Developing as assay to screen inhibitors for various ATP-dependent ligases.

Loveleen Kaur

School of Life Sciences

Developing an Assay to Screen Inhibitors for various ATP-dependent Ligases

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy
University of Westminster

Loveleen Kaur

December 2009

ACKNOWLEDGEMENTS

I am heartily thankful to my supervisor, Mark Odell, for his encouragement, guidance and support from my initial enrolment through to the completion of my thesis. I would also like to express my appreciation to Prof. Taj, Diluka and Suresh, who have assisted and encouraged me in the completion of this thesis. Thanks are also due to technical staff at School of Biosciences for their generous support.

I would like to express my gratitude to Claire Brooke for her assistance in SPR analysis. Anatoily, Anu and Julien are thanked for their assistance with all types of technical problems - at all times.

Finally, I am grateful to my family for their support, patience and encouragement.

ABSTRACT

DNA ligases (EC.6.5.1.1) are key enzymes that catalyze the formation of phosphodiester bonds at single-stranded or double-stranded breaks between adjacent 5'-PO_4 and 3'-OH groups of DNA. These enzymes are essential guardians of genomic integrity and have recently been drawing a lot of attention as novel therapeutic targets in anti-bacterial and anti-cancer therapies.

A novel, non-electrophoretic assay method, based on the strength of interaction of the oligonucleotides with Q-sepharose (a strong anion exchanger), was developed to screen inhibitors of DNA ligases from natural product pools as well as chemical libraries. The binding affinities to Q-sepharose resin of a nicked DNA substrate (created from a 30-mer hairpin oligonucleotide and complementary ^{32}P-labelled 6-mer oligonucleotide) and its sealed, ligated product (36-mer) were determined. Initial optimisation studies were performed with T4 DNA ligase, PBCV-1 DNA ligase and a catalytically active form of human DNA ligase I in the presence of doxorubicin (inhibitor of ATP-dependent ligases). These results when analysed in parallel between the conventional electrophoretic assay and the labelled nick-sealing assay showed that the newly developed assay is a reliable non-electrophoretic method in identifying potent DNA ligase inhibitors. The feasibility of the assay was tested in screening a collection of whole cell mass extracts, obtained from a natural product library from *Basidiomycetes*, in 96-well format.

A novel single DNA ligase was identified, expressed and characterised from *Trichomonas vaginalis* (TV), a pathogenic protozoan parasite. Protein sequence analysis of TV DNA ligase indicates a strong sequence similarity to DNA ligase I homologues. The activity of recombinant TV DNA ligase I (TVlig) was investigated using protein expressed in *E.coli* cells. The TVlig gene product is ~76 kDa and showed optimal ligation activity on a nicked DNA substrate at pH 7-8 in the presence of 1 mM ATP and (8- 20) mM $MgCl_2$ at 30-38^{o}C. The inhibition of the only DNA ligase present in *T. vaginalis* might suggest for a rational approach to stop replication and hence propagation of the parasite during infection.

List of Figures:

List of Tables:

List of Abbreviations

Amp	Ampicillin
AppDNA	DNA-adenylate intermediate
ATP	Adenosine-5'-triphosphate
Bam HI	Restriction enzyme from the bacterium *Bacillus amyloliquefaciens H*
bp	Base pair(s)
BSA	Bovine serum albumin
BER	Base excision repair
°C	Degrees centigrade
cDNA	Complementary DNA
Ci	Micro-Curie (unit of measurement of radioactivity)
cpm	Counts per minute (radioactivity)
(d)ATP	(Deoxy)adenosine-5'-triphosphate
CTP	(Deoxy)cytidine-5'-triphosphate
GTP	Guanosine-5'-triphosphate
DNA	Deoxyribonucleic acid
dNTP	N = A (adenosine) + G (guanosine) + C (cytidine) + T (thymidine)
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EpA	Ligase-adenylate intermediate
EDTA	Ethylendiamintetraacetic acid
g	Gravity force (9.81 m/s2)
His	Histidine
HR	Homologous recombination
HuLigI	Human DNA ligase I
Kan	Kanamycin
kDa	Kilodalton (= 1000 Dalton)
LB medium	Luria Bertani medium
MHEJ	Micro homology end joining
mRNA	Messenger RNA
NAD	Nicotinamide adenine dinucleotide
Nde I	Restriction enzyme from the bacterium *Neisseria denitrificans*
Ni-NTA	Nickel-nitrilotriacetic acid (Ni2+ ions are chelated by four positions)
NHEJ	Non-homologous end joining
OD_{595}	Optical density measured at a wavelength of 595 nm
[γ^{32}P]	Labelled with the γ-radiation-emitting ^{32}phosphorus isotope
PBCV-1	*Paramecium bursaria Chlorella* virus 1
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
Pfu-polymerase	DNA polymerase from the bacterium *Pyrococcus furiosus*
PMSF	Phenylmethylsulfonyl fluoride
Pol	DNA polymerase

Pflig	*P.falciparum* DNA ligase
PPi	Pyrophosphate
RF-C	Replication factor C
RNA	Ribonucleic acid
RPA	Replication protein A
SDS	Sodiumdodecyl sulphate
SDS-PAGE	Sodiumdodecyl sulphate-polyacrylamide gel electrophoresis
ssDNA	Single-stranded DNA
TAE	Tris-acetate EDTA (40 mM Tris-acetate (pH 8.0), 1 mM EDTA)
TBE	Tris-borate EDTA (45 mM Tris-borate (pH 8.0), 1 mM EDTA)
TBS	Tris-buffered saline (10 mM Tris-HCl (pH 7.5), 150 mM NaCl)
TCA	Trichloroacetic acid
Taq polymerase	DNA polymerase from *Thermus aquaticus*
Tris	Tris-(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
TV	*Trichomonas vaginalis*
TVlig	TV DNA ligase
U	Units of enzyme activity
v/v	Volume per volume
w/v	Weight per volume
Xho I	Restriction enzyme from the bacterium *Xanthomonas holci*

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

CHAPTER 1: INTRODUCTION

In the 1960s, biologists were seeking to identify the enzymatic activity that was responsible for both rejoining breaks caused by ultraviolet radiation and completing the sealing of DNA that had moved as a result of recombination. Various groups, independently and simultaneously, identified and purified the enzymatic activity which formed phosphodiester bonds between DNA ends held by hydrogen-bond pairing in double-stranded configuration and named it as DNA ligase (Becker *et al.*, 1967; Gellert, 1967; Zimmerman *et al.*, 1967).

DNA ligases are ubiquitous cellular proteins that catalyse the formation of phosphodiester bonds between nucleotides at potentially damaging nicks or breaks in DNA (Figure 1.1; Weiss & Richardson, 1967; Modrich *et al.*, 1973; Lehman, 1974; Shuman, 1996; Timson *et al.*, 2000). They are essential for the accurate copying and maintenance of DNA by replication and repair. A nick can be described as a juxtaposed 3'hydroxyl (3'-OH) terminated DNA strand adjacent to a 5' phosphate (5'-PO_4) terminated DNA strand both annealed to a complementary templating DNA strand. Nicks or strand-breaks in DNA occur as a result of DNA replication within undamaged cells (Lehman, 1974).

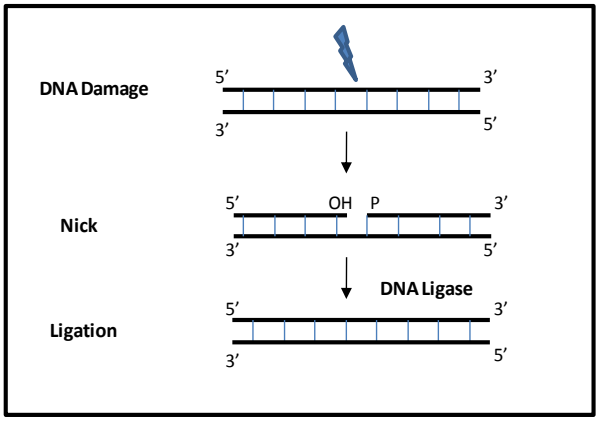


Figure 1.1: Repair of nick by DNA ligase

DNA damage caused by damaging agents, such as ionising radiations, leaves a nick at DNA strand. DNA ligases catalyse the synthesis of phosphodiester bond at the site of these nicks and restore the continuity of the repaired DNA strand.

DNA ligases catalyse the joining of Okazaki fragments, generated by discontinuous DNA synthesis on the lagging strand during such genome replication (Ogawa &

Okazaki, 1980). Ligases are also essential for the ligation step in variable (V) - diversity (D) – joining (J) recombination [V(D)J], a mechanism to generate antigen receptor diversity (Grawunder *et al.,* 1998; Verkaik *et al.,* 2002). They have also found widespread use as a tool for *in vitro* DNA manipulation, cloning techniques and ligase based diagnostic assays for mutation detection (Subryamanya *et al.,* 1996; Cao, 2001).

DNA ligases belong to a family of covalent nucleotidyl transferases that includes RNA ligases and GTP-dependent mRNA capping enzymes (Shuman & Schwer, 1995). The nucleotidyl transferase (NTases) is a superfamily of phosphotransferase enzymes are defined by a set of conserved sequence motifs and act through a lysyl-N-NMP intermediate (Shuman & Lima, 2004).

1.1 Classes of DNA ligases

Despite the occurrence of DNA ligases in all organisms, they show a wide diversity of molecular sizes, amino acid sequences and properties (Doherty & Suh, 2000; Timson *et al.,* 2000; Ellenberger & Tomkinson, 2008). They are grouped into two families according to their high-energy cofactor requirements for either ATP or NAD^+. ATP-dependent DNA ligases (EC 6.5.1.1) are more widely distributed in eukaryotic and archaeal cells but are also encoded by certain eukaryotic viruses, bacteriophage and eubacteria (Cheng & Shuman, 1997; Timson *et al.,* 2000; Martin & MacNeil, 2002). NAD^+-dependent DNA ligases (EC 6.5.1.2) are encoded predominantly by eubacteria (Wilkinson *et al.,* 2001) but also by certain eukaryotic viruses, such as the entomopoxviruses (Sriskanda *et al.,* 2001) and mimiviruses (Benarroch & Shuman, 2006).

The eukaryotic ATP-dependent DNA ligases vary greatly in distribution, size and their ability to ligate a range of nucleic acid substrates (Ellenberger & Tomkinson, 2008). For example, DNA ligase I appears to be conserved in all eukaryotes, as an orthologue has been identified and characterised in a diverse range of organisms, from yeast and plants to mammals (Martin & MacNeill, 2002). Mammals, however, encode for four distinct ATP dependent DNA ligases (I, IIIα, IIIβ and IV; DNA ligase II is now classified as DNA ligase IIIα) whereas lower organisms such as *Saccharomyces*

cerevisiae encode for only two DNA ligases - Cdc9, a homologue of human DNA ligase I (HuLigI) and DNL4, a homologue of DNA ligase IV (Tomkinson & Mackay, 1998; Tomkinson & Levin, 2005). Mammalian DNA ligases are large polypeptides (>900 amino acid residues) whereas the virus encoded ATP-dependent DNA ligases, for example from bacteriophage T7 (359 amino acid residues) and *Paramecium bursaria Chlorella* virus 1 (PBCV-1, 298 amino acid residues) are much smaller (Subramanya *et al.,* 1996; Ho *et al.,* 1997; Tomkinson & Levin, 2005). An ATP dependent DNA ligase I has also been identified in the parasitic protozoan *Plasmodium falciparum* (912 amino acid residues; Buguliskis *et al.,* 2007) and *Trichomonas vaginalis* (679 amino acid residues; discussed in detail in Chapter 3).

The presumption that bacteria encode only NAD^+-dependent DNA ligases (referred to as LigA) was overturned by the demonstration of an ATP-dependent DNA ligase (268 amino acid residues) in the respiratory pathogen *Haemophilus influenzae* (Cheng & Shuman, 1997). ATP-dependent DNA ligase homologues coexist with NAD^+-dependent enzymes in several other bacterial species, including major human pathogens such as *Neisseria meningitidis*, *Vibrio cholerae* and *Mycobacterium tuberculosis* (Wilkinson & Bowater, 2001; Gong *et al.,* 2004).

Previously it was suggested that archaeal ligases exclusively utilize ATP as their nucleotide substrate (Keppetipola & Shuman, 2005) but this assumption has been more recently contested by the observations that the cofactor requirement of archaeal DNA ligases is not solely limited to ATP. Table 1.1 lists diverse nucleotide substrate specificities of various archaeal DNA ligases encoded by selection of organisms. *Haloferax volcanii (H. volcanii),* a halophilic archaeon, is found to have both ATP and NAD^+-dependent DNA ligases (Zhao *et al.,* 2006). Phylogenetic analysis suggested that the ligase was acquired by lateral gene transfer from eubacteria and may provide additional ligase activity under conditions of high genotoxic stress (Poidevin & MacNeill, 2006). Individually neither ATP nor NAD^+ enzyme is essential for cell viability, but the deletion of both ligases in *H. volcanii* is lethal, indicating that the enzymes share an essential function (Zhao *et al.,* 2006). The recently identified DNA ligase from *Sulfophobococcus zilligii* is reported to utilise any of the three nucleotide substrates; ATP, ADP and GTP (Sun *et al.,* 2008). The undifferentiated

nucleotide specificities of these archaeal ligases might be explained if they recognized only the ADP component common to all three nucleotides (Shuman, 2009).

Table 1.1: Nucleotide substrate specificities of various archaeal ATP-dependent DNA ligases

Archaeon	Nucleotide specificity
Acidithiolbacillus ferroxidans	ATP
Aeropyrum permix	ATP, ADP
Ferroplasma acidophilum	ATP, NAD^+
Pyrococcus horikoshii	ATP, ADP
Staphylothermus marinus	ATP, ADP
Thermplasma acidophilum	ATP, NAD^+

(Information compiled from Keppetipola & Shuman, 2005; Ferrer *et al.,* 2008; Shuman *et al.,* 2009)

1.2 Reaction Mechanism of Nucleotidyltransferase Superfamily

The analysis of the biochemistry of the ligase reaction mechanism has revealed that all DNA ligases catalyse the synthesis of phosphodiester bonds in a very similar manner (Lee *et al.,* 2000). The enzyme mechanism (Figure 1.2) involves three main steps (Lehman, 1974). The initial step involves the nucleophilic attack on the α-phosphate of ATP (or adenyl-phosphate of NAD^+) by the active site lysine residue in motif I (Figure 1.3; Gumport & Lehman, 1971; Timson *et al.,* 2000). Pyrophosphate (PP_i) is released and a covalent ligase-adenylate intermediate (EpA) is formed in which AMP is linked to the lysine via a phosphoamide bond (Figure 1.2, panel D; Gumport & Lehman, 1971; Lehman, 1974). In 2002, Cherepanov & Vries demonstrated that step 1 in the reaction catalysed by T4 DNA ligase uses dimagnesium ATP-Mg_2 for transfer of the adenyl moiety to the ligase. In the second step, AMP is transferred to the 5'-phosphate terminated nicked DNA strand via a pyrophosphate bond, forming a DNA-adenylate intermediate (AppDNA; Olivera *et al.,* 1968; Yang & Chan, 1992). The final step involves the ligase catalysed attack by the 3'-OH group of the juxtaposed DNA strand on the adenylated strand of the

AppDNA resulting in the joining of the two polynucleotides and release of AMP (Shuman, 1996). One ATP molecule is utilised per nick-sealing event (Crut *et al.*, 2008). In NAD^+-dependent DNA ligases the enzyme-AMP adenylate is formed by the breakdown of NAD^+ and the release of NMN rather than PP_i (Martin & MacNeill, 2002).

RNA ligases join 3'-OH and 5'-PO_4 terminated RNA through the same series of catalytic steps as those employed by DNA ligases (Shuman & Lima, 2004). These steps involve the formation of a step 1 covalent ligase-adenylate and release of pyrophosphate followed by the formation of the step two RNA-adenylate intermediate and step three catalysis of strand joining and AMP release (Figure 1.2, panel B). The RNA ligases can be divided into two distinct families: RNA ligase 1 (Rnl1) and RNA ligase 2 (Rnl2). The Rnl1 family includes enzymes such as bacteriophage T4 RNA ligase 1 (the first RNA ligase to be identified; Silber *et al.*, 1972), and the tRNA ligases of plants and fungi. The Rnl2 family includes enzymes such as T4 RNA ligase 2 and the RNA editing ligases (RELs) found in the kinetoplast of the protozoa *Typanosoma* and *Leishmania* (Ho & Shuman, 2002). Rnl1s function in the repair of single strand RNA breaks in tRNA anticodon loops or breaks in mRNAs introduced by site-specific RNA endonucleases during the RNA-based antiviral response (Nandakumar *et al.*, 2006). Rnl2s play an important role in sealing nicks in duplex RNA, particularly during mRNA editing by RELs (Nandakumar *et al.*, 2006). Rnl1 and Rnl2 have very little sequence similarity outside of the conserved nucleotidyl transferase core (Ho & Shuman, 2002; Pascal, 2008).

RNA capping enzymes (previously termed RNA guanyltransferases) transfer GMP derived from GTP onto the 5'-diphosphate terminus of triphosphatase processed mRNAs to form a G(5')ppp(5')RNA cap structure. This plays an important role in the processing, protection and transporting of mRNAs to the ribosome (Shuman & Schwer, 1995; Shuman & Lima, 2004). The capping mechanism involves three separate reactions, catalysed either by three separate enzymes or by modular combinations between the various components creating either a single enzyme or two enzymes depending on the organism (Hakansson & Wigley, 1998, Shuman & Lima,

2004). The first step entails the removal of the terminal 5'-PO_4 from mRNA (pppRNA) by an RNA triphosphatase.

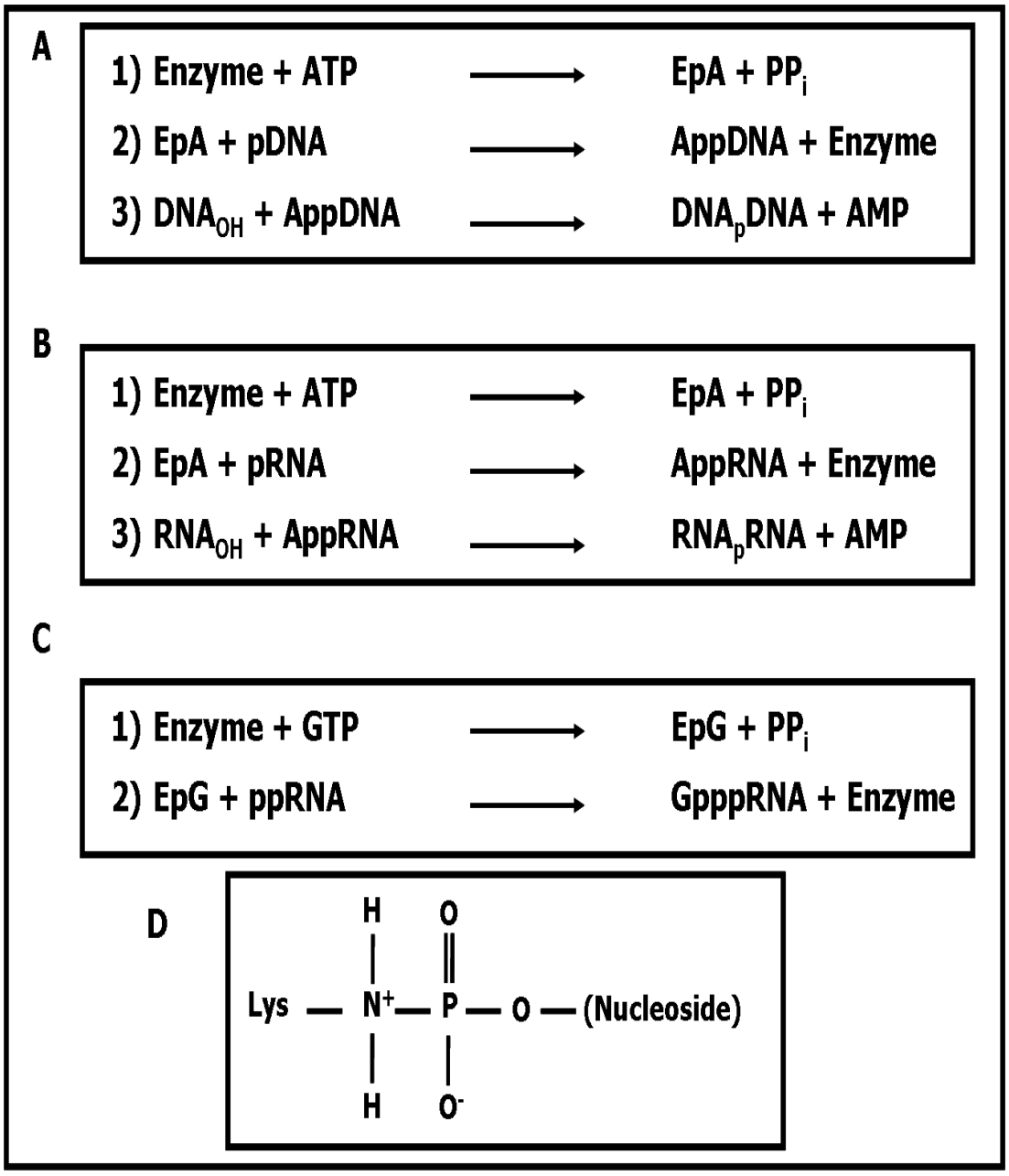


Figure 1.2: The conserved reaction mechanism of covalent nucleotide transferase enzymes

The reaction mechanism of DNA ligases (panel A), RNA ligases (panel B) and capping enzymes (panel C) is shown. Step one involves the nucleophilic attack by the active site lysine on the α-phosphate of the nucleotide substrate and the formation of a covalent enzyme adenylate linked by a phosphoamide bond. The chemical structure of the phosphoamide bond forming the covalent enzyme intermediate is shown in panel D. The abbreviated terms are described in the text (Adapted from Doherty, 1999; Doherty & Suh, 2000).

The second reaction, catalysed by RNA capping enzymes, involves the linkage of GMP to the 5' diphosphate terminated RNA via an intermediate step in which the GMP is covalently linked to the capping enzyme active site lysine (located in motif I which is also conserved in RNA and DNA ligases), through a phosphoamide bond. This step most resembles the catalytic mechanism of DNA and RNA ligases (Figure 1.2, panel C; Shuman & Hurwitz, 1981; Hakansson & Wigley, 1998). The third step involves a final enzyme activity, RNA guanine-7-methyltransferase, which catalyses the methylation at N7 of the guanine base (Shuman & Hurwitz, 1981; Hakansson & Wigley, 1998).

Accumulation of nicked-DNA adenylates (AppDNA) are observed during nick sealing by various ATP dependent bacterial DNA ligases, reflecting the dissociation of DNA ligase before step 3 (Gong *et al.*, 2004; Zhu & Shuman, 2007). For most ligases the ligation reaction is functionally unidirectional, i.e. it proceeds forward from step 1 to 2 and 3; for example, when the ligase apoenzyme is incubated with a pre-adenylated nicked DNA substrate, the reaction is predisposed towards step 3 and phosphodiester formation (Odell & Shuman, 1999; Nandakumar *et al.*, 2006; Crut *et al.*, 2008). Although Modrich *et al.* (1972) and Sekiguchi & Shuman (1997) have shown that steps in the nick ligation pathway may be reversible when forced. Modrich *et al.* (1972) demonstrated this with the incubation of supercoiled closed circular DNA with excess *E. coli* DNA ligase in the presence of AMP. This yielded DNA molecules with single-strand breaks and covalently-closed circular DNA. Supercoiled DNA was also converted to slower migrating forms by Vaccinia virus DNA ligase. The AMP and magnesium-dependent relaxation reaction required excess amounts of ligase in relation to the input DNA substrate, ratios which are obviously not present in the cell under physiological conditions (Sekiguchi & Shuman, 1997).

The adenylated form of DNA ligase is stable and enhances its nick sensing capability (Sriskanda & Shuman, 1998). Occupation of the nucleotide binding site by AMP, is crucial to nick recognition as the enzyme undergoes a conformational change which exposes the DNA binding surface of domain one (see section 1.3), allowing the 5'-PO_4 of the nicked DNA to make direct contact with the adenylation site of the enzyme (Sekiguchi & Shuman, 1997; Odell *et al.*, 2000). Mutations in the ligase active site

that prevent the binding of ATP, particularly mutations of the active site lysine and arginine in motif I; have been found to eliminate the binding of the enzyme to nicked DNA (Sekiguchi & Shuman, 1997). This suggests that only the catalytically competent form (EpA) is capable of high affinity DNA binding, although enzymes with mutations in the active site lysine are able to turn-over and ligate DNA when incubated with a pre-adenylated DNA substrate (Sekiguchi & Shuman, 1997; Sriskanda & Shuman, 1998).

1.3 The overall domain structure of DNA ligases

Sequence alignments of a number of enzymes that catalyse nucleotidyl transfers including all DNA ligases (ATP and NAD^+-dependent), RNA ligases (Rnl1 and Rnl2 families) as well as RNA capping enzymes, which share homologous reaction chemistry, have shown the presence of several regions of amino acid homology with six motifs of conserved residues with consistent spacing (Shuman, 1996; Shuman & Lima, 2004). These conserved motifs were previously numbered I, III, IIIa, IV and V (ATP-dependent DNA ligase and mRNA capping enzymes contain an addition homologous motif, VI), until recently when Shuman and co-workers proposed the existence of another conserved motif (Ia), downstream of motif I and first suggested to be such a motif in 1997 (Figure 1.3, panel A; Odell, PhD thesis, 1997; Nair *et al.*, 2007; Shuman, 2009). These motifs make up the common tertiary structure of the catalytic core and nucleotide binding pocket of DNA ligases (Figure 1.3, panel A; Subramanya *et al.*, 1996; Odell *et al.*, 2000).

The numbering of the motifs corresponds to the position within the primary sequence with motif I being the most proximal to the protein N-terminus and motif VI, to the C-terminus. Mutations of the conserved residues that make up these motifs have been shown to be critical for activity in nucleotidyltransferase enzymes (Subramanya *et al.*, 1996). There is virtually no sequence similarity between ATP and NAD^+-dependent DNA ligases except in this central DNA ligase catalytic core of motifs I through V (Timson *et al.*, 2000; Martin & MacNeill, 2002) as NAD^+-dependent ligases lack a recognisable counter-part of motif VI (Sriskanda *et al.*, 2001).

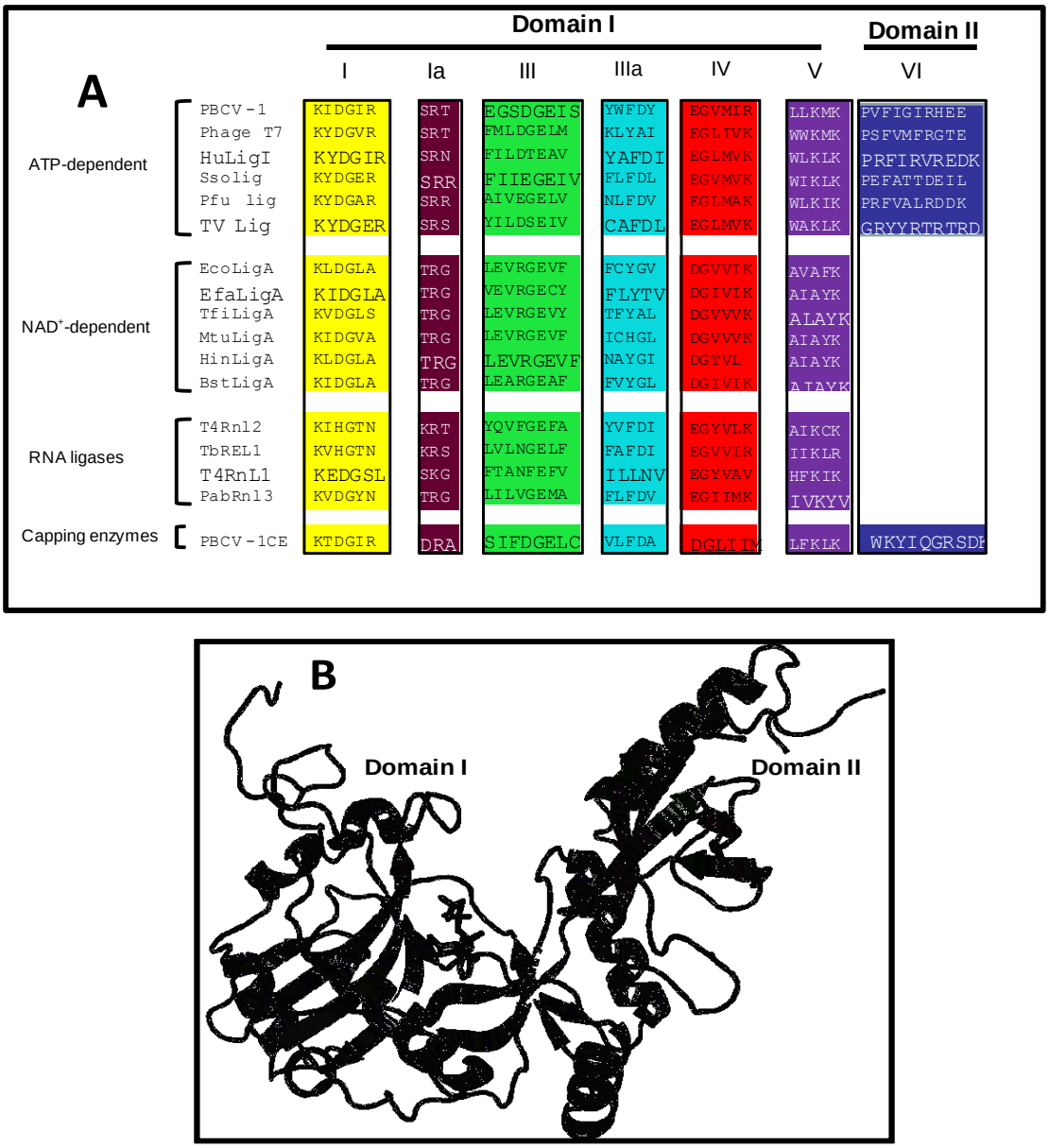


Figure 1.3: Sequence alignment of conserved sequence motifs and domain organisation in nucleotidyl transferases.

Sequence motifs (designated I, Ia, III, IIIa, IV, V and VI; highlighted in yellow, brown, green, cyan, red, magenta and blue respectively) conserved in DNA ligases, RNA ligases and RNA capping enzymes are shown. The alignment includes the ATP-dependent DNA ligases [encoded by: *Paramecium bursaria Chlorella* virus 1 (PBCV-1; NP_048900), Bacteriophage T7 (T7; P00969), human DNA ligase I (HuLig1; NP_000225), *S. Solfataricus* (Ssolig; AAK40535), *Pyrococcus furiosus* (Pfulig; NP_579364), *Trichomonas vaginalis* DNA ligase I (TV lig; XP_001581589)], NAD^+-dependent DNA ligases [encoded by: *E. coli* (*Eco*LigA; P15042), *Enterococcus faecalis (EfaLigA;* NP_814472*)*, *Thermus filiformis* (TfiLigA; Q9ZHI0), *Mycobacterium tuberculosis* (MtuLigA; NP_217530), *Haemophilus influenzae* (HinLigA; NP_439257), *Bacillus stearothermophilus* (BstLigA; NP_622271)], ATP-dependent RNA ligases [encoded by T4 (T4 Rnl2; NP_049790.), *Trypanosoma brucei* (TbREL1; XP_001219001), T4 Rnl1 (NP_049839), *Pyrococcus abyssi* (Pab Rnl3; CAB49026)] and GTP-dependent RNA capping enzymes from PBCV-1 (NP_048451). Panel B shows domain organisation of Bacteriophage T7 DNA ligase. The larger adenylation domain (AdD) is shown in grey and the smaller OB-fold domain in green. ATP located within the active site is coloured according to element.

The smaller, structurally less complex DNA ligases, those encoded by viruses and bacteriophages, have a minimal two-domain structure. This is highlighted by the atomic structures of the ATP-dependent DNA ligases from bacteriophage T7 (359 residues; Figure 1.3, panel B; Subramanya *et al.,* 1996) and PBCV-1 DNA ligase (298 residues; Odell *et al.,* 2000; Odell *et al.,* 2003). A large N-terminal nucleotide binding domain (adenylation domain; AdD), which binds ATP, and a smaller C-terminal oligonucleotide binding domain (OB-fold) are structurally analogous to the catalytic core of complex multidomain DNA ligases found in bacteria and higher eukaryotes (Figure 1.4; Martin & MacNeill, 2002).

The adenylation domain consists of two, twisted anti-parallel β-sheets and six α-helices (Figure 1.3, panel B; shown in grey). The adenylate binding pocket composed of the six peptide motifs I, Ia, III, IIIa, IV and V lies in the adenylation domain (Subramanya *et al.*, 1996; Nair *et al.*, 2007). The active site lysine residue to which AMP becomes covalently linked is located within conserved motif I (KxDGxR) (Cong & Shuman, 1995; Odell *et al.,* 2000). The OB-fold domain, consisting of a single five-stranded anti-parallel β barrel and an α helix, assists in the formation of the ligase-AMP intermediate during step 1 of ligation reaction (Figure 1.3, panel B; shown in green; Sriskanda & Shuman, 1998). The OB-fold domain binds across the minor groove of double-stranded DNA and interacts with the DNA strand adjacent to the 5' phosphate end of the nick (Nair *et al.,* 2007). In ATP-dependent ligases, the OB-fold domain closes over the adenylation domain during step I of the ligation reaction to contribute the residues of conserved residues of motif VI required for AMP transfer to the active site lysine (Odell *et al.,* 2000; Pascal *et al.,* 2006). RNA capping enzymes use a similar mechanism and two domain structure to form a capping enzyme-GMP intermediate, before transferring GMP to the 5' end of mRNA (Pascal, 2008).

After formation of the ligase-AMP reaction intermediate, the OB fold domain functions as a DNA-binding component during steps 2 and 3. For these steps the OB fold turns to orient DNA-binding residues for interaction with the DNA substrate and the motif VI residues are exposed on the protein surface away from the active site

(Nair *et al.*, 2007; Pascal, 2008). This conformational flexibility of the ligase is accommodated by an extended polypeptide linker, containing motif V, which serves as a bridging segment joining the adenylation and the OB fold domains and also assists in loading the enzyme onto DNA (Nair *et al.*, 2007).

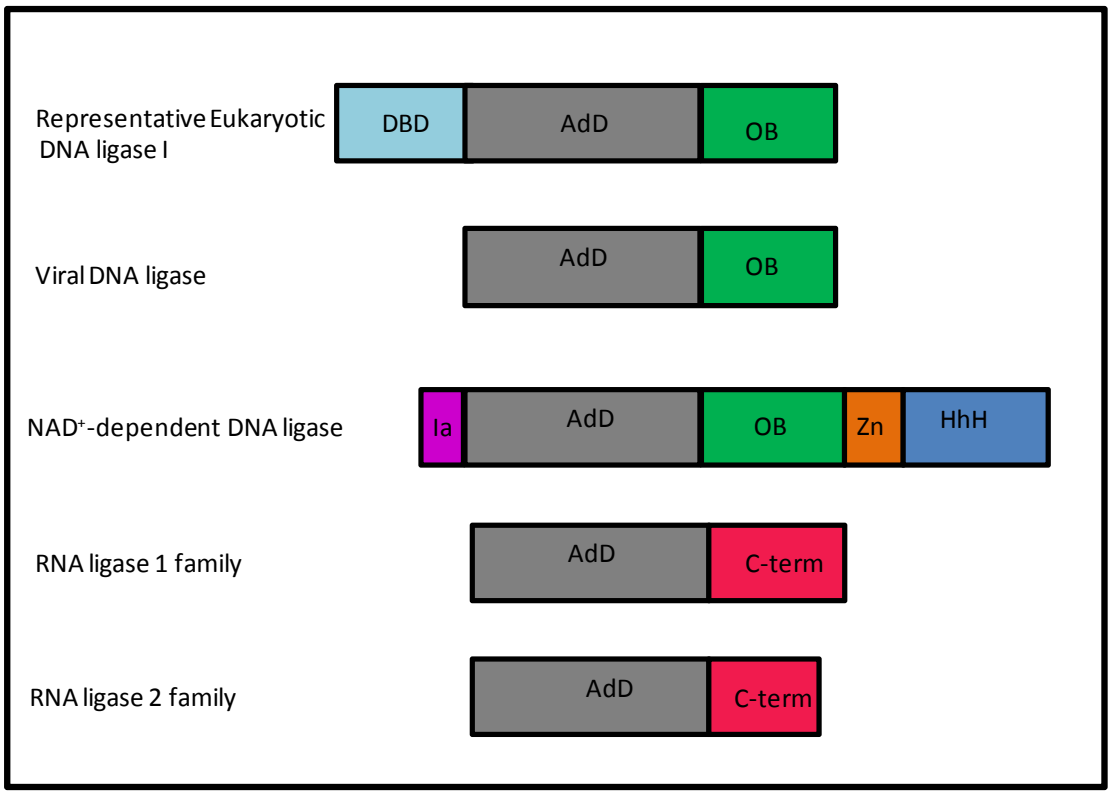


Figure 1.4: Domain architecture of DNA and RNA ligases

The adenylation domain, consisting of the conserved motifs (I, Ia, III, IIIa, IV and V) which catalyse the three step end-joining reaction, is conserved in all polynucleotide ligases. The AdD and OB fold domains represent the catalytic core of ATP and NAD^+-dependent DNA ligases. Viral DNA ligases are much smaller than cellular DNA ligases and are devoid of any additional N or C-terminal domains. Accessory domains extending from the adenylation domain promote the ligation reaction or provide special substrate binding properties for specific ligases. RNA ligase 1 (Rnl1) and RNA ligase 2 (Rnl2) have a conserved adenylation domain and unique C-terminal domain which is distinct in two families and is unrelated to the C-terminal OB-fold domains of DNA ligases and RNA capping enzymes. Abbreviations: DNA binding domain (DBD), OB fold (OB) domain, Helix-hairpin-helix (HhH), Domain Ia (Ia), Zinc-binding domain (Zn) and C-terminal domains (C-term) (Information taken from Nandakumar *et al.*, 2007; Ellenberger & Tomkinson, 2008).

In addition to the nucleotide binding and OB-fold domains described above ligases of multicellular organisms, such as human DNA ligase I (HuLigI), have a dedicated N-terminal DNA-binding domain (DBD; Pascal *et al.*, 2004). The DBD domain makes protein-protein interactions with the adenylation and the OB-fold domain and allows

ligase I to encircle its DNA substrate and hence contributes a substantial DNA-binding affinity of the enzyme. This domain also serves to distort the DNA structure, widening the minor groove that exposes the ligatable ends of the DNA substrate to conserved residues of the catalytic core (Tomkinson *et al.,* 2006; Pascal, 2008). Pascal *et al.* (2004) also identified two divalent metal binding sites in the active site of the nucleotide-binding domain. The OB-domain in HuLigI also appear to aid substrate selection and hence prevent untimely ligation of Okazaki fragments before the removal of the RNA primer thereby adding to the fidelity of the DNA ligation event (Pascal *et al.,* 2004).

NAD^+-dependent bacterial DNA ligases have a highly conserved adenylation domain and OB-domain but lack the N-terminal DBD found in mammalian DNA ligases. This DBD is substituted for by a C-terminal, helix-hairpin-helix (HhH) domain (Figure 1.4; Lee *et al.,* 2000). The HhH domain binds DNA in the minor groove and serves as a functional mimic of the mammalian DBD (Lee *et al.,* 2000). Domain Ia is unique to NAD^+-dependent DNA ligases**.** It is required for the formation of the ligase-AMP intermediate during step I of the ligation reaction and its deletion leads to a loss in the adenylation capability of the enzyme (Sriskanda *et al.,* 2001; Gajiwala & Pinko, 2004). The Zn domain appears to provide structural support to the ligase architecture in bridging the OB and HhH domains (Nandakumar *et al.*, 2007).

RNA ligases (Rn1 and Rnl2) use only the AdD domain to catalyse the first step of their reaction chemistry, but have a unique C-terminal domain, distinct in the two ligase families, required for AMP-RNA formation for the step 2 reaction (Figure 1.4; Ho *et al.,* 2004).

1.4 Structure of DNA ligases

The progress in structural studies of nucleotidyl transferase enzymes has added to our understanding of the mechanism of action of these enzymes. The structures of nucleotide bound DNA and RNA ligases (Odell *et al.,* 2000; Nandakumar *et al.,* 2006; Nishida *et al.,* 2006; Pascal *et al.,* 2006) and nucleic acid bound DNA and RNA ligases (Pascal *et al.,* 2004; Nair *et al.,* 2007; Nandakumar *et al.,* 2007) have provided molecular insights into different steps of the reactions catalysed by these enzymes.

1.4.1 ATP-dependent DNA ligases

The crystal structure of several ATP- dependent DNA ligases has been solved. These include: bacteriophage T7 DNA ligase in complex with ATP (Figure 1.3, panel B; Subramanya *et al.,* 1996), enzyme-AMP covalent complexes of PBCV-1 DNA ligase (Odell *et al.,* 2000; Odell *et al.,* 2003), N-terminally truncated HuLigI complexed with an AMP-DNA reaction intermediate (Pascal *et al.,* 2004), *Sulfolobus solfataricus* DNA ligase in complex with heterotrimeric Proliferating Cell Nuclear Antigen (PCNA; Pascal *et al.,* 2006) and PBCV-1 DNA ligase-AMP intermediate bound to phosphorylated DNA substrate (Nair *et al.,* 2007). These studies suggest that all DNA ligases except the simplest viral ligases completely encircle their DNA substrates (Ellenberger & Tomkinson, 2008).

The crystal structure of human DNA ligase I (HuLigI; 232-919 amino acid residues; the first structure of a DNA ligase complexed to nucleic acid substrate), revealed the function of a novel N-terminal α-helical extension as a DNA binding domain (DBD, Figure 1.5, panel B; Pascal *et al.,* 2004). The AdD and OB domains interact to form a continuous protein surface that engages the minor groove of DNA (Pascal *et al.,* 2004). This interaction of enzyme with DNA alters the substrate conformation resulting in adoption of an RNA-like A form helix, partially unwinding the DNA duplex and positions the DNA ends in the active site for the joining reaction (Pascal, 2008). The HuLigI-DNA structure revealed that a dramatic conformational change between AdD and OB domain is necessary to proceed the nick joining activity of the enzyme. The DBD interacts with AdD and OB domains establishing a ligase protein clamp which fully encircles the DNA substrate (Figure 1.7, panel A; Pascal *et al.,* 2004).

Recent studies have shown that the DBD of HuLigI also interacts with PCNA and the heterotrimeric cell cycle checkpoint clamp, hRad9-hRad1-hHus1 (9-1-1), where it preferentially binds to the hRad1 subunit (Song *et al.,* 2009). This was previously reported in DNA ligases from *P. furiosus* and *S. solfataricus* using DBD residues for interaction with PCNA sliding clamps (Pascal *et al.,* 2006; Pascal, 2008). The DBD-PCNA interactions stimulate ligation activity in these organisms (Pascal, 2008).

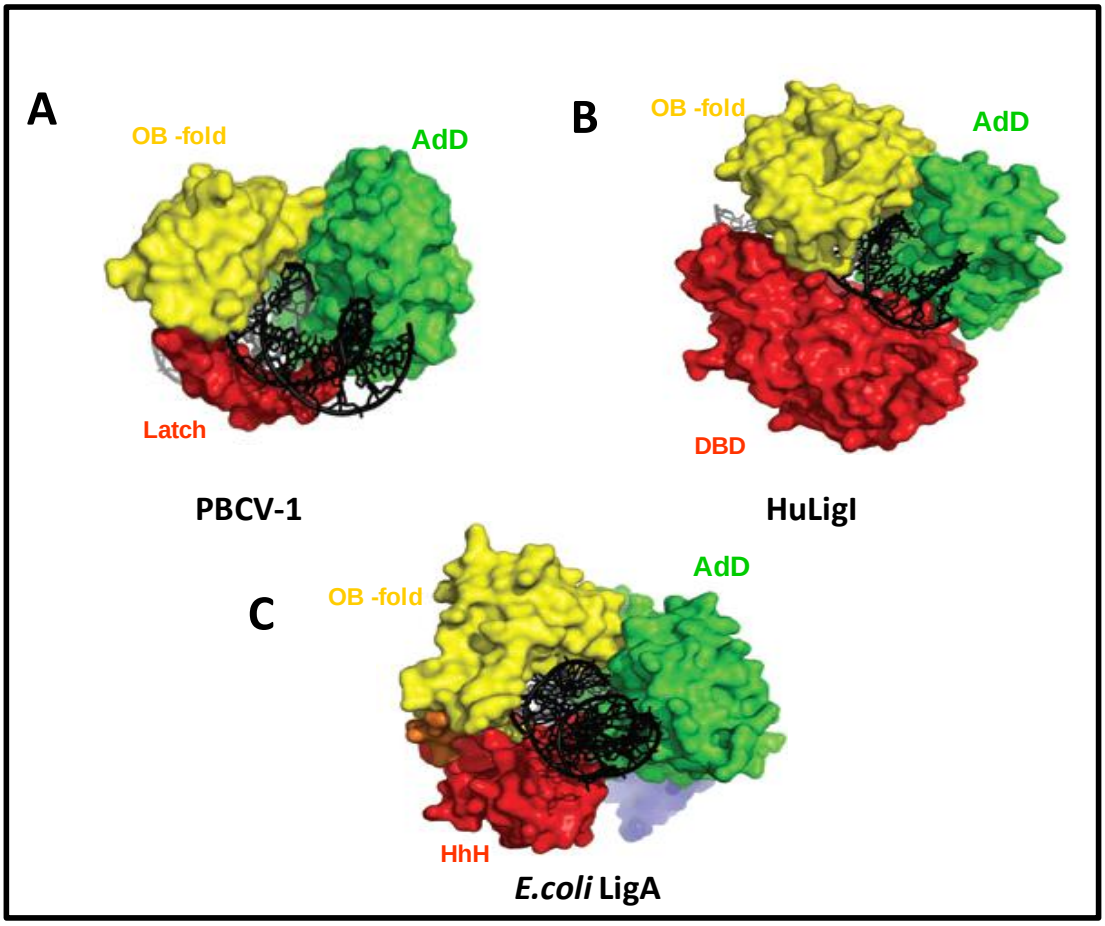


Figure 1.5: Structural configuration of viral, bacterial and mammalian DNA ligases bound to DNA

The DNA ligases from PBCV-1, *Escherichia coli* (LigA) and human ligase I (HuLigI) adopt a similar ring shaped structure in complex with a nicked DNA substrate. A conserved catalytic core consisting of AdB and OB fold domains contacts the DNA substrate during step 2 and 3 of end joining reaction. PBCV-1 has a latch like structure in the OB fold domain, *E. coli* LigA has a C-terminal helix-hairpin-helix (HhH) domain and HuLigI has an N-terminal DNA binding (DBD) domain all of which interact with DNA to complete the ring-shaped structure in conjunction with the catalytic core. (Lee *et al.*, 2000; Pascal *et al.*, 2004; Nair *et al.*, 2007).

The structure of PBCV-1 DNA ligase in complex with nicked DNA reflects the conformation of the enzyme prior to step two catalysis, as it shows the protein engaged at the nick before AMP is transferred to the 5'-PO_4 (i.e. before the AppDNA intermediate is formed; Nair *et al.*, 2007). The enzyme forms a 'C-shaped' protein clamp around the nicked DNA substrate with protein contacts in the AdD, OB-fold domain and a 30 residue lysine-rich surface loop, 'latch' region (Figure 1.5, panel A; Figure 1.6; Nandakumar *et al.*, 2007). Previous structures of the PBCV-1 enzyme-adenylate reveal that in this enzyme conformation the OB-fold domain is positioned away from the AdD to expose the AMP-binding pocket (Figure 1.7, panel A ; Odell *et*

al., 2003). In order to close the clamp around the nicked DNA substrate, 'kissing' contacts are formed between the N-terminal AdD and C-terminal 'latch' region (Figure 1.7, panel B) in the OB-fold domain.

The importance of the β-hairpin loop region was originally intuited from previous PBCV-1 DNA ligase biochemical analyses. The latch was disordered in the absence of DNA and only becomes ordered upon binding to DNA (Figure 1.7, panels A & B, shown in red; Odell *et al.,* 2000; Odell *et al.,* 2003). Replacement of the latch module with a smaller flexible loop (Gly-Ser-Gly-Ser-Gly) resulted in a 10-fold decrease in the specific activity and nick sealing when compared to the wild type enzyme (Odell *et al.,* 2003; Nair *et al.,* 2007). The peptide segment (residues 202-231) which encompasses the disordered loop region was found to be proteolytically sensitive after limited digestion with trypsin and chymotrypsin; although the latch was resistant to proteolysis in the presence of DNA (Odell & Shuman, 1999). In the new structure of the PBCV-1 ligase, upon binding to DNA, following a 63 Å movement of the OB-fold domain, the now-ordered latch encircles the DNA (Figure 1.6, panel A). Structural comparison of HuLigI and PBCV-1 DNA ligase reveals that the AdD and OB domains adopt similar conformations while the α-helical DBD and β-hairpin latch occupy almost the same angular position when docked on the nicked DNA substrate (Shuman, 2009).

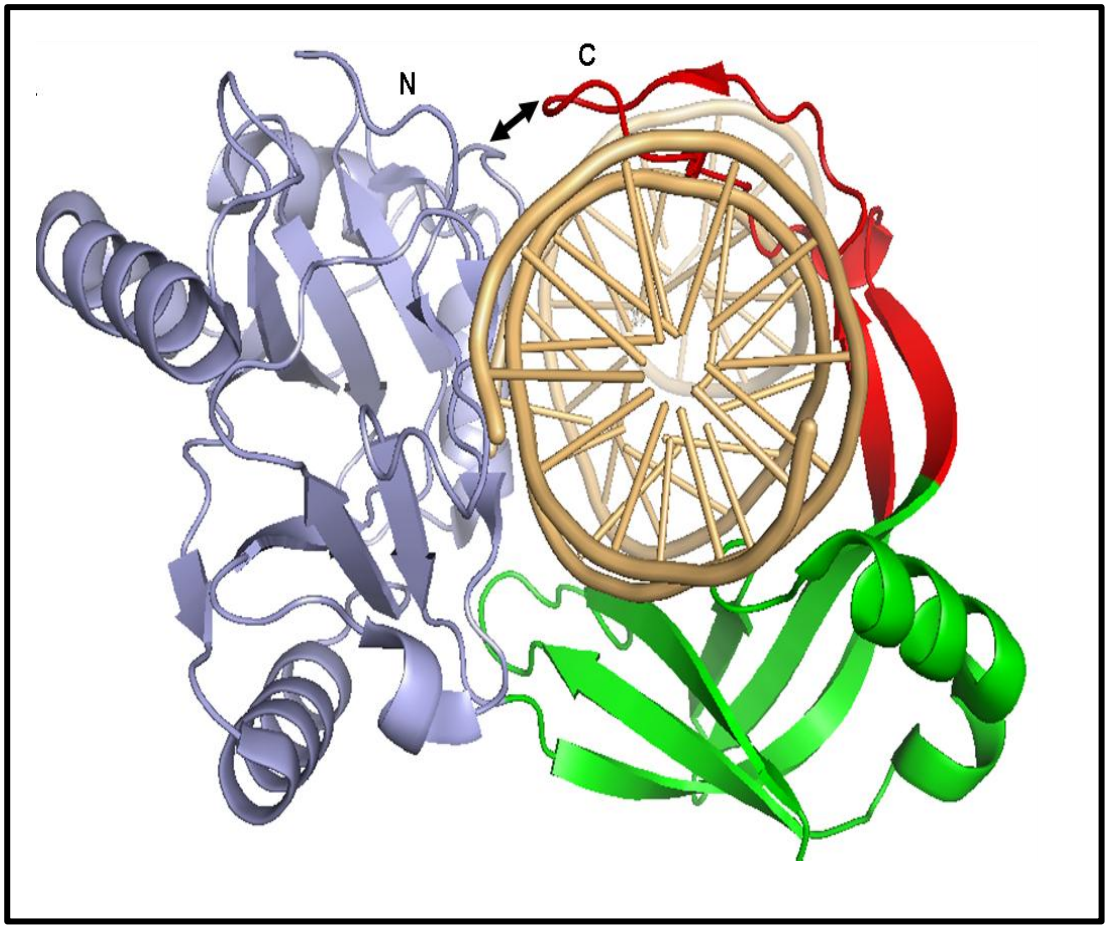


Figure 1.6: Fold structure of PBCV-1 DNA ligase upon DNA binding

Ribbon diagram shows the DNA clamp formed by PBCV-1 DNA ligase as it encircles the nicked DNA substrate. The latch (red) encircles and clamps the DNA substrate. The arrow represents the domain shift by the OB-fold domain in order to position the now structured C-terminal 'latch' region for DNA stabilisation and interdomain contacts with the AdD (PDB 2QT). The adenylation domain is shown in grey, the OB-fold domain in green and the latch region in red (adapted from Nair et al., 2007). The image was prepared with PyMol (DeLano, 2002).

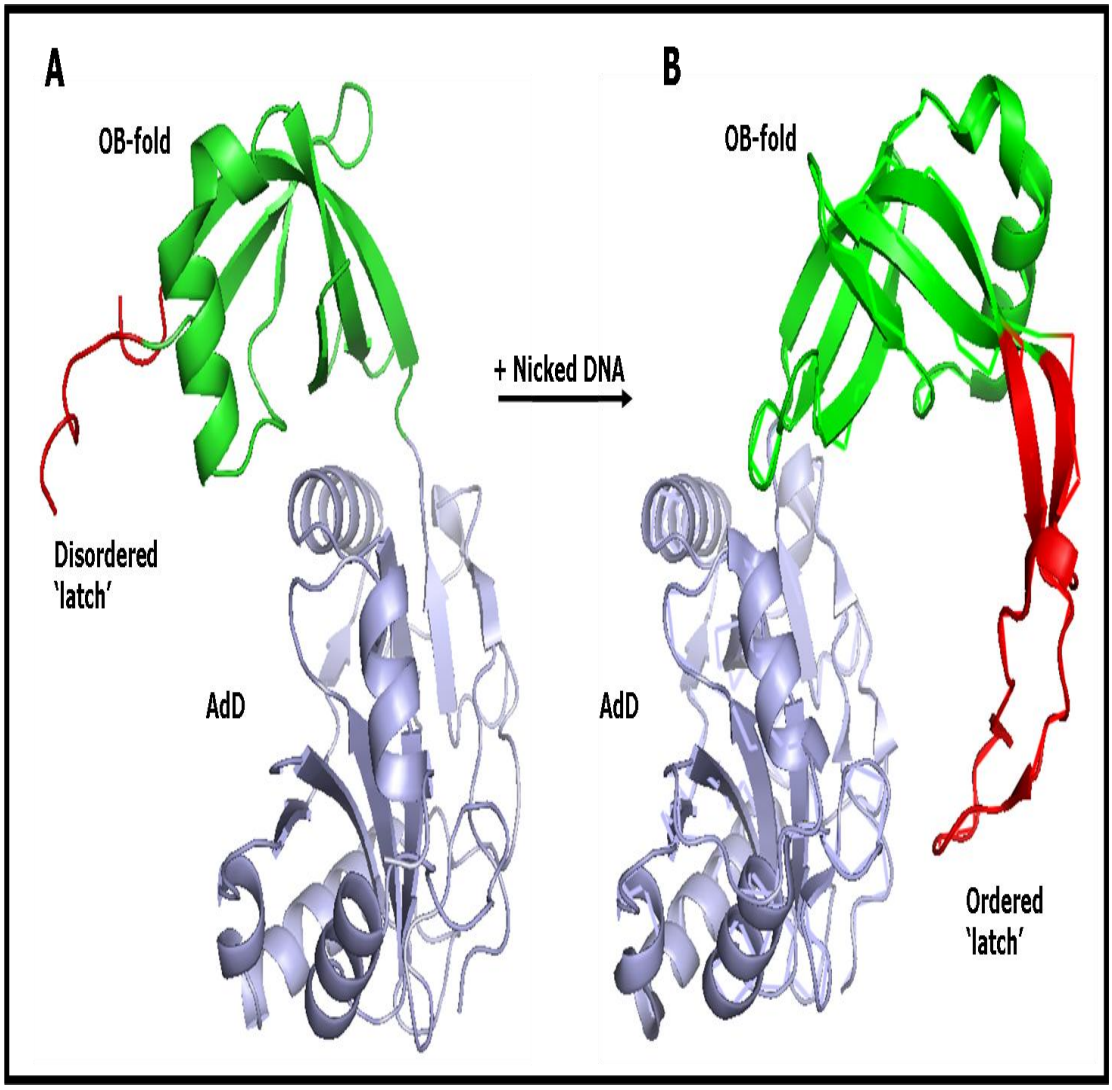


Figure 1.7: Conformational changes occurring in the structure of PBCV-1 DNA ligase upon DNA binding

Ribbon diagrams showing the DNA ligase, panel A: without DNA (PDB: 1P8L), panel B: with the addition of nicked DNA substrate (DNA not shown) (PDB: 2Q2T). Upon addition of DNA, the disordered latch (red) becomes ordered and encircles and clamps the DNA substrate (Odell *et al.*, 2000; Nair *et al.*, 2007). The images were prepared with PyMol (DeLano, 2002).

1.4.2 NAD^+-dependent DNA ligases

The crystal structures of full-length adenylated *Thermus filiformis* DNA ligase (Lee *et al.*, 2000), *Enterococcus faecalis* ligase (Gajiwala & Pinko, 2004), adenylation domain of the NAD^+-dependent *Mycobacterium tuberculosis* DNA ligase (Srivastava *et al.*, 2005) and *E. coli* ligase in complex with DNA (Nandakumar *et al.*, 2007) have been determined.

The structure of *E. coli* LigA (*Eco*LigA) in complex with an AMP-DNA reaction intermediate revealed that the topology and connectivity of the circumferential clamp formed by LigA is distinct from the arrangement adopted by HuLigI (Figure 1.8, panels A & B; Pascal *et al.*, 2004; Nandakumar *et al.*, 2007).

*Eco*LigA encircles the nicked DNA as a 'C-shaped protein clamp' with multiple residues in the AdD, OB and HhH domains contacting the DNA and a 19 bp 'footprint' over the centrally placed nick (Nandakumar *et al.*, 2007). The HhH was found to interact in the minor groove of both DNA strands close to the nick. The zinc finger motif and domain Ia made no physical DNA contacts. Domain Ia is known to play a role in ligase adenylation and the zinc finger module was found to stabilise the OB and HhH domains and it remains in a fixed position in relation to the OB-fold domain (Nandakumar *et al.*, 2007). The HhH domain undergoes a large shift relative to the OB-fold domain in order to facilitate the formation of contacts with the DNA backbone (Nandakumar *et al.*, 2007).

The HuLigI DBD and *Eco*LigA HhH domain have different protein folds, but occupy similar positions on the DNA substrate (see Figure 1.8, panels A & B). They appear to play analogous roles contacting and stabilising the DNA in co-ordination with the conserved AdD and OB-fold domains (Tomkinson *et al.*, 2006; Nandakumar *et al.*, 2007). HuLigI closes the protein clamp around the nicked DNA substrate via 'kissing' contacts between the OB-fold domain and DBD. *Eco*LigA closes the clamp via kissing contacts between the AdD and HhH domains in the major groove of the DNA, opposite the nick.

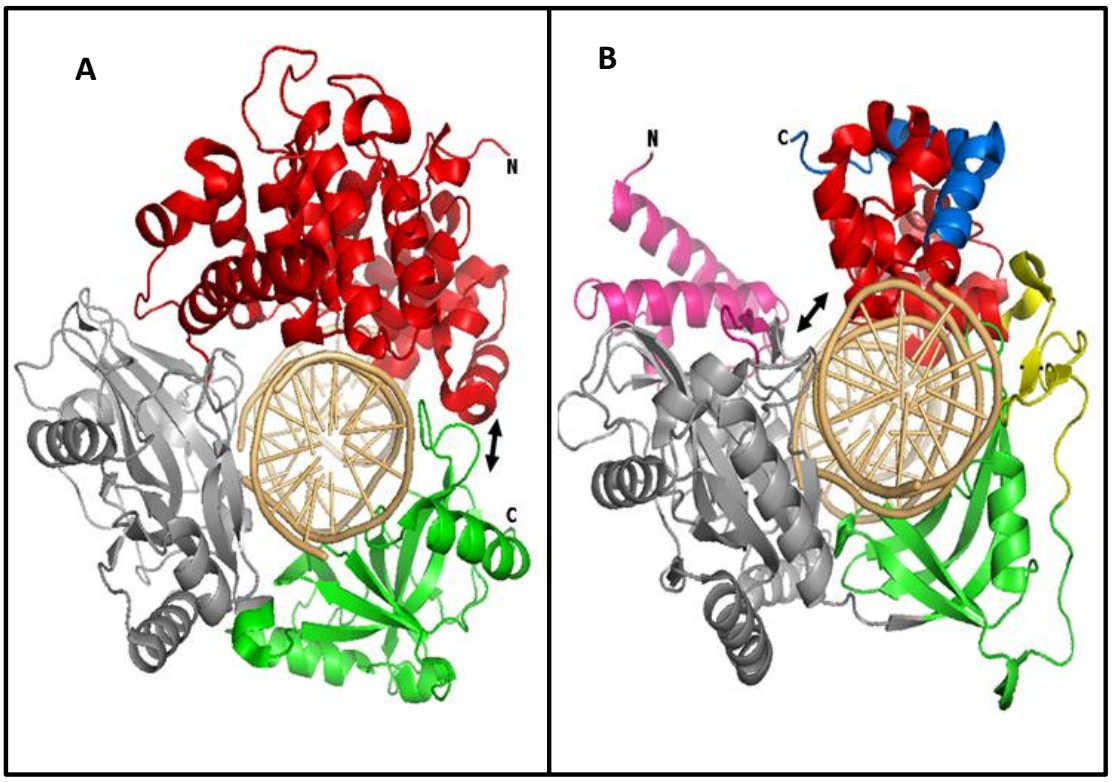


Figure 1.8: Comparison of HuLigI ligase I and *E. coli* LigA fold structure

Ribbon diagrams comparing the DNA clamp of: HuLigI (PDB 1X9N) and *Eco*LigA (PDB 2OWO). The adenylation domains of the two proteins were aligned. The arrows indicate the movement of the relevant domains encircling the DNA and closing the circumferential protein clamp. In both structures the AdD is coloured in grey and the OB-fold domain in green. In HuLigI the DBD is highlighted in red. In *Eco*LigA, domain Ia is shown in magenta, the zinc-finger domain in yellow, HhH in red and the C-terminal BRCT domain in blue (adapted from Nandakumar *et al.*, 2007).

1.5 Mammalian DNA ligases

Mammalian cells have *LIG1, LIG3 and LIG4* genes encoding for four distinct DNA ligases (I, IIIα, IIIβ and IV; Montecucco *et al.*, 1992; Tomkinson & Mackay, 1998; Tomkinson *et al.*, 2006). Homologues of *LIG1* and *LIG4* genes are reportedly present in all eukaryotes but LIG3 genes are restricted to vertebrates (Tomkinson & Mackay, 1998). The catalytic core of mammalian DNA ligases (comprising of AdD and OB-fold domains) is embellished with additional N- and C- terminal domains, which are essential for biological activity and protein-protein interactions (Timson *et al.*, 2000, Pascal *et al.*, 2006; Tomkinson *et al.*, 2006).

1.5.1 DNA ligase I

DNA ligase I is the main replicative ligase of eukaryotes where its function is to join Okazaki fragments during lagging strand DNA synthesis (Ogawa & Okazaki, 1980; Martin & MacNeil 2002). It also plays a role in the base excision repair pathway (BER; described in depth in section 1.7.1.3) where it seals single-strand nicks after repair of DNA damage (Martin & MacNeil, 2002; Almeida & Sobol, 2007). The human cell line 46BR.1G1, which is defective in DNA ligase I function, exhibits abnormal joining of Okazaki fragments during S phase of the cell cycle, a defect that can be complemented by addition of exogenous DNA ligase I protein (Doherty & Suh, 2000). This cell line is also hypersensitive to simple DNA-alkylating agents (*e.g.* methyl methanesulphonate) because of a defect in the long-patch sub-pathway of base excision repair (Levin *et al.,* 2000).

DNA ligase I is the product of the *LIG1* gene, is composed of a carboxyl-terminal catalytic core, an N-terminal DNA binding domain and a PCNA binding motif (Figure 1.9; Tomkinson *et al.,* 1990; Pascal *et al.,* 2004). The recruitment of HuLigI to the site of DNA damage is dependent on its interaction with PCNA (Levin *et al.,* 1997; Tomkinson & Mackey, 1998). Human PCNA (HuPCNA) has been shown to interact with and stimulate the activity of HuLigI, although the specific mechanism of stimulation is unknown (Levin *et al.,* 1997; Pascal *et al.,* 2006). PCNA is a homotrimer that non-specifically binds to and encircles DNA at the site of a nick. It is termed a member of the DNA 'sliding clamp family' – a group of enzymes that act as a scaffold for the assembly of DNA processing factors (Tom *et al.,* 2001; Dionne *et al.,* 2003). It associates with, and facilitates the interaction of more than 25 cellular proteins involved in DNA synthesis and repair (Pascal *et al.,* 2006). PCNA-interacting proteins contain a consensus sequence called the PCNA-interacting protein box (PIP). The conserved motif QXXhXXaa is located in this sequence; where X represents any amino acid, h represents hydrophobic residues (*e.g.* Leu, Ile, or Met), and 'aa' represents aromatic residues (*e.g.* Phe, Tyr, or Trp; Vivona & Kelman, 2003). HuLigI contacts PCNA, via the PIP motif, at the PCNA inter-domain connector loop (IDCL; Pascal *et al.,* 2006). Recent structural studies involving the yeast DNA ligase I homologue, Cdc9, have revealed another site of interaction for the ligase PIP motif

within the C-terminal domain of PCNA (Vijayakumar *et al.*, 2007). Vijayakumar *et al.* (2007) also showed that the nick-joining ability of Cdc9 was inhibited by replication factor C (RFC), the protein responsible for loading PCNA onto DNA (Dionne *et al.*, 2003). They found that nick joining activity was recovered when PCNA was loaded onto DNA by RFC and the RFC/PCNA complex had dissociated following loading. As neither PCNA nor DNA ligase I has specific nick-sensing ability, it is hypothesised that RFC holds the PCNA at the nick generated by the action of DNA polymerase and Flap endonuclease-1(FEN-1; FEN-1 is the structure specific nuclease involved in cleavage of over-lapping nucleotides – 'flaps' that occur during DNA replication and repair).

HuLigI also interacts with hRad9-hRad1-hHus1 (9-1-1), a heterotrimeric DNA sliding clamp, involved in cell cycle checkpoints and known to be a sensor of DNA damage (Wang *et al.*, 2006; Song *et al.*, 2009). The 9-1-1 clamp complex is loaded onto DNA by hRad17-RFC in response to replication blockage and DNA damage to initiate a signalling cascade (Wang *et al.*, 2006). Recent studies have shown interactions between the DBD of HuLigI with both PCNA and 9-1-1 clamps (Pascal *et al.*, 2006; Song *et al.*, 2009). The interactions between HuLigI and PCNA linked to DNA initially occur via the HuLigI PIP box, thereby facilitating a subsequent interaction between the HuLigI DBD and PCNA (Song *et al.*, 2009).

HuLigI also encodes an N-terminal nuclear localisation signal (NLS; Figure 1.9). NLS's contain lysine and/or arginine clusters immediately preceded by a proline (Dingwall & Laskey, 1991). Localisation of the enzyme to the nucleus allows it to be targeted to DNA replication and repair protein complexes (Montecucco *et al.*, 1995).

Recently, Vijayakumar *et al.* (2009) reported that phosphorylation of the N-terminal region of HuLigI in post-translational modifications is critical for its participation in different DNA transactions. Their group expressed a mutant version of HuLigI in the 46BR.1G1 (HuLigI-deficient) cell line, in which four serine residues (Ser51, 66, 76, and 91), phosphorylated *in vivo*, were replaced with either alanine or aspartic acid. The cell lines expressing these phosphorylation site mutants exhibited a dramatic

reduction in proliferation and DNA synthesis and were also hypersensitive to DNA damage (Vijayakumar *et al.*, 2009).

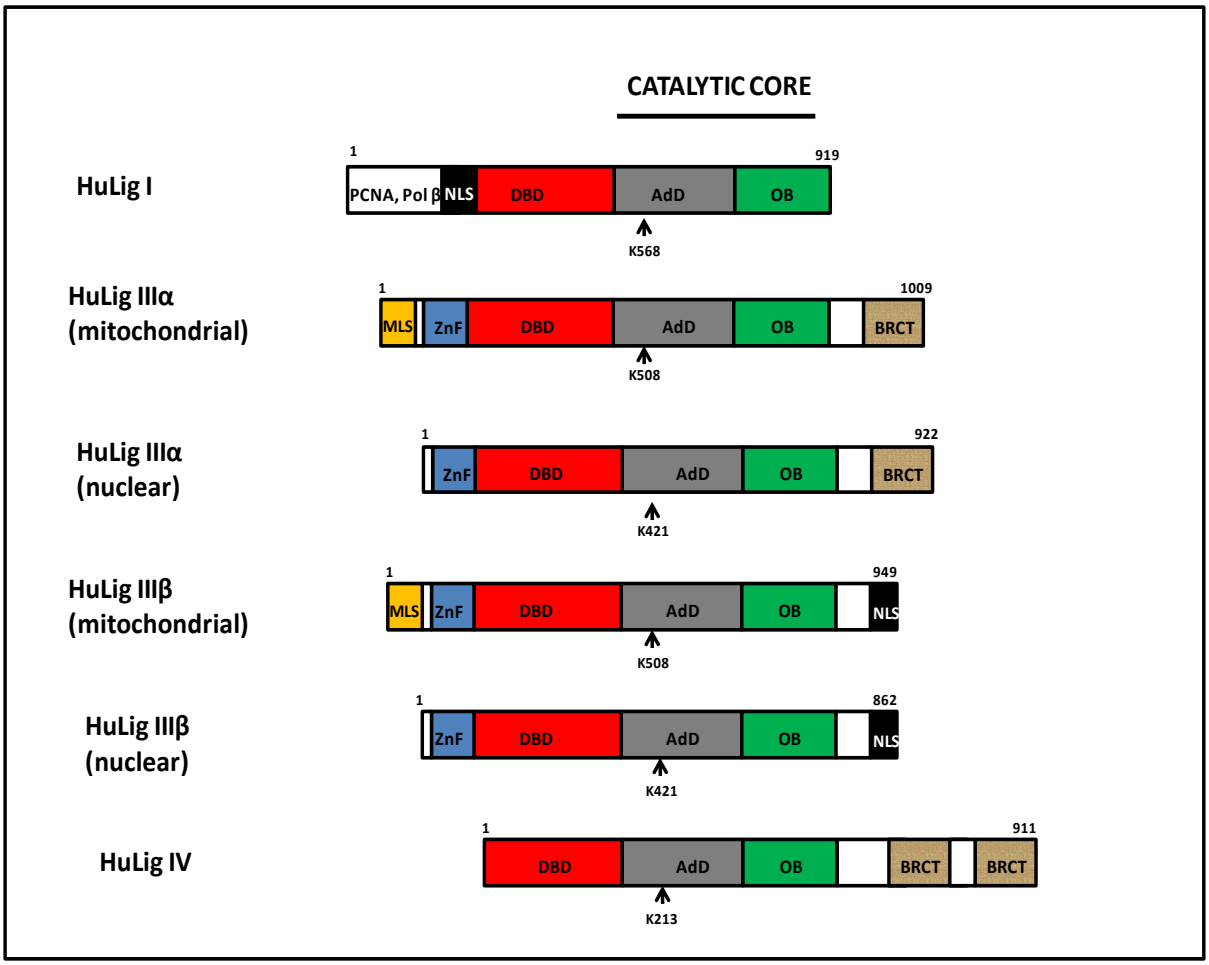


Figure 1.9: Domain structures of mammalian DNA ligases

Schematic representation of human DNA ligase I, IIIα, IIIβ and IV with some regions of the proteins and approximate binding sites shown for some of their main, known partners. The DNA binding domain (DBD; *red)*, catalytic core (composed of adenylation; *grey* and OB-fold domain; *green*) are conserved in ligases (I-IV). The positions of the active site lysine residues (K568, K508, K421, K508, K421 and K213) within the adenylation domain that form the covalent bond with AMP are shown. Abbreviations: PCNA - Proliferating cell nuclear antigen, NLS - nuclear localisation signals, BRCT - Breast and ovarian cancer susceptibility protein 1 C-terminal, ZnF - Zinc finger, MLS - mitochondrial leader sequence.

1.5.2 DNA ligase III

DNA ligase III is unique to vertebrates and is therefore absent in lower eukaryotes (Martin & MacNeill, 2002). The mammalian *LIG3* gene encodes for three different DNA ligase polypeptides (Figure 1.9). The nuclear and mitochondrial versions of

DNA ligase IIIα are synthesised in somatic cells by the use of different in-frame alternative translational initiation sites from DNA ligase IIIα mRNA (Lakshmipathy & Campbell, 1999; Perez-Jannotti *et al.,* 2001). DNA ligase IIIβ is generated by a germ-cell specific alternative splicing mechanism, in which the terminal 3'-coding exon in DNA ligase IIIα mRNA is replaced by a different exon (Perez-Jannotti *et al.,* 2001). The two isoforms of DNA ligase III differ at their C-termini. DNA ligase IIIα contains a breast and ovarian cancer susceptibility protein (BRCT) motif (Figure1.8; Timson *et al.,* 2000). The BRCT domain, an autonomously folding protein module of about 95 amino acids was first identified in the carboxy-terminal region of the BRCA1 tumour suppressor protein but which has since been found in a range of proteins implicated in DNA replication, DNA repair and checkpoint functions (Moore *et al.,* 2000). In DNA ligase IIIβ, this motif is replaced by a short sequence which acts as a nuclear localization signal (Lakshmipathy & Campbell, 1999).

The BRCT domain of HuLigIIIα facilitates the enzyme's interaction with XRCC1, a multi-domain protein that functions as a 'molecular scaffold', stabilising DNA and interacting with other proteins involved in the base excision repair pathway (BER) such as DNA polymerase β (polβ) and poly(ADP-ribose) polymerase (PARP; Caldecott *et al.,* 1994; Doherty & Suh, 2000; Thompson & West, 2000). DNA ligase III also contains an N-terminally located zinc finger domain, containing four cysteine residues that are homologous to the zinc finger domain of PARP (Mackey *et al.,* 1999). The LigIII zinc finger is known to interact with, and mediate the binding of the HuLigIII-XRCC1 complex with DNA (Mackey *et al.,* 1999; Tomkinson *et al.,* 2006).

Cotner-Gohara *et al.* (2008) showed that the ZnF alone fails to bind tightly to nicked DNA and instead functions in cooperative manner with the DBD to form a DNA-binding module that interacts specifically with nicked DNA. The catalytic core of DNA ligase III also binds selectively to nicked DNA but nick joining activity is greatly accelerated by the ZnF-DBD module in full-length DNA ligase III**,** suggesting that the ZnF-DBD and catalytic core function as two separate DNA-binding modules during catalysis. The jack knife model proposed by Cotner-Gohara *et al.* (2008) suggests that the two DNA binding modules of DNA ligase III contact the DNA substrate at different times during the end-joining reaction. The ZnF-DBD domain

binds to the nick first and then subsequently disengages from the nick to allow access by the catalytic core for catalysis (Ellenberger & Tomkinson, 2008).

