Biochemical analysis of bacteriophage Orf family recombinases

IAIN DAVID WALLACE

Table of Contents

Chapter 1

Introduction

1.1 Summary

Genetic recombination is a vital feature of life on earth; a ubiquitous and essential damage-repair mechanism and a potent driving force behind evolution. In the case of pathogenic bacteria, struggling against the harsh selective pressures of an adaptive immune system and potent antibiotic chemotherapeutics, the need for effective and rapid adaptability is high. The bacterial armory for procuring pathogenic variants is diverse, including plasmid uptake, transposons, homologous gene swapping and the molecular switching of antigens. Temperate bacteriophages provide another avenue for adaptation through the direct transfer of genetic material between and within species through transferring useful genes from host to host or by causing recombination events that generate new gene combinations. The study of phage and bacterial recombination pathways is critical for understanding the way in which new pathogens emerge and how multiple drug resistance arises. This chapter aims to summarise the contribution of bacteriophages to bacterial evolution, the homologous recombination pathways of bacteriophage λ and *Escherichia coli* as well as introducing the Orf family of proteins.

1.2 Bacteriophages

Bacteriophages, or phages, are viruses that infect bacteria. Their name literally means “Eaters of Bacteria” (from the Greek *phagein:* “to eat”) due to the clear plaques seen on suitably infected bacterial lawns. They are the most abundant entity in the biosphere by a large margin, outnumbering their bacterial hosts by a factor of 10 (1) and may even have total numbers in excess of 10^{30} – outnumbering the totality of all other life forms (2). Phages were arguably first reported by Ernest Hankin in 1896 in water samples taken from the rivers Ganges and Jumna in India (3). The phenomenon of phages transferring pathogenically relevant information is well documented; examples include the neurotoxin produced by *Clostridium botulinum* and the causative toxin of diphtheria (4). In fact, phages are ubiquitous in the prokaryotic world and are responsible for a large proportion of the inter-strain

differences within bacterial species (4). Furthermore, prophages can be found in two-thirds of the γ-proteobacteria and low G+C gram-positive bacteria, where they can constitute up to 20% of the total genome (5).

1.2.1 Lysis, lysogeny and lysogenic conversion

Temperate bacteriophage infection of a bacterial host yields either a lytic or lysogenic infection (6). The infection path depends on the type of phage, the prevailing environmental conditions and on the type of bacterium (7), although, given the number of integrated prophages in bacterial genomes, lysogeny is more likely the rule than the exception. A lytic infection does not require phage DNA integration, only DNA injection into the host, phage protein synthesis, DNA replication, capsid formation and virion release (8).

During lysogeny, however, phages express genes that lead to the integration and repression of the phage genome, where the integrated genetic material is termed the prophage (6). The transcriptional repression of the prophage provides the host bacteria with lytic immunity to the phage that they carry; that is, they do not lyse upon infection by other phage particles of the same type (8). Studies of bacterial genomes have revealed that lysogeny is likely the most common outcome of phage infection, with many bacteria containing multiple prophages; such as *Escherichia coli* K-12 with its 9 assimilated, cryptic, prophages contributing to 3.6% of the total genome (1).

Phages have the capacity to rearrange their genomes at an accelerated rate, allowing the acquisition of genes with potentially valuable functions for both phage and recipient bacteria; such genes are termed morons or lysogenic conversion genes (LCGs) (4). LCGs increase the reproductive fitness of the phage indirectly by contributing positively to the survival of the host, for example by providing a protein or toxin that allows the host to outcompete its uninfected peers or competitors (4). Alternatively, LCGs can provide advantages in the face of adverse environmental conditions; whether it be molecular weaponry or resistance to oxidative stress, acidic

conditions or osmotic stress (like many of the cryptic prophages found in the previously mentioned *E. coli* K-12 study) (1).

1.2.2 Bacteriophage λ

Bacteriophage λ infects the Gram-negative prokaryote *Escherichia coli* and related Enterobacteria (9). Given the status of *E. coli* as the pre-eminent prokaryotic model organism, observing recombination within λ is a rational strategy for investigation of temperate phage recombination. λ has a 48,502 base pair genome encoding about 67 proteins (10). Phages with high degrees of similarity to λ are termed the lambdoid phages – specifically defined as phages able to form productive recombination with λ (2). Most of these are infectious agents of *E. coli*, although a small number also infect *Salmonella enterica* Serovar Typhimurium (11). As with other phages, λ displays a significant degree of mosaicism; on comparing two lambdoid genomes, areas of sequence homology will be evident, with transition to areas of marked heterology (2) caused by horizontal exchange events.

Like many other bacteriophages, λ has its own recombination system, the Red pathway, to promote genetic exchanges important for DNA repair, replication and generating genetic variability (12). Phage-stimulated recombination events are responsible for a number of phenomena: the formation of branched DNA junctions as a consequence strand exchange or annealing (usually occurring at a break in the genome), the reshuffling of phage genes and the reshuffling of host genes and integration of lysogenic conversion genes.

Recombination of any DNA results in the formation of atypical DNA structures. In order to resolve 3-stranded (D-loops) and 4-stranded (Holliday junctions) structures, incisions need to be made to allow separation and reformation of the two duplex DNA strands. Failure to separate joint molecules results in the blockage of phage DNA packaging (13) and resolution of such artifacts is vital for allowing transcription and replication

1.3 Homologous Recombination

Recombination was defined by Clark in 1971 as "any of a set of pathways in which elements of nucleic acid interact with a resultant change of linkage of genes or parts of genes" (14). Homologous recombination involves the physical interaction of two similar DNA double-helices through the exchange of single strands (15) or the invasion of single stranded DNA into an intact duplex (16) and often occurs after interruption, breakage or collapse of a replication fork (17). Strand transfer at these sites allows an intact chromosome to provide a template for the damaged chromosome by initiating DNA synthesis. Migration of initial 3-stranded intermediates leads to the formation of a 4-branched DNA junction known as a Holliday junction. Holliday junctions are subsequently resolved by a family of structure-specific endonucleases, returning the recombined DNA into nicked duplexes that can be sealed by DNA ligase. Recombination is intimately connected to the process of DNA replication – in fact the very existence of systems of homologous recombination are necessitated because DNA replication stalls upon encountering DNA breaks. Homologous recombination allows the cell to effect repairs and permit replication continue.

Homologous recombination does not just act to preserve genomic integrity – paradoxically it can also promote variability (18). For example, if a chromosome that has suffered a mutation is used in the process of homologous recombination to repair a damaged but previously non-mutated chromosome, then that mutation would be transferred to the repaired chromosome. In addition, various rearrangements can occur at legitimate or illegitimate locations by crossover or splice recombinational exchanges that may also involve incoming DNA from other sources (18).

Specific groups of proteins are able to initiate, manipulate and terminate homologous recombination (19). Those that function to execute homologous recombination in *E. coli* and λ will be discussed next.

1.3.2 Recombination in *Escherichia coli*

Recombination in *E. coli* occurs through two main pathways – one initiated by RecBCD at dsDNA ends, the other by RecFOR at ssDNA gaps. Both pathways involve the processing of damaged DNA regions to facilitate assembly of RecA for homologous pairing and strand exchange and ultimately the formation of 4-stranded Holliday junction intermediates that are processed and removed by RuvABC or RecG. Components of these recombination pathways in *E. coli* are shown in Table 1.1. These proteins and their functions will be discussed first followed by a description of the current understanding of how RecBCD and RecFOR bring about recombination exchanges important for DNA repair, replication and recombination.

Protein	Main Activity
RecA	Homologous pairing, strand exchange (formation of joint molecules)
RecBCD	Chi site-modulated 5'-3' dsDNA exonuclease, initiation of repair at double-strand breaks, loading of RecA at 3' ssDNA tails
RecF	Binds single- and double-stranded DNA, together with RecR limits extension of RecA onto adjacent dsDNA
RecO	Binds SSB-coated ssDNA, together with RecR helps stabilise RecA filaments; helps RecA load onto SSB-coated DNA and possibly also assists in homologous pairing
RecR	Binds dsDNA; RecFOR together are thought to facilitate RecA loading onto SSB-coated DNA
RuvAB	DNA branch migration helicase, unwinding joint molecules
RecG	Branch migration DNA helicase; unwinds R loops
RuvC	Holliday junction endonuclease, resolution of joint molecules in combination with RuvAB
DNA polymerase II	DNA polymerase Replication-restart at template lesions
DNA polymerase I, RecJ, SbcB, SbcBD, RecQ	DNA polymerase, 5'-3' exonuclease Gap sealing
DNA ligase	DNA ligase; seals 5'-P and 3'-OH nicks in DNA
SSB	Single-stranded DNA-binding protein, binds single-stranded DNA for replication and loads multiple DNA replication, repair and recombination proteins onto ssDNA

Table 1.1: Components of recombination pathways in *E. coli*. Adapted from Cox et al. (2000) **(20)**

RecA

RecA is a 37,842 Da, 353 amino acid protein with a diverse range of activities that include ssDNA binding, ssDNA-dependent ATP hydrolysis, homologous pairing, strand exchange and an inducer of LexA autoproteolysis as part of the SOS response (16; 21). RecA is highly conserved and found in virtually all organisms, with the notable exception of certain endosymbiotic species that have lost much of their genomic content (22). Nevertheless, RecA is a vital protein involved in genome repair and rearrangement. Mutations in RecA are pleiotropic and exhibit various adverse phenotypes including effects on cell division, chromosome segregation, SOS induction, mutagenesis and repair, as well as drastically affecting homologous recombination (23; 24; 25). RecA is essential for almost all pathways of genetic recombination. A single RecA subunit consists of a large central domain, with two smaller domains at the N and C termini (see Fig 1.2).

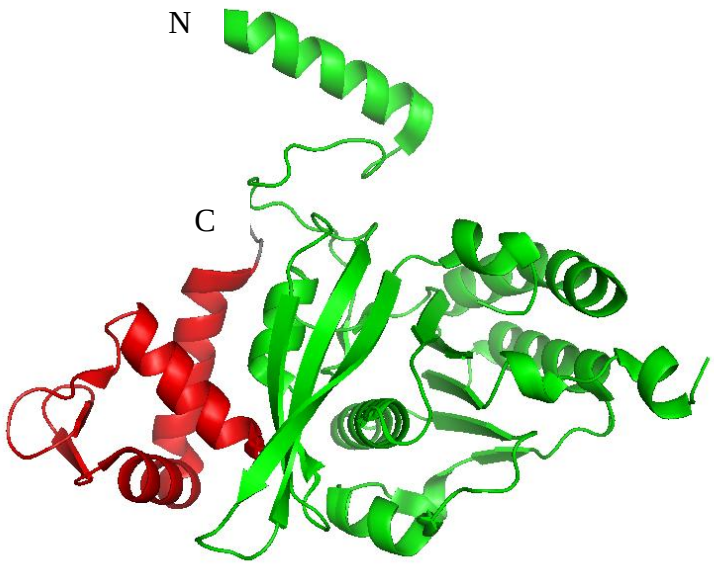


Figure 1.2: RecA monomer showing central core domain (green) and the ATPase domain (red).

