

# Chapter 1. Introduction

## 1.1 DNA-Protein Interactions

All cellular functions are initiated and controlled by proteins. DNA binding proteins are required for DNA replication, transcription, recombination, restriction, modification and repair. To perform their functions, these proteins must first find a DNA molecule, bind to it, and then locate their specific target site within that molecule.

Non-covalent interactions (electrostatic and hydrogen bonds) between exposed groups in the major and minor grooves of DNA and the active site of the protein, allow the protein to locate its specific target sequence. These proteins are often highly sequence specific and have two challenges in locating their target sequence: i) the kinetic challenge of finding the target sequence amongst a vast amount of DNA in a short space of time, and ii) the thermodynamic challenge of stability and specificity (Kampmann, 2004).

The rate at which proteins locate their target sequences is limited to the rates at which protein and DNA molecules move and collide. The rate at which they encounter each other by chance is by Brownian motion, and molecular diffusion is quantified by diffusion coefficients. The maximum rate of site location depends on the probabilities of the number of collisions between the two molecules ('the Einstein-Smoluchowski limit') (von Hippel *et al.* 1989). Studies of some DNA-binding proteins, such as restriction enzymes or the lac repressor, showed that the actual rate of target location within a DNA chain can be much faster than expected from random diffusion alone (Berg *et al.* 1985). These studies proposed that a protein is most likely to bind initially to non-specific sequences far away from the specific site, and then transfer along DNA to locate a target sequence. The protein therefore tests numerous nonspecific DNA locations until it finds its target sequence.

For a protein to locate its specific target sequence in a short space of time, the protein must have the ability to bind neither too tightly nor too weakly to non-specific sequences. The target sequence for any protein is very short in comparison to the entire genome. For example, a specific 10 base pair promoter sequence for a transcription factor would occur only three times within the genome of a typical bacteria like *E.coli*. Double-stranded DNA molecules generally exist in higher order conformational structures. The DNA can be highly supercoiled and compacted, or be present in a linear open reading frame conformation.

A DNA binding protein is unlikely to directly locate its target due to the high proportion of non-specific sequences. In 1970, Riggs *et al* (cited in Winter *et al.* 1981) noticed that the measured association rate of the *lacI* repressor for its operator sequence was much faster than the maximum rate achievable by diffusional collisions alone. Since then it has been accepted that DNA binding proteins locate their targets by a two-step mechanism. Firstly, at a given temperature, the protein diffuses in three dimensional space until it enters the targeting radius of a DNA chain, where it has a 50% probability of successfully colliding with non-specific DNA (Halford and Marko 2004). Secondly, the protein remains on or very close to the DNA chain and diffuses along the DNA towards the specific target site by intramolecular translocation events (Murugan, 2004).

## **1.2 Target site location**

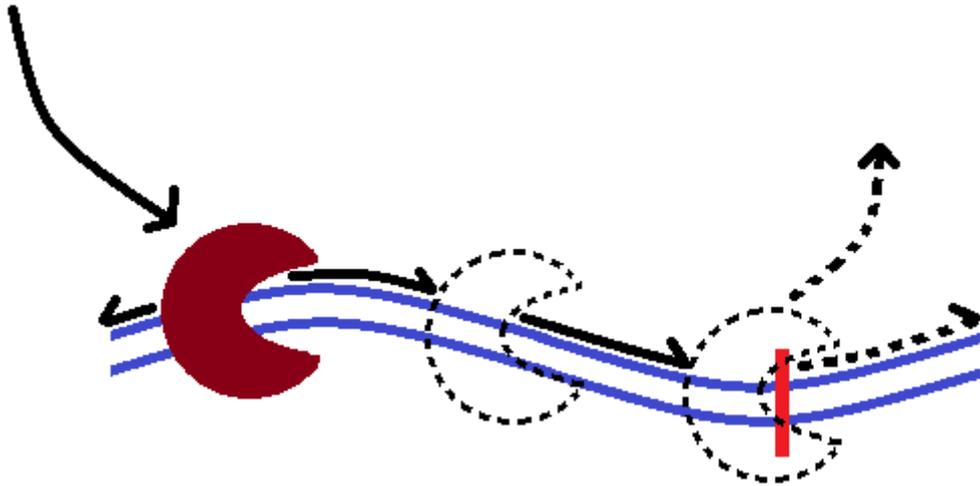
There are several mechanisms that a protein can adopt for intramolecular target site location. The first mechanism is simple random walk diffusion, in which the protein moves on and off the DNA molecule, randomly searching for its target site. The protein then tests this site and, if it is not the target site, completely dissociates from the DNA and continues searching. This is trial and error searching, and is also known as a macroscopic random diffusional search.

Another strategy that a protein might use for site location is facilitated diffusion, in which the protein does not completely dissociate from the domain of a DNA molecule, but instead uses the molecule as an aid to locate the target site. This mechanism involves dissociation and reassociation events and can happen by one of two possible routes: sliding or hopping. Figure 1.1 is a cartoon showing facilitated diffusion by one dimensional sliding or three dimensional hopping/jumping (Florescu and Joyeus 2010, Kolomeisky 2011).

