 208 nm minima (blue line). However, the minimum at 208 nm underwent the largest increase in intensity compared to the minimum at 222 nm. This indicated that the secondary structure of the tpol λ protein still remained predominantly α-helical however, the presence of the proline-rich domain resulted in an increase in the α-helical content in the sample.

The summation of the two individual spectra (proline-rich domain alone and core domain alone) resulted in a hypothetical spectrum (dotted line) different from the 1:1 mixture (blue line). Similar to the 1:1 spectrum, the summation spectrum also showed an increase in intensity about the 208 nm and 222 nm minima compared to the core domain spectrum; however the increase in intensity was not as great. This change in spectrum indicated that, if the two individual domains (proline-rich and core domains) were in solution and not interacting, you would see a spectrum representing an increase in α-helical content (dotted line) compared to the individual core domain (red line); however the increase would be to a lesser extent compared to the 1:1 mixture (blue line).

The difference between the spectrum of the summation of individual domains and the spectrum of the 1:1 mixture indicated that there was interaction between the proline-rich domain and the core domain of DNA polymerase Lambda that resulted in changes in secondary structure within the protein.

3.5 No interaction detected by His Pull-down assays

To further confirm the interactions between the core domain and the proline-rich domain, we carried out His pull-down assays using nickel-charged beads. The buffers had salt concentrations of 150 mM NaCl, similar to the protein sizing buffer. This salt concentration is close to the physiological level and would ensure that non-specific interactions were disrupted while allowing specific interactions to be maintained.

The core domain was used as the His-tagged protein while the proline-rich domain was used as the binding partner. The control experiments using the individual domains showed that the proline-rich domain did not bind strongly enough to the nickel beads, and majority of it was found in the flow-through fractions and was completely washed off the beads using His wash buffer that contained 40 mM imidazole. The core domain had strong binding to the nickel beads and had to be eluted using the His elution buffer that contained 300 mM imidazole (Figure 15A).

In brief, the experiment was carried out by incubating the core and proline-rich domains together, followed by incubating the mixture with the nickel beads while rocking the whole mixture gently, and then washing of any unbound/excess proteins, and finally eluting the bound proteins. Samples from each step of the experiment were analyzed using SDS gel analysis.

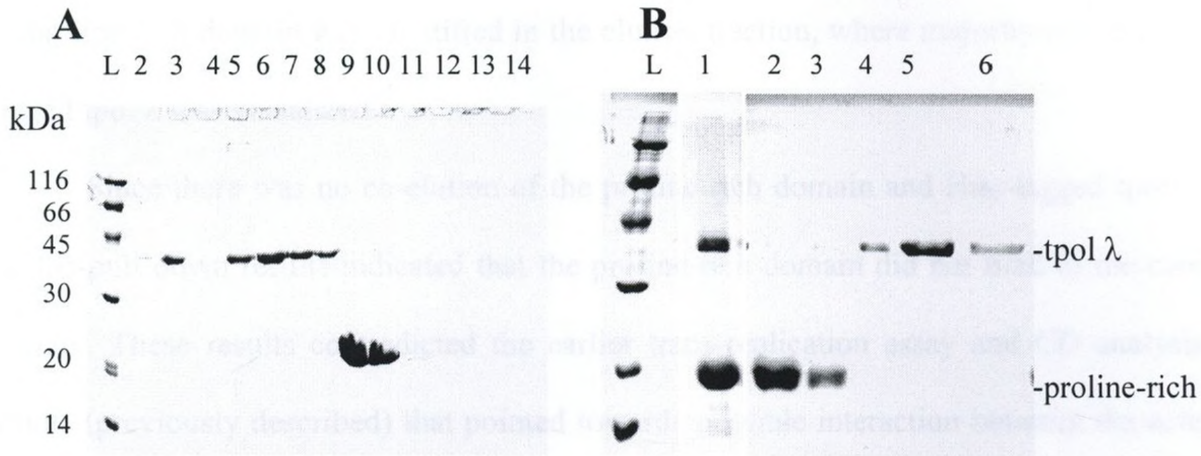


Figure 15. His-pull down experiment between the His_6-tagged tpol λ and the proline-rich domain. The experiment was carried out by incubating the His-tagged core domain (His_6-tpol λ) and excess proline-rich domain together for 1 hour, after which the mixture was incubated with nickel charged beads. The beads were subsequently washed thoroughly to remove any unbound/excess proteins and finally, the bound His-fusion protein with the interacting partner was eluted using the His Elution buffer. The proteins were detected using SDS gel analysis along with Coomassie Brilliant Blue staining.

A) Control experiment of the individual domains incubated with the nickel charged beads separately. Lanes 2-9 represent the tpol λ control experiment and Lanes 10-14 represent the proline-rich domain control experiment. Lane 'L' is the low molecular weight marker, Lane 2 is the flow through fraction, lanes 3-4 are the wash fractions and lanes 5-8 are the elution fractions for tpol λ. Lane 9 is the flow through fraction, lanes 10-11 are the wash fractions and lanes 12-14 are the elution fractions for the proline-rich domain.

B) His-pull down experiment of the His_6-tpol λ and proline-rich domain mixture incubated with the nickel charged beads. Lane 'L' is the low molecular weight marker. Lane 1 is the mixed proline-rich and core domain sample (load), Lane 2 is the flow through fraction, lane 3 is the wash fraction and lanes 4-6 are the elution fractions.

In the pull down experiment (Figure 15B), majority of the proline-rich domain was identified in the flow through fraction and some in the wash fraction. Furthermore, no proline-rich domain was identified in the elution fraction, where majority of the His_6-tagged tpol λ was visualized.

Since there was no co-elution of the proline-rich domain and His_6-tagged tpol λ, the his-pull down results indicated that the proline-rich domain did not bind to the core domain. These results contradicted the earlier trans-replication assay and CD analysis results (previously described) that pointed towards possible interaction between the core and proline-rich domains. This suggested that either there was an error in both the trans-replication and CD experiments that indicated there was domain-domain interaction between the core and proline-rich domain, or that there was no error and that the interaction was weak or transient, therefore being unable to detect it by the His-pull down assay.

3.6 Core domain and proline-rich domain interaction detected by NMR spectroscopy

To obtain more substantial information about the interaction between the core domain and the proline-rich domain, we decided to use NMR spectroscopy to identify possible domain-domain interaction and also determine how tight the binding was.

Using an ^{15}N-labelled protein of interest (proline-rich domain) and an unlabelled binding partner (tpol λ), we monitored the structural changes of the isotopically enriched protein caused by the presence of the unlabelled partner. This was done by performing a series of ^{1}H-^{15}N HSQC experiments and analyzing the different spectra that resulted from the addition of tpol λ (previously described in the methods and material section).

3.6.1 Overexpression and purification of Human ^{15}N-labelled Proline-rich domain

NMR spectroscopy is a powerful technique in identifying the structural and dynamic properties of proteins; however, large quantities of protein (milligram levels) with high purity are often required for these studies. By using the described expression protocol (section 2.8.1) and a purification protocol similar to the unlabelled proline-rich domain, we were able to obtain close to 7 mg (0.6 mM) of ^{15}N-proline-rich domain (Figure 16).

The expression and purification of the tpol λ was carried out as previously described in the methods section and the final yield was about 11 mg (0.3 mM) of the unlabeled protein. The expression and purification methods for both proteins were sufficient to provide enough protein samples to carry out the study, and further optimizations were not necessary.

3.6.2 Two-Dimensional ^{1}H-^{15}N HSQC Experiments

The ^{1}H-^{15}N HSQC spectrum is a contour map of the chemical shifts of ^{1}H and ^{15}N of each amide in a protein backbone, and because there is only one amide proton per amino acid, each HSQC signal represents one single amino acid and also reflects its chemical environment. Thus, the ^{1}H-^{15}N HSQC spectrum represents the protein "fingerprint" [66]. When a molecule/ligand binds to a protein, the chemical environment of the residues in the binding site is altered, hence resulting in a change in chemical shift signal of the residues involved. By comparing the ^{1}H-^{15}N HSQC spectra of an ^{15}N-labeled target protein in the absence and presence of ligand, we obtain information about protein–ligand interactions, where the spectrum of the chemical shifts of the amino acid residues in the ligand-binding site move or fade away if there is interaction [66].