1.5.3 DNA ligase IV

DNA ligase IV (911 amino acids), encoded by *LIG4*, is characterized by a carboxy–terminal extension comprising two BRCT domains (Figure 1.9), and the region between them that binds the 38 kDa XRCC4 protein (Grawunder *et al.*, 1998; Sibanda *et al.*, 2001). This interaction has been found to increase the stability of DNA binding by ligase IV and stimulate double-strand DNA break repair (Critchlow *et al.*, 1997; Grawunder *et al.*, 1998). Cells that lack XRCC4 or DNA ligase IV are hypersensitive to ionising radiation and are deficient in V(D)J recombination [variable (V)-diversity(D)-joining(J) - a site specific recombination process; Li *et al.*, 1995; Gao *et al.*, 1998]. The DNA ligase IV/XRCC4 complex is required for repairing DNA double- strand breaks by the non-homologous end joining pathway (NHEJ) pathway (described in detail in section 1.8.1.1; Martin & MacNeill, 2002). Biochemical and genetic analyses have revealed that DNA ligase IV is the primary ligase utilised for end joining as neither DNA ligase I nor III are able to complement for DNA ligase IV in cells deficient in the enzyme (Baumann & West, 1998).

1.6 PBCV-1 DNA ligase

PBCV-1 DNA ligase is the smallest (298 amino acid) eukaryotic ATP-dependent ligase that has been characterised to date (Ho *et al.*, 1997). Other eukaryotic DNA ligases such as human ligases I, III and IV are 919, 922 and 844 amino acid polypeptides respectively (Martin & MacNeil, 2002). PBCV-1 is a lytic protovirus of the family *Phycodnaviridae* (large polyhedral, plaque-forming viruses) of the genus *Chlorovirus*. It infects the *Chlorella* algae which are endosymbionts of the protozoan organism *Paramecium bursaria* (Yamada *et al.*, 2006). The PBCV-1 genome is a linear, double-strand DNA molecule with inverted terminal repeats and covalently closed hairpin telomeres. The sequence of the 330-kbp PBCV-1 genome encodes 380 polypeptides (Ho *et al.*, 1997; Yamada *et al.*, 2006). Many proteins encoded by

PBCV-1 are the smallest or among the smallest proteins in their class these include the PBCV-1 DNA ligase, RNA capping enzyme (guanylyltransferase) and topoisomerase (Ho *et al.,* 1997; Shuman & Lima, 2004).

Topoisomerases manipulate the topological state of cellular DNA by controlling the superhelicity and DNA strand interlocking. They are involved in the relaxation of supercoiled DNA by the cleavage of one strand and rotating it around the intact strand (type I) or by passing an intact DNA duplex through a DNA strand break in the same or a neighbouring DNA duplex (type II; Lavrukhin *et al.,* 2000). The 120 kDa, type II topoisomerase encoded by PBCV-1, cleaves double-stranded DNA thirty times faster than human toposiomerase II (Lavrukhin *et al.,* 2000). As with the PBCV-1 DNA ligase, it is used as a model enzyme to study the mechanism, biochemistry and structure-function analysis of topoisomerase activity (Lavrukhin *et al.,* 2000).

The PBCV-1 DNA ligase consists only of the conserved catalytic core without the flanking N and C terminal domains found in the larger cellular, eukaryotic DNA ligases (Figure 1.10, panel A; Ho *et al.,* 1997). The structure of the 34 kDa PBCV-1 DNA ligase-adenylate has been determined and reflects the same two-domain arrangement (a 188 amino acid N-terminal adenylation domain and a 110 residue OB-fold domain) and structural features of the larger 41 kDa T7 DNA ligase (Subramanya *et al.,* 1996; Odell *et al* 2000). This can be seen in superimposition of the active sites of the two enzymes where critically conserved residues in the key motifs I and III and their contacts with the ribose sugar of adenine ring align almost identically (Figure 1.10, panel B; Odell *et al.,* 2000).

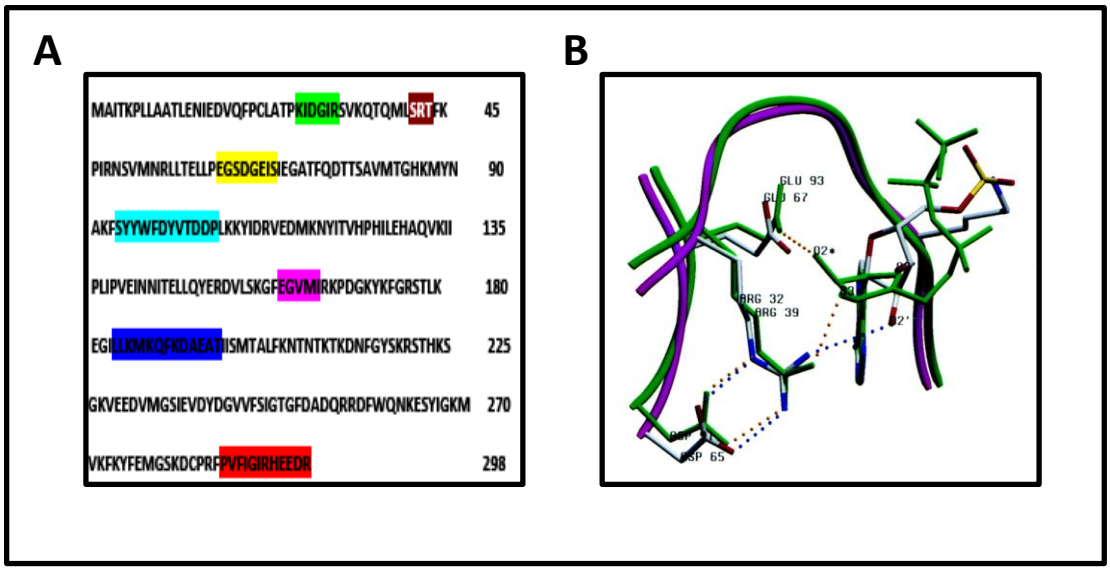


Figure 1.10: PBCV-1 DNA ligase amino acid sequence and superimposition of conserved motifs in PBCV-1 and T7 bacteriophage

The PBCV-1 DNA ligase consists only of the conserved catalytic core without the flanking N and C terminal domains. Panel A :Nucleotidyltransferase motifs comprising AdD [I (g*reen*), Ia (*brown*), III (*yellow*), IIIa (*cyan*), IV (*purple*), V (*blue*)] and OB-fold domain [VI (*red*)] are highlighted. Panel B: Superimposition of key residues of the active site in motifs I and III and their contacts with the ribose sugar of AMP in PBCV-1 (shown in green) and ATP in T7 bacteriophage ligase (shown in purple) (Odell *et al.*, 2003).

PBCV-1 DNA ligase is regarded by some as the catalytic core present in all ligases such that anything that applies to the catalytic core of this enzyme will apply to the larger ligases (Odell *et al.*, 2003). This 'minimal' enzyme is a fully functional ligase as it is able to sustain mitotic growth, excision repair and non-homologous end joining (NHEJ) in *S. cerevisiae* strains deleted for the endogenous ligases, the Cdc9 (the LigI homologue, essential for cell growth) and LIG4 (a DNA ligase IV homologue, not essential for cell growth). It is also able to repair yeast DNA damage induced by UV irradiation or treatment with methyl methane sulphonate (MMS; Sriskanda *et al.*, 1999).

In recent years, PBCV-1 ligase has emerged as an excellent model for structural and functional studies of eukaryotic ligases. PBCV-1 DNA ligase is the only DNA ligase with structures analysed for DNA bound and unligated states, providing instructive insights into the conserved mechanism of DNA ligation and conformational changes

that occur upon DNA binding (Figure 1.7; Odell *et al.,* 2000; Odell *et al.,* 2003; Nair *et al,* 2007). The fact that the recombinant PBCV-1 ligase is purified in high yield as ligase-adenylate offered an opportunity to solve the structure of the covalent reaction intermediates (Ho *et al.,* 1997; Odell *et al.*, 2000; Odell *et al.,* 2003).

1.7 DNA repair proteins as therapeutic targets

The genetic connections between DNA repair and human cancers have increased interest in the proteins that recognise and rectify specific sites of DNA damage (Madhusudan & Hickson, 2005; Damia & D'Incalci, 2007). An increase in the activity of proteins involved in DNA damage repair pathways is observed in early precursor lesions of human breast, colon, lung and urinary bladder tumours (Martin *et al.,* 2008). It is characterised by the phosphorylation of many DNA damage response proteins hence suggesting that activation of DNA damage response machinery is a consequence of oncogenic activity (Barktova *et al.,* 2005). The activation of DNA repair pathways in tumourigenesis suggests their function in protecting cells against deleterious genetic changes that contribute to cancer formation (Lord *et al.,* 2006).

Ionising radiation and chemotherapeutic drugs, currently used in the treatment of cancer, directly or indirectly damage DNA. Mammalian cells have highly conserved DNA damage sensor mechanisms that result in the activation of a number of signal transduction pathways leading to cell cycle arrest and thus allowing repair or the induction of apoptosis to eliminate heavily damaged cells (Hoeijmakers, 2001; Martin, 2001). Activation of repair pathways can lead to the survival of tumour cells and have a negative impact on the treatment efficacy, leading to resistance to therapies. The causes of resistance to DNA damaging agents can be associated with increased cellular repair activities while defects in DNA repair pathways results in hypersensitivity to these agents (Tentori & Graziani, 2005; Damia & D'Incalci, 2007). Therefore, effectiveness of the existing therapies, ionising radiation and cytotoxic drugs, may be improved in combination with the pharmacological inhibition of DNA repair pathways or DNA damage signalling pathways.

1.7.1 Overview of DNA repair pathways

DNA lesions can involve a single-strand of DNA (single-strand breaks) or both strands (double-strand breaks or DSBs). Single-strand breaks are repaired by the base excision repair (BER) pathway and nucleotide excision repair (NER) pathway whereas double-strand breaks are repaired by the non-homologous end joining (NHEJ) and homologous recombination (HR) pathways (Takata *et al.*, 1998; Costa *et al.*, 2003; Almeida & Sobol, 2007).

DSBs in DNA can arise on genome exposure to a variety of DNA-damaging agents, such as ionizing radiation (IR), ultra-violet light, reactive chemicals and reactive oxygen species (Lord *et al.*, 2006; Tomimatsu *et al.*, 2007). DSBs also occur from closely spaced single-strand breaks, during replication of single-strand breaks or in normal endogenous processes that demand that genome sequences be rearranged such as generation of genetic diversity in meiosis and V(D)J recombination during development of the immune system (Grawunder *et al.*, 1998; Hoeijmakers, 2001; Verkaik *et al.*, 2002). Cells respond to DSBs by activating a complex DNA-damage-response pathway that includes cell-cycle arrest, the transcriptional and post-transcriptional activation of a subset of genes (including those associated with DNA repair) or may stimulate apoptosis if the damage is too significant (Lord *et al.*, 2006; Wyman & Kannar, 2006).

The cellular response to DSBs is activated by ataxia-telangiectasia mutated (ATM) kinase, a phosphatidylinositol 3-kinase-related kinase (PIKK) defective in ataxia telangiectasia (AT - an autosomal recessive disorder; Hoeijmakers, 2001; Tomimatsu *et al.*, 2007). ATM is activated when it is recruited to sites of DSB damage by the MRE11-RAD50-NBS1 (MRN) complex (Wyman & Kannar, 2006). Following ATM activation, several DNA-repair and cell cycle checkpoint proteins, including P53, the kinases CHK1 and CHK2, are activated leading to cell cycle arrest and DNA repair (Tomimatsu *et al.*, 2007). During cell cycle arrest, DSBs are repaired by NHEJ and HR (Wyman *et al.*, 2004; Hefferin & Tomkinson, 2005).

HR dominates in late S and the G2 phase of the cell cycle and repairs breaks by copying genetic information from either homologous chromosomes or sister chromatids (Wyman *et. al.,* 2004; Lord *et al.,* 2006). NHEJ is a homology independent process for DSB repairs in the G0, G1 and early S phases of the cell cycle (Jayaram *et al.,* 2008).

1.7.1.1 Non-homologous end joining pathway (NHEJ)

Eukaryotic NHEJ is a multi-step reaction catalysed by a core set of conserved proteins, comprising the Ku70-Ku80 heterodimer (Ku) and a complex of DNA ligase IV, XRCC4 and XLF/Cernunnos (Figure 1.11, panel A; Hefferin & Tomkinson, 2005; Lord *et al.,* 2006). Ku has high affinity for DNA ends and binds as a ring to double- strand break ends, where it can translocate along the duplex (Wyman & Kannar, 2006). The crystal structure of Ku complexed with DNA revealed that the heterodimer has a toroidal shape with a central hole to accommodate duplex DNA (Hefferin & Tomkinson, 2005). The DNA-binding subunits of Ku70 and Ku80 bind to the ends of DSB and then recruit DNA-dependent protein kinase (DNA-PKcs; a member of the PIKKs family). Assembled together on a DNA end, Ku and DNA-PKcs constitute the DNA-dependent kinase (DNA-PK). Activated DNA-PK holoenzyme recruits other DNA-repair proteins including Artemis, XRCC4, DNA ligase IV along with XLF/Cernunnos and DNA polymerase μ to site of DNA damage to carry out the DNA end joining (Wyman & Kannar, 2006; Tomimatsu *et al.,* 2007). The Artemis protein, an endonuclease, is essential for processing damaged DNA ends such as heterologous loop and stem-loop DNA structures containing single-stranded DNA adjacent to double-stranded DNA (Wyman & Kannar, 2006). A DNA ligase IV monomer forms a stable complex with a XRCC4 dimer that appears to stabilise, activate and target the ligase to the double-strand break and this complex carries out the final ligation step (Jayaram *et al.,* 2008). XRCC4 possess DNA binding affinity and the residues involved in the interaction between XRCC4 and DNA ligase IV are highly conserved across eukaryotic species from human to yeasts (Hefferin & Tomkinson, 2005). Recently, an XRCC4-like protein, XLF (also known as Cernunnos), has been identified as an interaction partner of the DNA ligase

IV/XRCC4 complex, the function of which is not known in NHEJ (Lord *et al.*, 2006; Jayaram *et al.*, 2008).

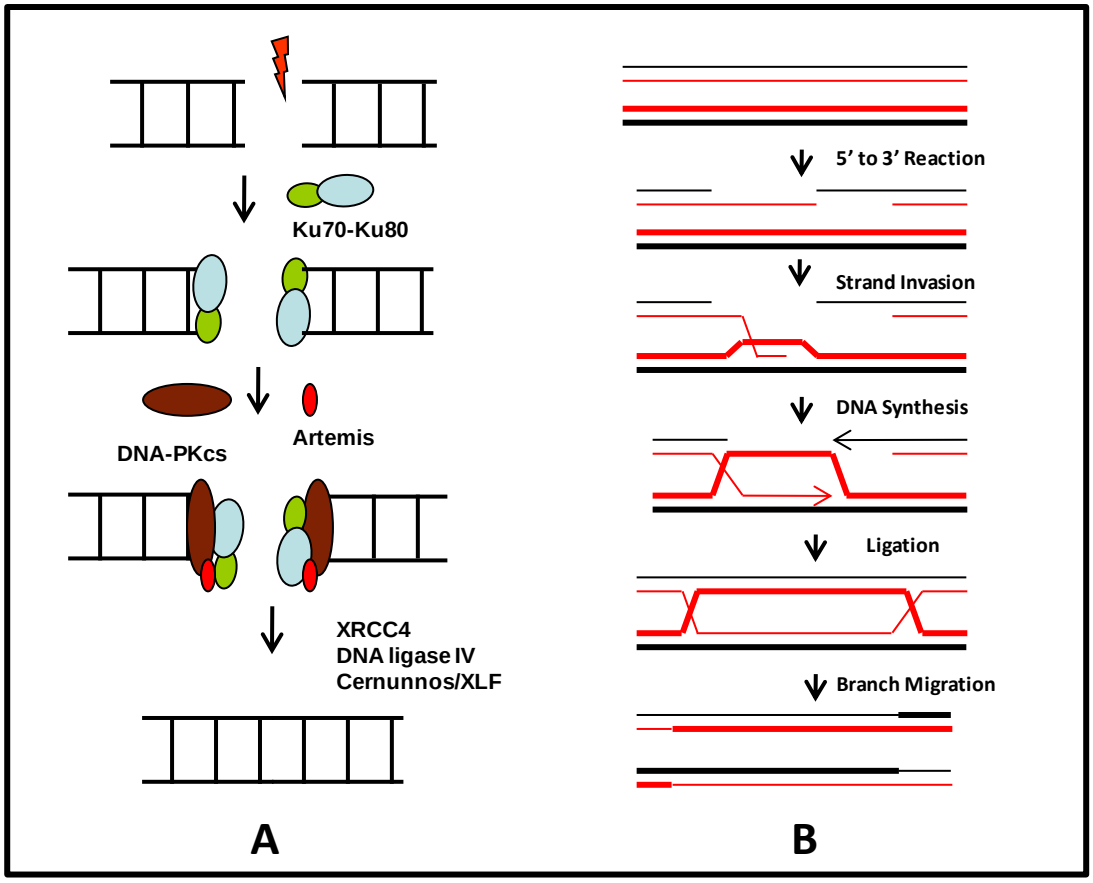


Figure 1.11: NHEJ and HR pathway for double-strand DNA repair

Double-strand breaks in DNA are repaired by non-homologous end joining pathway (NHEJ) or Homologous recombination (HR) pathway. Panel A: NHEJ mechanism of DSB repair. Induction of a double-strand break results in the recruitment of the Ku heterodimer to the lesion site to bind the free DNA. This is followed by the recruitment of the catalytic subunit of DNA protein kinase (DNA-PKcs). Assembly of DNA-PK results in the recruitment of the DNA ligase IV/XRCC4/XLF complex. After gap filling by DNA polymerase µ, DNA ligase IV seals the nick (Jayaram *et al.*, 2008). Panel B: DNA double-strand break repair through homologous recombination. The close pairs of parallel lines represent the two strands of duplex DNA, while the lighter coloured pair is the sister chromatid of the darker coloured pair. Upon double-strand damage, the ends are nucleolytically processed to result in 3' single-stranded tails. Processed broken DNA forms a joint molecule with the intact homologous repair template DNA using strand invasion and strand exchange activities of Rad51 protein. The damage isrepaired by resynthesis of DNA (Filippo *et al.*, 2008).

1.7.1.2 Homologous Recombination (HR)

Homologous recombination consists of processing of double-strand breaks (DSB) to give single-strand tails, formation of a recombinase filament on the resultant single-stranded DNA (ssDNA), strand invasion into a homologous sequence followed by DNA polymerase extension and ligation (Wyman *et al.,* 2004; Wyman & Kannar, 2006).

DNA breaks are first converted to 3' ssDNA tails by nuclease activities of the MRN complex consisting of Mre11, Rad50 and NBS1 (Nijmegen breakage syndrome 1) proteins (Figure 1.11, panel B; Wyman *et al.,* 2004; Filippo *et al.,* 2008). The single-stranded DNA is bound by replication protein A (RPA), which is later displaced by Rad51 (Wyman *et al.,* 2004). The heterotrimeric RPA is an abundant nuclear protein that binds to and removes secondary structures in the ssDNA. Rad51 is a eukaryotic recombinase protein that mediates the pairing and shuffling of DNA sequences during HR (Wyman & Kannar, 2006). The Rad51-ssDNA nucleoprotein filament assembly, referred to as the presynaptic filament, is a slow process and prone to interference by the ssDNA binding protein RPA. It requires the involvement of recombination mediator proteins, such as Breast Cancer Type 2 susceptibility protein (BRCA2), for enhancement (Filippo *et al.,* 2008). BRCA2 physically interacts with Rad51 and overcome the inhibitory effect of RPA on the assembly of the presynaptic filament. Cells deficient in BRCA2 are sensitive to DNA-damaging agents and mutations in the BRCA2 gene are known to predispose the affected individuals to breast, ovarian, and other cancers (Lord *et al.,* 2006). The Rad51 nucleoprotein filament mediates homology recognition, strand invasion and strand exchange during the repair process (Wyman & Kannar, 2006; Filippo *et al.,* 2008).

DNA polymerase uses the joint molecule between the broken DNA and the intact homologous template as a substrate to reincorporate missing nucleotides. The gaps are ligated at the end of the newly synthesised sequence and the process is completed by resolvases, which remove the links between sister chromatids (Wyman *et al.,* 2004).

In addition to HR and NHEJ, a non-conservative backup pathway called Micro homologous end joining (MHEJ) exists for double-strand break repair, which relies on micro homologies at repair junctions (Wang *et al.,* 2003). This was identified in NHEJ deficient cells where DNA substrates with non-matched ends were joined at much lower frequency than compatible-ended substrates (Wang *et al.,* 2003). This type of mechanism is also found to be involved in chromosomal translocations and is independent of the core proteins involved in NHEJ (Tsuji *et al.,* 2004). Recent studies have shown that MHEJ and NHEJ are two distinct pathways. NHEJ relies on the Ku heterodimer and the DNA ligase IV/XRCC4 complex but MHEJ requires DNA ligase I or III depending on the length of micro homology sequences (Liang *et al.,* 2008).

1.7.1.3 Base Excision Repair (BER)

Base excision repair (BER) is the predominant DNA damage repair pathway for the processing of small base lesions that do not distort the DNA helix (Krokan *et al.,* 2000; Hoeijmakers, 2001). These include alterations of DNA bases by deamination of cytosine to uracil, errors in DNA replication (misincorporation of dUTP or 8-oxo-dGTP) and from by-products of normal cellular metabolism such as reactive oxygen species (Izumi *et al.,* 2003). The reactive oxygen species include superoxide anions, hydroxyl radicals and hydrogen peroxide, derived from oxidative respiration and products of lipid peroxidation (Izumi *et al.,* 2003; Almeida & Sobol, 2007).

BER is normally initiated by a lesion-specific DNA glycosylase which recognises and removes the aberrant DNA base to leave an apurinic or apyrimidinc site (AP site; Figure 1.12; Krokan *et al.,* 2000; Fortini & Dogliotti, 2007). Once the AP sites in DNA are created by this excision, AP endonucleases (APE) cleave the phosphodiester bond immediately 5' to the lesion leaving behind a strand break with a normal 3'-hydroxyl group terminated deoxyribonucleotide strand and a 5'deoxyribose-phosphate terminated strand (5'dRP; Almeida & Sobol, 2007).

AP sites generated during BER are processed by two sub-pathways: short-patch BER or long-patch BER (Sattler *et al.,* 2003; Almeida & Sobol, 2007). Short-patch BER

involves replacement of a single nucleotide whereas in long-patch BER, 2-13 nucleotides are replaced (Hoeijmakers, 2001).

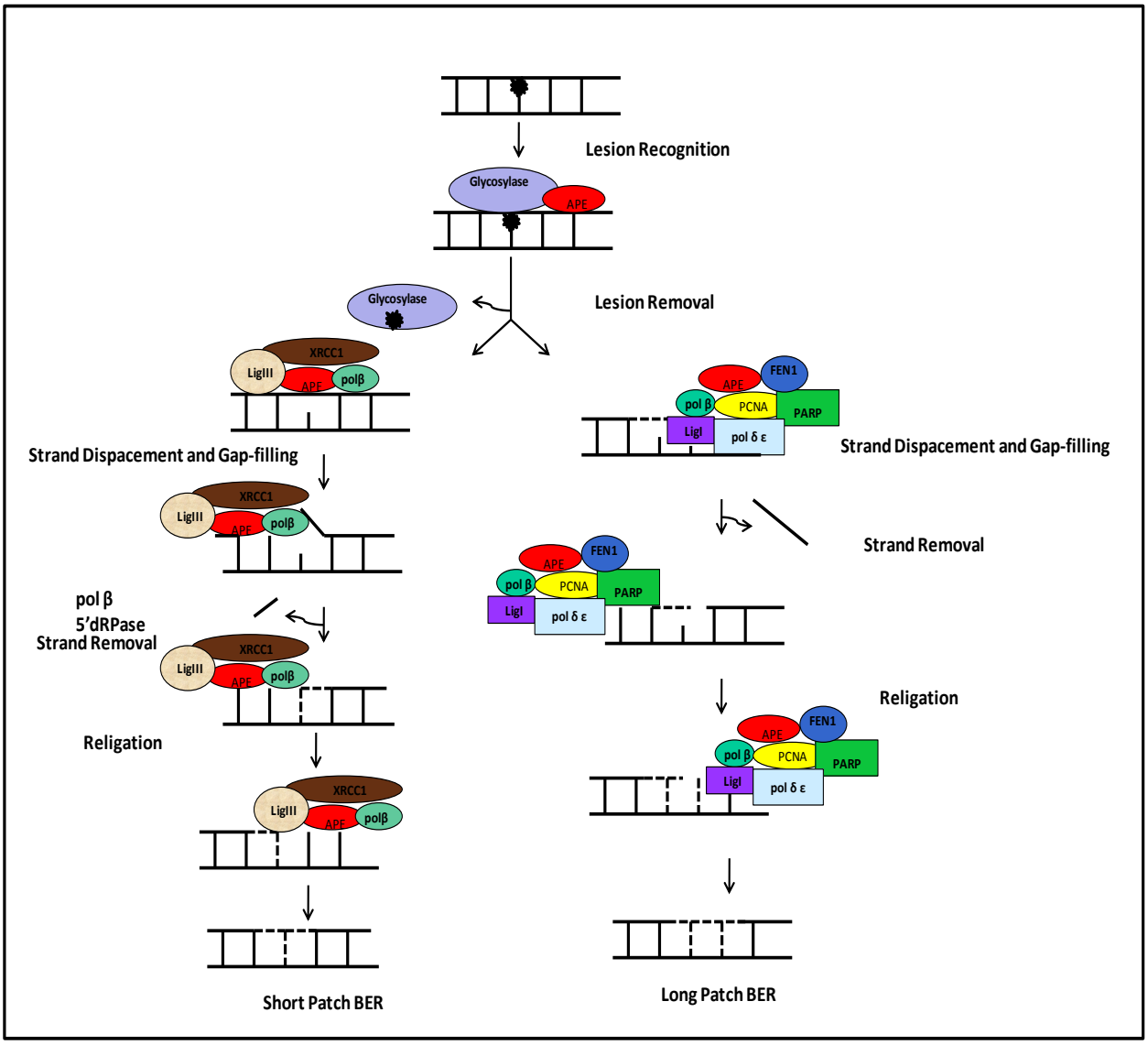


Figure 1.12: General model of the short-patch (left) and long-patch (right) DNA repair pathways.

BER is initiated by glycosylase activity followed by strand excision by APE. In short-patch BER DNA polymerase β incorporates the nucleotide and the resulting nick is sealed by a complex of XRCC1 and DNA ligase IIIα. Long-patch pathway results in the replacement of 2-13 nucleotides including the damaged base. In this type of sub-pathway, the 5' dRP moiety is removed as part of a flap of DNA by FEN1 and re-ligation is completed by DNA ligase I.

In short patch BER, the 5'-terminal dRP resulting from AP endonuclease incision is removed by intrinsic AP lyase activity of DNA polymerase β (pol β) leaving a single nucleotide gap (Krokan *et al.*, 2000). The gap is filled by pol β and the nick is usually sealed by a DNA ligase III/XRCC1 heterodimer to complete the repair (Caldecott *et*

al., 1996). X-ray Cross Complementation protein 1 (XRCC1), discovered by its ability to complement a hypersensitivity phenotype towards DNA damaging agents, acts as a scaffold protein in BER. It coordinates and facilitates single-strand repair by forming a tight complex with DNA ligase III and interacts with other DNA repair proteins including pol β (Caldecott *et al.,* 1996).

In long patch BER, the repair size is usually 2-13 nucleotides in length. This pathway (Figure 1.12) utilizes AP endonuclease for the 5'-incision but a flap endonuclease (FEN1) is required to remove the 5'-terminal dRP moiety along with at least one adjacent nucleotide to leave a gap of two or more nucleotides (Almeida & Sobol, 2007; Fortini & Dogliotti, 2007). This pathway is responsible for repairing damage caused by γ irradiation as these lesions are resistant to dRP elimination by DNA polymerase β and requires FEN1 for processing (Sattler *et al.,* 2003). Repair of such sites requires PCNA, a toroidal homotrimeric DNA binding protein that encircles template DNA, to form a holoenzyme complex with DNA polymerase δ or Ɛ in conjunction with replication factor C (RF-C) and PCNA loading factor (Izumi *et al.,* 2003). PCNA encircles duplex DNA in its central cavity and provides a platform for the attachment of the polymerase (Fortini & Dogliotti, 2007). DNA synthesis and strand displacement by pol β is stimulated by the combined presence of FEN1 and poly(ADP-ribose)polymerase 1 (PARP1; Krokan *et al.,* 2000; Hoeijmakers, 2001). PARP1 recognises the single-strand breaks in the sugar-phosphate backbone of DNA via an N-terminal zinc finger domain, translating the occurrence of DNA breaks into a DNA repair response by covalently transferring ADP-ribose moieties to a variety of nuclear proteins including PARP itself that are then activated or recruited to the site of damage (Tentori & Graziani, 2005). Long-patch repair pathway uses DNA ligase I as an end step to seal the nick (Sattler *et al.,* 2003).

1.7.1.4 Nucleotide excision repair (NER)

Nucleotide excision repair pathway, a multiprotein repair system, acts on lesions produced by UV irradiation [*e.g.* cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4 PPs)] and DNA adducts induced by chemicals like aflatoxinB1 and N-acetoxy-2-acetylaminofluorene (NA-AAF) that lead to helical distortion of the

DNA duplex, interfere with base pairing and obstruct transcription (Hoeijmakers, 2001; Fousteri & Mullenders, 2008). This process requires the combined action of a number of proteins that carry out damage recognition, local opening of the DNA duplex around the lesion, dual incision in the damaged DNA strand, gap repair synthesis and strand ligation (Figure 1.13).

There are two distinct NER sub pathways which differ only in the step involving recognition of the DNA lesion: transcription-coupled (TC-NER) repair, for the repair of actively transcribed genes and global genome (GG-NER) repair, for the removal of damage over the entire genome (Fousteri & Mullenders, 2008; Shuk *et al.,* 2008). Initial damage sensing is performed by the heterotrimeric complex XPC-HR23B in GG-NER and by RNA polymerase II complex (RNAPIIo) in TC-NER. The six core factors that have been implicated in the damage recognition and dual incision steps of GG-NER are the XPC-hHR23B complex, transcription factor IIH (TFIIH; a nine subunit complex that functions in both NER and transcription initiation), XPA (xeroderma pigmentosum complementation group A; a zinc metalloprotein that acts as a limiting factor in NER damage recognition), replication protein A (RPA; a heterotrimeric protein, that binds to single-stranded DNA and is involved in DNA replication and DNA damage checkpoints), XPG and XPF-ERCC1 (Hoeijmakers, 2001; Costa *et al.,* 2003; Shuk *et al.,* 2008).

After lesion recognition, all subsequent steps leading to assembly of a functional NER complex require the same NER core factors in GG-NER and TC-NER. Once XPA is bound to the lesions, it recruits the DNA repair transcription complex TFIIH and opening of the DNA helix is initiated by the XPD and XPB helicases (subunits of TFIIH; Hoeijmakers, 2001; Costa *et al.,* 2003; Fousteri & Mullenders, 2008). Once the DNA has been unwound, XPG and XPF-ERCC1 are sequentially employed to make the 3' and 5' incisions, respectively, on each side of the lesion. Following incision, the damaged oligomer of approximately 24 to 32 nucleotides is excised (Shuk *et al.,* 2008). PCNA assists DNA polymerase to fill the excision gap and the nick is sealed by DNA ligase I to complete the pathway (Hoeijmakers, 2001). Recent studies suggest to the XRCC1-Ligase III complex as the principal ligase involved in the ligation step of NER throughout the cell cycle in addition to DNA ligase I that is

mainly engaged in NER during S phase (Fousteri & Mullenders, 2008; Shuk *et al.*, 2008).

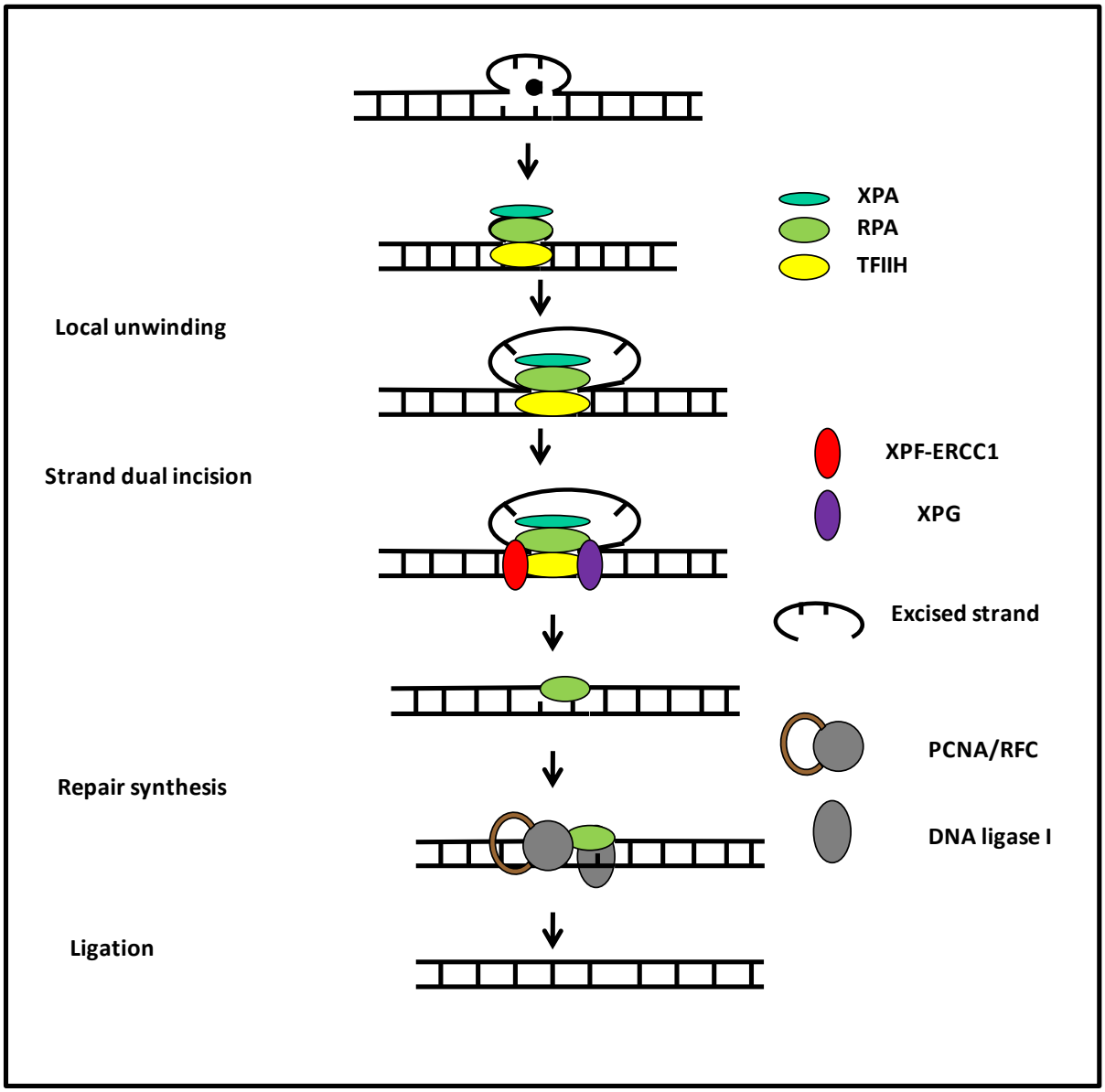


Figure 1.13: Simplified model of steps in Nucleotide excision repair (NER) after DNA lesion recognition.

NER is initiated with DNA damage recognition followed by the local unwinding of the DNA. Dual incision removes the damaged DNA fragment around the lesion; the single-strand gap is filled by DNA polymeraseδ, PCNA and RFC. The nick is sealed by DNA ligase I. Abbreviations: XPA- xeroderma pigmentosum complementation group A; RPA- Replication protein A; TFIIH - Transcription factor IIH; XPG and XPF-xeroderma pigmentosum complementation group G and F respectively; PCNA- Proliferating cell nuclear antigen; RFC- Replication factor C.

1.7.2 Inhibition of DNA repair proteins

Cell lines derived from patients with inherited DNA repair disorders are generally very sensitive to certain DNA damaging agents (alkylating agents, ionising radiations *etc.*). Various molecular genetics approaches are used to study phenotypic effects of disabling DNA repair function in mammalian systems. These include mice knock out studies, anti-sense and/or RNA interference approaches that abolish expression of gene of interest. Such studies have indicated that down regulation of DNA repair proteins sensitises cells to a specific set of damaging agents, thereby leading to the development of several potential drug targets and small molecule inhibitors of DNA repair.

1.7.2.1 Poly(ADP-ribose) polymerase 1 (PARP1)

Poly(ADP-ribose) polymerases (PARPs) constitutes a family of cell signalling enzymes which catalyses poly(ADP-ribosylation) of DNA-binding proteins. Poly(ADP-ribose) polymerase 1 (PARP1), encoded by the ADPRT (ADP-ribosyl transferase) gene, has emerged as a critical regulatory component of the immediate cellular response to DNA damage (Tentori *et al.,* 2005). PARP1 interacts with DNA ligase III, XRCC1, DNA polymerase and other components of single-strand break repair (Fortini & Dodliotti, 2007). On activation, PARP1 transfers ADP-ribose unit from NAD^+ to nuclear target proteins and itself forming long and branched polymers of poly (ADP-ribose) (PAR; Martin, 2001; Tentori *et al.,* 2005). Negatively charged polymers of ADP-ribose (PAR) attached to PARP1 itself and histones lead to chromatin relaxation, facilitating the access of base excision repair proteins and activating these repair enzymes. When DNA is moderately damaged, PARP1 is activated and participates in the DNA repair process and the cell survives. However in the presence of extreme DNA damage, PARP1 is over activated and induces depletion of cellular NAD^+ and ATP levels leading to cell dysfunction or necrosis (Madhusudan & Hickson, 2005; Tentori *et al.*, 2005).

Preclinical studies have confirmed the chemo-sensitising and radio-sensitising effects of PARP inhibitors. PARP inhibitor - AG014699 (Phase II clinical trial study) is used to potentiate cytotoxic effects of the antitumor drug, Temozolomide, (TMZ; Tentori *et*

al., 2005; Albert *et al.,* 2007). TMZ interacts with DNA and generates a wide spectrum of base adducts such a N7-methylguanine (N7-MeG) and N3-methyladenine (N3-MeA). N3-MeA and N7-MeG adducts are repaired by BER, which substitutes a single modified nucleotide (Figure 1.14). The enhancement of antitumor activity derives from the impairment of the co-ordinating function of PARP1 and the resulting nicks are left unrepaired due to the interruption of the BER process (Tentori *et al.,* 2005).

Cells with germline mutations in BRCA1 and BRCA2 genes have a strong tendency to transform to cancerous cells (Turner *et al.,* 2005). These cells are defective in homologous recombination (HR), which relies on BRCA2 to localise Rad51 to sites of DSBs and mediate DNA repair (Lord *et al.,* 2006). PARP inhibitors are lethal with BRCA2 deficiency as inhibition of PARP probably leads to persistence of DNA lesions that would normally be repaired by BRCA2 mediated HR (Turner *et al.,* 2005).

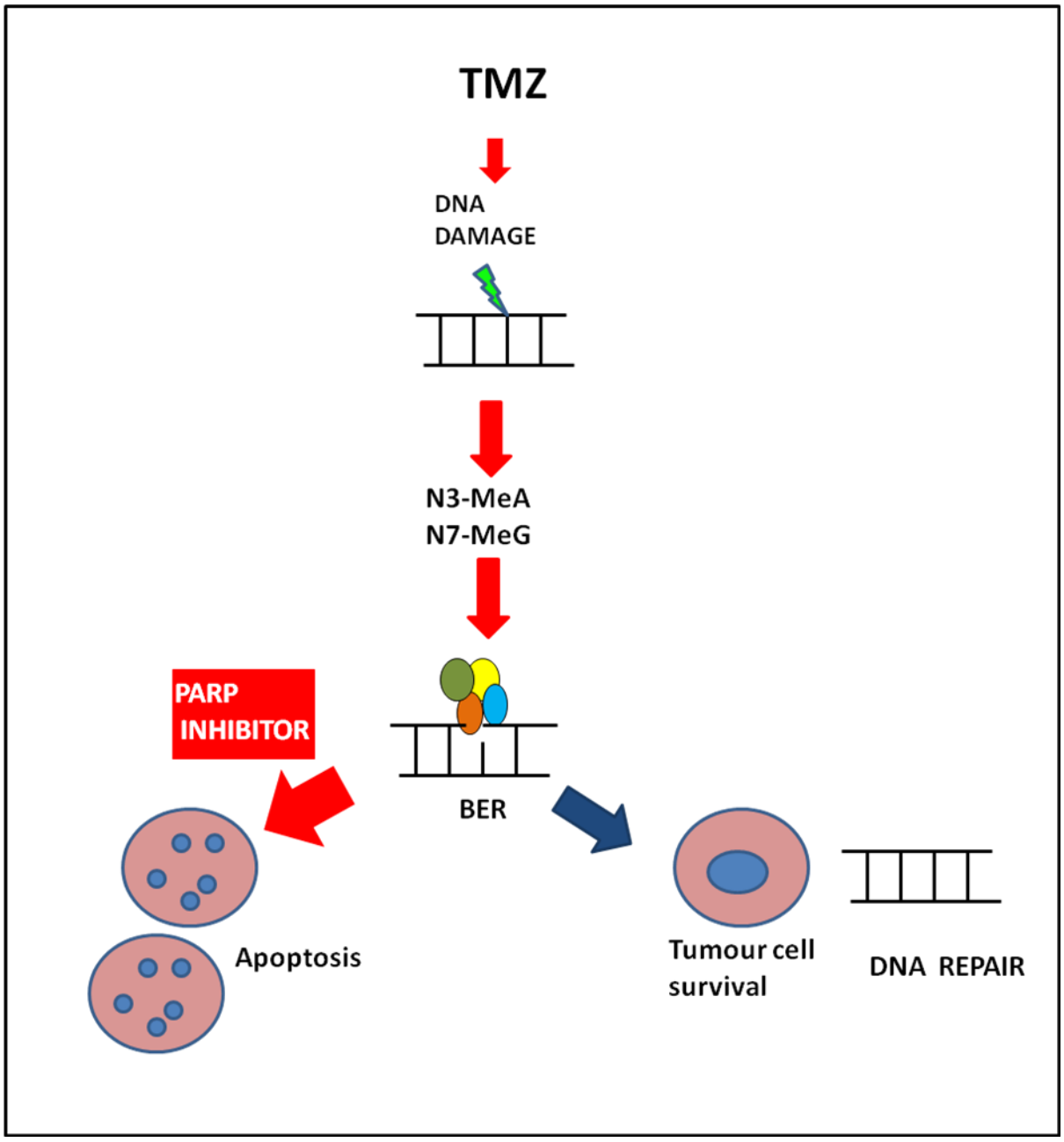


Figure 1.14: Molecular mechanisms underlying cytotoxicity induced by Temozolomode, as single agent or combined with PARP inhibitor

Temozolomode (TMZ) generates a variety of DNA adducts such as N7-methylguanine (N7-MeG) or *N*3-methyladenine (N3-MeA) which are both removed by base excision repair (BER). Inhibition of PARP prevents recruitment of BER components involved in the repair process of *N*-methylpurines; this results in generation of strand breaks and induction of apoptosis.

1.7.2.2 Base excision repair (BER) proteins

BER is able to recognise and repair the DNA damage caused by alkylating agents (Krokan *et al.,* 2000; Izumi *et al.,* 2003). The inhibition of BER sensitises cells to the cytotoxic effects of different alkylating agents. Methoxyamine (MX) is a small molecule inhibitor that binds to the AP sites and prevents their processing by APE1, leading to the accumulation of abasic sites (Figure 1.15; Damia & D'Incalci, 2007; Martin *et al*., 2008). Another small molecule inhibitor, E330 is able to inhibit the redox activity of APE1 and has been shown to increases the cytotoxic activitiy of alkylating agents in an ovarian cancer cell line (Madhusudan & Middleton, 2005). A synthetic compound palmoic acid inhibits the polymerase and lyase activities of polβ that is required to synthesise new DNA strand after the removal of damaged base (Damia & D'Incalci, 2007).

1.7.2.3 Double-strand break repair proteins

Inhibitors of ataxia-telangiectasia mutated (ATM) kinase and DNA dependent protein kinase (DNA-PKcs) are being developed as potential therapeutics for the treatment of cancer (Lord *et al.,* 2006; Salles *et al.,* 2006). A small molecule ATM inhibitor, KU55933, sensitises cells to ionising radiation, etoposide and doxorubicin (Figure 1.14; Lord *et al.,* 2006). Several small molecule inhibitors of DNA-PKcs including Vanillin and Salvicine have also been developed (Salles *et al.,* 2006). Figure 1.15 highlights the extent to which the DNA repair pathways inhibitors are being developed.

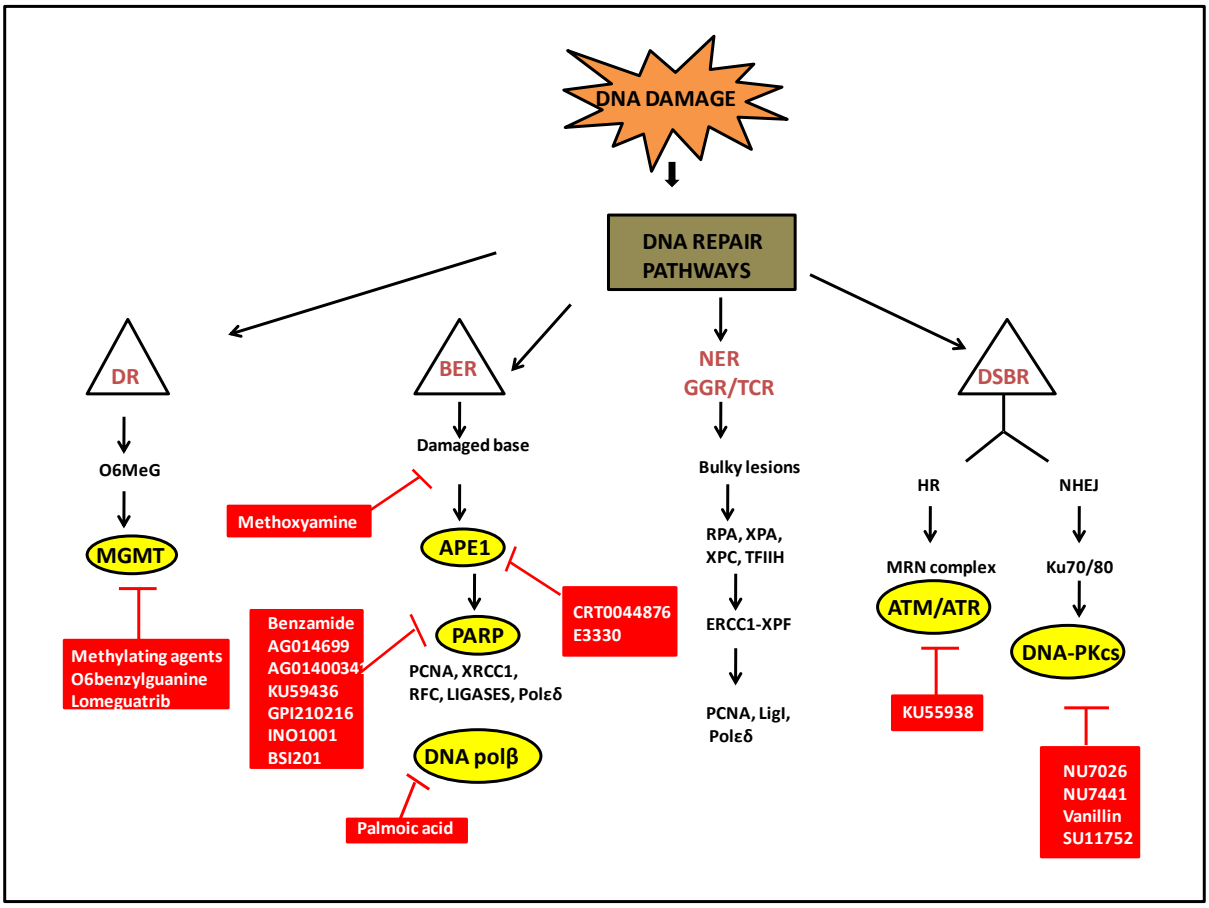


Figure 1.15: Schematic representation of the main DNA repair pathway and their inhibition under investigation

Abbreviations – DR: Direct repair; MGMT: O6-alkylguanine-DNA-alkyltransferase; BER: Base excision repair; APE1: apurinic endonuclease; PARP: poly(ADP-ribose)polymerase; NER: nucleotide excision repair; GGR: global genome repair; TCR: transcription coupled repair; DSBR: double-strand break repair; HR: homologous recombination; NHEJ: non-homologous end joining. Triangled pathways are the ones for which an inhibiting strategy has been pursued. The yellow highlighted proteins are the targets to which inhibitors (red boxes) are used for preclinical studies.

1.7.3 DNA ligases as drug targets

The inhibition of various DNA binding proteins, such as DNA gyrase (type II topoisomerase; an enzyme that unwinds double stranded DNA), is widely reported in literature (Bernard *et al.*, 1993; Chen *et al.*, 1996). Its inhibition by quinolones and aminocoumarins induces cell death by stimulating DNA damage and blocking replication processes (Chen *et al.*, 1996). Despite the predominant strategy in chemotherapy being focussed on substances that interfere with DNA metabolism, DNA ligases have received little attention as possible targets for anti-tumour agents.

After the lesions have been removed and gaps are filled with DNA polymerases by the DNA damage repair pathway, a DNA ligase finally seals the nicks into phosphodiester bonds, restoring the DNA fidelity (Timson *et al.*, 2000; Tomkinson *et al.*, 2006). Since DNA ligases in bacteria and eukaryotes use different cofactors, its binding site has been targeted by some groups to identify competitive inhibitors which can distinguish between the two kinds of DNA ligases based on the structural differences in the co-factor binding site (Gajiwala & Pinko, 2004; Srivastava *et al.*, 2005).

The presence of a distinctive N-terminal Ia domain in NAD^+-dependent DNA ligases (LigA) has made them an attractive target for the development of broad-spectrum antibacterial drugs. Brotz-Oesterhelt *et al.* (2003) have demonstrated remarkable specificity of Pyridochromanones for *E. coli* LigA compared to the ATP dependent HuLigI. Increased drug resistance in pathogenic bacteria such as *Mycobacterium tuberculosis* have lead researchers to seek to identify various small molecule inhibitors of LigA either by high-throughput screening of chemical libraries or virtual screening by computational docking of ligands into the LigA crystal structures (Srivastava *et al.*, 2005; Dwivedi *et al.*, 2008).

The unravelling of the crystal structure of HuLigI and the advances in the understanding of the molecular mechanisms involved in DNA damage repair pathways have made available potentially new targets suitable for new anti-cancer therapies or to enhance the anti-tumour activity of current DNA-damaging agents. The majority of the DNA ligase activity in proliferating cells is attributable to DNA ligase I as compared to other DNA ligases, which represent the majority of the ligase activity in resting cells (Signoret & David, 1986). Montecucco *et al.* (1992) observed a significant increase in the level of DNA ligase I after induction of cell proliferation from a resting state in HL-60 and NIH-3T3 cells whereas gene expression of the enzyme decreased in these cell lines after differentiation.

In 2001, Sun *et al.* studied the level of HuLigI in human malignant and benign tissues obtained from cancer patients, and peripheral blood lymphocytes obtained from healthy donors. The comparison was made between 29 different human tumour

specimens ranging from breast to neuroblastoma tissue and 3 benign tissues (2 breast and 1 lung). The studies concluded that the level of DNA ligase I in malignant tumours was considerably higher than in benign normal tissue and peripheral blood lymphocytes (Sun *et al.,* 2001). This study provided strong evidence that the level of DNA ligase I is considerably up-regulated in human tumour cells in comparison with normal tissues and that the activity of this enzyme may be essential for the survival of tumour cells.

Sun *et al.* (2002) studied the modulation of DNA ligase I expression by anticancer drugs that damage DNA directly or indirectly. Their studies showed that gemicitabine (a pyrmidine antimetabolite) increased the level of DNA ligase I in various human tumour cells either without affecting or with a concomitant decrease in levels of other DNA ligases (III and IV). Their study concluded that gemicitabine treatment arrests DNA replication through incorporation of gemicitabine triphosphate into replicating DNA and that this triggers an increase in DNA ligase I levels in cancer cells They observed the same effect in MiaPaCa cells treated with other anticancer agents (hydroxyurea and ara-C), which share a common mechanism of action with gemacitabine.

Recently, Chen *et al.* (2008) have conducted a virtual screen for small molecules that could dock into the N-terminal DNA binding domain (DBD) of HuLigI. They identified three compounds (L189, L67 and L82) inhibiting the second and third steps of the ligation reaction. L67 and L189 act as competitive inhibitors, preventing the interactions of the DBD of human ligase I with nicked DNA whereas L82 is an non-competitive inhibitor, analogous to campothecin (DNA topoisomerase I inhibitor). Campothecin binds to topoisomerase I and DNA (the covalent complex) resulting in a ternary complex. This prevents DNA re-ligation and therefore causes DNA damage (Bodley & Shapiro, 1995).

Other anti-tumour drugs like Doxorubicin are potent inhibitors of DNA ligases having an IC_{50} (μM) of 1.3-0.3 for *E. coli*. 2-0.4 for T4 ligase and 1.8-0.4 for HuLigI (Ciarrocchi *et al.,* 1999).

Tan *et al.* (1996) identified HuLigI inhibitors from a natural product library derived from plant extracts. They isolated certain inhibitors, such as ursolic and oleanolic acids, which disrupted the activity of enzyme by interfering with the initial adenylation step of the ligation reaction. The isolated molecules did not disrupt the DNA relaxation of supercoiled plasmid indicating that, although the same enzyme active site may be involved in both enzyme adenylation and DNA relaxation, inhibitors may exert allosteric effects by inducing conformational changes that disrupt only one of these activities (Tan *et al.*, 1996).

Whilst the inhibitory effect of several classes of natural-product, anti-leukaemic drugs, including anthracyclines, the *Catharanthus* (*Vinca*) alkaloids and podophyllotoxin against purified DNA ligase from both normal and leukaemic lymphocytes has already been described in the literature (Yang *et al.*, 1992; Tan *et al.*, 1996; Ciarrocchi *et al.*, 1999; Chen *et al.*, 2008), ligase inhibition has not been seen as the primary reason for the efficacy of these drugs.

The recent identification of a single replicative DNA ligase from parasitic protozoans - *Plasmodium falciparum* (*P. falciparum*; Buguliskis *et al.*, 2007) and *Trichomonas vaginalis* (*T. vaginalis*; discussed in detail in Chapter 3) suggests that these organisms are reliant on a single ligase for all DNA repair and replicative activities. This may be a potential drug target if a suitable compound could be found that would selectively inhibit the parasite enzyme and not those of the host.

1.9 Aims

The work described herein had three aims:

(i) To develop a high-throughput, non-electrophoretic ligase inhibitor screen that could be used to identify molecules that can selectively target DNA ligase enzymes.

(ii) To identify and produce ligase enzymes, for use in the above assay, whose inhibition could potentially have medical benefit. This would include the human replicative DNA ligase I as a potential cancer therapeutic and the single DNA ligase enzyme from a parasitic protozoan.

(iii) To validate the *Trichomonas vaginalis* DNA ligase as a drug target the enzyme required to be cloned, purified and characterised, prior to being put through into the inhibitor screen.

CHAPTER 2

GENERAL MATERIALS AND METHODS

CHAPTER 2: GENERAL MATERIALS AND METHODS

Standard molecular biology techniques and protocols used in this thesis, based on those described by Sambrook and Russell (2001), are summarised below. All chemicals were supplied by Sigma and were of analytic grade unless otherwise indicated.

2.1 DNA Isolation and Manipulation

2.1.1 Preparation of competent cells and transformation of DNA

Competent *E. coli* cells (DH5α and BL21 DE3) were prepared using a modified calcium chloride method (Dagert & Ehrlich, 1979). The cells were streaked onto Luria Agar (LA: 1% (w/v) bactotryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, 0.2% (w/v) bactoagar) and incubated overnight at 37^{o}C. A single colony was inoculated into Luria Broth (LB: 1% bactotryptone (w/v), 0.5% NaCl (w/v), 0.5% yeast extract (w/v), pH 7.4), and grown overnight at 37^{o}C with shaking. The overnight *E. coli* culture was diluted 1 in 100 into fresh LB and grown at 37^{o}C with shaking until early log phase (Optical Density at 595 nm of 0.4-0.5) was reached. The cells were harvested by centrifugation at 16,000 x g for 5 minutes at 4^{o}C (all subsequent steps were performed at 4^{o}C). The cells were re-suspended in one culture volume of sterile, ice-cold 0.1 M $MgCl_2$ and incubated for one minute before pelleting at 16,000 x g for 5 minutes. The cells were re-suspended in one half of the culture volume in sterile, ice-cold 0.1 M $CaCl_2$ and incubated on ice for 30 minutes. The cells were again harvested at 16,000 x g, re-suspended in one tenth of the culture volume of sterile, ice-cold, 0.1 M $CaCl_2$ and 30% glycerol then stored in 100 µl aliquots at -80^{o}C.

Plasmid DNA or ligation reactions were added to 100 µl of competent cells and incubated on ice for 20 minutes. The cells were heat-shocked at 42^{o}C for 90 seconds and returned to ice for 5 minutes. LB was added (0.5 ml) and the cells were incubated at 37^{o}C for at least one hour and subsequently plated onto LA plates containing the appropriate antibiotic (filter sterilised; 0.2 µm filter).

2.1.2 Small-scale preparation of plasmid DNA (Qiagen modified alkaline lysis)

Plasmid DNA was prepared using a modified alkaline lysis method (Birnboim & Doly, 1979) designed by Qiagen (USA). *E. coli* DH5α cells transformed with the plasmid of interest were grown overnight at 37^{o}C, with shaking, in 5 ml LB containing the appropriate antibiotic. The cells were harvested by centrifugation and the bacterial pellet was resuspended in 250 µl buffer P1 (50 mM Tris-HCl pH 8.0 and 10 mM EDTA) supplemented with RNase A 100 µg/ml. Buffer P2 (250 µl; 200 mM NaOH and 1% SDS, w/v) was added followed by buffer N3 (350 µl; the composition of the buffer is proprietary of Qiagen), the tubes were gently inverted to mix and centrifuged in a micro centrifuge for 10 minutes at 16,000 x g. The supernatant was applied to a QIAprep spin column (blue column) and centrifuged at 16,000 x g for 60 seconds. The columns were washed by application of 0.75 ml buffer PE (the composition of the buffer is proprietary of Qiagen) and centrifugation at 16,000 x g for 60 seconds followed by another 60 second centrifugation to remove any residual wash buffer. The column was placed in a clean 1.5 ml eppendorf tube and 30-50 µl of sterile distilled water was added to the centre of the column. The column was incubated at room temperature for one minute before elution of the DNA by centrifugation at maximum speed for 60 seconds.

2.1.3 Resolving of DNA by agarose gel electrophoresis (AGE)

Molecular biology grade agarose was melted in 1 x Tris-Borate-EDTA buffer (89 mM Tris, 89 mM boric acid and 2.0 mM EDTA) or 1 x Tris-acetate-EDTA buffer (40 mM Tris, 1.14% (v/v) glacial acetic acid and 1 mM EDTA) to a final concentration of 0.7% or 1% w/v agarose. The solution was cooled until it could be hand-held and ethidium bromide was added to a final concentration of 0.5 µg/ml. The solution was poured into a sealed gel platform and allowed to set. DNA loading buffer (6x loading buffer: 0.25% bromophenol blue, 30% v/v glycerol) was added to the DNA in a 1:5 ratio. A commercial DNA molecular weight ladder was loaded into the first lane of all

gels. The agarose gels were run at 100-110 V for one hour in 1 x TBE or 1 x TAE buffer and the DNA was visualised by UV light at 254 nm (long wave length UV light, 320 nm, was used when purifying DNA from agarose gels). DNA fragments smaller than 300 bp were resolved on 2% agarose gels.

2.1.4 Purification of DNA

2.1.4.1 QIAQUICK gel extraction

DNA fragments were resolved by 2% AGE and gel purified using the Qiagen gel purification kit. The DNA fragment was excised from the gel, three volumes of Buffer QG (the composition of the buffer is proprietary of Qiagen; contains guanidine thiocyanate) was added (three volumes of buffer: 1 volume of gel) followed by incubation at 50^{o}C until the gel slice had dissolved. One gel volume of isopropanol was added and the sample was mixed. 800 µl of the sample was applied to the QIAquick DNA affinity column (purple column) and this was centrifuged at 16,000 x g for 60 seconds. The flow through was discarded and the column washed with 750 µl of Buffer PE before centrifugation at 16,000 x g for 60 seconds. The flow through was again discarded and the column centrifuged at 16,000 x g for 60 seconds to remove any residual buffer. The column was placed in a clean 1.5 ml eppendorf tube and 30-50 µl of sterile distilled water was added to the centre of the column. The column was incubated at room temperature for one minute before elution of the DNA by centrifugation at maximum speed for 60 seconds.

2.1.4.2 QIAQUICK DNA purification following PCR or restriction digest

DNA from PCR (2.1.9) or restriction digest reactions (2.1.7) was purified using the QIAquick DNA purification protocol (Qiagen). Five volumes of Buffer PBI (the composition of the buffer is proprietary of Qiagen; contains guanidium chloride and propan-2-ol) was added, the sample was mixed and applied to the QIAquick column before centrifugation at 16,000 x g for 60 seconds. The flow through was discarded and the column washed with 750 µl of Buffer PE before centrifugation at 16,000 x g for 60 seconds. The DNA sample was processed as before (2.1.4.1) and eluted with 30 µl sterile distilled water.

2.1.5 Genomic DNA extraction

72-hours old *Trichomonas vaginalis* culture (an isolate from UCLH called UCLHI was obtained from Dr. Pamela Greenwell's laboratory who received these from Professor Akers from the London School of Hygiene and Tropical Medicine) was spun at 5000 x g for 15 minutes. The supernatant was discarded and the pellet resuspended in 1 ml lysis buffer (10 mM Tris-HCl pH 7.4, 5 mM EDTA, 10% SDS and 10 µg/ml proteinase K) preheated to 60^{o}C and then incubated at this temperature for 30 minutes (Rosl, 1992). The lysed material was then incubated with 1 ml of phenol: chloroform (1:1 v/v), inverted to mix and centrifuged in a microcentrifuge for 10 minutes at 16,000 x g. The top layer was retained, placed in a clean eppendorf tube and an equal volume of phenol: chloroform was added and processed as before. The top layer was again retained; three volumes of ice-cold 95% ethanol and a 1 in 10 volume of 3 M sodium acetate were added, thoroughly mixed and stored at -20^{o}C overnight. The re-suspended cells were thawed and centrifuged at 16,000 x g for 10 minutes. The supernatant was discarded and the DNA was washed with 100 µl 70% ethanol without mixing, the pellet and wash were then centrifuged at 16,000 x g for 5 minutes. The supernatant was discarded and the DNA pellet washed with 100 µl 95% ethanol and without mixing, these were centrifuged at 16,000 x g for 10 minutes. The supernatant was discarded and the pellet air-dried at 37^{o}C for 30 minutes before being dissolved in 20µl sterile distilled water.

2.1.6 cDNA Preparation

2.1.6.1 SV Total RNA isolation

HT-29 cells (metastatic colorectal cancer cell line) or MDA MB435 cells (breast carcinoma cell line), used to isolate total RNA using the SV RNA isolation kit (Promega), were a kind gift from Dr. M. Dwek's laboratory, University of Westminster. 175 µl of SV RNA lysis buffer (4 M guanidine thiocyanate, 10 mM Tris pH 7.5 and 0.97% β-mercaptoethanol) was added to 30 mg of frozen cells and mixed by inversion (Chirgwin *et al.,* 1979). 350 µl of SV RNA dilution buffer (blue buffer,

the composition of the buffer is proprietary of Promega) was added to the sample which was then placed in a water bath at 70^{o}C for 3 minutes, followed by centrifugation at 16,000 x g for 10 minutes. The clear lysate was transferred to a clean 1.5 ml eppendorf tube and mixed with 200 µl of 95% ethanol. The mixture was transferred to an RNA-affinity, spin column and centrifuged at 16,000 x g for 60 seconds. The flow through was discarded and column was washed with 600 µl of SV RNA wash solution (60 mM potassium acetate, 10 mM Tris pH 7.5 and 60% ethanol) before centrifugation at 16,000 x g for 60 seconds. The flow through was again discarded. The column was incubated for 15 min with 50 µl of DNase solution (yellow buffer, containing DNase I) at room temperature. The DNase I reaction was stopped by adding 200 µl of DNase stop solution and the column was centrifuged for 60 seconds at maximum speed. The column was washed twice with 250 µl of RNA wash solution before centrifugation at 16,000x g for 60 seconds. The column was placed in a clean 1.5 ml eppendorf tube and 50-70 µl of nuclease free water (supplied by Promega) was added to the centre of the column to elute the RNA by centrifugation at maximum speed for 60 seconds. The eluted RNA was stored at -80^{o}C.

2.1.6.2 PolyATtract® mRNA Isolation following total RNA isolation

The mRNA was extracted from the total RNA using the Poly A Tract® mRNA Isolation system (Promega). 500 µl of nuclease free water was added to 50 µl of previously isolated total RNA and incubated at 65^{o}C for 10 minutes. The sample was mixed gently with 3 µl of biotinylated-oligo(dT) probe and 13 µl of 20 X SSC (containing 87.7 gm NaCl and 44.1 gm sodium citrate) for annealing to the 3´ poly(A) region of the mRNA in a total RNA sample (Moyer & Henderson, 1983).

Streptavidin Paramagnetic Particles (SA-PMPs) were mixed gently by inversion, until the particles were completely dispersed. The SA-PMPs were captured by placing the tube in the supplied proprietary magnetic stand and the supernatant was carefully removed. The SA-PMPs were washed three times with 0.5 X SSC buffer (300 µl per wash); every wash followed by capturing the SA-PMPs using the magnetic stand, and removing the supernatant. The washed SA-PMPs were resuspended in 100 µl of 0.5 X SSC, mixed with the contents of the biotinylated-oligo(dT) annealing reaction and

incubated for 10 minutes at room temperature. The SA-PMPs were captured and washed four times with 0.1 X SSC (300 µl per wash) as before. The final SA-PMP pellet was resuspended in 100 µl of RNase-free water, magnetically re-captured and the eluted mRNA was transferred to a sterile 0.5 ml eppendorf tube.

2.1.6.3 The Universal RiboClone® cDNA synthesis following mRNA isolation

A reverse transcriptase (RT) reaction was performed with the previously isolated mRNA for cDNA synthesis using universal riboclone® cDNA synthesis (Promega). A sample was prepared by mixing 13 µl of isolated mRNA (~2 µg) with 2 µl of oligo (dT) primer or gene specific primers or random primers (0.5 mg/ml each) and heated at 70^{o}C for 5 minutes, followed by cooling on ice (Okayama & Berg, 1979). First strand synthesis was initiated by adding 5 µl of first strand 5 X buffer (contains 5 mM each of dATP, dCTP, dGTP and dTTP) and 1 µl of RNasin ribonuclease inhibitor to the sample, incubated at 42^{o}C for 5 minutes and followed by the addition of 2.5 µl sodium pyrophosphate (40 mM) and 1.5 µl of AMV reverse transcriptase. The reaction was incubated at 42^{o}C for 60 minutes and cooled on ice.

Second strand synthesis was initiated by mixing 40 µl of second strand 2.5 X buffer, 3 µl of DNA polymerase I and 0.5 µl of RNase H to 20 µl of previously synthesised first strand reaction and incubated at 14^{o}C for 2 hours**.** The sample was heated to 70^{o}C for 10 minutes before adding 0.5 µl of T4 DNA ligase. The reaction sample was again incubated at 37^{o}C for 10 minutes and the reaction was stopped by adding 10 µl of 200 mM EDTA. cDNA was extracted from the solution by adding an equal volume of phenol:chloroform:isoamyl alcohol (in a ratio 25:24:1), mixed thoroughly, and centrifuged at 16,000 x g in a microcentrifuge for 2 minutes at room temperature. The aqueous phase was transferred to a fresh 1.5 ml eppendorf tube, mixed with 0.5 volumes of 7.5 M ammonium acetate and 2 volumes of cold (-20°C) 100% ethanol and stored at -70°C for 30 minutes followed by centrifugation at 16,000 x g for 5 minutes. The supernatant was removed and pellet was mixed with 500 µl of cold (-20°C) 70% ethanol and centrifuged at 16,000 x g for 2 minutes. The supernatant was discarded and the pellet air-dried at 37^{o}C for 30 minutes before being dissolved in 20 µl sterile distilled water.

2.1.7 Restriction of plasmid DNA

Plasmid DNA was prepared (as per section 2.1.2) and restricted with restriction endonucleases (New England Biolabs, USA) for 2 hours (single digest - one restriction endonuclease) or 4 hours (double digest - two restriction endonucleases). Plasmid DNA (300 ng, estimated by agarose gel electrophoresis) was incubated with the appropriate restriction buffer (New England Biolabs, to a final concentration of 1 X) and 10 U of each restriction endonuclease, sterile distilled water was added to correct the final volume and the reaction was incubated at 37^{o}C for the appropriate time. When restricting plasmid DNA that was to be used for the insertion of DNA fragments, Calf intestine phosphatase (CIP; New England Biolabs) was added (1 U) thirty minutes prior to the end of the digest reaction to prevent re-circularisation of the plasmid vector. The restriction digest was resolved by 1% or 0.7% agarose gel electrophoresis as described in section 2.1.3. Following restriction digest, the plasmid DNA or the gene of interest was purified via gel extraction (2.1.4.1) or via the QIAquick DNA purification protocol (2.1.4.2).

2.1.8 Ligation of DNA fragments

Prior to ligation, DNA fragments generated by restriction digest were purified as above (2.1.4). The DNA fragments were ligated into the plasmid of interest using T4 DNA ligase (New England Biolabs). DNA fragments (500 ng; estimated by agarose gel electrophoresis) were incubated in ligase buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1.0 mM ATP, pH 7.5) with restricted plasmid DNA (200 ng; in an insert to vector ratio of 2.5:1), 0.175 U T4 DNA ligase (0.25 U/ μg of DNA) and sterile distilled water was added to correct the final volume. The ligation reactions were incubated over night at 4^{o}C (Olivera & Lehman, 1967) and the products were transformed into competent *E. coli* DH5α cells.

2.1.8.1 Ligation of PCR products into pGEM-T

PCR products were inserted into pGEM-T following the standard protocol (Promega). The PCR product (100 ng; estimated by agarose gel electrophoresis) was incubated

overnight at 4^{o}C with ligation buffer (2 X; 60 mM Tris-HCl pH 7.8, 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP, 10% polyethylene glycol), pGEM-T vector (50 ng), 3 U T4 DNA ligase and sterile distilled water to the correct final volume. The ligation reaction was transformed into *E. coli* JM109 cells (Promega) competent cells and plated onto LA containing the appropriate antibiotic as well as 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and 80 µg/ml of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal; dissolved in 2 ml N, N-dimethyl formamide). Colonies were selected by blue/white colony screening to distinguish recombinant (white) colonies from non-recombinant ones (blue). White colonies were selected for DNA isolation as previously described (2.1.2) and the DNA insert was released by digestion with *Nde* I and *Xho* I as previously described (2.1.7). Control reactions to assess the formation of background blue colonies were also carried out. A positive control ligation reaction was prepared with 50 ng of the control plasmid, pGEM-luc vector DNA (Promega). The control plasmid contains multiple stop codons in all six reading frames, thus a low background of blue colonies should be obtained. A negative control reaction was also performed with 50 ng of pGEM-T plasmid only.

2.1.9 Polymerase Chain Reaction (PCR)

PCR was applied for the exponential *in vitro* amplification of particular DNA sequences by using sequence-specific synthetic oligonucleotide primers obtained from Invitrogen. A reaction mixture containing PCR master mix (2x Promega: 50 U/ml Taq DNA polymerase in Promega proprietary reaction buffer, 400 µm dATP, 400 µm dGTP, 400 µm dCTP, 400 µm dTTP and 3 mM $MgCl_2$), template DNA (3 ng), forward and reverse primers (0.5 µM each) was prepared (Bartlett & Stirling, 2003). The PCR solution was overlaid with 200 µl of mineral oil and cycled in a thermocycler (Perkin-Elmer) as follows: an initial denaturation step of one cycle at 95^{o}C for one minute followed by 30 cycles of: denaturation, 94^{o}C for one minute, primer annealling, 55^{o}C-65^{o}C for one minute (varied with the annealing temperature of the specific primers used in amplification reaction), polymerase extension 72^{o}C for one minute and a final 7 minute extension at 72^{o}C. The reactions were cooled to 4^{o}C for overnight storage or to room temperature and gel purified as described in section 2.1.4.

2.1.10 Radiolabelling of DNA substrate

A labelling reaction was prepared as described by Shuman & Ru (1995), containing 50 pmole of radioactive label (^{32}P; GE Healthcare), 1 mM ATP (1800 pmole), 1300 pmole DNA substrate (18-mer; figure 3.15, panel B), kinase buffer (60 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10 mM DTT) and 10 U T4 polynucleotide kinase (New England Biolabs). The reaction was incubated at 37^{o}C for one hour and terminated by heating to 60^{o}C for 10 minutes. A small amount of 10% glycerol was added to the DNA solution and the radiolabelled DNA was separated from unincorporated label by 20% native polyacrylamide gel electrophoresis in 1 X TBE at 400 V for one hour. No DNA loading buffer (composition 2.1.3) was added to the DNA as the dye and unincorporated label bind irreversibly and can contaminate the labelled DNA sample, instead, DNA was mixed with 1 mM EDTA and 5 µl formamide. The density of the glycerol is sufficient to carry the DNA into the well. The labelled 18-mer was located by autoradiography of the wet gel. The developed autoradiograph was placed underneath the glass plate to allow the DNA to be excised from the gel in an acrylamide slice. The DNA/acrylamide slice was then incubated overnight, with rocking at 4^{o}C in distilled water containing 150 mM NaCl. The supernatant containing the DNA was then centrifuged at 16,000 x g. Three volumes of ice cold 100% ethanol were added and the DNA was incubated at -20^{o}C for two hours. Following this, the pellet was washed with ice cold 70% ethanol and the fractions were monitored for radioactive activity. The pellet was air-dried and dissolved in 100 µl sterile distilled water. A 1/100 dilution (1 µl label dilution in 99 µl distilled water) of the radioactive label was prepared at stored at -20^{o}C to determine the rate of decay of the radioactivity and to allow comparison with the incorporated material to guage the quantity of label incorporated.

The amount of labelled substrate was measured using a Beckman coulter LS6500, multipurpose scintillation counter. 1 µl of the 1/100 dilution of the radioactive label was counted to determine the counts per pmole of label. 1 µl of the excised labelled DNA was counted and the amount of labelled substrate was determined using the following equation:

$$\frac{\text{Count of 1µl of eluted substrate x substrate volume x 100/2.7*}}{\text{Counts per pmol of label}}$$

***(100/2.7 is the proportion of the label of the total amount of ATP used in the kinase reaction)**

2.1.11 Annealing of DNA substrates

In order to determine the ratio for annealing, the A_{260} of each DNA substrate was determined and the concentration calculated according to the following formula:

$$\mathbf{C = A_{260} \times fc}$$

C = concentration of DNA

A_{260} = absorbance of DNA at 260 nm

f = dilution factor

c = concentration (standard)/ absorption (standard) dsDNA = 0.05 µg/ml

ssDNA = 0.03 µg/ml

The DNA substrates were added in the appropriate ratio (4:1 unmodified: modified for radioactively labelled substrates) to ensure that every labelled substrate was part of a productive nick and could therefore be utilised in a ligation reaction. Each substrate was heated at 65^{o}C for five minutes in the presence of annealing buffer (50 mM Tris pH 7.5, 200 mM NaCl, 5.0 mM DTT, pH 7.5). The hairpin substrate was snap-cooled on ice before addition of the downstream 5'phosphate labelled oligonucleotide. The reaction was slowly cooled to room temperature and incubated on ice. The annealed DNA was then either immediately used or stored at -20^{o}C until required.