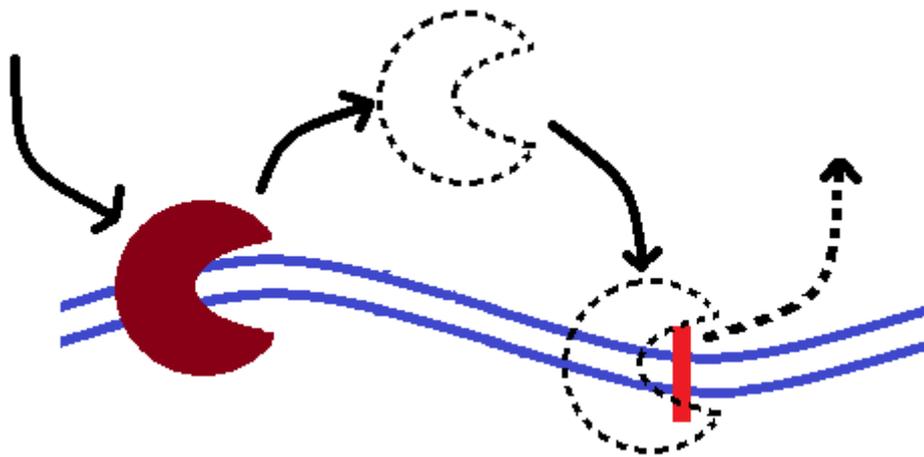
In one dimensional (1D) sliding (Figure 1.1A), the protein first collides with non-specific DNA. The protein remains non-specifically bound and slides along the linear contour of the DNA molecule, rarely losing direct electrostatic contact with the chain. By this method, protein tests adjacent sites in a strongly correlated manner – that is, thoroughly. The enzyme moves further and further from the initial landing point until the search is interrupted by finding the specific target site. Movement is achieved by following either the helical grooves of the DNA or the sugar phosphate backbone (Kampmann, 2004, Stanford *et al* 2000). It is thought that charge:charge interactions allow the protein to remain bound to non-specific sequences throughout translocation (Rau and Sidorova, 2010).

Studies in the 1980's assumed that DNA sliding was the main mechanism by which proteins found their sites (Fried and Crothers 1984). For example, the diffusion rate at which EcoRI was found to move along DNA was determined as  $3 \times 10^4$  bp/s, suggesting that this enzyme adopted a sliding technique to locate its target site. More recently, facilitated diffusion has been seen using fluorescent microscopy, confirming that the sliding mode is adopted by some enzymes (Biebricher *et al*, 2009). Two dimensional sliding is a similar mechanism to that of one dimensional sliding, but the protein is proposed to diffuse along the cylindrical surface of DNA. Diffusion is not restricted to following one of the duplex (Kampmann 2004). A variation of 1D or 2D

**A**



**B**



**Figure 1.1. Facilitated diffusion search mechanisms.**

**Panel A.** One dimensional sliding. The enzyme (red) locates DNA (blue) and binds non specifically before sliding in either a 5'-3' or 3'-5' direction, testing DNA sequences before locating a specific site (red bar) and subsequently dissociating from DNA.

**Panel B.** Three dimensional hopping. The enzyme (red) locates DNA (blue) and binds non-specifically, before then dissociating and reassociating with the same stretch of DNA. The enzyme remains within the DNA domain. This is repeated until a target is located or the enzyme diffuses a threshold distance away from the DNA (dotted arrow).

sliding is a correlated walk. This mechanism involves an asymmetric one dimensional walk along the DNA lattice. Each movement of the protein is determined by the energy correlation between the initial binding site and the final binding position. Positive correlations allow the protein to translocate the DNA towards the target site (Murugan, 2004).

In three dimensional (3D) hopping/jumping (Figure 1.1B), the protein once bound to the DNA undergoes multiple intradomain association and dissociation. The protein likely dissociates freely within the targeting radius in three dimensional space, and reassociates rapidly close by its initial binding site (around 4 bp away) on the same DNA strand (Halford and Marko 2004, Gowers *et al* 2005). Again, using fluorescent microscopy, a restriction enzyme (EcoRV) has been directly observed hopping along DNA whilst searching for its target site (Bonnet *et al.* 2008). It appears therefore that EcoRV and other DNA binding enzymes use a combination of both 1D sliding and 3D hopping to locate target DNA. This has been proposed to be true for several other enzymes (Porecha and Stivers, 2008; DeSantis *et al.*, 2011). If the distance of the target from the initial binding site is greater than 100 bp time will be wasted by 1D sliding (it will take too long for the protein to find its target site). But, if the distance is less than 100 bp then time will be wasted by 3D hopping. It is likely that proteins adopt one or both of these diffusion mechanisms (Gowers *et al.* 2005).

The third main mechanism of target site location is intersegment transfer (Hu and Shklovskii 2007). This mechanism is only possible for proteins with two DNA binding sites (e.g. the lac repressor). The protein binds to DNA via one of its active sites. As another segment of the same DNA molecule passes close to the protein, the remaining free active site on the protein binds to the second DNA chain, forming an intersegment loop. The protein binds both sites concurrently and is either exchanged

from one segment to the other or remains in its original position (Gowers and Halford 2003; Halford *et al*, 2004). This is not a mechanism relevant to the work in this thesis.

### **1.2.1 Experimental methods for investigating site location.**

Traditionally, kinetic studies are used to analyse the association rates of proteins with DNA. However these studies only provide information about rates of diffusion. They can be used to distinguish between random diffusion and facilitated diffusion. Such kinetic analyses involve length dependency analysis, and results in Chapter 4 in this thesis explore length dependency. These kinetic assays have been used mainly for restriction enzymes (Halford and Marko, 2004). Generally, the longer the DNA molecule the faster the initial association rate. However a longer DNA molecule will always provide a larger target for protein and it occupies a larger 3D volume. By varying the length of the flanking or antenna DNA evidence for facilitated diffusion can be derived by a further increase in rate. However, these studies do not provide information about the exact mechanism adopted by the protein.