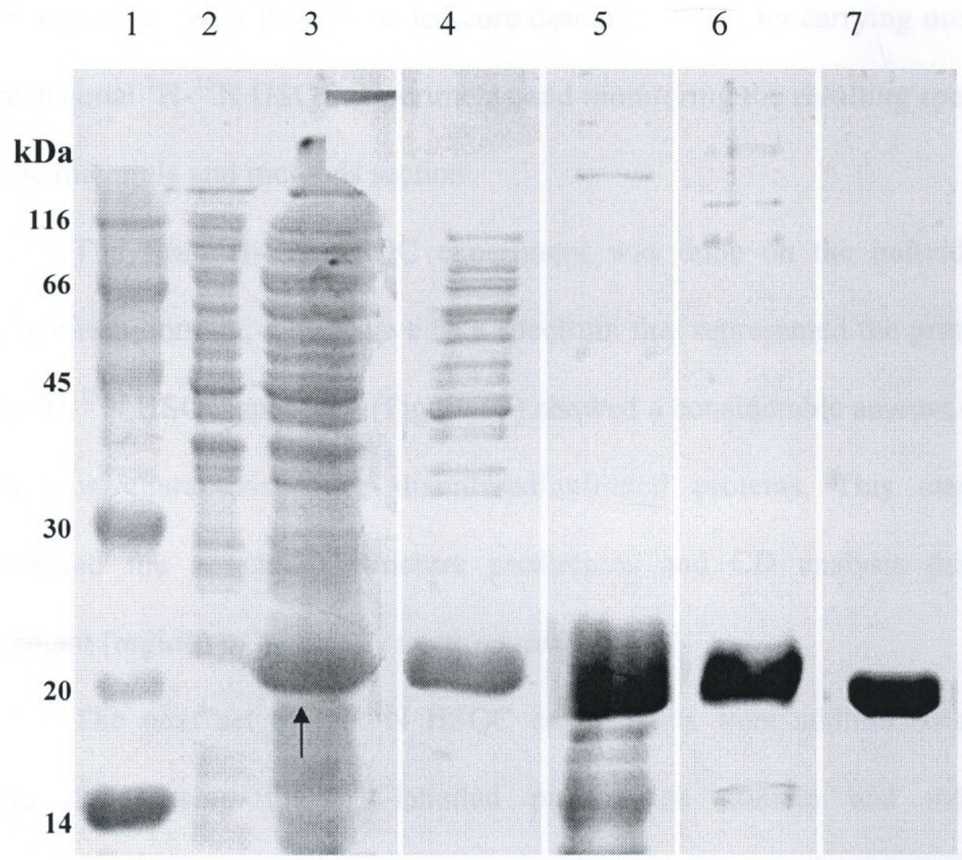


Figure 16. SDS-PAGE gel of the expression and purification of the ^{15}N-labelled proline-rich domain. The ^{15}N-labelled proline-rich domain was expressed in *E. coli* BL21(DE3) cells with M9 minimal media and induced using IPTG and was purified using a series of affinity (HiTrap Nickel column) and ion exchange (HiTrap Q column) chromatographies. The His_6-tag was eventually cleaved off using TEV protease.
The protein samples from each step of purification were analyzed by SDS-PAGE using a 20% polyacrylamide gel and stained with Coomassie Brilliant Blue R-250 dye. Lane 1 represents the low molecular weight marker with the corresponding sizes listed on the left. Lane 2 contains the uninduced *E.coli* BL21 (DE3) cells. Lane 3 contains the *E.coli* cell induced using 0.4 mM IPTG. The arrow points to the expressed protein. Lane 4 contains the protein sample loaded onto the HiTrap nickel column. Lane 5 contains the His_6-fusion protein eluted from the HiTrap nickel column. Lane 6 contains the protein sample purified through the HiTrap Q column. Lane 7 contains the final ^{15}N-labelled proline-rich domain with the His_6-tag cleaved off using TEV protease.

Using the ^{15}N-labelled proline-rich domain we set out to determine if there was any interaction with the unlabelled core domain (tpol λ) by carrying out a series of two-dimensional ^{1}H-^{15}N HSQC experiments and monitoring the resulting spectra as described in the materials and methods section.

The first ^{1}H-^{15}N HSQC experiment was done on the individual ^{15}N-labelled proline-rich domain, which gave us a spectrum that represented the protein 'fingerprint'. The ^{1}H-^{15}N HSQC spectrum (Figure 17) showed a considerable amount of signal overlap which is characteristic for disordered/unfolded proteins. This observation further supported the secondary structure predictions and CD analysis that indicated this fragment (region) to be mostly unstructured.

The next set of ^{1}H-^{15}N HSQC experiments were addition experiments, which were done using the ^{15}N-labelled proline-rich domain and increasing tpol λ concentrations (1:1 and 1:2 molar ratios of proline-rich domain to tpol λ). The ^{1}H-^{15}N HSQC spectra obtained in the addition experiments and the individual ^{15}N-labelled proline-rich domain spectrum were overlaid and analyzed to detect changes in the ^{1}H and ^{15}N chemical shift.

The overlaid spectra (Figure 18) revealed significant changes in some of the ^{1}H-^{15}N HSQC signals as indicated by the arrows. These regions of the ^{15}N-labelled proline-rich domain saw the ^{1}H-^{15}N HSQC signals reduce/completely disappear as the tpol λ protein was added into the sample. This change in ^{1}H-^{15}N HSQC signals clearly indicated that certain residues within the proline-rich domain saw a considerable change in chemical environment that was due to the presence of, or interaction with the tpol λ. Thus, confirming the domain-domain interaction between the proline-rich domain and core domain of DNA polymerase Lambda.