The central domain contains a conserved motif known as the RecA fold; a configuration also found amongst DNA helicases, DNA-transport proteins and notably the motor protein F1 ATPase, a subunit of ATP synthase (21). RecA functions in genetic recombination by forming a right-handed helical filament on ssDNA, consisting of 6 subunits per turn, 3 nucleotides per monomer. This binding

mode displays a preference for certain sequences, particularly G-T rich sequences and the χ-recombination hotspot (26). Within this nucleoprotein filament, the ssDNA is stretched and uncoiled in preparation for pairing with a homologous partner strand (21). Filamentous RecA can bind a total of 3 DNA strands, including the strand on which it forms a filament; this binding of DNA acts as a co-factor for ATPase activity that drives the exchange reaction. ATP hydrolysis leads to conformational changes in the filament that optimise the DNA-protein complex for exchange. When a region of dsDNA homologous to the bound ssDNA is located, the RecA-nucleoprotein filament drives exchange and inserts the ssDNA into the dsDNA, creating a D-loop structure. RecA will form DNA joints in this manner provided that a free end exists somewhere in the two DNA molecules (27). The numerous functions of RecA are summarized in figure 1.3

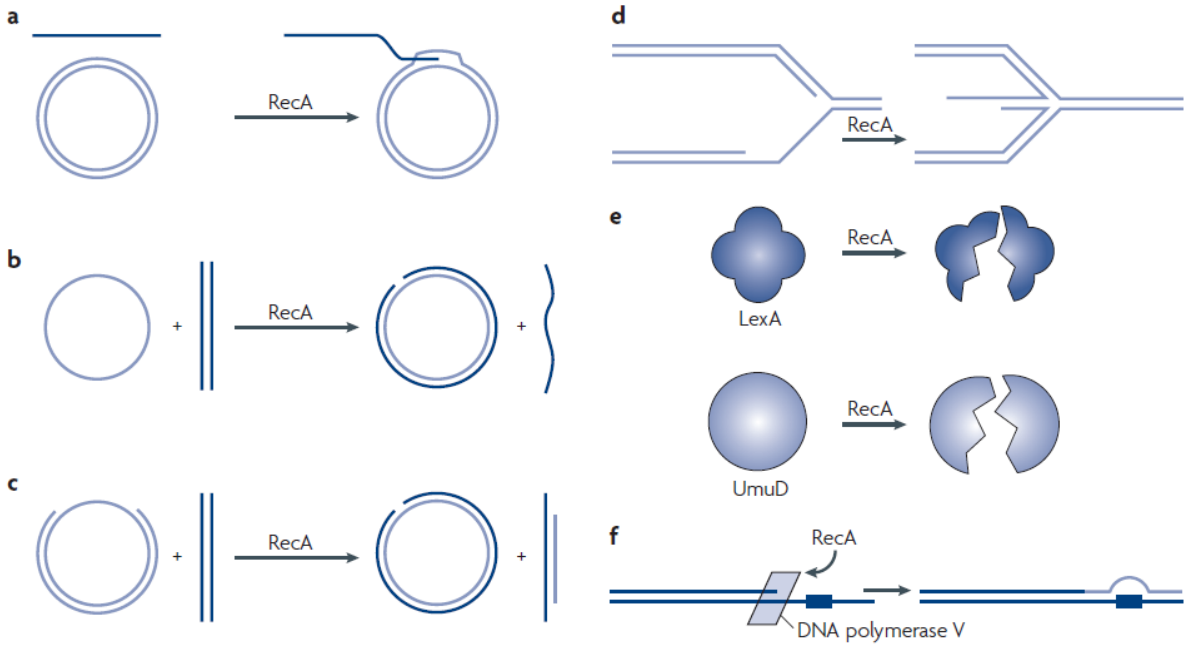


Figure 1.3 : Activities of *E. coli* RecA: a. strand invasion producing a D-loop structure, b. a three strand exchange reaction, c. a four strand exchange reaction, d. replication-fork-regression (28), e. RecA co-protease function f. RecA stimulation of DNA polymerase V translesion-replication activity (21).

Given RecA's central role in recombination, it is unsurprising that its function is tightly regulated (29). Three levels of regulation have been proposed: control at the

level of the SOS regulon, autoregulation by RecA itself and finally regulation by other proteins.

Crystal structures of RecA show significant disorder in the 25 C-terminal amino acids and further investigation has revealed that the terminal 17 residues constitute a regulatory motif (30). Various C-terminal deletions (ΔC6, ΔC13, ΔC17 and ΔC25) cause dramatic alterations in the pH-reaction curves for DNA strand exchange (30). *In vivo* these deletion mutants exhibit wild type behaviour when exposed to UV or ionizing radiation. However, treatment with the DNA cross-linking agent mitomycin confers markedly reduced survival with deletions of 13 or more C-terminal residues. These results indicate a crucial role for the C terminus in the repair of DNA interstrand cross links (30). Furthermore, RecA activity is seemingly enhanced substantially by removal of the final 17 residues (29), with the mutant no longer requiring divalent magnesium ions for DNA strand exchange (normally essential for wild type RecA); it shows improved binding to dsDNA relative to the wild type protein and displaces SSB more effectively (31; 30). The location of the C-terminal regulatory region within the folded protein may have implications on its interactions with other protein regulators (29), discussed below (figure 1.4).

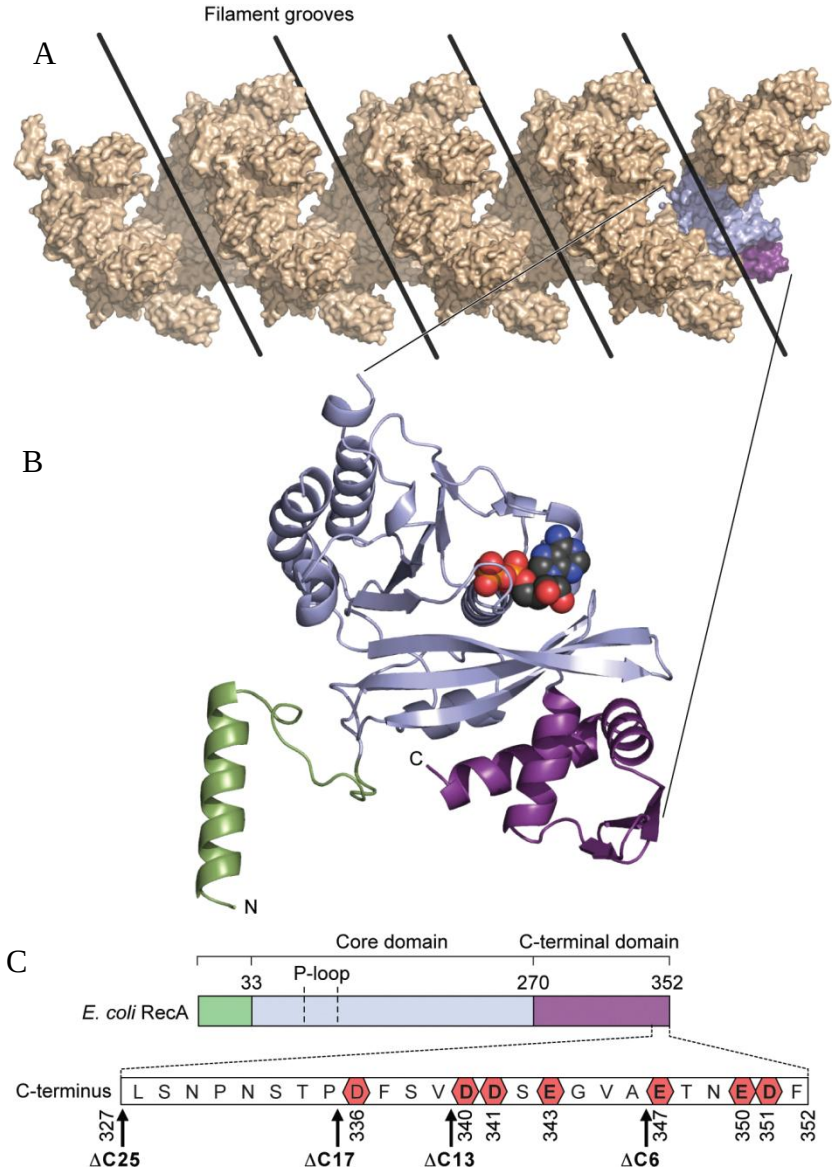


Figure 1.4 : A. RecA structure, showing filament (orange) with one monomer highlighted (light blue, purple) B. detail of one monomer showing C-terminus (purple) C. primary sequence map showing mutants studied. Figure taken from (Cox 2007)

RecA is subject to regulatory control by a number of cytosolic proteins. Those currently known include RecF, RecO, RecR, DinI, RecX, RdgC, PsiB and UvrD (29). UvrD is a DNA helicase (32) and as such is able to disrupt the formation of RecA filaments; the RuvAB branch migration helicase can also displace RecA filaments *in vitro* (33). The RecFOR proteins, discussed in more detail later,

modulate both the assembly and breakdown of RecA filaments. RecF in particular may have a complex impact on RecA regulation (34).

DinI protein, when present at the same, or slightly higher, stoichiometric levels as RecA, has a stabilizing effect on RecA DNA binding, almost entirely preventing its tendency to dissociate while leaving its LexA coprotease and ATPase activities unaffected (DinI also seems to have a complex role in the regulation of the SOS response) (35).

RecX acts as an effective inhibitor of RecA functions; *in vitro* relatively low levels of RecX are able to inhibit RecA-dependent strand exchange, ATPase and co-protease activities (36). *In vivo* RecX, as part of the SOS regulon through the LexA repressor, is up-regulated in response to DNA damage by RecA itself. Loss of function *recX* mutants display decreased UV radiation resistance, while overexpression leads to an even greater increase in UV sensitivity (36). RecX functions through a proposed filament capping mechanism, possibly acting to limit the extent of RecA filament formation, and therefore the extent of invasion, at sites of homologous recombination (37).