Processivity studies allow for a better understanding of the mechanism that can be adopted by the protein. These experiments involve two target sites in the same DNA molecule (Gowers *et al*. 2005). Processivity can reveal information about the spatial pathways to target sites, by measuring the proportion of reactions in which a protein visits the two sites within one binding event. This is generally done using DNA substrates whose products can be separated by PAGE. For example Stanford *et al* 2000, used the change in processivity ( $fp$ ) of EcoRV when the distance between two EcoRV sites was varied. The manner in which processivity decreased as the intersite distance increased was mathematically better fit by the theory of 3D hopping than by 1D sliding (Stanford *et al* 2000). A second form of processivity experiments from the Halford group studied the restriction enzyme BbvCI. When two asymmetric target sites were

directly repeated (that is they followed head to tail) the targets could be located by BbvCI in either a 1D sliding or 3D hopping mode. However, when two target sites are inverted (that is head to head), the protein was able to show a similar level of processivity to the directly repeated substrates. This indicated that the enzyme was able to sample sites on both strands of the DNA, corresponding to a 3D hopping/jumping mechanism with some evidence of local 1D sliding. Chapter 5 of this thesis explores processivity measurements on DNA Ligase.

An alternative experiment for target site location was carried out with plasmids and catenanes (Gowers and Halford 2003). The DNA used in this method was of three forms: supercoiled plasmid, supercoiled catenane (a two-ring form of the plasmid) and supercoiled minicircle (one ring of the catenane linearised). These three substrates were incubated with EcoRV in pairwise competitions. For example plasmid versus catenane showed whether EcoRV was able to move between catenane rings. In this competition, the single target site in the plasmid could be reached by 1D or 3D mechanisms. However, the single site in the small catenane ring could only be efficiently cut by 3D jumping pathways (1D sliding would never reach the target site). The initial rate of plasmid and catenane cutting were found to be nearly identical and therefore EcoRV must have used 3D pathways on route to its target site.

Experimental analyses of location mechanisms by proteins have been focused on *in vitro* studies. These studies do not account for the differences of the natural environment within a cell. DNA in the cell undergoes many conformational changes and can be complexed with many bound proteins. It cannot be determined *in vitro* whether roadblocks on DNA (transcribing polymerases, bound transcription factors etc) affect the diffusion pathways of sequence- or structure-specific proteins. For example, do protein roadblocks prevent long range 1D sliding? *In vivo* studies need to be carried

out to determine if the cell and nuclear environment have an effect on site location (Hannon *et al.*, 1986; Kampmann, 2007).

### **1.3 DNA Ligases**

DNA is damaged every day by cellular processes, by external environmental factors such as UV radiation, and can also occur spontaneously by oxidative damage. Nicks and double stranded breaks are also made in DNA due to processing of spontaneous lesions and as products of lagging-strand Okazaki fragments during replication. Cell death can occur if DNA damage is not repaired. Therefore DNA Ligases are essential enzymes for maintaining genomic integrity in cells. Their roles in nucleic acid metabolism include DNA replication, recombination and repair (Fabre and Roman, 1979; Wei *et al* 1995; Mossi *et al.* 1998).

DNA Ligase was first discovered in 1967 by the Lehman laboratory and other groups (Olivera *et al.*, 1967; Gefter *et al.*, 1967; Gellert, 1967, Zimmerman, 1967). DNA Ligases are present in all organisms (eukaryotes, prokaryotes and viruses). They catalyse the re-formation of a phosphodiester bond where DNA damage has occurred. (Lehman 1974). There are two classes of DNA Ligase; those that use ATP as a cofactor (Cheng and Shuman, 1997; Jackson *et al.*, 2007), and those that use NAD<sup>+</sup> as a cofactor (Shuman and Benarroch, 2006; Wilkinson *et al.*, 2003; Lee *et al.* 2000).