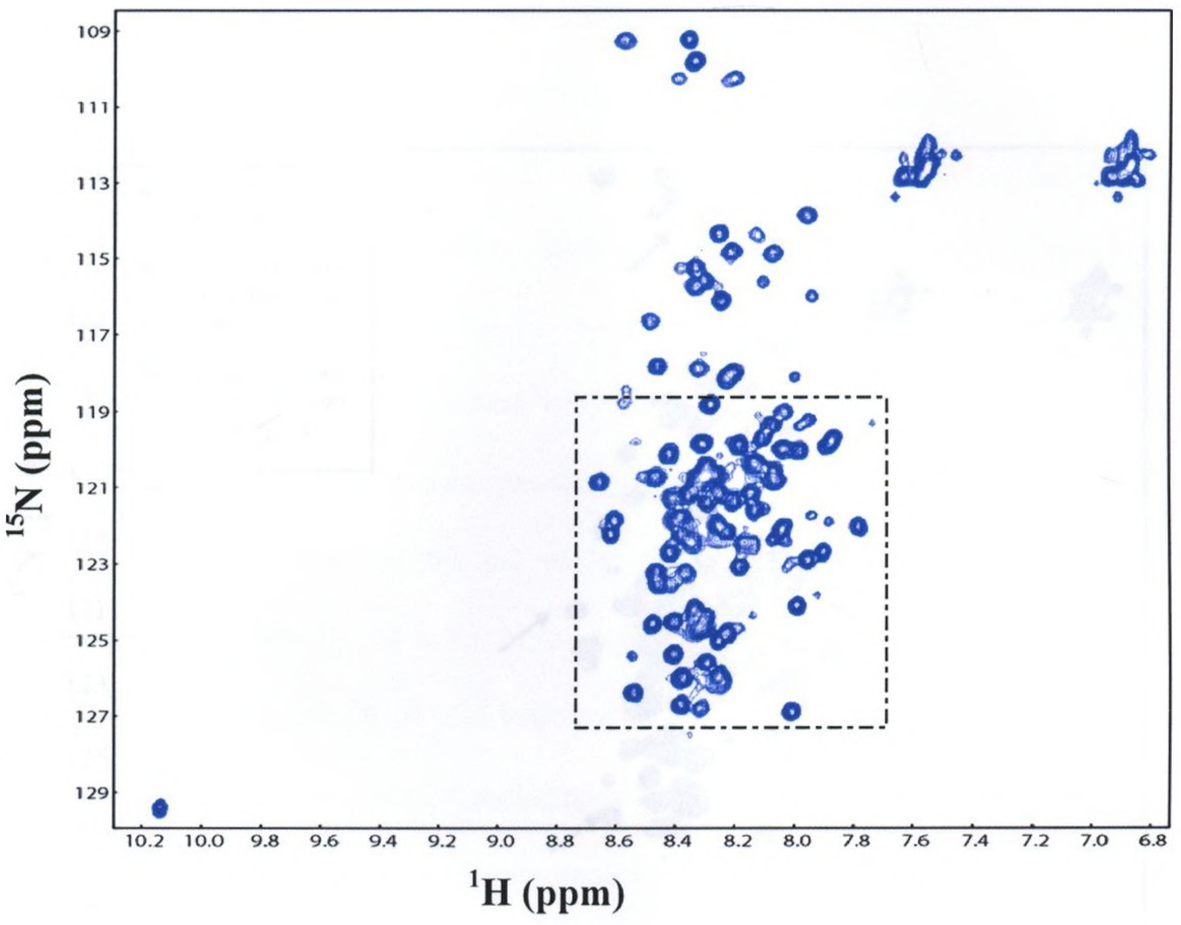


Figure 17. ^{1}H-^{15}N HSQC experiments showing the spectrum ('fingerprint) of the ^{15}N-labelled proline-rich domain.
^{1}H-^{15}N HSQC spectrum of 70 nM ^{15}N- proline-rich domain alone recorded at 600MHz in 50 mM phosphate buffer pH 7.0 and 100 NaCl. The boxed area highlights a region of considerable signal-overlap which is indicative of an unfolded region.

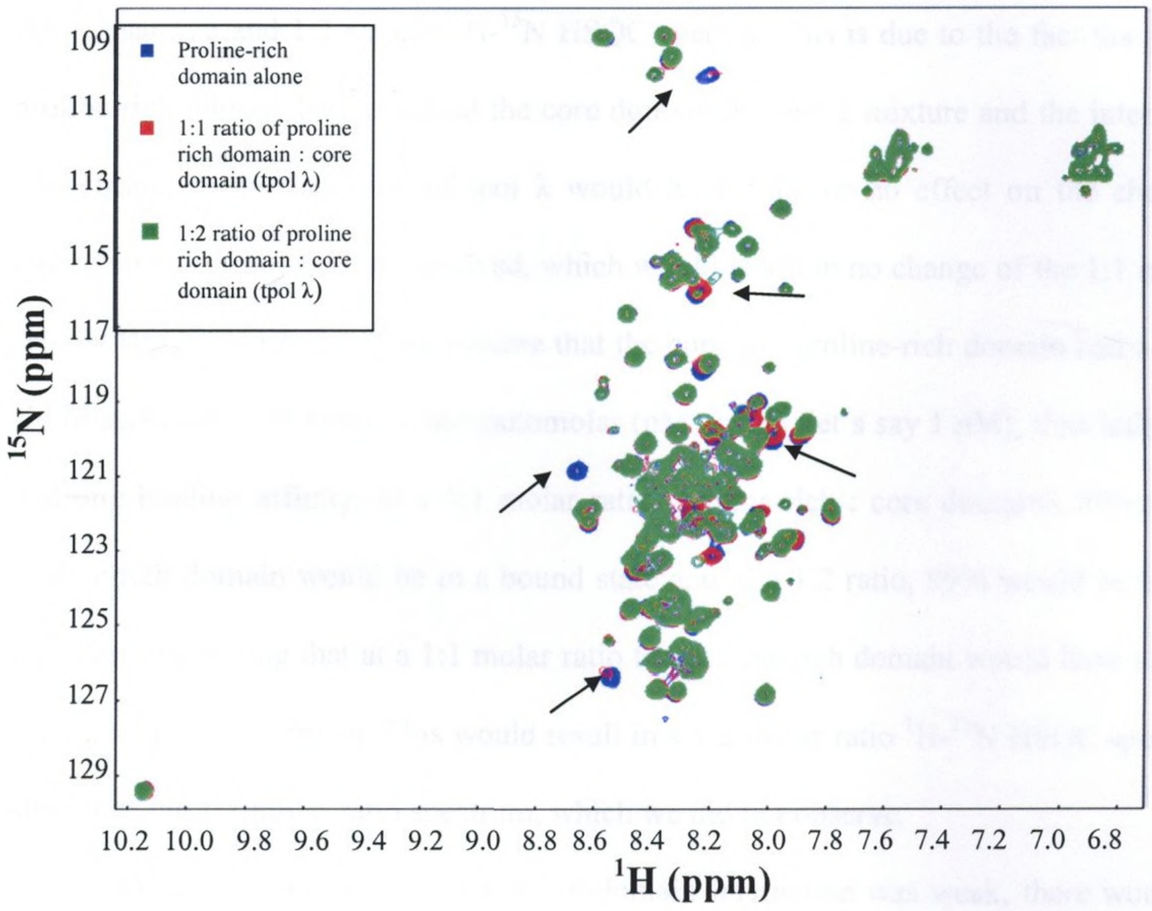


Figure 18. ^{1}H-^{15}N HSQC experiments showing change in $^{1}H_N$ and ^{15}N chemical shifts of the proline-rich domain of Pol λ induced by the core domain (tpol λ).
Overlaid ^{1}H-^{15}N HSQC spectra of 70 nM ^{15}N-proline-rich domain alone (Blue), 1:1 molar ratio of ^{15}N- proline-rich domain to core domain (Red) and 1:2 molar ratio of ^{15}N-proline-rich domain to core domain (green), recorded at 600MHz in 50 mM phosphate buffer pH 7.0 and 100 NaCl. The arrows point to regions with major changes in ^{1}H-^{15}N HSQC chemical shift signal.

The ^{1}H-^{15}N HSQC addition experiments also provide an insight into the strength of the interaction between the proline-rich and core domains of Pol λ. In the case of strong interaction between the two domains, we would not have seen any change between the 1:1 sample and 1:2 sample ^{1}H-^{15}N HSQC spectra. This is due to the fact that, if the proline-rich domain had saturated the core domain in the 1:1 mixture and the interaction was strong, further addition of tpol λ would have little or no effect on the chemical environment of the residues involved, which would result in no change of the 1:1 sample ^{1}H-^{15}N HSQC spectrum. If we assume that the core and proline-rich domain had a small Kd (dissociation constant), in the nanomolar (nM) range (let's say 1 nM), thus indicating a strong binding affinity; at a 1:1 molar ratio (proline-rich : core domain), 89% of the proline-rich domain would be in a bound state and at a 1:2 ratio, 99% would be bound; therefore suggesting that at a 1:1 molar ratio the proline-rich domain would have already saturated the core domain. This would result in a 1:2 molar ratio ^{1}H-^{15}N HSQC spectrum similar to the 1:1 molar ratio spectrum, which we did not observe.

On the other hand, if the domain-domain interaction was weak, there would be constant changes in the chemical environment of the residues involved that would result in changes of the ^{1}H-^{15}N HSQC spectra from the 1:1 sample to the 1:2 sample as the tpol λ protein was added. The overlaid spectra of the 1:0, 1:1 and 1:2 samples (Figure 18) clearly shows regions of the ^{15}N-labelled proline-rich domain (indicated by the arrows) that change as an equal molar ratio of tpol λ is added into the sample and change even further as more (twice the concentration) tpol λ is added, indicating that the interaction between these domains is weak at best. Assuming that the core and proline-rich domain had a large Kd (dissociation constant), in the micromolar (μM) range (let's say 1 μM),

thus indicating a weak binding affinity; at a 1:1 molar ratio (proline-rich : core domain), only 6% of the proline-rich domain would in a bound state and at a 1:2 ratio, 12% would be bound; therefore suggesting that even at a 1:2 molar ratio, the proline-rich domain would be far from saturating the core domain (a 1:20 molar ratio would result in 58% saturation while a 1:100 molar ratio would result in 87% saturation of the core domain by the proline-rich domain). This would result in a 1:2 molar ratio ^{1}H-^{15}N HSQC spectrum different from the 1:1 molar ratio spectrum, which is what we observe.