PsiB, unlike other RecA inhibitors, does not act to dismantle RecA filaments once they have formed, but instead binds RecA in free solution (38). Its principal function seems to be inhibition of the SOS response, as it shows less effect on RecA DNA binding or ATP hydrolysis. This might point to PsiB levels establishing a threshold that needs to be overcome by DNA-damage induced LexA-autocatalysis before full expression of the SOS response is achieved.

SSB

Single-stranded DNA binding (SSB) protein was first isolated and investigated by Sigal at al. in 1972 (39). The 19kDa subunits of SSB form complexes on ssDNA that fulfill four important functions. Firstly, it protects ssDNA in a transient manner, shielding it from nuclease attack. Secondly it helps eliminate intra-strand secondary structure to prevent replication skipping of repeat sequences. Thirdly, it prevents

reannealing of strands separated by the replicative machinery. Finally, it targets DNA replication, repair and recombination proteins to exposed ssDNA via specific protein:protein interactions (40). Under electron microscopy, SSB-ssDNA complexes appear as nucleosome-like beaded loops (41). Such ssDNA-SSB complexes show resistance to digestion by non-specific nucleases (41). Each individual SSB-ssDNA complex consists of a homotetramer of SSB that bind in different modes, differing in the number of bases occluded by each tetramer. Two major binding conformations protect 35 ±2 (known as $(SSB)_{35}$) and 65 ±3 (known as $(SSB)_{65}$) nucleotides per tetramer (42). The former binding mode predominates at low salt concentrations (<10 mM NaCl) and high protein:DNA ratios, and results in the development of long stretches of SSB complexes bound to DNA with only two subunits of SSB bound to DNA. Alternatively, the $(SSB)_{65}$ binding mode leads to a less organised pattern of binding to ssDNA and involves all four subunits contacting the DNA (42).

SSB physically limits the access of many of the enzymes of DNA metabolism, including the recombinase RecA. Hence RecA needs to be loaded onto ssDNA by DNA processing enzymes. In *E. coli,* recombination is executed by RecBCD at severed dsDNA ends and at gaps by RecFOR. These pathways ensure that RecA is loaded onto suitable substrates to promote exchange and ultimately result in a repaired, integrated, DNA duplex.

RecBCD Pathway

RecBCD recombination in *E. coli* can be divided into three steps: pre-synapsis, synapsis and post-synapsis/resolution (fig. 1.5) (43).

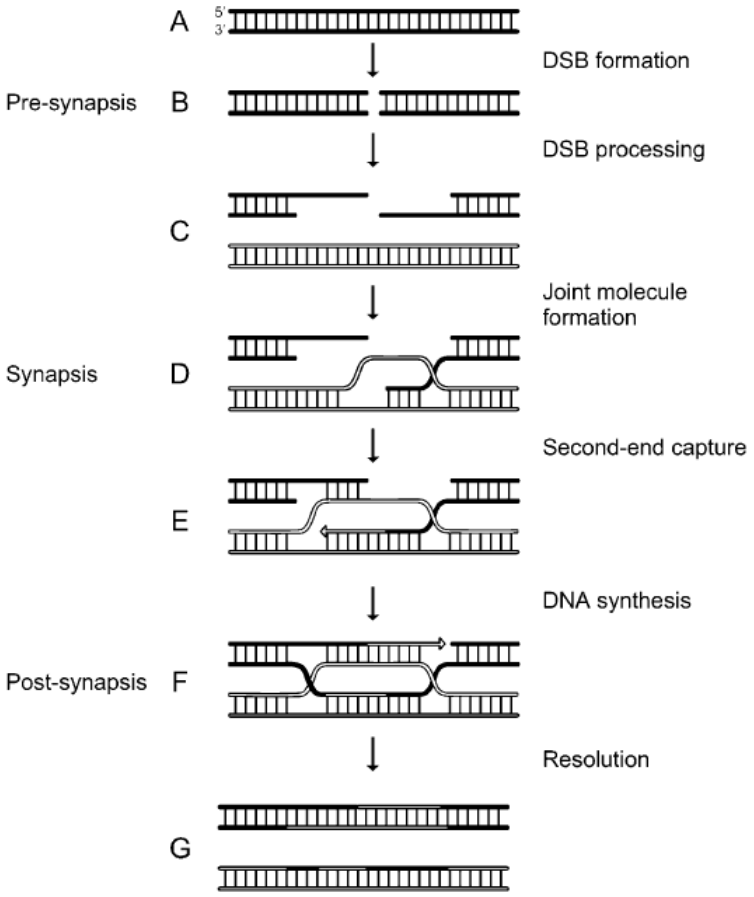


Figure 1.5 : Homologous recombination starting from a double-stranded break utilising the RecBCD pathway in *E. coli*. From Wyman, Ristic, and Kanaaret (2004).

Pre-synapsis involves the processing of broken dsDNA to produce ssDNA upon which the recombination machinery can assemble. In *E. coli*, end processing is primarily achieved through the multiple functions of the RecBCD protein complex (44), which incorporates twin helicases, Chi-sequence recognition and a processive exonuclease (45). Chi sequences consist of an 8 nucleotide consensus sequence (5'-GCTGGTGG-3') present on the *E. coli* genome that act to regulate RecBCD degradative activities. RecBCD loads at a dsDNA end and translocates along the dsDNA, preferentially cleaving the 3' strand. Upon reaching a Chi-site, the cutting of the 3' strand is mitigated, while a weaker nuclease activity on the 5' strand is stimulated and the complex proceeds at roughly half its previous rate (44; 45). RecA is loaded onto the resulting 3' ssDNA overhang at this point through an interaction with RecB (46).

Each of the components within the RecBCD complex contributes differing functions. Within the RecBCD complex, RecB acts as a 'slow' 3'-5' helicase and provides nuclease activity (47). RecD serves as the 'faster' 5'-3' DNA helicase (48; 49). Both helicases complement each other's function, as they act on the anti-parallel strands of DNA, by virtue of their opposite and unidirectional helicase activity, they drive the migration of the RecCD complex in a single direction (48). As with all helicases, RecB and RecD utilize the energy from ATP hydrolysis to drive holoenzyme translocation along duplex DNA (50). The χ-recognition function resides in the RecC subunit (51).

The post-synaptic phase of double-stranded DNA break repair recombination in *E. coli* uses the invaded DNA molecule as a template and the invading strands as a substrate for DNA replication to replace missing nucleotides (43). The RuvABC complex eliminates the Holliday junctions formed to release the restored heteroduplexes (43; 44).

RecFOR Pathway

RecFOR provide an alternative route for recombination and repair (52). While RecBCD functions to initiate recombinational repair at dsDNA breaks, the RecFOR pathway faciliates restoration at ssDNA gaps resulting from problems encountered during DNA replication (53). Like RecBCD, RecFOR, in combination with nucleases and helicases such as RecQ and RecJ, functions to process damaged DNA and load RecA to allow exchanges that provide a template for repair or to restore a replication fork. *E. coli recF, recO* and *recR* mutants are sensitive to DNA damage and show moderate defects in recombination (54). The structures of the RecFOR proteins have been determined from proteins obtained from *Thermus thermophilus* (55) and *Deinococcus radiodurans* (56). Although the precise roles of the three components have not been fully elucidated, RecFOR appear to primarily function to help load RecA onto ssDNA covered with SSB protein, thus stimulating synapsis and repair. Various combinations of the protein may also help in homologous pairing and in stabilisation or disassembly of the RecA filament.

The RecFOR pathway (fig. 1.6) is stimulated by the presence of a duplex region bordering a ssDNA region within the same molecule. If RecFR is exposed to DNA before the addition of RecO, recombination is stimulated, while incubation of RecF alongside RecO leads to substantial inhibition of RecFOR recombination (57).

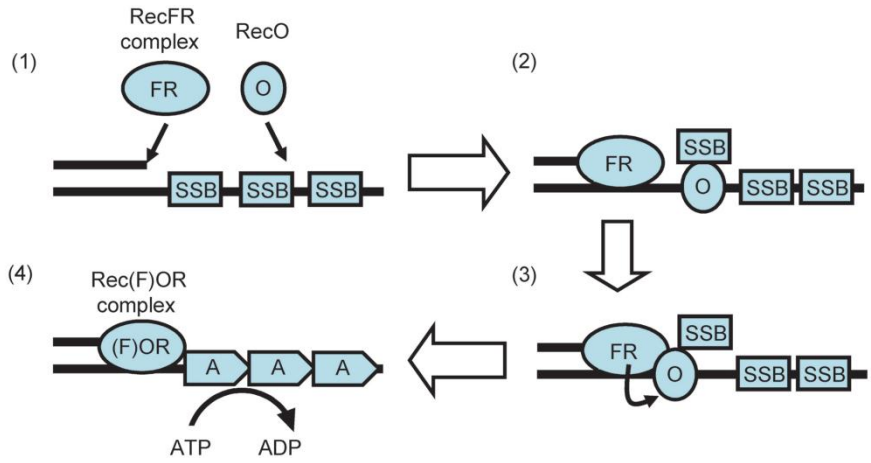


Figure 1.6 : Model of the RecFOR pathway: 1. RecFR detects and binds at the dsDNA/ssDNA interface, while RecO recognises SSB bound to ssDNA. 2. RecO displaces SSB and binds ssDNA. SSB remains associated with ssDNA indirectly via the interaction with RecO. 3. A portion of RecR within the RecFR complex is exposed and interacts with RecO. 4. A RecFOR complex is formed that can load RecA onto the ssDNA to mediate homologous recombination (58)

The 40 kDa RecF protein binds ssDNA, dsDNA and displays dsDNA-dependent ATPase activity (59; 60; 61). It also associates with with RecR in an ATP and DNA dependent manner (59). RecF is an ABC-type ATPase, undergoing ATP-dependent homodimerisation; it is likely that the ATPase activity plays a crucial regulatory role (62). Its tendency to localize to the ssDNA/dsDNA interface is not an inherent property of the protein, but likely to be a result of a matrix of DNA-protein and protein-protein interactions (62). RecF binds DNA in a crab-claw manner; encircling the dsDNA-ssDNA interface, although not completely (61) (fig. 1.7). It has been proposed that *E .coli* RecF bound to DNA specifically interacts with dimeric RecR and causing it to form a tetrameric RecR clamp on ssDNA which attenuates RecF ATPase activity as RecR is loaded (61).