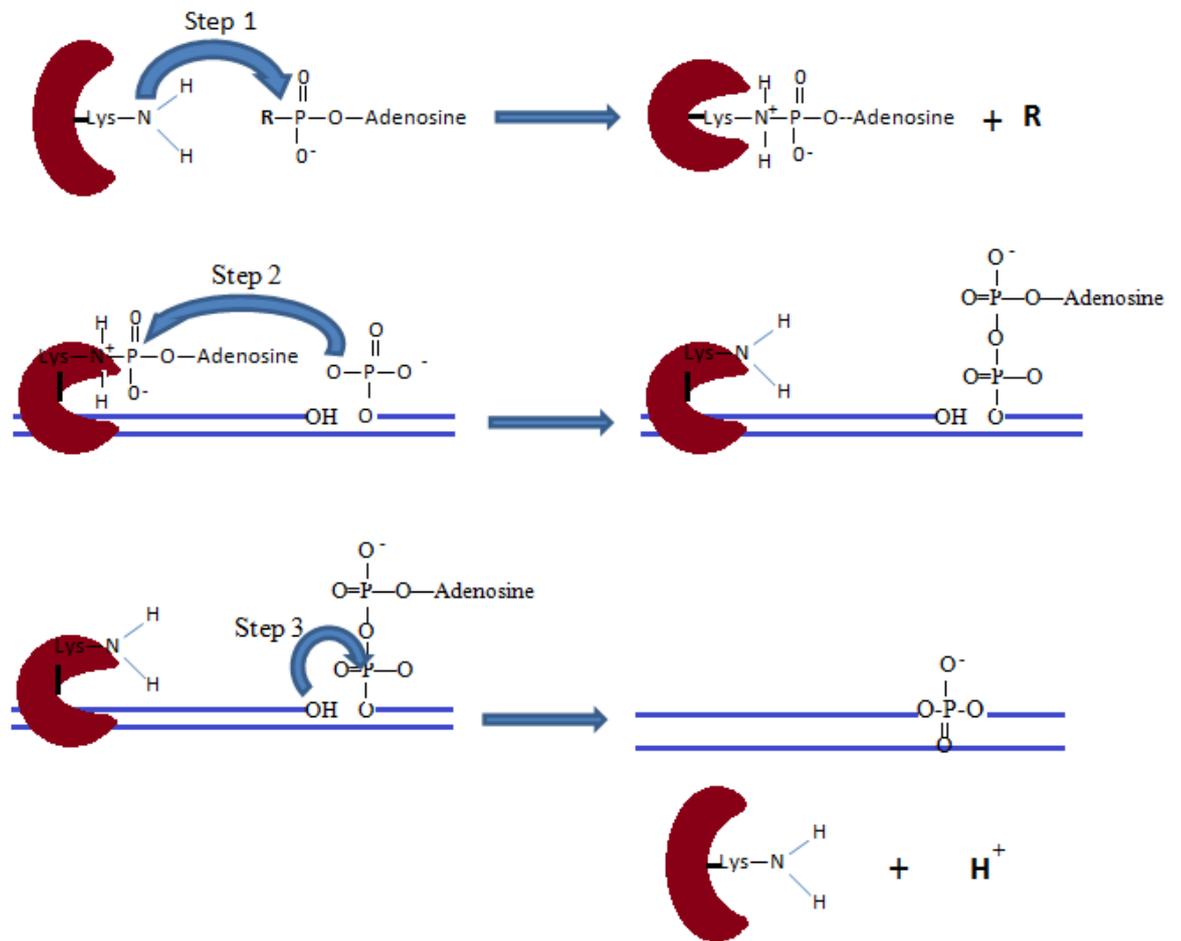
DNA Ligases vary greatly in size from organism to organism. ATP Ligases range from 30 to greater than 100 kDa, whilst NAD<sup>+</sup> Ligases are more conserved in size, ranging from 70 to 80 kDa. NAD<sup>+</sup>- Ligases have extensive sequence conservation – 35%-50% conservation between bacterial species. There is little sequence conservation between NAD<sup>+</sup>- and ATP- dependent Ligases at the amino acid sequence level, but their tertiary structures are similar (Doherty and Suh, 2000). To date, NAD<sup>+</sup> Ligases have not been found in eukaryotes. The difference in cofactor specificity make

NAD<sup>+</sup>-dependent Ligase a good potential target for producing a novel class of antibacterial drugs. Various chemical compounds have already been studied as possible inhibitors of the NAD<sup>+</sup> binding site in Ligase (Ciarrocchi *et al.*, 1999; Brotz-Oesterhelt *et al.*, 2003; Dwivedi *et al.*, 2008). High throughput screening has shown that pyridochromanones inhibit NAD<sup>+</sup>- Ligase specifically, but do not affect ATP- Ligase. However, further information is needed about NAD<sup>+</sup> recognition and binding by Ligase before a suitable antibacterial drug can be produced (Georlette *et al.*, 2003).

DNA Ligases belong within the superfamily called nucleotidyl transferases. Other binding proteins in this family include GTP-dependent mRNA capping enzymes and ATP-dependent RNA Ligase. These enzymes all have a common domain called the nucleotidyl transferase domain, which has six conserved motifs II, III, IIIa, IV, V and VI (Sriskanda *et al.* 2002; Shuman *et al.* 2004; Zhu *et al.* 2005). The transferase portion was first mapped in a viral mRNA capping enzyme (Fausnaugh and Shatkin, 1990). Residues within these conserved motifs are essential for enzyme function (Cong and Shuman, 1993).

### **1.3.1 Ligase Function**

DNA Ligases repair breaks in the phosphodiester backbone of DNA. As mentioned Ligases are essential for many cellular processes; DNA recombination, DNA replication and DNA repair (Billen *et al.*, 1975; Montecucco *et al.*, 1988). In all these processes, Ligase is required to join the 3'hydroxyl (3'OH) group on one end of a DNA chain, to the 5'phosphoryl (5'P) group on the other side of the nick. Bond formation is likely achieved by three separate reactions all catalysed by a single Ligase molecule (Lehman 1974). Figure 1.2 shows the three proposed stages of ligation, and these are described below:



**Figure 1.2. DNA Ligase three-step reaction mechanism.**

Step 1: Ligase (red) reacts with cofactor (ATP or NAD<sup>+</sup>) transferring AMP to form Ligase-adenylate intermediate and inducing a closed conformation of the enzyme. R represents the second product of Ligase reaction: PPi (ATP Ligase) or NMN (NAD<sup>+</sup> Ligase).

Step 2: AMP is transferred from Ligase-adenylate to the 5' P at the nick site.

Step 3: The 3' OH attacks the 5' P adenylated DNA, creating a phosphodiester bond, releasing the enzyme.