2.1.12 Ligation Assay using Radiolabelled DNA substrate

The standard DNA ligation assays were performed by incubating DNA ligase with the ^{32}P-labelled DNA substrate (1 pmole) for 2 min at 37^{o}C in ligation buffer, containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP in a total

volume of 20 µl. The reaction was terminated by the addition of 20 µl of loading buffer [98% (v/v) formamide, 10 mM EDTA, 0.01% (w/v) bromophenol blue]. Reaction samples were heated at 95^{o}C for 5 minutes and electrophoresed through 20% polyacrylamide 7 M urea gel in 1 X Tris-borate-EDTA buffer. The gel was run at 300-400 V for 2-3 hours and exposed to X-ray film or analysed with a Phosphorimager (GE Healthcare) and the intensity of the bands analysed with an Imagequant 5.2 (Typhoon scanner, GE Healthcare).

2.2 Protein Purification

2.2.1 Protein expression

E. coli BL21 (DE3) cells were transformed with the plasmid of interest as before and inoculated into LB containing the appropriate antibiotic. The culture was incubated overnight at 37^{o}C with shaking. 500 ml LB was inoculated with a 1 in 100 dilution of the overnight culture; adjusted to 200 µg/µl antibiotic and incubated at 37^{o}C with shaking. When the optical absorbance at a wavelength of 595 nm (A_{595}) reached 0.4-0.5, the cultures were adjusted to 0.4 mM IPTG and incubated for a further 2 hours at 37^{o}C. The cells were harvested by centrifugation with a Sorvall Super T21 centrifuge (SL250T rotor) at 10,000 x g for 10 minutes at 4^{o}C. The supernatant was discarded and the bacterial cell pellet re-suspended in Buffer A (25 ml/l of culture; 50 mM Tris-HCl pH 7.0, 300 mM NaCl and 10% glycerol) and stored at -80^{o}C.

2.2.2 Cell lysis and protein solubilisation

All steps were performed at 4^{o}C. The re-suspended bacterial cell pellets were thawed on ice, adjusted to 2.0 µg/ml lysozyme and 2.0 mM phenylmethylsulphonyl fluoride (PMSF) and incubated for 30 minutes. Triton X-100 was added to 0.1% (v/v) and the cells were incubated for a further 30 minutes. The lysate was sonicated with six, 30 second bursts at 40% power (Status 200, MS73 probe) to reduce viscosity followed by ultracentrifugation to remove cellular material in a Sorvall *Discovery* 90SE centrifuge and T865 rotor at 30,000 x g for one hour. The insoluble pellet fraction was resuspended in buffer A and retained for sodium dodecyl sulphate polyacrylamide gel

electrophoresis (SDS-PAGE) analysis and the enzyme in the soluble supernatant fraction was purified.

2.2.3 Protein purification using Ni-NTA agarose resin

The pre-charged nickel-nitrilotriacetic acid (Ni-NTA) agarose resin was prepared as per the manufacturer's instructions (Ni-NTA coupled to Sepharose CL-6B; QIAexpressionist Handbook, Qiagen, 2003). A 5.0 cm column was poured, the storage buffer was removed and the column washed with 5 column volumes of distilled water followed by 5 volumes of binding buffer (50 mM Tris-HCl pH 7.0 or pH 7.5, 300 mM NaCl and 10 mM imidazole). The soluble protein fraction was loaded onto the column and the flow through retained. The column was washed with 5 volumes of wash buffer (50 mM Tris-HCl pH 7.0 or pH 7.5, 300 mM NaCl and 30 mM imidazole); 5.0 ml fractions were collected. After optimising protein elution along an imidazole gradient (50 mM Tris-HCl pH 7.0 or pH 7.5, 300 mM NaCl and 100 – 500 mM imidazole), subsequent protein preparations were eluted with elution buffer (50 mM Tris-HCl pH 7.0 or pH 7.5, 300 mM NaCl and 300 mM imidazole).
After use, ionically-bound proteins were removed by washing the column with one volume of 2.0 M NaCl. Precipitated proteins, hydrophobically bound proteins and lipoproteins were removed by washing the column with three volumes of 0.5 M NaOH. The column was then washed with three column volumes of the elution buffer followed by three column volumes of binding buffer and re-used a maximum of three times. Following elution from Ni-NTA agarose, the protein was adjusted to 5.0 mM DTT and subjected to ion-exchange chromatography.

2.2.4 Western blotting of proteins

The transfer of proteins onto nitrocellulose was performed using the method described by Towbin *et al.* (1979) with a Bio-Rad western apparatus. The transfer was performed at 4^{o}C for one hour (100 V) using a freshly prepared transfer buffer (25 mM Tris, 192 mM glycine and 20% methanol). Proteins transferred onto the nitrocellulose membrane were visualised using Ponceau S stain (0.5% Ponceau S w/v

and 1% glacial acetic acid v/v). The membrane was de-stained by multiple washes in sterile distilled water. Following transfer of proteins and staining, the nitrocellulose membrane was washed with Tris-buffered saline (TBS)/Tween (TBS: 50 mM Tris pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween 20) and blocked overnight, at 4^{o}C with 2% bovine serum albumin (BSA) (dissolved in TBS/Tween). The membrane was incubated with 5 μg/ml biotinylated anti-His antibody (Sigma, Germany) for one hour at room temperature. Following this, the membrane was washed three times for five minutes with TBS/Tween to remove any unbound antibody and then incubated with 2 μg/ml horse radish peroxide (HRP) conjugated streptavidin (Pierce) for one hour. The detection of His-tagged proteins was performed by adding diaminobenzidine (DAB) in TBS/H_2O_2 using the method derived by Krajewski *et al.* (1995). The nitrocellulose membrane was again washed three times for five minutes in TBS/Tween.

2.2.5 Glycerol gradient sedimentation

An aliquot (70 μg) of the purified, wild type TV DNA ligase was mixed with catalase (50 μg), bovine serum albumin (BSA; 50 μg), and cytochrome *c* (50 μg). The mixture was applied to a 5 ml, 15-30% glycerol gradient containing 50 mM Tris–HCl pH 7.0, 500 mM NaCl, 2 mM DTT, 1 mM EDTA, and 0.01% Triton X-100 (Tanese, 1997). The gradient was centrifuged in a Sorvall SW55 rotor at 368,000 x g for 17 hours at 4°C. Fractions (~0.2 ml) were collected from the bottom of the tube. The polypeptide compositions of the gradient fractions were analyzed by SDS-PAGE.

CHAPTER 3

CHARACTERISATION OF *T. vaginalis* DNA LIGASE

CHAPTER 3: CHARACTERISATION OF *T. vaginalis* DNA LIGASE

3.1 Introduction

3.1.1 *Trichomonas vaginalis*

The protozoans, meaning "first animal," are an enigmatic group of single-celled eukaryotes forming the earliest-diverging lineages in eukaryotes. Although unicellular, they have a nucleus and membrane-bound organelles, making them functionally complex despite their small size (Honigberg, 1963). There are over 50,000 species of protozoa, of which ~10,000 are parasitic species. Examples of parasitic protozoa include *Trypanosoma* (causes Chagas disease and sleeping sickness); *Leishmania* (causes leishmaniasis); *Plasmodium falciparum* (causes malaria); *Entamoeba histolytica* (causes amoebic dysentery) and *Trichomonas vaginalis* (causes trichomoniasis; Sood & Kapil, 2008).

Trichomonas vaginalis (*T. vaginalis*)**,** a flagellated pear-shaped parasitic protozoan (Figure 3.1), is the etiological agent of trichomoniasis; a sexually transmitted disease (STD) of global concern leading to an estimated 180 million cases per annum (Petrin *et al.,* 1998). This disease is also linked to cervical cancer, atypical pelvic inflammatory disease infertility and predisposition to human immunodeficiency virus (HIV) infection (Petrin *et al.* 1998; Sood & Kapil, 2008). *T. vaginalis* is one of the most primitive eukaryotes. It lacks mitochondria in its cytoplasm but possesses a hydrogen-producing organelle called the hydrogenosome; of analogous function to the mitochondria of more advanced eukaryotes (Cudmore *et al.,* 2004).

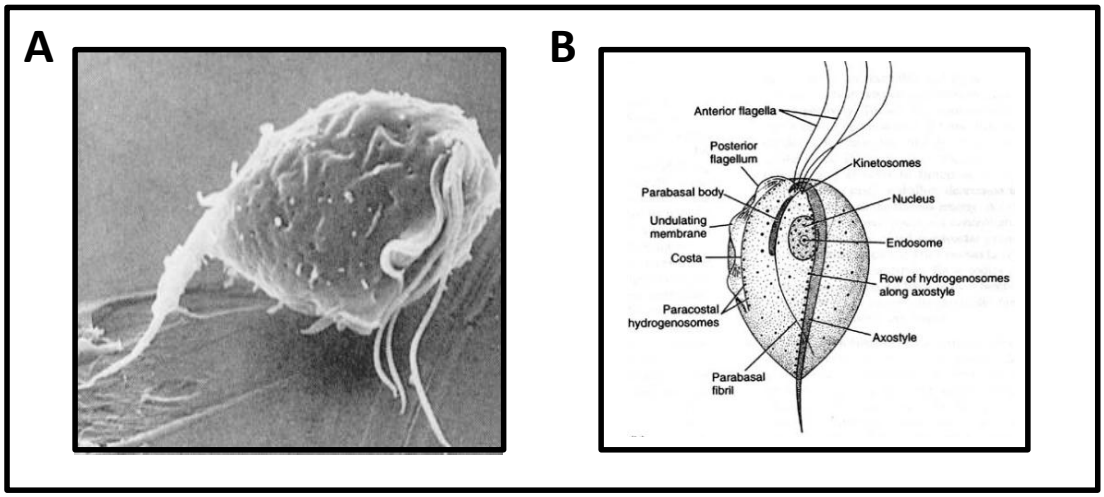


Figure 3.1: Microscopic image of *T. vaginalis* trophozite and its interior components

Panel A shows microscope image of *T. vaginalis* trophozite as seen in broth culture. The axostyle, undulating membrane and flagella are clearly visible. Panel B shows the picture of the interior components showing the position of hydrogenosomes and flagella (Petrin *et al.*, 1998; http://www.tigr.org/tdb/e2k1/tvg).

Trichomoniasis is treated with the antimicrobial agent 5-nitroimidazole, most usually metronidazole. Metronidazole is administered in an inactive form and diffuses into the hydrogenosome where it is modified to become cytotoxic by reduction via ferrredoxin-mediated electron transport (Schwebke & Burgess, 2004; Sood & Kapil, 2008).

Two hydrogenosomal proteins, pyruvate ferredoxin oxidoreductase and ferredoxin play a critical role in the reductive activation of metronidazole via the transfer of a single electron from ferredoxin to the nitro group of the drug, resulting in the formation of a free radical. The nitro radical is hypothesised to bind transiently to DNA, causing disruption and breakage of DNA strands and ultimately leading to cell death (Schwebke & Burgess, 2004).

However, strains of *T. vaginalis* resistant to metronidazole (at least 5% of clinical cases) are on the rise, emphasising the need for research into alternative antibiotics and drug targets (Sood & Kapil, 2008). In metronidazole resistant strains, the expression levels of pyruvate ferrodoxin reductase and ferrodoxin are reduced dramatically, eliminating the ability of the parasite to activate metronidazole (Cudmore *et al.*, 2004).

3.1.2 DNA damage and repair in parasitic protozoans

DNA repair processes are vital to all living organisms. Failure to remove and replace damaged bases and nucleotides results in accumulation of multiple lesions which might lead to total genome degradation and loss of vital genetic information. DNA lesions take many forms, including single-strand and double-strand breaks, in addition to inter- and intra-strand crosslinks and modified bases (Hoeijmakers, 2001). Various pathways that operate in a manner to minimise information loss during DNA lesions are conserved throughout eukaryotes, and many eukaryotic enzymes have homologues in prokaryotes (Wang *et al.*, 2003).

Depending upon the type of DNA damage involved, cells employ a specific mechanism to fix the damage. When a single-strand break arises in the genome, the Base excision repair (BER) and Nucleotide excision repair (NER) pathways are involved (Sattler *et al.*, 2003). When double-strand break arises, non-homologous end joining (NHEJ) appears to be the major pathway of repair (Wyman & Kannar, 2006).

Various reports suggest that parasitic protozoa such as *Plasmodium falciparum* (*P. falciparum*), utilise long-patch BER for single-strand break repairs but interestingly, a NHEJ mechanism for double-strand break repair has not been described in any protozoal parasites to date (Haltiwanger *et al.*, 2000; Burton *et al.*, 2007). Conway *et al.* (2002a) observed that deletion of genes encoding Ku70/ Ku80 proteins (major proteins of NHEJ pathway) had no detectable effect on DNA damage repair in the closely related parasitic protozoan, *Trypanosoma brucei*.

The recent publication of the *T. vaginalis* (hereafter referred to as TV) genome sequence has helped broaden the current understanding of the cellular and molecular pathways of the parasite (Carlton *et al.*, 2007). However, still very little is known about the mechanisms that TV employs to maintain genomic integrity and stability during replication and repair. DNA replication is essential for parasitic proliferation and a DNA ligation capability is clearly important for DNA damage repair but to our current knowledge, no reports have been published on these enzymes in Trichomonads. To date there is only a single report (published contemporaneously

with this study) of a DNA ligase identified from a member of the parasitic protozoan family that from *P. falciparum* (Buguliskis *et al.*, 2007).

In this study, we report the identification, recombinant expression, purification and biochemical characterization of a DNA ligase I from TV. TV DNA ligase (hereafter referred to as TVlig) is a 679 amino acid polypeptide with a predicted molecular weight of 76430 Da. The presence of only one ligase in TV genome suggests this as a uniquely focal point for therapeutic intervention in this organism, disruption of which will stop replication and hence propagation of the parasite during infection.

3.2 Material and methods

3.2.1 Cloning of TVlig

Genomic DNA was extracted from TV culture, obtained from Dr. Pamela Greenwell's laboratory (University of Westminster, London), as described in Chapter 2 (section 2.1.5). The coding region of TVlig was amplified by polymerase chain reaction (PCR) as described in Chapter 2 (section 2.1.9), except the PCR master mix containing Taq polymerase was replaced with Phusion master mix (Finnzymes, UK) containing *Pfu* DNA polymerase.

Two oligonucleotide primers that flanked the open reading frame of the TVlig were designed with *Bam* HI restriction enzyme sites on either side (5'-GATC*GGATCC*GATGACTCAACAGGGAATTGC and 5'-GATC*GGATCC*TTATTTCTCAAATTGTTGATGATAGAG) (Accession number: XP_001581589; Invitrogen). Amplification was performed for 30 cycles, consisting of 1 minute at 94^{o}C, 2 minutes at 55^{o}C and 3 minutes at 72^{o}C followed by final extension at 72^{o}C for 7 minutes.

The products were resolved on 0.7% agarose gel (as described in Chapter 2; section 2.1.3) and the 2050 bp product was purified from the gel using the Qiagen DNA purification kit (as described in Chapter 2; section 2.1.4.1) and eluted in 30 µl-distilled water.

3.2.2 Subcloning in pGEM-T vector

The primers obtained from Invitrogen were not phosphorylated, so the purified PCR product was phosphorylated with T4 polynucleotide kinase (New England Biolabs; NEB) prior to an A-tailing reaction required to ligate the insert with the T-overhangs of the p-GEMT vector. 20 µl of PCR product was mixed with 1 X T4 kinase buffer (the composition of the buffer is proprietary of NEB), 1mM ATP solution and 0.02 U T4 kinase enzyme. The reaction mixture was incubated at 37^{o}C for 30 minutes and the enzyme was heat inactivated at 65^{o}C for 2 minutes. Taq polymerase (NEB) was used to add an A-tail (single A nucleotides to the 3' terminus of the PCR product) to the phosphorylated product. The reaction mixture (5 units of Taq polymerase, 0.2 mM dATP, 1 X Taq polymerase buffer and 10 µl of phosphorylated PCR product) was incubated at 72^{o}C for 30 minutes and was purified using a Qiagen PCR purification kit (described in section 2.1.4.2).

A ligation reaction was set up between the pGEM-T vector and the purified, A-tailed PCR product and transformed into *E. coli* JM109 cells (Chapter 2; 2.1.1 & 2.1.8) Plasmid DNA was extracted using a Qiagen plasmid mini-prep kit (Chapter 2; 2.1.2) from overnight cultures of transformed cells grown in liquid LB culture, containing 100 µg/ml ampicillin, inoculated with a single white colony. The plasmid was digested with *Bam* HI and analysed on a 0.7% agarose gel. The released insert was extracted from the gel and then purified by using a Qiagen gel purification kit (as per Chapter 2; 2.1.4.2). The purified DNA fragment was ligated into *Bam* HI digested, CIP treated pET-16b vector (Chapter 2; 2.1.8). The ligated mixture was transformed into *E. coli* DH5α cells and plated onto LB-Ampicillin plates. Plasmid DNA was extracted using a Qiagen plasmid mini-prep kit from overnight cultures, inoculated with a single, transformed bacterial colony. The resulting clones were screened for the orientation of the *Bam* HI insert by using the plasmid as a template in a PCR reaction and T7 forward primer and reverse primer (3') of the gene under the same conditions as described earlier. The identities of the constructs were confirmed by sequencing the plasmid at the Advanced Biotechnology Centre (ABC; Imperial College, London).

3.2.3 Over-expression and Purification of TVlig

An overnight culture of *E. coli* BL21(DE3) competent cells transformed with the plasmid pET-16b/TVlig was added to fresh LB medium supplemented with 100 µg/ml ampicillin. The protein was expressed as previously described (Chapter 2; 2.2.1) and the pellet was re-suspended in buffer A containing 50 mM Tris-HCl pH 7.0, 300 mM NaCl and 10% glycerol. The re-suspended pellet was lysed as described earlier (Chapter 2; section 2.2.2) followed by centrifugation at 36,000 x g for 1 hour at 4^{o}C. The lysate was applied to Ni-NTA nickel-chelate resin (Qiagen) the protein was eluted with buffer A containing 300 mM imidazole (Chapter 2; 2.2.3). The elution profile was monitored via SDS-PAGE. The peak fraction of TVlig, determined by SDS-PAGE, was dialysed against buffer B containing 50 mM Tris-HCl pH 7.0, 10% glycerol, 5 mM DTT and 25 mM NaCl. The dialysed fraction was applied to S-Sepharose and Blue-Sepharose (GE Healthcare) resins, equilibrated with buffer B containing 25 mM NaCl. The columns were washed with buffer B containing 15 mM NaCl and the protein was step eluted with buffer B containing 0.75, 0.1, 0.2, 0.3, 0.5 and 1.0 M NaCl. The elution profile was monitored via 12% SDS-PAGE. The best protein recovery was obtained from S-Sepharose in the 0.2 M NaCl fractions. The peak fraction was dialysed against buffer B containing 100 mM NaCl. The purified protein was aliquoted in small volumes and stored at -80^{o}C until required.

3.2.4 *Hin*d III DNA ligation assays with TVlig

The purified, recombinant TVlig was used to ligate *Hin*d III digested λ DNA in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, 25 µg/ml bovine serum albumin, 5% Polyethylene glycol (PEG) 8000 at 16^{o}C for 1 hour.. The concentration of PEG 8000 required for optimal ligation by the recombinant ligase was measured by adding 0% and 5% PEG 8000 to the reaction mixture. The ligated mix was resolved on 0.8% agarose gel, stained with ethidium bromide and viewed under UV illumination.

3.2.5 DNA ligation assays using ^{32}P labelled substrate

The 18-mer oligonucleotide (Figure 3.15; panel B) obtained from Invitrogen was labelled with ^{32}P at 5'-end and the amount of labelled oligonucleotide was assessed as per section 2.1.10.

The nicked duplex DNA substrate (Figure 3.15; panel C) formed by mixing the 18-mer strand labelled with ^{32}P at its 5' end with the complementary 42-mer strand at a molar ratio of 1:4 (18-mer : 42-mer) in 0.2 M NaCl (section 2.1.11). The standard DNA ligation assays were performed by incubating TVlig (0.1 pmole) with the ^{32}P-labelled DNA substrate (1 pmole) for 2 minutes at 37^{o}C in buffer C containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT) and 1 mM ATP (section 2.1.12). The effect of pH on the extent of ligation was examined by performing assays in reaction buffer C as indicated above except that it contained Tris-HCl (pH 7-8.5) or Sodium citrate (pH 4-6.5) or MOPS [3-(N-morpholino) propanesulfonic acid (pH 6-7.5)] as the buffering agent. Reactions were also performed with a variety of divalent metals and nucleotides at the described concentrations in the reaction buffer C.

3.2.6 DNA ligase-DNA binding gel shift assay

DNA binding activity of TVlig was tested by a protein-DNA complex gel shift assay. TVlig-DNA binding assay was performed by incubating up to 2.5 pmole of TVlig with 1 pmole of nicked DNA substrate in 50 mM Tris-HCl pH 7.5 and 5 mM dithiothreitol (DTT) at 22^{o}C for 10 minutes. Glycerol (5% v/v final) was added to the samples before electrophoresing them through a non-denaturing 10% polyacrylamide gel in TBE buffer at 100 V for 2.5 hours at 4^{o}C. The wet gel was wrapped in cling film and exposed to a phosphorimager plate at -80^{o}C. The gel image was analysed with Imagequant (5.2) after scanning with a Typhoon phosphorimager (GE Healthcare).

3.2.7 Construction of active site mutant (K338A) of TVlig

A PCR-based method was used to substitute the lysine (K) in the active site of motif I (Figure 1.3) with alanine (A). Two oligonucleotide primers that overlap the lysine residue in the active site motif I of TVlig were designed and synthesised (Invitrogen) - see table 3.1.

PCR was performed in two steps in order to generate two halves of the TVlig gene with each half carrying an alteration in the codon for the predicted active site lysine, residue 338 (Figure 3.2). For each reaction, 25 µl reaction samples were prepared using 2 X Taq master mix (final concentration 1 X, Promega), 0.5 µM primers (Table 3.1) and 3 ng of pET-16b plasmid encoding the wild type (WT) TVlig.

Table 3.1: Primers used to make mutations in active site of TVlig

Primers	PCR 1	PCR 2
Forward primers	GATC*GGATCC*GATGACTCAACAGGGAA TTGC	GAAGAATCACAGGCGAATAC*G CT*TACGATGGCGAAAGAGC
Reverse Primers	GCTCTTTCGCCATCGTA*AGC*GTATTCGC CTGTGATTCTTC	GATC*GGATCC*TTATTTCTCAAA TTGTTGATGATAGAG

Underlined nucleotides represent *Bam* HI restriction sites and nucleotides in red show the mutated position encoding for alanine.

Amplification was performed for 30 cycles, consisting of 1 minute at 94^{o}C, 2 minutes at 55^{o}C and 1 minute at 72^{o}C followed by final extension at 72^{o}C for 7 minutes. The products were resolved on a 0.7% agarose gel and the ~1 Kb products were purified from the gel using the Qiagen DNA purification kit and eluted in 30 µl distilled water.

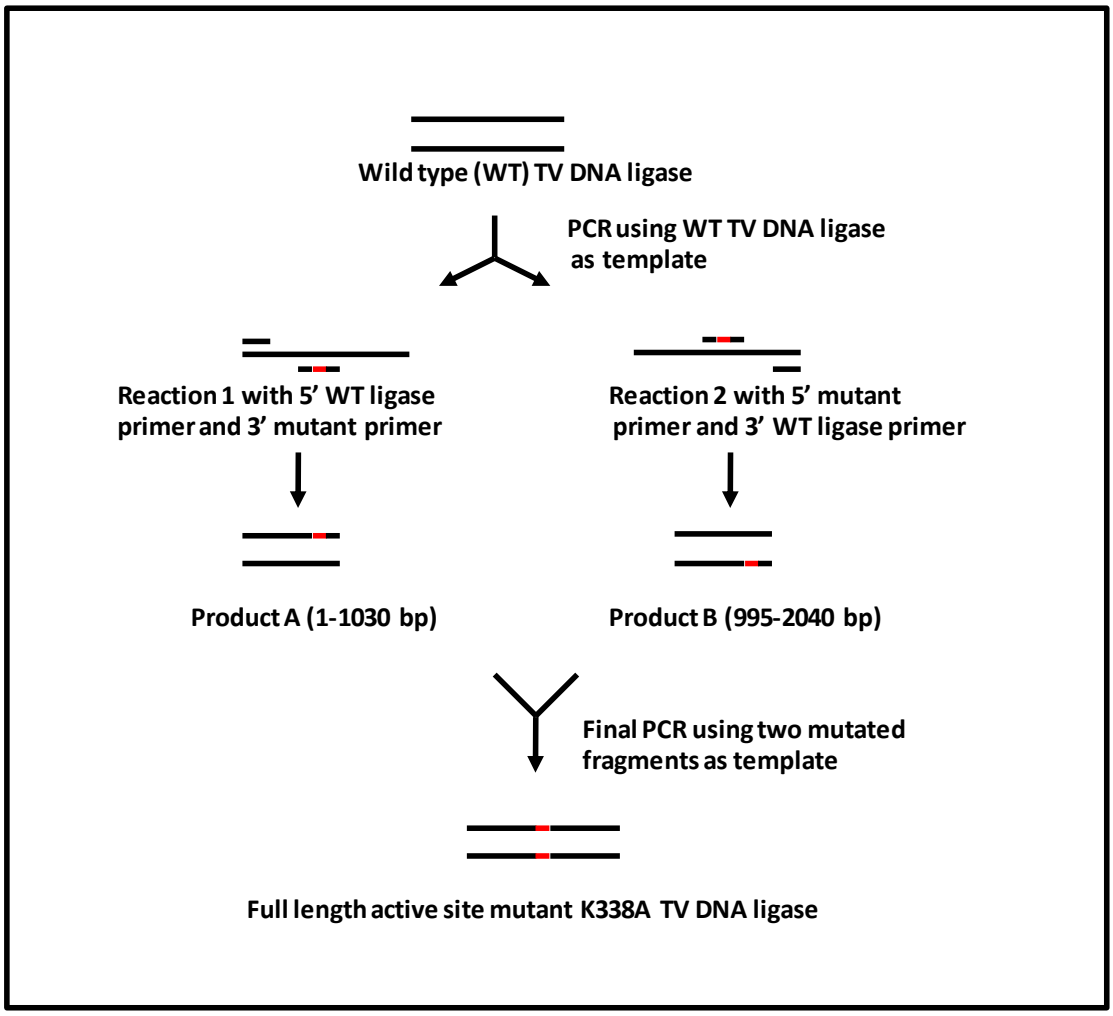


Figure 3.2: PCR based method to introduce active site mutation in TV DNA ligase.

The eluted DNA from the first two PCR reactions were mixed in the ratio 1:1 and used as a template for the final PCR amplification of full-length active site, K338A mutant of TVlig. In this reaction 25 µl samples were prepared using 2 X Taq master mix (final concentration 1 X, Promega), 0.5 µM forward and reverse primers for wild type TVlig and 3 ng of the DNA fragment mix (products A and B, Figure 3.2) obtained in the above PCR reactions using conditions as described above in section 3.2.3.

The full-length TVlig encoding product was resolved on a 0.7% agarose gel and purified from the gel using the Qiagen DNA purification kit and eluted from the DNA

affinity resin in 30 µl of sterile distilled water. The purified PCR product was digested with *Bam* HI restriction enzyme (New England Biolabs) for 1 hour at 37^{o} C and then ligated with *Bam* HI digested CIP-treated pET-16b vector for 16 hours at 4^{o}C. The ligated mixture was transformed into *E. coli* DH5α cells (Invitrogen) and plated onto LB-Amp (50 µg/ml) plates. Plasmid DNA was extracted using a Qiagen plasmid mini-prep kit from overnight cultures, inoculated with randomly chosen single, transformed bacterial colonies. The identities of the constructs containing mutants were confirmed by sequencing the plasmid at the Advanced Biotechnology Centre (ABC), London.

The confirmed mutants were transformed into *E. coli* BL21 (DE3) competent cells and protein was expressed and purified under similar conditions to those described in section 3.2.3. The purified protein was recovered from S-sepharose by elution with 0.2 M NaCl. This peak fraction was then dialysed against buffer B containing 0.1 M NaCl. The purified protein was aliquoted in small volumes and stored at -80^{o}C until required.

3.2.8 Surface Plasmon Resonance (SPR) analysis

All surface plasmon resonance (SPR) analysis was performed on the Biacore 2000 automated system (Biacore AB; Uppsala, Sweden). Prior to substrate immobilisation the Biacore system was desorbed with BIAdesorb solutions 1, 0.5% (w/v) SDS in distilled water) and 2 (50 mM glycine-NaOH pH 9.5; Biacore AB; Uppsala, Sweden) and sanitised with BIAdisinfectant solution (1.0% sodium hyperchlorite (v/v) in distilled water; Biacore AB; Uppsala, Sweden). The system was then primed with degassed and filtered (with a 0.22 µm filter) Biacore buffer 1: 50 mM Tris pH 7.5, 50 mM NaCl, 5.0 mM DTT, 5.0 mM EDTA pH 7.5 and 0.005% Surfactant p20 (Biacore AB; Uppsala, Sweden). The chip was docked and the priming protocol was repeated, the chip was then treated with 3 – 4 one minute pulses of 1.0 M NaCl in 50 mM NaOH to pre-condition the surface and normalised with BIAnormalising solution (70% glycerol v/v in distilled water; Biacore AB; Uppsala, Sweden). All experiments were performed at 20^{o}C. HPLC-purified and desalted DNA substrates were obtained from Sigma-Aldrich (Poole) and used without any further purification. The biotinylated 21 bp (21-mer) oligonucleotide formed the bottom strand and was

annealed to various oligonucleotides (details given in Table 3.2 and Figure 3.24) to form double-stranded DNA substrates: 1. Gap – 10-mer and 5' phosphorylated 10-mer oligonucleotide with templating 21-mer strand in a 1:2:3 ratio; 2. Nick – 10-mer and 5' phosphorylated 11-mer oligonucleotide with templating 21-mer strand (1:2:3 ratio) and 3. Duplex – 21-mer oligonucleotide with templating 21-mer strand (1:1 ratio). The experiments were performed with help from C. Brooke, University of Westminster.

Table 3.2: Sequence information of DNA substrates immobilised for Biacore analysis

Oligonucleotide	Sequence
5' phosphorylated 10-mer	5'ACTATCGGAA
10-mer	5'ATTGCGACC
5' phosphorylated 11-mer	5'CACTATCGGAA
21-mer	5'ATTGCGACCCCACTATCGGAA
Biotinylated 21-mer	5'TTCCGATAGTGGGGTCGCAAT

The DNA substrates were annealed as described in Chapter 2 (2.1.11) in annealing buffer containing 5.0 mM EDTA. The cooled biotinylated substrates were diluted appropriately in Biacore buffer 1 (50 mM Tris pH 7.5, 50 mM NaCl, 5.0 mM DTT, 5.0 mM EDTA pH 7.5 and 0.005% Surfactant p20) and immobilised in one of four flow cells of a streptavidin (SA) – coated sensor chip (Biacore AB; Uppsala, Sweden) in a Biacore 2000 (Biacore AB; Uppsala, Sweden). Immobilised DNA surfaces were equilibrated with 20 µl/min Biacore buffer 1. The gapped DNA substrate was immobilised in cell 4, the double-stranded (ds) duplex DNA substrate was immobilised in cell 3 and the nicked DNA substrate was immobilised in cell 2 (Figure 3.24). Approximately 200 Resonance Units (RU) of each DNA substrate was immobilised in cells 2, 3 and 4.

Prior to injection, the concentration of each protein was determined by Bradford protein assay. The units of protein concentration were converted to µM using the following formula:

$$\mu M \text{ protein} = \frac{\mu g \text{ protein}}{\text{size of protein (kDa)}} \times \frac{kg}{10^{9}\ \mu g} \times \frac{mol}{10^{12}\ \mu M}$$

The proteins were diluted to the appropriate concentration (as indicated in results section) with Biacore buffer 1 or 2 (50 mM Tris pH 7.5, 150 mM NaCl, 5.0 mM DTT, 5.0 mM EDTA and 0.005% Surfactant p20) and injected into the system at 30 µl/min for three minutes. Each injection flowed over all four cells and the raw data was corrected by subtraction of cell 1 (control cell without DNA substrate). All DNA binding analysis was performed with the enzymes diluted to 50 – 60 mM NaCl unless otherwise specified.

3.3 Results

3.3.1 Identification of the putative TVlig gene

TV has a genome of ~160 Mb which is suggested to encode for ~60,000 candidate proteins (Carlton *et al.*, 2007). Annotation of the genome identified a homologue of an ATP dependent DNA ligase, accession number XP_001581589. In all other single cell eukaryotic organisms, for example *Neurospora* and *Schizosaccharomyces,* two ligase activities can be identified by a simple BLAST analysis. We attempted to identify other ligase sequences by searching the TV genome database using BLAST (NCBI/TIGR) with a number of ATP-dependent ligase proteins including the fission yeast (*Schizosaccharomyces pombe*) ligase I and IV sequences, small viral DNA ligase sequences (PBCV-1) and with an archael ATP-dependent ligase (*Sulfolobus solfataricus;* detailed sequence alignment attached in appendix 7.2.1). Not one of these searches identified a second ligase open reading frame (ORF). The only ORF in TV predicted to be a functional DNA ligase was 85859.m00330. Furthermore searching the TV database with this ORF also failed to identify any other ligase-

related sequences. The putative TVlig ORF encodes for a 679 amino acid polypeptide with a calculated molecular weight of 76,430 Da and an isoelectric point of 7.8 (http://www.expasy.org/cgi-bin/pi_tool; Gasteiger *et al.*, 2005).

BLAST analysis of the TVlig at the NCBI database showed that the putative polypeptide sequence had 51 and 53% similarity to human DNA ligase I (HuLigI) and *Saccharomyces cerevisiae* cdc9 (an orthologue of human DNA ligase I in the budding yeast) respectively. Table 3.3 shows the percentage homology of putative TVlig with an array of ATP dependent DNA ligases used in BLAST searches to identify any potential ligase ORFs in the TV genome. The homology shown by TVlig with DNA ligase III and IV is only limited to the catalytic core region, which is conserved in ligases throughout the eukarya and archaea. The amino acid sequence alignment of the putative TVlig with DNA ligase I enzymes encoded by a higher and lower eukaryote, *H. sapiens* and *S. cerevisae* respectively, is shown in figure 3.3.

Our analysis suggests that the polypeptide is indeed a true homologue of the human replicative DNA ligase I and does not contain additional domains such as Zinc finger motif and BRCT domains present in DNA ligase III and IV respectively (detailed alignment studies are attached in the appendix 7.2.2). DNA ligases III and IV form stable complexes with accessory proteins (XRCC1 and XRCC4) to catalyse the ligation reaction (Ellenberger & Tomkinson, 2008). BLAST searches revealed that TV genome is devoid of any homologues of such accessory proteins. Further suggesting that the TVlig polypeptide is the homologue of the human/yeast replicative enzyme and that the repair pathways associated with these other ligases are absent in TV.

Table 3.3: Percentage homology of TVlig with known ATP dependent ligases

DNA ligases	Percentage identity	Percentage similarity	Accession number
DNA LIGASE I			
Plasmodium falciparum DNA ligase I (*Pf*LigI)	39	57	XP_001349768.1
Sulfolobus solfataricus DNA ligase	35	54	NC_002754.1
Arabidopsis thaliana DNA ligase I	36	56	NP_172293.2
Saccharomyces cerevisiae DNA ligase I (cdc9)	31	53	NP_010117.1
Drosophila melanogaster DNA ligase I	32	54	NP_611843.2
Human DNA ligase I (HuLigI)	31	51	NP_000225.1
DNA LIGASE IV			
Human DNA ligase IV (HuLigIV)	27	46	NP_001091738.1
Drosophila melanogaster DNA ligase IV	27	45	NP_572907.1
Saccharomyces cerevisiae DNA ligase IV (DnL4)	30	49	NP_014647.1
DNA LIGASE III			
Human DNA ligase III (α) (HuLigIII)	32	50	NP_039269.2

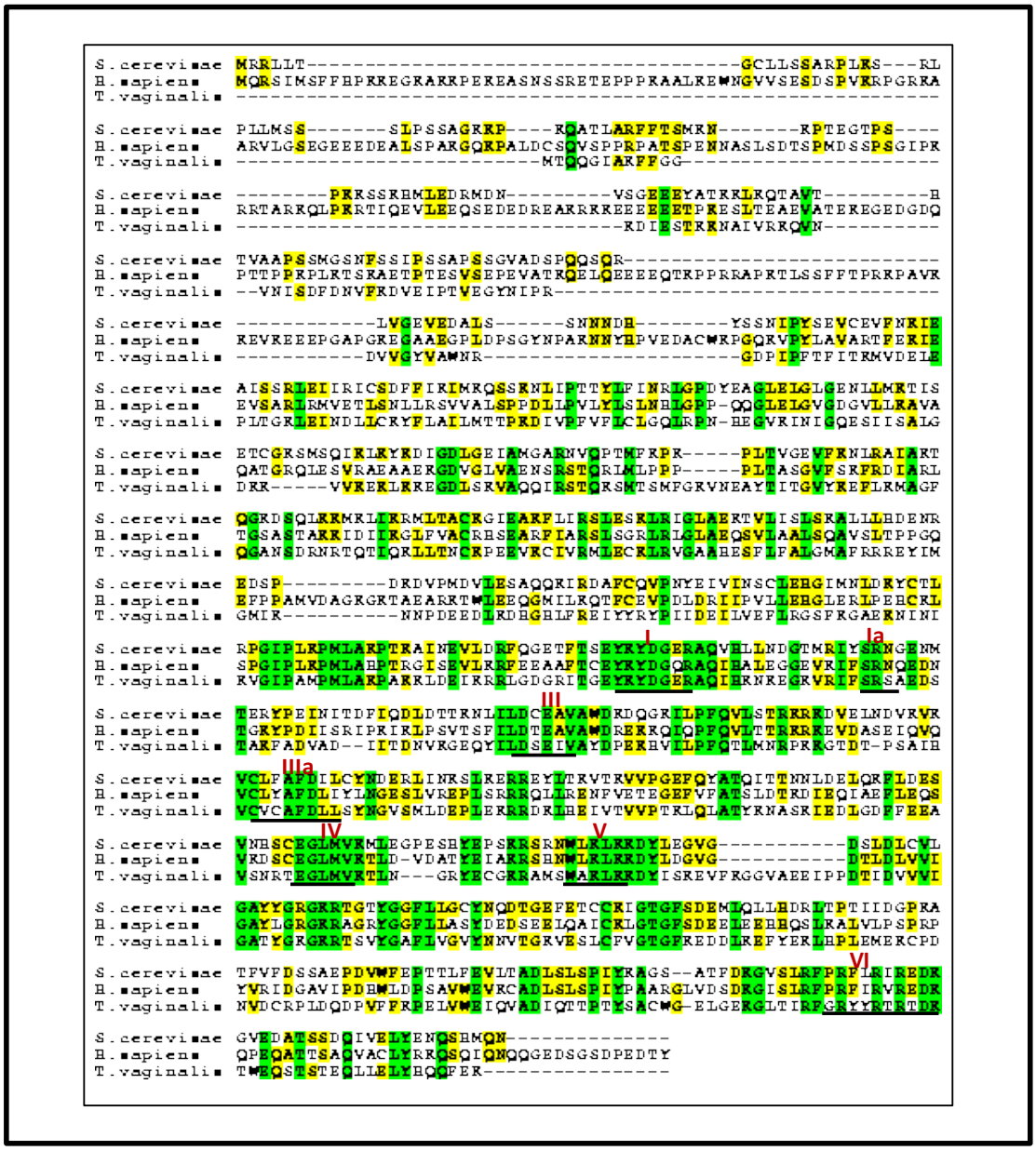


Figure 3.3: Amino acid sequence alignment of DNA ligase I from *T. vaginalis, H. sapiens* and *S. cerevisiae*.

Amino acid sequence of TVlig (*T. vaginalis;* obtained from the TIGR database; 85859.m00330), was aligned with the amino acid sequence of HuLigI (*H. sapiens*; obtained from NCBI; NP_000225.1) and cdc9 of *Sacharomyces cerevisiae* (an orthologue of human DNA ligase I; NP_010117.1). Conserved residues are shaded with green and identical residues with yellow. Positions of conserved motifs (labelled I – VI) are shown with a black bar underneath. The alignment was generated by ClustalW (Chenna *et al.*, 2003).

All three domains of life, Eukaryota, Archaea and Bacteria (prokaryotes) possess sliding clamps to enhance the processivity of their replicative DNA polymerases (Stoimenov & Helleday, 2009). In prokaryotes, this activity is brought about by the actions of the β clamps, whereas in eukaryotes and archaeal organisms the analogous functions are provided by the PCNA (Warbrick, 2000). In addition to enhancing processivity of the polymerase, PCNA has been observed to act as a scaffold for the assembly of accessory processing factors such as DNA ligase I and the flap endonuclease, FEN1 (Vivona & Kelman, 2003; Stoimenov & Helleday, 2009). Most of the PCNA-interacting proteins, including DNA ligase I, contain a conserved PCNA Interacting Peptide (PIP) box whose sequence consensus is Q**XX***h***XX**F(F/Y), where *h* is either M/L/I, to interact with the Inter domain connector loop (IDCL) and other residues located at the C-terminus of PCNA (Warbrick, 2000; Stoimenov & Helleday, 2009). Inactivation of the PIP motif in HuLigI reduces its stable binding with PCNA *in vitro* and abolishes its targeting to replication foci (Pascal *et al.*, 2006). Closer examination of the TVlig sequence revealed the presence of a potential PIP motif located at its N-terminus whose sequence (^{3}QQGIAKFF^{10}) exactly fits the consensus and whose location is consistent with that found in other replicative ligases (Figure 3.4; panel A).

NLS, usually a stretch of basic amino acids- lysine and arginine, is also suggested in the N-terminal region of TVlig. The sequence of ligase NLSs shows considerable conservation in higher eukaryotes such as human (^{119}PKRRTARKQLPKR^{131}), *Mus musculus* (^{116}PKRRTARKQLPKR^{128}) and *Xenopus laevis* (^{185}PKRKTARKQLPKR^{197}). Although no sequence homology is reported in lower eukaryotes such as *Plasmodium falciparum* and *Saccharomyces cerevisiae*, the TV sequence in this region (^{142}KTDKKVVKEKLKK^{153}) is consistent with the NLS of higher eukaryotes (Figure 3.4; panel B; Cardoso *et al.*, 1997). Identification of a PIP box and NLS signal in TVlig sequence further supports the view that the identified ORF encodes for a DNA ligase I.

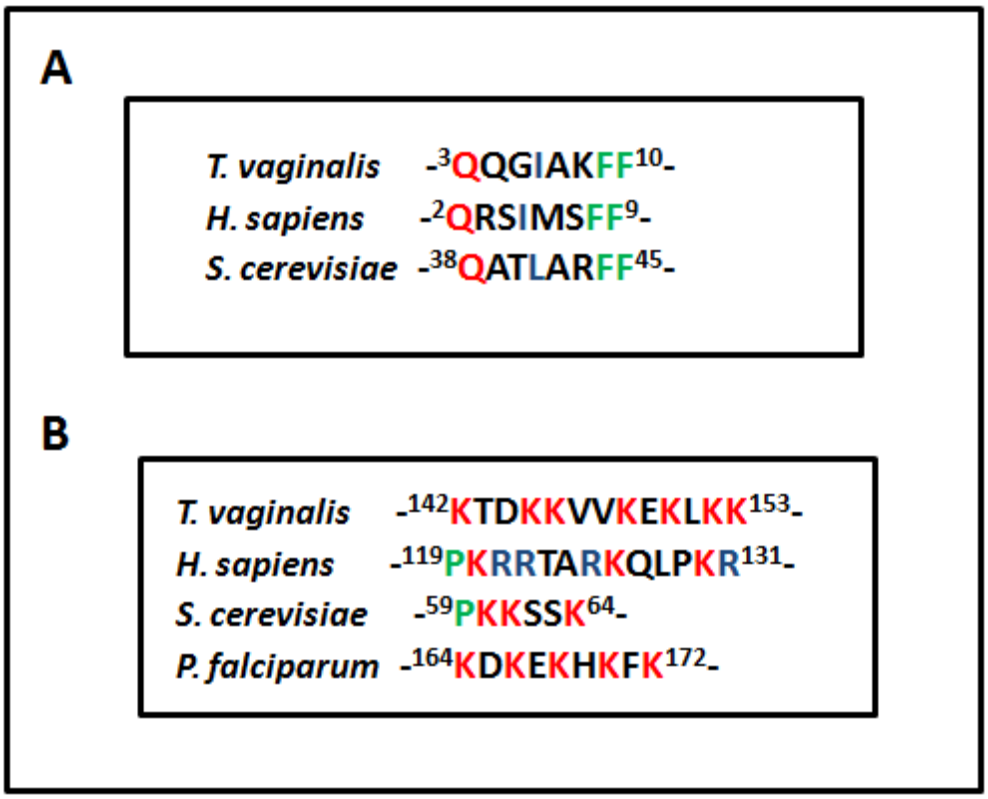


Figure 3.4: Amino acid sequence alignment of PIP motif and NLS of DNA ligase I from *T. vaginalis, H. sapiens, S.cerevisiae* and *P. falciparum*

The presence of potential PIP (PCNA interacting polypeptide) motif and NLS (Nuclear localisation signal) is suggested at the N-terminal of TVlig. Panel A shows the alignment of PIP motif and panel B shows the alignment of NLS located at N-termini of TVlig (T.vaginalis; 85859.m00330), human DNA ligase I (H.sapiens; NP_000225), cdc9 of *Saccharomyces cerevisiae* (S.cerevisae; NP_010117.1) and *Plasmodium falciparum* DNA ligase I (P.falciparum; AAL59668). The PIP motif is absent in *Plasmodium falciparum* DNA ligase I. The alignment was performed with ClustalW (Chenna *et al.*, 2003). The search results were evaluated and the sequences were manually realigned.

Biochemical analyses, especially targeted mutagenesis in a variety of ligases and capping enzymes, has enhanced our understanding of the role of critical residues in each conserved motif. The residues within the catalytic core of PBCV-1 DNA ligase, the smallest known ATP dependent DNA ligase, have been mutated to assess their role in the catalytic mechanism of DNA binding and repair (Odell *et al.*, 2000; Sriskanda & Shuman, 2002a & 2002b; Nair *et al.*, 2007). The role of essential amino acid residues residing within the catalytic core of TVlig was analysed by comparing with that of PBCV-1 DNA ligase and other ATP dependent DNA ligases (Figure 3.5, panel A).

The catalytic core of TVlig is composed of an adenylation domain (AdD) and oligomer binding (OB) domain (Figure 3.5, panel B). Within the AdD is an adenylate-binding pocket composed of six motifs (I, Ia, III, IIIa, IV and V; Figure 3.5, panel A; Figure 3.30; panel B) that define the polynucleotide ligase/mRNA capping enzyme superfamily of covalent nucleotidyltransferases (Shuman & Lima, 2004; Tomkinson *et al.,* 2006). The motif I (**K**YDGE**R**) includes the lysine (Lys338) to which it is probable that AMP becomes covalently linked in the first step of the ligase reaction and an arginine (Arg343) which will contact the 3'OH of the nucleotide ribose sugar and are thus critical in step one catalysis. Motifs Ia, III, IIIa, IV and V contain conserved amino acids that contact AMP and play essential roles in one or more steps of the ligation pathway (Doherty & Suh, 2000; Shuman, 2006). Motif IIIa normally contains a conserved aromatic base, here we find a phenylanine, Phe438 (Phe98 in PBCV-1 DNA ligase) that acts to sandwich the purine base of the ATP nucleotide with a conserved hydrophobic side chain from motif IV, here Met492 (Met164 in PBCV-1 DNA ligase). The significance of this aromatic/purine/hydrophobic arrangement to support the nucleotide base is observed in the structure of adenylated ATP dependent DNA ligases such as T7 (Subramanya *et al.,* 1996) and PBCV-1 DNA ligase (Odell *et al.,* 2000), NAD^+-dependent ligases such as *Thermus filiformis* (*Tfi*; Luo & Baranay, 1996) and RNA capping enzymes *e.g.* PBCV-1 mRNA capping enzyme (Hakansson & Wigley, 1998).

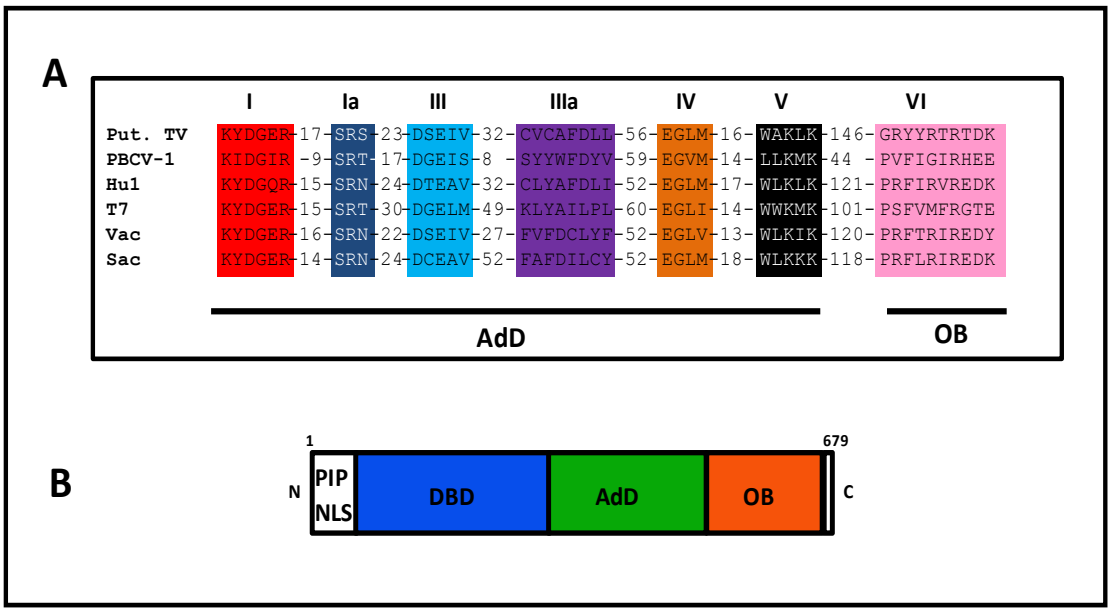


Figure 3.5: Sequence alignment of conserved motifs and domain organisation in predicted TVlig structure

Panel A shows a ClustalW alignment analysis (Chenna *et al.*, 2003) of sequence motifs conserved in characterised and putative ATP-dependent DNA ligases from *Trichomonas vaginalis* (Put. TV), *Paramecium Bursaria Chlorella Virus* -1 (PBCV-1) DNA ligase, Human DNA Ligase I (HuI), Bacteriophage T7 DNA ligase (T7), *Vaccinia* virus DNA ligase (Vac) and *Saccharomyces cerevisiae* DNA ligase (Sac). The spacings between the motifs are indicated. Panel B shows domain organisation in predicted TVlig structure. Abbreviations: PIP - PCNA interacting polypeptide; NLS - Nuclear localisation signal; AdD - Adenylation domain; OB - Oligomer binding domain; DBD - DNA binding domain.

Motifs III and IV include conserved glutamate residues (here, Glu389 and Glu488) involved in metal binding, vital for catalysing conformational changes which drive forward steps one and three of the ligation reaction (Odell *et al.*, 2000). A pair of conserved lysines (here, Lys512 and Lys513) that form contacts with the ribose sugar hydroxyl group of AMP are located in motif V (Sriskanda & Shuman, 2002b). These residues are also involved in stabilising the β and α-phosphate of ATP in a mechanism analogous to the stabilisation of GTP by capping enzymes (Sriskanda & Shuman, 2002b). Located near the C-terminus of the OB domain is motif VI containing an acidic residue (here, Asp657), essential for forming contacts with the β-phosphate of ATP during step I of the reaction and thus required to form ligase-adenylate intermediate (Nair *et al*, 2007). Two positively charged residues (here, Arg655 and Lys658) located in motif VI also serve to orientate the β and α-phosphate of ATP

during the adenylation reaction in step 1 of catalysis in a mechanism analogous to the mRNA capping enzymes (Doherty & Suh, 2000).

In a manner similar to the domain organisation present in HuLigI, an additional DNA binding domain (DBD) and an N-terminal PCNA binding domain are present in TVlig (Figure 3.5, panel B; Figure 3.6). The DBD is present in all three mammalian DNA ligases (DNA ligase I, III and IV) and its deletion abolishes DNA binding and repair in HuLigI (Pascal *et al.*, 2004).

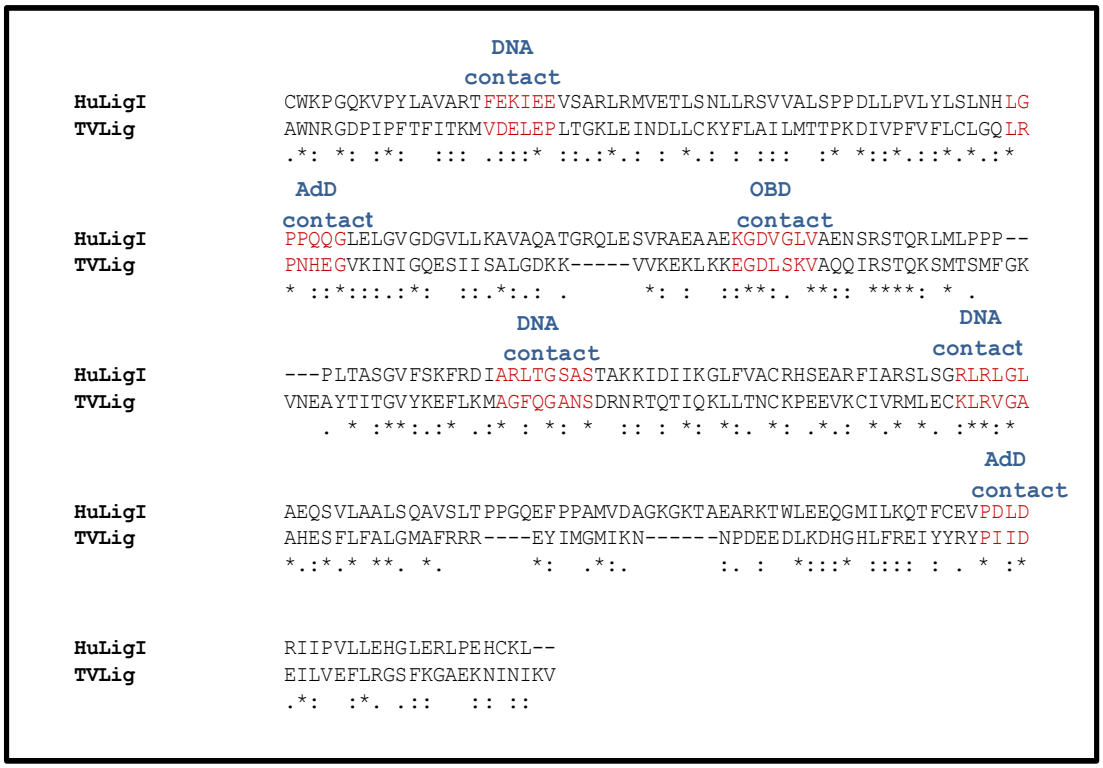


Figure 3.6: Amino acid sequence alignment of DNA binding domains of HuLigI and TV DNA ligase

The alignment of amino acid sequences of DNA binding domain (DBD) of HuLigI and TVlig is shown. The DBD contacts and stabilises the DNA substrate in co-ordination with the conserved adenylation domain (AdD) and oligomer binding domain (OBD). The positions of protein interactions with AdD, OBD and DNA substrate are highlighted in red. The alignment was performed with ClustalW (Chenna *et al.*, 2003). The `*` indicates identical residues between the two ligases while conserved and semi-conserved substitutions are marked as `:` and `. ` respectively (Pascal *et al.*, 2004).

The DBD of HuLigI closes the protein clamp around the nicked DNA substrate via contacts between the N-terminal DBD and the C-terminal OB-fold domain. The

crystal structure of HuLigI bound to adenylated DNA highlighted the peptide regions of DBD involved in making these interactions (Figure 3.6; Pascal *et al.,* 2004). Recent studies suggest that the DBD of HuLigI also interacts with PCNA and the heterotrimeric cell cycle checkpoint clamp, hRad9-hRad1-hHus1 (9-1-1; Song *et al.*, 2009). This was previously reported in archaeal DNA ligases from *P. furiosus* and *S. solfataricus* using DBD residues for interaction with PCNA sliding clamps (Pascal *et al.*, 2006; Pascal, 2008). The DBD-PCNA interactions stimulate ligation activity in these organisms. The protein sequence of the potential DBD of putative TVlig, when aligned with HuLigI using ClustalW, suggest the presence of peptide regions involved in making contacts with DNA, AdD and OB domain (Figure 3.6; Chenna *et al.*, 2003).

Trichomonas vaginalis represents one of the earliest branching lineages among the eukaryotes and has acquired many prokaryotic genes through lateral gene transfer (LGT; Carlton *et al.,* 2007). Based on known DNA ligase sequences, a phylogenetic tree was constructed (organisms were randomly chosen from different origins) with the neighbour-joining method using Phylip (Figure 3.7; Felsenstein, 1993). Phylogenetic analysis suggests that TVlig is aligned more closely with lower unicellular eukaryotes such as *S. pombe* than higher eukaryotic species such as *H. sapiens*. Higher eukaryotes are known to have at least two DNA ligases, three in mammals, which play a role in DNA repair, replication and recombination however the number of ligases in lower eukaryotes appears somewhat variable. The TV polypeptide shares its highest indentity (39%, Table 3.3) with the recently identified DNA ligase I from *P. falciparum* and viral DNA ligases such as PBCV-1 DNA ligase. In these organisms there is a single copy of DNA ligase, which may not be surprising for viruses but it is certainly interesting that the only other fully sequenced parasitic organism (a single-celled eukaryote) also appears to have only a single ligase. When we searched the *P. falciparum* database, with the same sequences used to trawl the TV database to find ligases, again only a single ligase was evident. The evidence, therefore suggests, that either these parasitic organisms are overly sensitive to double-strand breaks (DSB) in DNA or have evolved to repair DSB with DNA ligase I.

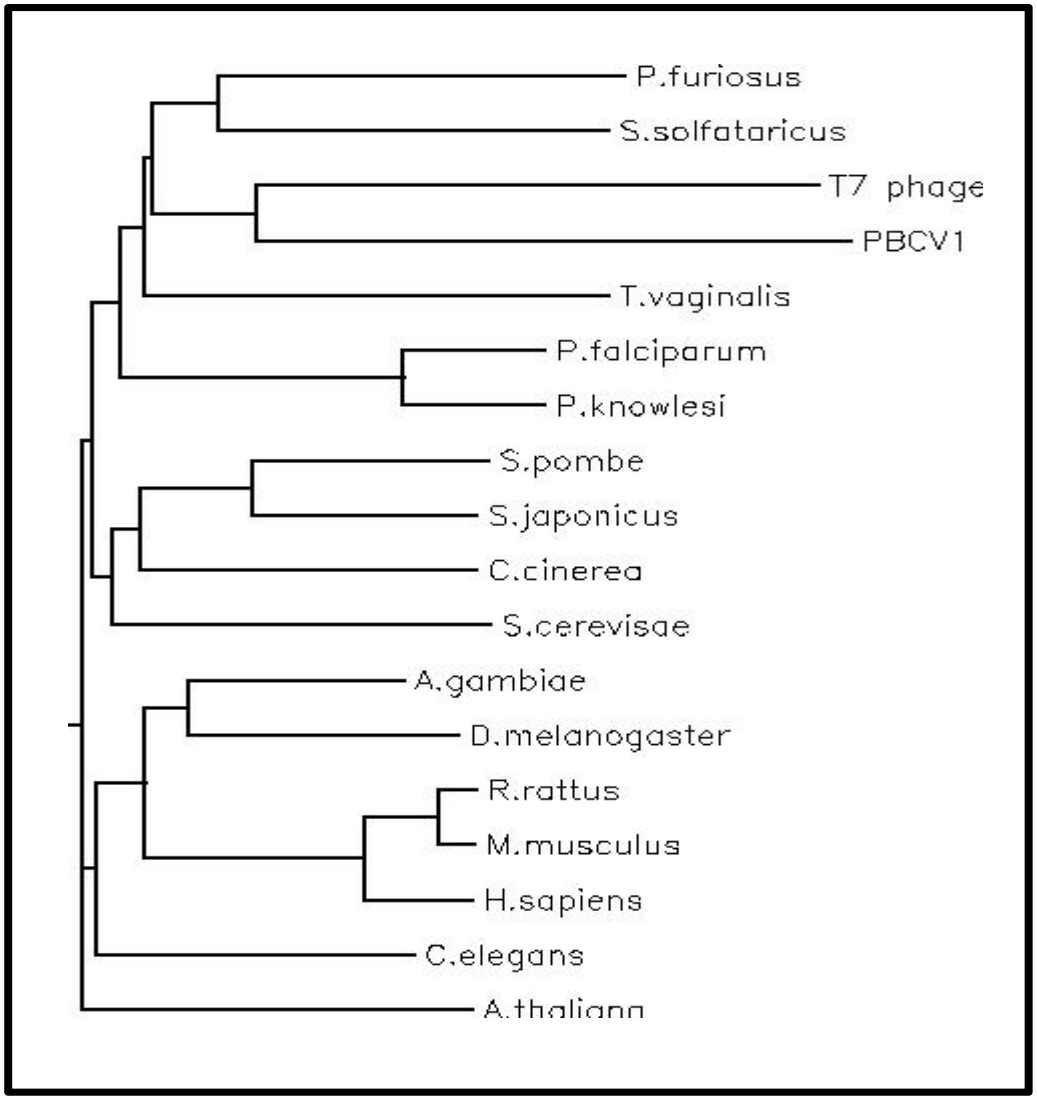


Figure 3.7: Phylogenetic analysis of ATP dependent DNA ligases.

Phylogenetic tree showing the evolutionary relationship of some randomly chosen ATP-dependent DNA ligases was constructed using Phylip (Felsenstein, 1993) from sequence alignment data generated using ClustalW (Chenna *et al.*, 2003). DNA sequences of ATP-dependent DNA ligases were obtained from NCBI. The full names of the hosts of DNA sequences are: *Schizosaccharomyces japonicas* - XP_002171493, *Coprinopsis cinerea* - BAC76762, *Saccharomyces cerevisiae* - NP_010117.1, *Schizosaccharomyces pombe* - NP_593318.2, *Homo sapiens* - NP_000225.1, *Mus musculus* - NP_001076657.1, *Rattus rattus* - NP_001019439.1, *Anopheles gambiae* - XP_319999.2, *Arabidopsis thaliana* - NP_175351.1, *Caenorhabditis elegans* - NP_741625.2, *Drosophila melanogaster* - NP_611843.2, *Sulfolobus solfataricus* - AAK40535, *Paramecium bursaria Chlorella* virus-1 - NP_048900, Bacteriophage T7 - NP_041963, *Pyrococcus furiosus* - NP_579364, *Plasmodium falciparum* - AAL59668, *Plasmodium knowlesi* - XP_00226193.

3.3.2 Cloning of the putative TVlig

Following identification of the open reading frame (ORF) of a putative DNA ligase from the TV genome, primers complementary to the 5' and 3' ends of the nucleotide sequence of the polypeptide were designed (sequences given in section 3.2.1). Genomic DNA was extracted from a 72-hour TV culture by the phenol-chloroform method (as described in section 2.1.5) and used as a template for amplification of the putative ligase gene by PCR. The amplification was performed using *Pfu* polymerase, a proof reading enzyme, to exclude (or at least minimise) the introduction of mutations in the course of the reaction. A product was obtained by the PCR reaction that was consistent with the expected size for the TV ORF of 2040 bp when analysed by electrophoresis (Figure 3.8) and gel purified.

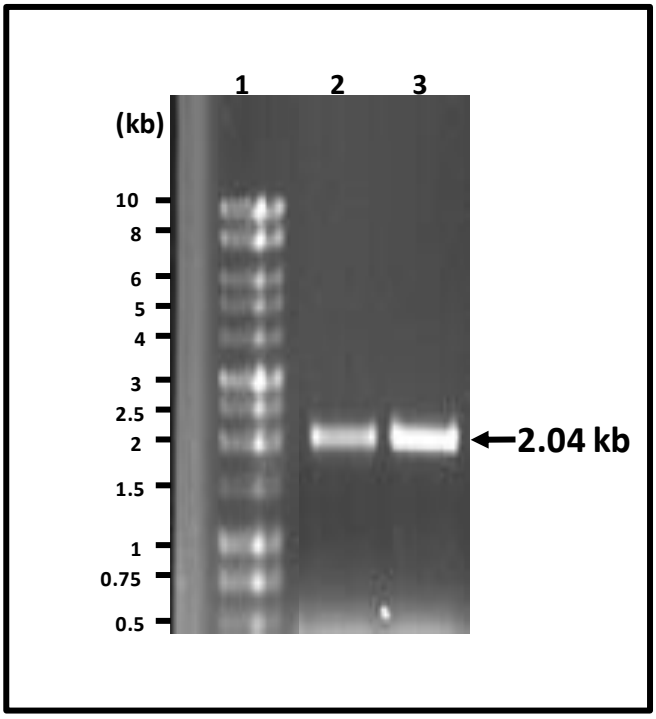


Figure 3.8: PCR product of TVlig resolved on a 0.7% agarose gel

Genomic DNA was extracted from a TV culture by the phenol-chloroform method and used as a template in PCR amplification. The PCR consisted of 30 cycles of 1 minute denaturation at 94^{o}C, 2 minute annealing at 55^{o}C and 3 minute extension at 72^{o}C with TVlig specific primers. DNA products were resolved on a 0.7% agarose gel run in 0.5 X TBE buffer visualised by staining with ethidium bromide (0.5 µg/ml for 20 minutes). Lane 1 shows 1 Kb ladder molecular weight markers (Promega) and Lanes 2 and 3 shows separately amplified 5 µl aliquots of PCR reactions. A black arrow indicates the position of the putative TVlig product.

The gel-purified**,** A-tailed PCR product was inserted into pGEM-T and the putative positive (white) clones were screened by restriction digest of mini-prepped plasmid DNA to observe any insert release by agarose gel electrophoresis (Figure 3.9, panel A). A number of positive clones were obtained from which larger scale ultra-pure plasmid mini preps were made and sent for sequencing. Once sequenced we were able to verify that the strain of TV we had obtained in the UK had exactly the same sequence as the US strain used to generate the genomic DNA sequence. This is somewhat surprising to us that there was not even a silent base pair change and either suggests that the two sources of TV were very similar or that the gene is strongly conserved.

To characterise the putative TVlig, a fully sequenced ligase gene was sub-cloned into the expression plasmid pET-16b (Novagen). Figure 3.9, panel B; shows the insert release after the restriction digest of one of these plasmids. The sequence of the final clone again exactly matched that of the open reading frame published by TIGR.

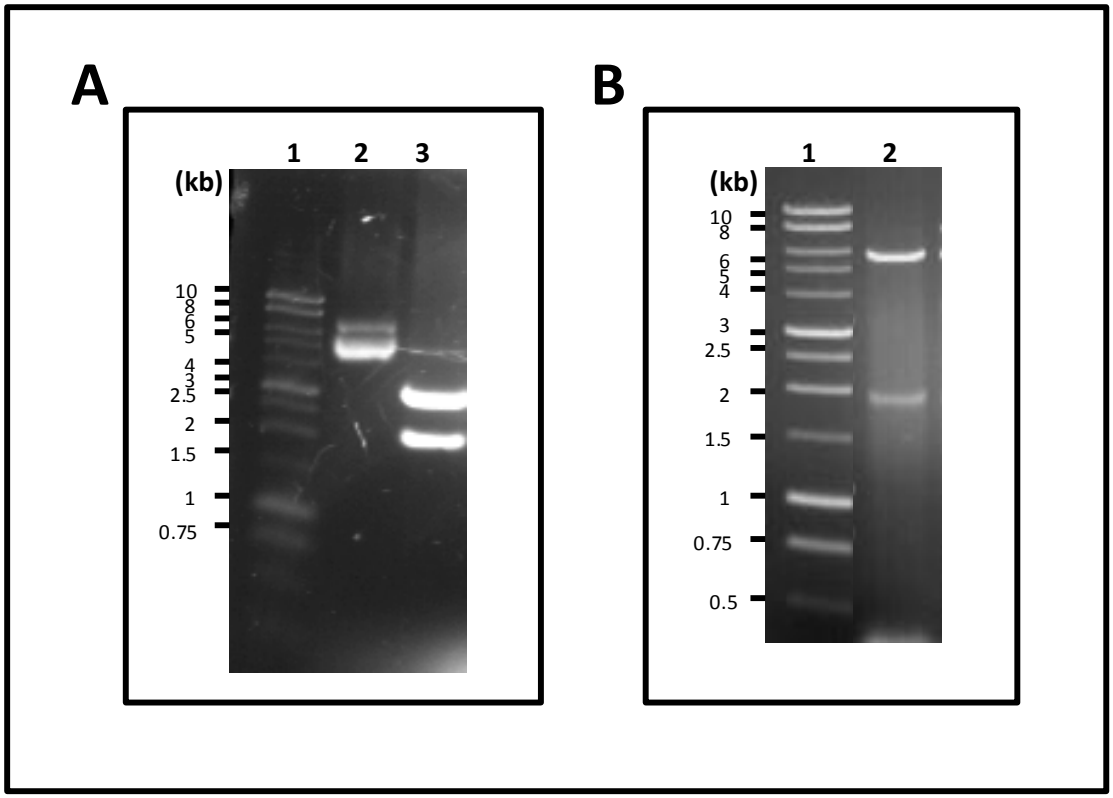


Figure 3.9: Restriction digest confirmation of TVlig clone in pGEM-T and pET-16b.

The A-tailed PCR product was ligated into pGEM-T. *E. coli* DH5α cells transformed with the ligation mixture apparently containing plasmid with an insert (white colonies) were selected and the gene was released from the vector by restriction with *Bam* HI.
Panel A: molecular weight markers are resolved in lane 1; Lane 2: Uncut plasmid; Lane 3: Plasmid digested with *Bam* HI. The TVlig was then sub-cloned into pET-16b and the success of the insertion confirmed by restriction with *Bam* HI as shown in panel B. Lane 1: molecular weight marker; Lane 2: *Bam* HI digested plasmid.

3.3.3 Over-expression and Purification of TVlig

The recombinant pET-16b vector carrying the TVlig gene was transformed into *E. coli* BL21(DE3) cells for inducible control of the expression of the protein. Following induction with IPTG, cells harbouring the recombinant plasmid produced a prominent ~75 kDa polypeptide (Figure 3.10). The size of this species is consistent with the 76 kDa estimated for the TVlig on the basis of the amino acid sequence. It is however somewhat smaller than might be expected from the addition of the ten histidine tag (~3 kDa) at the amino terminus of the protein. Anomalous electrophoretic mobility of DNA ligases has been reported earlier, for example, PBCV-1 DNA ligase (predicted size of 34 kDa) migrates slowly and corresponds to ~40 kDa polypeptide on SDS-

PAGE (Odell & Shuman, 1999). Takano *et al.* (1988) suggest that nearly a 10-kDa difference in apparent size between the deduced mass and the relative SDS-PAGE-determined mass is within the error attributable to anomalous migration of proteins on SDS-PAGE.

When the lysate of the induced cells was centrifuged, this protein remained soluble. The recombinant His-tagged protein was purified from the soluble extract by Ni-NTA affinity chromatography. The Ni-NTA column was washed with 25 mM imidazole to remove loosely bound contaminating proteins. Histidine-tagged proteins were eluted with buffer A containing 300 mM imidazole (as described in section 3.2.3; Figure 3.10; panel A). The peak fraction, determined by inspection of the Coomassie stained gel, was then dialysed against 25 mM NaCl and applied separately to Blue-sepharose (cation exchange resin) and S-sepharose (cation exchange resin with sulphonyl exchange group; $-SO_3^-$) resins. Blue-sepharose is a strong cationic ion exchanger with cibacron blue (a triazine dye) covalently attached to the sepharose matrix by the triazine coupling method. Cibacron blue binds most enzymes requiring adenyl-containing cofactors (*e.g.* NAD^+; Lascu *et al.,* 1984). It acts as a strong cationic ion exchanger because of its negative charge and binds other proteins at pH-values below their isoelectric points (Lascu *et al.,* 1984).

The salt gradient applied to Blue-sepharose helped remove some of the higher and lower contaminating protein bands but a lot of the TVlig protein was lost during the process (estimated by SDS-PAGE analysis; Figure 3.10; panel B). Column chromatography on S-sepharose and elution along a NaCl gradient, yielded an apparently homogenous protein preparation (Figure 3.10; panel C), with better recovery as compared to Blue-sepharose and was used as the final purification step.
Approximately 5 mg of TVlig was obtained per 1 litre of culture as estimated by a Bradford protein assay (calibration curve attached in the appendix 7.4). The size of the polypeptide was estimated by comparing with marker proteins on Coomassie stained SDS gel (when present in small amounts; Figure 3.10; panel C; lane 3 or 5). The over expressed protein is bigger than the 75 kDa marker protein consistent with its estimated size (78 kDa). We therefore analysed by western blot to confirm that it was indeed a his-tagged protein that must have come from our pET-16b plasmid.

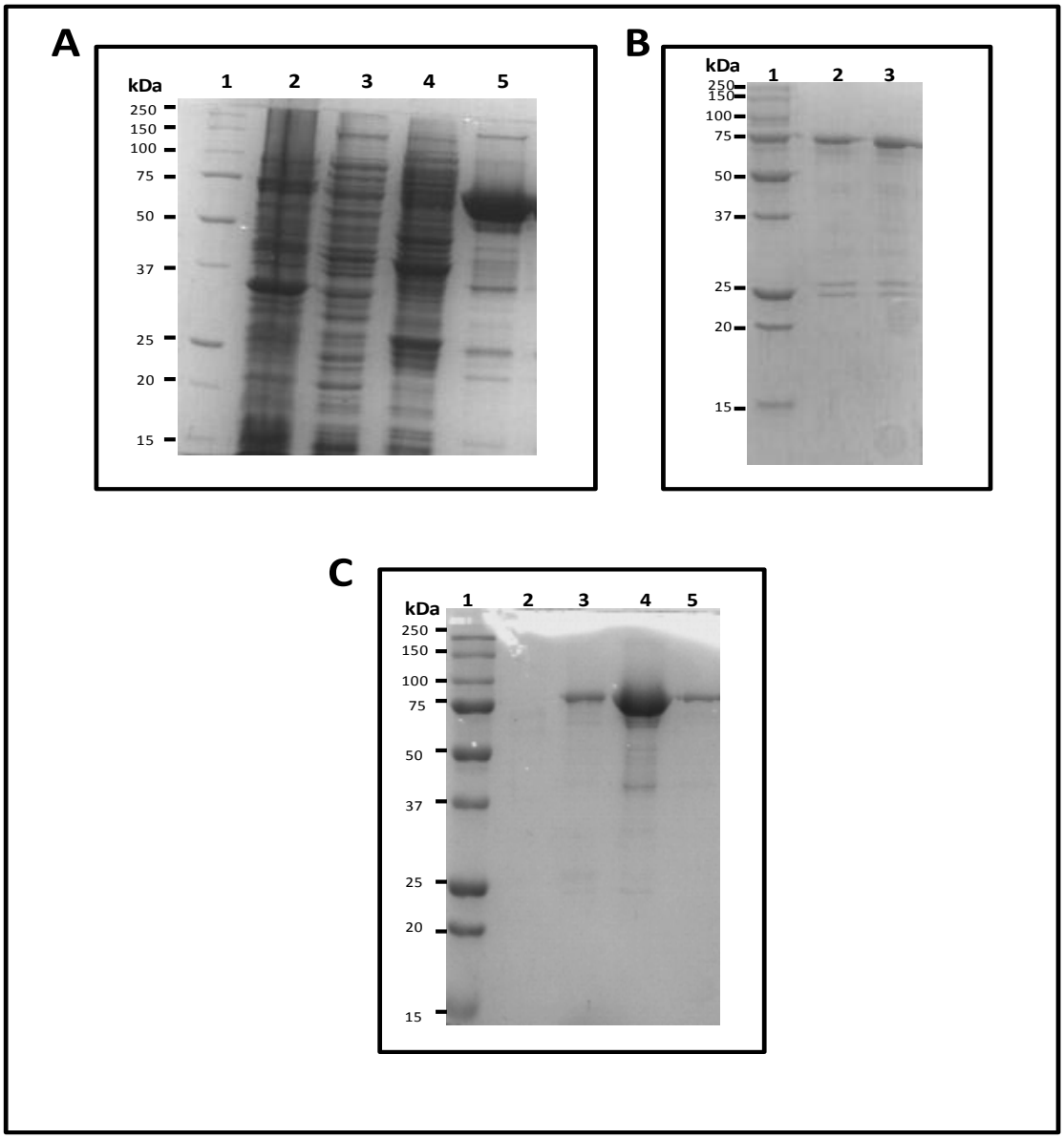


Figure 3.10: Elution profile of TVlig from Ni-NTA, Blue-Sepharose and S-Sepharose

TVlig was expressed in BL21(DE3) cells by induction with 0.4 mM IPTG for two hours at 37^{o}C. The lysed, soluble extract was applied to Ni-NTA resin and eluted with buffer A (50 mM Tris-HCl pH 7.0, 300 mM NaCl and 10% glycerol) containing 300 mM imidazole. The polypeptide composition was analysed by SDS-PAGE (12% resolving gel) stained with Coomassie Brilliant Blue.