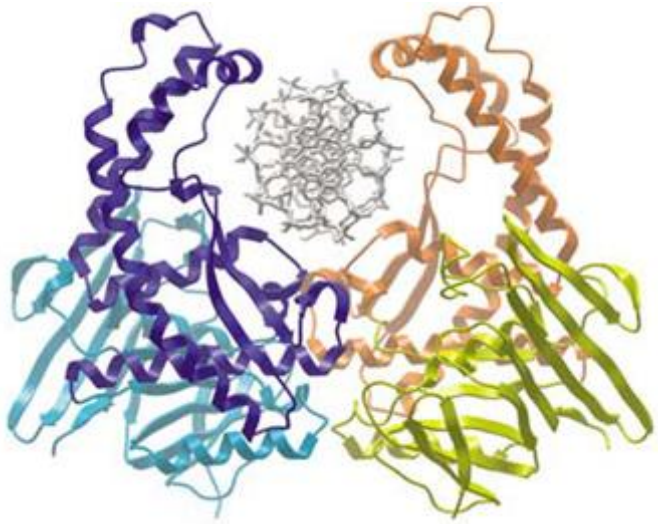


Figure 1.7: RecF (Monomer 1 lobes I and II cyan and dark blue, monomer 2 lobes I and II orange and yellow) bound to DNA (grey) (61)

The 22kDa RecR protein interacts with both RecF and RecO (63). It most likely fulfils a regulatory role in distinct RecF and RecO-mediated recombination, likely interacting with RecA as well (62). In solution, RecR forms a dimeric complex on its own, a hexoheteramer with RecF (2 RecF: 4 RecR) and a heterotetramer with RecO (2 RecO: 2 RecR). Studies have shown that a modified domain with notable similarity to topoisomerase and primase-type proteins is actually responsible for protein-protein interactions with both RecF and RecO, and that RecO and RecF compete for this binding site (64).

There is evidence that RecR may act as a regulator of RecA. In the presence of both RecO and SSB, indirectly associated with RecO, the ssDNA-dependent ATPase function of RecA is usually inhibited, but can be restored by the inclusion of RecR (55). It is possible that RecR helps target RecA function specifically to the ds/ssDNA junctions found in gapped-duplexes.

RecO, a 27kDa polypeptide, displays ssDNA and dsDNA binding properties (co-ordinated through separate binding sites), along with ssDNA annealing activity that is stimulated in the presence of SSB (56; 65). RecO is a dimer in solution and forms protein-protein interactions with RecR (66). In *E. coli* these interactions are believed

to result in a RecOR heterotetramer, formed from dimers of RecO and RecR (56). In *D. radiodurans* a heterohexamer is assembled from four RecR and two RecO monomers. The combination of RecO and RecR may constitute a recombination pathway in the absence of RecF that requires distinct protein-protein interactions, the presence of different DNA substrates and unique properties (fig 1.8) (57). For example, RecOR is more efficient than RecFOR at loading RecA onto ssDNA substrates. RecOR pathway recombination appears to be dependent upon the interaction between RecO and the C-terminus of SSB (67). This interaction distinguishes the RecOR pathway since no such interaction is required in RecFOR recombination (57).

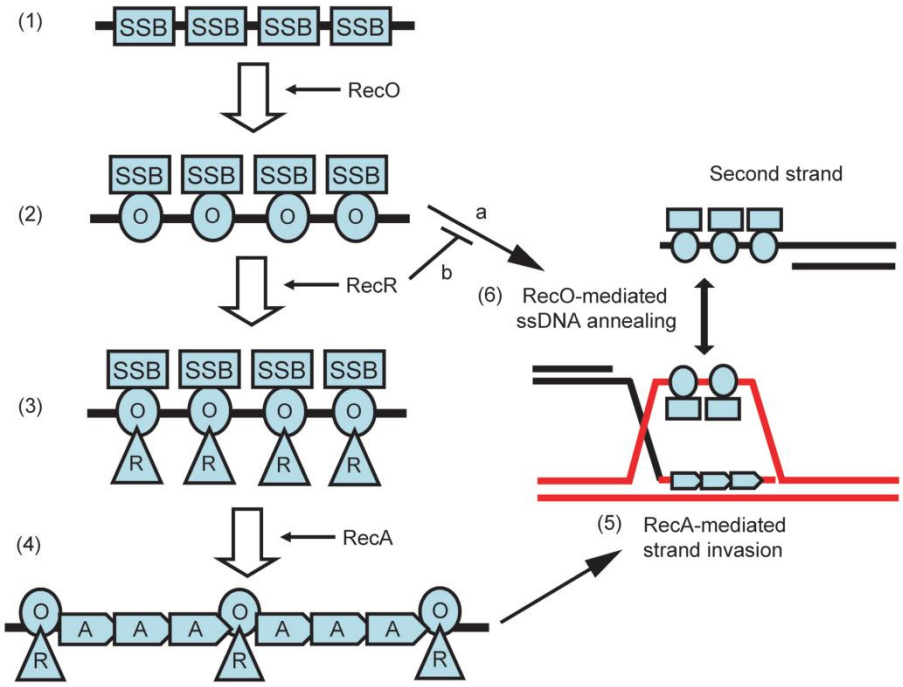


Figure 1.8: Model for RecOR pathway recombination. 1. ssDNA is coated with SSB 2. RecO recognizes SSB bound to ssDNA and displaces SSB, while maintaining the protein-protein interaction. In the absence of RecR, RecO can initiate strand annealing if a complementary single strand is present (a). 3. In the presence of RecR, RecO-mediated single strand annealing is inhibited 4. The RecOR complex can load RecA onto ssDNA with RecA-nucleoprotein filament formation causing displacement of SSB. 5. RecA promotes strand invasion. The displaced strand can be captured and processed as per the preceding steps (58).

RuvABC

The RuvABC complex is responsible for the ATP-dependent migration and resolution of Holliday junctions. Connected duplexes, generated by RecA strand exchange, are pumped through the RuvAB machinery until the junction is ultimately resolved by RuvC endonuclease cleavage.

The *ruv* locus consists of three genes: *ruvA, ruvB* and *ruvC*. Mutations in any of the *ruv* genes confer sensitivity to UV radiation, ionizing radiation and certain antibiotics (68). Although these mutant strains are only moderately defective in homologous recombination, combinations of *ruv* with *recBC sbcBC*, *recBC sbcA* and *recG* show 30- to 500-fold reductions in recombination (69).

Within the RuvABC complex, RuvA is responsible for targeting RuvB to the Holliday junction (70). In *E. coli*, RuvA is a 22 kDa protein that forms stable tetramers in solution (69; 71). RuvA is able to bind both supercoiled and relaxed DNA, ssDNA and dsDNA but binds preferentially to Holliday junction DNA; an affinity that is increased four-fold by the presence of RuvB (72). Holliday junction binding is structure specific and independent of DNA sequence (69). Upon Holliday junctions binding, RuvA imposes an open, unfolded square-planar conformation rather the stacked-X conformation favoured in the presence of metal ions (figure 1.9) (73). This is important because the spontaneous migration of Holliday junctions within the stacked-X conformation is radically (approximately 1000-fold) slower than that observed in the open, square-planar configuration (74).

As in solution, RuvA assembles as tetramers on the Holliday structure. One or two tetramers can form on each Holliday junction – the presence of two tetramers likely inhibits junction cleavage by RuvC by sandwiching the junction and physically blocking RuvC access (75). A RuvA octamer is required for effective migration of branched DNA molecules *in vitro* and for replication fork reversal *in vivo* (76).

The flower-like RuvA tetramer has concave and convex sides – the former being positively charged and acting as a platform for DNA interaction (69). Conserved

residues on the concave surface include four negatively-charged pin structures (two on each monomer), which ensure Holliday junction recognition over duplex DNA binding and may assist branch migration (77).

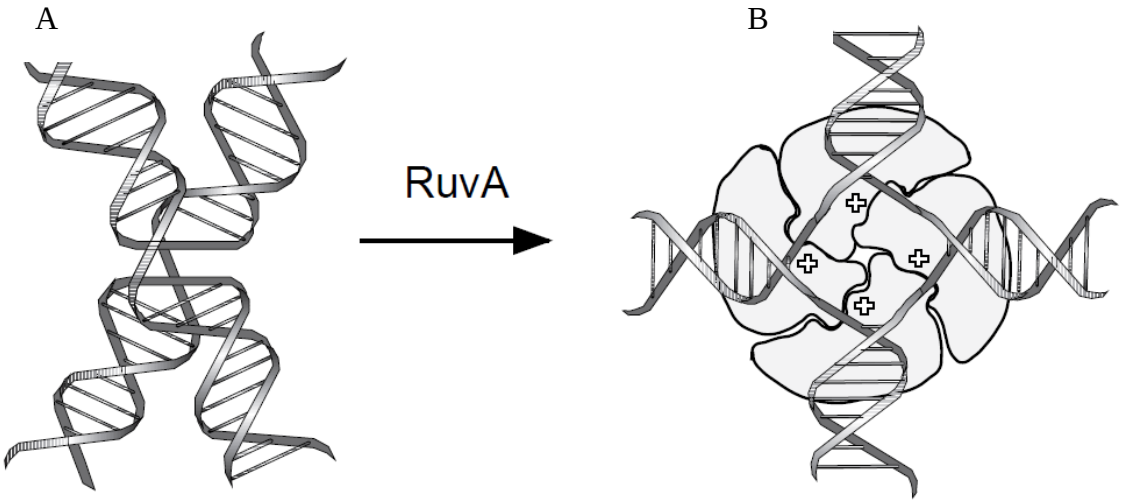


Figure 1.9: Holliday junction conformation is influenced by RuvA binding. The junction changes from a co-axial stacked X structure (A) to a square planar conformation (B). Negatively charged 'pin' structures on RuvA are highlighted at the junction centre (69)

RuvB is a 37 kDa protein containing an ATP binding domain and degenerate DNA helicase motifs (78; 69). In the presence of Mg^{2+} and adenosine 5'-γ-thiotriphosphate, RuvB can bind DNA in two forms; a single homohexameric ring or dual two-fold symmetric rings arranged as a dodecamer (79). RuvB functions as an ATP-dependent molecular motor, specifically targeted to Holliday junctions by RuvA with hexameric rings assembled on opposing arms acting to drive branch migration (80).