Firstly, Ligase covalently reacts with a suitable cofactor (NAD<sup>+</sup>/ATP). The 5<sup>th</sup> carbon in a lysine residue (the ε-carbon) attacks the cofactor, forming a covalent adenylated Ligase-(lysyl-N)-AMP intermediate. ATP-dependent Ligases release pyrophosphate (PP<sub>i</sub>) after the AMP is transferred, whilst NAD<sup>+</sup>-dependent Ligases release nicotinamide mononucleotide (NMN). The binding of the cofactor induces conformational rearrangement within the active site of Ligase (Georlette *et al.* 2003), in which it becomes catalytically active and adopts a more compact structure (Cherepanov and Vries 2002, Georlette *et al.* 2004).

When Ligase is adenylated it becomes catalytically active and is then able to bind DNA. This compaction method is known as the 'open-close' process; before Ligase is adenylated, the DNA-binding face of the Oligomer Binding domain (OB; described in the next section) is rotated away from the active site. This is the open conformation of the enzyme (Georlette *et al.* 2004). Once adenylated, the OB domain changes conformation and rotates around exposing the DNA-binding face towards the active site. This is the closed conformation (Murzin *et al.*, 1993, Lee *et al.*, 2000, Kaczmarek *et al.* 2001; Lavesa-Curto *et al.* 2004). The change in conformation is a switch that allows only the adenylated, active Ligase to bind to DNA (Georlette *et al.*, 2003).

Secondly, the 5'P at the nick attacks the phosphoryl group of the Ligase-AMP intermediate, so that the AMP moiety is transferred from the ε-lysine to the 5'P. This forms a bisphosphate intermediate on the 5' side of the nick. Much less is known about this step of the reaction in terms of conformational change or mechanism.

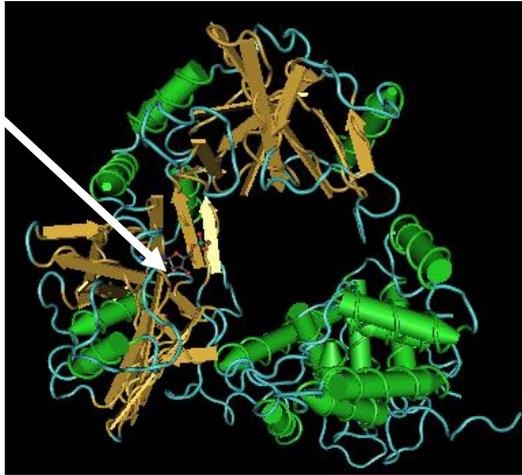
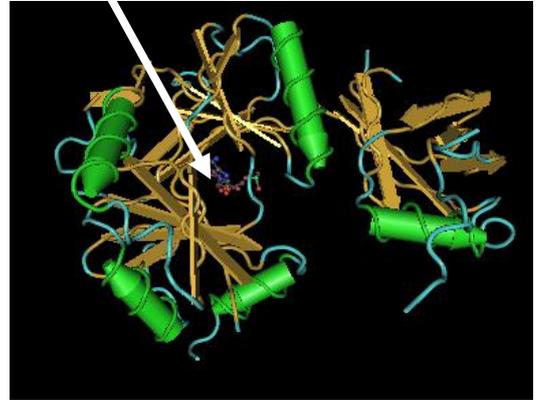
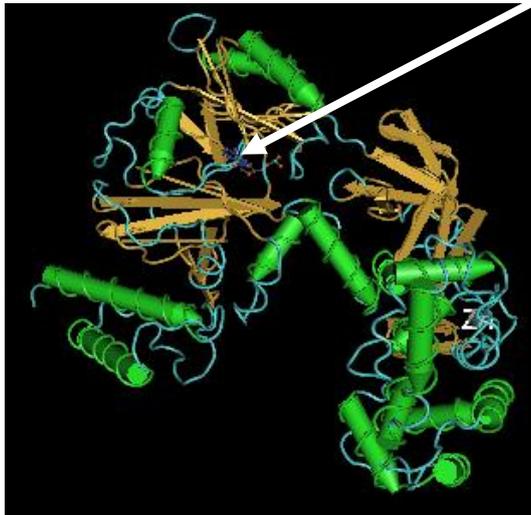
Finally, Ligase catalyses a nucleophilic attack by the 3'OH on the high energy bisphosphate 5' bond. The two polynucleotide strands are joined, forming a phosphodiester bond in the DNA backbone, and AMP is released into solution. Ligase is then free to become re-adenylated for another round of ligation (Lehman 1974).

### 1.3.2 Ligase Structure

Most DNA Ligases have a similar tertiary structure and share common domains. It is thought that all Ligases came from a common ancestor and therefore it is not surprising that they share common tertiary structures. Several crystal structures have been determined from different species, and Figure 1.3 shows the monomeric form of three of these: Human DNA Ligase I, *Chlorella* virus and *Thermus filiformis* (Lee *et al* 2000, Odell *et al* 2000, Pascal *et al* 2004). These structures were derived from the Protein Data Bank (PDB) using PDB-IDs of 1X9N, 1FVI and 1DGS and visualised with the PubMed Cn3d viewer.

Studies of several crystal structures have determined that they all contain a catalytic core that consists of two parts: the adenylation domain (also known as the nucleotidyl transferase or NTase domain), and the Oligomer Binding (OB) fold (Lee *et al.* 2000; Georlette *et al.* 2004). Further domains, for example the Ia sub-domain, the zinc finger (Zn) motif, the Helix-hairpin-Helix (HhH) motif and the BRCT domain may be present or absent in a DNA Ligase. Figure 1.3D shows the domain structure of the three Ligases above. A further domain classification is used in the literature as follows: Domain 1 consists of Ia and the Ntase domain, Domain 2 is the OB fold, Domain 3 consists of the Zn and the HhH motifs and Domain 4 is the BRCT domain. The function of each domain is outlined below.