CHAPTER 4: DISCUSSION

4.1 BRCT and Proline-rich domains of Pol λ affect nucleotide incorporation

Our protein activity studies illustrated that the proline-rich domain functionally suppressed the polymerase activity of Pol λ, but at the same time, the BRCT domain appeared to slightly limit this inhibition. Previous studies have also determined that the proline-rich domain serves as a suppressor of DNA polymerase activity (SDPA) that functionally suppresses the Pol λ activity by about 17 fold [10]; and additional *in vitro* experiments using a chimeric protein composed of the N-terminal domains (proline-rich and BRCT domains) of Pol λ and Pol β, revealed about a 50 fold decrease in Pol β activity, further supporting this role. In this study, the presence of only the proline-rich domain (dpol λ) resulted in a major reduction (approximately 50 %) in polymerase activity of the core domain while the presence of the BRCT domain (fpol λ) resulted in a 25 % recovery of the activity. This demonstrated the ability of the BRCT domain to influence nucleotide gap filling efficiency, however, the proline-rich domain appeared to be the primary regulator of nucleotide incorporation. One interesting observation was the effect of deleting the first 22 N-terminal residues of the proline-rich domain, which seemed to greatly reduce the domain's ability to limit nucleotide incorporation. To date, features in this region have not been characterized and we propose that it may contain key residues involved in protein regulation.

With the limited knowledge about the Pol λ proline-rich domain, its actual role in different DNA repair pathways and cellular responsiveness to DNA damage has yet to be determined. It is generally assumed to be dispensable for polymerase activity based on the idea that it simply couples polymerase action to the protein-protein or protein-DNA

interactions during the gap filling process, however [4], studies have shown that it contains a high number of serine and threonine residues [48] that could serve as possible phosphorylation sites during regulation of DNA repair processes [21] or cell cycle check point upon DNA damage [49].

Other studies on the tpol λ activity have found that it makes up to a 100 times more mistakes than the similar repair protein, DNA polymerase beta [1,4]. Therefore, with the high affinity for dNTPs and low fidelity of Pol λ, we speculate that the proline-rich domain functions as a regulator of polymerase activity and nucleotide incorporation, to reduce hypermutations within the DNA. However, the understanding about how the presence of a proline-rich region influences the protein activity when it comes to DNA repair, still remains ambiguous, and could be that it functions to promote enzyme localization through protein-protein interactions and be a target for post-translational modifications [21].

4.2 Fidelity regulation of Pol λ by its BRCT and Proline-rich domains

The DNA polymerase Lambda is a protein that exhibits unusually low fidelity and unique error specificity [8], where it is the only human DNA polymerase studied to date that displays an average single base deletion error rate that exceeds its average single base substitution error rate [32]. It deletes single nucleotides at a rate that is much higher than for Pol β but generates single base substitution errors at rates that are only slightly higher [6, 32], reflecting its capability to use template–primers with limited base pair homology at the primer terminus.

In this study, we showed that the presence of the N-terminal BRCT and proline-rich domains have significant effects on the fidelity of Pol λ. The tpol λ (core domain)

construct displayed a significant (up to 30%) loss in fidelity compared to the fpol λ protein. Earlier data from single-turnover kinetic assays [1,4] measured the fidelity of fpol λ to be 10 to 100 fold higher than the fidelity of tpol λ, further supporting the observations made in our study that indicate the N-terminal domains upregulate the protein fidelity. Prior experiments by Dr. Z. Suo's lab (Ohio State University) have also shown that the absence of the N-terminal domains does not cause any type of misfolding of tpol λ that contributes to its low fidelity. This is because they observed that, although tpol λ possessed slightly higher correct nucleotide incorporation efficiency than fpol λ, it bound to all correct and incorrect nucleotides with similarly high affinity [4]. While the presence of the proline-rich domain did increase the protein fidelity (dpol λ) the presence of the BRCT domain (fpol λ) seemed to synergistically contribute to this enhancement and further limit the misincorporation of the wrong nucleotide. This further supports the theory that the DNA polymerase λ fidelity is controlled not by an accessory protein or a proofreading exonuclease domain but by an internal regulatory domain that works by significantly lowering incorporation rate constants of incorrect nucleotides [4]. It is worth noting that in these fidelity assays, we observed the loss of fidelity by the proline rich domain when the N-terminal 22 residues were deleted (tdpol λ), and thus, further supporting our speculations that this region contains key residues involved in protein regulation.

This general observation that fpol λ (and dpol λ) were slower yet more faithful than tpol λ in filling single-nucleotide-gapped DNA goes against the general trend summarized from a survey of the A-, B-, X-, and Y-families [50] that correlated a more catalytically efficient polymerase to having a lower base substitution rate. It is not clear

what contributes to this correlation between the DNA polymerase lambda high fidelity and low nucleotide incorporation rate but it is thought that the presence of the proline-rich domain in both fpol λ and dpol λ may either tighten the polymerase active site to achieve a better geometric selection [51,52] or shield it (active site) from solvent, thus leading to greater desolvation of a nascent base pair and amplifying the free energy differences between matched and mismatched nucleotide incorporations [53].

Correct dNTP incorporation (high fidelity) during DNA synthesis is essential for preserving genetic information over many generations and avoiding mutations that can initiate and promote human diseases, at the same time, low fidelity DNA synthesis is beneficial for the evolution of species, for generating diversity leading to increased survival of viruses and microbes when subjected to changing environments, and for the development of a normal immune system [43]. The crucial enzymes that replicate and maintain genomic stability, have evolved through different mechanisms to adjust their polymerization fidelity to best perform diverse physiological functions and the enhancement of the fidelity of polymerase lambda by its N-terminal domains is a good example.

4.3 Roles of BRCT and Proline-rich domains in DNA binding

The BRCA1 C- terminal (BRCT) domain is a common feature of polymerases involved in non-homologous end joining (NHEJ), where it plays a role in mediating protein-protein or protein-DNA interactions [47]. Enzymes such as Pol λ, yeast Pol IV, Pol μ, and TdT all posses an N-terminal BRCT domain and have all been linked to the NHEJ repair of DNA double strand breaks [49, 54].

In our DNA binding experiments, we wanted to observe how the different Pol λ constructs bound to DNA and determine if the various N- and C-terminal domains had any effect on the protein's ability to bind DNA. The aim was to test the idea that some of the activity and fidelity regulation demonstrated by the N-terminal domains could perhaps be due to their influence on the core domain's ability to interact with DNA. We observed that the proline-rich domain by itself (dpol λ and tdpol λ) greatly reduced the core domain's (tpol λ) ability to bind DNA. No other studies have been done so far to examine the effects of the proline-rich domain on DNA binding and the exact mode by which it significantly limits the core domain's ability to bind DNA is still unknown. However, it is very likely that due to the great negative charge (pI = 4.3) of the proline-rich domain, we have a case where the domain exerts a repelling force on the DNA (due to like charges) or even compete for the DNA binding site on the core domain, which happens to contain mostly positively charged residues.

More interestingly, the DNA binding studies showed that the BRCT domain in context with the proline-rich domain (fpol λ) fully restored the DNA binding ability of the core domain (tpol λ). Similar observations have been made in Pol μ, which belongs to the X-family DNA polymerases and also contains a BRCT domain. A deletion mutant of Pol μ that lacked the BRCT domain but retained full polymerase activity was shown to lack the ability to shift DNA, indicating that the BRCT domain was required for DNA-protein interaction [46]. The manner by which the BRCT domain upregulates DNA binding is still ambiguous, but we can speculate that it possibly reduces the inhibitory effects of the proline-rich domain or even play a compensatory role.