The helicase function of RuvB is Mg^{2+} and ATP dependent and is stimulated by the presence of RuvA (81). Increasing the concentration of Mg^{2+} correlates with increased RuvB ATPase activity in line with increased DNA affinity in the absence of RuvA (69). Circular duplex DNA stimulates the ATPase activity more than that induced by ssDNA, despite linear dsDNA and circular dsDNA affinities being similar (81). Despite the hexameric nature of RuvB, the rate of ATP hydrolysis within this annular structure is lower than expected compared to the rate observed with RuvB monomers (82). This seems to indicate a functional asymmetry within the hexamer (69). Crystals of *Thermus thermophilus* RuvB contained two RuvB molecules arranged in an asymmetric unit (83) (see figure 1.10). There are known

examples of other annular, hexameric helicases that interact via three pairs of asymmetric dimers (69) including phage T7 gp4 and Rho and DnaB from *E. coli*.

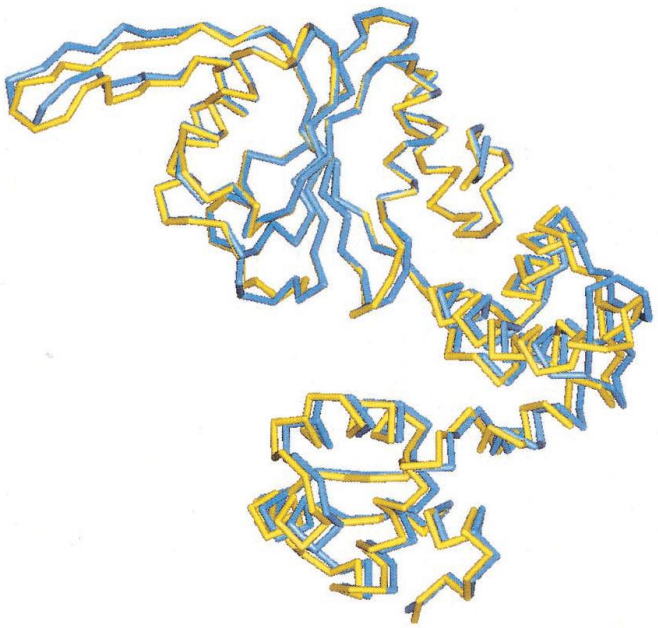


Figure 1.10: Structural differences between the two asymmetric RuvB molecules from RuvB crystals. Only the N-domain (nucleotide binding pocket) C_α backbone is shown taken from Yamada et al. (2001).

RuvAB interactions create a heteromeric helicase of two RuvA tetramers forming an octomeric core around the Holliday junction, with two hexameric RuvB rings enveloping opposing DNA duplexes. The action of RuvB rotating the bound DNA is thought to drive migration of the Holliday junction (figure 1.11). Interestingly, RuvAB can act to drive recombination through regions of considerable heterology or act as an 'anti-recombinase' through reversal of stalled recombination attempts (69). There is good evidence that RuvAB interacts directly with RecA, inducing its dissociation from DNA (84; 85); in this scenario RuvAB acts to promote recombination by increasing the bioavailability of RecA within the cell (69).

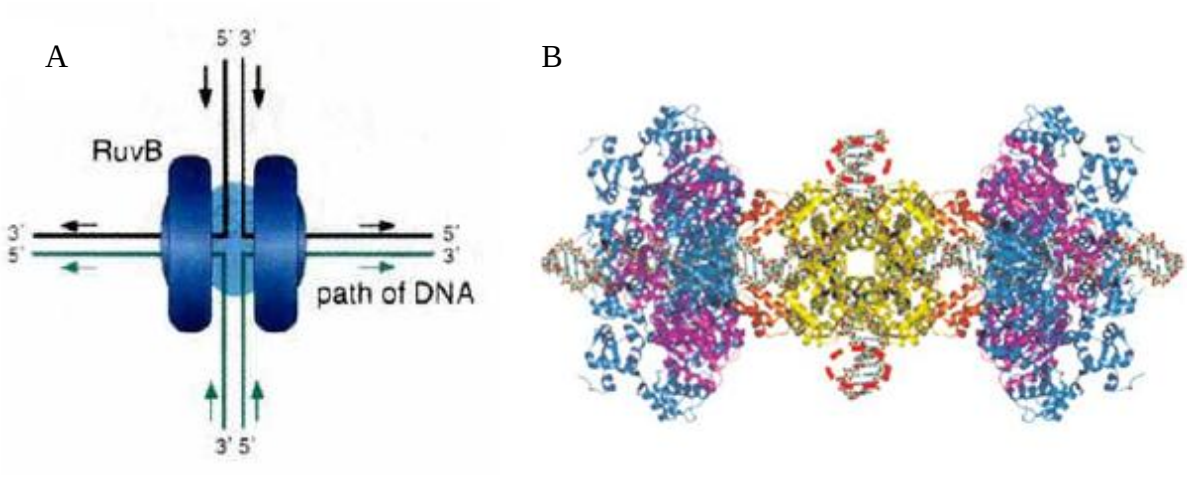


Figure 1.11 : A. Cartoon representation of the active RuvAB complex (69) B. The RuvAB complex as modelled by X-ray crystallography (86)

The 19 kDa RuvC endonuclease is the major junction resolution enzyme of *E. coli* (87; 88; 15; 89; 90; 16). It forms a homodimer in solution and displays a clear affinity for Holliday junction structures over duplex or single stranded DNA (91) (figure). RuvC functions by making paired incisions at the junction crossover (18; 69; 92). This reaction yields nicked duplexes that can be repaired by DNA ligase (92). RuvC belongs to a class of branch-specific endonucleases known as the Holliday junction resolvases; a group characterized by homodimerism, branched DNA recognition and cleavage, high structure selectivity, generally containing <200 residues, with a high level of positively charged residues and often displaying sequence-specific cleavage, while inducing specific junction folding on the bound DNA (92). The active sites of these enzymes typically contain 3-4 acidic residues for divalent cation sequestration that are essential for cleavage, but not for binding (93). In *E. coli* RuvC, the catalytic domain is composed of Asp-7, Glu-66, Asp-138, and Asp-141 (94) (figure 1.12). In terms of cleavage specificity, RuvC shows a preference for cleaving on the 3' side thymine dinucleotides with the following consensus, 5'-$^{A}/_{T}$TT↓$^{G}/_{C}$- 3' (90).

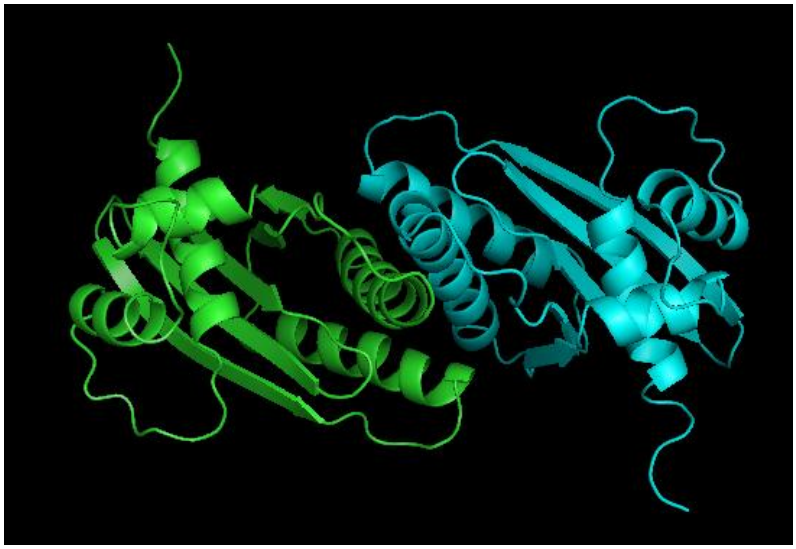


Figure 1.12: Ribbon structure of a RuvC endonuclease dimer. Monomers are coloured in green and cyan.

Assembly of RuvA, RuvB and RuvC on the X structure (95) results in the formation of a complex known as the 'resolvasome' (96) (fig. 1.13). RuvC assembly on the junction is stabilised by interactions with RuvB with the final assembly consisting of a dimer of RuvC, a tetramer of RuvA and two hexamers of RuvB (figure). The resolvasome forms sequentially (RuvA → RuvB → RuvC) upon joint DNA molecules generated by RecA and takes over extension of the joint to form heterologous DNA duplexes. Alternatively, it can function to reverse the joint and 'undo' stalled recombination intermediates depending on which duplexes RuvB can access for assembly. As the branch migration proceeds the RuvC dimer 'scans' the crossover for its preferred cleavage recognition sequences, whereupon the junction is resolved, the complex dissociates and the resulting nicked duplexes can to be repaired by DNA ligase.

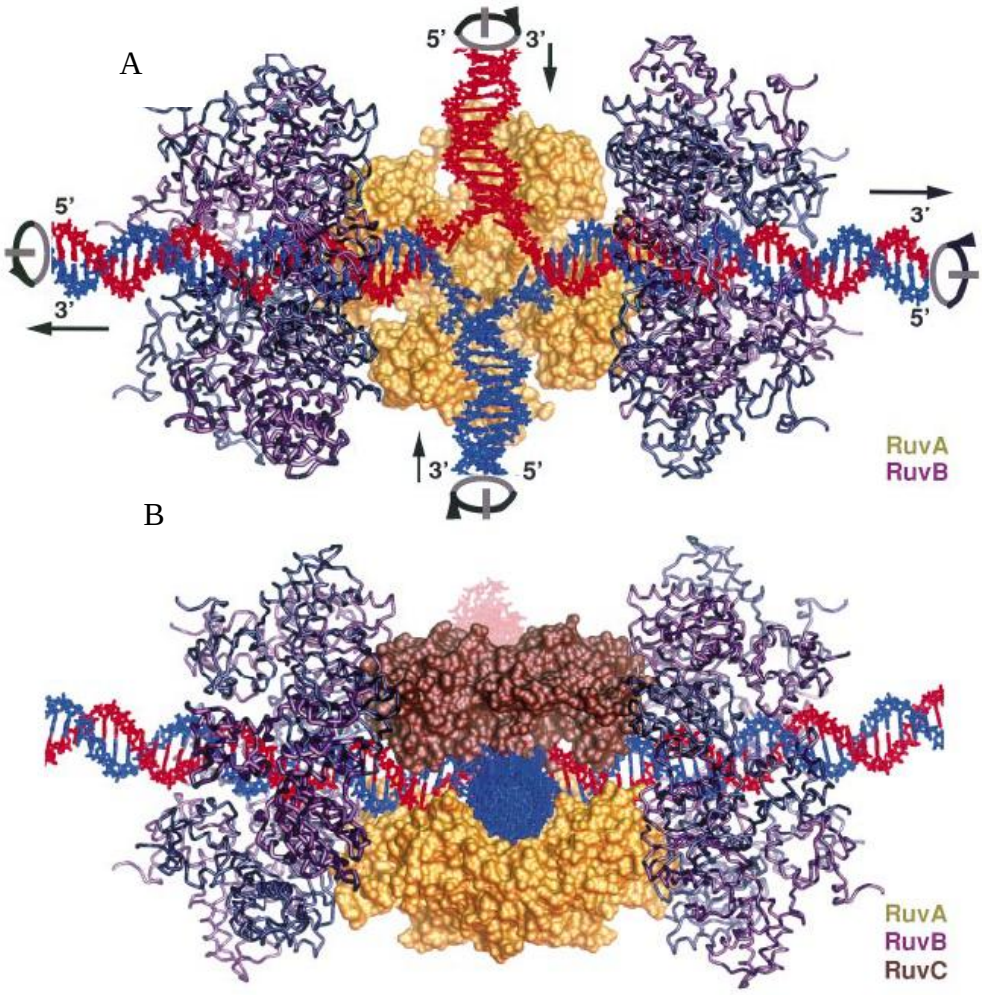


Figure 1.13 : (A) Molecular model of the RuvAB helicase complex incorporating two hexamers of RuvB (dark blue, purple) sandwiching a single RuvA tetramer (yellow). The direction of DNA translocation is indicated by arrows in which strands from two duplexes (red and blue) are exchanged during branch migration. (B) Representation of the RuvABC resolvasome complex based on molecular models of RuvC (red) and biochemical data indicating a RuvB-RuvC protein-protein interaction necessary for RuvABC to function in vivo. The Holliday junction has been rotated 90° into the plane of the page relative to that depicted in A.