The adenylation/NTase domain contains the lysine residue essential for cofactor binding (Singleton *et al.* 1999, Zhu *et al.* 2005). This lysine must be adenylated for Ligase to bind to DNA and is in a solvent accessible hydrophilic pocket (Shuman and Lima, 2004; Sriskanda and Shuman, 2002; Wilkinson *et al.*,2005). The NTase domain also contains the sequence motifs I-V (motif VI is present in the OB fold domain). These five motifs line the nucleotide-binding pocket and contribute amino acid side chains that are essential for Ligase catalytic activity. The secondary structure of the

**A****B****C****D**

Human DNA Ligase I

N	DBD	NTase	OB
---	-----	-------	----

*Chlorella virus Ligase*

N	NTase	OB	Zn	HhH	BRCT
---	-------	----	----	-----	------

*Thermus filliformis Ligase*

NTase	OB	LM
-------	----	----

**Figure 1.3. Known crystal structures of three representative DNA Ligases.**

**Panels A, B and C:** Views of the tertiary structure of Human DNA Ligase I (A), *Chlorella Virus Ligase* (B) and *Thermus filliformis Ligase* (C), from PDB-IDs 1X9N, 1FVI and 1DGS. Alpha helices are shown as green cylinders beta sheets as gold arrows and unstructured linker sequences in blue. The white arrow shows the AMP group bound to the reactive lysine. Therefore these structures show the Ligases adenylated prior to the second stage of catalysis as described in the text. White arrows indicate bound AMP.

**Panel D:** Representation of domains of Ligases in Panel A, B and C. N-terminal (N), DNA binding domain (DBD), Nucleotidyl transferase domain (NTase), Oligomer binding fold (OB), Zinc finger motif (Zn), Helix hairpin helix (HhH), BRCT domain, Latch module (LM). Diagram is not to scale.

adenylation domain consists of anti-parallel  $\beta$ -sheets surrounded by  $\alpha$ -helices, though the structure varies in different species. The NTase domain is essential for binding the cofactor (NAD<sup>+</sup> or ATP). Alanine scanning mutational studies have identified various other amino acid residues in the adenylation domain that are essential for Ligase function (Singleton *et al.*, 1999).

The OB fold (Domain 2) is attached to the adenylation domain. This domain is required for polynucleotide recognition in order for the Ligase to recognise and bind DNA. Crystal structure studies have shown that Ligase uses the OB fold to bind around DNA (Lee *et al.*, 2000). Domains 1 and 2 are present in all Ligases, probably because they play such an essential role for Ligase activity. In NAD<sup>+</sup>- Ligases these domains are also flanked by other domains: Ia, Zn, HhH and BRCT domain.

The Ia domain is a subdomain of Domain 1. It is a short N-terminal region with a secondary structure that is mainly  $\alpha$ -helical. It is essential for the reaction of Ligase with NAD<sup>+</sup> and therefore is required for Ligase adenylation. Alanine substitutions of residues within Ia domain of *E.coli*LigA have identified five amino acid residues that are involved specifically in adenylate transfer: Y22A, H23A, D32A, Y35A and D36A (Sriskanda and Shuman, 2002a). These residues are conserved in thirty other bacterial Ligases, and may be involved with the NMN binding site.

The OB fold is followed by three C-terminal domains: the Zn and the HhH domains (Domain 3) and the BRCT domain (Domain 4) (Doherty and Suh, 2003). The Zn motif consists of four cysteine residues that co-ordinate a zinc ion. This structure is conserved among all known bacterial NAD<sup>+</sup>- Ligases. This domain is required for recognition of the nick site and DNA interaction. With aid of the HhH, motif the Zn motif forms a complex with double-stranded DNA (Lavesa-Curto *et al.* 2004). The HhH motif consists of two  $\alpha$ -helices connected by a type II  $\beta$ -turn. The hairpin of the HhH motif is highly positive and is thought to be required for non-specific DNA-binding by

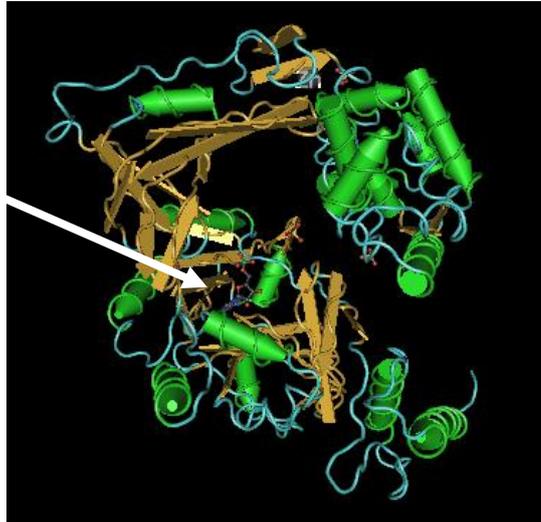
the Ligase (Doherty and Suh, 2003). Domain 4 is called the BRCT domain and is found in all NAD<sup>+</sup>- and some ATP- Ligases. The BRCT domain was originally identified at the C-terminus of the protein encoded for by the breast cancer susceptibility gene *BRAC1*. Its secondary structure consists of parallel  $\beta$ -sheets and  $\alpha$ -helices. It is about 80-100 amino acids in length and has a hydrophobic nature to the fold. As the BRCT domain is not present in all Ligases it is not considered to be an essential component to Ligase activity (Timson and Wigley, 1999). However, deletion analysis indicates that removing the BRCT domain, together with Domain 3, eliminates Ligase activity (Feng *et al.*, 2004). Domains 3 and 4 are not essential for Ligase adenylation.