Panel A: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker was resolved in lane 1 (molecular weight of marker proteins are shown to the left of the gel); Lane 2: pellet fraction; Lane 3: flow through; Lane 4: wash fractions (25 mM imidazole in buffer A); Lane 5: elution fractions (300 mM imidazole in buffer A). The protein peaks eluted from Ni-NTA resin were diluted to 25 mM NaCl and applied separately to Blue-sepharose and S-sepharose resins and eluted along an NaCl gradient (NaCl in buffer A: 50 mM Tris-HCl pH 7.0, 5.0 mM DTT, 10% glycerol). The elution profile from Blue-sepharose and S-sepharose columns are analysed in panels B & C respectively.

Panel B - Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker; Lane 2: 150 mM NaCl elution fraction; Lane 3: 200 mM NaCl. Panel C - Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker; Lane 2: Flow-through; Lanes 3, 4 & 5: 150, 200 & 250 mM NaCl in buffer A respectively.

3.3.4 Western blot for recombinant TVlig

Sequence analysis had confirmed the identity of the ORF cloned into pET-16b as that encoding the putative TVlig. *In silico* analysis gives this polypeptide a mass of 76.4 kDa which with the additional amino acids of the histidine tag results in a protein of mass 78.5 kDa. Whilst the recovery of an abundant protein species of similar molecular weight is strongly supportive of this being the recombinant TVlig, to confirm the identity of the ~76-78 kDa species as the TVlig a western blot procedure was employed (as described in Chapter 2; section 2.2.4). An aliquot of the peak fraction eluted from the S-sepharose ion-exchange column was electrophoresed and transferred to nitrocellulose membrane. The extent of transfer was determined by staining the membrane with Ponceau S followed by washing and blocking the membrane for non-specific binding. The membrane was probed with biotinylated anti-His antibody (primary antibody) followed by horse radish peroxide (HRP) conjugated streptavidin. A single polypeptide corresponding to the abundant ~76-78 kDa polypeptide species was detected on addition of the peroxidase substrate diaminobenzidine (DAB; Figure 3.11). The exact size of the species is impossible to discern from this blot because of the overloading of the recombinant protein on the gel. This provided good supporting evidence that polypeptide, eluting from the S-sepharose was the product of the cloned of TVlig gene.

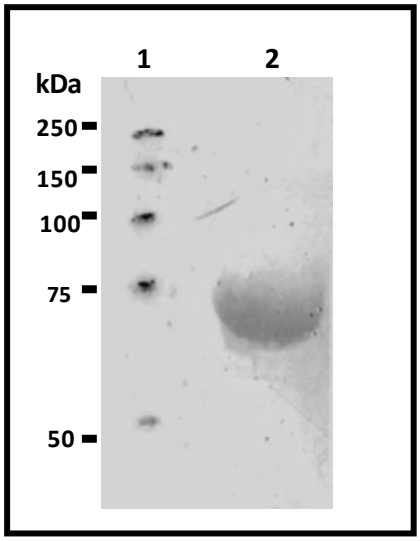


Figure 3.11: Western blot of the purified TVlig with anti-his antibody

A 15 µg aliquot of the purified protein eluted from S-sepharose was separated by 12% SDS-PAGE along with pre-stained broad range molecular weight markers (BioRad) and transferred to a nitrocellulose membrane in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol). After determining the transfer efficiency, by staining of the membrane with Ponceau red stain (0.5% Ponceau S, 1% acetic acid), the membrane was blocked with 2% BSA. Proteins blotted onto the nitrocellulose membrane were probed with anti-his antibody (Sigma) followed by HRP-conjugated streptavidin (Pierce). The detection of his–tagged protein was performed by adding diaminobenzidine in TBS/H_2O_2 and photographing the gel. Lane 1: Prestained broad range molecular weight markers; Lane 2: Purified TVlig.

3.3.5. Preliminary ligation assays with TVlig

The catalytic activity of the putative TVlig was first examined on *Hin*d III digested λ DNA. It was observed that the activity of the enzyme requires addition of macromolecular crowding agents such as polyethylene glycol (PEG 8000) to facilitate DNA ligation. When PEG is present at 5% (w/v) with the TVlig, ligation products (higher molecular weight species, >23 kb in size) are observed migrating in the agarose gel (Figure 3.12).

Addition of PEG stimulates blunt end ligation by Vaccinia virus DNA ligase but is not normally thought to be required for cohesive end ligation as that of *Hin*d III digested λ DNA fragments (Odell *et al.,* 1996). The volume-excluding effect of PEG stimulates ligation activity of T4 DNA ligase and is known to increase its affinity for its DNA substrate (Hayashi *et al.,* 1986). Similar adaptive effects are noticed for DNA

polymerases where the presence of macromolecular crowding agents results in an overall increase in enzymatic activity and enhances their binding to DNA (Zimmerman & Harrison, 1986).

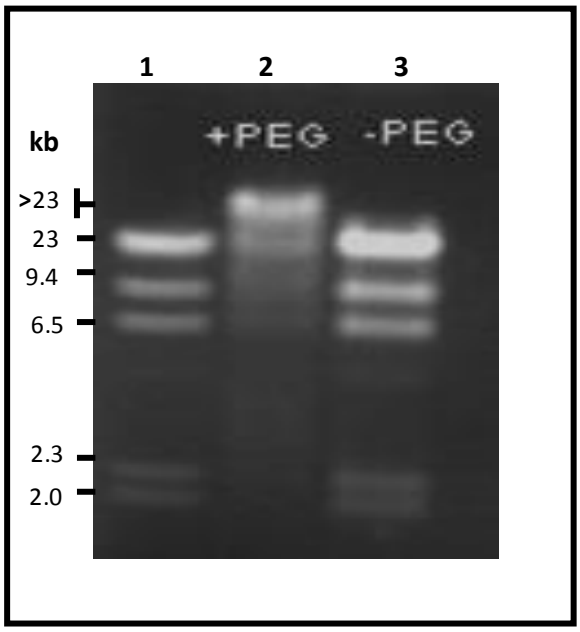


Figure 3.12: Preliminary ligation assays for TVlig with *Hin*d III digested λ DNA

Purified TVlig was used to ligate *Hin*d III digested λ DNA in ligation buffer (containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol and 25 mg/ml bovine serum albumin). The buffer was supplemented with 5% PEG and 1 mM ATP. The ligation reaction products were analysed on a 0.8% agarose gel, stained with ethidium bromide and viewed under UV illumination. Lane 1 contains *Hin*d III digested λ DNA; Lane 2 contains ligation reaction carried out in the presence of 5% PEG; Lane 3 contains the ligation reaction in the absence of PEG.

Crude assessment of the ATP-requiring nature of the enzyme preparation was also analysed by similar ligation of *Hin*d III digested λ DNA fragments (data not shown). The enzyme displayed a low but detectable activity as compared to T4 ligase, but it gave enough evidence that it is an ATP dependent ligase, which requires stringent conditions, like the addition of PEG for the efficient ligation of DNA molecules.

3.3.6. Construction of active site mutant (K338A) of TVlig

When this work was initiated our purified TVlig was the first ligase characterised from a parasitic protozoan (*P. falciparum* DNA ligase was published contemporaneously with this study). Moreover, TV at this point was the only eukaryote known to possess a single DNA ligase apparently capable of both repair and replication. To be absolutely certain that the ligase did not differ from other ligases we decided to create an active site, alanine substitution mutant of TVlig. The ligation of *Hin*d III digested λ fragments strongly supports that the purified ~76-78 kDa polypeptide is a DNA ligase but by purifying an active site mutant it could be confirmed that the ligation activity of the preparation is not contributed to or indeed solely as a result of co-purification of the his-tagged recombinant protein with *E. coli* host DNA ligase (75 kDa) whose size is comparable to TVlig (Kodama *et al.*, 1991).
DNA ligases and other members of the covalent nucleotidyl transferase superfamily are defined by an essential lysine residue situated within motif I (KxDGxR). Previous studies have shown that during the first step of the DNA ligation reaction, the ligase is activated through the formation of a phosphoamidate bond between the amino group of this lysine residue and AMP (Doherty & Suh, 2000). The contribution of the motif I (KYDGER) residue Lys338 to the activity of TVlig was surmised from the effect of substituting the residue with an alanine (A) by a PCR-based method in two steps (Figure 3.2). Forward and reverse primers complementary to the motif I region of the protein but encoding a base alteration resulting in a substitution of Lys338 to Ala were designed**.** The altered ligase was amplified from the pET-16b plasmid bearing wild type TVlig as a template (as described in materials and methods, section 3.2.7). The first step PCR reactions amplified two fragments approximately 1 kb in size (1-1034 bp and 996-2040 bp) bearing the alanine encoding mutation in both the upper and lower strands respectively (Figure 3.13; panel A).

The fragments obtained above were gel purified and used as a template in a single PCR reaction to amplify the full-length active site mutant using forward and reverse primers (see materials and methods, 3.2.7) complementary to the start and terminal portions of the full-length wild type TVlig. Figure 3.13, panel B shows the PCR

product obtained from this experiment producing the 2.02 kb DNA fragment, coding for the full-length active site mutant Lys338Ala (K338A).

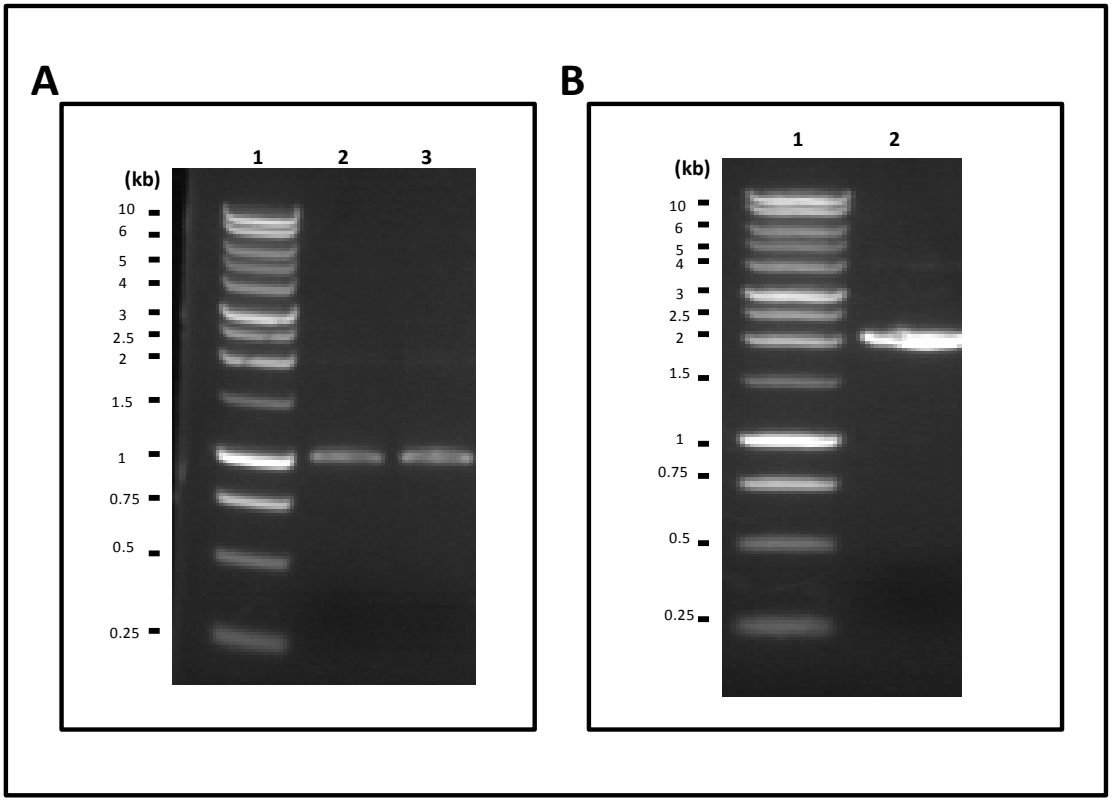


Figure 3.13: Producing an active site mutant (Lys338Ala) of TVlig by PCR

A plasmid bearing full-length wild type TVlig was used as a template in the reaction amplified over 30 cycles, consisting of 1 minute at 94^{o}C, 2 minutes at 55^{o}C and 1 minute at 72^{o}C. The reaction products were analysed on a 0.7% agarose gel, stained with ethidium bromide and viewed under long-wavelength UV illumination. Primer sequences for the mutagenising PCR are detailed in table 3.1; section 3.2.7.

Panel A shows two amplified fragments approximately 1 kb in size (1-1034 bp and 996-2040 bp) bearing the alanine mutation in both the upper and lower strands respectively. Lane 1: 1kb DNA ladder molecular weight markers; Lane 2: PCR product 1-1034; Lane 3: PCR product from 996 – 2040 bp.

Panel B shows full-length PCR product of the active site mutant of TVlig. The two PCR fragments obtained after introducing the point mutation in the active site of TVlig were used as a template to amplify the full-length ORF by 30 cycles consisting of 1 minute at 94^{o}C, 2 minutes at 55^{o}C and 3 minutes at 72^{o}C. Lane 1: 1 kb DNA ladder molecular weight markers (Promega); Lane 2: full-length active site mutant of TVlig (2040 bp).

The full-length product was gel purified, digested with *Bam* HI restriction enzyme and ligated with *Bam* HI digested and CIP treated pET-16b expression vector DNA. The ligation reaction was transformed into electrocompetent *E. coli* DH5α cells. Some

of the transformed colonies were randomly chosen and plasmid DNA was purified and sequenced. A positive mutant, confirmed after sequencing the isolated plasmids to confirm no other mutations had been introduced by the PCR reaction, was then transformed into *E. coli* BL21(DE3) cells. Expression of the active site mutant was induced by addition of IPTG during growth at 37^{o}C in LB broth as described for the wild type protein (section 3.2.3). After centrifugal separation of the crude lysate, an over expressed 76 kDa species the K338A TVlig was recovered in the soluble supernatant fraction (Figure 3.14).

The expected 76 kDa polypeptide was purified from the soluble bacterial extract by sequential Ni-NTA affinity chromatography and S-sepharose column chromatography steps. The recombinant protein eluted from Ni-NTA resin with buffer A containing 300 mM imidazole (Figure 3.14; panel A). The eluted fraction was dialysed against buffer B containing 25 mM NaCl and applied to the ion exchange resin, S-Sepharose (Figure 3.14; panel B). The peak fraction of the 76 kDa protein eluted from this column with 0.2-0.25 M NaCl.

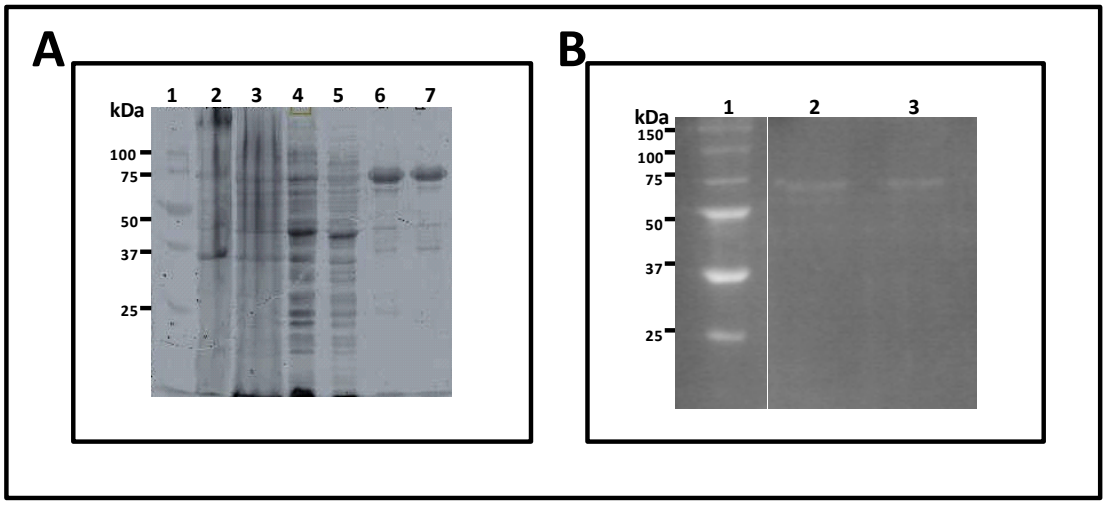


Figure 3.14: Purification of the K338A active site mutant of TVlig over Ni-NTA and S-Sepharose chromatography

Mutant TVlig was expressed in *E. coli* BL21(DE3) by induction with 0.4 mM IPTG for two hours at 37^{o}C. The lysed *E. coli*, soluble extract was applied to Ni-NTA resin, washed and then eluted in buffer A (50 mM Tris-HCl pH 7.0, 300 mM NaCl, 10 mM imidazole and 10% glycerol) containing increasing amounts of imidazole. Aliquots of column fractions were analysed by 12% SDS-PAGE stained with Coomassie Brilliant Blue.

Panel A: Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker (molecular weight of marker proteins are shown to the left of the gel image); Lane 2: pellet fraction; Lane 3: soluble load fraction; Lane 4: flow through; Lane 5: wash fractions (buffer A with 25 mM imidazole); Lanes 6 & 7: elution fractions (buffer A with 300 mM imidazole). The 300 mM imidazole protein peak eluted from Ni-NTA agarose was diluted to 25 mM NaCl and applied to S-sepharose. Panel B: Proteins that bound to the S-sepharose were then eluted along an NaCl gradient in buffer A **(**50 mM Tris-HCl pH 7.0, 5.0 mM DTT, 10% glycerol). The elution profile analysed by 12% SDS-PAGE stained with Coomassie Brilliant Blue. Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker; Lane 2: 0.2 M NaCl; Lane 3: 0.25 M NaCl.

3.3.7. Catalytic properties of K338A and wild type TVlig

Preliminary ligation assays were performed using *Hin*d III digested λ DNA fragments (Figure 3.12). However, this method is difficult to quantify. Earlier studies used a radioactively labelled oligo-dT oligomer annealed to a poly-dA polymer strand to measure extent of ligation. The drawback with this method was that it was difficult to know the exact proportion of accurately juxtaposed dT strands on the poly-dA backbone, i.e. it can't be assessed if the two substrates are really next to each other or if ligation is being seen across the gaps (Tomkinson *et al.,* 1991; Odell *et al.,* 1996). Shuman and co-workers have pioneered the use of defined oligonucleotides which have allowed the construction of synthetic nicks *in vitro* where the influence of 3' and

5’ mismatches can also be assessed (Sekiguchi & Shuman, 1997; Sriskanda & Shuman, 1998; Odell & Shuman, 1999). Use of such specific sequences allows exact positioning of the substrate nick and by radiolabelling an oligonucleotide at its 5’ phosphate position the extent of ligation can be quantified by using phosphorimagers. Phosphorimagers are quantitative imaging devices used to localise and quantify radioactive signals and are more sensitive than X-ray films (Robertson *et al.*, 2001).

The ligase activity of the recombinant TVlig was assessed through its ability to seal a single-strand nick in a duplex oligonucleotide. A 42 residue oligonucleotide was synthesised with internal base complementarity to allow folding back and the creation of a hairpin as shown in Figure 3.15, panel A. An 18 residue oligonucleotide (Figure 3.15; panel B) complementary to the single-stranded region of the hairpin oligonucleotide was subjected to ^{32}P end labelling (as described in Chapter 2; section 2.1.10) before being annealled together with the unlabelled 42-mer (ratio 1:4; 18-mer : 42-mer) and used as a duplex substrate in the ligation reactions (as described in Chapter 2; section 2.1.11).

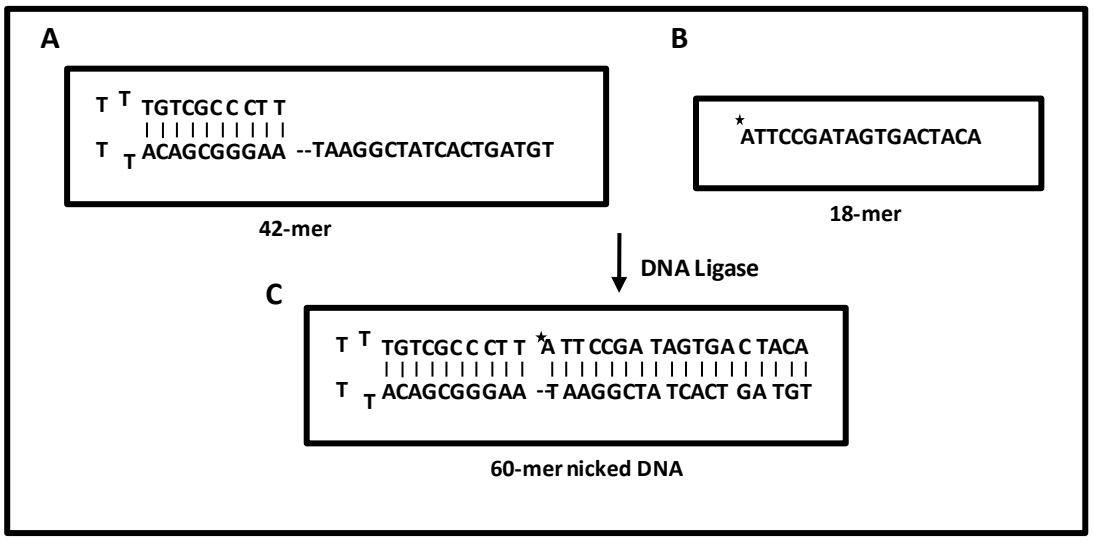


Figure 3.15: Substrates used for standard DNA ligation assays

18-mer oligonucleotide (Panel B) was labelled with ^{32}P in a 20 µl reaction for 60 min at 37^{o}C and gel purified to remove unincorporated nucleotide (described in Chapter 2; section 2.1.10). The quantified labelled strand was annealed with a complementary 42-mer DNA strand (Panel A) in the ratio 1:4 to create the substrate a nicked duplex DNA (60-mer; Panel C). The asterisk indicates the position of the ^{32}P labelled group.

The purified K338A protein and wild type TVlig were tested for strand joining activity on the singly nicked radiolabelled duplex DNA. T4 DNA ligase was used in control ligation reactions as shown in figure 3.16. The reaction products were run on a denaturing polyacrylamide gel, resolved by exposure of the wet gel to a radiation-sensitive screen, and the intensity of the bands was quantified using Imagequant software (Figure 3.16).

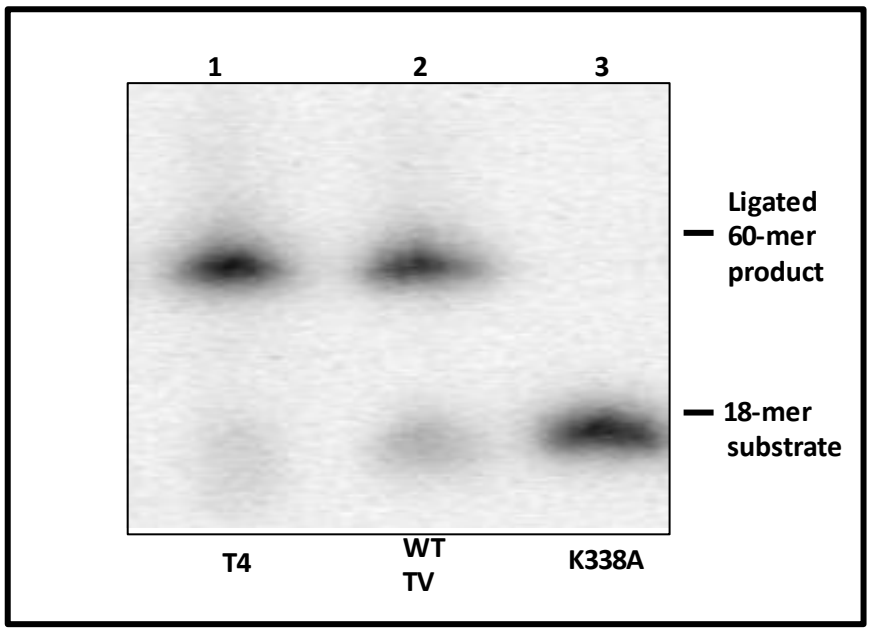


Figure 3.16: Ligation activity of TVlig

The ligase activity assay was carried out using 1 pmole of annealed DNA (18-mer + 42-mer) substrate and 0.1 pmole of the indicated DNA ligase in reaction buffer containing 50 mM Tris-HCl pH 7.5, 5 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP for 2 minutes at 37^{o}C. The reaction products were analysed by electrophoresis through a denaturing 20% polyacrylamide gel and scanned using a Typhoon phosphorimager (GE Healthcare). Lane 1: Positive control with T4 DNA ligase shows the formation of 60-mer ligated product; Lane 2: ligated 60-mer product formed with wild type TVlig; Lane 3: No ligation is seen with active site mutant of TVlig.

Incubation of the purified wild type TVlig with the labelled synthetic nick showed conversion of the labelled 18-mer to a DNA species of higher molecular weight shown by its retardation on a denaturing acrylamide gel (Figure 3.16). The mobility of this higher molecular weight species coincided exactly with that obtained by incubating the same substrate with commercially obtained T4 DNA ligase. The active site mutant, K338A showed no ligation when compared to wild type TVlig and T4 DNA ligase. The loss of ligation activity of the K338A mutant was not unexpected as

the alanine change at Lys338 altered the conserved lysine within what we believed would be the critical, nucleotidyltransferase motif I. This would abolish ligase activity presumptively through preventing the chemical linkage of the enzyme to AMP during step I of the ligation reaction. Similar mutational effects at the motif I lysine have been noted for the NAD^+-dependent *Thermus thermophilus* and *E. coli* DNA ligases, the ATP-dependent PBCV-1 DNA ligase and T4 RNA ligase (Luo & Barnay, 1996; Sriskanda & Shuman, 1996; Sriskanda *et al.,* 1999). From this experiment it was concluded that the ligation activity observed with the TV preparation is as a result of the purified recombinant protein and not as a result of contamination with host, *E. coli* ligase and that this confirms the earlier observation made with lambda *Hin*d III fragment sealing that the enzyme is a DNA ligase using the same constellation of catalytic residues as other polynucleotide ligases.

To understand the biochemical properties of the recombinant TVlig it was decided to profile the enzyme. Initially we decided to define an enzyme concentration where alterations in the reaction conditions would directly affect the formation of reaction products. Previous studies have performed such analyses where the enzyme has been in excess thus analysing effects on single turnover kinetics (Sriskanda & Shuman, 1998). A more accurate reflection of the impact of cofactor changes on ligase activity is revealed where the enzyme is undergoing cycles of ligation incorporating all three steps of ligase chemistry. To assess this, an experiment was carried out incubating TVlig (0.1 pmole) with an excess of DNA substrate (1 pmole). The time course of this reaction is shown in figure 3.17.

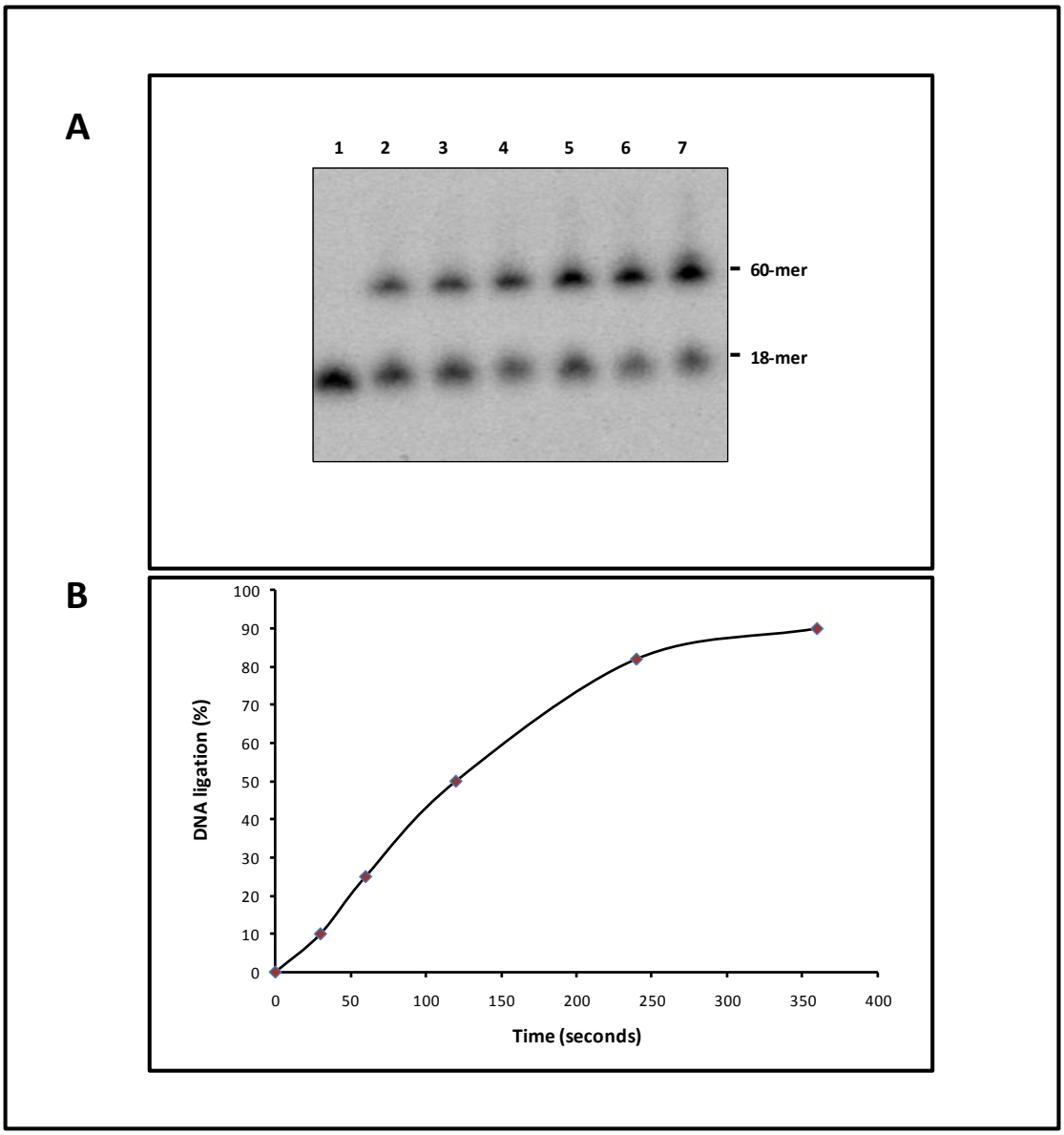


Figure 3.17: Ligation time course of TVlig sealing a 32P-labelled nicked substrate

The time course of ligation reaction was studied by incubating TVlig (0.1 pmole) with an excess of DNA substrate (1 pmole) in reaction buffer containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP and incubated at 37^{o}C. At each time-point, 10 µl of the sample was removed and the reaction was quenched by adding 20 µl loading buffer [98% (v/v) formamide, 10 mM EDTA, 0.01% (w/v) bromophenol blue] followed by incubation at 95^{o}C for 3 minutes. Panel A: Ligation reaction separated on a 20% acrylamide-8M urea gel. Lanes 1 to 7 represent 0, 30, 60, 90,120, 240 and 360 seconds of incubation at 37^{o}C. The sample at time zero was taken before adding the enzyme. Panel B: TVlig activity, measured as the amount of ligated product (60-mer) as a percentage of the total radioactivity in each lane. The gel in panel A was scanned using a phosphorimager (GE Healthcare) and bands were quantified using Imagequant software (GE Healthcare).

The reaction reveals a doubling of the amount of sealed DNA over the first three, thirty second intervals. This suggests that the rate of ligation here is directly determined by the concentration of enzyme. To confirm this finding we then incubated the TVlig at increasing concentrations with a constant amount of ligase substrate. The extent of ligation of the nicked substrate increased linearly with the concentration of TVlig from 0.1 pmole to 2 pmole (Figure 3.18). In the experiment it can be seen that by 30 seconds 0.1 pmole of enzyme has completed sealing of 10% of the substrate - 0.1 pmole. In effect each enzyme molecule completes a single ligation cycle. This is borne out by analysis of the 0.5 pmole enzyme aliquot where 50% of the substrate (0.5 pmole) is sealed in 30 seconds.

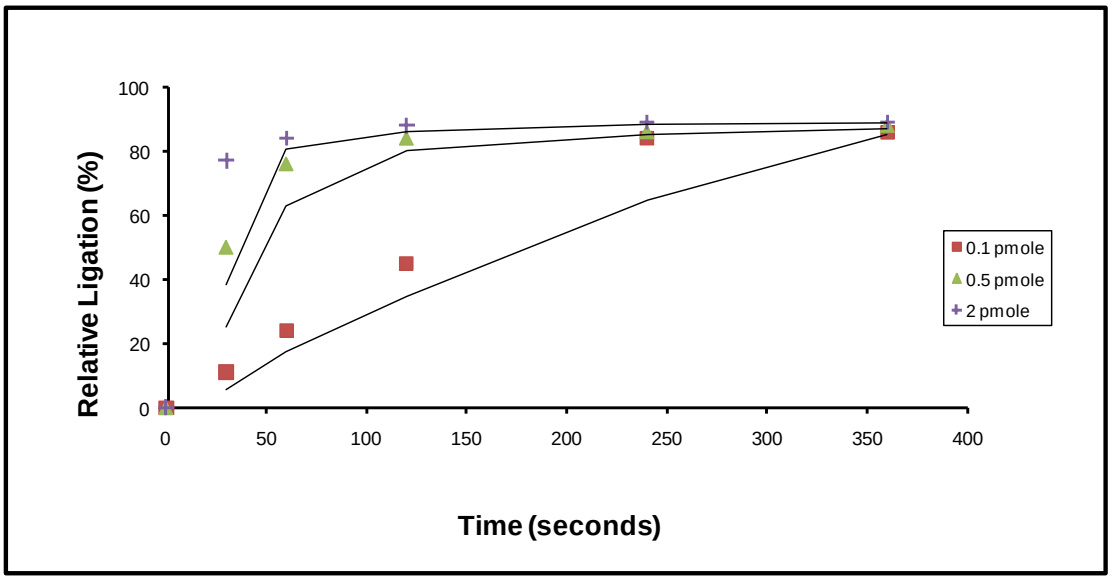


Figure 3.18: Effect of recombinant TVlig concentration on DNA ligation

The DNA ligation assay was carried out to seal 1 pmole of substrate using a range of enzyme concentrations (0.1 – 2 pmoles) in standard reaction buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP). A 20 µl of sample was taken from each reaction mixture at the time indicated and the reaction was quenched by adding 20 µl loading buffer [98% (v/v) formamide, 10 mM EDTA and 0.01% (w/v) bromophenol blue], followed by heating the reaction mixture to 95°C. The samples were resolved by denaturing 20% PAGE. The gel was scanned using a phosphorimager and bands were quantified using Imagequant. The extent of ligation is plotted as a function of time and enzyme concentration.

Approximately, 75% of ligation was completed within 30 seconds when the concentration of the enzyme used (2 pmole) was double the concentration of the substrate (1 pmole) in the reaction mixture. The reaction effectively reached completion by the end of 1 minute for both the 0.5 pmole and 2 pmole enzyme aliquots, when 75-85% of the substrate was converted into product (Figure 3.18).

The time course study of TVlig suggested that the biochemical characterisation of the enzyme could be studied using a concentration of substrate and enzyme in the ratio of 1:10 (enzyme: substrate) when approximately, 50% of ligation is completed in 2 minutes. Such an enzyme to DNA ratio will allow the reaction to be dissected in respect of the effect of changing the reaction conditions, for example temperature or pH, over the initial two minutes of the reaction where the rate is directly dependent on the enzyme.

3.3.8 Properties of the TVlig

3.3.8.1. Effect of temperature on ligation activity of TVlig

The activity of the TVlig was expected to be directly influenced by temperature. The activity of the TV enzyme was determined at different temperatures ranging from 5^{o}C-60^{o}C in the standard ligase assays using the labelled hairpin substrate as described above. The recombinant TVlig activity measured in this assay peaked between 30 - 40^{o}C with optimal activity demonstrated at 35^{o}C (Figure 3.19).

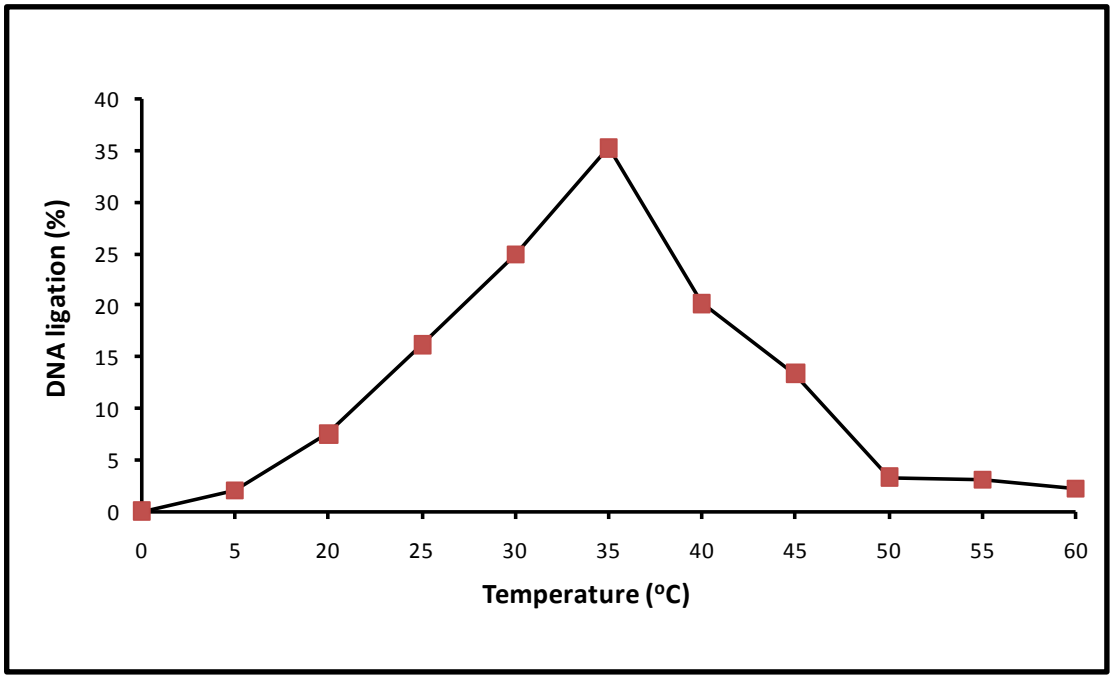


Figure 3.19: Temperature dependence of strand joining activity of TVlig

2 pmole of nicked DNA substrate and 0.2 pmole of TVlig were incubated for 2 minutes at indicated temperatures (^{o}C) in ligation buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM DTT and 1 mM ATP). The extent of ligation in each sample was quantified from densitometric quantification of the sample, once electrophoresed on a denaturing 20% polyacrylamide gel.

Trichomonas vaginalis grows at 30^{o}C - 35^{o}C and the optimal enzyme activity at these temperatures clearly reflects the physiological growth conditions of the parasite serving as source for the ligase. Temperatures higher and lower than the optimal physiological growth conditions of the host and the parasite result in lower activity of the enzyme. Activity of the enzyme was observed to be half that of the 35^{o}C optimal temperature when the temperature was reduced by 10^{o}C.

A similar temperature profile has been observed for *Plasmodium falciparum* DNA ligase (hereafter referred to as Pflig; another protozoan DNA ligase; published contemporaneously with this report), showing optimal activity at 37°C, which is reduced by half when the temperature is lowered to 25°C (Buguliskis *et al.*, 2007). Most of the ATP-dependent DNA ligases exhibit temperature optima between 22°C-37°C, for example, Vaccinia virus DNA ligase works optimally at 32°C (Odell *et al.*, 1996). The exception for ATP-dependent enzymes are those derived from thermophilic archaeons, such as *M. thermoautotrophicum* and *P. horikoshii* that show optimal *in vitro* activity at 60°C and 90°C respectively (Sriskanda *et al.*, 2000;

Keppetipola & Shuman, 2005). It can be concluded that the temperature requirement of these DNA ligases reflects the optimal growth temperature for the source organism.

3.3.6.2 Effect of pH on ligation activity of TVlig

To determine the pH optimum for the TVlig, ligation reactions were carried out over a broad pH range at 37°C with the enzyme in modified reaction buffer C (1 mM DTT, 1 mM ATP and 10 mM $MgCl_2$) with either 50 mM Tris–HCl or 50 mM MOPS or 50 mM sodium citrate buffers. It was found that activity of the enzyme was essentially unchanged over a pH range from 7.0 to 7.5 (Figure 3.20). This pH optima closely parallel those for other DNA ligases; pH 7.8–8.2 for *Drosophila* DNA ligase I (Rabin & Chase, 1987), pH 7.2–7.8 for T4 DNA ligase (Weiss *et al.,* 1968), and pH 7.4–8.0 for mammalian DNA ligase I (Lindahl & Barnes, 1992). A number of ATP-dependent ligases show tolerance of higher pH levels, Pflig has maximal activity at pH 8.5 (Buguliskis *et al.,* 2007) while Vaccinia virus DNA ligase retains >80% of maximal activity at pH 9.0 (Odell *et al.,* 1996). TVlig does not appear tolerant of such high pH levels, retaining only 56% of maximal activity at pH 8.

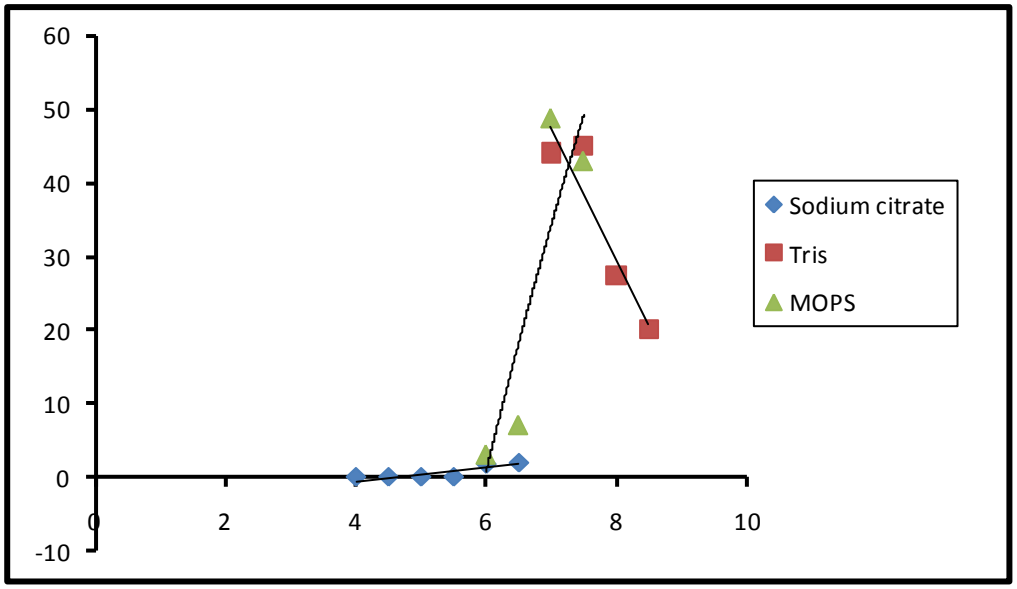


Figure 3.20: Effect of pH on ligation activity of TVlig

TVlig activity profile determined over a broad range of pH (4-8.5) using Tris–HCl (pink line), MOPS buffers (yellow line) and sodium citrate (blue line) with 1 pmole of nicked DNA substrate incubated with 0.1 pmole of TVlig for 2 minutes at 37^{o}C. The reaction products were electrophoresed on a denaturing 20% polyacrylamide gel and scanned using a Phosphorimager (GE Healthcare). The band intensities were quantified using Imagequant software. The extent of ligation is plotted as a function of pH of the buffer used in the assay.

The TVlig showed significantly reduced activity in the acidic pH range (5% at pH 6 compared to 45% percent at pH 7) probably due to instability of the acid labile phosphoamide bond formed in step I of the ligation reaction mechanism. The acid lability of TVlig is extremely similar to the profile of a number of other ATP-dependent DNA ligases such as T4 DNA ligase (Rossi *et al.*, 1997). This is contrary to the pH profile observed for recently identified acid tolerant DNA ligase from archaeon *Ferroplasma acidiphilum* which shows highest activity at pH 2.5–3.0 (Ferrer *et al.*, 2008). This 'pH optimum anomaly' observed for *F. acidiphilum* DNA ligase is an adaptation of the organism to inhabit acidic environments (Ferrer *et al.*, 2008).

3.3.6.3 Divalent cation specificity of TVlig

DNA ligases are metal-dependent enzymes and hence, the nick-joining activity was also determined in the presence of a variety of cations, using the standard buffer conditions for TVlig, but an extended incubation time to reveal all possible divalent metals capable of supporting ligation (Figure 3.21, panel A). Efficient nick-joining was observed in the presence of 10 mM Mg^{2+} (Figure 3.21, panel A & B). Ligation was less efficient with Mn^{2+} and Ca^{2+} at the same concentrations (Figure 3.21, panel A) and other divalent metals, Cu^{2+}, Co^{2+} and Ni^{2+} did not support ligation of the nicked substrate. The enzyme was consistently more active with Mg^{2+} than with Mn^{2+} over a range of concentrations from 2 to 20 mM (Figure 3.21, panel B). Mn^{2+} was observed to enable ligation with an optimal efficiency at 12 mM (Figure 3.21, panel B). Mg^{2+} is considered as the most common metal associated with enzymes as cofactor probably due to its high (millimolar) concentration inside the cell (Andreini *et al.*, 2008). Magnesium has a strong association with phosphate-containing substrates, including ATP (Andreini *et al.*, 2008).

The ability of ligases to use of Mn^{2+} as a cofactor in the absence of magnesium has been reported for a number of enzymes including the T4 DNA ligase, mammalian DNA ligase I, PBCV-1 DNA ligase and Vaccinia virus DNA ligase (Weiss *et al.*, 1968; Lindahl & Barnes, 1992; Odell *et al.*, 1996; Ho *et al.*, 1997). However, the use of Ca^{2+} as a divalent cofactor has only been restricted to the ATP-dependent DNA ligase from *Plasmodium falciparum* (Pflig) and an NAD^{+}-dependent DNA ligase from *Thermus thermophilus* (Tong *et al.*, 1999; Buguliskis *et al.*, 2007). It has been suggested that Ca^{2+} can compete with Mg^{2+} for the essential metal binding site located in the active site of DNA ligases and results in the accumulation of DNA-adenylate intermediates (Sriskanda *et al.*, 2000). This has been observed with *M. thermoautotrophicum* DNA ligase, where Ca^{2+} supported the adenylylation of the 5' pDNA strand in step 2 of the ligation reaction but did not suffice for step three of phosphodiester bond formation; the ligase catalysed attack of the 3'-OH on the activated 5'-PO4 (Sriskanda *et al.*, 2000). Such stalled DNA-adenylate intermediates were not observed for TVlig and Pflig leading to conclusion that ATP-dependent DNA ligases encoded by parasitic protozoans (*P. falciparum* and *T. vaginalis*) have a

unique ability to utilize calcium as a cofactor for the catalysis of phosphodiester bonds. We can be certain that the ligation in the presence of calcium is not due to magnesium that is being retained by the enzyme during purification as the enzyme shows no discernible activity when in the presence of cobalt, nickel or copper.

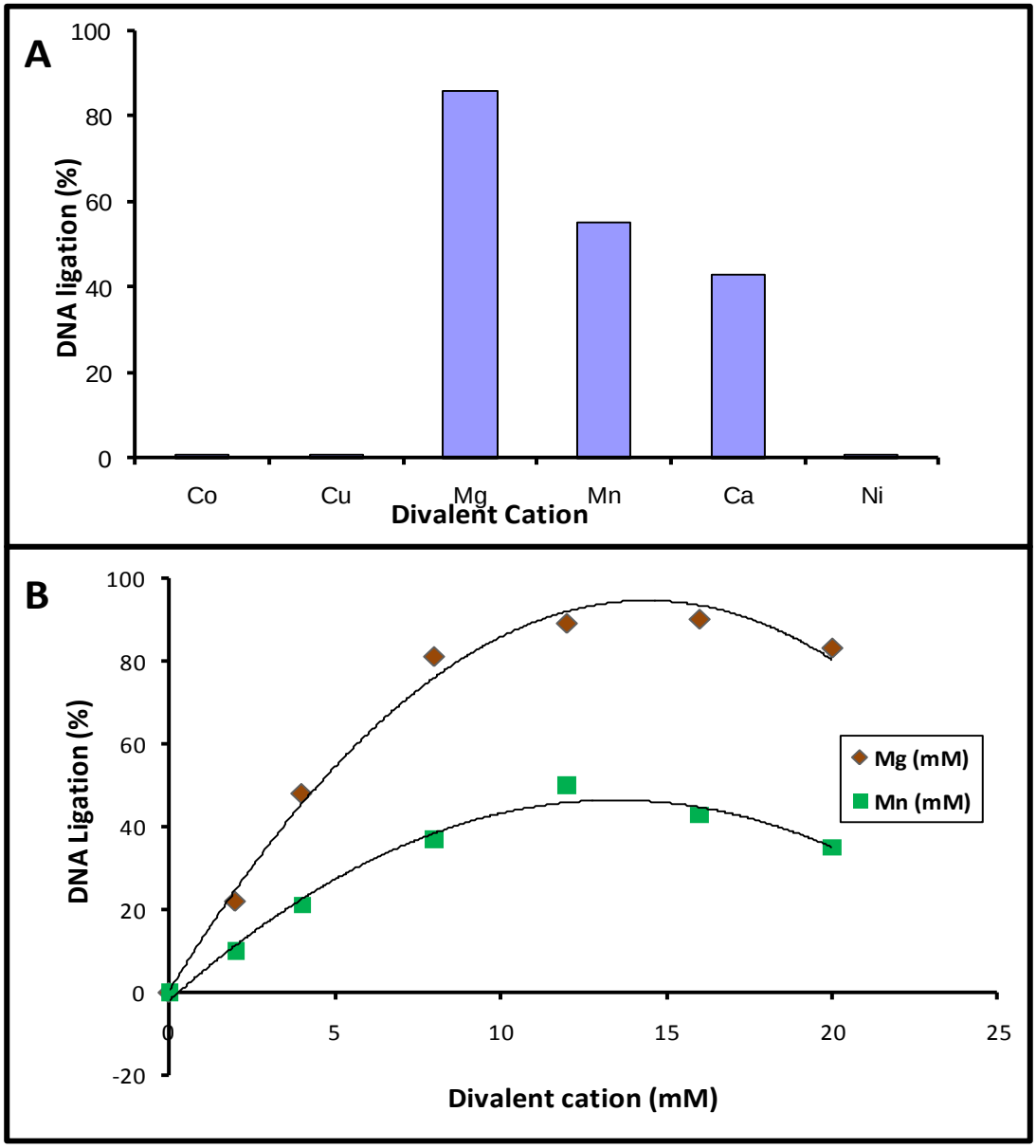


Figure 3.21: Metal cofactor requirement of TVlig

The radioactively labelled nicked 60-bp substrate was used to analyse the cation requirement for *in vitro* ligation activity of TVlig. Panel A : A ligation reaction was carried out with 1 pmole of DNA substrate incubated with 0.1 pmole of TVlig for 5 minutes at 37°C in the presence of 10 mM of one of the following divalent metals, $MgCl_2$, $MnCl_2$, $CaCl_2$, $CuCl_2$, $CoCl_2$ and $NiSO_4$. Ligation products were resolved by denaturing PAGE and quantified using a Phosphorimager and Imagequant software. The extent of ligation is plotted as a function of divalent cation concentration. Panel B: A ligation reaction was carried out with 1 pmole of DNA incubated with 0.1 pmole of TVlig for 5 min at 37°C using increasing concentrations (2 mM, 4 mM, 8 mM, 12 mM, 16 mM & 20 mM) of $MgCl_2$ and $MnCl_2$. Ligation products were resolved by denaturing PAGE and quantified using a Phosphorimager and Imagequant software. Ligation activity is plotted as a function of divalent cation concentration.

3.3.6.4 Nucleotide cofactor specificity of TVlig

In addition to divalent metal ions, DNA ligation reactions also require a nucleotide cofactor. Ligases are generally classified by which of the cofactors, ATP or NAD^+, support their activity. The nucleotide specificity of the TVlig was determined by examining which of the following nucleotides, ATP, NAD^+, GTP (the nucleotide cofactor of mRNA capping enzymes) and dATP were capable of ligation (Figure 3.22). It was observed that like many other DNA ligases recombinant TVlig was able to perform DNA ligation even in the absence of added cofactor indicating that at least part of the purified recombinant enzyme was in the adenylated form (Figure 3.22). When enzyme concentration equalled or exceeded the substrate concentration, it was not necessary to add additional ATP for efficient ligation. This is a remarkable finding that virtually 100% of the enzyme molecules purified from *E. coli* are in the adenylated state and thus one round of nick-sealing is possible without the need for additional cofactor (Figure 3.22). Nick sealing activity in the absence of added ATP has similarly been reported for a number of other enzymes for example in PBCV-1 DNA ligase. However, the level of adenylation has been more modest, studies have indicated that only 11% of enzyme molecules have AMP bound at the active site (Sriskanda & Shuman, 1998).

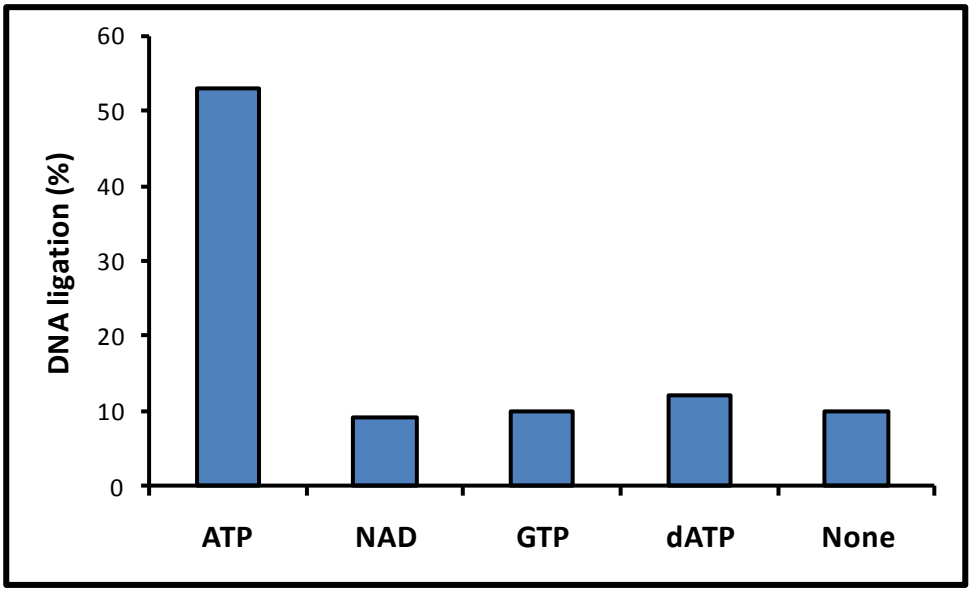


Figure 3.22: Nucleotide cofactor specificity of TVlig

The nucleotide cofactor specificity of TVlig was studied by incubating 0.1 pmole of TVlig with 1 pmole of DNA substrate in reaction buffer containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT) and 1 mM NTP (ATP, NAD^+, GTP) or dNTP (dATP) and incubated at 37^{o}C for 2 minutes. The extent of ligation in each sample was quantified from analysis of the sample on a denaturing polyacrylamide gel using a Phosphorimager and Imagequant software.

The ligation activity of TVlig was not stimulated by the addition of 1 mM dATP nor the same concentration of NAD^+ or GTP, thereby indicating that the enzyme has an absolute requirement for ATP in nick sealing reactions (Figure 3.22). The ligase fraction used in the above assay was not de-adenylated prior to incubating with substrate, hence the activity observed in the presence of NAD^+ and the other cofactors is contributed by the pre-adenylated TVlig and not by the cofactor. This is evident by the same amount of ligation taking place in the absence of any co-factor at all (Figure 3.22).

Nick-sealing reactions of TVlig were then performed with a range of ATP concentrations (0.4-1.2 pmole) in order that the K_m value for ATP could be determined. At each ATP concentration the formation of ligation products was plotted against time and the initial velocities [V] were determined for these graphs (as attached in appendix 7.3; Figure 7.3). A K_m of 0.458 µM ATP was calculated from the double-reciprocal plot of these values (Figure 3.23).

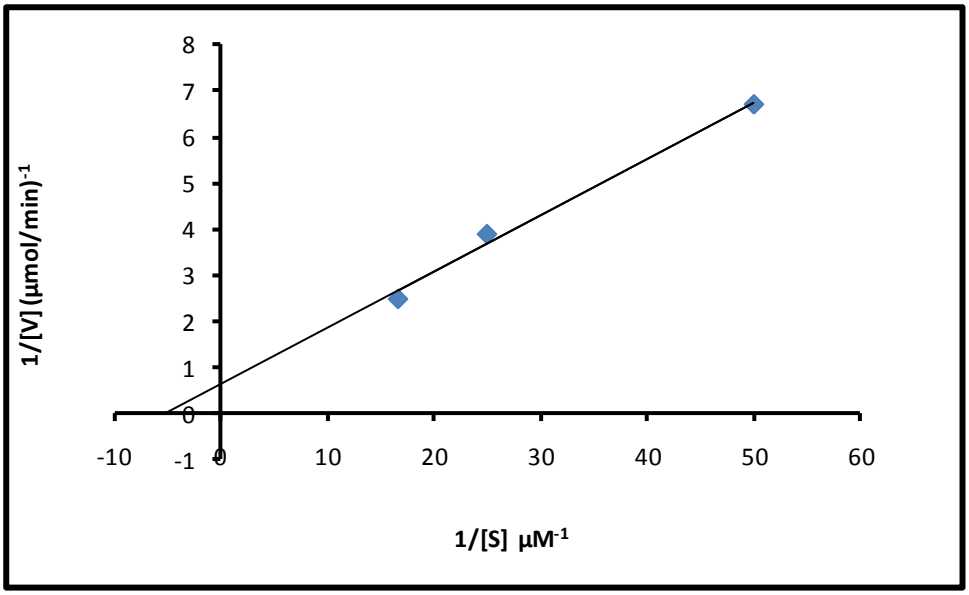


Figure 3.23: Determination of the Km for ATP of the TVlig

Different concentrations of ATP (0.4 pmole to 1.2 pmole) were used to determine the binding affinity of the TVlig using 1 pmole of nicked labelled substrate and 0.1 pmole of the ligase. Each ligation reaction was performed in a reaction buffer containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 5 mM DTT and incubated at 37°C (Appendix 7.3). At each time-point, 10 µl of the sample was removed and the reaction was quenched by adding 20 µl loading buffer [98% (v/v) formamide, 10 mM EDTA, 0.01% (wt/vol) bromophenol blue] followed by incubation at 95°C for 3 minutes. The reaction products were electrophoresed on a denaturing 20% polyacrylamide gel and scanned using a Phosphorimager (GE Healthcare). The change in the extent of ligation for each concentration of ATP used was determined and used as measure of the velocity of reaction [V]. The double-reciprocal plot of these values was used to determine Km of TVlig for ATP.

This K_m value is small as compared to 75 µM of PBCV-1 DNA ligase (Ho *et al.*, 1997) and 95 µM of Vaccinia virus DNA ligase (Odell *et al.*, 1996). The K_m value of TVlig indicates that its ATP binding affinity is more similar to ligases of multicellular organisms, for example mammalian DNA ligase I (1 µM, Lindahl & Barnes, 1992) and Pflig (1.36 µM, Buguliskis *et al.*, 2007).

3.3.6.5. Nucleic acid binding specificity

Earlier footprinting studies, based on PBCV-1 DNA ligase and T4 DNA ligase, have revealed an asymmetric binding profile of DNA ligases (Odell & Shuman 1999; Doherty & Dafforn, 2000). Physical binding analysed by minimal substrate analysis have suggested that the ligase needs only 6 nucleotides to seal DNA efficiently (Odell *et al.*, 2003). Unpublished observations (Odell & Brooke, 2007) from our laboratory suggest that a 21 nucleotide dimer is able to support the binding of two PBCV-1 (34

kDa) DNA ligase molecules. Therefore, native gel mobility assay provides a means of determining whether the enzyme can in fact bind ends as well as nicks. The nick, in the annealed 60-mer DNA substrate, is located 18 nucleotides from the end of the molecule (Figure 3.15) a distance that all available data suggests will allow a second ligase molecule to bind to the terminus should it be capable. Obviously by virtue of having the hairpin only one DNA end will bind DNA. By using a simple gel shift analysis we can suggest whether the enzyme can recognise ends. If the enzyme recognises both nicks and ends then we will expect two shifted products however only a single species might be expected if the ligase is binding centrally.

A native gel mobility shift assay was employed to examine the binding of purified recombinant TVlig to the ^{32}P-labeled nicked duplex DNA (Odell & Shuman, 1999). Binding reactions were performed in the absence of magnesium so as to preclude conversion of substrate to product during the incubation (Odell & Shuman, 1999). Mixing the wild type TVlig with nicked substrate resulted in the formation of a single discrete protein–DNA complex that migrated more slowly than the free DNA during electrophoresis through a 10% native polyacrylamide gel (Figure 3.24). The abundance of this complex increased in proportion to the amount of input ligase. In order to estimate binding affinity, the gel was scanned using a phosphorimager. The dependence of protein–DNA complex formation on input ligase is shown in figure 3.24.

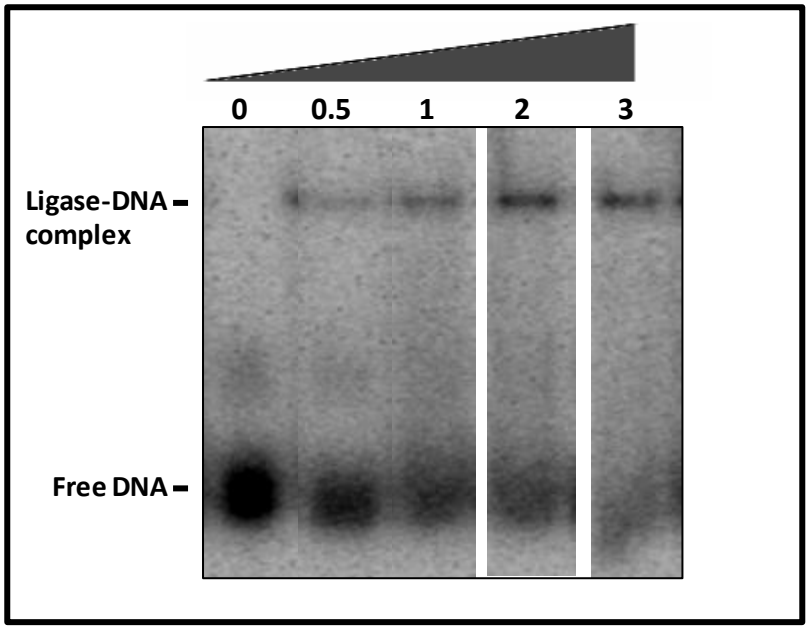


Figure 3.24: DNA binding native gel assay

Binding of TVlig to nicked substrate with a 5' phosphate group at the nick is shown. Binding reaction mixtures (20µl) containing 50 mM Tris-HCl pH 7.5, 5 mM DTT, 1 pmole of nicked ligand and 0.5, 1, 2 and 3 pmole of TVlig were incubated for 10 min at 22^{o}C. Lane 1: Control reaction contained no ligase; Lane 2-5 shows effect of increased ligase concentration from 0.5 – 3 pmole respectively. Glycerol was added to 5% and the samples were electrophoresed through a 10% polyacrylamide gel in TBE buffer at 100 V for 2.5 hours at 4^{o}C. An autoradiogram of the gel scanned using a Typhoon Phosphorimager is shown (Figure 7.3; Appendix 7).

3.3.6.6 DNA binding by TVlig

Classical analysis of DNA ligases interacting with DNA substrates, including PBCV-1 DNA ligase, has been undertaken by gel-shift (Ho *et al.*, 1997; Sriskanda & Shuman, 1998). The technique has allowed a number of insights into ligase-DNA interaction (Odell *et al.,* 2000) however it is limited to study protein-DNA interactions that result in stable binding. Alternative methods such as optical biosensor techniques like SPR can be employed to analyse real-time biomolecular (dynamic) interactions, such as those of proteins with DNA; even if those interactions do not result in the formation of stable complexes. Therefore, it was decided to analyse DNA binding by TVlig via SPR using Biacore system. The Biacore experiments were performed with help from C. Brooke, University of Westminster.

The Biacore 2000 instrument was used for all experiments (Biacore AB; Uppsala, Sweden). The biotinylated DNA substrates (Nick, Gap and Duplex; Figure 3.25) were annealed (as described in Chapter 2, 2.1.11), diluted to 0.5 µM with Biacore buffer 1 and manually injected at a flow rate of 20 µl/min over individual cells of a Biacore SA chip (as described above in section 3.2.8).

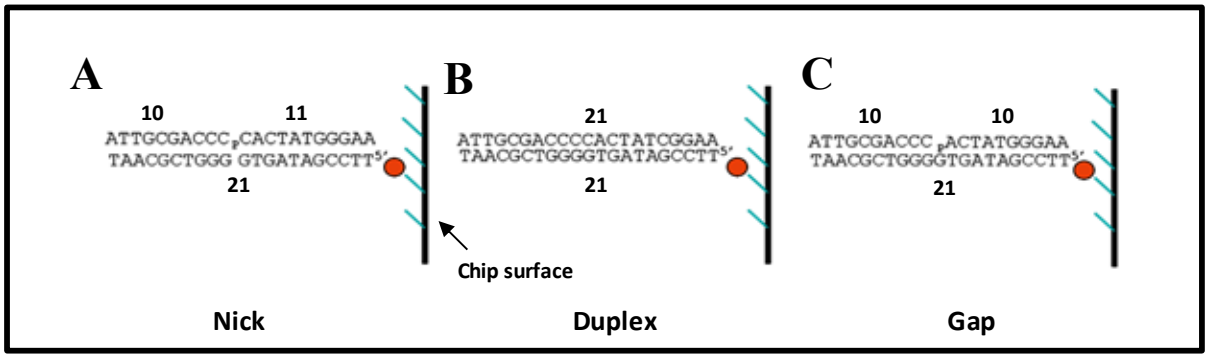


Figure 3.25: Covalent attachment of biotinylated substrates to strepavidin-coated SA sensor chip

Biotinylated dsDNA substrates were immobilised onto a streptavidin-coated SA chip. Panel A shows the nicked DNA substrate (Nick; top strand composed of the 10-mer + phosphorylated 11-mer) immobilised in cell 2, panel B shows the double-stranded duplex DNA substrate (Duplex; top strand 21-mer) immobilised in cell 3 and panel C shows the substrate with a 1 nucleotide gap (Gap; top strand composed of the 10-mer + phosphorylated 10-mer) immobilised in cell 4. The biotin group is indicated by the red circle attached to the 5'terminus of the bottom strand, the streptavidin groups are indicated by the green lines.

Following the immobilisation of approximately 200 Resonance Units (RU) of each substrate, a buffer wash was carried out and the baseline was monitored to ensure substrate stability. The DNA substrates, dsDNA with a centrally placed nick (Nick), dsDNA with a 1 nucleotide gap (Gap) and un-nicked duplex DNA (Duplex), utilised in this study were chosen as they had been previously applied to analyse DNA binding by PBCV-1 DNA ligase via gel shift analysis (Ho *et al.*, 1997; Sriskanda and Shuman, 1998).

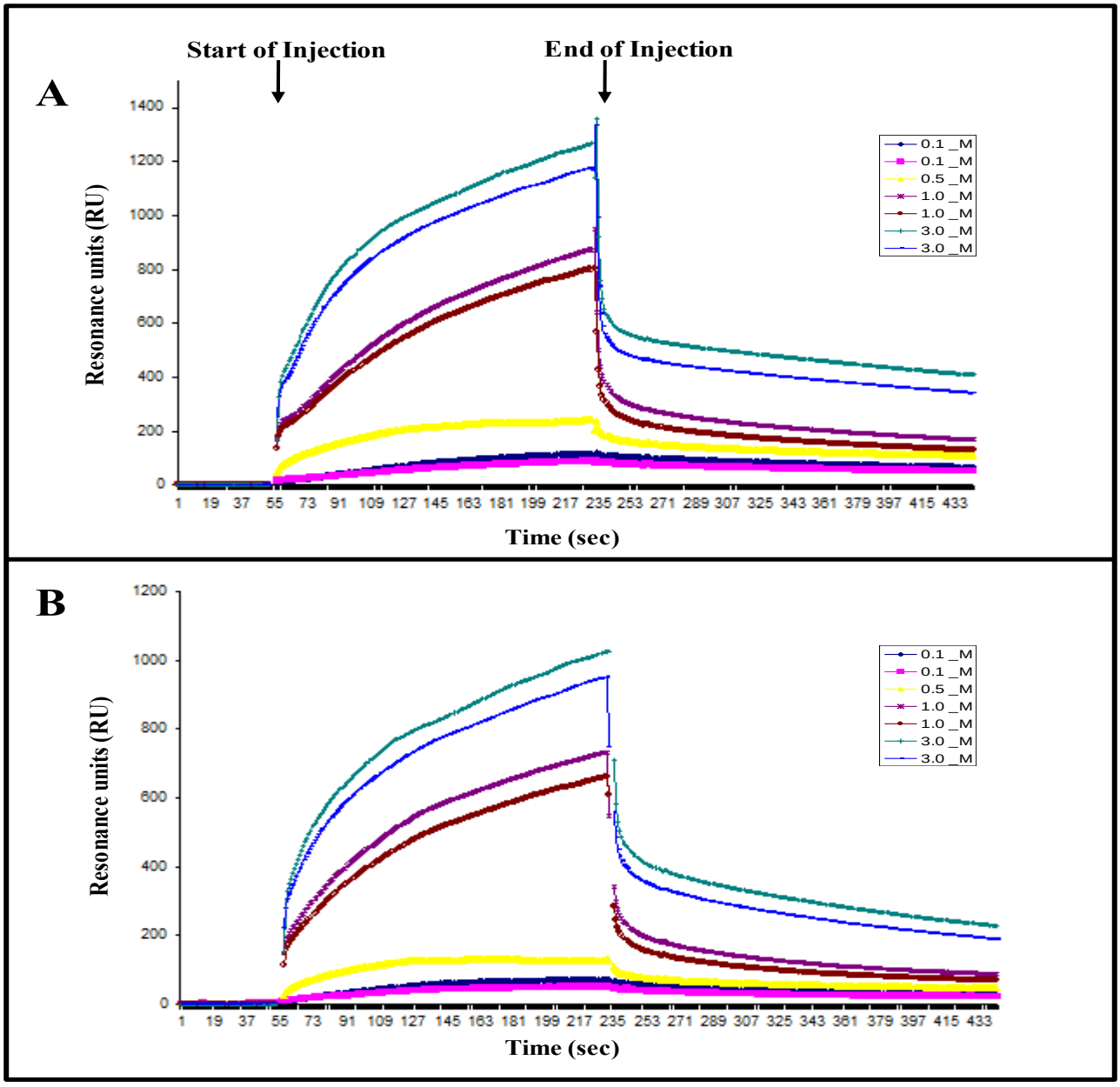


Figure 3.26: Biacore analysis of TVlig binding to duplex and gap DNA substrate

The binding profile of TVlig (0.1 – 3.0 μM) was assessed by 3 minute injections (30 μl/min flow rate) over the immobilised DNA substrate (200 RU). Panels A & B show the binding profile of TVlig to duplex DNA and the gap DNA substrate respectively. The start and end of the injection are indicated by black arrows. The values from the underivatised cell were subtracted and the baseline zeroed. The experiment was performed with help of C. Brooke, University of Westminster.

Multiple dilutions of TVlig (0.1 – 3 μM protein, in duplicates except 0.5 μM) were injected over the derivatised Biacore SA chip and the binding profile of the enzyme over each substrate was observed. Analysis of the DNA binding profile of TVlig reveals that the enzyme has a fast on-rate, highlighted by a rapid increase in RU following the injection of the protein (Figure 3.26). The ligase does not seem to show substrate discrimination over gap and duplex DNA as the binding profile and off-rate showed over these substrates was quite similar (Figure 3.26; panels A and B). The RU values increased with increasing protein concentration. Binding of approximately 150 RU and 100 RU (for 0.5 μM) was observed for duplex and gap substrate respectively (Figure 3.26). This is in contrast with the DNA binding profile observed for PBCV-1 DNA ligase. The PBCV-1 DNA ligase shows greatest association with the intact, duplex DNA substrate (RU480 for 0.5 μM) and the lowest binding was observed with the substrate containing the one nucleotide gap (RU of approximately 110 for 0.5 μM; Figure 3.27, panel B; Brooke, PhD thesis, 2007).