Depending on which pair of strands are targeted, Holliday junction resolvase-induced cleavage can yield either 'patch' or 'splice' DNA duplexes. Due to the conformation imposed on the Holliday junction by RuvA, this no longer affects the products of RuvC cleavage - the orientation of the resolvasome is the only dictator of whether the products are splice or patch recombinants (97) (figure 1.14).

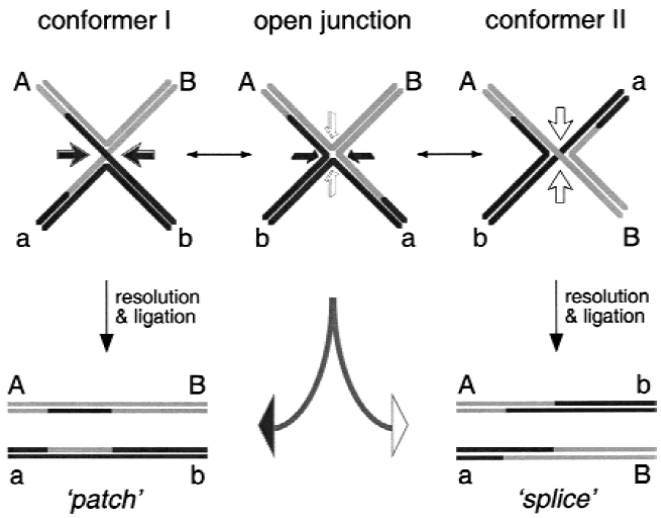


Figure 1.14 : Effect of DNA conformers on recombination products (Conformer I and Conformer II) and effect of RuvA binding (open junction)

1.3.3 Recombination in bacteriophage λ

During the lytic cycle, λ not only imposes its own recombination machinery on the host but also actively inhibits the host's capacity for autonomous recombination, achieving this with only a few proteins. Recombination in λ is an important field of study for a number of reasons. In *E. coli,* and other bacteria, the Red system induces a hyper-recombinational state where recombinational events involving as little as 40 base-pairs of homology occur with high incidence (98). The recombination efficiency of Red systems components makes it a potent biomolecular tool (99). Furthermore, the activities and functions of the Red pathway may have implications on how other phages execute homologous recombination

Exo, Bet and Gam

The Gam protein is encoded by the *gam* gene and expresses either a 138 (GamL, large) or 95 (GamS, small) residue product depending on which one of two possible translational start sites are utilised (100; 101). Gam is a homodimer is solution (102) and inhibits host RecBCD exonuclease activity by blocking RecBCD from associating with DNA ends, possibly by mimicking nucleic acid ends (101).

The *exo* gene encodes the 24 kDa λ-exonuclease Exo (98; 103), which functions in an analogous manner to RecBCD by recognising a double strand break and degrading the 5' strand to create a 3' overhang (98; 104; 105). Structurally, Exo forms a trimeric, tapered ring, with the channel averaging 30Å at its widest and 15Å at the narrowest (105). The wider opening in the ring accommodates dsDNA, while only ssDNA can traverse the aperture (105) (see fig 1.15). The nuclease active site of Exo is arranged to ensure that only the 5' end is targeted for destruction, utilising a magnesium ion for phosphodiester bond catalysis (105).

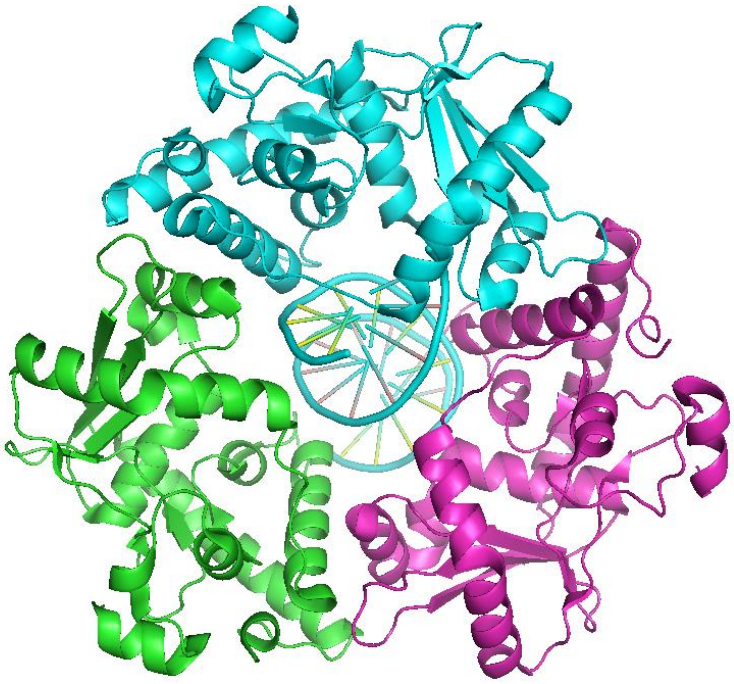


Figure 1.15: Exo trimer (green, cyan and purple) binding 12-bp DNA artifact (cyan and yellow)

β is part of an evolutionarily distinct superfamily of single-strand annealing proteins (SSAPs) that includes the *E. coli* prophage protein RecT, known as the RecT/Redβ superfamily (106). The 28 kDa β is encoded by the *bet* gene, whose role is similar to that of RecA in *E. coli* recombination, although β does not bind or hydrolyse ATP. β binds ssDNA at 3' overhangs and catalyses strand exchange with a homologous single DNA strand (103). Furthermore, β seems capable of driving strand exchange with mismatched base pairs of up to 8-16 bases (103). In the absence of any DNA substrates, β tends to form annular structures of approximately 12 subunits (107). On dsDNA products of annealing β forms a left handed helical filament, with a turn once

every 100 base pairs, and protects the duplex from nucleolytic attack (107). On ssDNA, β forms a polymer ring of between 15-18 subunits (107).

Exo and Redβ do not act independently, they utilise specific protein-protein interactions to coordinate degradative and annealing activities (108). The functionally equivalent protein pairing RecE/RecT in *E. coli* also work together (106; 108). These partnerships were confirmed by experiments whereby RecE/RecT and Exo/Redβ pairs were functional but the RecE/Redβ and Exo/RecT combinations were not (108), indicating protein-specific interactions between and that coupled reactions are necessary for DSB repair.

λ Rap

λ encodes its own activity for resolving branched DNA intermediates generated by RecA or by the Red system. The product of the *rap* (*ninG*) gene functions to resolve Holliday junctions like the RuvC resolving enzyme. However, it has a relaxed structure-specificity and cleaves, fork, D-loop and other branched structures more typical of phage resolving enzymes such as T4 endonuclease VII and T7 endonuclease I. λ *rap* mutants exhibit a 100-fold decrease in RecBCD-catalyzed λ-plasmid recombination (109).

λ Orf

Another important component of λ recombination is the Orf (NinB) protein (110). Orf appears to provide a functional equivalent of the host *E. coli* RecFOR complexes, catalyzing the loading of RecA onto SSB-coated ssDNA. In 2004 Poteete (111) discovered that bacteria expressing the Red system but lacking *recF*, *recO* or *recR*, have their sensitivity to ultra-violet radiation and recombinational deficiency suppressed if carrying the *orf* gene fused to a promoter (112). Orf could also partially suppress the UV sensitivity of *ruvABC* mutants in this genetic background, perhaps indicating that Orf can divert recombination down a pathway that does not require Holliday junction resolution (111).

Purified Orf protein interacts with both DNA and SSB in keeping with some of the activities found with RecF, RecO and RecR. It shows a distinct preference for ssDNA over dsDNA and binds equally well to gapped and 5' tailed or 3' tailed substrates. The crystal structure of Orf has been determined and shows that it forms an asymmetrical, homodimeric, toroid. A shallow cleft across the top of the central channel is lined with basic residues and DNA could potentially bind along this cleft or through the central hole (113; 114). The central hole is narrow and it is possible that the annular structure opens like a clamp and binds DNA in a manner analogous to RecF manner (115).

If Orf substitutes for RecFOR, it could stimulate the formation of RecA filaments on ssDNA substrates coated with SSB (113). Alternatively, Orf may promote the loading of β onto ssDNA by negating the inhibitory effect of SSB and promoting strand invasion (113) since β has been found to promote D-loop formation in superhelical DNA as well as exchanges between colinear ssDNA and dsDNA *in vitro* (116). The presence of a functional connection between the components of the Red system may explain why recombination supported by Orf works better with the λ rather than bacterial recombinases (113). Orf could encourage λ to effect chromosomal repair through strand invasion of a homologous region of a sister chromosome if λ-exonuclease has created an overhang and Redβ is unable to find a homologous ssDNA filament for synapsis (figure 1.16) (111).