Generally, BRCT domains consists of ~95 amino acid residues that may occur as a single copy or tandem repeat at the amino or carboxyl terminus of numerous proteins such as; BRCA1, RAD9, XRCC1, terminal deoxynucleotidyltransferases, DNA-ligases III and IV, and although they differ considerably in their biochemical properties, participation in DNA damage-responsive checkpoints appears to be the common theme. Therefore, the BRCT domain is likely to carry out critical functions in the DNA repair and cell cycle control of organisms from bacteria to humans [47, 49]. In yeast, Pol IV interacts with Dnl4, a subunit of the Dnl4-Lif1 complex, which is the homologue of the human ligase IV-XRCC4 complex through the BRCT domain [44]. This may suggest that the BRCT domain is involved both in the recruitment of the polymerase to the site of damage and in the interactions that link the gap-filling process and the ligation steps [44]. Co-immunoprecipitation studied have also found that Pol λ is able to co-localize to oxidative DNA lesions in situ with one of the oxidized base DNA glycosylases (SMUG1) [23]. Furthermore, the human Ku factors (Ku70/Ku80) have been shown to recruit Pol μ and λ via their BRCT domains [46]. All these observations demonstrate that the BRCT domains of Pol μ and λ are essential for their participation in an *in vitro* reconstitution of NHEJ proteins, which could be due to its ability to greatly enhance DNA binding.

4.4 Interaction between the Proline-rich domain and core domain of Pol λ

Based on the studies mentioned so far (activity, fidelity and DNA binding assays), the proline-rich domain has demonstrated that it has the ability to regulate the activity of the core domain when it comes to various processes such as nucleotide incorporation, fidelity regulation and DNA binding. Earlier studies have demonstrated that the effects

are not due to conformational changes as a result of having or lacking the proline-rich domain [4] but may be as a result of domain-domain interaction with the core domain.

Our observations from the trans-replication, CD and NMR experiments all indicated that there was interaction between the core and proline-rich domains of Pol λ. The trans-replication experiment indicated that the change in nucleotide incorporation induced by the proline-rich domain was a result of direct interaction with the core domain. In addition, the CD analysis also supported the idea that there was interaction between the proline-rich and core domains, which perhaps resulted in some change in secondary structure of the proline-rich domain. The NMR experiment further characterized the interaction as being weak in nature, which helps explain the negative results obtained from the His-pull down experiment.

Here, we were able to conclude that there was weak interaction between the core and proline-rich domains *in vitro* and that this interaction was possibly responsible for the regulation of DNA polymerase lambda's activity. Since it is not fully understood how exactly the presence of the proline-rich domain influences protein activity, we propose that the interaction with the core domain is mostly responsible for the regulation and that it may even possibly occur *in vivo*.

Proline-rich regions (PRRs) of proteins occur commonly in both prokaryotes and eukaryotes and are frequently found as multiple tandem repeats or even non-repetitive sequences (Table 5). These protein regions have often been referred to as protein-protein interaction domains; with the best known examples being the proline-rich ligands bound by SH3 (Src homology 3) domains [69-71], proline-rich tails on RNA polymerase II [73] that have been implicated in interactions with various transcription factors, proline-rich

regions found in some transcription factors [75] and the proline-rich ligands of the actin-binding protein profiling [74]. These regions tend to be flexible and form extended structures, which make them hard to crystallize. For this reason, there are very few crystal structures of PRRs and most structural information on such regions has come from solution-state NMR and CD spectrometry and from modeling studies using secondary structure predictions [75]. PRRs in proteins have also been shown to display fast but non-specific interactions, which makes them good candidates for multiple protein associations [75,76]; as seen in the RNA polymerase II initiation complex [73] and SH3 domain-binding proteins [70,71]. In addition, these region also display a different mode of interaction from the classic lock and key (or induced fit) mechanism that is exhibited by most enzyme-substrate reactions; where the PRRs often rely on multiple weak binding sites [75]. This multiple weak binding makes the PRRs difficult to study through biochemical methods such as site-directed mutagenesis, however, large scale deletions or additions can be used to analyze these regions together with physiochemical studies such as CD and NMR [75,77]. Despite the wide distribution (Table 5) and extensive study of the functions of PRRs, the manner by which these regions operate still remains unclear.

4.5 The Proline-rich domain of Pol λ is intrinsically disordered.

A significant number of proteins have recently been identified as being unstructured and yet still be biologically functional [55,56]. Furthermore, studies have also predicted that a considerable fraction of proteins in various organisms are either completely disordered or contain long disordered regions [55,57], which have been found to play essential roles in various reactions *in vivo*.

Proline-rich domain sequence

IFIPSRYLDH PQPSKAEQDA SIPPGTHEAL
LQTALSPPPP PTRPVSPPQK AKEAPNTQAQ
PISDDEASDG EETQVSAADL EALISGHYPT
SLEGDCEPSP APAVLDKWVC AQP [113]

Amino acid composition:		
Ala (A)	**14**	12.40%
Arg (R)	**2**	1.80%
Asn (N)	**1**	0.90%
Asp (D)	**8**	7.10%
Cys (C)	**2**	1.80%
Gln (Q)	**8**	7.10%
Glu (E)	**9**	8.00%
Gly (G)	**4**	3.50%
His (H)	**3**	2.70%
Ile (I)	**5**	4.40%
Leu (L)	**8**	7.10%
Lys (K)	**4**	3.50%
Met (M)	**0**	0.00%
Phe (F)	**1**	0.90%
Pro (P)	**20**	17.70%
Ser (S)	**11**	9.70%
Thr (T)	**6**	5.30%
Trp (W)	**1**	0.90%
Tyr (Y)	**2**	1.80%
Val (V)	**4**	3.50%

Figure 19. The amino acid sequence and composition of the proline-rich domain. The sequence is composed of 113 amino acid residues, with 17.7% of them being proline residues. It also possesses a significant number of serine (9.7%) and threonine (5.3%) amino acids, which have been shown to be possible phosphorylation sites [48]. 52% of the residues are hydrophobic, 15% are acidic, 5% are basic, 6.3% are aromatic and 1.8% are sulfur-containing residues. This amino acid composition results in a high net charge (pI ~ 4.3), which might explain its unfolded state.

Table 5. Different proteins containing proline-rich regions (PRRs).
The table shows the wide distribution of PRRs in various organisms, different organization of PRRs, the function of proteins containing PRRs and the roles these proteins play.