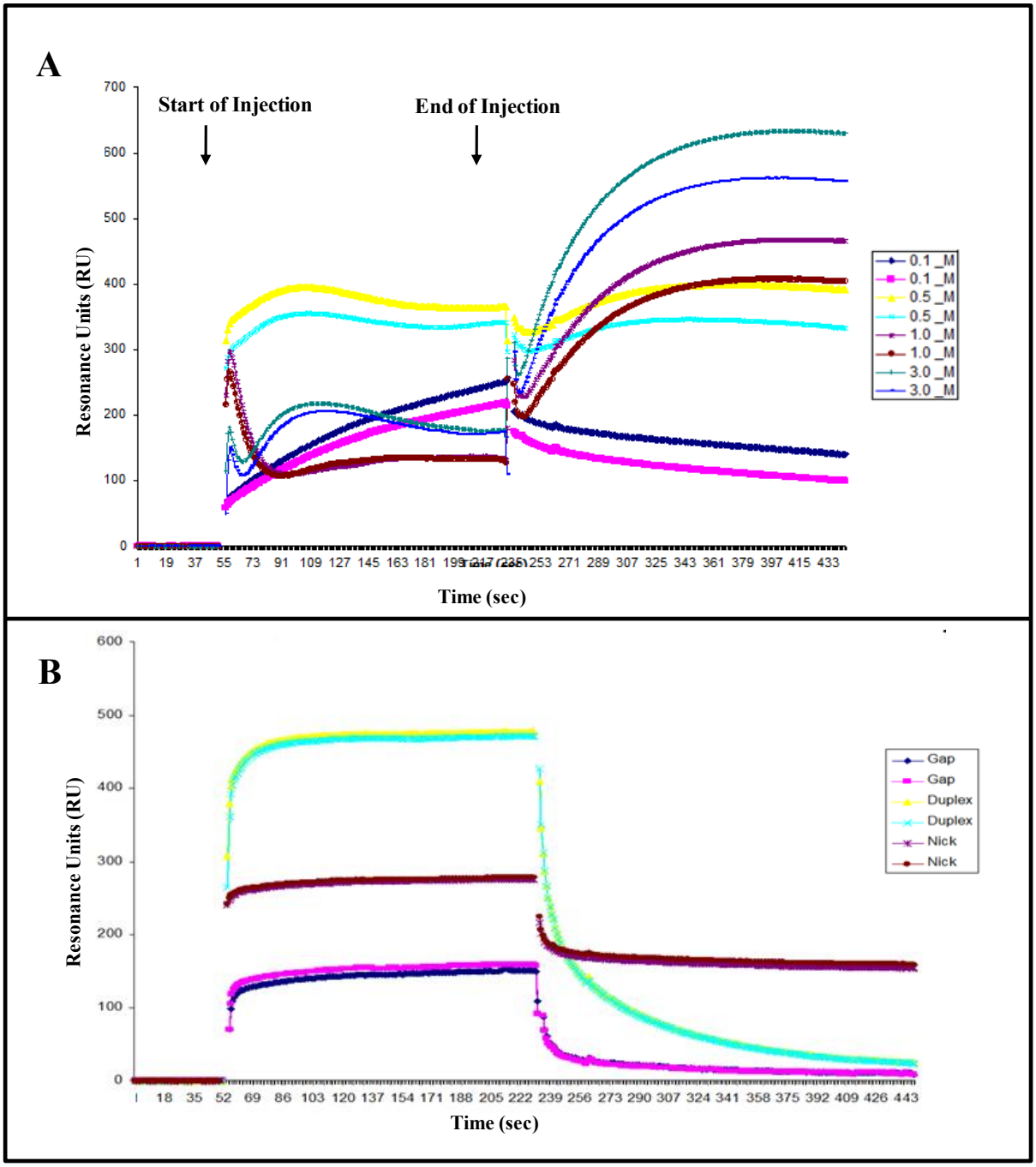


Figure 3.27: Comparison of DNA binding of TVlig and PBCV-1 DNA ligase

Panel A shows the DNA binding curve for TVlig (0.1 – 3.0 µM) assessed by 3 minute injections (30 µl/min flow rate) over the immobilised nicked DNA substrate (200 RU). In panel B, the binding of the PBCV-1 DNA ligase (0.5 µM) was assessed by duplicate 3 minute injections (30 µl/min flow rate) over the immobilised substrates. The enzyme was diluted to approximately 50 mM NaCl. The Gap substrate is represented by dark blue and magenta, Duplex by yellow and cyan and the Nick by dark purple and brown. The start and end of the injection are indicated by black arrows. The values from the underivatised cell were subtracted and the baseline zeroed. The figure in panel B is reproduced by kind permission of C.Brooke (PhD thesis, 2007), University of Westminster.

The binding profile for TVlig to the nicked substrate (Figure 3.27, panel A) is significantly different to that observed for PBCV-1 ligase (Figure 3.27, panel B; Brooke, PhD thesis, 2007) and HuLigI-Δ232 (Figure 3.28; Brooke, PhD thesis, 2007). At lower concentrations, 0.1 and 0.5 μM, it showed a maximum binding of approximately 190 and 380 RU respectively (Figure 3.27, panel A) as compared to the 10 and 100 RU observed for the HuLigI-Δ232 (Figure 3.28). At the end of the injection with PBCV-1 and HuLig1 biphasic dissociation was observed for 0.1 μM – an initial rapid dissociation of material that was not stably bound, followed by a second phase showing very little decrease in RU; indicative of stable binding to the substrate. Unlike the PBCV-1 ligase and the truncated human enzyme, a change in the shape of the binding curve was observed at higher concentrations of TVlig (1-3 μM). The cause of the change in shape of binding curve at these ligase concentrations is unknown although it may have occurred as a result of a conformational change in the structure of the enzyme.

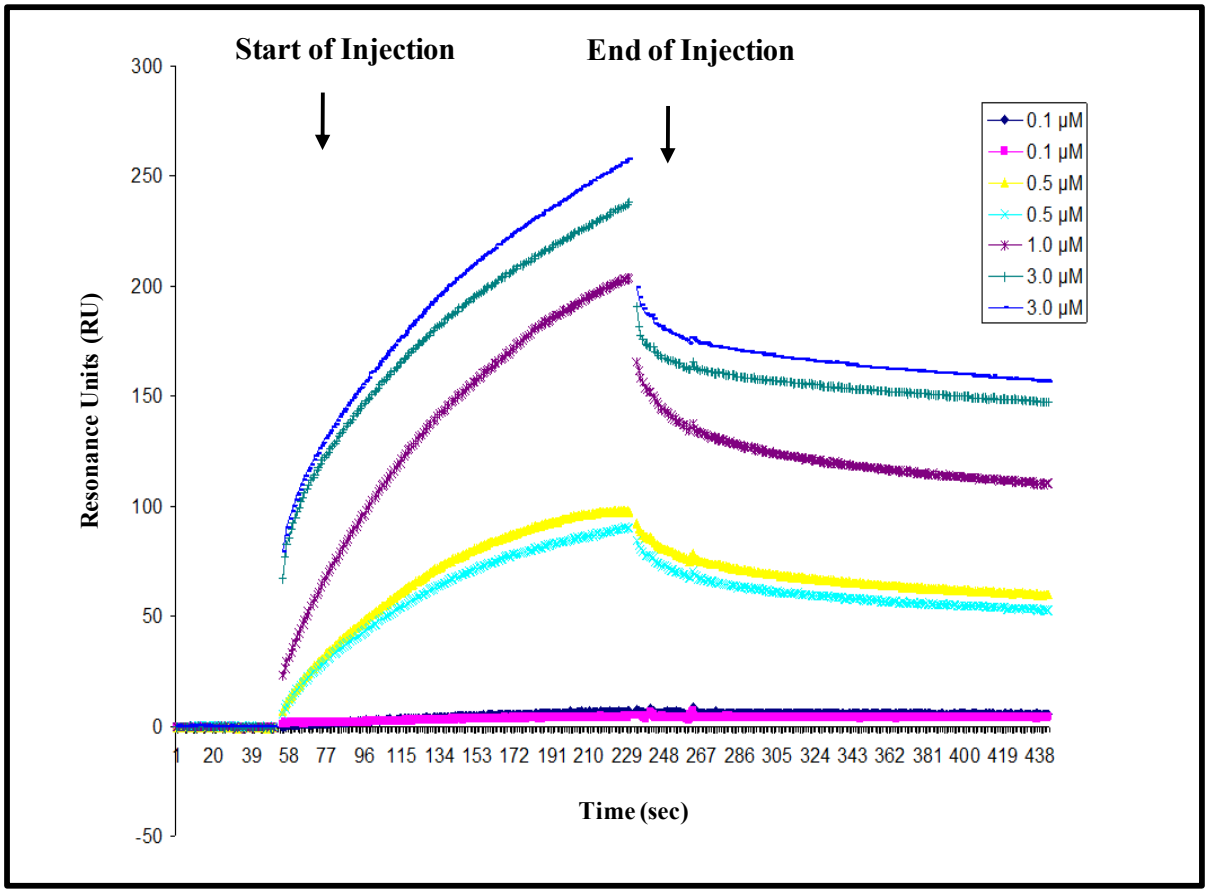


Figure 3.28: The DNA binding curve observed for HuLigI-Δ232 binding on the nick

HuLigI-Δ232 (0.1 – 3.0 µM) was injected for 3 minutes at a flow rate of 30 µl/min over the nicked DNA substrate (described in Chapter 3, 3.2.8), averaged duplicate runs are shown: The values from the underivatised cell were subtracted and the baseline zeroed. The start and end of the injection are indicated by black arrows. The figure is reproduced by kind permission of C.Brooke (PhD thesis, 2007), University of Westminster.

The above experiments confirms that TVlig binds to singly nicked duplex DNA but it is not clear how DNA ligase I might bind to double-strand DNA breaks. Double-strand breaks (DSBs) in DNA are generated when the two complementary stands of the DNA double helix are broken simultaneously at sites that are insufficiently close to one another, such that base-pairing and chromatin structure are insufficient to keep the two DNA ends juxtaposed (Takata *et al.*, 1998). One of the major pathways employed by the cells to repair DSBs is non-homologous end joining (NHEJ). The main proteins required for NHEJ are the Ku heterodimer (Ku70/80), DNA-PKcs (the catalytic subunit of DNA-PK, the DNA-dependent protein kinase), XRCC4 (X-ray-cross complementation gene 4; LifI in yeast), DNA ligase IV and XLF (an XRCC4-like factor; also called Cernunnos; Wyman *et al.*, 2004; Hefferin & Tomkinson,

2005). The Ku heterodimer binds to ends of DNA at DSBs and initiates the recruitment process of additional components of the NHEJ pathway, including ligase IV. A type IV ligase is found in all eukaryotic organisms including unicellular species such as yeast however there was no evidence of such an enzyme when we analysed the TV genome.

As previously stated, our bioinformatics analysis suggests that a homologue of DNA ligase IV is absent in the TV genome. Furthermore, other core members of the NHEJ pathway, especially, Ku70/Ku80 also appear absent. BLAST searches of the TV genome with human, yeast and drosophila Ku proteins failed to identify any homologue. We hypothesised that the TVlig might function in the parasite genome as a dimer, which binds to and aligns DNA ends at DSBs, thereby, compensating for the loss of Ku heterodimer. Such a protein species would probably only allow one dimer to bind to our DNA substrate and as such result in only a single species under gel-shift analysis.

To test this hypothesis, an additional step of zonal velocity sedimentation was employed. This sedimentation method involves the separation of particles by centrifugation through a fluid of increasing density. The protein components in a mixture, sediment through the gradient in separate zones based on their sedimentation rates which are determined by differences in their size, shape and density (Tanese, 1997). This analytical approach has been employed in the analysis of various enzymes including, DNA ligases and topoisomerases (Cheng & Shuman, 1997; Keppetipola & Shuman, 2005). The native size of TVlig was gauged by sedimentation through a 15–30% glycerol gradient containing 0.5 M NaCl (as described in Chapter 2; section 2.2.5). Marker proteins - catalase (248 kDa), bovine serum albumin (BSA; 66 kDa) and cytochrome *c* (12 kDa) were included as internal standards during sedimentation. Aliquots of the even-numbered gradient fractions were analyzed by SDS-PAGE (Figure 3.29).

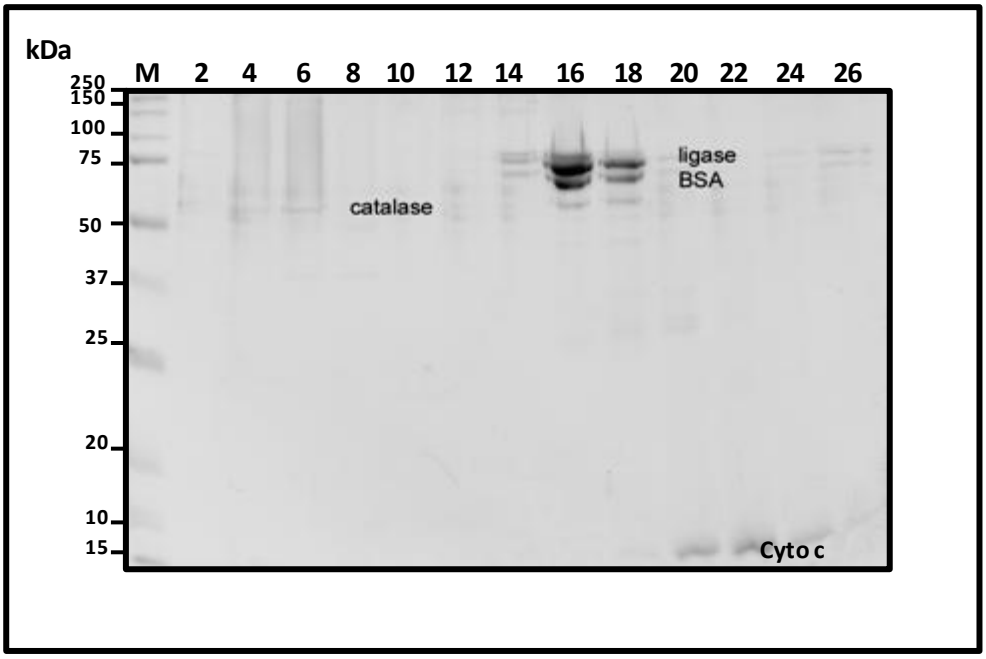


Figure 3.29: Gradient sedimentation profile of TVlig

An aliquot (70 µg) of the purified, wild type TVlig was mixed with catalase (50 µg), bovine serum albumin (BSA; 50 µg), and cytochrome *c* (50 µg). The mixture was applied to a 5-ml 15–30% glycerol gradient containing 50 mM Tris–HCl pH 7.0, 500 mM NaCl, 2 mM DTT, 1 mM EDTA, and 0.01% Triton X-100. The gradient was centrifuged at 36,800 x g for 17 hours at 4°C. Aliquots (18 µl) of the even-numbered gradient fractions (2-26) were analysed by denaturing 12% SDS-PAGE. The Coomassie blue-stained gel is shown. Bio-Rad broad range pre-stained protein standard molecular weight marker proteins were resolved in the extreme left lane (M; the position and sizes of the marker proteins are indicated to the left of the gel). The positions of the ligase polypeptide and the internal standards catalase, BSA, and cytochrome (cyto) *c* are indicated.

The 76 kDa TVlig polypeptide sedimented as a single component, coincident with the BSA peak, in fractions 16–18. It was concluded from the sedimentation chromatography that the TVlig, similar to other eukaryotic DNA ligases, exists as a monomer.

3.4 Discussion

3.4.1 *Trichomonas vaginalis* DNA ligase

This work began as an attempt to establish Ligase enzymes as targets for therapeutic intervention. DNA ligases inhibitors have been previously proposed as antibacterial agents to selectively target the NAD^+-dependent replicative ligase enzyme of bacteria. DNA ligases from the protozoal family had never been analysed and so their

biochemistry and properties had not been explored. We chose this opportunity to characterise a DNA ligase from the parasitic protozoan *Trichomonas vaginalis*. We report here the identification, expression and characterisation of the single DNA ligase encoded by the *Trichomonas vaginalis* (TV) genome.

DNA ligases belong to a superfamily of covalent nucleotidyltransferases which include the polynucleotide ligases and the mRNA capping enzymes. These proteins show a wide variation in size, ranging from for the smallest described 298-amino acid PBCV-1 DNA ligase to >900 amino acid for some of the mammalian DNA ligases such as 919 amino acid human DNA ligase I (Tomkinson *et al.,* 2006). However, all these proteins contain a catalytic domain which includes six motifs that are conserved in both sequence and spacing (Odell *et al.,* 2000). The active site contains a reactive lysine residue in the sequence -KYDGER- which is common to these enzymes from a wide range of organisms including bacteriophages, yeasts and human (Sriskanda & Shuman, 2002a). In the ligase proteins, large portions of the regions N terminal and C terminal to the catalytic domain can be deleted without a subsequent loss in enzymatic activity (Tomkinson *et al.,* 2006). Since the PBCV-1 DNA ligase contains only 26 amino acid upstream of the lysine motif I at the N-terminus and no amino acids downstream of motif VI at the C-terminus, it represents the catalytic core of the nucleotidyltransferase supefamily (Figure 1.8, panel A; Odell *et al.,* 2000).

Eukaryotes can have at least three DNA ligase species (I, III and IV), which show sequence identity with each other but DNA ligase III has been identified only in vertebrates (Timson *et al.,* 2000; Martin & MacNeill, 2002). In other species or lower eukaryotes, for example in plants, Drosophila and fungi, only two species of DNA ligase (I and IV) are found. In each case a single replicative ligase is present. *Saccharomyces cerevisiae* encodes two different DNA ligases, Cdc9 and ligase IV, which play distinct roles in DNA replication, repair and recombination. Cdc9 catalyses the joining of Okazaki fragments during DNA replication and it functions in the base excision and nucleotide excision repair pathways. DNA ligase IV in yeast plays a specialized role in the repair of double-strand DNA breaks via the NHEJ pathway. The presence of multiple DNA ligases in eukaryotic cells suggest that theses

enzymes are assigned different functions and deletion of any of these genes in mammalian cell lines have exhibited different phenotype (Tomkinson *et al.,* 2006).

Sequence analysis of the entire *Trichomonas vaginalis* genome identified a gene encoding a protein with significant homology to the replicative DNA ligase. The open reading frame for the ligase-like protein was identified through alignment studies of TV genome from the TIGR database with the yeast ligase I and IV sequences (*Schizosaccharomyces pombe*), small viral DNA ligase sequences (PBCV-1) and with archael ATP-dependent ligases (*Sulfolobus solfataricus*)**.**

The ORF of the putative DNA ligase bears 54% similarity with *Drosophila melanogster* DNA ligase I, 53% with yeast DNA ligase I and 51% with HuLigI (Table 3.3). Alignment studies suggested that identified ORF has the six key nucleotidyltransferase motifs (I-V) that constitute the catalytic core of DNA ligases. The order and spacing of conserved motifs in the putative ligase is similar, as its sequence conservation, to that seen in other ATP-dependent ligase family members such as Vaccinia virus DNA ligase, PBCV-1 DNA ligase and HuLigI (Figure 3.5; panel A). The domain architecture of TVlig as deduced from the sequence alignment studies consists of a N-terminal PCNA interacting domain, DNA binding domain and catalytic core - comprising of AdD and OB-fold domains (Figure 3.5, panel B). The predicted tertiary structure of TVlig suggests close structural alignment with known structure of HuLigI (Figure 3.30).

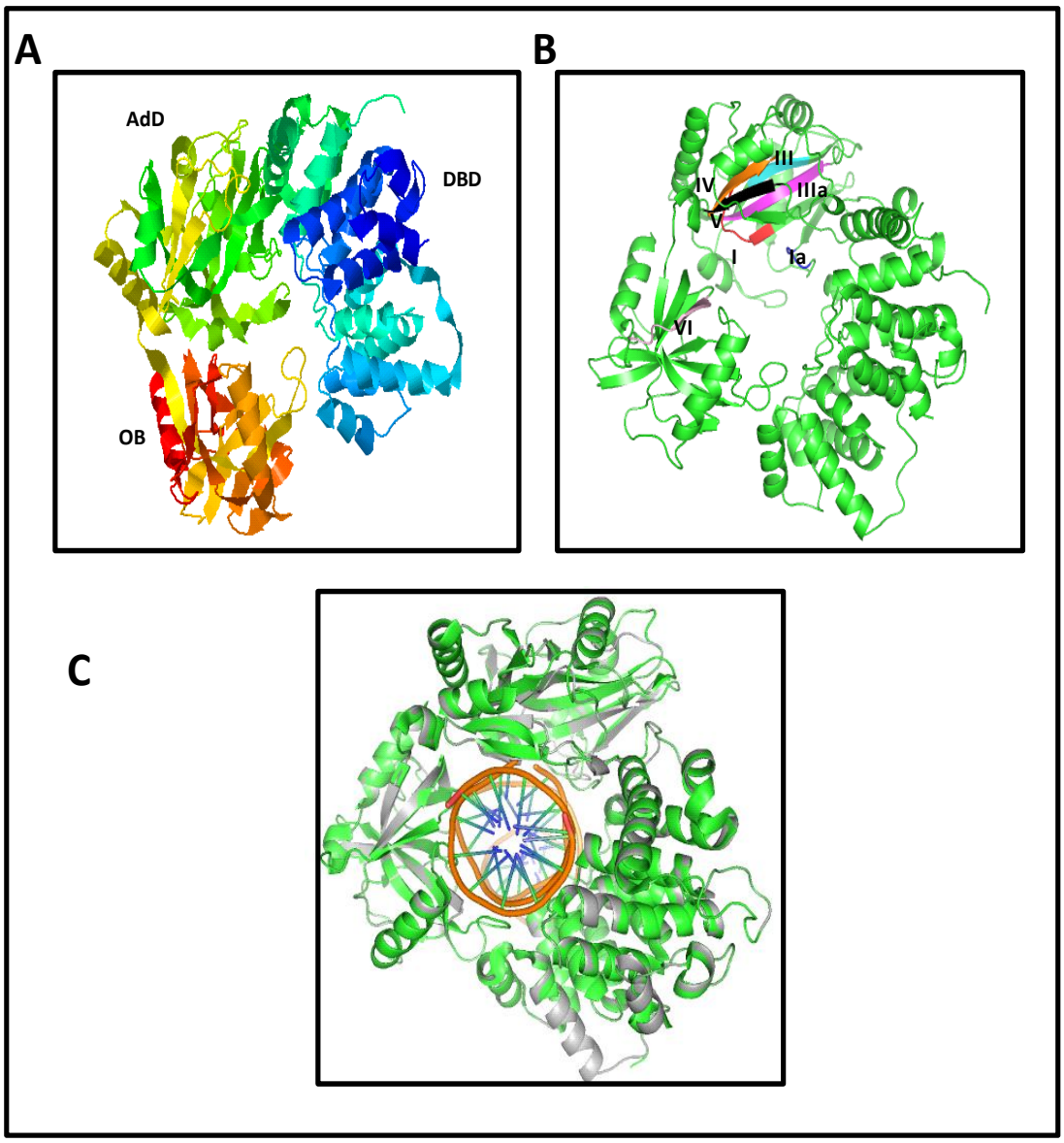


Figure 3.30: Predicted tertiary structure of TVlig

TVlig structure was predicted using the structural predictive programme from *Phyre* (Panel A; Kelley & Sternberg, 2007). Panel B shows the position of six signature motifs of DNA ligases that form the adenylation domain (AdD; I- red; Ia-blue; III-cyan; IIIa-purple; IV-orange; V-black) and moif VI (pink) present in OB domain. Panel C: Ribbon diagram showing TVlig structure (grey; as predicted from *Phyre*) aligned to structure of human DNA ligase I enzyme in complex bound to adenylated DNA (green; 1X9N; Pascal *et al.*, 2004;). The image in panel A was prepared with Rasmol (Bernstein, 2000) and images in panel B and C were prepared PyMol (DeLano, 2002). Abbreviations: AdD - Adenylation domain; OB - Oligomer binding domain; DBD - DNA binding domain.

The bioinformatic analysis suggests that TVlig, similar to HuLigI, uses DBD to encircle and stabilise the nicked DNA substrate in co-ordination with the conserved AdD and OB-fold domains (Figure 3.6; Figure 3.30; Pascal *et al.*, 2004). The catalytic core of TVlig, comprising of adenylation domain (AdD) and oligomer binding (OB)

domain, has conserved amino acid residues; functions of which have been highlighted earlier in structural and mutational studies of other DNA ligases (Figure 3.5, panel B). For example, motif I (KYDGER) encompass the conserved lysine residue (Lys338) that constitutes part of the ATP binding pocket of the enzyme and motif III has conserved aspartate (Asp387) and glutamate (Glu389) residues required for step I and III of catalysis (Odell *et al.*, 2000). Similarly, motif V has a pair of conserved lysines (Lys512 and Lys513) essential to form contacts with the ribose sugar hydroxyl group of AMP (Figure 3.5, panel A; Sriskanda & Shuman, 2002). The predicted size of TVlig (76 kDa) is closer to yeast DNA ligase I (83 kDa) but smaller than higher species such as human DNA ligase I (101 kDa) which could be due to the evolutionary complexity of the organism. To better understand the TVlig and to confirm its function, the polypeptide was cloned, sequenced and expressed from *E. coli.* as a recombinant His-tagged fusion protein which was purified to near homogeneity by a two-step column chromatographic procedure.

Initial characterization of the ligation activity of this enzyme was performed by using *Hin*d III digested λ DNA. Preliminary results confirmed that the enzyme required ATP consistent with the bioinformatic analysis of the sequence of the polypeptide. Ligation of the compatible ends of the *Hin*d III digested λ DNA required the addition of macromolecular crowding agents such as polyethylene glycol (PEG 8000; Figure 3.12). PEG is known to function as a volume exclusion factor, increasing the apparent concentration of DNA and thereby possibly facilitates the ligation. The surprising observation here was that the efficient ligation of cohesive ends should require such a crowding agent. Previous studies with T4 DNA ligase have suggested that macromolecular crowding is only required for blunt-ended DNA fragments (Hayashi *et al.*, 1986).

A more detailed, quantified analysis of TVlig reaction chemistry was performed using a synthetic singly nicked duplex DNA consisting of a 42-nucleotide 3'-OH terminated hairpin DNA and a 5' ^{32}P-labeled 18-mer strand annealed to the 5'-tail of the hairpin strand (Figure 3.15). The results obtained during biochemical characterisation of TVlig suggested that the purified protein was able to catalyse phosphodiester bond formation on nicked DNA substrates and observed that the product formation was

similar when compared to equimolar T4 DNA ligase (Figure 3.16). The active site mutant K338A TVlig, constructed by substituting the lysine (Lys338; K) residue with alanine (A) in the active site motif I, failed to show any ligation activity. This lysine residue is the active site of AMP transfer by the ligases and GMP transfer by the capping enzymes (Odell *et al.,* 2000). Absence of ligation by the active site mutant highlights the importance of the residue during the step I of the ligation reaction mechanism (Figure 3.16) and confirmed that the ligation activity purified was as a result of our recombinant protein expression.

Analysis of the activity of TVlig in the presence of different co-factors identified optimal ligation in the presence of ATP with K_m of 0.458 µM and a divalent metal. The ligation of the nicked DNA substrate was observed in the absence of added ATP which suggested (Figure 3.22) that a fraction of purified TVlig I was pre adenylated. Adenylated form of DNA ligases in the cell are stable and also enhance their DNA nick sensing capability (Odell *et al.,* 2000; Ellenberger & Tomkinson, 2008). The lack of ligation in the presence of NAD^+ was consistent with our summation that the TV genome elaborates an ATP-dependent replicative ligase as is found in all other eukaryotic organisms. The essential divalent cation cofactor, was found to be most readily satisfied by Mg^{2+}; however, Mn^{2+} and Ca^{2+} were found to sustain 50% and 40% ligation respectively at the same concentrations. Structural studies of DNA ligases suggest that the divalent metals presumably orient the phosphate atoms of AMP and the DNA as well as helping to neutralize the charge during phosphoryl transfer reactions (Nair *et al.,* 2007). Previous studies have suggested that other ligases can use of Mn^{2+} as their divalent cofactor, including T4 DNA ligase, mammalian DNA ligase I, PBCV-1 DNA ligase and Vaccinia virus DNA ligase (Weiss *et al.,* 1968; Lindahl & Barnes, 1992; Odell *et al.,* 1996; Ho *et al.,* 1997). However the use of Ca^{2+} for efficient nick sealing has only been reported in the NAD^+-dependent DNA ligase from *Thermus thermophilus* (Tflig) and ATP-dependent DNA ligase from *P. falciparum* (Pflig; Tong *et al.,* 1999; Buguliskis *et al.,* 2007). This behaviour is attributed to the large ionic radius of Ca^{2+} (1.00 Å) than Mg^{2+} (0.72 Å) or Mn^{2+} (0.80 Å) requiring more flexibility within the catalytic core of DNA ligase (Buguliskis *et al.,* 2007).

The location of Ca^{2+} in the active site of T4 RNA ligase I structure suggests that it competes with Mg^{2+} for an essential metal-binding site in these enzymes (RNA ligases are a member of the nucleotidyltansferase superfamily and share the same catalytic core with DNA ligases; Wang *et al.,* 2006). DNA ligases are able to use Ca^{2+} to catalyse the formation of DNA-adenylate intermediate but are unable to convert the nick into a phosphodiester bond (except in the cases of TVlig, Pflig and Tflig). We conclude that protozoal DNA ligases (TVlig and Pflig) have evolved to catalyse phosphodiester bond synthesis in the presence of Ca^{2+}, although at least 50% less efficiently than when compared to Mg^{2+}, but without the accumulation of DNA-adenylate intermediates in step 2 of catalysis.

TVlig showed optimal ligation activity on a nicked DNA substrate in the presence of ATP and $MgCl_2$ at 30^oC-38^oC (Figure 3.19). The temperature optima of TVlig I suggests that the enzyme is adapted to functioning in physiological temperature of the host similar to that observed in Pflig (Buguliskis *et al.,* 2007). The pH profile of TVlig is similar to that observed for T4 DNA ligase. The ligase shows optimal activity at pH 7-7.5 and loses its activity by 80% in slightly acidic pH (Figure 3.20). Whilst previous authors have commented that this reflects the acid lability of the phosphoamidate lys-AMP bond there have not been extensive studies to confirm that this is the case.

Alignment of the amino acid sequence of the TVlig to that of HuLigI suggest that this is the homologue of the DNA ligase I and as such is likely to be responsible for ligase events associated with DNA replication. TVlig shares 32% and 27% identity with human DNA ligase III and IV respectively, with highest homology restricted to DNA binding domain and common catalytic core shared by all members of nucleotidyltransferase family. Detailed alignment studies are attached in appendix 7.2.1. The presence of additional DNA ligase III and IV specific domains such as BRCT domain that target these ligases to sites where their action is required, seem to be absent in the TVlig. Instead the amino acid sequence of TVlig revealed a probable NLS in the N-terminal region of the enzyme (Figure 3.4, panel B). NLS are required for active transport of enzymes into the nucleus through nuclear pores and are characterised by the presence of lysine (K) and/or arginine (R) clusters which in some

cases are preceded by a proline (Dingwall & Laskey, 1991). A NLS has been identified in HuLigI at residues 119–131, deletion of which abolished nuclear trafficking of the protein (Montecucco *et al.*, 1995; Cardoso *et al.*, 1997).

DNA ligase I has a conserved PCNA-interacting polypeptide (PIP) motif. This is essential for proteins that interact with PCNA and is found in DNA polymerase, FEN-1 (flap endonuclease-1), the clamp loader RFC (replication factor C), CAF1 (chromatin assembly factor), DNA repair factors such as MutS, XPG (xeroderma pigmentosum complementation group G) and UDG (uracil DNA glycosylase) and a number of cell-cycle-regulatory proteins such as p21; (Vivona & Kelman, 2003; Vijayakumar *et al.*, 2007). The archaeal PCNA binding proteins, *e.g.* archaeal DNA polymerase B, generally interact with their PCNA via a conserved PIP motif located in their C-terminal region except for the DNA ligases from *Sulfolobus solfataricus* and *Pyrococcus furiosus* (PfuLig; Dionne *et al.*, 2003; Kiyonari *et al.*, 2006). The PIP motif of PfuLig is instead located in the loop structure connecting α helices in the N-terminal DNA binding domain (Kiyonari *et al.*, 2006). TVlig, similar to HuLigI (2-9 amino acid residues) and yeast DNA ligase I (cdc9; 18-25 amino acid residues), apparently has a PIP motif located in the N-terminal portion (3-10 amino acid residues) of its DNA binding domain (Figure 3.4, panel A; Warbrick, 2000; Vijayakumar *et al.*, 2007).

It is interesting to note that *P. falciparum's* genome (another parasitic protozoan) encodes for two homologues of PCNA gene but the recently characterised DNA ligase I (published contemporaneously with this report) from this organism lacks an obvious PIP motif. Instead this polypeptide has an apicoplast targeting signal located in its N-terminus (Buguliskis *et al.*, 2007). The apicoplast is a plastid organelle, homologous to the chloroplasts of plants, that occurs throughout the Apicomplexa; a group of protists characterized by the presence of apicoplast (Foth & McFadden, 2003). The absence of the PIP motif raises interesting questions about the mode of interaction of Pflig with PCNA that is usually required during the process of replication and repair in eukaryotes. However, not all PCNA-interacting proteins possess a PIP motif, as reported for proteins like the histone acetyl transferase (HAT) p300, and histone deacetylase 1 (HDAC1; Stoimenov & Helleday, 2009). It could be

hypothesised that *P. falciparum* DNA ligase I uses some unusual PCNA interacting motif similar to that recently reported in the archaeal organism *Pyrococcus furiosus* (Kiyonari *et al.,* 2006). PCNA interacting differences have been reported for *P. falciparum*. The recently identified *P. falciparum* FEN-1 has a PIP motif located internally as opposed to C-terminal location observed in other FEN-1 homologues (Casta *et al.,* 2008).

3.4.2. DNA Repair in *Trichomonas vaginalis*

DNA repair processes are vital to all living organisms; loss of which invariably results in accumulation of mutations and/or cell death. Many of the DNA lesions repair pathways have been conserved throughout eukaryotes and many eukaryotic enzymes have homologues in prokaryotes (Taylor & Lehman, 1998). There are no reports of DNA repair proteins or pathways that exist in TV or other related parasitic protozoans such as *Toxoplasma gondii* or *Trypanosoma cruzi.* We attempted to identify putative TV homologues for DNA repair proteins using tBLASTn to shed light on whether we would expect that there would be a requirement for additional ligases, such as a type III or IV enzyme. Table 3.4 lists some of the putative genes present in TV genome that might be involved in DNA repair pathways (when compared to DNA repair proteins in *S.cerevisiae*). Single-strand DNA (ssDNA) breaks are repaired either by short-patch base excision repair (BER) or long-patch BER whereas double-strand breaks (DSB) are repaired by non-homologous end joining (NHEJ) or homologous recombination (HR). Mammalian cells uses short-patch BER for single ssDNA repairs and NHEJ for DSB repairs. Short patch BER utilises DNA glycosylases, AP endonucleases, XRCC1 and DNA ligase III whereas long patch BER is PCNA dependent pathway involving FEN-1, replication protein A (RPA) and DNA ligase I (described in depth in Chapter 1; section 1.6.1.3**).**

Table 3.4: Hypothetical TV proteins that may participate in different DNA repair pathways

Gene Name	Present in TV	*S.cerevisiae* Protein ID	*T.vaginalis* Protein ID
Long patch base excision repair			
PCNA	Y	AAS56041	XP_001329442
Rad1	Y	P06777	XP_001329017
Cdc9 (DNA ligase I)	Y	CAA48158	XP_001581589
Rad27 (FEN-1)	Y	CAA81953	XP_001294157
Short patch base excision repair			
DNA ligase III	**N**	-	-
XRCC1	**N**	-	-
Homologous recombination (HR)			
Mre11 (Rad32)	Y	BAA02017	XP_0013291917
Rad50	Y	CAA65494	XP_001311080
Rad51	Y	CAA45563	XP_001303202
RPA1	Y	NP_009404	XP_009404
Non-homologous end joining (NHEJ)			
Ku70	**N**	**NP_014011**	**-**
Ku80	**N**	**NP_013824**	**-**
DnL4 (DNA ligase IV)	**N**	**CAA99193**	**-**
Lif I (XRCC4)	**N**	**NP_011425**	**-**

Amino acid sequences of DNA repair proteins from *S.cerevisiae* were collected from NCBI GENBANK, and compared to the TV protein and nucleotide database using BLASTP and tBLASTN (significance was defined arbitrarily as an e-value of 10^{-5} or less). Presence (Y) or absence (N) of repair proteins in TV is indicated, along with the genbank accession numbers for both *S.cerevisiae* and TV proteins. Proteins lacking obvious homologues are indicated in bold type. Pathway components were largely compiled from the following sources: Base excision repair from Boiteux & Guillet, 2004; HR from Filippo *et al.*, 2008; NHEJ from Hefferin & Tomkinson, 2005.

Haltiwanger *et al.* (2000) observed that cell-free lysates prepared from *P. falciparum* repaired regular and synthetic AP sites in various DNA substrates by long-patch BER. This mechanism was different from mammalian cells, yeast, *Xenopus* and *E. coli* which predominantly repair such AP sites by removal of a single nucleotide or small-patch BER. This could be because of the absence of proteins involved in short-patch BER, specifically XRCC1 and DNA ligase III in the *P. falciparum* genome (Buguliskis *et al.*, 2007). Recent characterisation of FEN-1 and other genes such as

DNA ligase I and AP endonucleases from *P. falciparum* suggests that the malarial parasite has the entire core enzymes required for long-patch BER pathway and hence uses the above mechanism for repairing damages in single-strands of DNA (Casta *et. al.,* 2008). The TV genome also appears to encode for homologues of FEN-1, PCNA and DNA ligase I (characterised in this study) the enzymes required for long-patch BER but is devoid of DNA ligase III and XRCC1 required for short-patch BER. This lead us to speculate that *T. vaginalis* being related to *P. falciparum* and having only one DNA ligase in its genome is likely to use long-patch BER for single-strand DNA break repair.

Double-strands breaks (DSB) are the most serious forms of DNA damage and a single unrepaired DSB can lead to cell death. These types of damage are repaired predominantly by either non-homologous DNA end joining (NHEJ) or homologous recombination (Burton *et al.,* 2007). NHEJ pathway relies on DNA ligase IV, XRCC4 (LifI in yeast) and the Ku70/Ku80 heterodimer (Figure 1.9; panel A). Extensive BLAST searches revealed that the TV genome is devoid of the core machinery required for DSB repairs by NHEJ. We searched the genomes of three other related parasites: *Trypanosoma cruzi*, *Leishmania major* and *Plasmodium falciparum* but orthologues for DNA ligase IV and XRCC4/LifI were not detected. However, homologues of DNA ligase I were apparent in *Trypanosoma cruzi* (EAN97122) and *Leishmania major* (CAJ02559) genome (DNA ligase I is already characterised in *Plasmodium falciparum*). The lack of Ku proteins in the TV genome is highly interesting. These proteins play a pivotal role in NHEJ; they bind to and align DNA DSB sites and recruit other repair factors to the break site and are present in Archaebacteria, Prokaryotes and Eukaryotes. The core of Ku proteins is highly conserved from prokaryotes to eukaryotes (Hefferin & Tomkinson, 2005).

A native gel mobility shift was used to demonstrate the formation of stable complex between the TVlig and nicked DNA (Figure 3.24). A more sensitive, Biacore SPR analysis - a solution study that facilitates the investigation of the dynamics of DNA-ligase interactions, was used to study TVlig binding to different, immobilised DNA substrates (Figure 3.25). It was observed that TVlig binds equally well to the duplex as to the gap DNA substrates (Figure 3.26) but its binding affinity was different for nicked DNA substrate. This suggests that TVlig can differentiate nicks from duplex

DNA. At low concentrations, the enzyme rapidly reached equilibrium over the nicked substrate and following the end of the injection, dissociation of some, unbound ligase molecules was observed. The shape of the binding and dissociation curve observed at higher TVlig concentrations was quite unusual. It was initially thought that this behaviour of TVlig bound to nicked DNA substrate is a result of enzyme oligomerisation. We hypothesised the existence of TVlig as a dimer which could bind to and align DNA ends at DSB sites but velocity sedimentation analysis clearly demonstrated that TVlig exists as a monomer, similar to other eukaryotic DNA ligases. One additional experiment that might be attempted would be to repeat the velocity sedimentation with a blunt-ended DNA duplex. It may be that the enzyme can self-associate when such DNA is present and that by this means the organism can seal DSBs.

Verkaik *et al.* (2002) demonstrated that the inactivation of the factors involved in the NHEJ pathway (Ku80, DNA-PK, DNA ligase IV and XRCC4), in mammalian cells, results in increase of micro homology mediated end joining (MHEJ) of the DNA substrate. This DNA repair pathway uses short stretches of homology or micro homology near the DNA ends (Glover *et al.*, 2008). Although MHEJ had been previously reported to help NHEJ-mediated repair of DSBs, the MHEJ pathway is distinct from NHEJ because it does not rely on the Ku, DNA ligase IV and XRCC4 proteins for completion (Decottignies, 2007). It has been suggested that micro homology end joining (MHEJ) can act as a back-up pathway where DSBs were not repaired by HR or NHEJ (Nussenzweig & Nussenzweig, 2007). MHEJ is dependent on Rad1, Mre11 and Rad50 in yeast and FEN-1 in mammalian cells (Glover *et al.*, 2008).

Trypanosoma brucei's (*T. brucei)* genome, a parasitic protozoan, does not encode homologues of DNA ligase IV or XRCC4, suggesting that NHEJ is also absent in these protozoans (Conway *et al.,* 2002b). Conway *et al.* (2002a) created KU null mutants of *T. brucei* by deleting both alleles of Ku70 and Ku80 (components of NHEJ). However, no increased sensitivity of this mutant to methyl methane sulphonate (MMS; causes both single and double-strand breaks) and to phleomycin (causes double-strand breaks) was detected, indicating the existence of a repair pathway independent of the Ku proteins (Conway *et al.*, 2002a). Burton *et al.* (2007)

have observed that DSBs in *T. brucei* cell extracts are repaired by sequence micro homology which is distinct from NHEJ and is independent of the Ku –heterodimer. These is similar to findings of Glover *et al.* (2008), who demonstrated that chromosomal DSB repairs in *T. brucei* were predominantly repaired by MHEJ and homologous recombination.

Liang *et al.* (2008) studied the role of DNA ligase I in end joining of DNA double-strand breaks by micro homology mediated end joining (MHEJ) in mammalian cells and demonstrated that human DNA ligase IV is not required in MHEJ. Comparisons were made by down regulating DNA ligase I in human HTD114 cells which lead to impaired joining of DNA ends mediated by micro homology and treatment of HTD114 nuclear extracts with antiserum against DNA ligase I significantly reduced MHEJ (Liang *et al.,* 2008).

All the above findings and lack of evidence for the core machinery of NHEJ (to date) in the TV genome suggests that double-stranded breaks are most likely repaired by micro homology mediated end joining pathway (MHEJ). It is difficult to imagine how the organism could exist without any means to mend DSB DNA. TV must, therefore, utilize some other form of double-strand break repair, or contain such highly divergent copies of most NHEJ and HRR proteins that were impossible to identify in this study. A more detailed study of *Trichomonas vaginalis* DNA repair proteins will determine whether controlling the disease caused by these pathogenic parasites can be undertaken through inhibiting these components.

CHAPTER 4

PRODUCTION OF VARIOUS ATP-DEPENDENT DNA LIGASES FOR USE IN DEVELOPING A LIGASE-INHIBITOR ASSAY

CHAPTER 4: PRODUCTION OF VARIOUS ATP-DEPENDENT DNA LIGASES FOR USE IN DEVELOPING A LIGASE-INHIBITOR ASSAY

4.1 Introduction

Research over the past few years has provided ample evidence that genome instability caused by germline mutations in DNA repair genes can cause a predisposition to carcinogenesis. For example, defects in the BRCA1 and BRCA2 proteins strongly predispose individuals to breast and ovarian cancer (Turner *et al.,* 2005). DNA repair mechanisms also play a critical role in sensitivity and resistance of tumour cells both during and after anti-cancer drug treatment and irradiation. Commonly used chemotherapeutics such as, doxorubicin, cisplastin and etoposide induce DNA damage, which in turn generates a complex cascade of events leading to cell cycle arrest, transcriptional and post-transcriptional activation of a subset of genes including those associated with DNA repair and ultimately the triggering of apoptosis (Christmann *et al.,* 2003). The cytotoxicity of the majority of these drugs is directly related to their ability to cause DNA damage (Madhusudan & Hickson, 2005).

The range of DNA damage induced by chemotherapeutic drugs and clinically employed radiation can activate multiple DNA repair pathways such as base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end joining (NHEJ; Hoeijmakers, 2001). Tumour cells that are deficient in one or more DNA repair pathway depend upon the remaining pathways for damage repair (Turner *et al.,* 2005). Therefore, therapeutic inhibition of those remaining pathways or particular components of these pathways should make a crucial impact on the survival of the tumour cells. This cytotoxicity may in fact be at a sub-lethal level for normal human cells, where all the DNA repair processes are intact.

Although a number of approaches have undergone intensive and close scrutiny as potential approaches to treat and kill cancer (targeting signalling pathways and components of multi-drug resistance, etc.), much less work has focussed on blocking the ability of a cancer cell to recognise and repair the damaged DNA which results

primarily from the front line cancer treatments; chemotherapy and radiation. More recent studies on a number of DNA repair targets have produced proof-of-concept results showing that selective targeting of these DNA repair enzymes has the potential to augment the currently used chemotherapeutic agents and radiation as well as overcoming drug resistance (Salles *et al.,* 2006). A key protein responsible for the terminal step of all DNA repair pathways is DNA ligase. Whilst human DNA ligase has been a possible target of the chemotherapeutic doxorubicin (Ciarrocchi *et al.,* 1999), at the outset of our study no previous experimental data was available where human ligase was being deliberately targeted as a means of treating cancer.

DNA ligase I is required for ligation of Okazaki fragments during replication and ligases are the last step of almost all DNA repair pathways. DNA ligase deficient cell lines exhibit sensitivity to a wide range of DNA damaging agents (Tomkinson *et al.,* 2006; Vijayakumar *et al.,* 2009). Thus DNA ligase inhibitors would be predicted to have pleiotropic effects on cell proliferation and sensitivity to DNA damage. Among the three DNA ligases (I, III and IV) found in humans, expression of DNA ligase I is elevated in tumour cells as compared to normal cells whereas the expression of DNA ligase III and IV does not significantly change in proliferating and resting cells (Sun *et al.,* 2001). Cisplastin, an anticancer drug, reacts to link covalently with DNA molecules and blocks the replication process by forming inter and intrastrand DNA crosslinks (Selvakumaran *et al.,* 2003). This damage is repaired by the NER pathway (Sun *et al.,* 2002). Increased NER activity has been observed in cisplastin-resistant tumour cell lines, whereas cells defective in NER are more sensitive to cisplastin (Selvakumaran *et al.,* 2003; Sun *et al.,* 2002). These studies suggest that combining a therapeutic such as cisplatin with ligase inhibitors could possibly augment the lethal effect of the initial chemotherapeutic.

The catalytic core structure of DNA ligases is maintained between diverse species. At only 298 amino acids, PBCV-1 DNA ligase is the smallest characterised eukaryotic ATP-dependent DNA ligase (Ho *et al.,* 1997; Odell & Shuman, 1999). The PBCV-1 enzyme consists only of the ligase catalytic core without the flanking N and C-terminal domains found in the larger cellular eukaryotic DNA ligases (Odell *et al.,* 2000). The structure of the 34 kDa PBCV-1 DNA ligase enzyme-adenylate

intermediate shows the same two-domain arrangement (an N-terminal adenylation domain and a C-terminal OB-fold domain) as the larger 41 kDa T7 DNA ligase (Subramanya *et al.*, 1996; Odell *et al.*, 2000). The critically conserved catalytic residues of both the T7 and the PBCV-1 ATP-dependent DNA ligases are superimposable when one compares their three dimensional structures (Figure 1.8, panel B).

All ATP-dependent DNA ligases retain critical residues that can be shown to be involved in ligase reaction chemistry (Shuman, 1996). The linear spacing of these residues is very closely conserved and structural studies suggest that their linear arrangement reflects their conservation in the three dimensional arrangement of residues involved in DNA ligation (Subramanya *et al.*, 1996; Odell *et al.*, 1999; Odell *et al.*, 2000).

HuLigI is a 919 amino acid polypeptide (102 kDa) that can only be expressed in a recombinant form in very small amounts. Our initial strategy was to clone and express this enzyme and by modification of the second amino acid and use of appropriate bacterial strains to obtain sufficient material to use in an inhibitor assay. Ultimately this approach did not yield sufficient protein material however a truncated version of this protein, Δ232 (232-919 amino acid), used by Pascal *et al.* (2004) to study the detailed structure of the catalytic and DNA binding core, is smaller in size and is better expressed in *E. coli* cells. More importantly we know from the crystal structure of this truncated form bound to its substrate that it is capable of recognising and binding to DNA nicks (Pascal *et al.*, 2004). It was not only shown to be capable of stably interacting with DNA furthermore it was catalytically active in DNA sealing (Pascal *et al.*, 2004). It is our opinion that the truncated form of the human ligase represents a defined version of the human enzyme that can therefore be used to identify inhibitors of the human enzyme that block both key catalytic residues and moreover those residues critical for DNA recognition. This truncated version, containing the DNA binding domain, the adenylation domain and the OB fold domain, was therefore purified for use in inhibitor assays (Chapter 5) as a primary screen in place of full-length HuLigI.

In this study we also used PBCV-1 DNA ligase as a representative of the catalytic site of ATP-dependent DNA ligases and as a representative of a new class of potential targets found in the ATP-dependent ligases of eubacteria.

4.2 Materials and methods

4.2.1 Cloning and expression of full-length human DNA ligase I

Forward and reverse primers were designed for the mRNA sequence of HuLigI available at www.ncbi.nlm.nih.gov under accession number NM_000234 to amplify the coding region of the gene. Repeated attempts to amplify the full-length ORF of DNA ligase I using purified cDNA library obtained from HT-29 or MDA MB-435 cells (as described in Chapter 2; 2.1.6) yielded no product. New primers, as listed in Table 4.1, were designed to amplify the N-terminal and C-terminal portions of the gene in two different reactions whose products could later be ligated to yield the full-length ORF (Teraoka *et al.,* 1993). The reaction samples were prepared as described in section 2.1.9, and these underwent an amplification reaction using different set of primers as listed in table 4.1.

The full-length clone (1-3421 bp) of HuLigI in vector pCMV-SPORT6 was purchased from Gene Services Ltd, Cambridge (MGC: 117397; IMAGE: 6141535). Forward (5'-GATC*CTCGAG*ATGCAGCGAAGTATCATGTC-3') and reverse (5'-GATC*CTCGAG*TTAGTAGGTATCTTCAGGGTCAGAG-3') primers were designed with *Xho* I restriction sites; italicised and underlined in the primer sequence; flanking the ORF (421-3180 bp) of the full-length clone.

PCR to amplify full-length ORF was performed for 20 cycles (consisting of 1 minute at 94^{o}C, 1 minute at 55^{o}C and 3 minutes at 72^{o}C) after initial denaturation at 95^{o}C for 2 minutes. The products were resolved on a 0.7% agarose gel, purified using the Qiagen DNA purification kit (as described in Chapter 2, 2.1.4.1) and eluted from the DNA affinity resin in 30 µl of sterile distilled water. The purified PCR product was digested with *Xho* I restriction enzyme (New England Biolabs) for 3 hours at 37^{o}C and then ligated with *Xho* I digested pET-16b vector for 16 hours at 4^{o}C. The ligated

mixture was transformed into *E. coli* (DH5α) cells (Invitrogen) and plated onto LB-Amp (100 µg/ml ampicillin) plates. Plasmid DNA was extracted using a Qiagen plasmid mini-prep kit (Chapter 2, 2.1.2) from overnight cultures, inoculated with a single, transformed bacterial colony and digested with *Xho* I as described earlier to check for insert release. The integrity of the constructs was confirmed by sequencing purified plasmid at the Advanced Biotechnology Centre (ABC), London.

The pET-16b plasmid containing full-length HuLigI was transformed in *E. coli* BL21(DE3) and *E. coli* BL21(DE3)RP strains. The *E. coli* cells harbouring the plasmid containing the human DNA ligase gene were grown at 37^{o}C either in Luria-Bertani broth or 2YT media (8 g Bacto tryptone, 5 g Yeast extract, 5 g NaCl per 1 litre) containing 100 µg/ml ampicillin and/or 34 µg/ml chloramphenicol. The expression of HuLigI was induced by the addition of 0.2-1 mM isopropyl β-D-thiogalactopyranoside (IPTG) when the culture density reached 0.4-0.5 at OD_{595} and incubation was continued either overnight at 18^{o}C or for 3 hours at 37^{o}C. The cells were harvested by centrifugation for 10 minutes at 3,000 g and resuspended in buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol and 300 mM NaCl). The cells were lysed, centrifuged and the soluble fraction was loaded onto Ni-NTA resin as previously described (Chapter 2; 2.2.1; 2.2.2; 2.2.3). The expression of the protein was checked by electrophoresis using a 10% sodium dodecyl sulphate (SDS)-polyacrylamide gel and a western blot was performed with anti-his antibody as described in Chapter 2; 2.2.5.

4.2.2 Expression and purification of Δ232 HuLigI

10 ml LB media was inoculated with a single colony of *E. coli* BL21(DE3)RP harbouring plasmid encoded Δ232 HuLigI (obtained from Dr. T. Ellenberger, Washington University, USA). The antibiotic concentration was adjusted to 50 µg/µl kanamycin and 34 µg/µl chloramphenicol and the culture grown overnight at 37^{o}C with aeration. This overnight culture was diluted (1 in 100) into 500 ml LB, adjusted to 50 µg/µl kanamycin and 34 µg/µl chloramphenicol and grown at 37^{o}C with shaking until the absorbance measured at A_{595nm} reached 0.4-0.5. The protein expression was induced by the addition of IPTG to a final concentration of 0.4 mM and the culture

was incubated at 37^{o}C for further two hours. The cells were harvested by centrifugation and processed as described above for full-length HuLigI. The soluble extract was loaded onto Ni-NTA resin as previously described (Chapter 2; 2.2.1) and protein was eluted in buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol and 300 mM NaCl) containing 300 mM imidazole. The elution profile was monitored via SDS-PAGE.

The peak fraction eluted from Ni-NTA (judged by SDS-PAGE) was diluted to 25 mM NaCl with buffer B (50 mM Tris-HCl pH 7.5, 10% glycerol, 5 mM DTT) and applied to Q-Sepharose resin (GE Healthcare), pre-equilibrated with buffer B containing 25 mM NaCl. The column was washed with buffer B containing 15 mM NaCl and the protein was step eluted with buffer B containing 0.075, 0.1, 0.2, 0.3, 0.5 and 1.0 M NaCl. The elution profile was monitored via 12% SDS-PAGE. The peak fraction was dialysed against buffer B containing 100 mM NaCl. The purified protein was aliquoted in small volumes and stored at -80^{o}C until required.

4.2.3 Expression and purification of PBCV-1 DNA ligase

LB media containing 100 μg/ml ampicillin was inoculated with a single colony of *E. coli* BL21(DE3) cells transformed with pET-16b containing the PBCV-1 ligase gene (obtained from Professor S. Shuman, Sloan-Kettering, New York) and grown overnight in LB broth at 37^{o}C. Fresh, pre-warmed LB was inoculated from this static overnight culture (at a 1/100 dilution), adjusted to 100 μg/ml ampicillin and incubated at 37^{o}C with shaking. When the optical density of the cultures reached 0.4-0.5 (OD_{595}), cells were induced with IPTG (0.4 mM final) and incubated at 37^{o}C for 2 hours with constant shaking. The cells were processed as described above, the soluble fraction was applied to Ni-NTA resin as described previously (Chapter2, 2.2.3) and protein was eluted in buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol and 300 mM NaCl) containing 300 mM imidazole. The elution profile was monitored via SDS-PAGE.

The peak protein fractions from the Ni-NTA column were diluted to 50 mM NaCl with buffer B (50 mM Tris-HCl pH 7.5, 5.0 mM DTT and 10% glycerol) and applied

to an S-sepharose column pre-equilibrated with buffer B. The column was washed with buffer B containing 25 mM NaCl and the protein was step eluted with buffer B containing 0.1, 0.15, 0.2, 0.3, 0.5 and 1.0 M NaCl. The elution profile was monitored via 12% SDS-PAGE. The peak fraction was dialysed against buffer B containing 100 mM NaCl. The purified protein was aliquoted in small volumes and stored at -80^{o}C until required.

4.2.4 *Hin*d III DNA ligation assays with Δ232 HuLigI

The purified Δ232 HuLigI was used to ligate *Hin*d III digested λ DNA in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP and 25 μg/ml bovine serum albumin at 16^{o}C for 5 minutes. The ligated mix was resolved on a 0.8% agarose gel, stained with ethidium bromide (0.5 μg/ml) and viewed under UV illumination.

4.3 Results

Our desire to develop inhibitors to the HuLigI enzyme required that we clone and express the full-length human protein. Whilst we altered our strategy during the cloning of the full-length HuLigI - deciding to use the better expressed, truncated Δ232 HuLigI to identify inhibitors - potential inhibitors from our assay that will need to be screened against the full-length enzyme to substantiate the idea that can be developed for *in vivo* cell inhibition.

4.3.1 Amplification and cloning of HuLigI

Based on previous findings of Sun *et al.* (2001) that human DNA ligase I (HuLigI) is up-regulated in proliferating cancer cells, we chose HT-29 (a metastatic colorectal cancer cell line) and MDA MB-435 (a breast carcinoma cell line) cells to make human cDNA libraries from which to isolate a human DNA ligase I clone. Total RNA was isolated from these cells lines using SV total RNA isolation kit (Promega) followed by mRNA isolation using the Poly A Tract® mRNA Isolation system

(Promega; as described in Chapter 2; section 2.1.6). A reverse transcriptase (RT) reaction was performed with this mRNA to allow cDNA synthesis using a universal riboclone® cDNA synthesis kit (Promega). Random primers, gene specific primers and oligo-dT primers were used to prepare separate cDNA libraries from the purified mRNA obtained from the above cell lines (as described in Chapter 2; 2.1.6). Use of oligo-dT primers is recommended as they can initiate first strand synthesis by annealing to the 3' end of any polyadenylated RNA molecule whereas the gene specific primers (Table 4.1; set1) should only initiate the first strand synthesis by annealing to a gene specific region on appropriate RNA molecules.

Initial attempts made to amplify the full-length HuLigI gene by PCR using gene specific sets of primers (Table 4.1; set 1 & 2), in independent reactions containing any of the cDNA libraries isolated above, were largely unsuccessful. Various attempts including varying annealing temperature of the primers and changing DNA polymerase used in the amplification reaction failed to yield a PCR product of required length. A commercially available human pancreatic cDNA library (Invitrogen) was also used as a template in a PCR reaction using primer set 1 and 2 (Table 4.1) but failed to amplify the gene. Table 4.1 lists various primer sequences used in an attempt to amplify full-length HuLigI.

Table 4.1: Oligonucleotide primer sequences used for the amplification of the HuLigI

Primer Set used		Sequence of the primers used (5'-3')	Region corresponding to mRNA of HuLigI	Expected size of amplified PCR product
Set 1	Forward	GCCTGGCGGATCTGAGTTG	10-29	~ 3 kb
	Reverse	TGAAAGAAATGACGGGATGAATC	3040-3062	
Set 2	Forward	ATGCAGCGAAGTATCATGTC	121-140	~ 2.8 kb
	Reverse	GCGAGGGCTTAGTAGGTATC	2868-2888	
Set 3	Forward	ACCCAGCGAAGTATCATGTCATTT	35-55	0.75 kb (N- terminal)
	Reverse	GAAGTGAAGGAAGAGGAGCC	755-774	
Set 4	Forward	AGAGCTCCCAAGACGCTCAG	698-717	2.1 kb (C-terminal)
	Reverse	GATACCTACTAAGCCCTCGC	2780-2799	
Set 5	Forward	CAGAGGGCGAGTTTGTCTTC	2186-2205	0.2 kb
	Reverse	AGCCAGTTGTGCGATCTCTT	2329-2348	

Teraoka *et al.* (1993) cloned full-length HuLigI into *E. coli* by separately amplifying the N-terminal and C-terminal ends of the gene, followed by their ligation into a single expression vector. Based on their experimental design, we designed a new set of primers (Table 4.1; set 3) with which DNA ligase I could be amplified separately as an N-terminal fragment (0.75 kb) and a C-terminal fragment (2.1 kb).

Only the N-terminal region of the gene (0.75 kb; Figure 4.1) could be successfully amplified while the C-terminal region failed to amplify even after trying all available

cDNA libraries as the template and varying the primer annealing temperatures using gradient PCR over a 20^{o}C range.

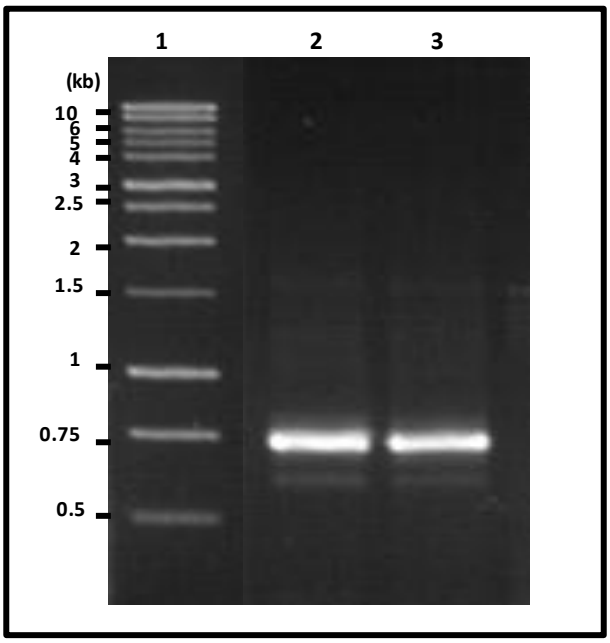


Figure 4.1: Amplification of N-terminal of Human DNA ligase I

A PCR reaction performed for 30 cycles, consisting of 95^{o}C denaturation for 1 minute, 55^{o}C annealing and 72^{o}C extension for 1 minute, using primers (Table 4.1; set 3) designed to amplify from 35-774 bp of the N-terminal region of the HuLigI gene. The resultant PCR product was analysed by 0.7% agarose gel electrophoresis stained with ethidium bromide and viewed under UV illumination. Lane 1: 1kb ladder molecular weight marker; Lanes 2 & 3: 0.75 kb N-terminal PCR product.

Different components of the PCR reaction that can effect the amplification reaction were addressed separately. The role of Mg^{2+} in PCR is to promote DNA/DNA interactions and form complexes with dNTPs (Eckert & Kunkel, 1991). The total Mg^{2+} concentration (usually 1.5 mM) used in the PCR reaction must exceed the total dNTP concentration so that after forming complexes with dNTPs, free magnesium ions are available to serve as an enzyme cofactor for the DNA polymerase. At the same time, too high a concentration of Mg^{2+} results in strong base pairing such that very high temperature ($\geq$94^{o}C) is required to denature the double helical structure of DNA (Eckert & Kunkel, 1991). The $MgCl_2$ concentration was varied from 0.5-8 mM in PCR reactions and different DNA polymerases ranging from the widely used *Taq* DNA polymerase to proof-reading *Pfu* DNA polymerase and high fidelity Phusion DNA polymerase failed to amplify the full-length or C-terminal region of HuLigI. A PCR reaction was also tried in the presence of dimethylsulfoxide (DMSO). The use of

such solvents helps increase the yield, efficacy, and specificity of PCR amplifications by effectively lowering the melting and strand separation temperatures (Eckert & Kunkel, 1991). Several independent PCR reactions tried with increased template concentration failed to amplify the full-length ORF of HuLigI.

The amplification of a PCR product clearly also depends on the presence of the mRNA of the gene in the corresponding cDNA library. We designed new primers (Table 4.1; set 5) that would amplify a region from 2186-2348 bp of the HuLigI mRNA to check expression of the C-terminal region in the various cDNA libraries used as templates. Figure 4.2 shows that a PCR product corresponding to ~200 bp was obtained with all the different template reactions, although, the intensity of the product formed varied with the source of cDNA. The size of the product (200 bp) was bigger compared to expected size (162 bp) and there is no clear rationale for this discrepency.

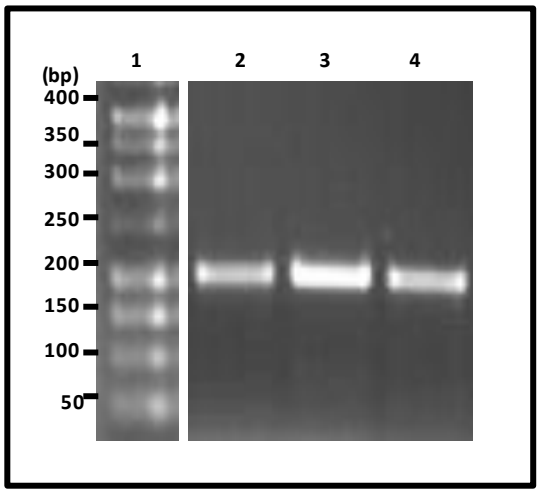


Figure 4.2: Amplification of a C-terminal region of HuLigI using various cDNA libraries as a template.

Expression of mRNA corresponding to the C-terminal region of HuLigI in cDNA libraries obtained from different sources was checked by performing a PCR reaction for 30 cycles consisting of 1 minute denaturation at 95^{o}C, 1 minute annealing at 55^{o}C and 30 seconds extension at 72^{o}C. Reactions were performed using a set of primers specific for the 2186-2348 bp region of the full-length HuLigI. Lane 1: 50 bp step ladder molecular weight marker; Lane 2: cDNA obtained from HT-29 cells used as template; Lane 3: commercially obtained pancreatic cDNA library used as template; Lane 4: cDNA obtained from MDA MB-435 cells used as a template in amplification reaction.

In order to obtain a full-length polypeptide of the HuLigI a source of the mature, full-length ligase I gene was required. A shuttle vector containing a cDNA clone of HuLigI was obtained from Gene Services Ltd., Cambridge. The vector contained the

genomic region for HuLigI and this allowed the use of better quality DNA in greater abundance in the amplification reaction as compared to the cDNA obtained from different cell lines.

The primers complementary to the 5' and 3' ends of the coding region of HuLigI incorporating *Xho* I restriction sites at both ends of the gene allowing cloning into the pET-16b expression vector were employed (sequence given above in section 4.2.1). PCR was performed for 20 cycles (consisting of 1 minute at 94^{o}C, 1 minute at 55^{o}C and 3 minutes at 72^{o}C) after initial denaturation at 95^{o}C for 2 minutes. This time the PCR was successful. An amplified PCR product of ~3 kb was obtained (Figure 4.3; panel A). The PCR product was gel purified, digested with *Xho* I restriction enzyme and ligated into *Xho* I digested CIP treated pET-16b expression vector. The ligation reaction was transformed into electro competent *E. coli* (DH5α) cells.

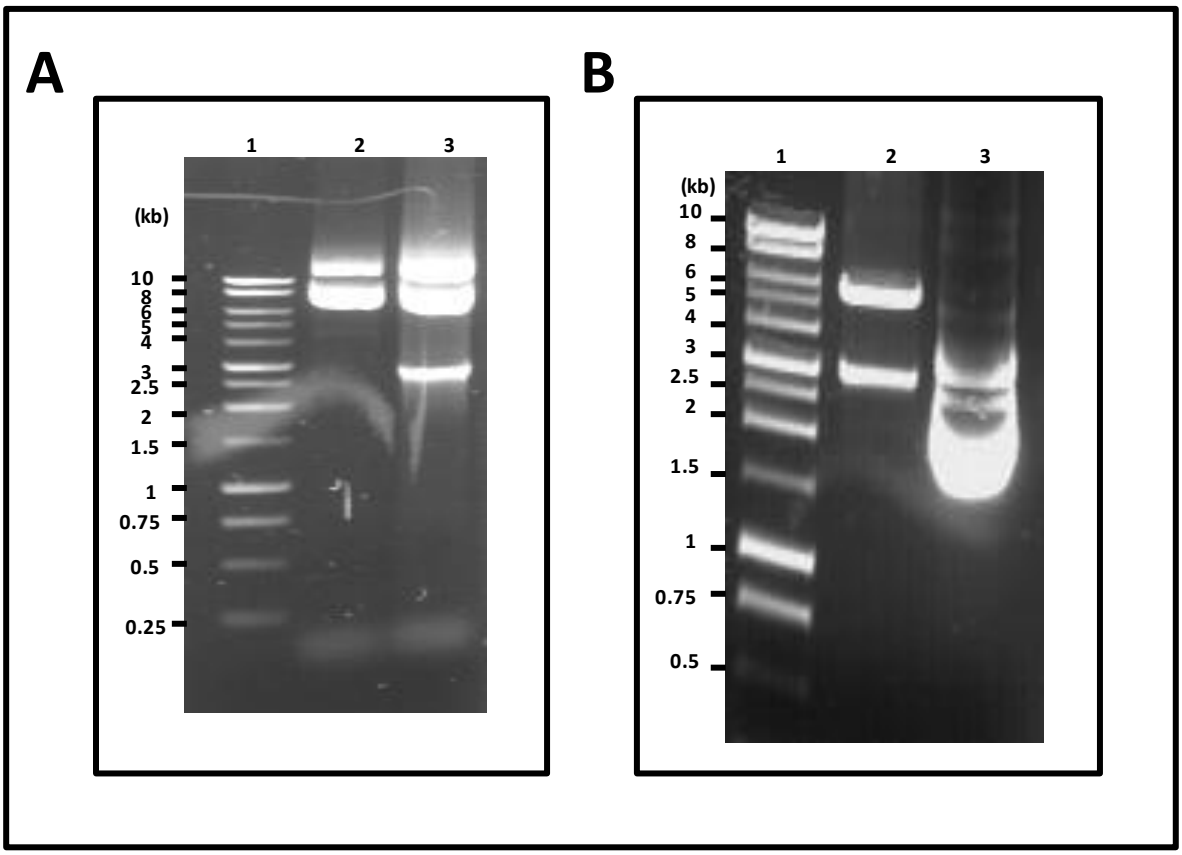


Figure 4.3: PCR-cloning of the full-length ORF encoding HuLigI

The pCMV-SPORT6 vector harbouring the cDNA of HuLigI was used as a template in a PCR reaction to amplify the coding region of the human ligase. The reaction was performed for 20 cycles consisting of 1 minute at 94^{o}C, 1 minute at 55^{o}C and 3 minutes at 72^{o}C and products were analysed on a 0.7% agarose gel, stained with ethidium bromide and viewed under UV illumination as shown in panel A. Lane 1: 1kb DNA ladder molecular weight markers (Promega); Lane 2**:** Negative control (only template); Lane 3: product of amplification reaction.

Panel B: The gel purified PCR product was digested with *Xho* I and ligated into *Xho* I digested and CIP treated pET-16b expression vector and the success of the insertion was confirmed by restriction with *Xho* I. Lane 1: 1 kb DNA ladder molecular weight markers (Promega); Lane 2: Insert release (corresponding to 2759 bp of HuLigI) & linearised pET-16b vector; Lane 3: Uncut plasmid before digestion with *Xho* I.

Some of the transformed colonies were randomly chosen and screened for the presence of the appropriate clone. Figure 4.3 (panel B) shows the insert release after the restriction digest of one of these plasmids. The clones that were positive for insert were sequenced to confirm that no changes had been introduced during the PCR procedure. An accurate, fully sequenced clone was then used to study expression of full-length human DNA ligase. The initial expression of HuLigI carried out in *E. coli* BL21(DE3) cells, induced for 3 hours with 0.4 mM IPTG at 37^{o} C, failed to give a

visible band at 102 kDa (Figure 4.4). This could be due to poor yield of the protein associated with its high molecular weight. The expression conditions were varied by changing IPTG concentration from 0.4 mM to 0.8 and 1 mM, and independently increasing the induction time period from 3 hours at 37^{o} C to 18 hours at 18^{o}C with the various IPTG concentrations. The change in the induction conditions failed to improve the expression of the protein.

Several attempts were made by varying all the possible parameters to improve the expression of the full-length protein including change of growth media from LB to richer 2YT media (twice the amount of yeast and tryptone as of LB media). Kane (1995) suggests that if the gene of interest contains codons not commonly used in *E. coli*, low expression may result due to depletion of tRNAs for the rare codon. This is more concerning when the sequence encodes a protein >60 kDa or when rare codons are found at high frequency or multiple rare codons are found over a short distance of the coding sequence. We looked at rare codon usage in HuLigI and found that HuLigI gene encodes for 12 rare arginine (AGG, AGA and CGA) and 14 rare proline (CCC) codons. Moreover, rare codons for arginine (AGG, AGA) were found in tandem at different positions (185, 575, 745, 878, 1019, 1765 and 2051 base pairs). Such sequences can create a recognition sequence for ribosome binding (such as –AGGAGG) that closely approximates a Shine-Dalgarno sequence (UAAGGAGG). The Shine-Dalgarno sequence helps recruit the ribosome to the mRNA to initiate protein synthesis by aligning it with the start codon. This AGGAGG sequence may bind ribosomes non-productively and block translation from the functional ribosome binding site at the initiator codon (Kane, 1995). One such rare arginine codon is present next to the initiation codon (ATG) in HuLigI. Chen & Inouye (1990) have shown that the number and the position of the arginine codons in an mRNA play an important role in modulating gene expression at the level of translation. The presence of rare arginine codons in a close proximity to the initiation codon exhibit severe negative effects on the expression of that gene (Chen & Inouye, 1990).

Based on these observations we decided to change the host strain to *E. coli* BL21(DE3)RP, which contains a plasmid that encodes copies of the rare arginine and proline tRNA genes and is able to rescue expression of genes restricted by either

AGG/AGA codons or CCC codons. Various attempts were made to improve the expression of HuLigI in *E. coli* BL21(DE3)RP cells, including varying inducer concentration from 0.2-0.4 mM and opting for longer hours of induction (24 or 48 hours) at 16^{o}C or 20^{o}C. All of these experiments were largely unsuccessful with no production of significant amounts of appropriately sized, recombinant protein. A typical profile for expression of the full-length HuligI is shown in Figure 4.4. Whilst a number of proteins appear to be bound to the column (being released by the 300 mM imidazole wash) none compared to the predicted weight of HuligI (101.7 kDa). These may reflect proteolytic breakdown products of the full-length ligase or *E. coli* proteins that have been able to bind in the absence of an over-expressed His-tagged protein (Figure 4.4).

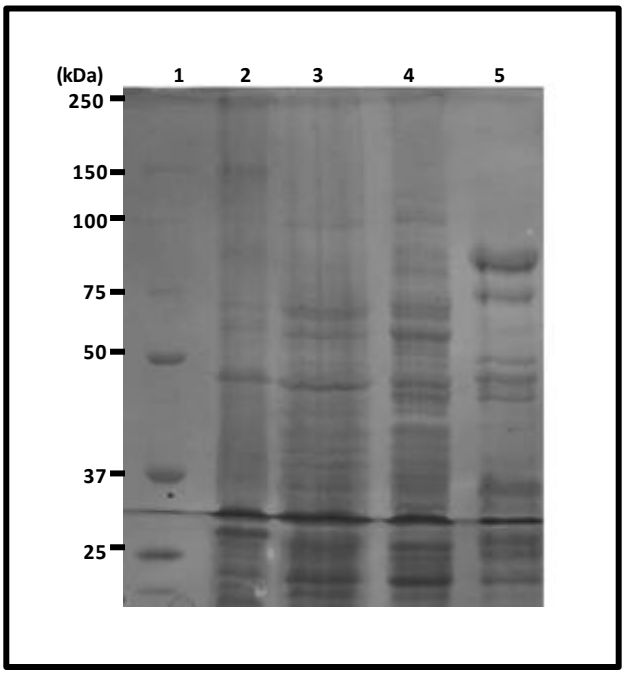


Figure 4.4: Analysis of recombinant HuLigI using Ni-NTA chromatography

The pET-16b plasmid containing full-length human DNA ligase was transformed into *E. coli* BL21(DE3)RP. Cultures were grown and induced with 0.4 mM IPTG for three hours at 37^{o}C. The lysed, soluble extract was applied to Ni-NTA resin and eluted with buffer A (50 mM Tris-HCl pH 7.5, 300 mM NaCl and 10% glycerol) containing 300 mM imidazole. The polypeptide composition was analysed by SDS-PAGE (10% resolving gel). Lane 1 contained Bio-Rad precision plus prestained molecular weight marker proteins (sizes shown in kilodaltons); Lane 2: cell pellet; Lane 3: Flow through collected from column after loading the soluble fraction; Lane 4: wash with buffer A containing 25 mM imidazole; Lane 5: elution with buffer A containing 300 mM imidazole.