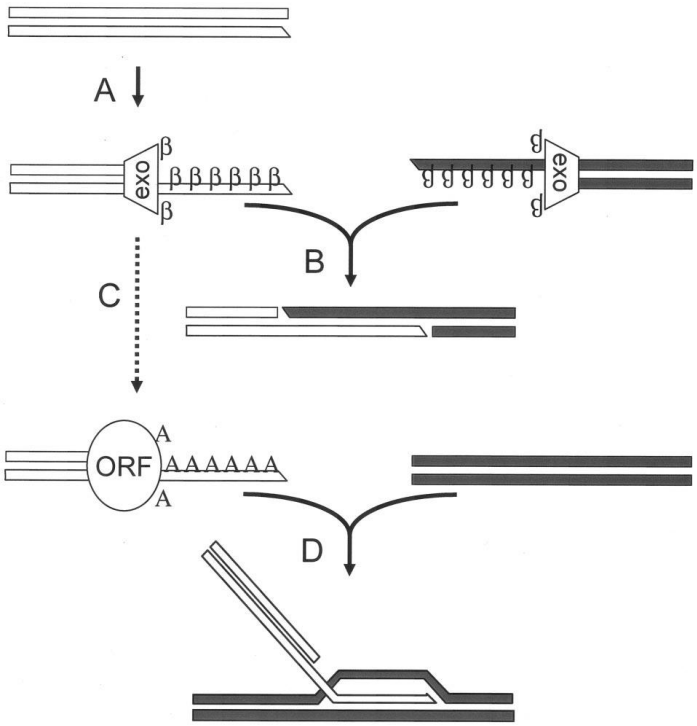


Figure 1.16 : Model for the role of Orf in λ rcombination. A. Free dsDNA is processed by λ-exonuclease. B. Redβ can anneal complementary ssDNA overhangs. C. If an appropriate ssDNA is unavailable, Orf could load RecA onto the ssDNA generated by λ-exonuclease. D. RecA catalyses strand invasion of a homologous chromosome.

Mutagenesis of Orf has been undertaken in an effort to identify which parts of the protein specify the DNA and SSB binding sites. A severely truncated version of Orf (ΔN126), retaining only the C-terminal tail, no longer associates with SSB, while conversely ΔC6 and ΔC19 mutants maintain an SSB interaction, consistent with the C-terminus being dispensible for SSB recognition (115). However, when these mutants were tested for DNA binding activity the ΔC6 mutant showed reduced ssDNA binding, while the ΔC19 mutant was unable to bind DNA at all (115; 117). The C-terminus does, therefore, appear to be important for DNA recognition. The C-terminus in wild-type Orf forms an alpha helix (αE) in one of the crystllographic subunits (figure 1.17) that constitutes 19 residues absent in the CΔ19 mutant. This region is most likely flexible, able to adopt differing conformations, as it is largely disordered in the electron density map of the other subunit (113).

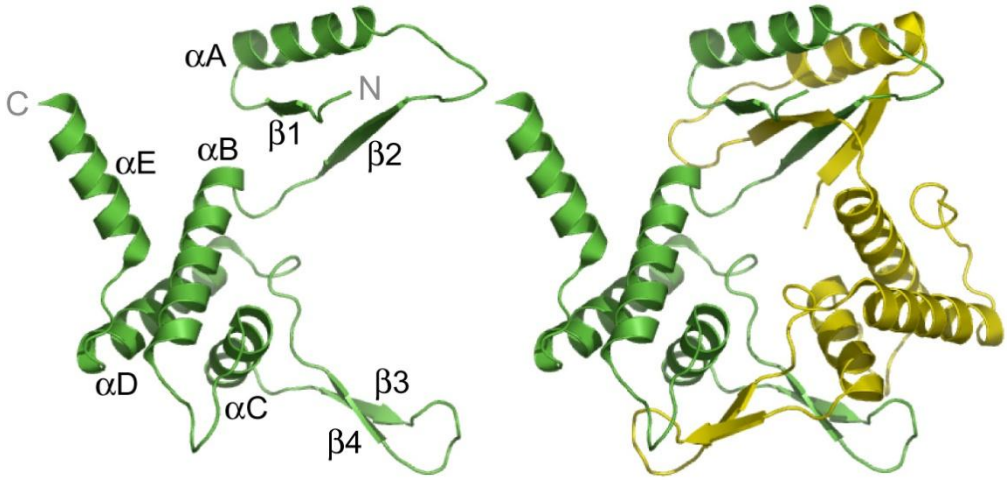


Figure 1.17:Quaternary structure of the asymmetric Orf homodimer (113). The C-terminus of Orf is a helix in one subunit (green) but disordered

1.4 Conclusion

Recombination is a cornerstone of genetic maintenance and evolution. The pathways responsible for homologous recombination in *E. coli* and the lamboid phages act to process damaged DNA or stalled DNA artifacts through the removal of SSB and the loading of RecA to promote homologous strand invasion. Numerous proteins act together to facilitate homologous recombination, many forming annular or helical structures, with several protein-protein interactions underlying important processes; such as RecBCD or RecFOR. λ Orf provides similar function to the RecFOR complex, but its protein-protein interactions have yet to be clearly defined. Rajagopala1 et al. (2011) (118) utilized yeast two-hybrid screens across approximately 93% of the lambda genome to look for prtein protein interactions. It was found that Orf may produce protein-protein interactions with 5 other proteins; however, none of these were other recombinationases. Orf may undergo conformational changes upon ssDAN binding that allow these interactions.

Chapter 2

Materials and Methods

2.1 Software

This thesis was written in Microsoft Word 2007. Figures were edited and amended in Microsoft Paint Version 6 and Adobe Photoshop Elements 9. Gel shift assays were analysed with Image J (W. Rasband, National Institutes of Health) and the numerical data processed by Microsoft Excel 2007. Protein and DNA sequences were obtained from the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov).

2.2 Chemicals and Reagents

Research-grade reagents and chemicals were provided by Bio-Rad, New England Biolabs and Sigma-Aldrich.

2.3 Proteins

The phage Orf, Orf151, ETA20 and NinH proteins used in this study were purified to homogeneity by Patricia Reed (2006) (119), Laura Bowers (2008) (117), Fiona Curtis (unpublished results) and Gary Sharples (unpublished results). Concentrations of Orf protein are expressed as moles of dimer. T4 polynucleotide kinase was obtained from Invitrogen.

2.4 Electrophoresis

2.4.1 Agarose Gel Electrophoresis

DNA samples were separated on 1% agarose-TBE gels. Gels were prepared by melting analytical grade agarose in Tris-Borate EDTA buffer (90mM Tris-borate pH 7.5, 2 mM EDTA) in a microwave. Molten agar was then poured into the gel casting apparatus and allowed to cool. DNA samples were mixed with loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficoll type 400) to a final concentration of 20% prior to electrophoresis in Tris-Borate EDTA buffer. A 1 kb DNA ladder (Invitrogen?) was used as a reference marker.

2.4.2 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS polyacrylamide gels (10-15%) made with 40% 29:1 acrylamide-bisacrylamide (Sigma) were run on BioRad (Mini Protean III) electrophoresis tanks. Protein samples were mixed with SDS-loading dye (100 mM Tris-HCl pH 6.8, 200 mM dithiothreitol, 4% SDS, 0.2% bromophenol blue, 20% glycerol) at 20-25% final concentration and heated at 95°C for 3-4 minutes. Gels were subjected to electrophoresis at 140V for 40-50 minutes in Tris-Glycine running buffer (25 mM Tris-HCl, 250 mM glycine, 0.01% SDS) Proteins were visualized by staining with Coomassie blue (200 ml methanol, 200 ml dH_2O, 80 ml acetic acid, 0.48 g Coomassie Blue R-250) for 15 minutes and destained in 20% methanol, 10% acetic acid for 1-2 hrs as required.

2.4.3 Low Ionic Strength Gel Elecrophoresis

Low Ionic Strength (LIS) polyacrylamide gels (6-8%) made with 40% 29:1 acrylamide-bisacrylamide (BioRad) were run on BioRad (Protean II xi) electrophoresis tanks. Gels were subjected to electrophoresis in LIS running buffer (6.7 mM Tris.HCl pH8, 3.3 mM sodium acetate, 2 mM EDTA pH8) at 160V for 1.5-2 hours.

2.5 Protein Concentration Estimations

Protein concentrations were initially estimated with a modified Bradford Assay kit from BioRad using Bovine Serum Albumin (BSA) as a standard and monitoring absorbance at 595nm in a spectrophotometer (name of machine?). More accurate protein concentrations were determined latterly using a NanoDrop 2000 (Thermo Scientific) micro-volume spectrophotometer; distilled water was used as a blank and absorbance measured at 280nm.

2.6 DNA Substrates

The 50-nt ssDNA substrate consisted of strand 1 (5'-
GGCGACGTGATCACCAGATG
ATTGCTAGGCATGCTTTCCGCAAGAGAAGC-3') which was annealed to its

complement to produce a 50 bp dsDNA. Strand 1 was labelled with [γ^{32}P] ATP (Perkin-Elmer) at the 5' end using T4 polynucleotide kinase (10 units); samples were incubated at 37°C for 60 min and enzyme inactivated by incubation at 65°C for 10 minutes. Labelled DNA was separated from unincorporated nucleotide using MicroBioSpin columns (BioRad). Annealed DNA substrates were further purified by separation on 10% polyacrylamide gels in 90 mM Tris-borate, 2 mM EDTA. Gels were subjected to electrophoresis at 190V for 1.5 hours and labelled products identified by autoradiography. Bands were excised from the gel and allowed to diffuse into distilled water overnight at 4°C. ϕX174 Virion DNA was obtained from New England Biolabs.

2.7 Biochemical Assays

2.7.1 DNA binding assays

The DNA-binding capability of Orf, Orf151, ETA20 and NinH proteins was assessed using ^{32}P-labelled DNA incubated with serial half dilutions of protein. Protein and 0.3nM ^{32}P-labelled DNA (20 µl mixtures) were incubated in gel binding buffer (50 mM Tris.HCl pH8, 5 mM EDTA, 1 mM DTT, 100 µg/ml BSA, 5% glycerol) for 15 minutes on ice. 12µl of each DNA binding assay was then subjected to electrophoresis on 4-8% low ionic strength (LIS) polyacrylamide gels. Gels were then transferred to filter paper and dried for 20-30 minutes on a BioRad gel drier with a dry ice condenser. Dried gels were exposed to phosphoimager screens and detected with a Fuji Film FLA-3000 phosphoimager. Images in tiff format were subsequently analysed with ImageJ software and the data transferred to Microsoft Excel 2007. Raw data from the density of the bands on the gels was used to calculate meaningful data using the formula:

$$\% \text{ DNA}_{\text{bound}} = [\text{Density}_{\text{shifted band}} / (\text{Density}_{\text{shifted band}} + \text{Density}_{\text{unshifted band}})] \times 100$$

2.7.2 Nuclease Assays

The nuclease activity of ETA20 proteins was investigated using ϕX174 Virion DNA and protein fractions collected from gel filtration [include details of column and buffer used] in nuclease buffer (50 mM Tris.HCl pH8, 1 mM DTT, 100 µg/ml BSA) with 1-10 mM $MgCl_2$. Reactions (20 µl) were incubated at 37°C for 30 minutes, terminated by the addition of 5 µl stop buffer (20 mM Tris-HCl pH8, 0.5% SDS, 20 mM EDTA, 2 mg/ml proteinase K) and incubated for a further 10 minutes at 37°C. DNA loading dye (5 µl; 0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficoll type 400) was added to each sample and then subjected to electrophoresis at 80V for 1-2 hours.