Proteins with repetitive short proline-rich sequences				
Name	***Source***	***Sequence***	***Protein function***	***Remark***
Light chain myosin kinase	Rabbit skeletal muscle	$(AP)_6$	Binds actin	PRR is at N-terminus
βB1 crystallin	Ox eye lens	$GP_3GPAPGSG(PA)_5Q(PA)_2$	Cytoskeletal binding?	PRR is at N-terminus
OmpA	E coli	$(AP)_4$	Major outer membrane protein	Mediates F-dependent conjugation
Proteins with tandemly repeated proline-rich sequences				
Mucins	Man	$(GSTAPPAHGVTSAPDTRPAP)_n$	Lubrication of epithelium	Glycosylated
Salivary PRPs	Man, mouse	$(PQGPPQQGG)_n$	Polyphenol binding	Most of the protein is PRR
Gluten	Wheat	GYYPTSPQQ, PGQGQQ; many repeats	Cereal storage protein	Small N- and C-terminal domains
RNA polymerase II	Man	YSPTSPS (26 times)	Transcription	Binds TFIID?
Rhodopsin	Squid	$(PPQGY)_{10}$	Vision	Organizes microvillar structure?
Synapsin	Man	PQPAGPPAQQVPPPQQG (x 3)	Regulates vesicle release?	Binds vesicle and cytoskeleton?
Non-repetitive PRRs				
CTF/NF-1 family	Man	PPHLNPQDPLKDLVSLACDPASQQPGPPTLRPTRPLQTVPLT	Transcription activator	Binds RNA polymerase II/TFIID
Calcineurin A	Man	$MAAPEPARAAP_{11}GA$	Calmodulin-regulated phosphatase	N-terminus binds calmodulin along helix?
Consensus SH3-binding	Mouse, rat	XPXXPPP-XP	Binds SH3	Signal transduction; cytoskeletal sequence regulation
Pyruvate dehydrogenase	E. coi	$GA_2PA_3PAKQEA_3PAPAAKAEAPA_3PA_2KA$	Dehydrogenase	Mobile linker
Vitelline	Drosophila melanogaster	$PYA_2(PA)_2YSAPA_2S_2GYPAP_2$	Eggshell structure	Membrane bound
Proteins rich in hydroxyproline				
Collagen	Man	$(GP\mathit{P})_{350}$ (with variations on P and *P*)	Stiff connective fibre	Triple helix
Clq	Man	$(\mathit{P}GX)3(ASXGX)_2(FGX)_2PGX\mathit{P}$	Blood defence	system Kinked triple helix
Extensin P1	Tomato, carrot etc.	$[S\mathit{PPPP}(VKPYHP)T\mathit{P}IKY]_n$	Cell wall constituent	Protects plant against damage?
Cucumber peel cupredoxin	Cucumber	PPPSSSPPSSVM*PPP*VMP*PPS*P*S*	Electron transfer	PRR is C-terminal extension

(Adapted from Williamson, P.M. 1994)

In this study, we were able to identify the proline-rich domain of DNA polymerase lambda as a disordered region of the protein using CD and NMR analysis. Other than structure predictions, no actual data/analysis about the secondary structure of the proline-rich domain had been published. And although we carried out the secondary structure analysis of the proline-rich domain without the rest of the protein, we believe that it gave us a relatively accurate picture of its organization. We believe the observed disordered (unstructured) nature of the proline-rich domain may be responsible of the domains ability to regulate the protein activity. The flexibility and accessibility of the domain may perhaps coordinate or mediate protein-protein interactions, protein-DNA interaction or even conformation changes in Pol λ.

For a long period of time, proteins have been thought to carry out various functions via a stable three-dimensional structure; however, disordered (unstructured) proteins have recently been shown to undergo disorder-to-order transitions as a way of carrying out their function [55-57]. This sort of transition was not observed in the proline-rich domain as the core domain was added, during the NMR experiment. The ^{1}H-^{15}N HSQC spectrum of the mixed sample still showed significant amount of signal overlap, which is indicative of an unstructured protein. Although we did not observe any disorder-to-order transitions, we believe that the proline-rich domain exerted its effect while remaining in an unfolded state. As it happens, some proteins have been found to require being unfolded so as to be functional; such as p27 (cyclin-dependent kinase (CDK) inhibitor), which is mostly disordered in its free state thus allowing it to interact with cyclin A and CDK2 [58,59,67]. Some of the features that contribute to the unfolded nature of disordered proteins include higher net charge with extreme isoelectric points (pI), which leads to charge repulsions

between residues and lower mean hydrophobicity in their sequences, compared to globular proteins, which reduces the driving force in protein folding [60]. The high charge distribution is a characteristic also shared by the proline-rich domain, which has an isoelectric point of ~4.3 and even though it does not display low hydrophobicity, the large number of proline residues may contribute to its lack of a stable secondary/tertiary structure under physiological conditions (Figure 20). Unstructured proteins are often found to be involved in processes such as signaling and cell cycle control [61-65]. Their ability to be flexible, allows their entire primary structure to be accessible to solvents, ligand and substrates [64,65]. In the case of the Pol λ protein, the proline-rich domain is the unfolded region of the protein and may be the solvent/ligand/substrate accessible region. Moreover, it contains multiple targets for phosphorylation and glycosylation. This is supported by studies that found Pol λ to be phosphorylated *in vitro* by several Cdk/cyclin complexes, including Cdk2/cyclin A, in its proline-rich domain and also *in vivo* in human cells during the cell cycle progression [48]. Furthermore, this was confirmed by the identification of three serines (amino acids 167, 177 and 230, respectively), which could be part of a consensus site for Cdk phosphorylation that are all located in the proline-rich domain of Pol λ [48].

The unfolded (disordered) nature of the proline-rich domain may be the basis by which it regulates various protein activities *in vivo* and *in vitro*. And by being able to study this region of DNA polymerase lambda, we might better understand how the protein is regulated as well as what role it plays *in vivo*.

4.6 Conclusion

Most structural studies performed thus far have been done on the C-terminus of DNA polymerase λ (core domain), however, very little structural information is known about the rest of the protein (BRCT and proline-rich domains), which has led to the limited understanding of the roles these domains play *in vivo*. To better understand the kind of fold the protein adopts and also determine what sorts of interactions are involved between the different domains within the protein, we used a series of structural and biochemical approaches to study the protein.

The protein activity assays supported earlier research data [10] that indicate the proline-rich domain functions as a suppressor domain of DNA polymerase activity (SDPA); furthermore it also found that deleting the first 22 N-terminal amino acids seemed to limit the suppression of this domain. This study also saw the BRCT domain slightly compensate for, or inhibit the effects of the proline-rich domain.

In addition, the protein fidelity assays showed that the proline-rich domain did increase the protein fidelity by limiting the misincorporation of the wrong nucleotides. This increase was further enhanced by the presence of the BRCT domain, which seemed to work synergistically with the proline-rich domain in preventing mismatches in the gap-filling process of DNA repair.

The proline-rich domain also limited the core domain's ability to bind DNA but when combined with the BRCT domain we saw a recovery of the DNA binding ability, in which case, the BRCT domain was playing a compensatory or inhibitory role for the proline-rich domain.

The CD and NMR experiments identified the proline-rich domain as a mostly unfolded protein and that it did interact with the core domain of the protein. Additionally, trans-replication experiments established that the domain-domain interaction between the core and proline-rich domains was responsible for the inhibitory effects of the proline-rich domain on protein activity.

4.7 Significance

This study has only been able to paint a small picture on the functioning of the human DNA polymerase lambda in regards to its proline-rich domain, however, it has provide some important groundwork for future studies in a few areas:

1) Having being able to clone, express and purify the individual proline-rich domain in high yield we can continue to perform NMR and other biophysical studies to extend our knowledge about this domain. The aim would be to assign the residues of this fragment and obtain information about its secondary structure to further support the CD experiments.

2) The positive interaction between the proline-rich domain and core domain determined in this study can be expanded to be more specific. The CD, NMR and trans-replication experiments were only able to confirm that there was domain-domain interaction that did influence the protein activity; however, the residues involved could not be determined. The assigning of the residues of the proline-rich and perhaps the core domain and more NMR experiments, would further provide detailed descriptions of these two domains upon interacting with each other.

3) The BRCT domain also appeared to play a role in regulating the protein activity. This observation paves way for an approach similar to the one used in this study, to be carried out on the BRCT domain. And when combined with the results in this experiment, we can better understand the functioning of the DNA polymerase Lambda.

4.8 Future direction

Having determined that the DNA polymerase Lambda's proline-rich domain does interact with the core domain, and that this interaction has an effect on various protein activities such as; nucleotide incorporation, fidelity regulation and DNA binding, understanding/determining the specific residues involved would provide greater insight into the catalytic involvement of this domain. By taking the NMR experiment further, we can attempt to label and assign (identify) the residues of the proline-rich domain and the core domain and perform similar "addition" experiments that will identify the exact amino acids involved in catalytic regulation. Furthermore, the identified residues can then be mutated and the effects of these mutations studied using a similar approach to the one employed in this study (DNA extension and Binding assays).

The BRCT domain was also shown to be a great contributor in regulating the Pol λ activity possibly by exerting its effects through the proline-rich domain or through the core domain of the protein. By extending an approach similar to that of the study of the proline-rich domain, we can determine what actual role the BRCT domain plays in context with the proline-rich and the core domains of the Pol λ protein.