The sensitivity of SDS PAGE (50 ng/band) is much less than the sensitivity of Western blotting (1 ng/band). Therefore, it was decided to probe uninduced and induced soluble and insoluble fractions with anti-His antibody to check whether we were getting even minimal expression (Figure 4.5). We further hoped that if any of the polypeptide species that bound to the Ni-NTA resin derived from the N-terminus of the expressed full-length human enzyme that these too would be detected by the western blot. No detectable signal was observed corresponding to a protein species of molecular weight 102 kDa (without taking account of the additional 2kDa as a result of the His-tag), the expected size of full-length HuLigI. The presence of lower molecular weight species (very faintly staining bands at ~80, 62 and 45 kDa) could be due to cross-reaction artefacts from host cell proteins. A number of native *E. coli* proteins have been identified, which are over-expressed in response to cellular stress, for example the 29 kDa *E. coli* protein SlyD is histidine-rich and is a common contaminant following over-expression of recombinant proteins in *E. coli* (Bolanas-Garcia & Davies, 2006). The presence of the 80, 62 and 45 kDa species in both the induced and uninduced samples suggests that these species are not derived from the expression of human HuligI. The presence of a unique ~42 kDa species in the induced insoluble fraction (Figure 4.5) may represent a degraded product from the expression of the full-length HuLigI enzyme.

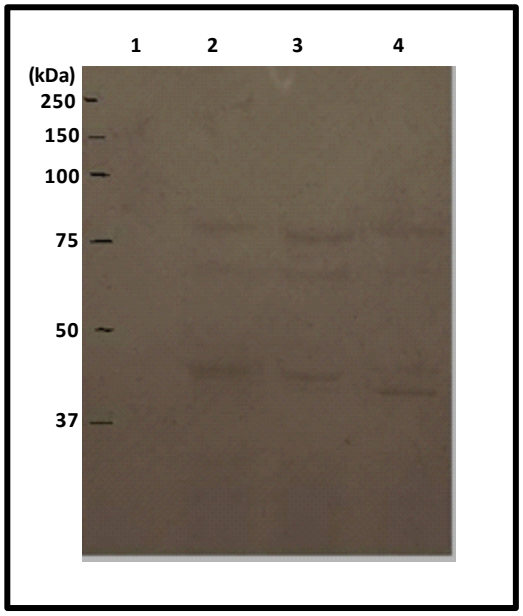


Figure 4.5: Western blot analysis with anti-his antibody of proteins expressed in *E. coli* cells harbouring a plasmid expressing full-length HuLigI

Soluble and insoluble extracts from cells grown in the presence or absence of IPTG were prepared in reducing sample buffer and separated by SDS-PAGE. Proteins were then transferred to a nylon membrane and incubated with anti-his antibody. Lane 1 shows the position of broad range molecular weight markers (kilo Daltons determined by staining the membrane with Ponceau-S), Lane 2 is a sample of soluble protein from uninduced *E. coli* cultures harbouring the HuligI plasmid and lanes 3 & 4 are the soluble and insoluble sample obtained after 3 hour induction with 0.5 mM IPTG.

After varying all the possible parameters, full-length HuligI protein was not expressed in detectable levels. This might be due to its size, full-length HuLigI is 102 kDa as compared to the well-expressed 34 kDa PBCV-1 DNA ligase. A low yield of pure protein production has always been the main obstacle in studying this enzyme recombinantly. Similar results with expression of full-length HuLigI have been reported by Kodama *et al.* (1991) and they concluded that this was probably due to degradation of the recombinant enzyme by *E. coli* encoded proteases (Kodama *et al.*, 1991). Harcum & Bentley (1992) reported enhanced protease activity in *E. coli* cells in a response to over-expression of recombinant proteins, for example, induction of recombinant chloramphenicol acetyltransferase (CAT) increased intracellular protease activity and triggered stress responses. Kodama *et al.* (1991) observed that a truncated HuLigI (250-919 amino acids) is the optimal size for expression of the HuLigI protein in *E. coli*. This truncated version of HuLigI polypeptide was found to be catalytically active *in vitro* and *in vivo*, shown by complementation assays (Kodama *et al.*, 1991).

4.3.2 Purification of Δ232 HuLigI

Since Kodama *et al.* (1999) had defined an optimum size for the function of human DNA ligase I in *E. coli***;** truncated HuLigI has been studied in place of the full-length protein. When this work was in progress, Pascal *et al.* (2004) solved the structure of a similarly truncated HuLigI (232-919 amino acid) bound to adenylated DNA. This structure, for the first time, revealed the role of the HuligI N-terminal DNA binding domain which stabilises and encircles the nicked DNA substrate in coordination with the adenylation and oligomer binding domain**s** (Pascal *et al*., 2004).

After numerous attempts to express full-length HuLigI, the plasmid pET-28a harbouring a truncated HuLigI [232-919 amino acid; used for structural studies by Pascal *et al.* (2004)] was obtained from Professor T. Ellenberger (Washington University, USA). The structure of Δ232 HuLigI has been solved in complex with DNA and furthermore the authors had shown that the truncated enzyme was active as a ligase in *in vitro* sealing assays (Pascal *et al*., 2004). We knew therefore that the enzyme could function as a model both for enzyme binding inhibition and for catalytic inhibition and furthermore the conditions for the expression and purification of this construct were well documented. The pET-28a plasmid was transformed into *E. coli* BL21(DE3) cells to over-express the N-terminal His-tagged protein. The soluble fraction containing the protein of interest was applied to Ni-NTA resin and the protein was eluted with buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol and 300 mM NaCl) containing 250 mM imidazole (Figure 4.6, panel A). The electrophoresis made estimation of the molecular weight difficult, of the predominant protein species retained on the nickel resin. We decided to proceed with the purification and make a better estimate of the size when the protein was more pure. The eluted fractions were pooled, diluted to 25 mM NaCl with buffer B (50 mM Tris-HCl pH 7.5, 10% glycerol, 5 mM DTT and 25 mM NaCl) and applied to Q-sepharose (anion exchange resin with a quaternary amine exchange group). A prominent polypeptide of ~80 kDa was recovered during step elution from Q-sepharose in a fraction containing 0.2 M NaCl (Figure 4.6, panel B), as judged by the SDS gel electrophoresis. The molecular weight of the Δ232 HuLigI recombinant truncated protein was estimated to be 78.2 kDa on the basis of the amino acid sequence with the inclusion of the histidine tag.

The 0.2 M NaCl fraction containing purified Δ232 HuLigI was then dialysed extensively against buffer B containing 100 mM NaCl and stored at -80^{o} C for use in further assays.

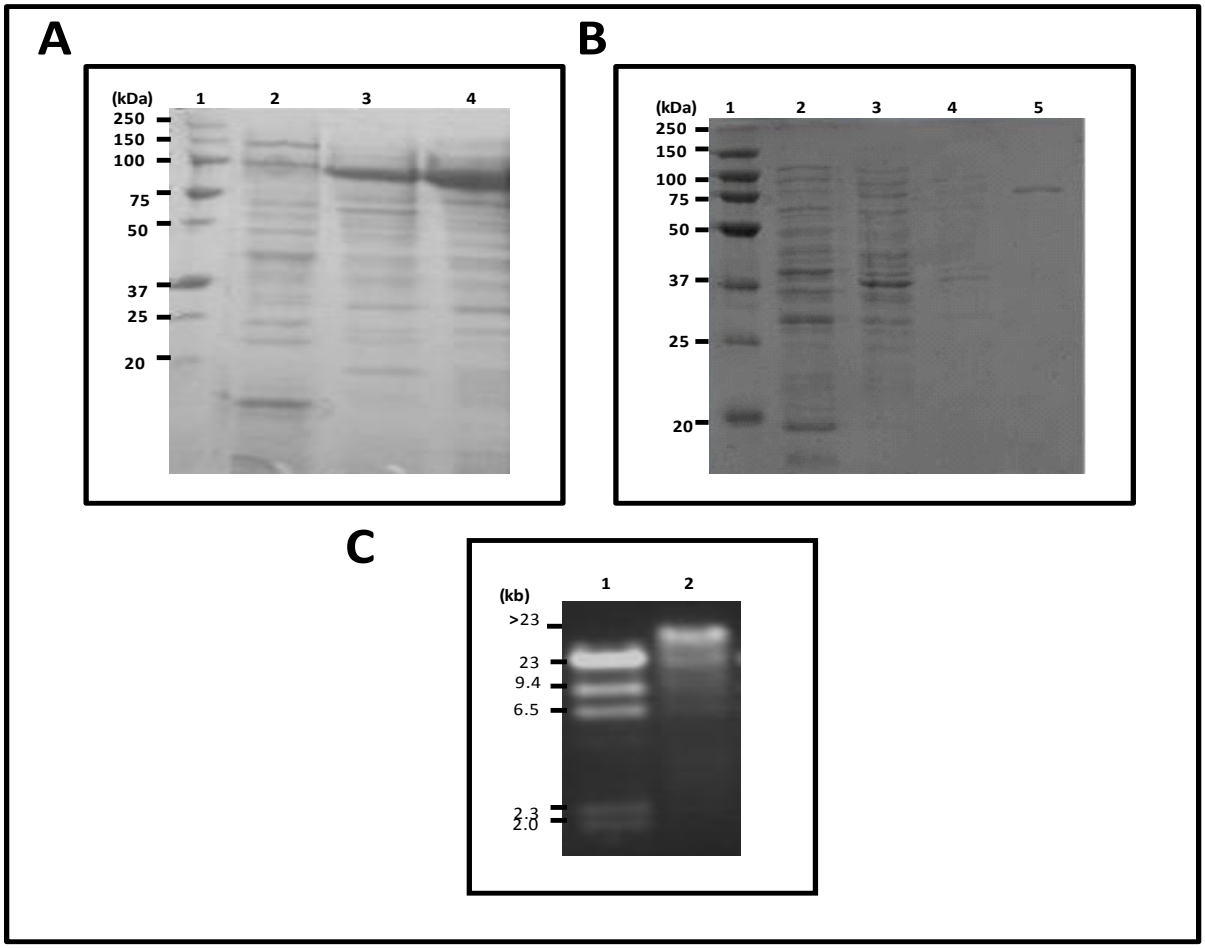


Figure 4.6: Elution profile and ligation activity of Δ232 HuLigI

The pET-28a plasmid bearing Δ232 HuLigI was expressed in *E. coli* BL21(DE3) cells by induction with 0.4 mM IPTG for two hours at 37^{o}C. The lysed, soluble extract was chromatographed on Ni-NTA resin, aliquots were analysed by 12% SDS-PAGE stained with Coomassie Brilliant blue. Panel A - Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker (size of the molecular weight of marker proteins are shown to the left of the gel); Lane 2: flow through; Lane 3: wash fraction (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 25 mM imidazole and 10% glycerol); Lane 4: elution fraction (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM imidazole 10% glycerol).

The protein peaks eluted from Ni-NTA agarose were diluted to 25 mM NaCl and applied to Q-sepharose and eluted along an NaCl gradient (NaCl in buffer B: 50 mM Tris-HCl pH 7.5, 5.0 mM DTT, 10% glycerol). The elution profile is analysed in panel B. Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker; Lane 2: wash (50 mM NaCl in buffer B), Lanes 3 & 4: 0.1 M NaCl in buffer B; Lane 5: 0.2 M NaCl in buffer B.

Purified Δ232 HuLigI was used to ligate *Hin*d III digested λ DNA in ligation buffer (containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP and 25 mg/ml bovine serum albumin). The ligation reaction products were analysed by 0.8% agarose gel (panel C), stained with ethidium bromide and viewed under UV illuminator. Lane 1 contains *Hin*d III digested λ DNA; Lane 2 contains ligation reaction product when carried out in the presence of Δ232 HuLigI.

To verify that the purified human protein had retained activity during the purification we incubated the enzyme with λ DNA digested with *Hin*d III (as described in section 4.2.4; Figure 4.6; panel C). From the alteration in the migration of these cohesive fragments we determined that the truncated human enzyme retained DNA ligation activity.

4.3.3 Purification of PBCV-1 DNA ligase

All ATP-dependent DNA ligases, including the PBCV-1 DNA ligase, that of TV and the human DNA ligases, share six motifs that are in large part responsible for the enzyme catalytic core (Odell *et al.,* 2000). Using the known 3-dimensional structures of these enzymes it is possible to superimpose the key catalytic residues that contact the ATP co-factor (Figure 1.8; panel B). PBCV-1 DNA ligase, the smallest characterised eukaryotic DNA ligase, is regarded as the minimal catalytic unit present in all ligases, such that anything that applies to this enzyme will apply to the larger ligases (Odell *et al.,* 2003). This is the only ligase with its structure analysed for both DNA bound and unbound adenylated states, providing instructive insights into the mechanism of DNA ligation and conformational changes that occur upon DNA binding (Odell *et al.,* 2000; Odell *et al.,* 2003; Nair *et al.,* 2007). The fact that the recombinant PBCV-1 ligase is expressed and purified easily in high yield from *E. coli* makes it a useful tool to screen for inhibitors of larger DNA ligases that are more difficult to express in large quantities and whose behaviour has not been so closely studied. Hence it was our initial intention to use this enzyme as a first screen to identify ligase catalytic inhibitors. Once a potential candidate compound had been identified it could then be tested for its activity against the purified target ligase enzymes such as the TV or HuLigI.

The 34 kDa PBCV-1 DNA ligase encoded on a pET-16b vector was obtained from Professor Stewart Shuman (Sloan-Kettering, New York), then transformed and expressed in *E. coli* BL21(DE3) under the inducible control of a T7 promoter as previously reported (Ho *et al.,* 1997). The PBCV-1 enzyme was purified from a soluble bacterial extract via affinity chromatography from Ni-NTA resin. The protein

was eluted in buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol and 300 mM NaCl) containing 300 mM imidazole. A significantly over-expressed polypeptide species with an apparent molecular weight of 40 kDa elutes from the column as analysed by SDS-PAGE (Figure 4.7, panel A). Although PBCV-1 DNA ligase is a 34 kDa polypeptide, it migrates unusually slowly on SDS-PAGE as has been previously noted (Odell *et al.,* 1999). The peak protein fractions eluted from Ni-NTA resin were diluted to 50 mM NaCl with buffer B (50 mM Tris-HCl pH 7.5, 5.0 mM DTT and 10% glycerol) and applied to an S-sepharose column pre-equilibrated with buffer B. The protein was step-eluted along an NaCl gradient (0.1–1.0 M NaCl in 50 mM Tris-HCl pH 7.5, 5 mM DTT and 10% glycerol) and the protein elution profile analysed by SDS-PAGE (Figure 4.7, panel B). The analysis of the peak fractions from S-sepharose chromatography showed that apparently homogenous PBCV-1 DNA ligase elutes at 0.2 M NaCl (Figure 4.7, panel B). Approximately 5 mg of wild type PBCV-1 DNA ligase was obtained per litre of culture as estimated by a Bradford protein assay (Sigma) using BSA as a standard (standard curve attached in the appendix 7.4).

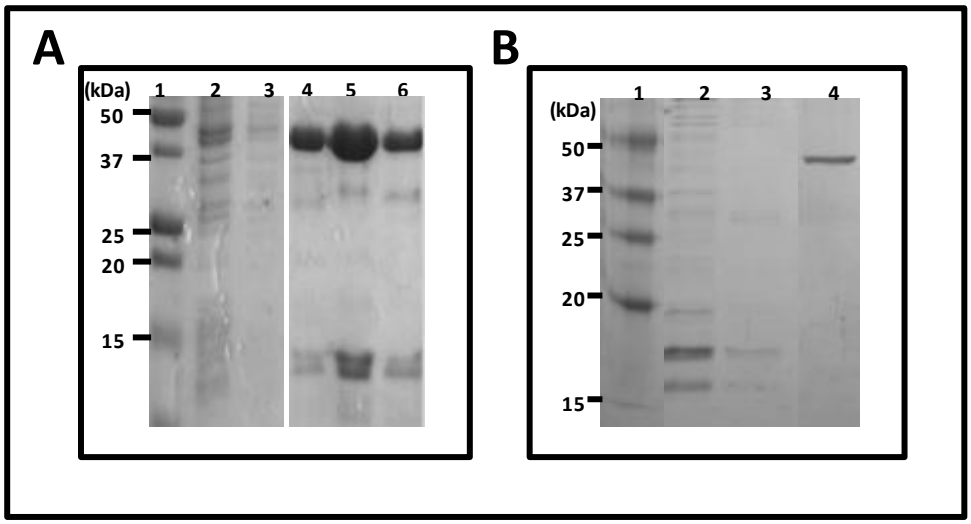


Figure 4.7: Chromatography profile of PBCV-1 DNA ligase over Ni-NTA and S-sepharose resins

The pET-16b plasmid bearing PBCV-1 DNA ligase was transformed into *E. coli* BL21(DE3) cells, grown overnight in LB media. The overnight culture was diluted 1 in 100 into fresh LB media and grown at 37^{o}C until A_{595} reached 0.4-0.5. The cells were adjusted to a final concentration of 0.4 mM IPTG and incubated at 37^{o}C for 2 hours. The soluble protein was applied to Ni-NTA resin, aliquots were analysed by 12% SDS-PAGE stained with Coomassie Brilliant Blue. Panel A, Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker (molecular weight of marker proteins are shown to the left of the gel); Lane 2: flow through; Lane 3: wash fraction (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 30 mM imidazole); Lanes 4-6: elution fractions (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM imidazole).
The protein peaks eluted from Ni-NTA agarose were diluted to 50 mM NaCl and applied to S-sepharose and eluted along an NaCl gradient (50 mM - 0.2 M NaCl) in buffer B (50 mM Tris-HCl pH 7.5, 5.0 mM DTT, 10% glycerol). The elution profile is analysed in panel B. Lane 1: Bio-Rad Precision Plus All Blue pre-stained molecular weight marker; Lane 2: flow through; Lane 3: wash (50 mM NaCl in buffer B); Lane 4: 0.2 M NaCl in buffer B.

4.4 Discussion

An increased repair of drug-induced DNA damage is an important mechanism of chemotherapy resistance in many human tumours (Madhusudan & Hickson, 2005). Several reports suggest that inhibiting the activity of major components of DNA damage repair pathways such as Poly(ADP-ribose)polymerase 1 (PARP 1) and replication protein A (RPA) could potentially circumvent resistance to certain chemotherapeutic agents (Andrews & Turchi, 2004; Curtin *et al.*, 2004). PARP is a zinc-finger DNA binding enzyme that is activated by binding to DNA breaks. DNA damage is converted into intracellular signals via poly(ADP)-ribosylation of nuclear proteins which activates DNA repair by the base excision repair pathway (BER;

Curtin *et al.*, 2004). PARP inhibitors (such as NU1025) have been shown to enhance the anti-tumour activity of temozolomide (a DNA methylating agent) and are now under clinical evaluation (Madhusudan & Hickson, 2005). The type of DNA damages caused by the various drugs used in chemotherapy and the corresponding repair mechanism triggered as a part of cell response to fix these damages are summarised in figure 4.8. NHEJ, BER and NER are some of the principal repair pathways, which, ultimately converge on a common terminal step in restoring the continuity of the repaired DNA strand by DNA ligases.

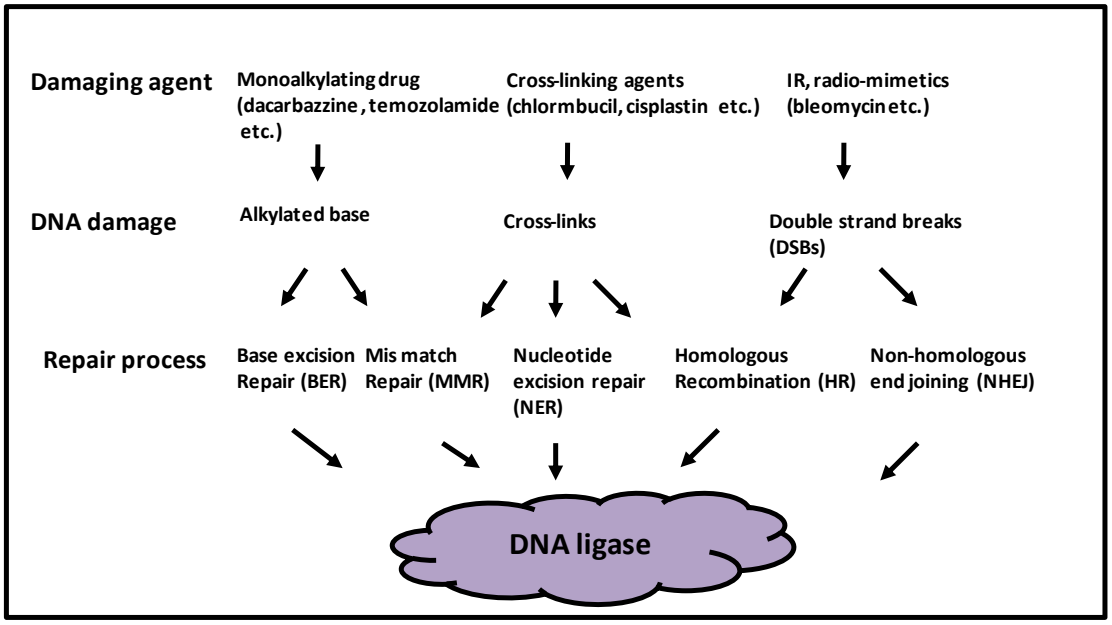


Figure 4.8: DNA damage induced by anti-tumour agents and the various pathways involved in repairing such lesions

Interferring with human DNA ligase's ability to bind DNA, ultimately interfering with DNA repair and replication, could potentiate DNA damage induced by other agents and also prevent the uncontrolled cell division that characterises cancer pathology.

Several lines of evidence suggest that proliferating cells contain a substantially higher level of DNA ligase I than other non-dividing or resting cells. Such a correlation does not exist for DNA ligases III and IV (Signoret & David 1986, Montecucco *et al.*,

1992; Sun *et al.,* 2001). Sun *et al*. (2002) suggested that an increase in expression of DNA ligase I in a human pancreatic cell line after exposure to cisplastin, an anticancer agent, might be required for aiding the cells to recover from DNA damage as a result of the cisplatin treatment by facilitating the repair process.

Other work by the same group, Sun *et al.* (2001) has shown that DNA ligase I in human tumours is present at higher levels than in normal tissues and normal peripheral lymphocytes. In their studies, using the HT-29 cell line (a metastatic colorectal cancer cell line) to investigate the kinetics of DNA ligase I expression, they showed that expression appeared constitutive and increased with the size of the tumour. Since, the HT-29 cell line is known to express high levels of DNA ligase I, it was used to construct a cDNA library from which we attempted unsuccessfully to isolate a HuLigI clone.

The amplification of DNA by PCR is dependent on a number of variables including the source of DNA template, primers, annealing temperature and concentration of $MgCl_2$ used in the reaction. All these variables were extensively analysed but failed to allow the production of a full-length HuLigI clone. HT-29 cell lines are known to have high level expression of HuLigI but cDNA obtained from these cell lines failed to amplify either the C-terminus or full-length HuLigI. Teraoka *et al.* (1993) constructed a recombinant plasmid, pTD-T7N, for expression of full-length HuLigI by separately amplifying the N-terminal region (0.7 kb) and the C-terminal region (2.1 kb) of the gene. We attempted to clone the same HuLigI enzyme fragments from cDNA obtained from the HT-29 cell line and two other cDNA libraries using the same primers (set 3 & 4, table 4.2) as those published by Teraoka *et al.* (1993), however no PCR reaction was successful for the C-terminal PCR product (2.1 kb). The N-terminal region of the gene (750 bp) was, however, successfully amplified. The cell lines used in this study were available in-house but we considered that the expression of the HuLigI gene in this passage of the cell line might have been different to that of the previous study. We therefore probed for the presence of mRNA showing expression of the HuLigI C-terminus in the cDNA library by performing a PCR amplification of a region (Figure 4.3; primer set 5) corresponding to 200 base pairs this region of the gene. Two different cDNA libraries; pancreatic cDNA library

(Invitrogen) and cDNA obtained from MDA MB-435 (a breast carcinoma cell line), were also used as a template in the control reaction. The PCR product was obtained with all the three templates, however, none of them yielded the 2.1 kb sized C-terminal product. Despite trying the PCR under several conditions, amplification of full-length or the remaining C-terminal region of the gene was unsuccessful. The primers used for the PCR reaction were same as published by Teraoka *et al.* (1993), and furthermore we used the same cell line which suggested to us that an alternate splicing of the ligase was not responsible for the PCR failure. Possible reasons for the failure to amplify the gene could be our cDNA template, though the success of the C-terminal fragment and the N-terminal amplification argue against this.

Because of the PCR difficulties a full-length cDNA clone of the HuLigI, mRNA sequence from nucleotide 1 through to the poly (A) tail in a shuttle vector pCMV-SPORT6 was purchased from Gene Services Ltd, Cambridge. The coding region of the gene (421-3180) was amplified using this plasmid. The amplified product was ligated into the pET-16b expression vector and the sequence of the clone obtained was confirmed by sequencing.

The initial expression studies of full-length HuLigI were performed using the *E. coli* BL21(DE3) strain. The recombinant protein failed to express under these conditions. Expression of heterologous proteins depends on various parameters such as induction conditions, host strain, growth temperature and media etc. All these above mentioned factors, which could affect the expression of full-length HuLigI, were addressed separately to optimise conditions for expression of the protein. Efficient production of heterologous proteins in *E. coli* can also be limited by the use of tRNAs by the human protein that are either absent or are less common in *E. coli* (Kane, 1995). The HuLigI nucleotide sequence encodes for 12 rare arginine (AGG, AGA and CGA) codons and 14 rare proline (CCC) codons. One of the rare arginine (AGG) codon is present in close proximity to start codon, methionine (ATG). This codon-bias could be the main obstacle in the expression of difficult target proteins as previously described by Cheng & Inouye (1990). In order to circumvent this problem the *E. coli* strain, BL21(DE3)RP, was used for expression studies which expresses the rare tRNAs from additional plasmid copies of the arginine and the proline tRNA genes. Unfortunately,

the use of this different *E. coli* host strain failed to improve the expression of full-length HuLigI. It was not possible to demonstrate significant expression of full-length HuLigI in these *E. coli* cells, after varying different growth and induction parameters.

Sequence comparisons of eukaryotic DNA ligases have suggested that a catalytic domain, common to all ATP-dependent ligases, is embellished by additional specific protein segments situated at their amino or carboxyl termini (Odell & Shuman, 1999). These protein segments are not essential for catalysis such as in HuLigI, which retained activity after removal of 249 amino acids from the N-terminus and 16 amino acids from the C-terminus (Kodama *et al.,* 1991). Further deletions of HuLigI resulted in loss of function, so it is understood that the 654 amino acid segment of ligase I represents the catalytic core (Kodama *et al.,* 1991). During the progress of this work, Pascal *et al.* (2004) published the crystal structure of truncated version of HuLigI (232-919 amino acids) bound to adenylated DNA. This confirmed that the small truncated version of HuLigI is stable and functional enzyme with required catalytic domains.

We decided to use a truncated version of HuLigI in our screening assays. The Δ232 HuLigI (232-919 aa) has been well studied and shown to not only catalyse DNA nick-sealing but also to bind stably to DNA (Pascal *et al.,* 2004). The Δ232 HuLigI has the ligase DNA binding domain, required to encircle the substrate in cooperation with the catalytic domain consisting of the adenylation and OB-fold domains. This construct was used to express a truncated version of HuLigI, which was purified to near homogeneity after two chromatography steps. DNA ligation assays confirmed the earlier reports that this fragment retained DNA sealing capability.

The 298 amino acid PBCV-1 DNA ligase is the smallest characterised eukaryotic DNA ligase containing all six motifs that characterise the nucleotidyltransferase superfamily to which DNA ligases belong. It contains no additional amino acid sequences downstream of motif VI, which is located at its extreme C-terminus and only 26 amino acids N-terminal of the active site (lysine 27) within motif I (Odell & Shuman, 1999). Cellular DNA ligases are much larger, for example, human ligases I, III and IV are 919, 922 and 844 amino acid polypeptides respectively whereas at 34

kDa, PBCV-1 DNA ligase is the smallest and most well researched eukaryotic DNA ligase. This size makes it an attractive choice to use as a paradigm for all ATP-dependent enzymes. Structural comparison of HuligI and PBCV-1 DNA ligase reveal that the adenylation (AdD) and oligomer binding (OB) domains adopt similar conformations while the DNA binding domain of HuLigI and β-hairpin latch of PBCV-1 DNA ligase occupies almost the same angular position when docked on the nicked DNA substrate (Shuman, 2009). This suggested that PBCV-1 would be a useful representative of ATP-dependent DNA ligases for the large volume, high-throughput screening step of our *in vitro* assays. The small size of the PBCV-1 enzyme and lack of rare codons made it an ideal candidate for high-level expression.

The recombinant PBCV-1 DNA ligase was expressed from the plasmid pET-16b, a construct that adds a ten histidine, amino acid tag to the amino terminus of the mature recombinant polypeptide. This enabled the use of Ni-NTA resin (that has affinity for multiple histidine residues) to purify recombinant His-tagged PBCV-1. Although this step helped remove many contaminating polypeptides some low molecular weight contaminating polypeptides were still retained in the protein sample. These additional protein species may have been due to cleavage of the PBCV-1 ligase by cellular proteases during protein synthesis; some of the cleavage products retaining the histidine tag would thus co-purify with the full-length ligase during Ni-NTA chromatography. For use in an enzyme-based inhibitor assay further processing of the enzyme preparation was needed, therefore the protein was further purified, to near homogeneity, using S-sepharose ion-exchange resin.

In recent years the essential NAD^+-dependent DNA ligases of eubacteria have been proposed and are under development as potential antibacterial targets on the basis of their cofactor requirement and sequence differences between these enzymes and their eukaryotic host (Brotz-Oesterhelt *et al.,* 2003; Srivastava *et al.,* 2005). These differences include an absence of motif VI in the OB domain of the catalytic core of NAD^+-dependent DNA ligases and different accessory domains extending from the catalytic core of these ligases (Wilkinson & Bowater, 2001; Sriskanda & Shuman, 2002). However, in addition to NAD^+-dependent DNA ligases, some bacteria also encode for ATP-dependent DNA ligases; for instance, *Mycobacterium tuberculosis*

codes for three different such enzymes: LigB, LigC and LigD in addition to an NAD^+-dependent ligase, LigA (Gong *et al.*, 2004) whereas *Bacillus subtilis* harbours two ATP-dependent ligases (YkoU and YoqV) and an essential NAD^+-dependent ligase (YerG). All three ATP-dependent ligases of *M. tuberculosis* were found to be capable of nick-sealing, although their catalytic efficiency was lower than that of LigA (Gong *et al*., 2004). Neither LigB, LigC or LigD are essential for Mycobacterial growth, although deletion of LigD renders the bacteria defective in NHEJ (Gong *et al.*, 2004). The crystal structure of *M. tuberculosis* LigD has been solved. The enzyme is composed of a basic DNA ligase catalytic core (AdD and OB domain), equivalent to that of the PBCV-1 ligase, fused to a polymerase and phosphoesterase module (Akey *et al*., 2006). LigD functions in NHEJ, but unlike PBCV-1 ligase, is inefficient at *in vitro* nick joining and is unable, alone, to sustain cell viability in yeast or *M. tuberculosis* cells (Gong *et al*., 2004; Akey *et al*., 2006).

Cheng & Shuman (1997) identified and characterised a 31 kDa (268 amino acids) ATP dependent DNA ligase from *Haemophilus influenzae.* It was referred to as URF1 before its characterisation as an ATP-dependent ligase, and its deletion had lethal effects on the organism; suggesting that the ATP-dependent ligase is critical for *H. influenzae* cell growth (Cheng & Shuman, 1997). The *H. influenzae* enzyme at 31 kDa is markedly smaller than the multi-domain ligase proteins found in Mycobacteria, possessing only a nucleotide binding domain and an OB-fold domain. The PBCV-1 enzyme also lacks the DNA binding domain of the larger cellular enzymes and has been shown to bind to DNA by a different mechanism involving a short nucleotide loop (Nair *et al.*, 2007). We searched for other small ATP-dependent DNA ligases, in size comparable to PBCV-1 DNA ligase, to identify the type of organisms that have such a DNA ligase. BLAST searches of the NCBI database, performed using PBCV-1 DNA ligase protein as the query sequence, revealed that PBCV-1 shared sequence similarity with a number of other small, putative ATP-dependent DNA ligases (all between 30 and 40 kDa in size). It was interesting to note that most of the protein sequences of the small, ATP-dependent DNA ligases homologous to PBCV-1 DNA ligase, were encoded by bacteria including many pathogenic organisms for example *Vibrio cholerae*, *Haemophilus influenzae*, *Burkholderia pseudomallei* and *Neisseria meningitidis* (Figure 4.9, panel A). Many of these enzymes are uncharacterised except

those of *Haemophilus influenza* (31 kDa) and *Neisseria meningitides* (31.5 kDa) which have been studied and confirmed to catalyse DNA end-joining, and are thus functional ligases (Cheng & Shuman, 1997; Magnet & Blanchard, 2004). Sequence alignment of these proteins suggests that they share the same catalytic core and conserved motifs as PBCV-1 ligase. It was noted from the sequence alignment of PBCV-1 DNA ligase and other small ATP-dependent ligases that the region homologous to the essential DNA binding surface loop ('latch') motif present in the PBCV-1 ligase (Figure 4.9, highlighted in orange) is truncated in most of these enzymes except the 306 amino acid DNA ligase from *Burkholderia pseudomallei.* A gram negative bacterium *B. pseudomallei* is the causative agent of melioidosis; a fatal disease causing an acute pneumonia or acute septicaemia. The organism is resistant to a number of antibiotics thereby creating a need for the development of other therapeutic strategies, including the identification of novel therapeutic targets (O'Carroll *et al*., 2003). The use of the PBCV-1 enzyme in the assay will hopefully provide information on compounds that can have inhibitory effects on smaller ligases but particularly the *B. pseudomallei* enzyme where the enzyme retains an almost identical DNA binding loop to the PBCV-1 enzyme.

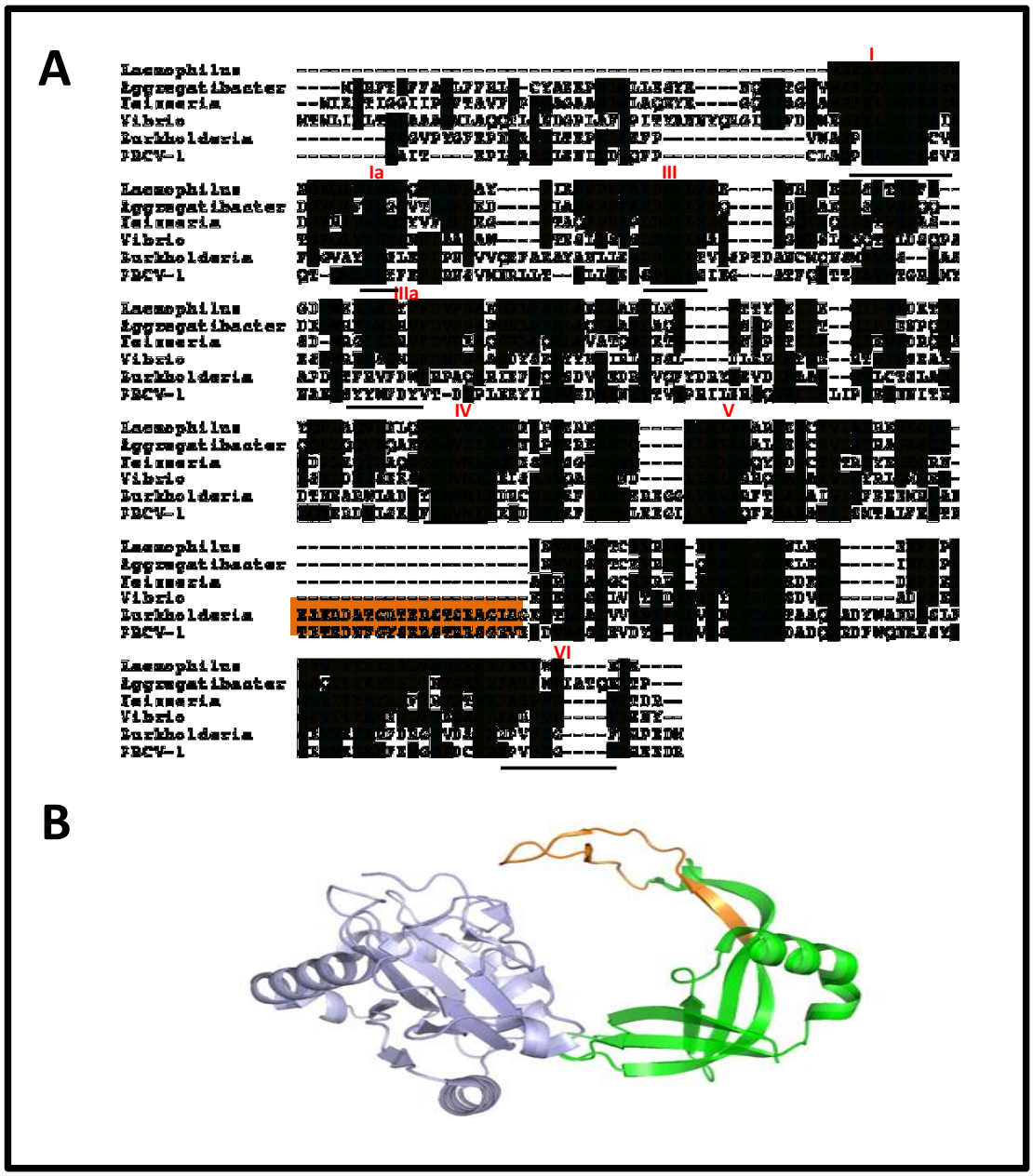


Figure 4.9: Sequence and structural comparison of PBCV-1 DNA ligase with a selection of small ATP-dependent ligases and structure of the DNA binding loop

Panel A shows a ClustalW alignment analysis (Chenna *et al.*, 2003) of PBCV-1 ligase with a selection of ATP-dependent DNA ligases from bacteria, with their accession number indicated in brackets - *Haemophilus influenzae* (Haemophilus; YP_001290961), *Aggregatibacter aphrophilus* (Aggregatibcater; YP_003007537), *Neisseria meningitidis* (Neisseria; AAQ72358), *Vibrio cholerae* (Vibrio; ZP_04418809) and *Burkholderia pseudomallei* (Burkholderia; YP_333052). Completely conserved residues are shown in green, identical residues in yellow and similar residues are shown in cyan. The position of conserved motifs I- VI is shown with a black bar underneath. The residues that make up the surface loop 'latch' in PBCV-1 and *Burkholderia pseudomallei* ligase are highlighted in orange. Panel B shows the ribbon structure of PBCV-1 DNA ligase (PDB: 2QT), highlighting in orange the residues in the latch region which are absent from the sequences of the other putative DNA ligases except *Burkholderia pseudomallei* in panel A. The figure was prepared using Pymol (Delano, 2002).

B. pseudomalleus encodes for a 75.6 kDa NAD^+-dependent DNA ligase, homologous to that of *E. coli* (LigA; accession number: Q3JR23-1). Although the exact function of the ATP-dependent DNA ligase in *B. pseudomallei* is not known, previous studies suggest that ATP-dependent DNA ligases in bacteria are not required for DNA replication but participate in DNA repair pathways (Gong *et al.,* 2004; Korycka-Machala *et al.,* 2005). Weller *et al*. (2002) observed that deletion of the ATP-dependent DNA ligase (YkoU) in *B. subtilis* is not lethal, confirming that this pathway is non-essential for growth but the mutant strains were sensitive to ionizing radiation in stationary phase (Weller *et al.,* 2002).

We propose that such small ATP-dependent DNA ligases in pathogenic bacteria especially, *B. pseudomallei,* might be dispensable for cell growth but their inhibition may render the repair mechanism inefficient and could be exploited as potential target in combinatorial drug therapy. Our belief is that any small molecule or chemical compound inhibiting PBCV-1 DNA ligase would also have inhibitory effect on the ATP-dependent DNA ligase of *B. pseudomallei.* Furthermore the complementation by additional ATP-dependent DNA ligases encoded by eubacteria might complicate potential strategies to target NAD^+-dependent DNA ligases with antibiotics. We therefore realised that the inclusion of a small ATP-dependent ligase would extend the screening assay to identify inhibitors for DNA ligases encoded by eubacteria. The main focus of this study has been to develop an assay to identify inhibitors for eukaryotic ligases, exemplified by the TV enzyme and HuLigI, using PBCV-1 as an initial catalytic inhibitor screen. However, it is clear that once potential inhibitors are identified their efficacy can also be tested as potential anti-bacterial agents. Furthermore, the conditions under which the wild-type PBCV-1 DNA ligase crystallises are known, in fact they can be grown from three different conditions, so it would hopefully be possible to grow such crystals with the inhibitor or to soak the inhibitor into the wild type crystals. Therefore its use can be advanced as a screening agent, through the use of the crystal structure of the PBCV-1 in complex with the inhibitor to enable modifications to be made to the compounds structure chemically to enhance its activity. Inhibitors identified from this screen with PBCV-1 will then be analysed against the Δ232 HuLigI and the TV enzyme in the same *in vitro* screen to confirm (hopefully) the inhibitory potential against these enzymes.

CHAPTER 5

DEVELOPMENT OF AN ASSAY TO SCREEN FOR INHIBITORS OF DNA LIGASES

CHAPTER 5: DEVELOPMENT OF AN ASSAY TO SCREEN FOR INHIBITORS OF DNA LIGASES

5.1 Introduction

Chemotherapy and radiation during treatment of cancer has the ability to induce DNA damage and hence different mechanisms are employed by the cells to detect and repair such damage. The modulation of these pathways has a great impact on clinical outcome and is frequently responsible for therapeutic resistance to anti-cancer agents (Martin, 2001; Damia & D'Incalci, 2007). Such studies have lead to the identification of several potential drug targets and consequently the development of small molecular inhibitors of DNA repair (Madhusudan & Middleton, 2005). Given the unique and critically important roles of DNA ligases in DNA replication, DNA damage repair and maintenance of genomic stability, the efficacy of anti-cancer drug treatments may be improved in certain cases by targeting human DNA ligases.

DNA ligases utilize either ATP or NAD^+ as cofactors to catalyse the formation of phosphodiester bonds in nicked DNA. Differences in the cofactor specificity of bacterial DNA ligases have made them an attractive target for broad-spectrum antibiotics (Korycka-Machala *et al.,* 2007). Recent studies have identified various small molecules such as pyridochromanones that inhibit NAD^+-dependent DNA ligases (Gong *et al.,* 2004). Although, the NAD^+-dependent ligases are restricted to eubacteria some bacteria encode for both NAD^+ and ATP-dependent DNA ligases (Doherty & Suh, 2000). *Mycobacterium tuberculosis* encodes for at least three different types of ATP-dependent DNA ligases and an NAD^+-dependent ligase (Gong *et al.,* 2004) whereas *Bacillus subtilis* harbours two ATP-dependent ligases and an essential NAD^+-dependent ligase (Petit & Ehrlich, 2000). Targeting NAD^+-dependent DNA ligases might not be immediately successful as additional ATP-dependent DNA ligases encoded by these organisms will influence the efficacy of the NAD^+-inhibitory compounds. This might lead to drug resistance or the opportunity for some of these strains to survive assault on the NAD^+-dependent DNA ligases. Therefore, there is a need to develop a compound or series of compounds that can be used in conjunction with NAD^+-dependent DNA ligase inhibitors to overcome possible drug resistance in these organisms.

This potential of ligase inhibition in cancer therapeutics or otherwise will only become apparent once potent and selective inhibitors become available for pharmacological studies over the next few years. DNA ligase inhibitors are likely to be anti-proliferative and potentiate the cytotoxicity of DNA damaging agents. Recent efforts by various independent research groups have resulted in the identification of many classes of ligase inhibitors by screening chemical and natural product libraries (Tan *et al.*, 1996; Chen *et al.*, 2008; Dwivedi *et al.*, 2008; Zhong *et al.*, 2008). The assays used in these efforts however, have not been well suited to high throughput screening of natural product extracts. This prompted us to develop a 96-well plate based assay that can be used for high throughput screening to identify inhibitors of ATP-dependent DNA ligase enzymes, human, protozoal and bacterial. The key feature of the rapid DNA ligation, inhibition assay is the relative binding of affinities of the components of the unligated DNA substrate and ligated DNA product to the ion exchange resin, Q-sepharose. Knowledge of the smallest size of DNA that ligases can efficiently ligate on the 5'-PO_4 side of the DNA nick was utilised to provide differential binding affinity between the unligated, labelled 5'-PO_4 strand and the sealed ligation product (Odell *et al*., 2003).

Crude extracts from the culture of white rot fungus, *Basidiomycetes* - already known for anti-tumour and anti- bacterial properties, were used as test extracts in developing the screening of the inhibitors (Russell & Patterson, 2006; Silva, 2006). Various ATP-dependent ligases (produced and purified in-house) were used to set up the assay. Various salt conditions and other steps required to optimise the assay are described here.

5.2 Materials and methods

5.2.1 Optimisation of oligonucleotides binding to Q-sepharose

A 30-mer hairpin strand with a 10-nucleotide 5' single-strand tail (Figure 5.1) and 6-mer oligonucleotide complementary to the tail were designed and then obtained from Invitrogen.

The oligonucleotides were diluted to a final concentration of 0.2 mg/ml with buffer A (50 mM Tris-HCl pH 7.5, 5 mM DTT and 10 mM $MgCl_2$) and applied separately to 1 ml of Q-Sepharose resin (Sigma-Aldrich) pre-equilibrated with buffer A. The bound fraction was step eluted with 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 and 1 M NaCl in buffer A. The DNA elution profile was monitored by checking the absorbance of the eluate at 260 nm.

5.2.2 Ligation inhibition assay methodology

5.2.2.1 Preparation of the nicked DNA substrate

A 36-mer (30 + 6-mer) nicked duplex was synthesised commercially as two separate oligonucleotides by Invitrogen and used as a substrate for the ligase reaction (see figure 5.1).

The 5'-end of the 6-mer oligonucleotide was labelled with ^{32}P using T4 polynucleotide kinase (as described in Chapter 2; 2.1.10). The nicked duplex DNA substrate (Figure 5.1) was formed by mixing the 6-mer strand labelled with ^{32}P at its 5' end with the complementary 30-mer strand at a molar ratio of 1:4 [1 (6-mer) : 4 (30-mer)] in 0.2 M NaCl as previously described (Chapter 2; 2.1.11).

5.2.2.2 Inhibition assay

The assay was carried out in 96-well microtitre plates. All potential inhibitor extracts (~8 µg) were supplied in solvent 30% Dimethylsulfoxide (DMSO): Methanol (1:1; v/v). To preclude the action of the solvent being responsible for the inhibition, DNA ligase activity of all test ligases (T4, PBCV-1 and HuLigI) was tested in a standard ligation assay (Chapter 2; 2.1.12) in the presence of 10%, 5%, 2.5%, 1.25%, 1.0%, 0.5% and 0.25% solvent and 0-4 µM Doxorubicin (Sigma-Aldrich) as a positive control for ligase inhibition in separate reactions.

The final ligation assay reaction mixture (20 µl) included 50 mM Tris pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 5 mM DTT, 1 pmole annealed (30 + 6-mer) nicked duplex DNA,

1 pmole DNA ligase and 0.5 µl of potential inhibitor (in 0.5% solvent). The reaction was incubated at 22^{o}C for 5 minutes and terminated by the addition of 5 µl of 0.1 M EDTA.

100 µl of Q-sepharose resin, pre-equilibrated with buffer A (50 mM Tris pH 7.5, 10 mM $MgCl_2$ and 5 mM DTT) was suspended in each well of a 96 well filter*EX* filter plate (Corning). The filter plate containing Q-sepharose resin was stacked on top of a 96-well receiver plate and placed into the top chamber of a vacuum manifold (Figure 5.7). The vacuum manifold unit was connected to a pressure pump via a waste bottle covered with an appropriate filter to trap liquid.

The completed ligation reactions were heated at 65^{o}C for 5 minutes prior to loading onto the resin. The filter plate (Q-sepharose) was washed with either 300 µl or 3 x 100 µl of pre-warmed (37^{o}C) buffer A containing 0.2 M NaCl. Loosely bound annealed DNA substrates were eluted with buffer A containing 0.6 M NaCl and ligated DNA substrate was eluted with buffer A containing 1 M NaCl. The contents of the receiving plates were freeze dried for 8 hours, dissolved in 20 µl water and spotted onto nitrocellulose membrane (Hybond NX, GE Healthcare). The membrane was exposed to a phosphorimager plate that was subsequently scanned using Typhoon imager (GE Healthcare).

5.3 Results

5.3.1 Determination of oligonucleotide binding affinity to Q-sepharose

The oligonucleotides used in this assay were designed based on a previous study that defined the minimum DNA length requirement for strand sealing on the 5'-PO_4 side of the nick by PBCV-1 DNA ligase (Odell *et al.*, 2003). These authors used a 30-mer oligonucleotide with internal base complementarity which allowed folding back and creation of a hairpin (as shown in figure 5.1; panel A) and annealing of a short complementary oligonucleotides (Figure 5.1; panel B) to this hairpin structure creates a synthetic double-stranded DNA (dsDNA) substrate with a single-strand nick (Figure 5.1; panel C). Odell *et al.* (2003) showed that a 30-mer hairpin strand (Figure 5.1)

with a 10 nucleotide (nt) 5' single-strand tail when annealed with an excess of complementary 10-mer, 8-mer, 6-mer and 4-mer 5'-PO_4 strands can be joined *in vitro* to yield a 40-mer, 38-mer, 36-mer and 34-mer product respectively. However the ligation of the 4-mer strand was less successful, possibly due to the absence of a duplex segment distal to the annealed 4-mer strand at the nick or more likely the thermal instability of such a construct. Identical rates and extents of sealing was observed for the 10-mer, 8-mer and 6-mer substrates, leading to the conclusion that a 6 nucleotide 5'-PO_4 strand sufficed for efficient ligation (Odell *et al.*, 2003). We therefore set out to devise a ligase inhibition assay method using a 30-mer hairpin strand with a 10 nt 5' single-strand tail and 6-mer oligonucleotide [^{32}P-labelled *in vitro* at its 5' terminus; (Figure 5.1)] complementary to the tail as the DNA substrate.

To follow the course of sealing of this nicked dsDNA substrate the 6-mer will be pre-incubated with T4 polynucleotide kinase and gamma-labelled ATP resulting in the attachment of a radioactive phosphate moiety to the 5' terminus of the oligonucleotide. This radioactive 6-mer then becomes incorporated into the backbone of the hairpin on successful DNA ligation creating a 36 nucleotide labelled product.

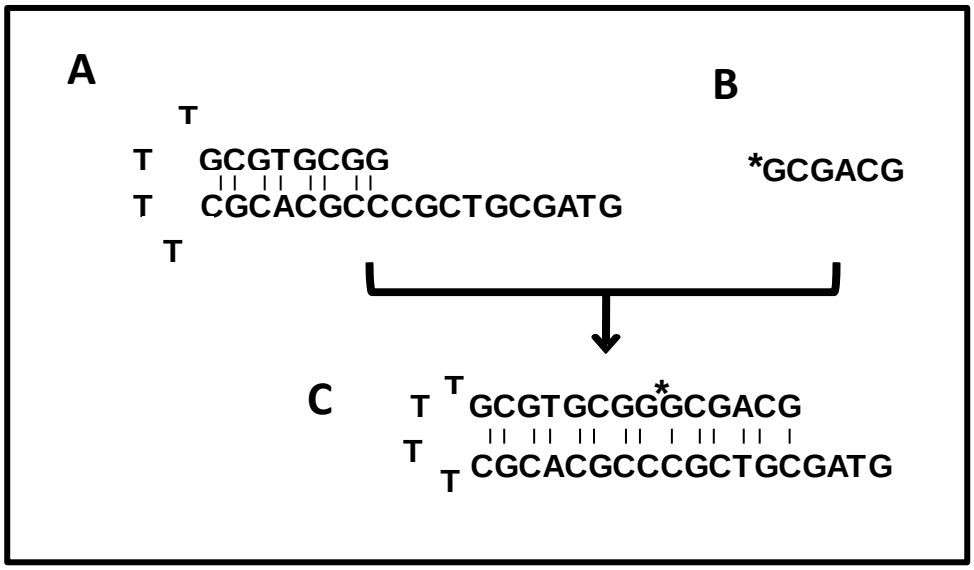


Figure 5.1: Structure of the radiolabelled synthetic substrate used in DNA ligation assays

The oligonucleotides were designed as described by Odell *et al.* (2003). A: Unlabelled 30-mer hairpin strand with an unpaired 10 nucleotide 5' single-strand tail; B: Radiolabelled 6-mer oligonucleotide complementary to the tail of hairpin structure (the asterisk indicates the position of the labelled nucleotide); C: Final annealled labelled DNA substrate.

The key principle behind the DNA ligation/inhibition assay is the relative binding affinities of the DNA substrates (the 6-mer and the 30-mer and the product 36-mer) to Q-sepharose (a sepharose resin with attached quaternary ammonium groups functioning as an anion exchanger) resin. The negatively charged phosphate groups of DNA are fully ionized at a pH above 4 as a result of which DNA binds to positively charged groups such as the quaternary amine group (Tseng & Ho, 2003).

To test the assumptions that underpin the assay, solutions of both the 30-mer and the 6-mer at 0.2 mg/ml were applied independently to a Q-sepharose column in ligation buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT). The binding affinity of the 30-mer hairpin structure and the 6-mer oligonucleotide to the Q-sepharose resin was expected to be different as the 30-mer will have more potential sites for attachment to the matrix and will hence show different affinity to the 6-mer when eluted with salt. Figure 5.2 shows the salt concentrations where the 30-mer and 6-mer are released from the Q-sepharose. The 6-mer oligonucleotide eluted at 0.25-0.3 M NaCl whereas the 30-mer hairpin structure eluted at 0.5 M NaCl. As the final product of a successful ligation would be a 36-mer, it was concluded that this would demonstrate tighter binding to the Q-sepharose resin than the 30-mer allowing differentiation between the unligated 6-mer and the ligation product by affinity to the ion exchange resin.

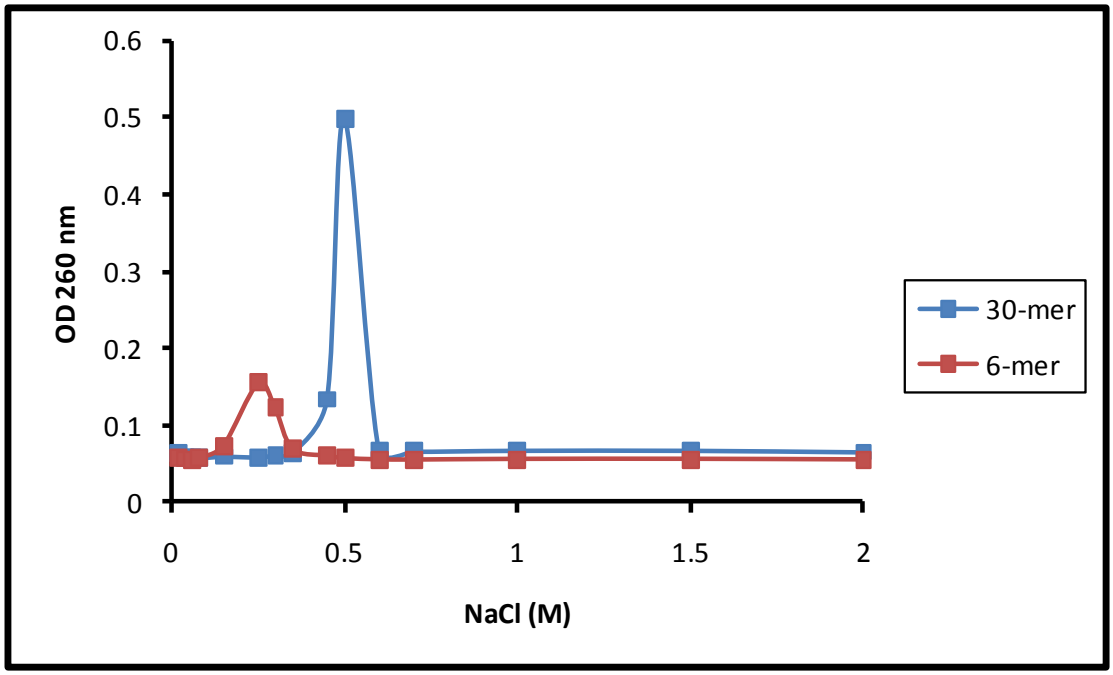


Figure 5.2: Elution profile of oligonucleotide DNA substrates, hairpin 30-mer and linear 6-mer from Q-sepharose under NaCl treatment

The 30-mer and 6-mer oligonucleotides (0.2 mg/ml) were applied independently to a Q-sepharose column in buffer A (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT) and step-eluted with 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 and 1 M NaCl in buffer A. The DNA elution profile was monitored by checking the absorbance of the eluate at 260 nm.

5.3.2 Inhibitory effect of Doxorubicin

Before the search for inhibitors could be undertaken it was necessary to show that the physical manipulations of the assay and the buffer conditions would not affect the normal function of a ligase enzyme. Furthermore it was important to demonstrate that a known ligase inhibitor, doxorubicin, could show activity under these conditions.

Doxorubicin is a highly effective anthracycline chemotherapeutic agent known to inhibit the ATP-dependent DNA ligase of bacteriophage T4 and NAD^+-dependent DNA ligase of *E. coli* with a similar potency (Ciarrocchi *et al.*, 1999). The inhibitory effect of doxorubicin was studied by titrating doxorubicin against HuLigI (Δ232), PBCV-1 and T4 DNA ligases in a standard ligation reaction as described previously (Chapter 2, 2.1.12).

The 6-mer oligonucleotide was end-labelled with ^{32}P in a kinase reaction and annealed with the 30-mer hairpin (as described in Chapter 2; 2.1.10, 2.1.11) to create a synthetic nicked dsDNA substrate. The annealed product was incubated at 22^{o}C for 5 minutes with the test ligases (T4, PBCV-1 and Δ232 HuLigI) with and without doxorubicin. The extent of ligation in each sample was quantified from analysis of the sample on a denaturing polyacrylamide gel. It was found that 0.5 µM had no discernible effect on DNA ligation under the reaction conditions used however 3 µM doxorubicin was sufficient to inhibit >90% of the ligation activity of all the above enzymes (Figure 5.3). These values closely parallel those reported by Ciarrocchi *et al.* (1999) for the T4 enzyme.

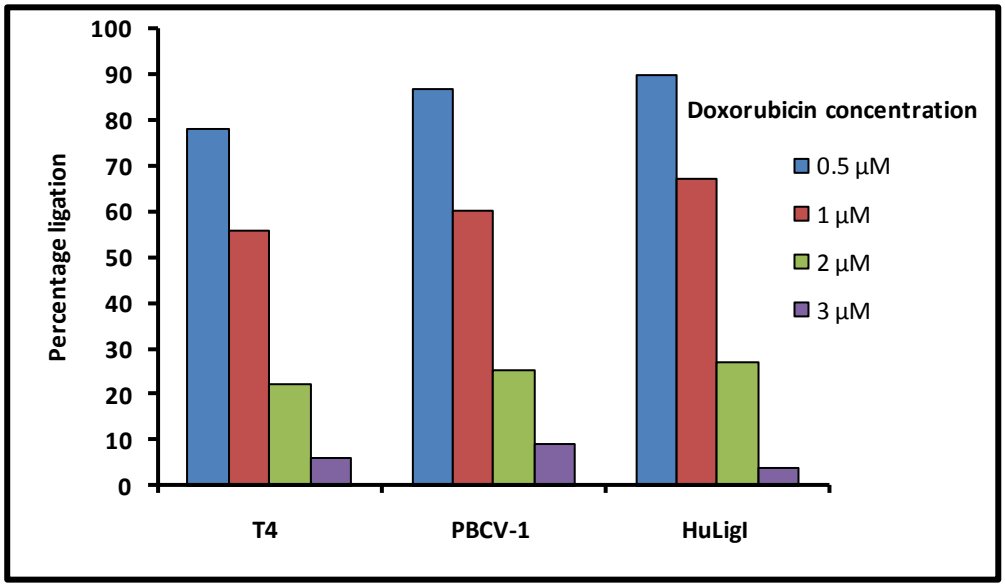


Figure 5.3: The effect of Doxorubicin on ATP-dependent DNA ligation monitored with a radiolabelled, nicked DNA substrate.

The activity of T4, PBCV-1 and HuLigI were tested in the presence of different concentrations of doxorubicin (0-3 µM). The standard ligation reaction was carried out at 22^{o}C for 5 minutes and stopped by the addition of 5 µl of 0.1 M EDTA and heating the sample at 95^{o}C for 5 minutes. The extent of ligation in each sample was quantified from analysis of the sample on a denaturing polyacrylamide gel analysed using a Typhoon Phosphorimager (GE Healthcare) and Imagequant software.

5.3.3 Validation of assay conditions

The elution profile of the oligonucleotides (30-mer and 6-mer; Figure 5.2) confirms a difference in their binding affinity to Q-sepharose and reveals that a higher salt concentration is required to remove bound 30-mer oligonucleotide compared to bound

6-mer. The binding affinites of annealed (30+6)-mer substrate and ligated product (36-mer; using ^{32}P-labelled 6-mer) to Q-sepharose resin were then determined to optimise the washing protocol of the assay. A ligation reaction (50 µl) was carried out with T4 DNA ligase, PBCV-1 DNA ligase and (Δ232) HuLigI on annealed (30+6 mer) substrate and loaded independently onto Q-sepharose resin. The resin was step eluted with 0.25, 0.6 and 1 M NaCl and the fractions obtained were spotted onto a nitrocellulose membrane and exposed to a phosphorimager screen (Figure 5.4; panel A).

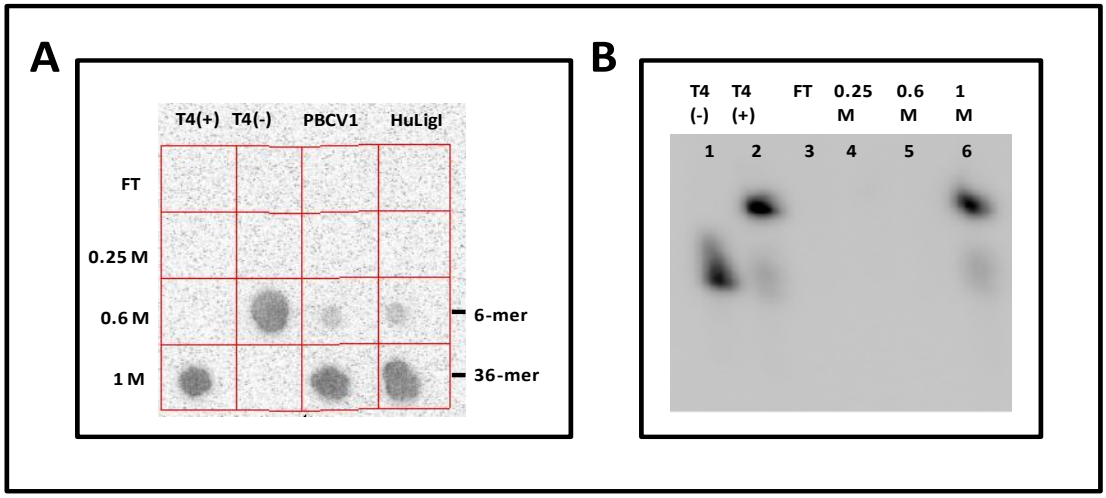


Figure 5.4: DNA ligation assay products analysed by spotting onto a nitrocellulose membrane and by denaturing gel electrophoresis

Panel A: A 50 µl ligation reaction was carried out with 1 pmole of annealed (30+6 mer) substrate in ligation buffer (containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT**)** in the presence of 0.1 pmole of either T4 DNA ligase, PBCV-1 DNA ligase or human (Δ232) DNA ligase I and loaded independently onto Q-sepharose resin. The resin was step eluted with 0.25, 0.6 and 1 M NaCl and the fractions obtained were spotted onto a nitrocellulose membrane and exposed to a phosphorimager screen. Lane1: control reaction with T4 DNA ligase; Lane 2: control reaction without ligase; Lanes 3 & 4: NaCl elution fractions obtained from the PBCV-1 and HuLigI reactions respectively.

Panel B**:** The reaction products from the same reaction shown in panel A were also analyzed using 15% denaturing polyacrylamide gel electrophoresis. The gel was exposed to a radioactive detecting screen and scanned using a phosphorimager (GE Healthcare). The autoradiogram is shown. Lane 1: DNA substrate with no ligase; Lane 2: control reaction with T4 ligase before loading onto the column; Lanes 3-6: different fractions collected after washing the Q-sepharose column loaded with the ligation reaction carried out with the hairpin substrate in the presence of T4 DNA ligase. Lane 3: FT- Flow-through, Lane 4: 0.25 M NaCl, Lane 5: 0.6 M NaCl, Lane 6: 1 M NaCl.

The preliminary wash was performed with 0.25 M NaCl to confirm that all the 6-mer had annealed to the 30-mer hairpin creating ds DNA, which in turn elutes at 0.6 M NaCl. In the absence of any ligase all the radiolabelled DNA species eluted from the resin at 0.6 M NaCl confirming that the substrate had formed correctly (Figure 5.4; panel A). Incubation with commercial T4 DNA ligase converted the radiolabelled

DNA into the final product of a successful ligation, the 36-mer. This demonstrated even higher affinity for the Q-sepharose resin than the 30-mer alone and hence eluted at 1 M NaCl. The absence of any radioactive signal in the resin flow through, 0.25 M and 0.6 M NaCl wash fractions indicates that all the available DNA substrate has been ligated. In reactions performed with PBCV-1 and the truncated HuLig I a very faint signal in the 0.6 M salt fraction suggests the presence of a very small amount of unligated DNA substrate (Figure 5.4; panel A). This may have occurred if the DNA ligase disengaged from the substrate after second step of the reaction i.e. after transferring the AMP moiety to the end of the DNA before completing the ligation reaction. This fraction would then be resistant to DNA ligation, but would remain annealled to the hairpin. This error occurs *in vivo* in human DNA metabolism as the cells produce a protein, aprataxin, which can reverse such stalled ligation intermediates (Ahel *et al.*, 2006). Aprataxin repairs such lesions by catalyzing the release of 5′-adenylate groups from the single-strand breaks, resulting in the restoration of 5′ phosphate termini that can be efficiently rejoined (Ahel *et al.,* 2006).

The results from the dot blot analysis were confirmed by electrophoresing, on a denaturing polyacrylamide gel, the various fractions obtained from the NaCl step-gradient elution of a ligation reaction. T4 DNA ligase was incubated with the radiolabelled substrate and the products were applied to Q-sepharose before washing with NaCl (Figure 5.4; panel B). The input size of the DNA in the absence of ligase represents the migration of the labelled 6-mer. Upon ligation the 6-mer converts to a slower migrating product whose size increase we take to indicate as conversion to a 36 nucleotide product. The larger, ligated product elutes from the Q-sepharose resin only upon washing with 1 M NaCl (Figure 5.4; panel B; lane 6).

To confirm that 0.6 M NaCl wash distinguishes between ligated and unligated DNA, T4 DNA ligase was used to seal the nick in the preformed DNA duplex in the presence of an inhibitory concentration of doxorubicin. The ligation reaction mixture was loaded onto Q-sepharose resin and step eluted with 0.25, 0.6 and 1 M NaCl. The eluted fractions were run on a 15% denaturing polyacrylamide gel (Figure 5.5). The reaction performed in the presence of doxorubicin, gave an intense signal at 0.6 M, indicating that unligated substrate elutes at this salt concentration.

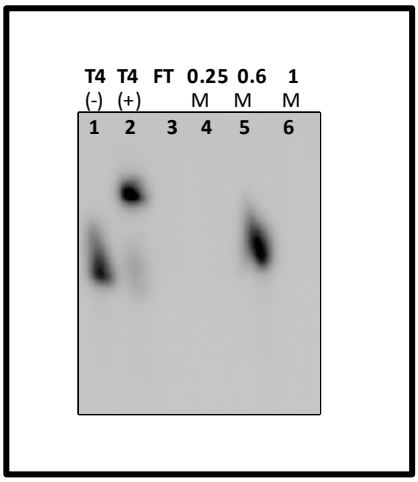


Figure 5.5: T4 DNA ligase sealing in the presence of doxorubicin analysed by conventional electrophoresis

A ligation reaction performed with 0.1 pmole T4 DNA ligase, 1 pmole DNA substrate in buffer A (containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT) in the presence of 3 µM doxorubicin, loaded onto Q-sepharose resin. The resin was then washed with 0.25, 0.6 and 1 M NaCl in buffer A and the different fractions were collected and run on 15% denaturing gel. Lane 1: control reaction with no ligase; Lane 2: control reaction with 1 pmole T4 DNA ligase in the absence of doxorubicin; Lanes 3-6: different elution fractions collected after washing the column loaded with the ligation reaction carried out in the presence of doxorubicin. Lane 3: FT- Flow-through, Lane 4: 0.25 M NaCl, Lane 5: 0.6 M NaCl, Lane 6: 1 M NaCl.

This experiment validated that the assay scheme that we wanted to employ would identify inhibitors of ATP-dependent DNA ligases in a manner comparable to electrophoretic analysis. In order that the assay could be automated and carried out rapidly we designed a procedure that would use 96-well microtitre plates. The sequence of steps involved in the inhibitor screening assay is shown in figure 5.6.

The standard ligation reaction, in the presence of potential inhibitors, is carried out in a 96-well microtitre plate (reaction volume 25 µl) and stopped by the addition of 5 µl of 0.1 M EDTA to each well and by incubating at 65^{o}C for 5 minutes. Each reaction sample is then pipetted (using a multi-channel pipette) into a second microtitre plate containing Q-sepharose resin, pre-equilibrated with ligation buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT). The samples and sepharose in the filtration plate lined with 0.25 mm glass fibre filter are placed onto 96-well collection plates. The resin was washed with warm 0.2 M NaCl to remove enzyme and inhbitors that might affect the elution of the DNA. The unligated DNA bound to

Q-sepharose elutes with the 0.6 M NaCl wash whereas ligated DNA duplex, binds more tightly to the resin and elutes at 1 M NaCl.

Since, the 0.6 M elution fraction contains only unligated species, the extent of inhibition by different compounds could be measured by measuring the intensity of radioactive signal in this fraction as a fraction of the total radioactivity contained in both the 0.6 and 1 M fractions.

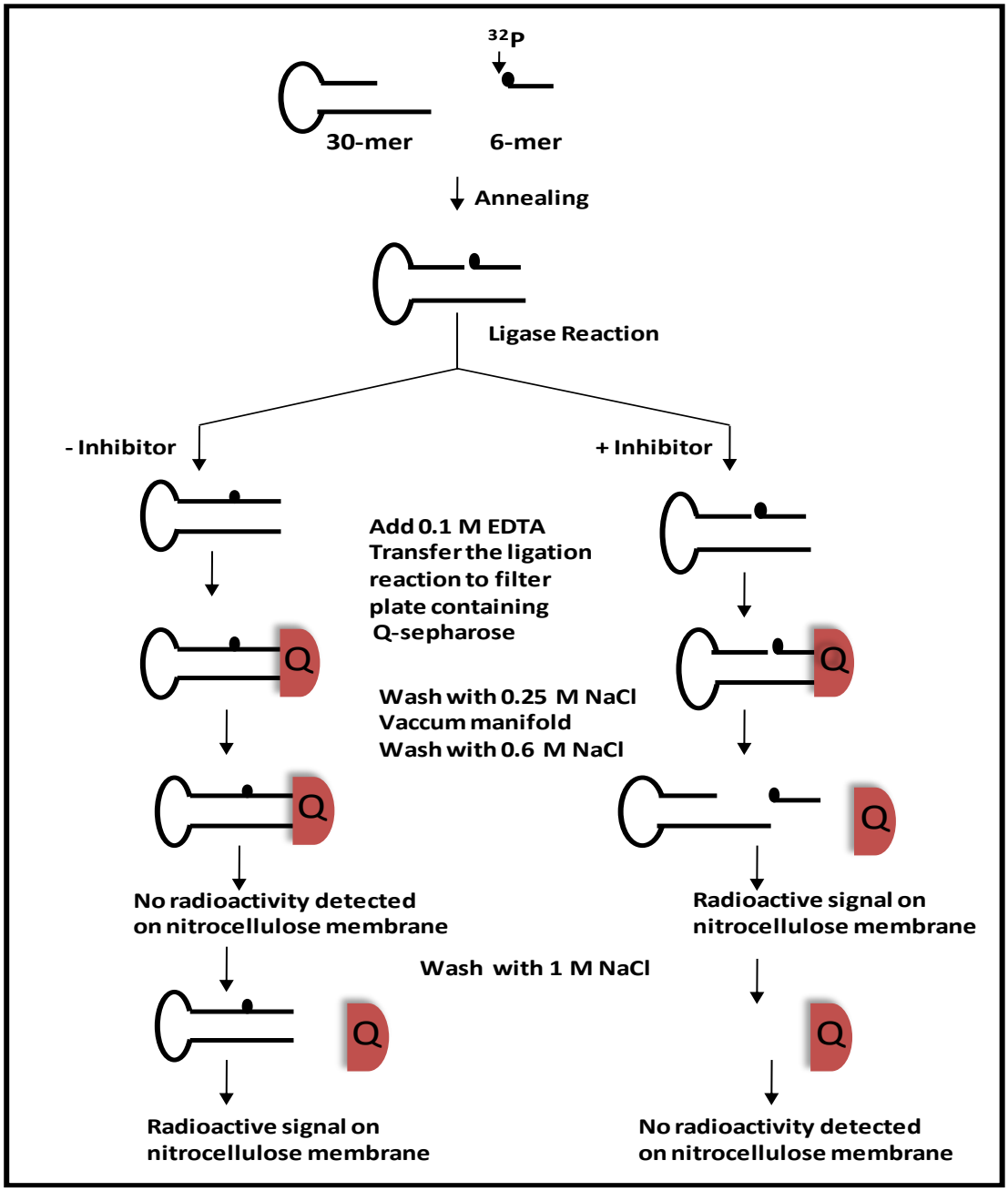


Figure 5.6: A schematic illustration of the various steps involved in the DNA ligase assay

The annealed double-stranded DNA duplex formed with a single-stranded nick is used as a substrate in DNA ligation reaction that can be performed in the presence of potential inhibitors. The ligation reaction is stopped by the addition of 0.1 M EDTA and incubating at 65°C for 5 minutes and mixed with Q-sepharose resin in the filter plate. The resin is washed with 0.25 M NaCl in buffer A (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT), followed by a 0.6 M NaCl wash in buffer A. A further 1 M NaCl wash in buffer A is performed to release ligated DNA. Differences in the radioactive signal intensity in 0.6 M and 1 M wash is calculated to determine the percentage inhibition of the ligation reactions.

Due to our desire to create a high-throughput assay, the lower collection plate was replaced with nitrocellulose in intimate contact with the base of the Q-sepharose resin microtitre plate. This assembly was then put onto a vacuum manifold designed to draw the various elution fractions through onto the nitrocellulose. At each step a new nitrocellulose membrane was inserted so that each part of the wash and elution could be quantified. By this means it was hoped that the degree of ligation inhibition could be monitored, by quantifying the total radioactivity eluted from the column in the 0.25, 0.6 and 1 M NaCl fractions. The proportion present in the 1 M fraction would represent the percentage of successful ligation achieved. Various efforts were made to place the nitrocellulose membrane in intimate contact with the filtration plate to avoid any loss of signal. Despite many attempts, a sufficiently tight seal could not be obtained between the vacuum and the nitrocellulose membrane. It was found that the volume of elution fraction (300 µl) obtained from each well was more than the vacuum pump could extract before cross-contamination occurred between samples. In order to prevent sample overflow, the volume of the buffer used to elute DNA was reduced but it resulted in inefficient washing of the resin. Figure 5.7 shows the signal overlap due to sample diffusion when the nitrocellulose membrane was placed in contact with filtration plate and we had adjusted all possible physical factors to increase the vacuum and reduce cross-contamination of the samples.