### 1.3.3 EcoLigA

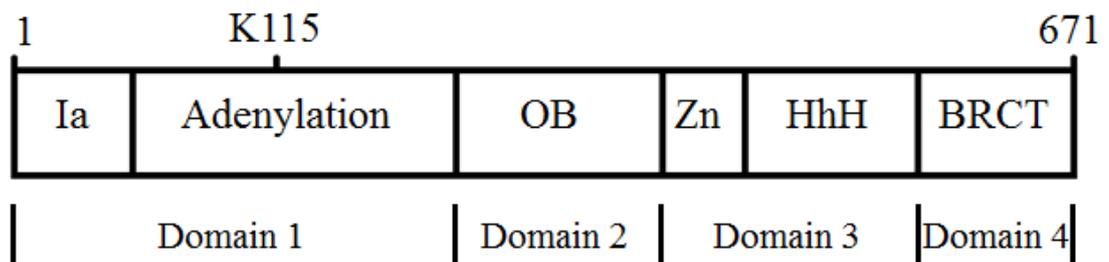
*Escherichia coli* Ligase A (*EcoLigA*) is a NAD<sup>+</sup>-dependent Ligase and is the product of the gene *LigA*. It was the first Ligase to be discovered (Olivera and Lehman 1967) and since then has been widely studied because every bacterial genome that has been sequenced so far encodes this Ligase. Inactivation of the *LigA* gene causes non-viability in most bacteria (Kaczmarek *et al.* 2001). The coding frame of the gene was confirmed to be correct by the determination of the amino acid composition of the purified Ligase product. *EcoLigA* is 671 amino acids in length and has a molecular weight of 73,690 Da (Ishino *et al.* 1986). In comparison to *EcoligA*, conservation of sequences between bacterial species is between 35% and 50%.

In 2007, the Shuman group published a 2.3 Å crystal structure of *EcoLigA* bound to adenylylated DNA (Nandakumar *et al.*, 2007). This is shown in Figure 1.4. It can be seen that *EcoLigA* adopts a closed conformation when bound to DNA as previously observed for human DNA Ligase I (Pascal *et al.*, 2004). Mutational substitutions and deletions have identified essential amino acid residues required for

**A**



**B**



**Figure 1.4 Structure of *Escherichia coli* Ligase A.**

Panel A: Crystal structure of *Escherichia coli* LigA bound to an adenylated nick (DNA not shown), adapted from Nandakumar 2007, using PDB-ID 2OWO. The adenylated Ligase is at the second step of the Ligase reaction, prior to transfer of AMP to the 5'P at the DNA nick. The white arrow shows AMP in the binding pocket of Ligase.

Panel B: Domain structure of *EcoLigA*. The essential lysine, K115, is located within the adenylation domain of Domain 1. The other domains are named according to the text.

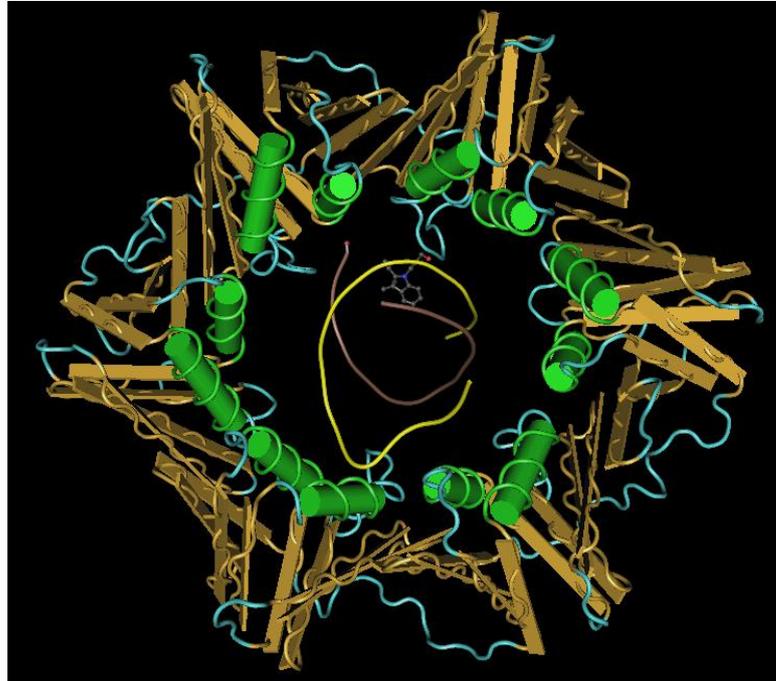
Ligase function (Timson and Wigley, 1999, Wilkinson *et al.*, 2005). Lysine residue 115 is the  $\epsilon$ -lysine required for Ligase adenylation and activation. In terms of domain organization, alanine substitutions within the Ia domain of *EcoLigA* either reduced (at His23 or Tyr35) or completely abolished (at Tyr22, Asp32 or Asp36) the Ligase nick-sealing function (Sriskanda and Shuman, 2002b). These amino acid side chains are required for NMN binding. Deletion of the BRCT domain also reduced *in vitro* ligation by up to 3-fold (Wilkinson *et al.* 2003). Deletion mutagenesis in the same paper identified that the BRCT domain is required for the nick joining reaction.