By continuing with the crystallization attempts of fpol λ, we may be able to finally get a 3D representation of the entire protein that when combined with the existing published data, and the results obtained in this study, we can be able to put everything

together and finally understand the functioning of DNA polymerase lambda and how it relates to its role *in vivo*.

REFERENCES

1. Fiala KA, Abdel-Gawad W, Suo Z. (2004) Pre-steady-state kinetic studies of the fidelity and mechanism of polymerization catalyzed by truncated human DNA polymerase lambda. *Biochemistry.*43(21):6751-62

2. Friedberg CE. (2005). Suffering in Silence: the tolerance of DNA damage. *Nat Rev Mol Cell Biol.* 6(12): 943-955

3. Wilson HS, Beard AW. (2006). Structure and Mechanism of DNA polymerase Beta (β). *Chem. Rev.* 106, 361-382

4. Fiala KA, Duym WW, Zhang J, Suo Z. (2006) Upregulation of the fidelity of Human DNA polymerase lambda by its Non-enzymatic Proline-Rich Domain. *J Biol Chem.* 281 (28): 19038-19044

5. Pastwa E, and Blasiak J., (2003). Non-homologous DNA End Joining. *Acta Biochem Pol.* 50(4):891-908.

6. Garcia-Diaz M, Bebenek K, Sabariegos R, Dominguez O, Rodriguez J, Kirchhoff T, Garcia-Palomero E, Picher AJ, Juarez R, Ruiz JF, Kunkel TA, Blanco L. (2002). DNA polymerase lambda, a novel DNA repair enzyme in human cells. *J Biol Chem.*277(15):13184-91

7. Garcia-Diaz M, Bebenek K, Krahn JM, Blanco L, Kunkel TA, Pedersen LC. (2004). A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol Cell.* 27; 13(4):561-72.

8. Garcia-Diaz M, Bebenek K, Gao G, Pedersen LC, London RE, Kunkel TA. (2005) Structure-function studies of DNA polymerase lambda. DNA Repair. A*mst* 8;4(12):1358-67

9. Kobayashi, Y., Watanabe, M., Okada, Y., Sawa, H., Takai, H., Nakanishi, M., Kawase, Y., Suzuki, H., Nagashima, K., Ikeda, K., and Motoyama, N. (2002). Hydrocephalus, Situs inversus, Chronic sinusitis and male infertility in DNA polymerase Lambda deficient mice: possible implication for the pathogenesis of immotile cilia syndrome. *Mol. Cell. Biol.* 22, 2769-2776.

10. Shimazaki N, Yoshida K, Kobayashi T, Toji S, Tamai K, Koiwai O. (2002). Over-expression of human DNA polymerase lambda in E. coli and characterization of the recombinant enzyme. *Genes Cells*;7(7):639-51

11. Showalter AK, Lamarche BJ, Bakhtina M, Su MI, Tang KH, Tsai MD. (2006). Mechanistic comparison of high-fidelity and error-prone DNA polymerases and ligases involved in DNA repair. *Chem. Rev.* 106(2):340-60

12. Filee, J.; Forterre, P.; Sen-Lin, T.; Laurent, J. (2002). Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins. *J. Mol. Evol.* 54, 763.

13. Dominguez, O.; Ruiz, J. F.; Lain de Lera, T.; Garcia-Diaz, M.; Gonzalez, M. A.; Kirchhoff, T.; Martinez, A. C.; Bernad, A.; Blanco, L. (2000) DNA polymerase mu (Pol μ), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J.* 19, 1731.

14. Ruiz, J. F.; Dominguez, O.; Lain de Lera, T.; Garcia-Diaz, M.; Bernad, A.; Blanco, L. (2001). DNA polymerase mu, a candidate hypermutase? *Philos. Trans. R. Soc. London, Ser. B*, 356, 99

15. Wang, Z.; Castano, I. B.; Adams, C.; Vu, C.; Fitzhugh, D.; Christman, M. F. (2002). Structure/function analysis of the Saccharomyces cerevisiae Trf4/Pol sigma DNA polymerase. *Genetics*, 160, 381

16. Carson, D. R.; Christman, M. F. (2001). Evidence that replication fork components catalyze establishment of cohesion between sister chromatids. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8270

17. Oliveros, M.; Yanez, R. J.; Salas, M. L.; Salas, J.; Vinuela, E.; Blanco,L. (1997). Characterization of an African Swine Fever Virus 20-kDa DNA Polymerase Involved in DNA Repair *J. Biol. Chem.* 272, 30899.

18. Bebenek, K.; Garcia-Diaz, M.; Patishall, S. R.; Kunkel, T. A. (2005). Biochemical properties of Saccharomyces cerevisiae DNA polymerase IV. *J. Biol. Chem.*, 280, 20051

19. Hirose, F.; Hotta, Y.; Yamaguchi, M.; Matsukage, A. (1989). Difference in the expression level of DNA polymerase beta among mouse tissues: high expression in the pachytene spermatocyte. *Exp. Cell Res.*, 181, 169

20. Garcia-Diaz, M.; Dominguez, O.; Lopez-Fernandez, L. A.; de Lera, L. T.; Saniger, M. L.; Ruiz, J. F.; Parraga, M.; Garcia-Ortiz, M. J.; Kirchhoff, T.; del Mazo, J.; Bernad, A.; Blanco, L. (2000). DNA polymerase lambda (Pol λ), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J. Mol. Biol.*, 301, 851

21. Braithwaite, E. K.; Prasad, R.; Shock, D. D.; Hou, E. W.; Beard, W. A.; Wilson, S. H. (2005). DNA polymerase lambda mediates a back-up base excision repair activity in extracts of mouse embryonic fibroblasts. *J. Biol. Chem.*, 280, 18469.

22. Uchiyama, Y.; Kimura, S.; Yamamoto, T.; Ishibashi, T.; Sakaguchi, K. (2004). Plant DNA polymerase lambda, a DNA repair enzyme that functions in plant meristematic and meiotic tissues. *Eur. J. Biochem. 271*, 2799.

23. Braithwaite, E. K.; Kedar, P. S.; Lan, L.; Polosina, Y. Y.; Asagoshi, K.; Poltoratsky, V. P.; Horton, J. K.; Miller, H.; Teebor, G. W.; Yasui, A.; Wilson, S. H. (2005). DNA Polymerase λ Protects Mouse Fibroblasts against Oxidative DNA Damage and Is Recruited to Sites of DNA Damage/Repair *J. Biol. Chem.*, *280*, 31641

24. Lee, J. W.; Blanco, L.; Zhou, T.; Garcia-Diaz, M.; Bebenek, K.; Kunkel, T. A.; Wang, Z.; Povirk, L. F. (2004). implication of DNA Polymerase λ in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts *J. Biol. Chem.*, 279, 805

25. Fan, W.; Wu, X. (2004). DNA polymerase lambda can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex. *Biochem. Biophys. Res. Commun.*, 323, 1328.

26. Maga, G.; Villani, G.; Ramadan, K.; Shevelev, I.; Tanguy Le Gac, N.; Blanco, L.; Blanca, G.; Spadari, S.; Hubscher, U. (2002). Human DNA polymerase lambda functionally and physically interacts with proliferating cell nuclear antigen in normal and translesion DNA synthesis. *J. Biol. Chem.*, 277, 48434

27. Ramadan, K.; Shevelev, I. V.; Maga, G.; Hubscher, U. (2002). DNA Polymerase λ from Calf Thymus Preferentially Replicates Damaged DNA. *J. Biol. Chem.* 277, 18454.