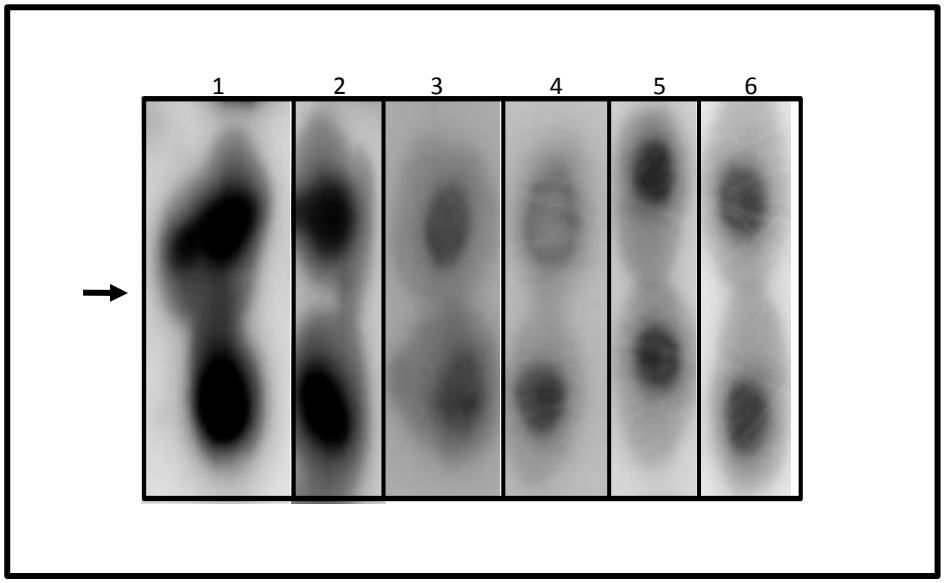


Figure 5.7: DNA ligation assay processed through Q-sepharose onto nitrocellulose using a vacuum manifold

Various panels showing diffusion of signal across the nitrocellulose membrane, when placed in intimate contact with filtration plate during elution. Arrow indicates the overlapping zone between samples. Panels 1-3 represents overlapping of radioactive signal observed after washing the resin with 300 µl of elution buffer whereas panels 4-6 represents overlapping of radioactive signal observed after reducing the volume of elution buffer (3 x 100 µl).

Therefore, we decided to allow the various washes of the Q-sepharose to pass by gravity flow into separate receiving plates and then to freeze dry these elutions. The samples were then dissolved in 20 µl of distilled water and spotted onto nitrocellulose membranes and exposed to a phosphorimager plate without mixing of the samples taking place (see figure 5.8).

Once we had established that the assay was as reliable as the standard electrophoretic assay used to measure DNA ligase activity, potential inhibitors were screened from the Mycodiverse^{TM} library of Hypha Discovery, London. The Hypha Discovery chemical library contains dried whole cell mass and extracts derived from cultures of *Ganoderma* species. *Ganoderma* is a basidiomycete white rot fungus whose extracts contain many bioactive compounds, including polysaccharides and terpenoids, some of which are known for their biomedical importance including anti-tumour and anti-bacterial activities (Russel & Patterson, 2006; Silva, 2006).

The dried extracts of white rot fungal cultures present in a specially selected bioactive subset, EC10B were dissolved in solvent (Dimethylsulphoxide: Methanol in a 1:1 ratio). The activity of HuLigI, PBCV-1 and T4 DNA ligase were tested in the presence of 0.25-10% v/v solvent in a standard ligation assay. 1% of solvent had no effect on the ligation activity of any of these enzymes (data not shown). In order to be certain that the solvent would not have a significant impact on ligation reaction, its concentration was further halved (0.5%) in screening assays.

After performing the ligation reaction in the presence of potential inhibitors from the EC10B collection (Hypha Discovery), the assay procedure was followed, as detailed in Figure 5.7. The control reactions (data not shown) were performed with T4 DNA ligase and PBCV-1 DNA ligase (with and without Doxorubicin). The different wash fractions, 0.6 M (Figure 5.8; panel A) and 1.0 M NaCl (Figure 5.8; panel B), of the Q-sepharose resin were collected and spotted onto nitrocellulose membranes. The extent of inhibition by different compounds was measured by measuring the intensity of radioactive signal in 0.6 M fraction as a fraction of the total radioactivity contained in both the 0.6 and 1 M fractions. It was observed that almost all the potential inhibitors pools from the EC10B collection were able to inhibit the activity of PBCV-1 DNA ligase (Figure 5.8) but the extent of inhibition varied in each reaction.

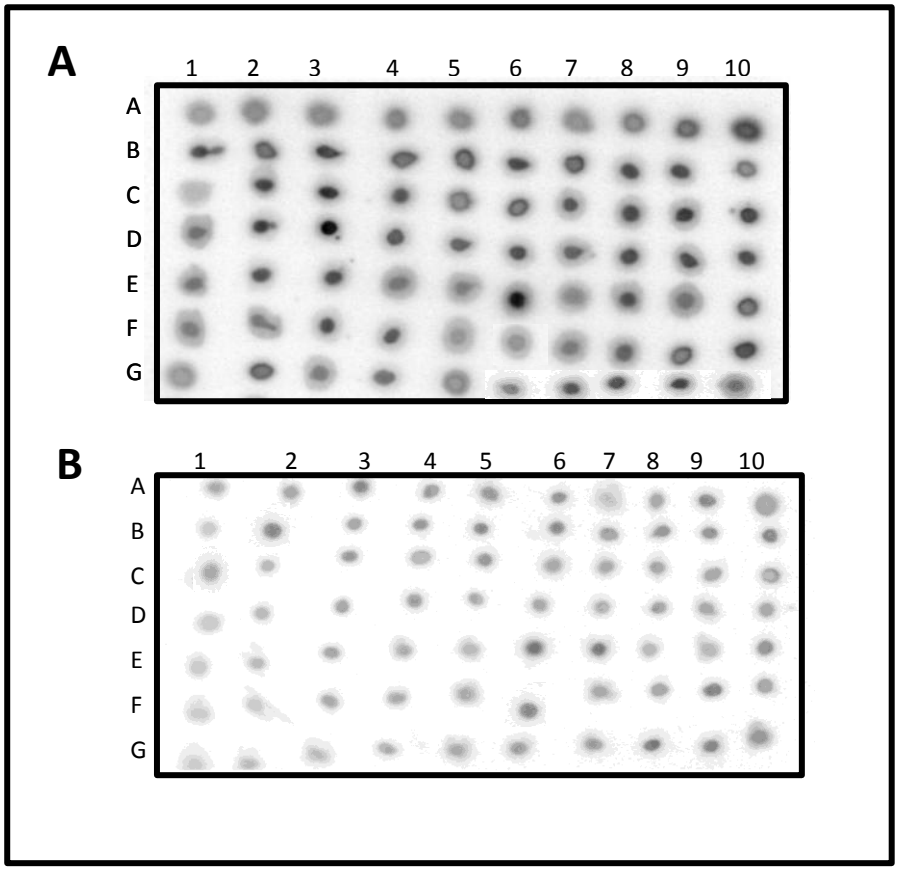


Figure 5.8: PBCV-1 ligase inhibition by EC10B *Ganoderma* extracts revealed in the wash fractions of the ligase assay

The ligation reaction using PBCV-1 DNA ligase performed in the presence of *Ganoderma* extracts was applied to Q-sepharose resin. The resin was washed with 0.25 M, 0.6 M and 1 M NaCl in buffer A (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 5 mM DTT).
Panel A: The 0.6 M wash fraction collected in collection plates, after vacuum manifold, was freeze-dried and spotted onto a nitrocellulose membrane. The membrane was exposed to a Phosphorimager radioactive detection screen and read using a Typhoon scanner (GE Healthcare).
Panel B: The membrane was washed with 1 M NaCl in buffer A to elute the 36-mer ligated product from the resin. The 1 M wash fraction was collected, freeze-dried and spotted onto a nitrocellulose membrane. The membrane was exposed to a Phosphorimager radioactive detection screen and read using a Typhoon scanner (GE Healthcare).

The EC10B collection contains dried cell mass extracts where each pool contains several chemical elements. As such many of these may have a toxic effect on the action of an enzyme. The only way to be certain of the presence of specific ligase inhibitor is to titrate the amount of toxic component. This could be achieved by setting an arbitrary lower limit of inhibition, for example 90%, and then decreasing the percentage of *Ganoderma* extract being added by a sequential series of two-fold

dilutions. If inhibition continues at a level of 90% over more than three such diltuions then the presence of a specific inhibitor may be inferred.

The rapid assay could be adapted to give more reliable results by using pure chemical or purified natural product libraries which would reduce the background signal and provide more precise information in terms of the specificity. Although, the first screening with the assay shows almost total inhibition of the reaction, serial dilution of the pools, used for screening, would lower the concentration of toxic contaminants and give a true measure of their inhibitory potential.

5.4 Discussion

The earliest stages of drug discovery involve identification of an appropriate biological target followed by development of an assay capable of rapid high throughput screening for effective, novel inhibitors. There is an emerging interest in developing small molecule inhibitors that can block the activity of some of the key proteins in the DNA repair processes to improve the effectiveness of current cancer treatments. Various chemicals that inhibit human DNA ligase I have been reported**,** including some widely used anti-tumour drugs such as doxorubicin**,** but little work has been done to develop a dedicated, rapid, non-electrophoretic assay to screen for other, ligase-specific inhibitors.

The widely used *in vitro* DNA ligase activity assay involves the monitoring of DNA ligation using a radiolabelled DNA substrate and analysing the results by gel electrophoresis. In this method of analysis, aliquots of the reaction are withdrawn at various times and quenched immediately with SDS to denature the enzyme and EDTA to chelate the essential metal cofactor, magnesium. The reaction products are analysed on a denaturing polyacrylamide gel. This method is time consuming and unsuitable for high throughput screening.

After the beginning of this work, Sun & Urrabaz (2004) published a non-electrophoretic assay for screening of inhibitors for DNA ligases. This assay was based on the use of radioactive label in addition to a bulky biotin group. The presence

of bulky group might affect the binding of inhibitors to the enzyme or DNA substrate used *in vitro* whereas the use of radioactively labelled phosphate alone rules out such a possibility.

The footprinting of PBCV-1 DNA ligase as studied by Odell & Shuman (1999; using an alanine substitution mutant of the active site Asp29 to inhibit strand joining) shows that the enzyme contacts 19–21 residues when bound to nicked duplex DNA. It extends 8–9 nucleotides on the 3'-OH side of the nick and 11–12 nucleotides on the 5'-PO_4 side of the nick. A similar asymmetric footprint was observed for T7 DNA ligase extending 9 nucleotides on the 5'-PO_4 side of the nick and 3–5 nucleotides on the 3'-OH side (Doherty & Dafforn, 2000). However, Odell *et al.* (2003) reported that PBCV-1 DNA ligase requires a minimum of 6 nucleotides at ligatable 3'-OH and 5'-PO_4 strands at the nick for efficient *in vitro* sealing. The discrepency between the footprint and the physical DNA size required for efficient ligation is probably explained by stearic hindrance. The PBCV-1 footprint was obtained by binding the ligase protein and then digesting with an exonucleolytic enzyme. Collision between the external surface of the two enzymes naturally overexaggerates the actual size of the ligase footprint on the DNA. These studies formed the basis of the design of the DNA substrates used in this assay system. The DNA consisted of a 30-mer strand with a stretch of nucleotides that were complementary, folding back on their Watson and Crick partner bases to create a hairpin with a 10 nucleotide 5' single-strand tail. To this tail a complementary 6-mer oligonucleotide (Figure 5.1) can be annealled that juxtaposes its 5'-PO_4 moiety adjacent to the 3'-OH of the 30-mer. On sealing this substrate a continuous 36-mer oligonucleotide is created (Figure 5.1).

The rapid assay method is based on the strength of interaction of the oligonucleotides with Q-sepharose, a strong anion exchanger. The binding affinities of the ligated 36-mer and the unligated substrate to Q-sepharose resin are different and hence have shown different elution profiles when the bound oligonucleotides are eluted with the salt NaCl. The 36-mer ligated product binds more tightly to the resin eluting with 1 M NaCl as compared to 0.6 M NaCl for the unligated 6-mer oligonucleotide when part of the annealed DNA substrate. The validity of our conclusions was confirmed by

comparing our fractionated ligation reaction on a denaturing gel in direct comparison with a complete ligation reaction using the same substrate.

The ligation assay we designed to be carried out in volumes that could be accommodated in a 96-well format to allow development of the technology as a rapid, robotic-based high throughput assay. After completion of the ligation step the DNA was added to a 96-well filtration plate containing Q-sepharose resin. The reaction mixture contained a mixture of ligated 36-mer product, unligated 6-mer substrate and annealed (30 + 6)-mer DNA substrate, all of which have different binding affinity for the resin. The resin was washed with warm NaCl buffer and the eluate collected in microtitre plates. With the vacuum apparatus available to us nitrocellulose membrane could not be kept in intimate contact with the filtration plate, causing a measure of cross contamination between the wells. Since the volume of wash buffer used was more than the retaining capacity of the membrane, the eluates were freeze-dried and later spotted onto individual nitrocellulose membranes.

A test ligation reaction was performed with T4 DNA ligase followed by PBCV-1 DNA ligase and HuLigI with our synthetic oligonucleotides as a substrate in the presence of doxorubicin, an anti-cancer drug and a known inhibitor of ATP-dependent ligases. Our data supported the findings (Ciarrocchi *et al.*, 1999) that doxorubicin inhibits ATP-dependent ligases. Significantly the sensitivity to doxorubicin determined by our assay was comparable to that determined by the previous study. Thus our assay has been demonstrated to be as reliable as conventional electrophoresis-based methods in identifying DNA ligase inhibitors. The method has been adapted to analyse multiple reactions in a 96-well plate format. This will allow for a high throughput version of this assay to be used that will be faster than the conventional gel-based analysis and suitable for high volume screens.

The assay procedure was tested using a collection of whole cell extracts obtained as part of a natural product library from *Ganoderma* species. Several *Ganoderma* species have been described having antibacterial activities for example; basidiocarp extracts of *G. lucidum* and *G. recinaceum* inhibit the growth of *Bacillus subtilis* and show additive effects on the activity of other antibiotics (Russel & Patterson, 2006). Culture

extracts of *G. recinaceum* can also inhibit *Staphylococcus aureus* (Russel & Patterson, 2006). Some of the biologically active compounds isolated from *Ganoderma* species have anti-tumour properties such as the triterpenes from *G .lucidum* that suppress growth and invasive behaviour of cancer cells (Silva, 2006). Aqueous extracts from *G. tsugae* inhibit cell proliferation in human breast cancer cell lines MCF-7 and MDA MB-231 without exerting any significant cytotoxic effects on normal human mammary epithelial cells (Yue *et al.*, 2006).

Recently, Lee *et al.* (2009) reported on the anti-tumour effect of another white rot fungus, *Inonotus obliquus***,** which inhibited the growth of HT-29**,** human colon cancer cells, by inducing apoptosis. It is believed that this apoptosis was as a result of the inactivation of the DNA repair protein, poly(ADP-ribose) polymerase (PARP; Lee *et al.*, 2009). Mizushina *et al.* (1998) isolated cerebrosides such as glycosphingolipids from the fruiting body of *G. lucidum* which inhibited eukaryotic DNA polymerase in a dose dependent manner, suggesting a possible chemotherapeutic role through the inhibition of DNA replication.

Screening with the crude *Ganoderma* extracts revealed inhibitory compounds present in many of the natural product pools. Further dilution analysis of these pools would have to be used to confirm whether the inhibition is due to ligase specific inhibitors rather than protein denaturants or toxic elements present in the crude extracts. This was not possible to undertake in this study because of recurrent and then long-term loss of the departmental phosphorimager.

Once a hit/lead compound pool has been identified from the natural product library the individual components of the pool can be separated by chromatography and their structures determined using gas chromatography/mass spectroscopy (GC/MS). The compounds may then be screened, using a computer-aided drug design (CADD) method to identify those with a high probability of binding to the target site in the chosen ligase.

The exact nature of the target in the case of the ligase is multifactorial. The assay can be used to identify inhibitors that inhibit both substrate recognition and the second

and third steps of ligase catalysis. But will in its current format have limited use in identifying inhibitors of the first, adenylation step of the DNA ligation reaction. This is because a mixed preparation of DNA ligase is used in the assay; a portion of the ligase enzyme is already adenylated in the cell, i.e. AMP had bound to the enzyme before purification and this form of the enzyme will be resistant to inhibition affecting this step (Odell *et al.*, 2000). One means of circumventing this problem would be through the employment of a deadenylation step. Incubation of adenylated ligase enzymes with sonicated salmon sperm DNA in the presence of a divalent metal provides numerous substrate molecules that exhaust the adenylated enzyme molecules that can then be purified using an affinity/ion exchange resin. In order that inhibition of catalysis can be separated from substrate recognition we propose that a surface plasmon resonance (SPR) step should be incorporated. Work in this laboratory (Figure 3.27; Brooke, PhD thesis, 2007) has shown that ligase binding to DNA can be directly monitored using SPR. Once inhibition is identified the pool/component can be added into the binding step of purified enzyme under SPR analysis.

The interest in DNA ligase inhibitors has increased since the beginning of our study particularly in those compounds that perturb normal DNA substrate recognition. It is at this step that the behaviour of the enzymes appears to be most different and thus where a selective inhibitor may be best targetted. The catalytic core of ligases is very strongly conserved but their substrate recognition is clearly different (Odell *et al.*, 2000; Pascal *et al.*, 2004 and this study - Figure 3.27). From the structure of HuLigI in complex with DNA, Chen *et al.* (2008) used a computer-aided drug design (CADD) method to screen small molecules as potential inhibitors of human DNA ligases. Using this ligand-based *in silico* screening technique they identified three inhibitors, L67, L82 and L189 (Figure 5.10), specific to human DNA ligases. These inhibitors inhibit cell proliferation and at sub-toxic concentrations, specifically potentiate the killing of cancer cells by DNA-damaging agents.

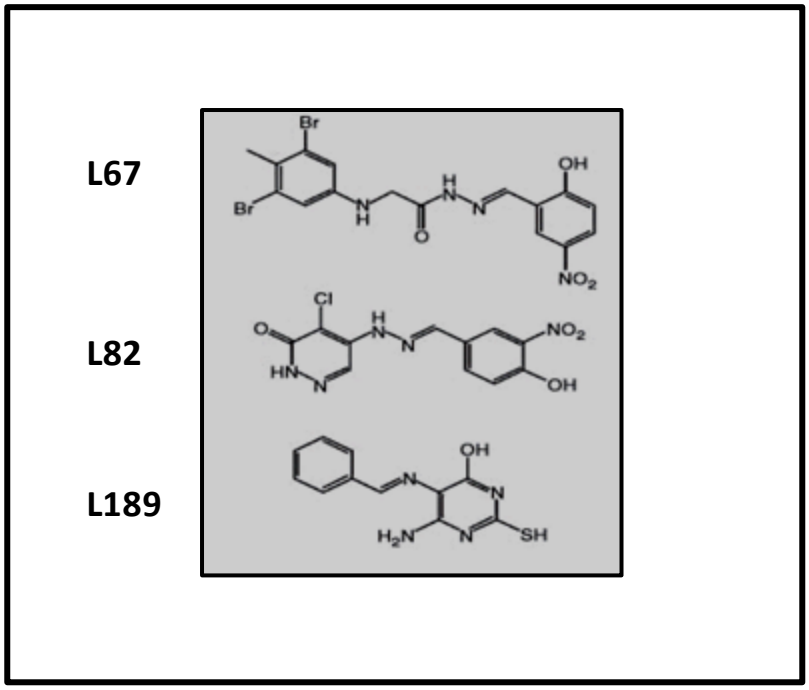


Figure 5.9: Chemical structure of HuLigI inhibitors identified by CADD

Chen *et al.* (2008) used CADD to identify human DNA ligase I inhibitors based on the structure of HuLigI complexed with nicked DNA and identified three inhibitors that could complex with the DNA binding domain of HuLigI and inhibit the DNA ligation reaction. L189 inhibits step 2 whereas L67 and L82 inhibits step 3 of the ligation reaction (adapted from Chen *et al.,* 2008).

The limitation of the CADD approach is that it can only screen libraries of known chemicals, where the structure of all the compounds to be tested has been established, and not natural product pools. It is our belief that the CADD will be more efficient if used in conjunction with our rapid assay system. Once the hit/lead compounds are identified after screening of natural product libraries CADD could help in studying interactions of the identified lead compounds with DNA ligases to suggest synthetic modifications to increase their potency.

This rapid assay method once combined with efficient vacuum methodologies will be able to analyse multiple reactions in a 96-well plate much faster than conventional gel based analysis. Since this method does not include additional steps, involving other enzymatic treatment**s** or signal amplification using other enzymatic methods, it eliminates some of the indirect assay errors obtained from false positive or negative data resulting from. Due to time constraints, further studies could not be carried to improve washing protocols which would make the assay more efficient than the current method in quantitatively analysing DNA ligase activity. Use of pure chemical

or natural product libraries, instead of crude extracts, will help in the rapid identification of inhibitor molecules. Further optimisation of resin quantity, wash buffer volumes and retention capacity of DNA affinity membranes to be used will help in screening larger numbers of compounds. Undoubtedly, the utility of the compounds identified will have to ultimately be judged against whole cells and tissues rather than the isolated enzyme to determine their specificity in treating cancers and their effectiveness in combinational therapies.

CHAPTER 6

CONCLUSIONS, GENERAL DISCUSSION AND FUTURE WORK

CHAPTER 6: CONCLUSIONS, GENRAL DISCUSSION AND FUTURE WORK

6.1 MAIN CONCLUSIONS

6.1.1. Identification and characterisation of a DNA ligase from *Trichomonas vaginalis*

DNA ligases have been studied in depth from various eukaryotic and prokaryotic organisms except protozoans. This study has identified**,** cloned, expressed and characterised the single DNA ligase from *Trichomonas vaginalis* (*T. vaginalis*), a pathogenic protozoan parasite.

Prior to beginning of the characterisation of *T. vaginalis* DNA ligase (TVlig), there were no reports of DNA ligases from protozoan protists. The sequence analysis of the *T. vaginalis* genome revealed an open reading frame (ORF) encoding a protein with a 51% and 53% homology to human DNA ligase I and Cdc9 of *Saccharomyces cerevisiae* respectively. Amino acid sequence alignment studies of TVlig with ATP-dependent DNA ligases from human, yeast and *Plasmodium falciparum* suggest the presence of a number of features in common with replicative DNA ligases: A PIP motif (PCNA interacting polypeptide); a nuclear localisation signal; a DNA binding domain N-terminal to the catalytic core, of adenylation and oligomer binding domains. The adenylation domain of TVlig, as with other members of nucleotidyltransferase superfamily members including DNA and RNA ligases, consists of six conserved motifs (I, Ia, III, IIIa, IV and V). The oligomer binding domain in TVlig possesses a seventh conserved amino acid motif VI as present in other ATP-dependent DNA ligases – for example those encoded by PBCV-1, T7 bacteriophage, Vaccinia virus, yeast and human. The spacing of the conserved motifs in TVlig is consistent with other previously reported DNA ligases. Further bioinformatic analyses were carried on the sequence of TVlig and its tertiary structure was predicted which superimposes well on the known structure of human DNA ligase

I bound to adenylated DNA, hence leading to the conclusion that identified gene is indeed a DNA ligase I.

The TVlig was successfully amplified and cloned from genomic DNA extracted from a *T. vaginalis* culture. The recombinant gene was expressed and the encoded protein was purified to near homogeneity from *Escherichia coli* cells. The purified polypeptide was used to study the biochemical properties of the TVlig.

In vitro nick-sealing experiments demonstrated that TVlig showed optimal ligation activity on a nicked DNA substrate in the presence of 1 mM ATP and (8- 20) mM $MgCl_2$ at 30-38^{o}C. The ligase showed optimal activity at pH 7-8 which is similar to that observed for DNA ligase from T4 bacteriophage (pH 7.2–7.8; Weiss *et al.,* 1968). Further, a mutant of TVlig (K338A) was prepared by replacing the putative active site lysine residue, located within a stretch of amino acids consistent with the ligase motif I (KYDGER), with alanine. The K338A mutant version of the TVlig lacked catalytic activity. This result strongly suggests that the lysine at position 338 is the catalytic nucleophile in TVlig and is involved in the formation of covalent ligase-adenylate intermediate during step 1 of reaction mechanism. The TVlig activity was shown to be dependent on a divalent cation. Significantly, in addition to magnesium and manganese, divalent calcium was also found to support ligation. The ligase also functions as a monomer which raises questions about how the *T. vaginalis* is able to seal double-strand breaks in DNA.

In summary, characterisation of the single DNA ligase encoded by the genome of *T. vaginalis* represents an important step towards understanding replication and repair mechanisms for maintaining genomic integrity in the pathogen, *T. vaginalis*.

6.1.2 Development of a high throughput assay to screen inhibitors for DNA ligases

DNA ligases have an essential role in DNA repair and the substantially higher level of DNA ligase I in proliferating cells makes this replicative ligase an attractive therapeutic target for the treatment of unregulated cell growth as in a cancerous state. Three classes of chemicals (pyridochromanones, glycosyl derivatives and aryl-amino

compounds) that inhibit DNA ligases of either bacteria or humans have been reported (Ciarrocchi *et al.,* 1999; Brotz-Oesterhelt *et al.,* 2003; Srivastva *et al.,* 2005) however a rapid, non-electrophoretic assay is not routinely used to isolate such chemicals. This study sought to develop a high throughput assay that could be used to screen inhibitors against various ATP dependent DNA ligases.

The rapid, non-electrophoretic assay method, based on the strength of interaction of the oligonucleotides with Q-sepharose (a strong anion exchanger), was developed to screen inhibitors from natural product pools as well as chemical libraries. The binding affinities to Q-sepharose resin of a nicked DNA substrate (created from a 30-mer hairpin oligonucleotide and complementary ^{32}P-labelled 6-mer oligonucleotide) and its sealed, ligated product (36-mer) were determined. It was concluded that the 36-mer ligated product binds more tightly to the resin as compared to unligated 6-mer oligonucleotide and annealed DNA substrate. Initial optimisation studies were performed with T4 DNA ligase, PBCV-1 DNA ligase and a catalytically active form of human DNA ligase I in the presence of doxorubicin, an anti-cancer drug and a known inhibitor of ATP-dependent ligases. Doxorubicin inhibition of ligation was analysed in parallel between a conventional electrophoretic assay and our labelled nick-sealing assay. This showed that the newly developed assay is as reliable as conventional electrophoresis-based methods in identifying potent DNA ligase inhibitors. The feasibility of the assay was tested in screening a collection of whole cell mass extracts, obtained from a natural product library from basidiomycetes, in 96-well format.

We conclude that our new non-electrophoretic assay method greatly improves the efficiency of the DNA ligase assay and can be employed in research aimed at identifying DNA ligase inhibitors for new therapeutic intervention against diseases.

6.2. GENERAL DISCUSSION AND FUTURE WORK

DNA ligases play an essential enzymatic role in the survival of cells by taking part in DNA transactions such as DNA replication, recombination and repair. The enzymes are divided into groups depending on their energy co-factor specificity - either NAD^+-dependent (eubacterial ligases and certain eukaryotic viruses and mimiviruses) or ATP-dependent (eukaryotic, viral and archaebacterial; Timson *et al.*, 2000; Ellenberger & Tomkinson, 2008). There is a structural and mechanistic commonality between the different groups of DNA ligases. The ligase basic catalytic core contains areas with conserved amino acid sequence and spacing (motifs I – V) and is composed of an N-terminal adenylation/nucleotide binding domain (AdD) and a C-terminal domain that adopts an oligonucleotide binding fold (OB-fold; Subramanya *et al.*, 1996). The smallest ATP-dependent enzymes (eubacterial and viral) are composed of only the basic catalytic core however the larger cellular enzymes have the catalytic core adorned with additional functional domains. The overall objectives of this work were to develop a high throughput assay that could be used to screen for inhibitors against ATP dependent DNA ligases from different origins and to characterise various target ligase polypeptides that might then be included in such an assay.

Many current non-surgical cancer treatments target the integrity of cellular DNA by directly (caused by ionising radiations) and/or indirectly damaging DNA (caused by chemotherapeutic drugs; Damia & D'Incalci, 2007). However, the anti-cancer effectiveness of these agents is limited by the activation of DNA-repair machineries in the target cell. Indeed, the enhancement of DNA repair is one of the crucial mechanisms governing intrinsic or acquired cellular resistance to DNA-damaging agents and defects in DNA-repair pathways result in hypersensitivity to these agents (Madhusudan & Hickson, 2005; Lord *et al.*, 2006). Thus, there is growing interest in the development of DNA-repair interference as an adjuvant to current therapies and approaches in the treatment of cancer.

DNA ligases represent an attractive therapeutic target because DNA ligation is the last step of almost all DNA repair pathways; except for direct repair pathways, for

example the repair of O6-alkyl-guanine occurs without generating DNA breakage (Martin, 2001). DNA ligase–deficient cell lines exhibit sensitivity to a wide range of DNA-damaging agents due to the inefficient repair of DNA damage in these cell lines (Vijayakumar *et al.*, 2009). Moreover, most proliferating cells including human cancer cells express high levels of DNA ligase I (HuLigI) when compared to normal forms of the same cells in resting and differentiated states (Montecucco *et al.*, 1992; Sun *et al.*, 2001). Chen *et al*. (2008) have recently used *in silico* docking of compounds into the HuLigI structure and identified three inhibitors that show differential inhibition of the three human DNA ligases; one compound in particular inhibits HuLigI alone. The inhibitors were also found to increase the sensitivity of cancerous cells to DNA damage (Chen *et al.*, 2008).

The initial approach undertaken in this study was to clone, express and purify full-length HuLigI enzyme that could be used in the inhibitor screen to identify lead compounds that might have applications in cancer treatment. Previous reports identified an elevated expression of HuLigI in proliferating cells (Sun *et al.*, 2001 & 2002), thus cDNA was isolated from both HT-29 cells (metastatic colorectal cancer cell line) and MDA MB-435 cells (a breast carcinoma cell line) and used to isolate a HuLigI clone. HuLigI is 2757 base pairs in length making it a relatively difficult target to amplify by polymerase chain reaction (PCR). Attempts to amplify full-length HuLigI from a single PCR reaction were unsuccessful. Thus, different sets of forward and reverse primers (Table 4.1) were designed to amplify the gene in separate N- and C- terminal specific reactions that could be later ligated together to form a full-length clone. Although this strategy failed to produce a full-length clone the successful amplification of the N-terminal portion of the gene suggested the cDNA libraries were viable (Figure 4.1). Finally, a full-length of the gene was successfully amplified from the pCMV-SPORT6 vector, harbouring a full-length cDNA clone of HuLigI (Figure 4.3, panel A). This ORF was then cloned into an expression vector pET-16b to allow us to produce recombinant polypeptide that could be used in the *in vitro* assay (Figure 4.4, panel B). The presence of a rare (for *E. coli*) arginine (AGG) codon at the second position to the start codon, methionine (ATG), we believe may have ultimately influenced the expression of the relatively large full-length polypeptide (102 kDa) in *E. coli* cells. Other authors have also reported difficulties in expressing full-length

HuLigI. For example, Brotz-Oesterhelt *et al*. (2003), in their study to identify inhibitors of the NAD^+-dependent DNA ligase from *E. coli* cells used a truncated version of HuLigI (250–919 amino acid). Similarly, due to the difficulty of isolating large quantities of active HuLigI, Ciarrocchi *et al.* (1999) used T4 DNA ligase to study the inhibition of ATP-dependent ligases. Kodama *et al.* (1991) suggested that the recombinant full-length HuLigI is susceptible to degradation by *E. coli* encoded proteases, which may be an additional problem with such a large polypeptide. We believe that the presence of low molecular weight species during our western blot analysis of full-length HuLigI may represent such of protein degradation (Figure 4.5). Teraoka *et al.* (1999) suggested the presence of an extra ACC (threonine) codon adjacent to the ATG (methionine) initiation codon prevents nonrestricted proteolysis and results in high levels of expression of full-length HuLigI in *E. coli* cells. We have successfully amplified the full-length HuLigI gene with an extra ACC codon (data not shown) which could be cloned into an expression vector and expressed for further analysis of the inhibitors identified with Δ232 HuLigI.

During the course of this work, the crystal structure of a truncated version of HuLigI (232-919; Δ232) was reported (Pascal *et al.,* 2004); with conditions for the expression and purification of this construct being fully understood. This protein was shown to be active in the ligation of DNA and furthermore was sufficiently stable in its interaction with DNA substrate to allow protein:DNA crystals to form. A plasmid bearing Δ232 HuLigI was obtained from Dr. T. Ellenberger's group and used to purify the truncated version of the protein to near homogeneity**,** after a two column purification over Nickel (affinity) and Q-sepharose (ion exchange) resins (Figure 4.6, panels A & B). Kodama *et al.* (1991) had previously demonstrated that truncated HuLigI (250-919 amino acids) represents the catalytic domain of the enzyme. They showed this polypeptide was capable of forming enzyme adenylate complex and joining nicked DNA substrate in their *in vitro* assays with same ability as the full-length polypeptide. This construct was also able to complement a DNA ligase mutant of *S. cerevisiae* indicating that the truncated HuLigI (250-919 amino acids) is sufficient for biological activity in yeast and is thus able to fold into a functional protein without the presence of the amino-terminal region of the intact protein (Kodama *et al.,* 1991). The availability of the crystal structure of HuLigI (232-919 amino acids) bound to

adenylated DNA further confirmed that the small truncated version of HuLigI is stable and a functional enzyme with sufficient catalytic domains (Pascal *et al.,* 2004). The catalytic fragment of HuLigI is thus composed of a DNA binding domain (232-532 amino acids) and catalytic core (533-919 amino acids) comprising adenylation domain and oligomer binding domain. Therefore, we decided to use this catalytically active, easily expressed and stable form of HuLigI to identify inhibitors in our rapid high throughput assay. Using this form of ligase allows the identification of inhibitors that bind either to the DNA-binding domain of HuLigI, thereby inhibiting DNA joining and also the catalytic pocket of the enzyme. Chen *et al.* (2008) showed that compounds inhibiting DNA recognition can be selective for individual human ligase enzymes. This may well be a more productive area to identify anti-human ligase inhibitors as the structure of the catalytic core in ligase enzymes is very highly conserved in respect of the three-dimensional location of conserved residues (Odell *et al.,* 1999).

The conventional method of screening for inhibitors of DNA ligases involves analysing reaction products by gel electrophoresis and is highly unsuitable for rapid processing of samples in a 96-well format. When this work was begun there were no reports of use of non-electrophoretic approaches for high throughput screening of DNA ligase inhibitors. This work aimed to develop a high throughput assay, based on the binding affinities of DNA oligonucleotides to Q-sepharose (an anion exchange resin), to screen chemical and natural product libraries for small molecule inhibitors against ATP-dependent DNA ligases from different origins.

Odell *et al.* (2003) showed that a 30-mer hairpin strand with a 10 nucleotide 5' single strand tail when annealed with an excess of a 6-mer oligonucleotide (Figure 5.1) complementary to the tail can be efficiently joined *in vitro* to yield a 36-mer product where the ^{32}P-label is incorporated into the backbone of the DNA strand. The assay method we have designed for DNA ligases incorporates this idea as a central part of a high-throughput screen for inhibitors. The different binding affinities of a 30-mer, 6-mer and 36-mer to Q-sepharose were thus determined by their elution profile with different salt concentrations. Initial optimisation studies, performed with T4 DNA ligase, PBCV-1 DNA ligase and the truncated, catalytically active form of HuLigI,

demonstrated that the newly developed assay is as reliable as conventional electrophoresis-based methods in identifying potent DNA ligase inhibitors.

The *in silico* docking approach, as adopted by Chen *et al.* (2008), to screen inhibitors for human DNA ligases, seems quite promising but it can only be used to screen synthetic libraries where the structure of the chemicals is already known. The rapid assay developed in this study can be used to screen natural product libraries, which provide greater structural diversity and will offer significant opportunities for finding novel low molecular weight lead compounds.

Our preliminary screen was performed with crude cell mass extracts derived from *Ganoderma* - a white rot fungus rich in bioactive compounds with potential to develop anti tumour and anti bacterial compounds (Russell & Patterson, 2006; Silva, 2006). We showed that the screen can be adapted for rapid processing of samples in a 96-well format. The source library used in this study, being crude in nature, contains a mixture of peptides, organic compounds, denaturants and toxic host (*Ganoderma*) proteins. Further refinement in washing protocols and serial dilutions of the crude sample used for screening will be required to lower the background signal to identify accurate hit or lead compounds. Once potentially anti-ligase compound pools have been identified, mass spectrometry can be employed to characterise the molecules contained therein. These can then be further dissected by appropriate chromatography to identify the bioactive component. In order that compounds affecting the catalytic and DNA binding steps can be separated Biacore can be used to show whether the component affects DNA binding. Thus far the anti-cancer compounds identified by Chen *et al.* (2008) have been molecules that interfere with DNA binding by the human ligases. Future work will then include co-crystallisation of identified lead compounds with DNA ligases to determine modes of binding and to facilitate medicinal chemistry lead optimization.

We also developed the inhibitor assay such that the screen can be used to identify potential anti-bacterial agents. There are reports of inhibitors being developed to target the essential bacterial NAD^+-dependent ligases, such as those encoded by *E. coli*, *S. aureus*, *M. tuberculosis* and *S. pneumoniae* (Ciarrocchi *et al.*, 1999; Brotz-

Oesterhelt *et al.*, 2003; Srivastava *et al.*, 2005; Meier *et al.*, 2008). The potential for antibiotics to target these ligases relies on the fact that these enzymes are vital for all bacteria due to their participation in DNA replication. Genetic studies of NAD^+-dependent DNA ligases from *B. subtilis, M. tuberculosis, S. typhimurium, E. coli, S. pneumoniae* and *S. aureus* have established the essential nature of these enzymes for the growth of the organisms (Park *et al.*, 1989; Petit & Ehrlich, 2000; Wilkinson & Bowater, 2001; Gong *et al.*, 2004; Meier *et al.*, 2008). Moreover, the differential cofactor requirement and structural differences between prokaryotic and eukaryotic representatives make them an attractive choice to be exploited as a potential antibacterial target (Ciarrocchi *et al.*, 1999; Brotz-Oesterhelt *et al.*, 2003; Srivastava *et al.*, 2005).

Of the three classes of NAD^+-ligase inhibitors identified so far; pyridochromanones and glycosyl derivatives (ureides and amines) appear to inhibit NAD^+-dependent DNA ligases by competing with NAD^+ binding, whereas the aryl-amino compounds shows non-competitive inhibition for the NAD^+ binding site (Figure 3.31; Ciarrocchi *et al.*, 1999; Brotz-Oesterhelt *et al.*, 2003; Srivastva *et al.*, 2005).

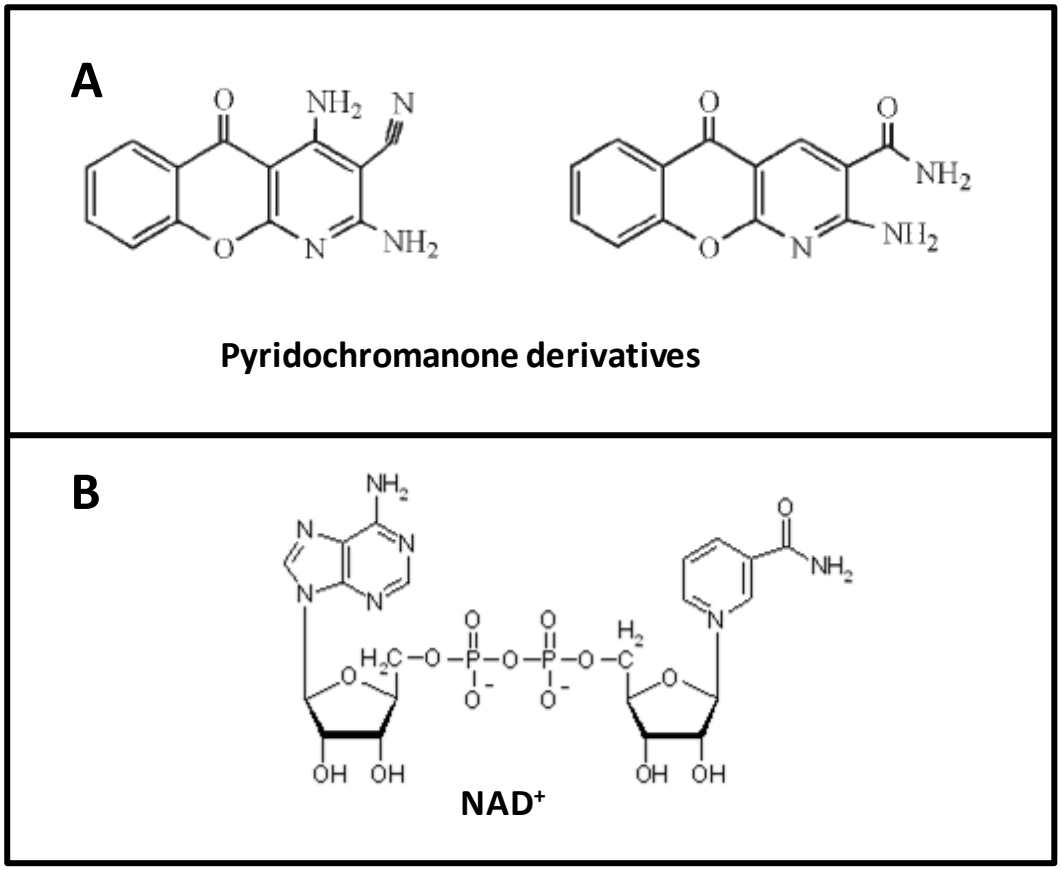


Figure 6.1: Chemical structures of inhibitors of NAD^+-dependent DNA ligases

Panel A shows the chemical structure of the compounds that bind competitively to the NAD^+ binding site and specifically inhibits the *E. coli* NAD^+-dependent DNA ligase. Panel B shows the chemical structure of NAD^+ (Dwivedi *et al.*, 2007).

Pyridochromanones and glycosyl derivatives specifically inhibit NAD^+-dependent DNA ligases, while not affecting ATP-dependent DNA ligases. The presence of additional ATP-dependent ligases in these pathogens however may contribute to the survival of cells treated with an NAD^+-dependent DNA ligase inhibitor and could be crucial in the subsequent development of antibacterial drug resistance to such new antibiotics.

In order that we could identify lead compounds that inhibit the action of ATP-dependent DNA ligases we elected to use PBCV-1 DNA ligase. PBCV-1 DNA ligase is one of the smallest characterised eukaryotic ATP-dependent DNA ligases. Many aspects of its structure and mechanism have been determined and at 34 kDa, consisting of only the basic ligase catalytic core, it represents a minimal, functional ATP-dependent DNA ligase (Ho *et al.*, 1997; Sriskanda & Shuman 1998, 2002; Odell

et al., 1999, 2000 & 2003; Nair *et al.*, 2007; Figure 1.8, panels A & B). In regard to this close catalytic resemblance at the structural level it is important to note that the DNA binding mode employed by the human and PBCV-1 ligases differs. The HuLigI employs a dedicated DNA binding domain (DBD) of nearly 200 amino acids that is absent in the smaller ATP-dependent ligases found in bacteria and exemplified by the PBCV-1 enzyme. Structural analyses of HuLigI and PBCV-1 ligases in complex with DNA show that the DBD and the PBCV-1 ligase latch module function co-operatively with the AdD and OB domains and fulfil a similar role, encircling and clamping the DNA substrate (Pascal *et al.* 2004; Nair *et al.* 2007). The PBCV-1 ligase latch module is thus critical for stable DNA binding by this enzyme. Protein sequence alignment of other small ATP-dependent DNA ligases encoded by pathogenic bacteria reveals that the latch module is not always present in an analogous location within the protein (Figure 4.9). In some polypeptides it is not clear if the latch module is present thereby raising interesting questions about the ability of these enzymes to stably bind DNA. However, the putative ATP-dependent DNA ligase from *Burkholderia pseudomallei*, a human and animal pathogen and causative agent of melioidosis, has a region homologous to the essential DNA binding surface loop ('latch') motif present in the PBCV-1 ligase. Although, *B. pseudomallei* possesses an NAD^+-dependent DNA ligase we believe inhibition of the ATP-dependent DNA ligase will have therapeutic value. The role of this loop will need to be studied in more detail to get an insight into the properties/DNA binding of these additional small ATP-dependent DNA ligases encoded by pathogenic bacteria. We believe that the inhibitors identified from a PBCV-1 DNA ligase screen will possess high specificity for the ATP-dependent DNA ligase of *B. pseudomallei.*

Whilst analysing/searching for other organisms where their DNA ligase might represent viable targets for therapeutic intervention, we observed that the recently annotated parasite *Trichomonas vaginalis*'s genome contained only one ligase. The identified open reading frame (ORF; XP_001581589), with conserved key nucleotidyltransferase motifs (I-VI), bears significant homology to the replicative DNA ligases of human, yeast and drosophila (Figure 3.3; Appendix 7.2, Figure 7.1). The order and spacing of conserved motifs in the putative ligase is similar, as is its sequence conservation, to that seen in other ATP-dependent ligase family members

(Figure 3.4; panel A). The sequence analysis of TVlig revealed an N-terminal NLS, a conserved PCNA interacting polypeptide (PIP) motif, DNA binding domain and a C-terminal catalytic core (Figure 3.4, panel B). The predicted tertiary structure of TVlig suggests close alignment with known structure of HuLigI bound to adenylated DNA (Figure 3.30; Pascal *et al.*, 2004). Our in-depth bioinformatics analysis of TV genome concluded that firstly the TV DNA ligase polypeptide represents a true homologue of HuLigI and secondly that its genetic complement is devoid of DNA ligase III and IV homologues and the accessory proteins (XRCC1 and XRCC4) required for their catalytic function. These findings mirrored those observed from analysis of the *Plasmodium falciparum* genome during the study of DNA ligase I from this organism (Buguliskis *et al.*, 2007; published contemporaneously with our study).

The inhibition of such a single replicative, DNA ligase I will profoundly affect the repair, replication and recombination of the TV organism. Disruption of such a biochemical link might suggest for a rational approach to stop replication and hence propagation of the parasite during infection. Moreover, the trichomonads adhere to vaginal epithelial cells, thereby raising the possibility of a therapeutic that can be administered topically without much effect on the host organism. In this regard it is important to note that the single DNA ligase of *P. falciparum* may in fact not be a good target. The parasite resides intra-cellularly within red blood cells of the host and disruption of the ligase through inhibition of the catalytic domain may cause coincidental inhibition of the human enzymes. It remains to be seen whether the mode of DNA binding by the parasite and the human are the same. The initial study reported here suggests that the TVlig does not bind to its substrate (nicked DNA) in the same manner as the human enzyme (Figures 3.26 & 3.27).

DNA repair mechanisms are aimed at preserving the genomic integrity of an organism. Our BLAST analysis suggests that the TV genome lacks core proteins of the non-homologous end joining pathway; a principal mechanism to repair double-strand DNA (dsDNA) breaks in eukaryotes. The apparent absence of DNA repair genes involved in double-strand break repairs raises questions about the ability of TV and other similar protozoans (*P. falciparum*) to repair such damages – do they utilise a unique, previously uncharacterised method or are they unable to repair dsDNA

breaks? Do they instead function in co-ordination with other enzymes, such as endonucleases and convert double strands to single strands? And even then how this is accurately joined to the other broken end and which enzyme do they use to align the DNA ends?

An absence of DNA repair genes has also been reported in *Encephalitozoon cuniculi*, a single celled parasitic eukaryote belonging to the group of microsporidia (Gill & Fast, 2007). The *E. cuniculi* genome has undergone massive genomic reduction resulting in a loss of genes from diverse biological pathways, including those that act in double strand DNA repair (Gill & Fast, 2007). This is clearly not the case in TV. The protein-coding complement of the TV genome numbers almost three times as many as the human genome and is the largest (approximately 60,000) of any eukaryotic genome sequenced to date (Carlton *et al.*, 2007).

Our literature survey suggests that the common feature in successful therapies employed against *T. vaginalis* and *P. falciparum* involves DNA damage of these parasites. For example, quinine (an anti-malarial drug) complexes with double-stranded DNA to prevent strand separation and blocks DNA replication (Ibezim & Odo, 2008). Similarly, the reduced (active) form of metronidazole (a drug used for the treatment of trichomoniasis) disrupts the helical structure of the DNA and prevents nucleic acid synthesis which eventually leads to cell death (Cudmore *et al.*, 2004; Sood & Kapil, 2008). However, these drugs do not seem to affect host (human) cell's DNA replication. We believe that the sensitivity of these parasites towards DNA damaging drugs may be as a result of their inability to repair such damage.

Based on previous reports of DNA damage repair in *T. brucei* and *P. falciparum* and their phylogenetic relationship with TV, I would speculate that the repair of single-strand breaks in TV DNA are performed by long-patch base excision repair (BER) pathway and double-strand repair is performed by the microhomology end joining method (MHEJ). Both of these pathways function with DNA ligase I as the terminal DNA sealing enzyme, as observed by *in vitro* experiments carried on mammalian cell lines and *T.brucei* and *P.falciparum* extracts (Haltiwanger *et al.*, 2000; Burton *et al.*, 2007; Casta *et al.*, 2008; Liang *et al.*, 2008).

It remains to be seen what enzymes TV can utilise for such repair. One method to isolate any interacting partners might be to utilise the recombinant TVlig. Larger cellular ATP-dependent DNA ligases such as DNA ligase I, III and IV do not function in isolation, but interact with other proteins that target them to lesions requiring sealing (Tomkinson *et al.*, 2006; Ellenberger & Tomkinson, 2008). Surface Plasmon Resonance (SPR), using Biacore, could be applied to assess the effect of protein interactions by analysing TVlig DNA binding in the presence of TV protein extracts. SPR is a solution study which facilitates the direct measurement of real-time interactions, even if these interactions do not result in the formation of stable complexes. SPR further allows the investigation of the dynamic nature of the ligase DNA interaction making it possible to examine the importance of the contribution of individual residues and different modules implicated in DNA binding. Leppard *et al.* (2003) used Biacore to assess the binding of Human DNA ligase III on nicked DNA duplex. However the concentration of DNA ligase that they used to analyse binding (20,000 RU) was far in excess of the 200 RU that is recommended for such studies. This may have led to mass transfer effects that may have represented non-specific binding. Certainly their ligase disengaged from DNA as soon as the enzyme was no longer being passed over the substrate. This is in stark contrast to the more stable forms of binding that we have observed using Biacore analysis (Figures 3.26 & 3.27). This technique has also been used to study DNA binding by PBCV-1 DNA ligase (Brooke, PhD Thesis, 2007).

The presence of a PIP motif in the TVlig made us search for a PCNA homologue in the TV genome. PCNA is a well conserved protein found in all eukaryotic species from yeast to humans (Stoimenov & Helleday, 2009). It belongs to the family of DNA sliding clamps (β clamps of *E. coli*) and is a toroidal molecule essential for processive DNA synthesis in Archaea and Eukaryotes. A BLAST search of the TV genome using PCNA gene sequences from human, yeast and drosophila identified an ORF encoding a putative PCNA polypeptide (XP_001329442). The ORF of TvPCNA when aligned with human PCNA shows 42% identity and 62% similarity (Figure 3.32).

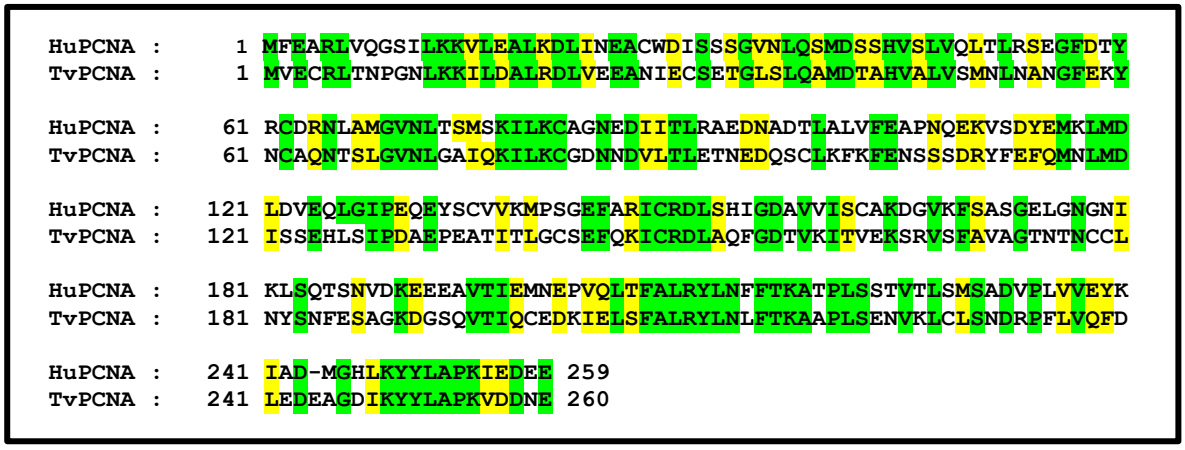


Figure 6.2: Sequence alignment of putative TvPCNA and HuPCNA

A BLAST search of TV genome using amino acid sequence of human PCNA gene (HuPCNA; NP_002583.1) revealed the presence of a putative TvPCNA (XP_001329442). Conserved residues are shown in green and identical residues are shown in yellow. The alignment was performed with ClustalW (Chenna *et al.*, 2003).

The putative TvPCNA could be characterised and its binding with TVlig could then be assessed using Biacore. Such an analysis may help to unravel the strange DNA binding we have observed by Biacore of the TVlig (Figures 3.26 & 3.27). Unlike the PBCV-1 enzyme the TVlig shows little discrimination for the gapped versus duplex DNA in its binding step (Figure 3.26). It bound very rapidly to all types of DNA substrate, nick, gap and duplex (the on-rate was too fast to measure accurately) however it showed rapid dissociation from both the duplex 'product' and the gapped 'non-productive substrate' DNA's. The binding to the true ligase substrate – nicked DNA – was unusual. At all concentrations the amount of enzyme bound (reflected by the total response units; RU) was half that of the duplex DNA substrate, the product of the ligation reaction. Previously this discrimination has been observed for the PBCV-1 enzyme where the conclusion was drawn that if the nick is centrally placed then specific nick recognition would restrict further enzyme binding (Figure 3.27, panel B).

At the lower enzyme concentrations (0.1 - 0.5 µM) studied we observed a concentration dependent binding and subsequent slow release from the DNA. At higher concentrations the enzyme showed a biphasic binding; the initial binding was weaker than at lower enzyme concentrations but this altered when the ligase enzyme was no longer passed over the surface of the chip. A mechanism has yet to be

developed to explain the observations. We initially wondered if the enzyme was oligomerising at higher concentrations (1-3 μM) in a manner that might restrict DNA binding. When the enzyme was no longer being added to the chip surface the concentration would drop rapidly potentially allowing the enzyme to bind to the DNA.When we analysed the behaviour of the TVlig at similar concentrations by sucrose density gradient it behaved as a monomer (Figure 3.29). This suggests that the altered nick binding profile does not reflect a simple oligomerisation of the protein at higher concentrations. In the future it will be desirable to extend the assay with TV extracts or recombinant proteins to study the interaction of DNA ligase I with PCNA during long-patch DNA repair, thereby providing valuable information about the DNA repair processes of the parasite. It would also be informative to compare the DNA damage repair mechanism of TV with that of *Trypanosoma brucei* using cell free extracts as described by Conway *et al.* (2000a & 2000b). Future work will also include determining the 3-D structure of TVlig alone or in complex with a DNA ligand by X-ray crystallography. The determination of the structure of this enzyme either bound to its partner proteins or ligand could help in understanding how the ligase functions. It will also provide a means whereby computer aided drug design (CADD) can be utilised to identify molecules that will specifically inhibit TVlig and not the host cellular ligases.

This work has developed a novel assay to screen for inhibitors of ATP-dependent DNA ligases which can be adapted for commercial usage in screening libraries of small molecules. This study has also enhanced the understanding of DNA ligase and repair mechanisms in parasitic protozoa, particularly the Trichomonad *T. vaginalis*.

CHAPTER 7

APPENDICES

CHAPTER 7: APPENDICES

7.1 STRAINS OF *E. coli* AND PLASMIDS USED

Table 7.1: Genotypes of *E. coli* used in this thesis

E.coli strains	Genotype
DH5α (Invitrogen)	F^- ϕ80d*lac*Z Δ M15 Δ (*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_k^-, m_k^+) *pho*A *sup*E44 $^-$ λ*thi*-1 *gyr*A96 *rel*A1
BL21 (DE3) (Novagen)	F– *ompT hsdS*B(rB– mB–) *gal dcm* (DE3
JM109 (Promega)	endA1 glnV44 thi-1 relA1 gyrA96 recA1 $mcrB^+$ Δ(lac-proAB) e14- [F' traD36 $proAB^+$ $lacI^q$ lacZΔM15] hsdR17($r_K^-m_K^+$)

Table 7.2: Plasmids used in this thesis

Plasmid	Function	Gene encoded	Other
pET-16b	T7 promoter-based expression vector. N-terminal 6xHis tag, Factor Xa site, 3 cloning sites (*Nde* I, *Xho* I and *Bam* HI), *lacI* coding sequence, pBR322 origin, *bla* coding sequence	PBCV-1 DNA ligase; TV DNA ligase and K338A mutant; HuLigI (full-length)	Original plasmid – Novagen pET-16b-latch Kind gift from Prof. S. Shuman (Sloan Kettering Institute, New York, USA).
pET-28a	T7 promoter-based expression vector. N-terminal 6xHis tag, C-terminal 6xHis tag, thrombin tag, T7 tag, multiple cloning sites (11: *Bam* HI – *Xho* I), *lacI* coding sequence, pBR322 origin, f1 origin, Kan coding sequence.	HuLigI-Δ232	Original plasmid (Novagen) pET28a- HuLigI-Δ232, Kind gift from Prof. T. Ellenberger (Washinton University, Washington, USA)
pGEM-T		Cloning of PCR products	Promega (USA)

7.2 SEQUENCE ALIGNMENTS GENERATED DURING THE COURSE OF THE THESIS

7.2.1 Sequence alignment of ATP-dependent DNA ligase I from *T. vaginalis* and its orthologues from other organism.

```
                                                                  I
S.pombe              EHG-LGTLRETCKLTPGIPTKPMLAKPTKQISEVLNTFDQAAFTCEYKYDGERAQVHFTE 428
S.cerevisae          EHG-IMNLDKYCTLRPGIPLKPMLAKPTKAINEVLDRFQGETFTSEYKYDGERAQVHLLN 431
D.melanogaster       KYD-IKELQERCPMHPGMPLRPMLAQPTKGVHEVFERFGGMQITCEWKYDGERAQIHRNE 407
A.gambiae            EHG-VQRLSEHCPMEPGTPLKPMLAHPTKGVQEVLQRFDGIDFTCEWKYDGERAQIHLLA 322
H.Sapiens            EHG-LERLPEHCKLSPGIPLKPMLAHPTRGISEVLKRFEEAAFTCEYKYDGQRAQIHALE 580
A.thaliana           SGG-VWNLPKTCNFTLGVPIGPMLAKPTKGVAEILNKFQDIVFTCEYKYDGERAQIHFME 456
P.knowlesi           NGDDMNTLSKKCTVKAGLPVQPMLAKPTKGIQEVLDRFNNVTFTCEYKYDGERAQIHYID 587
P.falciparum         NGDDMNTLSKKCTVKTGIPVQPMLAKPTKGVQEVLDRFNNVTFTCEYKYDGERAQIHYID 574
T.vag                RGS-FKGAEKNINIKVGIPAMPMLAKPAKKLDEIKRRLGDGRITGEYKYDGERAQIHKNK 351
P.furiosus           LEG--NEGLAKVQVQLGKPIKPMLAQQAASIRDALLEMGG-EAEFEIKYDGARVQVHK-D 260
                                    :       : :
                            Ia                          III
S.pombe              DGKFYVFSRNSENMSVRYPDISVSVSKWKKPDARSFILDCEAVGWDRDENKILPFQKLAT 488
S.cerevisae          DGTMRIYSRNGENMTERYPEINITDFIQDLDTTKNLILDCEAVAWDKDQGKILPFQVLST 491
D.melanogaster       KGEISIFSRNSENNTAKYPDLIARSTALLKGDVKSYIIDSEIVAWDVERKQILPFQVLST 467
A.gambiae            DGSVQIFSRNQENNTSKYPDVIARLEFTRTEPVASAILDCEAVAWDTDKRQILPFQVLST 382
H.Sapiens            GGEVKIFSRNQEDNTGKYPDIISRIPKIKLPSVTSFILDTEAVAWDREKKQIQPFQVLTT 640
A.thaliana           DGTFEIYSRNAERNTGKYPDVALALSRLKKPSVKSFILDCEVVAFDREKKKILPFQILST 516
P.knowlesi           KDNIKIFSRNLETMTEKYPDVIQIVRDQIISGATECIIDSEVVAYDIENKKILPFQVLTT 647
P.falciparum         KDNIKIFSRNLETMTEKYPDVIQIIKDQIGENVKECIIDSEVVAYDIVNNKILPFQVLTT 634
T.vag                EGKVRIFSRSAEDSTAKFADVADIITDNVKG--EQYILDSEIVAYDPEKHVILPFQTLMN 409
P.furiosus           GSKIIVYSRRLENVTRAIPEIVEALKEAIIP--EKAIVEGELVAIG-ENGRPLPFQYVLR 317
                            .:          :                           .      : .
                                      IIIa
S.pombe              --RKRKDVKIGDIKVRACLFAFDILYLNGQPLLETPLNERRKLLYSMFQPSTGDFTFAKH 546
S.cerevisae          --RKRKDVELNDVKVKVCLFAFDILCYNDERLINKSLKERREYLTKVTKVVPGEFQYATQ 549
D.melanogaster       --RKRKNVDIEEIKVQVCVYIFDLLYINGTALVTKNLSERRKLLLEHFQEVEGEWKFATA 525
A.gambiae            --RKRKDANEADIKVQVCVFMFDLLYLNGEPLVERPFVERRELLYRHFREIEGQWQYATR 440
H.Sapiens            --RKRKEVDASEIQVQVCLYAFDLIYLNGESLVREPLSRRRQLLRENFVETEGEFVFATS 698
A.thaliana           --RARKNVNVNDIKVGVCIFAFDMLYLNGQQLIQENLKIRREKLYESFEEDPGYFQFATA 574
P.knowlesi           --RKRKDVDIENIKVKICLFPFDLICCNGVPVIKEPLEIRRKLLYSLLKCKEGVLCYATH 705
P.falciparum         --RKRKDVDIENIKVKVCLFPFDLICCNGIPVIKEPLAVRRKLLYSLLKSKDGVLSYATH 692
T.vag                --RPKKGTDTP-SAIHVCVCAFDLLSYNGVSMLDEPLEKRRDKLHEIVTVVPTKLQLATY 466
P.furiosus           RFRRKHNIEEMMEKIPLELNLFDVLYVDGQSLIDTKFIDRRRTLEEIIKQNE-KIKVAEN 376
                                    :  :  :  .       :
                                      IV                          V
S.pombe              SDQKSIESIEEFLEESVKDSCEGLMVKMLEGPDSHYEPSKRSRHWLKVKKDYLSGVG--- 603
S.cerevisae          ITTNNLDELQKFLDESVNHSCEGLMVKMLEGPESHYEPSKRSRNWLKLKKDYLEGVG--- 606
D.melanogaster       LDTNDIDEVQQFLEESIKGNCEGLMVKTLD-EEATYEIAKRSRNWLKLKKDYLSNVG--- 581
A.gambiae            LDTGDLDELQRFLDEAVRGNCEGLMVKTLA-REATYEIAKRSRNWLKLKKDYLTGVG--- 496
H.Sapiens            LDTKDIEQIAEFLEQSVKDSCEGLMVKTLD-VDATYEIAKRSHNWLKLKKDYLDGVG--- 754
A.thaliana           VTSNDIDEIQKFLDASVDVGCEGLIIKTLD-SDATYEPAKRSNNWLKLKKDYMDSIG--- 630
P.knowlesi           SEMNNVEDMDIFLQDAIENNCEGLMVKTLL-DNASYEPSRRSLNWLKVKKDYIEGLS--- 761
P.falciparum         SEMNNIEDIDMFLQDAIENNCEGLMVKTLV-ENASYEPSRRSLNWLKVKKDYVEGLS--- 748
T.vag                KNASKIEDLGDFFEEAVSNRTEGLMVKTLN---GRYECGKRAMSWAKLKKDYISKEVFKG 523
P.furiosus           LITKKVEEAEAFYKRALEMGHEGLMAKRLD---AVYEPGNRGKKWLKIKPT--------- 424
                            .      :
S.pombe              --------DSLDLIVIGAYYGKGKRTSVYGAFLLGCYDPDTETVQSICKLGTGFSEEHLE 655
S.cerevisae          --------DSLDLCVLGAYYGRGKRTGTYGGFLLGCYNQDTGEFETCCKIGTGFSDEMLQ 658
D.melanogaster       --------DSLDLVVIGGYKGKGRRTGTYGGFLLACYDTENEEYQSICKIGTGFTDEDLQ 633
A.gambiae            --------DSLDLVVIGGYRGRGKRTGTYGGFLLACYDEENEEYQTICKIGTGFSDDDLQ 548
H.Sapiens            --------DTLDLVVIGAYLGRGKRAGRYGGFLLASYDEDSEELQAICKLGTGFSDEELE 806
A.thaliana           --------DSVDLVPIAAFHGRGKRTGVYGAFLLACYDVDKEEFQSICKIGTGFSDAMLD 682
P.knowlesi           --------DSVDLVPIAGYYGKGKRSGVYGAFVLATYNSETENFQTVCKAGTGFSDEILG 813
P.falciparum         --------DSVDLVPIAGYYGKGKRSGVYGAFVLAAYNSETENFQTVCKAGTGFSDEILS 800
T.vag                GVAEEIPPDTIDVVVIGATYGKGKRTSVYGAFLVGVYNNVTGKVESLCFVGTGFKEDDLK 583
P.furiosus           -------MENLDLVIIGAEWGEGRRAHLFGSFILGAYDPETGEFLEVGKVGSGFTDDDLV 477
                                                                  * *
```

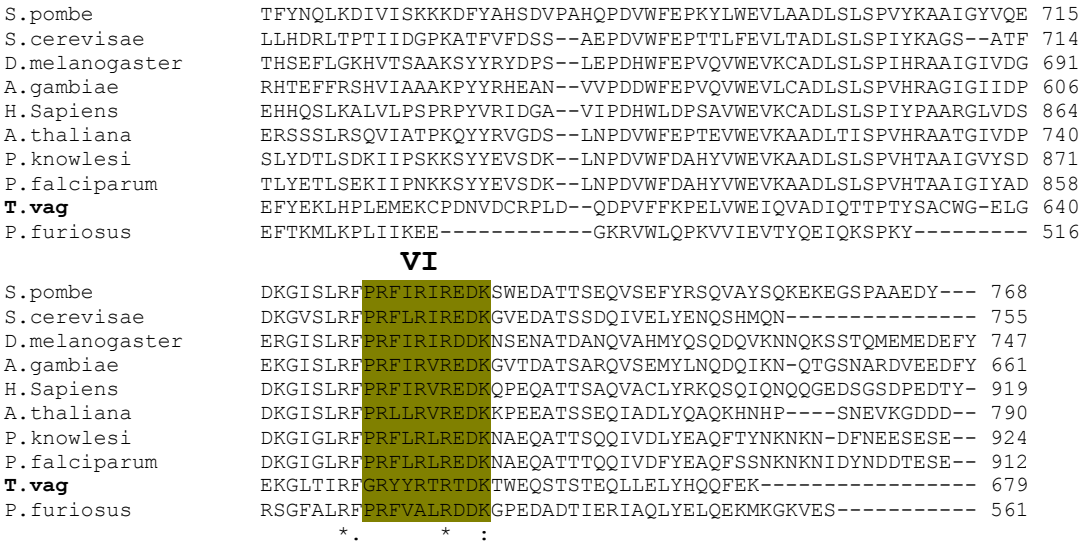


Figure 7.1: Amino acid sequence alignment of TVlig with other ATP-dependent DNA ligases.

Sequence alignment (from motif I) of a selection of ATP-dependent DNA ligases with amino acid sequence of TVlig (Tvag; obtained from TIGR database; 85859.m00330. The accession number of the DNA ligases used to generate above alignment are as follows:
Saccharomyces pombe (S.pombe; NP_593318.1*), Saccharomyces cerevisiae (S. cerevisae;* NP_010117.1*), Drosophila melanogster (D.melanogster;* NP_611843.2*), Anopheles gambiae (A.gambiae;* NP_319999.1*), Homo sapiens (H. sapiens;* NP_000225.1*), Arabidopsis thaliana (A.thaliana;* NP_175351.1*), Plasmodium knowlesi (P.knowlesi;* XP_00226193*), Plasmodium falciparum (P. falciparum;* AAL59668*), Pyrococcus furiosus (P.furiosus;* NP_579364*).* The alignment was performed with ClustalW (Chenna *et al.*, 2003).

7.2.2 Alignment of TVlig with DNA ligases III & IV

```
HuLigIV           KD-MKHQSFYIETKLDGERMQMHKD--GDVYKYFSRNGYNYTDQFGASPTEGSLTPFIHN 317
DnL4              ICRTLHDDFLVEEKMDGERIQVHYMNYGESIKFFSRRGIDYTYLYGASLSSGTISQHLR- 327
TVlig             TG--VYKEFLKMAGFQG--------------ANSDRNRTQTIQKLLTNCKPEEVKCIVR- 225
HuLigIIIalpha     DGFHGDVYLTVKLLLPGVIKTVYNLNDKQIVKLFSRIFNCNPDDMARDLEQGDVSETIRV 356
                         :     : *                .*           .     :.  ::

HuLigIV           AFKADIQICILDGEMMAYNPNTQTFMQKGTKFD----IK-RMVE------DSDLQTCYCV 366
DnL4              -FTDSVKECVLDGEMVTFDAKRRVILPFGLVKG----SA-KEALSFNSINNVDFHPLYMV 381
TVlig             ---------MLECKLRVGAAHESFLFALGMAFRRREYIMG-------------------- 256
HuLigIIIalpha     FFEQSKSFPPAAKSLLTIQEVDEFLLRLSKLTK--EDEQQQALQDIASRCTANDLKCIIR 414
                               .: .       ::  .

HuLigIV           FDVLMVNNKKLGHETLRKRYEILSSIFTPIPGRIEIVQKTQAHTKNEVIDALNEAIDKRE 426
DnL4              FDLLYLNGTSLTPLPLHQRKQYLNSILSPLKNIVEIVRSSRCYGVESIKKSLEVAISLGS 441
TVlig             -----MIKNNPDEEDLKDHGHLFREIYYRYPIIDEILVEFLRGSFKGAEKNINIKVGIPA 311
HuLigIIIalpha     LIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLQDVVERVLHNAQEVEKEPGQRRALSVQ 474
                       :  ..     *        .         :::        :   .  :

HuLigIV           EGIMVKQPLSIYKPDKRGEGWLKIKPEYVSGLMDELDILIVGGYWGKGSRGGMMSHFLCA 486
DnL4              EGVVLKYYNSSYNVASRNNNWIKVKPEYLEEFGENLDLIVIGRDSGKKDSFMLGLLVLDE 501
TVlig             MPMLAKPAKKLDEIKRRLG---------DGRITGEYKYDGERAQIHKNKEGKVRIFSRSA 362
HuLigIIIalpha     ASLMTPVQPMLAEACKSVEYAMKKCP---NGMFSEIKYDGERVQVHKNGDHFSYFSRSLK 531
                    ::        :                 :  : .         *

HuLigIV           VAEKPPPGEKPSVFHTLSR-------------------VGSGCTMKELYDLGLKLAKYWK 527
DnL4              EEYKKHQGDSSEIVDHSSQEKHIQNSRRRVKKILSFCSIANGISQEEFKEIDRKTRGHWK 561
TVlig             EDSTAKFADVADIITDNVK--------------------------GEQYILDSEIVAYDP 396
HuLigIIIalpha     PVLPHKVAHFKDYIPQAFPG-------------------------GHSMILDSEVLLIDN 566
                         ..  . .                                .   :. :

HuLigIV           PFHRKAPPSSILCG-TEKPEVYIEPCNSVIVQIKAAEIVPSD-------MYKTGCTLRFP 579
DnL4              RTSEVAPPASILEFGSKIPAEWIDPSESIVLEIKSRSLDNTETNMQ---KYATNCTLYGG 618
TVlig             EKHVILPFQTLMNR----PKKGTDTPSAIHVCVCAFDLLSYN------------------ 434
HuLigIIIalpha     KTGKPLPFGTLGVH-KKAAFQDANVCLFVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIP 625
                        *  ::       .    :    :   :   .:   :

HuLigIV           RIEKIRDDKEWHECMTLDDLEQLRGKASGKLASKHLYIGGDDEPQEKKRKAAPKMKKVIG 639
DnL4              YCKRIRYDKEWTDCYTLNDLYESRTVKS------NPSYQAERSQLGLIRKKRKRVLISDS 672
TVlig             ------------GVSMLDEPLEKR-----------------RDKLHEIVTVVPTKLQLAT 465
HuLigIIIalpha     NRIMFSEMKRVTKALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKKDYLNEGA 685
                                  * :                              .


HuLigIV           IIEHLKAPNLTNVNKISNIFEDVEFCVMSGT--------------DSQPKPDLENRIAEF 685
DnL4              FHQNRKQLPISNIFAGLLFYVLSDYVTEDTG--------------IRITRAELEKTIVEH 718
TVlig             YKNASKIEDLGDFFEEAVSNRTEGLMVKTLN--------------GRYECGKRAMSWAKL 511
HuLigIIIalpha     MADTADLVVLGAFYGQGSKGGMMSIFLMGCYDPGSQKWCTVTKCAGGHDDATLARLQNEL 745
                    :  .   :  .                                              :

HuLigIV           GGYIVQNPGPDTYCV----IAGSENIRVKNIILSNKHDVVKPAWLLECFKTKSFVPWQPR 741
DnL4              GGKLIYNVILKRHSIGDVRLISCKTTTECKALIDRGYDILHPNWVLDCIAYKRLILIEPN 778
TVlig             KKDYIS-----------------KEVFKGGVAEEIPPDTIDVVVIGATYGKGKRTSVYGA 554
HuLigIIIalpha     DMVKISKDPSKIPSWLKVNKIYYPDFIVPDPKKAAVWEITGAEFSKSEAHTADGISIRFP 805
                      :                                 :

HuLigIV           FMIHMCPSTKEHFAREYDCYGDSYFIDTDLNQLKEVFSGIKNSNEQTPEEMASLIADLEY 801
DnL4              YCFNVSQKMRAVAEKRVDCLGDSFENDISETKLSSLYK-----SQLSLPPMGELEIDSEV 833
TVlig             FLVGVYNNVTGKVES--LCFVGTGFKEDDLKEFYEKLH---------PLEMEKCPDNVDC 603
HuLigIIIalpha     RCTRIRDDKDWKSATNLPQLKELYQLSKEKADFTVVAGDEGSSTTGGSSEENKGPSGSAV 865
                      :  .                 . .  .:                    .   .

HuLigIV           RYSWDCSPLSMFRRHT---------------VYLDSYAVINDLSTKNEGTRLAIKALELR 846
DnL4              RR----FPLFLFSNRI---------------AYVPRRKISTEDDIIEMKIKLFGGKITDQ 874
TVlig             RP---------LDQDP---------------VFFKP------------------------ 615
HuLigIIIalpha     SRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGEKLATKSSPVKVGEKRKAAD 925
                               .                .


HuLigIV           --FHGAKVVSCLAEGVSHVIIGEDHSRVADFKAFRRTFKRKFKILKESWVTDSIDK-CEL 903
DnL4              QSLCNLIIIPYTDPILRKDCMNEVHEKIKEQIKASDTIPKIARVVAPEWVDHSINENCQV 934
TVlig             ELVWEIQVADIQTTPTYSACWGELGEKGLTIRFGRYYRTRTDKTWEQSTSTEQLLELYHQ 675
HuLigIIIalpha     ETLCQTKVLLDIFTGVRLYLPPSTPDFSRLRRYFVAFDGDLVQEFDMTSATHVLGSRDKN 985
                    .    :             .  .               :       . : .  .
```

```
HuLigIV             QEENQYLI---------------- 911
DnL4                PEEDFPVVNY-------------- 944
TVlig               QFEK-------------------- 679
HuLigIIIalpha       PAAQQVSPEWIWACIRKRRLVAPC 1009
```

Figure 7.2: Sequence alignment of TVlig, Human DNA ligase III & IV and DnL4

Amino acid sequence alignment of *Trichomonas vaginalis* DNA ligase (TVlig) with Human DNA ligase IIIα (HuLigIII alpha; NP_039269.2), Human DNA ligase IV (HuLig4; NP_001091738.1) and DnL4 (orthologue of HuligIV in *Saccharomyces cerevisiae;* NP_014647.1) using ClustalW (Chenna *et al.*, 2003).