#### **1.4 Beta-clamp and gamma clamp-loader proteins.**

During replication, DNA polymerases are required to be highly processive enzymes, since they have to faithfully copy many millions of base pairs of DNA. Diffusion alone could never account for the high degree of processivity observed in DNA polymerases (Jeruzalmi *et al.*, 2002). The main factor which massively increases processivity at all replication forks is a sliding clamp whose purpose is to keep polymerases at the replication fork. In bacteria, this processivity factor is called the beta sliding clamp, and allows DNA PolIII to function for 1000's of base pairs rather than 10's, without dissociating from the DNA (Stukenberg *et al.* 1991). The beta clamp comprises a highly stable homodimeric ring, consisting of two identical monomers form a continuous ring that are head to tail in conformation (Kong *et al.*, 1992).

Each monomer consists of three domains as shown in Figure 1.5. When in homodimeric form, the outer surface comprises of 6 beta sheets which support a matrix of 12 alpha helices. The alpha helices provide a favourable electrostatic surface between the clamp and DNA, allowing the clamp to slide freely along the DNA chain. The beta clamp needs to be loaded onto the DNA. *In vivo* this action is catalysed by the gamma clamp-loader complex (Onrust *et al.*, 1995). The clamp-loader is a 5-subunit complex consisting of subunits  $\delta$  and  $\delta'$ , and three  $\gamma$  subunits (Jeruzalmi *et al.*, 2001a).

A



**Figure 1.5 Structure of *Escherichia coli* beta-clamp bound to DNA**

**Panel A:** Crystal structure of *Escherichia coli* beta-clamp. This structure was solved in the presence of DNA shown in the centre of the clamp. Adapted from Georgescu 2008, using PDB-ID 3BEP. Alpha helices are shown as green cylinders, beta sheets as gold arrows and unstructured linker sequences in blue.

A fluorescent-based clamp opening assay was used to determine that the clamp loader first binds to the clamp and then opens it (Paschall *et al.*, 2011). The  $\delta$  subunit is essential for opening the beta clamp and loading it onto DNA. The beta-clamp is forced open by the loader in a spring-loaded type mechanism (Jeruzalmi *et al.*, 2001b).

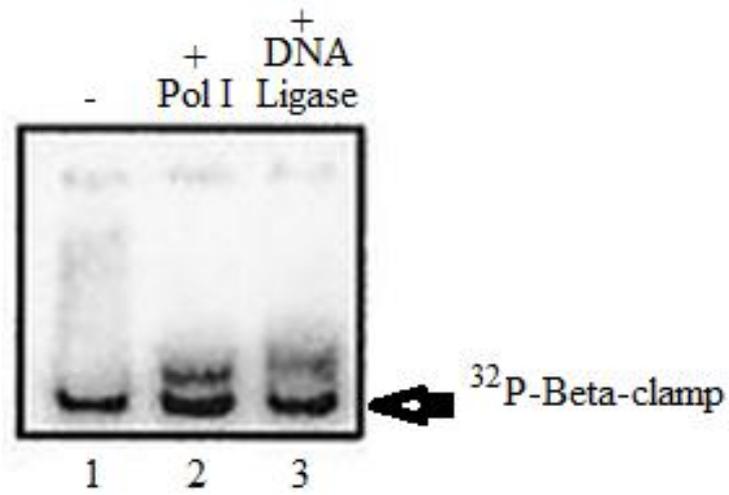
The beta clamp structure was solved directly bound to DNA in 2008 by the O'Donnell group (Georgescu *et al.* 2008). This structure showed that DNA passes through the beta clamp like a curtain ring on a pole, at a 22° angle.

### **1.5 Purpose of this work.**

The main aim of the work in this thesis was to bring together some of the experiments on target site location described earlier, to ask the question how does DNA Ligase locate a nick in a long DNA molecule. To address this problem a series of order-of-addition and cofactor studies were undertaken to firstly determine the order in which NAD<sup>+</sup> and nicked DNA were used by the enzyme (Chapter 3). This allowed length dependency (Chapter 4) and processivity (Chapter 5) experiments to be undertaken to identify the pathway that Ligase took when moving between nicks. The results of this work indicated that a 3D mode was operating but the processivity was lower than restriction enzyme processivity at around 32-35%.

Interestingly, in 2001 the O'Donnell group reported for the first time that *EcoLigA* interacts with the beta-sliding clamp (Lopez and O'Donnell 2001). They used a radioactively tagged form of beta-clamp and discovered a protein-protein interaction with unlabeled DNA Ligase. No nicked DNA was present. Figure 1.6 shows this single report of the interaction between beta-clamp and Ligase. No group since then has yet published a report on the kinetic mechanism of this observation. We therefore extended our processivity studies for preliminary experiments with the beta-clamp, gamma-loader and Ligase. These results are shown at the end of Chapter 5.

**A**



**Figure 1.6. Interaction between beta-clamp and Ligase.**

A gel shift on a polyacrylamide gel. Lane 1 is beta-clamp alone. Lane 2 is beta-clamp with addition of 3.5  $\mu\text{M}$  polymerase. Lane 3 is beta-clamp with addition of 3.5  $\mu\text{M}$  DNA Ligase. Adapted from Lopez 2001.