28. Garcia-Diaz M., Bebenek K., Krahn J.M., Kunkel T.A. and Pedersen L.C., (2005). A closed conformation for the Pol lambda catalytic cycle, *Nat. Struct. Mol. Biol.* 12(1):97-8

29. Steitz T.A., (1999). DNA polymerases: structural diversity and common mechanisms, *J. Biol. Chem.* 274. 17395–17398

30. Ramadan K., Maga G., Shevelev I.V., Villani G., Blanco L. and Hubscher U. (2003). Human DNA polymerase lambda possesses terminal deoxyribonucleotidyl transferase activity and can elongate RNA primers: implications for novel functions, *J. Mol. Biol.* 328 63–72

31. Shevelev I., Blanca G., Villani G., Ramadan K., Spadari S., Hubscher U. and Maga G., (2003). Mutagenesis of human DNA polymerase lambda: essential roles of Tyr505 and Phe506 for both DNA polymerase and terminal transferase activities, *Nucleic Acids Res.* 31, 6916–6925

32. Bebenek K., Garcia-Diaz M., Blanco L. and Kunkel T.A., (2003). The frameshift infidelity of human DNA polymerase lambda: implications for function, *J. Biol. Chem.* 278, 34685–34690.

33. Garcia-Diaz M., Bebenek K., Kunkel T.A. and Blanco L., (2001). Identification of an intrinsic 5′-deoxyribose-5-phosphate lyase activity in human DNA polymerase lambda: a possible role in base excision repair. *J. Biol. Chem.* 276, 34659–34663

34. Matsumoto Y. and Kim K., (1995). Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* 269, 699–702

35. Li S., Wilkinson MF. (1997). Site-directed mutagenesis: a two-step method using PCR and DpnI. *Biotechniques*. 23(4):588-90.

36. Sheffield P, Garrard S, Derewenda Z. (1999). Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr. Purif.* 15(1):34-9.

37. Sambrook, j., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor.

38. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR* 6, 135-140

39. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* 6, 277-293

40. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* 278, 313-352

41. Protein–Protein Interactions: Methods and Protocols. Methods in Molecular Biology, vol. 261

42. Greenfield, N. J. and Fowler, V. M. (2002). Tropomyosin requires an intact N-terminal coiled coil to interact with tropomodulin. *Biophys. J.* 82, 2580–2591.

43. Kunkel, T. A. (2004). DNA Replication Fidelity. *J. Biol. Chem.*, 279 (17), 16895-16898

44. Tseng, H. M., and Tomkinson, A. E. (2002). A Physical and Functional Interaction between Yeast Pol4 and Dnl4-Lif1 Links DNA Synthesis and Ligation in Nonhomologous End Joining *J. Biol. Chem.* 277, 45630–45637

45. Wilson, T. E., and Lieber, M. R. (1999). Efficient Processing of DNA Ends during Yeast Nonhomologous End Joining. *J. Biol. Chem.* 274, 23599–23609

46. Ma et al., 2004 Y. Ma, H. Lu, B. Tippin, M.F. Goodman, N. Shimazaki, O. Koiwai, C.-L. Hsieh, K. Schwarz and M.R. Lieber, (2004). A Biochemically Defined System for Mammalian Nonhomologous DNA End Joining *Mol. Cell* 16, . 701–713.

47. Callebaut, I. and Mornon, J.P., (1997). From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. *FEBS Lett.* 400. 25–30.

48. Frouin I., Toueille M., Ferrari E., Shevelev I., and Hübscher U., (2005). Phosphorylation of human DNA polymerase λ by the cyclin-dependent kinase Cdk2/cyclin A complex is modulated by its association with proliferating cell nuclear antigen. *Nucleic Acids Res.*; 33(16): 5354–5361.

49. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins, *FASEB J. 11*, 68-76.

50. Beard, W. A., Shock, D. D., Vande Berg, B. J., and Wilson, S. H. (2002). Efficiency of Correct Nucleotide Insertion Governs DNA Polymerase Fidelity *J. Biol. Chem.* 277, 47393–47398

51. Kool, E. T. (2002). Actice site tightness and substrate fit in DNA replication. Annu. Rev. Biochem. 71, 191–219

52. Goodman, M. F. (1997). Hydrogen bonding revisited: Geometric selection as a principal determinant of DNA replication fidelity *Proc. Natl. Acad. Sci. U. S. A.* 94, 10493–10495

53. Petruska, J., Sowers, L. C., and Goodman, M. F. (1986). Comparison of Nucleotide Interactions in Water, Proteins, and Vacuum: Model for DNA Polymerase Fidelity Proc. Natl. Acad. Sci. U. S. A. 83, 1559–1562

54. Takeuchi, T., Ishidoh, T., Iijima, H., Kuriyama, I., Shimazaki, N., Koiwai, O., Kuramochi, K., Kobayashi, S., Sugawara, F., Sakaguchi, K., Yoshida, H.i & Mizushina, Y. (2006) Structural relationship of curcumin derivatives binding to the BRCT domain of human DNA polymerase λ. *Genes to Cells* 11 (3), 223-235.

55. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. *J Mol Graph Model* 19, 26-59

56. Wright, P. E., and Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293, 321-331

57. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F., and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.*J Mol Biol* 337, 635-645

58. Toyoshima, H., and Hunter, T. (1994). p27, a novel inhibitor of G1 cyclin-Cdk protein kinase activity, is related to p21. *Cell* 78, 67-74

59. Bienkiewicz, E. A., Adkins, J. N., and Lumb, K. J. (2002). Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry* 41, 752-759

60. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41, 415-427

61. Minezaki, Y., Homma, K., Kinjo, A. R., and Nishikawa, K. (2006). Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *J Mol Biol* 359, 1137-1149

62. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic, Z., and Dunker, A. K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323, 573-584

63. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005). Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit* 18, 343-384

64. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry* 41, 6573-6582

65. Dyson, H. J., and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6, 197-208

66. Miller J B., (2000). "Why NMR is attracting drug designers". Today's Chemist at Work. 9 (1) : 44-49

67. Lacy, E. R., Filippov, I., Lewis, W. S., Otieno, S., Xiao, L., Weiss, S., Hengst, L., and Kriwacki, R. W. (2004). p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat Struct Mol Biol* 11, 358-364

68. Venot C., Maratrat M., Dureuil C., Conseiller E., Bracco L. and Debussche L. (1998). The requirement for the p53 proline-rich functional domain for mediation of apoptosis is correlated with specific *PIG3* gene transactivation and with transcriptional repression. *EMBO J.* 17,4668–4679.

69. Pawson, T. (1994). SH2 and SH3 domains in signal transduction. *Adv. Cancer Res.* 64:87–110.

70. Pawson, T. (1995). Protein modules and signalling networks. *Nature.*373:573–580

71. Schlessinger, J. (1994). SH2/SH3 signaling proteins. *Curr. Opin.Genet. Dev.* 4:25–30

72. Creamer T.P. (1998). Left-handed polyproline II helix formation is (very) locally driven. *Proteins* 33:218-226.

73. Sigler, P.B. (1988). Acid blobs and negative noodles. *Nature.* 333:210–212.

74. Mahoney, N.M., Janmey, P.A., Almo, S.C. Structure of the profilin-poly-L-proline complex involved in morphogenesis and cytoskeletal regulation. (1997). *Nat. Struct. Biol.* 4:953–960

75. Williamson, M.P. (1994). The structure and function of proline-rich regions in proteins. *Biochem. J.* 297:249–260.

76. Stapley B.J. and Creamer T.P. (1999). A survey of left-handed polyproline II Helices*Protein Science* 8:587-595.

77. Siligardi, G., Drake, A.F. (1995).The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. *Biopolymers* 37:281–292.

78. Kapust, R. B., Tözsér, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., and Waugh, D. S. (2001). Tobacco etch virus protease: Mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Prot. Eng.* 14: 993-1000.

79. Lucast, L. J., Batey, R. T., and Doudna, J. A. (2001). Large-scale purification of a stable form of recombinant tobacco etch virus protease. Biotechniques **30**: 544-550.

80. David S. Waugh, Ph.D. Chief, Protein Engineering Section Macromolecular Crystallography Laboratory, National Cancer Institute. *TEV Protease FAQ.* <http://mcl1.ncifcrf.gov/waugh_tech/faq/tev.pdf>