7.3 DETERMINATION OF K_M FOR TV DNA LIGASE

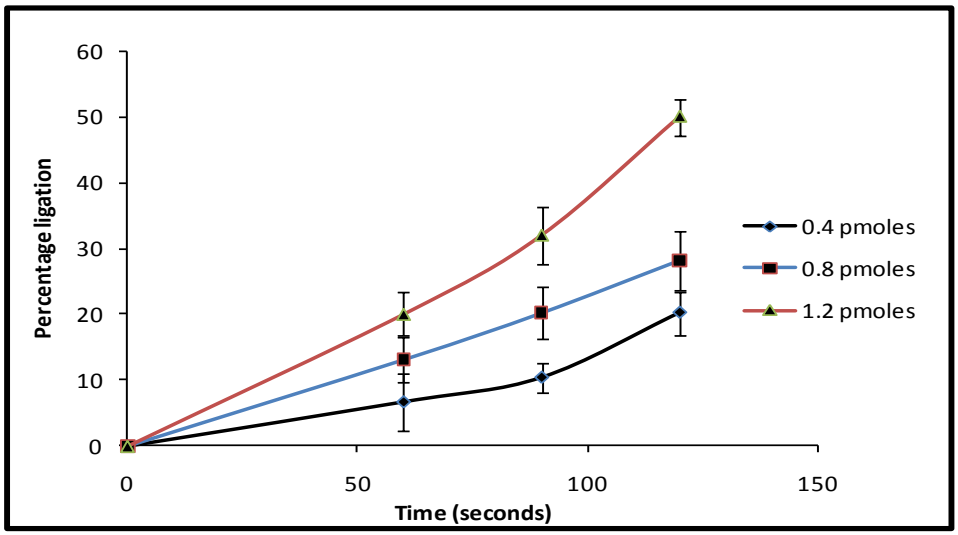


Figure 7.3: Effect of varying ATP concentration on strand joining activity of TVlig

Graph showig extent of ligation as a function of time when different concentrations of ATP (0.4 pmole to 1.2 pmole) were used to determine the binding affinity of the TVlig using 1 pmole of nicked labelled substrate and 0.1 pmole of the ligase. Each ligation reaction was performed in a reaction buffer containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 5 mM DTT and incubated at 37^{o}C. At each time-point, 10 µl of the sample was removed and the reaction was quenched by adding 20 µl loading buffer [98% (v/v) formamide, 10 mM EDTA, 0.01% (wt/vol) bromophenol blue] followed by incubation at 95^{o}C for 3 minutes. The reaction products were electrophoresed on a denaturing 20% polyacrylamide gel and scanned using a Phosphorimager (GE Healthcare).

\

7.4 BRADFORD STANDARD CURVE TO ESTIMATE PROTEIN CONCENTERATION

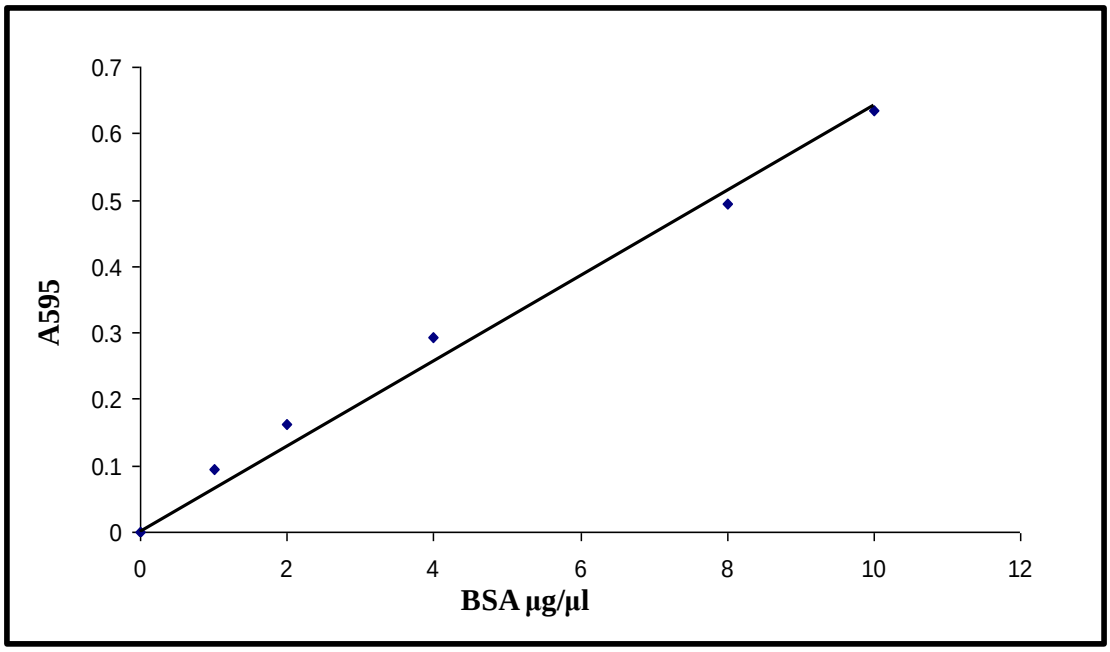


Figure 7.4: Sample standard curve for Bradford protein concentration assay

The protein used to determine this standard curve was 0 – 10 µg/µl BSA.
Bradford analysis provides a reasonable estimation of protein concentration within the linear portion of the curve. Spectrophotometric analysis of absolutely pure ligase and the employment of the PBCV-1 DNA ligase extinction co-efficient would be the ideal solution to this problem. In all instances the concentration of protein used in the DNA ligation assays was estimated in the Bradford linear range. As such all ligase preparations were referenced to these (or similar) values, thus the relative amounts of ligase are internally the same though they may not be accurately defined in molarity.

7.6 CONFERENCE PRESENTATIONS

7.6.1 Abstract and poster presented at Annual Biochemical Society Meeting held at Glasgow, Scotland, UK (July, 2006)

Characterisation of an ATP dependent DNA ligase from *Trichomonas vaginalis*

L.Kaur and M. Odell

Trichomonas vaginalis, a parasitic protozoan, is the etiological agent of trichomoniasis, a sexually transmitted disease (STD) of global concern leading to an estimated 180 million cases per annum. The 5-nitroimidazoles, especially metronidazole, are the most widely used antimicrobial agents for the treatment of trichomoniasis. The mechanism of the protozooicidal effect of the 5-nitroimidazoles is thought to involve generation of nitro radicals, which leads to subsequent DNA damage and cell death. However metronidazole resistant strains of *T. vaginalis* are on the rise, emphasising the need for research into alternative antibiotics and drug targets.

This laboratory is interested in the maintenance and metabolism of DNA in protozoa, particularly in *T. vaginalis.* DNA ligases are enzymes required to seal single-strand breaks in double-stranded DNA and maintain the integrity of an organisms' genome during various aspects of DNA metabolism, such as replication, excision repair and recombination. Most organisms' express at least one DNA ligase enzyme. In yeast two are elaborated, one responsible for replication the other intimately involved in DNA repair. Human cells encode at least three separate DNA ligase ORFs that are responsible for at least four distinct activities. These cellular DNA ligases utilise ATP as an energy co-factor. The analyses of DNA ligase 3-dimensional structures and their sequences have shown that these enzymes are built around a common catalytic core.

We have identified a single, hypothetical gene from *T. vaginalis* that shows similarity in sequence to cellular ATP-dependent DNA ligases, with the highest homology being to the replicative, DNA ligase I enzyme. Using PCR

we have successfully cloned, expressed and purified this enzyme from *E. coli* using Nickel-affinity and Ion exchange (S-sepharose) resins. The 78 kDa enzyme has been characterised confirming that it is indeed an ATP-dependent DNA ligase. We will present our detailed analysis of the biochemical and enzymatic capabilities of this enzyme that represents the first analysis of a protozoan DNA ligase.

UNIVERSITY OF WESTMINSTER

Cloning, Expression and Characterisation of an ATP dependent DNA ligase from *Trichomonas vaginalis*

L.Kaur and M.Odell

Department of Molecular and Applied Biosciences
University of Westminster, London (W1W 6UW)

Abstract

Recently the *Trichomonas vaginalis* genome sequence has been made available by The Institute for Genomic Research (TIGR). Our analysis of this organism has determined the presence of only a single ORF from *T. vaginalis* that shows similarity in sequence to cellular ATP dependent DNA ligases. The homology of 51 and 53% with the human and yeast replicative, DNA ligase I enzymes respectively suggests that this is the parasite replicative DNA ligase. We have cloned, over expressed and purified recombinant protein encoded by this ORF that we demonstrate to be an ATP-dependent DNA ligase.

Introduction

Trichomonas vaginalis (TV) is a parasitic protozoan organism that is the etiological agent of trichomoniasis; a sexually transmitted disease of global concern leading to an estimated 180 million cases per annum. This disease is linked to a predisposition to cervical cancer, atypical pelvic inflammatory disease, infertility and furthermore sufferers have been shown to have an increased susceptibility to human immunodeficiency virus (HIV) infection (Petrin *et al* 1998).

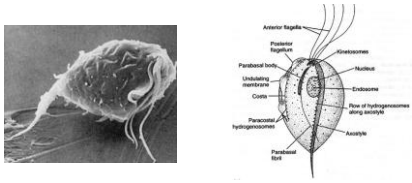


Figure 1:*Trichomonas vaginalis* trophozoite. Images taken from Petrin *et al* 1998.

The 5-nitroimidazoles, especially metronidazole, are the most widely used antimicrobial agents for the treatment of trichomoniasis. The mechanism of the protozooicidal effect of the 5-nitroimidazoles is thought to involve generation of nitro radicals, which leads to subsequent DNA damage and cell death. However metronidazole resistant strains of *T.vaginalis* are on the rise, emphasising the need for research into alternative antibiotics and drug targets.

DNA Ligase enzymes

DNA ligase enzymes are required to seal breaks in double-stranded DNA to maintain the integrity of an organisms' genome during various aspects of DNA metabolism, such as replication, excision repair and recombination. Most eukaryotic organisms express at least one DNA ligase; the human genome elaborates three such enzymes (types I, III and IV), Yeasts and Drosophila two (types I and IV). DNA ligase crystal structures and their sequences have shown that these enzymes are built around a common catalytic core of two polypeptide domains containing six conserved, co-linearly arranged motifs numbered I, III, IIIa, IV, V and VI, illustrated below by the structure of the PBCV-1 DNA ligase, the smallest characterized ATP-dependent ligase. Subsidiary domains are appended to this catalytic core facilitating the incorporation of the enzyme into cellular machines responsible for DNA replication and repai:

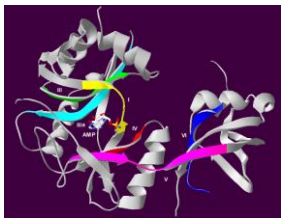
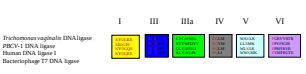


Figure 2: The structure of the PBCV-1 DNA ligase enzyme covalently linked to AMP. The essential catalytic ligase motifs are revealed by coloured ribbons with their appropriate numerical numbering illustrated (Odell *et al* 2000).

Results:

A BLAST search of the newly annotated TV genome revealed a putative polypeptide with 51% and 45% homology to human DNA ligase I and DNA ligase IV respectively. Further sequence analyses revealed the absence of any further conserved domains consistent with the polypeptide being either a type III or type IV DNA ligase but suggesting the putative enzyme to be the TV homolog of the type I, replicative DNA ligase, see Figure 3 below.

A



B

```
TV Lig  -150-KEGDLSKVAQQIRSTQR-337-EYEYDGEA-487-EGLMVKTLN-506-WAKLKKDYIS-532-DTIDVVVIGA
Hu LigI -374-EKGDVGLVAENSRSTQR-565-EYEYDGQR-719-EGLMVRTLD-743-WLKLKKDYLD-755-DTLDLVVIGA
             ::*:. .*::: ****:     ******:** ********:    * ******:.     **:*:*****
```

Figure 3: A. An alignment of the six conserved motifs of various ATP-dependent DNA ligases incorporating the putative TV DNA ligase.**B**. The sequence alignment of the DNA binding regions of human DNA ligase I with the putative TV DNA ligase (predicted expect value of 1e-23).

Expression and purification of TV DNA Ligase

The open reading frame of the TV ligase was amplified and cloned into the expression vector pET-16b between the *Nde* I and *Bam* HI restriction enzyme sites. The sequence of the final clone completely matched that of the open reading frame published by TIGR. The soluble protein with a predicted molecular weight of 78kD and p*I* of 7.8 was purified by nickel affinity and ion exchange chromatography (S-sepharose). The elution profiles, monitored via SDS-PAGE are shown in Figure 4.

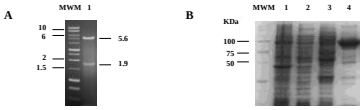


Figure 4: (A) Restriction digest confirmation of TV DNA ligase clone in pET -16b. **MWM;** molecular weight marker DNA. **1**; NdeI/BamHI restriction digest of TV clone. **(B)** Elution profile for TV ligase from Ni-NTA agarose: **MWM;** molecular weight marker proteins. **1**; Cell pellet. **2**; Column flow through. **3**; 25mM immidazole wash. **4**; 300mM immidazole elution.

The recombinant protein carried an N-terminal His-tag and the purified protein was confirmed by western blotting with anti-his antibody.

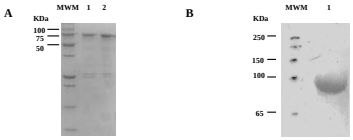


Figure 5: A. TV ligase purified from S-sepharose. **MWM;** molecular weight marker proteins. **1 and 2**; 0.2 M NaCl elution fraction. **B.** Western blotting of the purified TV DNA ligase with anti-his antibody. **MWM;** molecular weight marker proteins. **1** 0.2 M NaCl fraction.

Ligation assays

The purified recombinant protein was used to ligate *Hind*III digested λ DNA with T4 DNA ligase as a positive control. Figure 6, panel A, reveals a stringent requirement for the presence of polyethylene glycol (8000) to facilitate DNA ligation. When PEG is present at 5% with the TV DNA ligase ligation products (higher molecular weight species, >23 kb in size) are observed.migrating in the agarose gel, Figure 6 panel A. Crude assessment of the ATP-requiring nature of the enzyme preparation has also been analysed by similar ligation of lambda fragments. Figure 6 panel B.

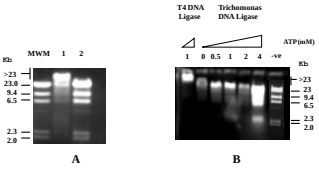


Figure 6: Preliminary ligation assays for TV DNA ligase with *Hind*III digested λ DNA. Briefly, λ DNA in 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol and 25 µg/ml bovine serum albumin buffer was supplemented with 5% PEG (experiment A) and 5% PEG with 0-4 mM ATP (experiment B).

This clearly indicates that the putative ORF from TV to function as an ATP dependent DNA ligase. Higher concentrations of ATP apparently inhibit the ligation which we deduce to be due to the chelation of the essential, divalent metal magnesium. The enzyme was active only in the presence of PEG which is believed to act as a macro-molecular crowding agent and hence, enhance the frequency of DNA ligation.

Conclusions and Future Work

We report here the cloning, expression and initial characterisation of the first protozoal DNA ligase. The ligase requires ATP and PEG for efficient DNA ligation. Analysis of the amino acid sequence of the enzyme suggests that this is the homolog of the human type I ligase and as such is likely to be responsible for ligase events associated with DNA replication. Our further analysis of the TV genome has not revealed the presence of any further ORFs that have homology with DNA ligases. This raises interesting questions about the nature of double-strand break repair in this organism, an event normally associated with a type IV DNA ligase.

Further experiments with γ-^{32}ATP labelled substrates are now being used to further characterise the K_m and V_{max} of the enzyme. The substrates to be used in the ligation assays are an oligonucleotide 30-mer (unlabelled) and a radioactively-labelled 10-mer oligonucleotide, as shown in Figure 7. Initial crystallisation screens of the purified enzyme are also underway.

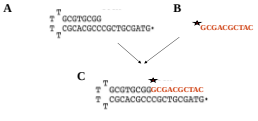


Figure 7: A: Unlabelled 30-mer hairpin **B:** Radiolabelled 10-mer **C:** Expected 40-mer product after ligation

References

- Cudmore,S.L., Delgaty,L.K., Hayward-McClelland,S.F., Petrin, D.P. and. Garber,G.E (2004) *Clin. Microbiol. Rev.* **(17):** 783-793
- Dunne, R. L., L. A. Dunne, P. Upcroft, P. J. O'Donoghue, and J. A. Upcroft. (2003) *Cell Research* **(13):**239-249
- Yarlett, N., C. C. Rowlands, N. C. Yarlett, J. C. Evans, and D. Lloyd. (1987) *Parasitology* **(94):**93-99.
- Odell, M., Sriskanda, V., Shuman, S. & Nikolov, D.B. (2000). *Molecular Cell*, 6, 1183 – 1193
- Odell,M., Malinina,L.,Sriskanda,V., Teplova,M. And Shuman,S.(2003) *Nucleic acid research*,**(31):** 5090-5100
- Poppe, W. A. J. (2001) *Eur. J. Obstet. Gynecol. Reprod. Biol.(* **(96):**119-120.
- Petrin, D., Delgaty, K., Bhatt, R. and Garber G. (1998). *Clinical Microbiology Reviews*. **11**(2), 300-317.

Acknowledgements

This piece of work is funded by the Cavendish Biosciences Scholarship from University of Westminster, London under the guidance of Dr. Mark Odell.

7.6.2 Abstract submitted for Gordon Research Conference on DNA Damage, Mutation & Cancer held in March 2008 at Ventura, USA.

Developing an assay to screen for inhibitors of ATP-dependent DNA ligases

L.Kaur and M.Odell

DNA ligases are enzymes required for the repair, replication and recombination of DNA. Most organisms express DNA ligases that utilise ATP as an energy cofactor; however eubacteria appear to be unique in having ligases driven by NAD^+. Sequence and structural analyses of DNA ligases have shown that these enzymes are built around a common catalytic core, which contains six conserved motifs. DNA ligases represent a novel therapeutic target for intervention in cancerous cells through the requirement of actively dividing cells to seal the discontinuous strand during semi-conservative DNA replication.

Our goal is to develop a high-throughput assay that can be used to screen natural product libraries to identify inhibitors of ATP-dependent DNA ligase enzymes.
The key feature of our rapid DNA ligation, inhibition assay is the relative binding affinities of the components of the unligated DNA substrate and ligated DNA product to the ion exchange resin, Q-sepharose.

The *Paramecium bursari Chlorella* virus, PBCV-1, DNA ligase is the smallest fully functional eukaryotic DNA ligase. This enzyme retains the catalytic core structure present in all DNA ligases. We used this enzyme as the primary screen for inhibitors in our *in vitro* assay. When suitable candidate inhibitors are identified their action will be evaluated on target ligases including human and protozoal enzymes. A hypothetical gene from a parasitic protozoan, *Trichomonas vaginalis*, showing similarity in sequence to other ATP-dependent DNA ligases, has been successfully cloned, expressed and purified from *E. coli* using Nickel-affinity and Ion exchange (S-sepharose) resins. This enzyme is biochemically characterised and to be used in assays to get an insight into the previously least studied protozoan DNA ligases.

Developing an assay to screen for inhibitors of ATP-dependent DNA ligases

L.Kaur, and M.Odell
Department of Molecular and Applied Biosciences
University of Westminster

Introduction

DNA ligases are the enzymes required to seal single-strand breaks in double stranded DNA and maintain the integrity of an organism's genome during various aspects of DNA metabolism such as replication, excision repair and recombination. Most organism express DNA ligases that utilise ATP as an energy co-factor, however eubacteria appear to be unique in having ligases driven by NAD+. The general DNA ligase catalytic mechanism involves 3 steps (figure 1):

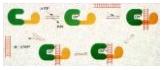


Figure 1: Mechanism of DNA ligation
i-ii)Formation of a covalent enzyme-AMP intermediate linked to a lysine side-chain in the enzyme;
iii-iv)Transfer of the AMP nucleotide to the 5' phosphate of the nicked DNA strand;
v-vi)Attack on the AMP-DNA bond by the 3'-OH of the nicked DNA sealing the phosphate backbone and releasing AMP (adapted from Subramanya *et al.*1996)

Structural and Mechanistic Conservation in DNA ligases Sequence and structural analyses of DNA ligases have shown that these enzymes are built around a common catalytic core. The *Paramecium bursaria Chlorella* virus 1 (PBCV1) ligase represents the most attractive eukaryotic enzyme to study as it contains the six key catalytic motifs found in all ATP-dependent DNA ligases with no extraneous protein sequences.

It is regarded by some as the catalytic core present in all ligases such that anything that applies to this enzyme will apply to the larger, cellular ligases (Odell *et al*. 2003). Figure 2 demonstrates the validity of this assumption by overlaying the key structural residues of motif I and III in the active sites of the PBCV1 (34 kD) and much larger 66kDa T7 ligases. The recently published structure of human DNA ligase I complexed with DNA validates this presumption (Pascal *et al*, 2004).

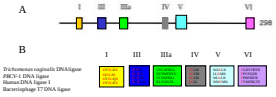


Figure 2: a) The 298 amino acid PBCV1 DNA ligase polypeptide is depicted as a straight line, with the positions of conserved motifs I, III, IIIa, IV, V and VI denoted by boxes. b)An alignment of the six conserved motifs of various ATP-dependent DNA ligases.

Human DNA Ligase I

Human encode for three different DNA ligases (I, III and IV) which are different from each other in their catalytic ,physical and serological properties.

DNA ligase I is responsible for majority of DNA ligase activity in proliferating cells. It is been reported that levels of DNA ligase I are up-regulated in cells from a human tumour versus normal tissues and the presence of DNA ligase I is essential for the survival of tumour cells (Sun *et al.* 2001).

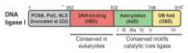


Figure3: Human Ligase I domain organisation (adapted from Pascal *et al*, 2004).

Trichomonas vaginalis DNA Ligase A single, hypothetical gene from *Trichomonas vaginalis* (TV) has been identified that shows similarity in sequence to cellular ATP-dependent DNA ligases, with the highest homology being to the replicative, DNA ligase I enzyme (figure 2b) . The biochemical and enzymatic capabilities of this enzyme will be the first protozoal DNA ligase to be so characterised.

Aims

The project aims to assess the potential of the essential enzymes DNA ligases as targets for chemotherapeutic intervention. A high-throughput assay is being developed for screening natural product libraries to identify inhibitors of ATP-dependent DNA ligase enzymes. The key feature of the assay is the relative binding affinities of the components of the unligated DNA substrate and the ligated DNA product to the ion exchange resin, Q-sepharose. The inhibition of a number of DNA ligase enzymes will be assessed as targets for intervention; those of human cells as a treatment for cancer and a novel DNA ligase from the parasitic protozoan organism *Trichomonas vaginalis*.

Materials and Methods

Expression and purification of PBCV-1 DNA ligase:
Wild type PBCV1 was over-expressed in *E.coli* BL21(DE3) cells. Protein expression was induced at 37ºC for 3 hours.
The protein was purified by S-sepharose(ion exchange) resin and Nickel affinity chromatography.

Oligonucleotides binding to Q-sepharose
The 30-mer and 6-mer oligonucleotides was applied to Q-sepharose (ion exchange resin) and step eluted with 0-2M NaCl.
The elution profile was monitored by checking the absorbance at 260nm.

Cloning and over expression of TV ligase
Two oligonucleotide primers that flanked the open reading frame of the TV ligase were designed with *Bam HI* restriction enzyme sites
5'-GATC*GGATCC*GATGACTCAACAGGGAATTGC-3'
5'-GATC*GGATCC*TTATTTCTCAAATTGTTGATGATAGAG-3'.
The oligonucleotides were used in a PCR and the amplified DNA fragment was cloned in the *BamHI* sites of the pet-16b plasmid vector·The plasmid was transformed into the *E. coli* strain BL21(DE3).A transformed colony was inoculated in to LB medium supplemented with 100 µg·mL^{-1} ampicillin. Protein expression was induced at OD_{595} = 0.5 with 0.5 mM isopropyl thio -D-galactoside (IPTG) for 3 h at 37 °C.

Purification of TV ligase
The over expressed protein was applied to Nickel-NTA resin and eluted with 300mM Imidazole. The eluted fraction was applied to S-sepharose and step eluted with 0.07-1.5M NaCl.

Ligation assays
The purified recombinant protein was used to ligate Hind III digested λ DNA in 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 1 mM ATP, 25 µg/ml bovine serum albumin with T4 and PBCV1 DNA ligases as positive controls.

Results and Discussion

Oligonucleotides binding to Q-sepharose

The non radiolabelled 30-mer and radiolabelled 6-mer oligonucleotides will be used as substrate in DNA ligation assays after annealing to form the DNA ligation substrate shown in Figure 4, panel A. After the completion of the reaction, the reaction mixture will be applied to Q-sepharose ion exchange resin. Figure 4, panel A shows the salt concentrations whereby 30-mer and 6-mer is unable to bind to the Q-sepharose

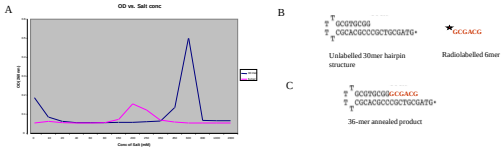


Figure4a) :Elution profile of profile of 30-mer and 6-mer from Q-sepharose under NaCl treatment b) substrate for DNA ligase inhibitory assay
c) 36-mer annealed product

Washing of the resin with 180mM NaCl elutes the unligated 6-mer oligonucleotide whereas 6-mer that has become incorporated into the 30-mer hairpin DNA through DNA ligation will be retained on the resin. By monitoring the labelled 6-mer passing through the column the percentage of DNA ligation can be determined.

Expression and purification of PBCV-1 DNA ligase

The ligase fraction was eluted using 0.2-0.3M NaCl from S-sepharose with some contaminating proteins (see figure 5a). The eluate from Nickel resin gave efficient purification of recombinant enzyme from all contaminants (figure 5b). This fraction was used in ligation assays (see figure8) which shows that protein is functional.

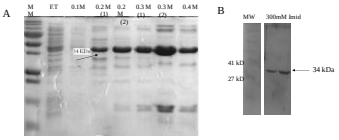


Figure5: a)S-sepharose Purification of Chlorella Virus Ligase b) Elution of PBCV1 ligase from Nickel resin

Cloning and over expression of TV ligase

Genomic DNA was used to amplify the putative TV ligase gene. The insert in pET-16b was sequenced and the plasmid then transformed into the *E. coli* strain BL21(DE3) to study expression of the protein.

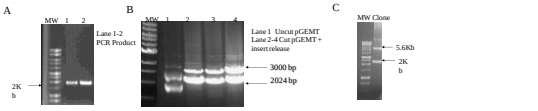


Figure 6: 0.7% Agarose gel electrophoresis for A) PCR product of *TV*. Ligase B)Insert release from p-GEMT after Bam HI digestion C) Insert release from pET -16b plasmid after Bam HI digestion

The soluble protein with molecular weight of 78kD and pI 7.8 was partially purified by nickel affinity chromatography (figure 7a). The eluate was further applied to S-sepharose resin and the putative ligase fraction was eluted with 0.2M NaCl (figure 7b).

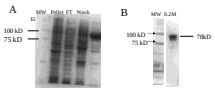


Figure 7:a) Elution profile of TV ligase from Nickel resin with 0.3MImidazole b)Elution of TV ligase from S-sepharose at 0.2M NaCl

Further attempts would be made to get pure protein to study detailed analysis of the biochemical and enzymatic capabilities of this enzyme which represents the first protozoal DNA ligase.

Ligation assays

Some preliminary ligation assays were conducted to ligate Hind III digested λ DNA to characterise the recombinant protein. Figure 8 shows some signs of higher molecular weight species formed when compared to–ve and +ve control reactions.

This clearly indicates that the putative ORF from TV can function as an ATP dependent DNA ligase

Figure8: Preliminary Ligation assay for TV ligase with Hind III digested λ DNA.

Future Work

•The enzymatic and biochemical properties of the TV DNA ligase enzyme will be characterised. The first such enzyme to be researched from the family of protozoan parasites.
•The Human DNA ligase I enzyme will be expressed and purified. Once this full battery of DNA ligases is available they will be screened in our *in vitro* assay to identify potentially useful inhibitory compounds.

References:
Odell,M., Malinina,L.,Sriskanda,V., Teplova,M. And Shuman,S.(2003) *Nucleic acid research*,**(31)**: 5090-5100.
Pascal,J.M, O'Brien,P.J., Tomkinson,A.E. and Ellenberger,T. (2004) *Nature* **(432)**: 473-478.
Subramanya,H.S., Doherty,A.J., Ashford,S.R. and Wigely,D.B. (1996) *Cell*. **(85)**: 607-615
Sun,D., Urrabaz,R., Nguyen,M., Martyn,J., Stringer, S., Cruz,E., Medina-Gundrum,L. and Weitman,S.(2001). *Clinical cancer research*,**(17)**: 4143-4148.

CHAPTER 8

REFERENCES

CHAPTER 8: REFERENCES

Ahel, I., Rass, U., El-Khamisy, S.F., Katyal, S., Clements, P.M., McKinnon, P.J., Caldecott, K.W. and West, S.C. (2006) The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature* **(7112):** 713-716.

Akey, D., Martins, A., Aniukwu, J., Glickman, M.S., Shuman, S. and Berger, J.M. (2006) Crystal structure and non-homologous end-joining function of the ligase component of *Mycobacterium* DNA ligase D. *The Journal of Biological Chemistry*, **(281):** 13412 – 13423.

Albert, J., Cao, C., Kim, K., Willey, C., Geng, L. and Xiao, D. (2007) Inhibition of poly(ADP-ribose) polymerase enhances cell death and improves tumor growth delay in irradiated lung cancer models. *Clinical cancer research* **(10):**3033–42.

Almeida, K.H. and Sobol, R.W. (2007) A unified view of base excision repair: lesion-dependent protein complexes regulated by post-translational modification. *DNA repair* **(6):** 695-711.

Andreini, I., Bertini, I., Cavallaro, G., Holliday, G.L. and Thornton, J.M. (2008) Metal ions in biological catalysis: from enzyme databases to general principles. *Journal of Biological Inorganic Chemistry* **(13):** 1205-1218.

Andrews, B.J. and Turchi, J.J. (2004) Development of high-throughput screen for inhibitors of replication protein A and its role in nucleotide excision repair. *Molecular Cancer Therapeutics* **(3):** 385-391.

Barktova, J., Horejsi, Z., Koed, K., Kramer, A., Tort, F., Zieger, K., Guldberg, P., Sehested, M., Nesland, J.M. and Lukas, C. *et. al.*, (2005) DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **(434):** 864-870.

Bartlett, M.S. and Stirling, D. (2003).A Short History of the Polymerase Chain Reaction Methods. Mol Biol. **(226):**3-6.

Barnes, D.E., Johnston,L.H., Kodama,K., Tomkinson,A.E., Lasko,D.D. and Lindahl,T.(1990) Human DNA ligase I cDNA: cloning and functional expression in *Saccharomyces cerevisiae*. *PNAS (USA)* **(87):**6679–6683.

Baumann, P. and West, S.C. (1998) DNA end-joining catalysed by human cell-free extracts. *PNAS USA* **(95):** 14066 – 14070.

Becker, A., Lyn, G., Gefter, M. and Hurwitz, J. (1967) The enzymatic repair of DNA, Characterisation of phage induced sealase. *PNAS (USA)* **(58):**1996-2003.

Benarroch, D. and Shuman,S. (2006) Characterization of mimivirus NAD^+-dependent DNA ligase. *Virology (1):* 133-143.

Bentley, D.J., Harrison, C., Ketchen, A.M., Redhead, N.J., Samuel, K., Waterfall, M., Ansell, J.D. and Melton, D.W. (2002) DNA ligase I null mouse cells show normal DNA repair activity but altered DNA replication and reduced genome stability. *Journal of Cell Science* **(115):** 1551-1561.

Bernard, P., Kezdy, K.E., Van Melderen, L., Steyaet, J., Wyns, L., Pato, M.L., Higgins, P.N. and Couturier, M. (1993) *Journal of Molecular Biology* **(3):** 534-541.

Bernstein, H.J. (2000) Recent changes to RasMol, recombining the variants. *Trends in Biochemical Sciences* **(25):** 453-455.

Birnboim, H. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **(7):** 1513-1523.

Bodley, A.L. and Shapiro, T.A. (1995): Molecular and cytotoxic effects of camptothecin, a topoisomerase I inhibitor, on trypanosomes and *Leishmania.* *PNAS (USA)* **(92):**3726-3730.

Boiteux, S. and Guillet,M. (2004) Abasic sites in DNA:repair and biological consequences in *Saacharomyces cerevisiae.* *DNA Repair* **(3):** 1-12.

Bolanos-Garcia, V.M. & Davies, O.R. (2006). Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta*, **(1760):** 1304 – 1313.

Brooke, C. (2007) Analysis of DNA binding by ATP-dependent DNA ligases. PhD Thesis, University of Westminster, London, UK.

Brotz-Oesterhelt, H., Knezevic, I., Bartel, S., Lampe, T., Warnecke-Eberz, U., Ziegelbauer, K., Habich, D. and Labischiinski, H. (2003) Specific and potent inhibition of NAD^+-dependent DNA ligase by pyridochromanones. *Journal of Biological Chemistry* **(278):** 39435-39442.

Buguliskis, J.S., Casta, L.J., Butz, C.E., Matsumoto, Y. and Taraschi, T.F. (2007) Expression and biochemical characterisation of *Plasmodium Falciparum* DNA ligase I. *Molecular and Biochemical Parasitology* **(155):** 128-137.

Burton, P., McBride, D.J., Wilkes, J.M., Barry, D. and McCulloch, R. (2007) Ku heterodimer-independent end joining in *Trypanosoma brucei* cell extracts relies upon sequence microhomology. *Eukaryotic Cell* **(10):** 1773-1781.

Caldecott, K.W., McKeown, C.K., Tucker, J.D., Ljungquist, S. and Thompson, L.H. (1994) An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Molecular and Cellular Biology* **(14):** 68 – 76.

Caldecott, K. W., Aoufouchi, S., Johnson, P. and Shall. S. **(**1996) XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADPribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Resarch. (***24):**4387–4394.

Cao, W. (2001) DNA ligases and ligase-based technologies. *Clinical and Applied Immunology Reviews.* **(2):** 33-43.

Cappelli, E., Taylor, R., Cevasco, M., Abbondandolo, A., Caldecotts, K. and Frosina, G. (1997) Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair.*The American Society for Biochemistry and Molecular Biology*.**(272):** 23970-23975.

Cardoso, M.C., Joseph, C., Rahn, H-P., Reusch, R., Nadal-Ginard, B. and Leonhardt, H. (1997) Mapping and use of a sequence that targets DNA ligase I to sites of DNA. *Journal of Cell Biology* **(139):** 579-587.

Carlton, J.M., Hirt, R.P., Silva, J.C., Delcher, A.L., Schatz, M., Zhao, Q., Wortman, J.R. and Bidwell, S.L. *et al.* (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis.* *Science* **(315):** 207 – 212.

Casta, L.J., Buguliskis, J.S., Matsumoto, Y. and Taraschi, T.F. (2007) Expression and biochemical characterization of the *Plasmodium falciparum* DNA repair enzyme, flap endonuclease-1 (PfFEN-1) *Mol Biochem Parasitol.* **(1):**1-12.

Cheng, C. and Shuman, S. (1997) Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae.* *Nucleic acid research* **(25):** 1369-1374.

Chen, G.F.T. and Inouye, M. (1990) Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Research* **(18):** 1465-1473.

Chen, C.R., Malik, M., Synder, M. and Drlica, K. (1996) DNA gyrase and topoisomerase IV on the bacterial chromosome: quinolone-induced DNA cleavage. *Journal of Molecular Biology* **(4):** 627-637.

Chen, X., Pascal, J., Vijayakumar, S., Wilson, G.M., Ellenberger, T. and Tomkinson, A.E. (2006) Human DNA ligases I, III and IV- purification and new specific assays for these enzymes. *Methods in Enzymology* **(409):** 39-52.

Chen, X., Zhong, S., Zhu, X., Dziegielewska, B., Ellenberger, T., Wilson, G.M., Mackerell, A.D. and Tomkinson, A.E. (2008) Rational design of human DNA ligase inhibitors that target cellular DNA replication and repair. *Cancer Research* **(68):** 3169-3177.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., Thompson, J. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* (**31):** 3497 – 3450.

Cherepanov A.V. and Vries S.D. (2002) Kinetic mechanism of the Mg^{2+}-dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases *European Journal of Biochemistry* **(277):**1695-1704.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **(18)**: 5294–9.

Christmann, M., Tomicic, M.T., Ross, W.P. and Kaina, B. (2003) Mechanism of human DNA repair: an update. *Toxicology* **(193):** 3-34

Ciarrocchi, G., Macphee, D.G., Deady, L.W and Tilley,L. (1999) Specific inhibition of the eubacterial DNA Ligase by arylamino compounds. *Antimicrobial Agents and Chemotherapy* **(43)**: 2766-2772.

Cong, P. and Shuman, S. (1995) Mutational analysis of mRNA capping enzyme identifies amino acids involved in GTP binding, enzyme-guanylate formation, and GMP transfer to RNA. *Moecular Cell Biology*. **(11):** 6222–6231

Conway. C., McCulloch, R., Ginger, M.L., Robinson, N.P., Browitt, A. and Barry, J.D. (2002**a**) Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *Journal of Biological Chemistry* **(277):** 21269-21277.

Conway, C., Proudfoot, C., Burton, P., Barry, J.D. and McCulloch, R. (2002**b**) Two pathways of homologus recombination in *Trypanosoma brucei. Molecular Microbiology*. **(45):**1687-1700.

Costa, R.M.A., Chigançcas,V., Galhardo,R.S., Carvalho,H. and Menck,C.F.M. (2003) The eukaryotic nucleotide excision repair pathway .*Biochimie* **(85):** 1083-1099

Cotner-Gohara, E., Kim, I-K., Tomkinson, A.E. and Ellenberger, T. (2008) Two DNA binding and nick recognition modules in human DNA ligase III. *Journal of Biological Chemistry* **(283):** 10764 - 10772.

Critchlow, S.E., Bowater, R.P. and Jackson, S.P. (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology* **(7):** 588 – 598.

Crut, A., Nair, P.A., Koster, D.A., Shuman, S. and Dekker, N.H. (2008) Dynamics of phosphodiester synthesis by DNA ligase. *PNAS (USA)* **(105):** 6894-6899.

Cudmore, S.L., Delgaty, K.L., Hayward-McClelland, S.F., Petrin, D.P. and Garber, G.E. (2004) Treatment of infections caused by metronidazole resistane *Trichomonas vaginali.s* *Clinical Microbiology Reviews* **(17):** 783-793.

Curtin, N.J., Wang, L-Z., Yiakouvaki, A., Kyle, S., Arris, C.A., Canan-Koch, S., Webber, S.E., Durkacz,B.W., Calvert, H.A., Hostomsky,Z. and Newell, D.R. (2004) Novel poly (ADP-ribose) polymerase –I inhibitor, AG14361, restores sensitivity to Temozolomide in mismatch repair deficient cells. *Clinical Cancer Research* **(10):** 881-889.

Dagert M. and Ehrlich, S.D. (1979) Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **(1):** 23-28.

Damia, G. and D'Incalci, M. (2007) Targeting DNA repair as a promising approach in cancer therapy. *European Journal of Cancer* **(43):**1791-1801.

Decottignies A. (2007) Microhomology-mediated end joining in fission yeast is repressed by pKu 70 and relies on genes involved in homologus recombination. *Genetics* **(176):** 1403-1415.

DeLano, W.L. (2002) The PyMol molecular graphics system (DeLano Scientific, San Carlos, California, USA)

Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences – a consensus? *Trends Biochemical. Sciences.*(**16**): 478 – 481.

Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J. and Bell,S.D. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Molecular Cell* (**11):** 275–282.

Doherty, A.J. (1999) Conversion of a DNA ligase into an RNA capping enzyme. *Nucleic Acids Research* **(27):** 3253-3258.

Doherty, A.J. and Dafforn, T.R. (2000) Nick recognition by DNA ligases. *Journal of Molecular Biology* (**296):** 43 – 56.

Doherty, A.J. and Suh,S.W.(2000) Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research* **(28)**: 4051-4058.

Dwivedi, N., Dube, D., Pandey, J., Singh, B., Kukshal, V., Ramachandran, R. and Tripathi, R.P. (2008) NAD^+ - dependent DNA ligase: A novel target waiting for the right inhibitor. *Medicinal Research Reviews* **(28):** 545-568.

Eckert, K.A. and Kunkel, T.A (1991) DNA polymerase fidelity and polymerase chain reaction. *Genome Research* **(1):**17-24.

Ellenberger, T. and Tomkinson, A.E. (2008) Eukaryotic DNA ligases: Structural and functional insights. *Annual Review of Biochemistry* **(77):** 313-338.

Filippo,J.S., Sung,P. and Klein, H. (2008) Mechanism of eukaryotic homologous recombination. *Annual Reviews of Biochemistry* (77): 229-257.

Ferrer, M., Golyshina,O.V., Beloqui, A., Bottgrt, L.H., Andreu, J.M., Polaina, J., De Lacey, A.L., Trautwein, A.X., Timmis, K.N. and Golyshin, P.N. (2008) A purple acidophilic di-ferric DNA ligase from *Ferroplasma PNAS (USA)* **(26):** 8878-8883.

Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fortini, P. and Dodliotti, E. (2007) Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair***(6):** 398-409

Foth, B.J. and McFadden, G.I.(2003) The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites. *International Reviews of Cytology* **(224):**57-110.

Fousteri, M. and Mullenders, L.H.F. (2008) Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Research* **(18):**73-84.

Franks, L.M. and Teich, N.M. (1995) Introduction to the cellular and molecular biology of cancer. (Second edition) Oxford University Press, Oxford.

Gajiwala, K.S. and Pinko,C. (2004) Structural rearrangement accompanying NAD^+ synthesis within a bacterial DNA ligase crystal. *Structure* **(12):** 1449-1459.

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fukiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J., Bronson, R.T., Malynn, B.A., Bryans, M., Zhu, C., Chauduri, J., Davidson, L., Ferrini, R., Stamato, T., Orkin, S.H., Greenberg, M.E. and Alt, F.W. (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell,* (**95):** 891 – 902.

Gellert, M. (1967) Formation of covalent circles of lambda DNA by *E. coli* extracts. *PNAS (USA)* **(57):** 148-155.

Gellert, M and Bullock, M.L. (1970) DNA ligase mutants of *Escherichia coli* *PNAS (USA)* **(67):** 1580-1587.

Gill, E.E. and Fast,N.M. (2007) Stripped-down DNA repair in a highly reduced parasite *BMC Molecular Biology*. doi:10.1186/1471-2199-8-24.

Glover, L., McCulloch, R. and Horn, D. (2008) Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Research* (**36):** 2608-2618.

Gong, C., Martins, A., Bongiorno, P., Glickman, M. and Shuman, S. (2004) Biochemical and genetic analysis of the four DNA ligases of Mycobacteria. *Journal of Biological Chemistry* **(279):** 20594-20606.

Grawunder. U., Zimmer,D., Fugmann, S., Schwarz, K. and Lieber, M. (1998) DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Molecular Cell* **(2):** 477-484.

Gumport, R.I., and Lehman, I.R. (1971) Structure of the DNA ligase-adenylate intermediate: lysine (ε-amino)-linked adenosine monophosphoramidate. *PNAS (USA)* **(68):** 2559-2563.

Haltiwanger, B.M., Matsumoto, Y., Nicolas, E., Dianov, G.L., Bohr, V.A. and Taraschi, T.F. (2000) DNA excision repair in human malaria parasites is predominately by a long-patch pathway. *Biochemistry* **(39):** 763-772.

Hakansson, K. and Wigley, D. B. (1998) Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. *PNAS (USA)* **(95):** 1505-1510.

Harcum, S.W. and Bentley, W.E. (2004) Response dynamics of 26-, 34-, 39-, 54-, and 80-kDa proteases in induced cultures of recombinant *Escherichia coli.* *Biotechnology and Bioengineering* **(42):**675 – 685.

Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N. and Obayashi, A. (1986) Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. *Nucleic Acids Research* **(14):** 7618-7631.

Hefferin, M.L. and Tomkinson, A.E. (2005) Mechanism of DNA double-strand break repair by non-homologous end joining.*DNA repair* **(6):** 639-48.

Ho, C.K., Van Etten, J.L. and Shuman, S. (1997) Characterization of an ATP dependent DNA ligase encoded by *Chlorella* Virus PBCV-1. *Journal of Virology* **(71):** 1931-1937.

Ho, C.K., Wang,L.K., Lima, C.D. and Shuman,S. (2004) Structure and mechanism of RNA ligase. *Structure* **(12):** 327-339.

Hoeijmakers, J.H.J. (2001) Genome maintenance mechanism for preventing cancer. *Nature* **(411):** 366-374.

Honigberg, B. M. (1963) Evolutionary and systematic relationships in the flagellate order Trichomonadida Kirby. *Journal of Protozoology* (**10)**: 20–63.

Ibezim, E.C. and Odo, U. (2008) Current trends in malarial chemotherapy. *African Journal of Biotechnology* **(4):** 349-356.

Izumi, T., Wiederhold, L.R., Roy, G., Roy, R., Jaiswal, A., Bhakat, K.K., Mitra, S. and Hazra, T.K. (2003) Mammalian DNA base excision repair proteins: their interactions and role in repair of oxidative DNA damage. *Toxicology* **(193):** 43-65.

Jayaram, S., Ketner, G., Adachi,N. and Hanakahi,L.A. (2008) Loss of DNA ligase IV prevents recognition of DNA by double-strand break repair proteins XRCC4 and XLF. *Nucleic Acid Research*. **(36):** 5773-5786.

Kaczmarek, F. S., Zaniewski, R. P., Gootz, T. D., Danley, D. E., Mansour, M. N., Griffor, M., Kamath, A. V., Cronan, M., Mueller, J., Sun, D., Martin, P. K., Benton, B., McDowell, L., Biek, D. and Schmid, M. B. **(**2001) Cloning and functional characterization of an NAD^+-dependent DNA ligase from *Staphylococcus aureus*. *Journal of Bacteriology* (**18)** : 3016-3024.

Kane, J.F. (1995) Effect of rare codon clusters on high level expression of heterologus proteins in *E. coli*. *Current Opinions in Biotechnology* **(6):** 494-500.

Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature protocols* **(4):** 363-371.

Keppetipola, N. and Shuman, S. (2005) Characterization of a thermophilic ATP-Dependent DNA ligase from the euryarchaeon *Pyrococcus horikoshii*. *Journal of Bacteriology* **(20):** 6902-6908.

Kodama, K., Barnes, D.E. and Lindahl, T. (1991) In vitro mutagenesis and functional expression in E.coli of a cDNA encoding the catalytic domain of human DNA ligase I. *Nucleic acid research.* **(19)**: 6093-6099.

Korycka-Machala,M., Rychta,E., Brzostek,A., Sayer, H.R., Rumijowska-Galewicz, A., Bowater, R.P. and Dziadek,J. (2007) Evaluation of NAD^+ dependent DNA ligase of Mycobateria as a potential target for antibiotics. *Antimicrobial agents and Chemotherapy* **(51):** 2888-2897.

Krajewski, S., Zapata, J.M. and Reed, J.C. (1995). Detection of multiple antigens on Western blots. *Analytical Biochemistry* **(236):** 221 – 228.

Krokan, H.E., Nilsen, H., Skorpen, F., Otterlei, M. and Slupphaug, G. (2000) Base excision repair of DNA in mammalian cells. *FEBS Letters* **(476):** 73–77.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T_4. *Nature* **(227):** 680-685.

Lakshmipathy, U. and Campbell, C. (1999) The human DNA ligase III gene encodes nuclear and mitochondrial proteins. *Molecular and Cellular Biology* **(5):** 3869-3876.

Lavrukhin, O.V., Fortune,J.M., Wood,T.G, Burbank,D.E.,Van Etten,J.L.,Osheroff,N. and Lloyd,R.S.(2000) Topoisomerase II from *Chlorella* virus PBCV-1. Characterization of the smallest known type II topoisomerase. *Journal of Biological Chemistry* **(10):** 6915-6921.

Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J.K.,Kim, H.K., Kwon, S.T. and Suh, S.W. (2000) Crystal struc ture of NAD+-dependent DNA ligase: modular architecture and functionalimplications. *European Molecular Biology Organization* **(19):** 1119–1129.

Lee, S.H., Hwang, H.S. and Yun, J.W. (2009) Antitumor activity of watern extract of a mushroom, Inonotus obliquus, against HT-29 human colon cancer cells. *Phytotherapy Research* : 10.1002/ptr.2836

Lehman, I.R. (1974) DNA ligase**:** structure, mechanism, and function. *Science* **(186):** 790-797.

Levin, D.S., Bai, W., Yao, N., O'Donnell, M. and Tomkinson, A.E.(1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining, *PNAS (USA)* **(94):** 12863–12868

Levin, D.S., Mckenna, A.E., Motycka, T.A., Matsumoto, Y. and Tomkinson A.E. (2000) Interactions between PCNA and DNA ligase I is critical for joining of Okazaki fragments and long-patch base excision repair. *Current Biology* **(15):** 919-922.

Liang, Li., Deng, L., Nguyen, S.C., Zhao, X., Maulion, C.D., Shao, C. and Tischfield, J.A.(2008) Human DNA ligase I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Research* **(36):** 1-14.

Li, Z., Otevrel, T., Gao, Y., Cheng, H-L., Seed, B., Stamato, T.D., Taccioll, G.E. and Alt, F.E. (1995) The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **(83):** 1079 – 1089.

Lindahl, T and Barnes, D.E. (1992) Mammalian DNA ligases. *Annual Review of Biochemistry* **(61):** 251–281.

Lord, C.J., Garrett, M.D. and Ashworth, A. (2006) Targeting the double-strand DNA break repair pathway as a therapeutic strategy. *Clinical Cancer Research* **(12):** 4463-4468.

Luo,J. and Barany,F. (1996) Identification of essential residues in *Thermus thermophilus* DNA ligase. *Nucleic Acid Research* **(24)**: 3079-3085.

Ma, J.L., Kim, E.M., Haber, J.E. and Lee, S.E. (2003) Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Molecular and Cellular Biology* **(23):** 8820–8828.

Madhusudan, S. and Hickson, I.D. (2005) DNA repair inhibition: a selective tumour targeting strategy. *Trends in Molecular Medicine* **(11):** 503-511.

Madhusudan, S. and Middleton, M.R. (2005) The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. *Cancer Treatment Reviews* **(31):** 603-617.

Mahler , H.C., Friess, W.C., Grauschopf , U. and Kiese, S. (2008) Protein aggregation: Pathways, induction factors and analysis. *Journal of Pharmaceutical Sciences*. **(98):** 2909-2934.

Martin, M.B. N. (2001) DNA repair inhibition and cancer therapy. *Journal of Photochemistry and Photobiology B: Biology* **(63):** 162-170.

Martin, I.V., and MacNeil, S.A. (2002) ATP- dependent DNA ligases. *Genome Biology* **(4)**: 3005.1-3005.7.

Meier, T.I., Yan, D., Peery,R.B., Mcallister, K.A., Zook, C., Peng, S-B. and Zhao, G. (2008) Identification and characterization of an inhibitor specific to bacterial NAD^+-dependent DNA ligases. *FEBS Journal* **(21):** 5258-5271.

Mizushina, Y., Hanashima, L., Yamaguchi, T., Takemura, M., Sugawara, F., Saneyoshi, M., Matsukage, A., Yoshida, S. and Sakaguchi, S. (1998) A mushroom fruiting body-inducing substance inhibits activities of replicative DNA polymerases. *Biochemical and Biophysical Research Communications* **(249):**17-22.

Modrich, P. and Lehman, I.R.(1971) Enzymatic characterisation of a mutant of *E. coli* with an altered DNA ligase. *PNAS (USA)* **(68):** 1002-1005.

Modrich, P., Lehman,I. R. and Wang, J.C. (1972) Enzymatic Joining of Polynucleotides XI. Reversal of *Eischerichia coli* deoxyribonucleic acid ligase reaction. *The Journal of Biological Chemistry* **(19):** 6370-6372.

Montecucco, A., Biamonti, G., Savini, E., Spadari, F.F.S. and Ciarrochi, G. (1992) DNA ligase-I gene expression during differentiation and cell proliferation. *Nucleic acid research* **(20):** 6209-6214.

Montecucco, A., Savini,E.,Weighardt,F., Rossi,R., Ciarrocchi,G., Villa,A. and Biamonti,G. (1995) The N-terminal domain of human DNA ligase I contains the nuclear localisation signal and directs the enzyme to sites of DNA replication. *EMBO Journal* (**14**):5379–5386.

Moore, D.J., Taylor, R.M., Clements, P. and Caldecott, K.W. (2000) Mutation of a BRCT domain selectively disrupts DNA single-strand break repair in noncycling Chinese hamster ovary cells. *PNAS (USA)* **(97):** 13649-13654.

Moreno, S.N.J. and Docampo, R. (2003) Calcium regulation in protozoan parasites. *Current Opinion in Microbiology* (4): 359-364.

Moyer, J.D and Henderson, J.F. (1983) Nucleoside triphosphate specificity of firefly luciferase. *Analaytical Biochemistry* **(131)**: 187–189.

Nair, P.A., Nandakumar, J., Smith, P., Odell, M., Lima, C.D. and Shuman, S. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nature structural and Molecular Biology* **(14):** 770-778.

Nandakumar, J., Shuman, S. and Lima, C.D. (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive the reaction forward. *Cell (***127)**:71–84.

Nandakumar, J., Nair,P.A. and Shuman,S. (2007) Last stop on the road to repair:structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Molecular Cell* **(26):** 257-271.

Nishida, H., Kiyonari,S., Ishino,Y. and Morikawa,K.(2006) The closed structure of an archaeal DNA ligase from Pyrococcus furiosus. *Journal of Molecular Biology*. **(360):** 956-967.

Nussenzweig, A. and Nussenzweig, M.C. (2007) A backup DNA repair pathway moves to forefront. *Cell* **(131):** 223-225.

O'Carroll, M., Kidd, T., Coulter, C., Smith, H., Rose, B., Harbour, C. and Bell, S. (2003) *Burkholderia pseudomallei*: another emerging pathogen in cystic fibrosis *Thorax* **(12):** 1087–1091.

Odell, M., Kerr, S.M. and Smith, G.L. (1996) Ligation of double-stranded and single-stranded [oligo(dt)] DNA by Vaccinia virus DNA ligase. *Virology* **(221):** 120–129.

Odell, M. (1997) Vaccinia virus DNA ligase. DPhil Tesis, Oxford.

Odell,M., Malinina,L.,Sriskanda,V., Teplova,M. and Shuman,S.(2003) Analysis of the DNA joining repertoire of Chlorella virus DNA ligase and a new crystal structure of the ligase-adenylate intermediate. *Nucleic acid research.* **(31):** 5090-5100.

Odell, M., Sriskanda,V., Shuman,S. and Nikolov, D.B.(2000) Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Molecular and cellular biology.* **(6)**: 1183-1193.

Odell, M. and Shuman,S.(1999) Foot printing of *Chlorella* virus DNA ligase bound at a nick in duplex DNA. *Journal of biological chemistry* **(274)**: 14032-14039.

Ogawa, T. and Okazaki,T. (1980) Discontinuous DNA Replication. *Annual. Reviews. Biochemistry.* **(49):**421-457

Okayama, H. and Berg, P. (1982) High-efficiency cloning of full-length cDNA. *Molecular and Cellular Biology* **(2):** 161-170.

Olivera,B.M and Lehman, I.R. (1967). Linkage of polynucleotides through phosphodiester bonds by an enzyme from *Escherichia coli*. Proc. Natl. Acad. Sci. (U.S.A.) **(5):** 1426-33.

Olivera, B.M., Hall, Z.W. and Lehman, I.R. (1968) Enzymatic joining of polynucleotides, V. A DNA-adenylate intermediate in the polynucleotide-joining reaction. *PNAS USA* (**61):** 237 - 244.

Pascal, J.M, O'Brien,P.J., Tomkinson,A.E. and Ellenberger,T. (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **(432)**: 473-478.

Pascal, J.M., Tsodikov,O.V., Hura, G.L., Song, W., Cotner, E.A., Classen,S., Tomkinson, A.E., Tainer, J.A. and Ellenberger,T. (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Molecular Cell* **(24):** 279-291.

Pascal, J.M. (2008) DNA and RNA ligases: structural variations and shared mechanisms. *Current Opinions In Structural Biology* **(18):** 96-105.

Panasenko, S.M., Alazard, R.J. and Lehman, I.R. (1978) A simple, three-step procedure for the large scale purification of DNA ligase from a hybrid λ lysogen constructed *in vitro*. *The Journal of Biological Chemistry* **(13)**: 4590-4592.

Park, U.E., Olivera, B.M., Hughes, K.T., Roth, J.R. and Hillyard, D.R. (1989) DNA ligase and the pyridine nucleotide cycle in *Samonella typhimurium*. *Journal of Bacteriology* **(171):** 2173 – 2180.

Petit, M-A. and Erlich, S.D. (2000) The NAD-dependent ligase encoded by *yerG* is an essential gene of *Bacillus subtilis*. *Nucleic Acids Research* **(28):** 4642 – 4648.

Perez-Jannotti, R. M., Klein, S. M. and Bogenhagen, D. F. **(**2001)Two forms of mitochondrial DNA ligase III are produced in Xenopus laevis oocytes. *The Journal of Biological Chemistry* **(276)**:48978–48987.

Petrin, D., Delgaty, K., Bhatt, R. and Garber, G. (1998) Clinical and microbiological aspects of *Trichomonas vaginalis.* *Clinical Microbiology Reviews* **(11):** 300-317.

Poidevin, L and MacNeill,S. (2006) Biochemical characterisation of LigN, an NAD^+-dependent DNA ligase from the halophilic euryarchaeon *Haloferax volcanii* that displays maximal *in vitro* activity at high salt concentrations. *Biomed Centeral Molecular Biology* doi:10.1186/1471-2199-7-44.

Rabin, B. A., and Chase, J. W. (1987) DNA ligase from *Drosophila melanogaster* embryos. Substrate specificity and mechanism of accumulation. *Journal of Biological Chemistry* **(262):** 14105–14111.

Robertson, J.D., Freeman, W.M. and Vrana, K.E. (2001) Phosphorimagers. Encyclopedia of life sciences, Macmillan Publishers Ltd.

Rossi, R., Montecucco,A., Ciarrocchi, G. and Biamonti, G.(1997) Functional characterisation of the T4 DNA ligase : a new insight into the mechanism of action. *Nucleic Acid Research* **(25):** 2106-2113.

Rossi, R., Villa,A., Negri,C., Scovassi, I., Ciarrochi,G., Biamonti,G. and Montecucco, A. (1999) The replication factory targeting sequence/PCNA-binding site is required in G_1 to control the phosphorylation status of DNA ligase I. *EMBO journal***(18):** 5745-5754.

Rosl, F. (1992). A simple method for detection of apoptosis in human cells. *Nucleic Acid Research (***201)**:5243.

Russell, R. and Paterson, M. (2006) *Ganoderma* – a therapeutic fungal biofactory. *Phytochemistry* **(67):**1985-2001.

Salles, B., Calsou, P., Frit, P. and Muller, C. (2006) The DNA repair complex DNA-PK, a pharmacological target in cancer chemotherapy and radiotherapy. *Pathologie- Biologie (Paris)* **(54):**185-193.

Sambrook, J. and Russell, D.W. (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory Press.

Sattler, U., Frit, P., Salles, B. and Calsou, P. (2003) Long-patch DNA repair synthesis during base excision repair in mammalian cells. *EMBO reports* **(4):**363–367.

Schwebke, J.R. and Burgess, D. (2004) Trichomoniasis. *Clinical microbiology reviews* **(17):** 794-803.
Skiguchi, J. and Shuman, S. (1997) Domain structure of Vaccinia DNA ligase. *Nucleic Acids Research* **(25):** 727-734.

Selvakumaran, M., Pisarcik, D.A., Bao, R., Yeung, A.T. and Hamilton, T.C. (2003). Enhanced cisplastin cytotoxicity by disturbing the nucleotide excision repair pathway in ovarian cancer cell lines. *Cancer Research* **(6):** 1311-1316.

Shuck,S.C., Short, E.A. and Turchi, J.J.(2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Research* **(1):** 64-72.

Shuman, S. (1996) Closing the gap on DNA ligase. *Structure* **(6):** 653-656.

Shuman, S. (2009) DNA ligases: progress and prospects. *Journal Of Biological Chemistry* **(284):**17365-17369.

Shuman, S. and Hurwitz, J.(1981) Mechanism of mRNA capping by Vaccinia virus guanylyltransferase: characterization of an enzyme-guanylate intermediate. *PNAS (USA)* **(78):**187-191.

Shuman, S. and Lima,C.D. (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Current Opinions In Structural Biology*. **(14)**:757–764

Shuman, S. and Schwer, B. (1995): RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Molecular Microbiology* **(3):** 405-410.

Shuman, S. and Ru, X-M. (1995) Mutational analysis of Vaccinia DNA ligase defines residues essential for covalent catalysis. *Virology* **(211):** 73 – 83.

Sibanda, B.L., Critchlow, S.E., Begun, J., Pei, X.Y., Jackson, S.P., Blundell, T.L. and Pellegrini, L. (2001) Crystal structure of an XRCC4-DNA ligase IV complex. *Nature* **(8):** 1015 – 1019.

Signoret, J. and David, J.C. (1986) Control of the expression of genes for DNA ligase in eukaryotes. *International Review Of Cytology* **(103):** 249–279

Silber, R., Malathi, V.G. and Hurwitz, J. (1972) Purification and properties of bacteriophage T4-induced RNA ligase. *PNAS (USA)* **(69):** 3009–3013.

Silva, D. (2003) *Ganoderma lucidum* (Reishi) in cancer treatment *Integrative Cancer Therapy* **(2):** 358-364.

Song, O.K., Kim, Y.S. and Rho, H.M. (1985) Studies on the optimal condition for ligation of blunt ended DNA fragments. *Korean Biochemistry Journal.* **(18):** 297-303.

Song, W., Pascal, J.M., Ellenberger, T. and Tomkinson, A.E. (2009) The DNA binding domain of human DNA ligase I interacts with nicked DNA and the DNA sliding clamps, PCNA and hRad9-hRad1-hHus1. *DNA repair* **(8):** 212-219.

Sood, S. and Kapil,A. (2008) An update on Trichomonas Vaginalis. *Indian Journal Of Sexually Transmitted Diseases* (29): 7-14.

Sriskanda,V. and Shuman,S. (1998) Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Resarch* **(26):**3536-3541.

Sriskanda,V., Schwer,B., Ho,C.K. and Shuman,S. (1999) Mutational analysis of *E. coli* DNA ligase identifies amino acids required for nick-ligation *in vitro* and for *in vivo* complementation of the growth of yeast cells deleted for *CDC9* and *LIG4. Nucleic Acids Resarch.*, **(27):** 3953–3963

Sriskanda, V., Kelman, Z., Hurwitz, J. and Shuman, S. (2000) Characterization of an ATP-dependent DNA ligase from the thermophilic archaeon *Methanobacterium thermoautotrophicum. Nucleic Acids Research* **(11):** 2221-2228.

Sriskanda, V., Moyer, R.W. and Shuman, S. (2001) NAD^+-dependent DNA ligase encoded by a eukaryotic virus. *Journal of Biological Chemistry*. **(276):**36100–36109.

Sriskanda, V. and Shuman, S. (2002a) Role of nucleotidyltransferase motifs I, III and IV in the catalysis of phosphodiester bond formation by *Chlorella* virus DNA ligase. *Nucleic Acid Research.* **(30)**: 903-911.

Sriskanda, V. and Shuman, S.(2002b) Role of nucleotidyl transferase motif V in strand joining by *Chlorella* virus DNA ligase. *Journal Of Biological Chemistry*. **(277):** 9661-9669.

Srivastava, S.K, Dube, D, Tewari, N, Dwivedi,N, Tripathi, R.P and Ramchandran, R. (2005) *Mycobacterium tuberculosis* NAD^+dependent DNA ligase is selectively inhibited by glycosylamines compared to human DNA Ligase I . *Nucleic Acids Resarch* **(33):** 7090–7100.

Stoimenov, I. and Helleday, T. (2009) PCNA on the crossroad of cancer. *Biochemical Society Transactions* **(37):** 605–613.

Subramanya, H.S., Doherty,A.J., Ashford,S.R. and Wigely,D.B. (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell.* **(85):** 607-615.

Sun,D., Urrabaz,R., Nguyen,M., Martyn,J., Stringer, S., Cruz,E., Medina-Gundrum,L. and Weitman,S.(2001) Elevated expression of DNA ligase I in human cancer cells. *Clinical Cancer Research* **(17):** 4143-4148.

Sun, D., Urrabaz,R., Kelly,S., Nguyen,M. and Weitman,S.(2002) Enhancement of DNA ligase I by Gemcitabine in human cancer cells. *Clinical Cancer Research*. **(8):** 1189-1195.

Sun, D. and Urrabaz, R. (2004) Development of non-electrophoretic assay method for DNA ligases and its application to screening of chemical inhibitors of DNA ligase I. *Journal Of Biochemical And Biophysical Methods* **(59)**: 49-59.

Sun, Y., Seo, M.S., Kim, J.H., Kim, Y.J., Kim, G.A., Lee, J.I., Lee, J.H., and Kwon, S.T. (2008) Novel DNA ligase with broad nucleotide cofactor specificity from the hyperthermophilic crenarchaeon *Sulfophobococcus zilligii*: influence of ancestral DNA ligase on cofactor utilization. *Environmental Microbiology* **(10):** 3212-3224.

Takata, M., Sasaki, M.S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A. and Takeda, S. (1998) Homologus recombination and non-homologus end joining pathways of DNA double strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO journal***(17):** 5497-5508.

Takano, E., Maki, M., Mori, H., Hatanaka, M., Marti, T., Titani, K., Kannagi, R., Ooi, T., and Murachi, T. (1988) Pig heart calpastatin: Identification of repetitive domain structures and anomalous behavior in polyacrylamide gel electrophoresis. *Biochemistry* **(27):** 1964–1972.

Tan, G.T., Lee, S., Lee, I., Chen, J., Letiner, P., Besterman, J.M. and Pezzuto, J.M.(1996) Natural-product inhibitorsof human DNA ligase I. *Journal Of Biochemistry* **(314)**: 993-1000.

Tanese, N. (1997) Small-Scale Density Gradient Sedimentation to Separate and Analyze Multiprotein Complexes *Methods***(12):**224-234

Taylor, E.M. and Lehman, A.R. (1998) Conservation of eukaryotic DNA repair mechanisms. *International Journal Of Radiation Biology* **(74):** 277-286.

Taylor, R.M., Whitehouse, C.J. and Caldecott, K.W. (2000)The DNA ligase III zinc finger stimulates binding to DNA secondary structure and promotes end joining. *Nucleic Acid Research* **(28):** 3558-3563.

Tentori, L. and Graziani, G. (2005) Chemopotentiation by inhibitors in cancer therapy. *Pharmacological Research.* **(52):** 25-33.

Teraoka, H., Minami, H., Iijima, S., Tsukada, K., Koiwai, O. and Date, T. (1993) Expression of active human DNA ligase I in Escherichia coli cells that harbor a full-length DNA ligase I cDNA construct *Journal Of Biological Chemistry* **(32):** 24156-24162.

Timson, D.J., Singleton, M..R. and Wigley, D.B. (2000) DNA ligases in the repair and replication of DNA. *Mutation Research*. **(460)**: 301-318.

Tom, S., Henricksen, L.A., Park, M.S. and Bambara. R.A. (2001) DNA ligase I and proliferating cell nuclear antigen form a functional complex *Journal Of Biological Chemistry***(276):** 24817–24825.

Tomkinson, A. E., Totty, N. F., Ginsburg, M . and Lindahl, T. (1991) Location of the active site for enzyme-adenylate formation in DNA ligases. *PNAS (USA)* **(2):** 400–40.

Tomkinson, A.E., Lasko, D.D., Daly, G. and Lindahl, T. (1990) Mammalian DNA ligases. Catalytic domain and size of DNA ligase I *Journal Of Biological Chemistry*. (**265):** 12611–12617.

Tomkinson, A.E. and Levin, D.S (2005) Mammalian DNA ligases. *Bioessays* **(19):** 893 – 901.

Tomkinson, A.E. and Mackey, Z.B. (1998) Structure and function of mammalian DNA ligases. *Mutation Resarch*, **(407):** 1–9.

Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. and Ellenberger, T. (2006) DNA ligases: structure, reaction mechanism and function. *Chemical review*. **(106):** 687-699.

Tomimatsu, N., Tahimic,C.G.T., Otsuki,A., Burma,S., Fukuhara,A., Sato,K., Shiota,G., Oshimura.M., Chen,D.J. and Kurimasa,A. (2007) Ku70/80 modulates ATM and ATR signaling pathways in response to DNA double strand breaks. *Journal Of Biological Chemistry* **(282):** 10138-10145.

Tong, J., Cao, W. and Barany, F. (1999) Biochemical properties of a high fidelity DNA ligase from Thermus species AK16D. *Nucleic Acids Resarch* **(27):**788–794.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *PNAS* **(76):** 4350 – 4354.

Tseng, W.C. and Ho, F.L. (2003) Enhanced purification of plasmid DNA using Q-Sepharose by modulation of alcohol concentrations. *Journal Of Chromatography.* **(791):** 263-272

Tsuji, H., Ishii-Ohba, H., Katsube, T., Ukai, H., and Arzawa, S. (2004) Involvement of illegitimate V(D)J recombination or microhomology mediated nonhomologous end joining in the formation of intragenic deletions of the Notch I gene in mouse thymic lymphomas. *Cancer Resarch* **(64):** 8882-8890.

Turner, N., Tutt, A. and Ashworth, A. (2005) Targeting the DNA repair defect of BRCA tumours. *Current Opinion In Pharmacology* **(5):** 388-393.

Verkaik, N.S., Esveldt-van Lange, R.E., van Heemst, D., Bruggenwirth, H.T., Hoeijmakers, J.H., Zdzienicka, M.Z and VanGent, D.C. (2002) Different types of V(D)J recombination and end-joining

defects in DNA double strand break repair mutant mammalian cells. *European Journal Of Immunology* **(32):** 701-709.

Vijayakumar,S., Chapados,S.R., Schmidt, K.H., Kolodner, R.D., Tainer, J.A. and Tomkinson, A.E. (2007) The C-terminal domain of yeast PCNA is required for physical and functional interactions with Cdc9 DNA ligase. *Nucleic Acid Research* **(5):** 1624–1637.

Vijayakumar, S., Dziegielewska, B., Levin, D. S., Song, W., Yin, J., Yang, A., Matsumoto, Y., Bermudez, V. P., Hurwitz, J. and Tomkinson, A. E. (2009) Phosphorylation of Human DNA Ligase I Regulates Its Interaction with Replication Factor C and Its Participation in DNA Replication and DNA Repair. *Molecular and Cellular Biology* **(29):** 2042-2052.

Vivona, J. B. and Kelman, Z. (2003) The diverse spectrum of sliding clamp interacting proteins. *Federation Of European Biochemical Societies Letters.* **(546):** 167–172

Wang, S.P., Ho, C.K. and Shuman,S(1997) Phylogeny of mRNA capping enzyme.*PNAS (USA)* **(94):** 9573-9578.

Wang, W., Lindsey-Boltz, L.A., Sancar, A. and Bambara, R.A. (2006) Mechanism of stimulation of human DNA ligase I by the Rad9-Rad1-Hus1 checkpoint complex *Journal Of Biological Chemistry* (**281**):20865–20872.

Wang, H., Perrault, A.R., Takeda, Y., Qin, W., Wang, H. and Iliakis, G. (2003)Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acid Research* (31): 5377-5388.

Warbrick, E. (2000) The puzzle of PCNA's many partners. *BioEssays : news and reviews in molecular, cellular and developmental biology* **(22):**997-1006.

Weller, G.R., Kysela, B., Roy, R., Tonkin, L. and Scanlan, E. (2002) Identification of a DNA non-homologous end-joining complex in bacteria. *Science.* **(297):**1686–1689.

Weiss, B. and Richardson, C.C. (1967) Enzymatic breakage and joining of deoxyribonucleic acid, repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage. *PNAS (USA)* **(57):** 1021-1028.

Wilkinson, A., Day, J. and Bowater, R. (2001) Bacterial DNA ligases. *Molecular Biology* **(6):** 1241-1248.

Wyman, C., Ristic, D. and Kannar, R. (2004) Homologus recombination-mediated double strand break repair *DNA Repair* **(3):** 827-833.

Wyman, C. and Kannar, R. (2006) DNA Double-Strand Break Repair: All's Well that Ends Well. *Annual Reviews of Genetics* **(40):** 363-383.

Yamada, T., Onimatsu, H. and Van Etten, J.L. (2006) Chlorella viruses. *Advances in Virus Resarch* (**66):** 293–336

Yang, S.W., Becker, F.F. and Chan,J.Y.(1992) Identification of a specific inhibitor for DNA ligase I in human cells. *PNAS (USA)* **(89):** : 2227-2231.

Yang, S.W. and Chan,J.Y.(1992) Analysis of the Formation of AMP-DNA Intermediate and the Successive Reaction by Human DNA Ligases I and II. *Journal of Biological Chemistry* **(12):**8117-8122.

Yue, G.G., Fung,K.P., Tse, G.M., Leung, P.C. and Lau, C,B. (2006) Comparative studies of various ganoderma species and their different parts with regard to their antitumor and immunomodulating activities *in vitro*. *Journal of alternative and complementary medicine* **(8):** 777-789.

Zimmerman, S.B. and Harrison, B. (1986) Macromolecular crowding increases binding of DNA polymerase to DNA: An adaptive effect. *PNAS (USA)* **(84):** 1871-1875.

Zimmerman, S.B., Little, J.W., Oshinsky, C.K and Gellert, M. (1967) Enzymatic joining of DNA strands: A novel reaction of diphospyridine nucleotide *PNAS (USA)* **(57):** 1841-1848.

Zhao, A., Gray, F.C. and MacNeill, S. (2006) ATP and NAD^+-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Molecular Microbiology*, **59**, *3*, 743 – 752.

Zhong,S., Chen, X., Zhu, X., Dziegielewska, B., Bachman, K.E., Ellenberger, T., Ballin, J.D., Wilson, G.M., Tomkinson, A.E and Mackerell, A.D (2008) Identification and validation of human DNA ligase inhibitors using computer-aided drug design. *Journal of Medicinal Chemistry* **(15):** 4553-4562

Zhu, H. and Shuman, S. (2007) Characterization of *Agrobacterium tumefaciens* DNA ligases C and D. *Nucleic Acid Research* **(35):** 3631-3645