2.8 Gel Filtration

Size exclusion chromatography of His-Orf was performed using a BioLogic DuoFlow system from BioRad. 1 mg/ml of protein was injected into an AKTA FPLC system with a 24 ml Superose 6HR 10/30 column (GE Healthcare). Molecular mass standards (BioRad) contained thyroglobulin (670 kDa), g-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa). Proteins were mixed in a 200 µl volume with 20mM Tris–HCl pH 8.0, 1mM EDTA, 0.5mM DTT, 250mM KCl and incubated on ice for 15 min prior to loading 100 µl onto the column at a flow rate of 0.3-0.5 ml/min.

Chapter 3

Mutations in λ Orf affecting ssDNA binding

3.1 Introduction

Within the Orf family of proteins, the most highly conserved residues lie close to the central channel (GJ Sharples, personal communication). In 39 of 124 Orf family proteins these include arginine-41 (R41), glutamine-45 (Q45), asparagine-46 (N46) and either histidine or tryptophan at position 50 (W50 in λ Orf). Since this cavity could potentially accommodate ssDNA binding site, point mutations in these conserved residues should help to confirm whether this is the case (fig 3.1)

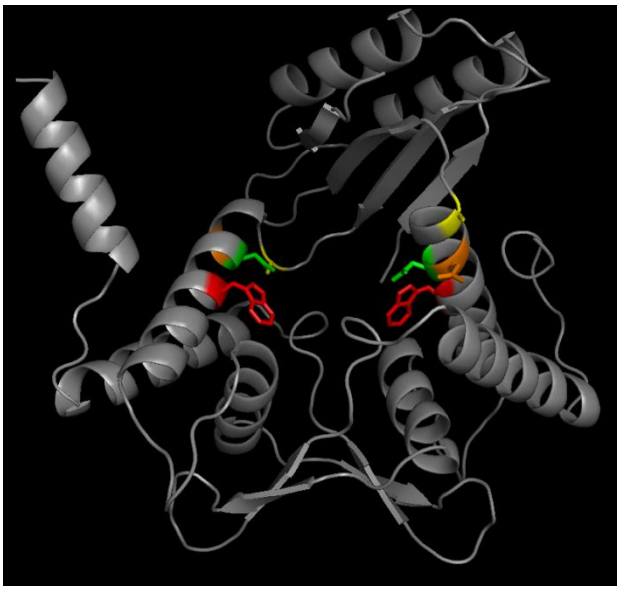


Figure 3.1: Ribbon diagram highlighting conserved residues around the central cavity of in dimeric λ Orf. R41 is yellow, Q45 is orange, N46 is in green and W50 is in red.

Furthermore, a short motif in Orf (residues 102-108) resembles a portion of RecA protein involved in intersubunit contacts that ultimately results in nucleoprotein filament formation on ssDNA (G.J. Sharples, unpublished results) (fig. 3.2) This motif also appears to be functionally conserved in the human homolog of RecA, Rad51, and also in the breast cancer susceptibility protein BRCA2. This motif in BRCA2 appears to facilitate loading of the Rad51 filament onto DNA by mimicking a Rad51 subunit (120; 121). Thus, Orf could facilitate RecA loading onto DNA by a similar mechanism (figure 3.2). The position of this motif in Orf, on the outer portions of the ring, supports the validity of this model (figure 3.3).

RecA	25-SIMRLGE-31
Orf	102-SRMVRGE-31
Rad51	85-GFTTATE-91
BRCA2	1523-GFHTASG-1529

Figure 3.2: Conservation of motifs in Orf and BRCA2 that match monomer-monomer interaction motifs in RecA and Rad51, respectively.

Mutation of Orf residues within this motif, if it is indeed involved in RecA loading, should affect any Orf-RecA interaction and ultimately, the formation of RecA nucleoprotein filaments.

A number of site-directed mutants were generated prior to the start of this project to evaluate the importance of conserved resides in the central cavity in DNA binding and also to test the possibility that there is a RecA-binding motif in Orf. The mutants made were Orf Q45A and W50A in the central cavity and R103E and V106E in the proposed RecA interaction motif (fig. 3.3) All mutants were purified as MBP fusions by amylose and heparin-agarose chromatography (GJ Sharples, unpublished results).

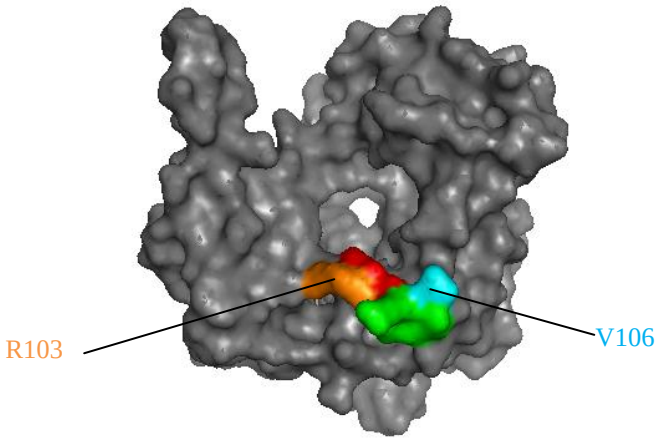


Figure 3.3: The RecA interaction motif of λ Orf

Orf protein-protein interactions have previously been the subject of study (see P. Reed (2006) and L. Bowers (2008)). Investigation of the intra-molecular mechanisms by which Orf binds ssDNA and how it performs its functions in λ-recombination is an area that has remained, until now, unexplored. The work described in this chapter aims to elucidate the effects of mutations in conserved residues of Orf in the suspected ssDNA binding and RecA recognition regions to explore its role in λ recombination.

3.2 Results

3.2.1 Mutations in the central cavity of Orf affect ssDNA binding

MBP-Orf Q45A and W50A proteins were investigated for their ability to bind 50 nucleotide ^{32}P-labelled DNA substrates. Wild-type MBP-Orf was used as a positive control and increasing amounts of protein were probed in a gel shift assay (fig 3.4)

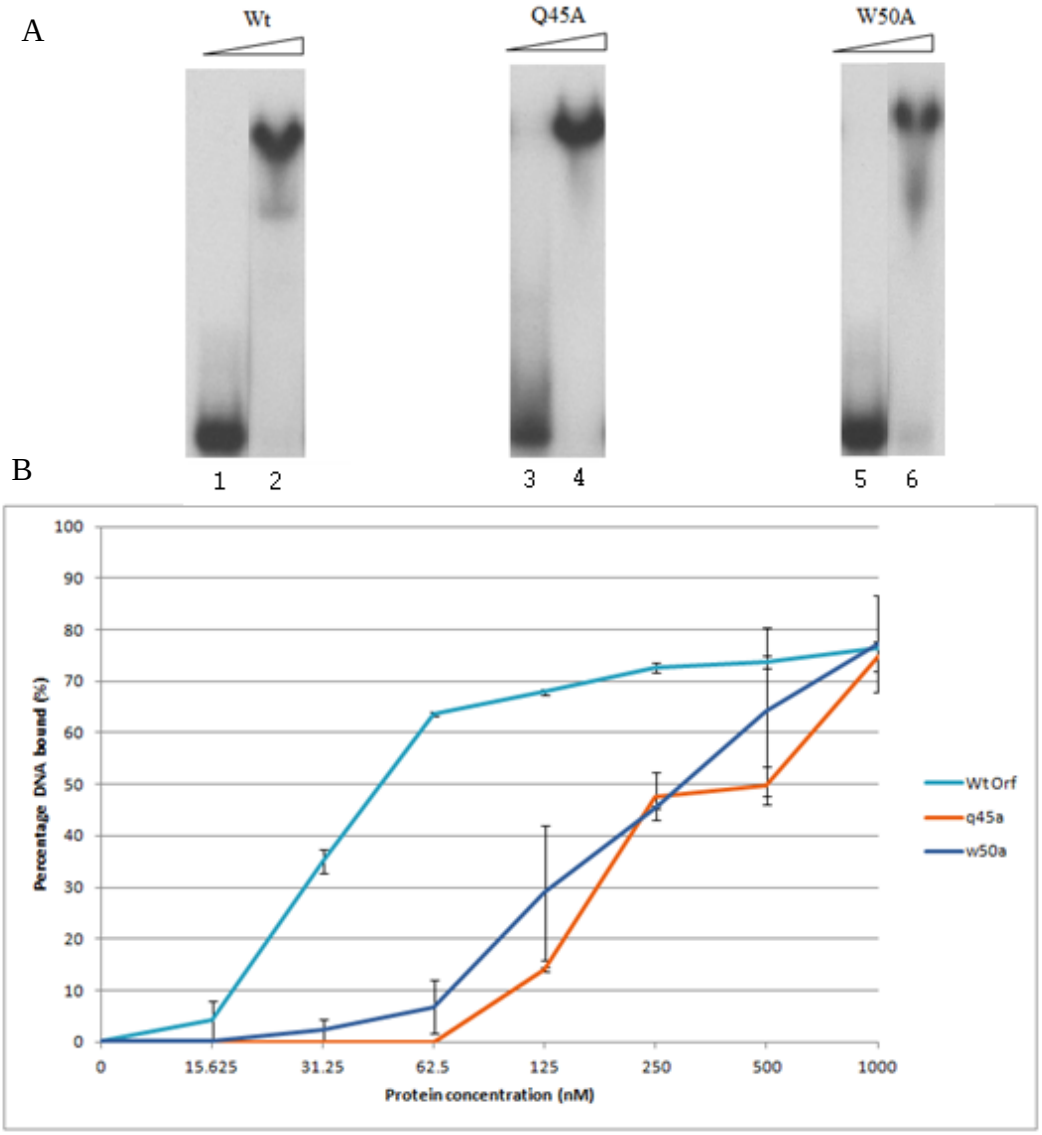


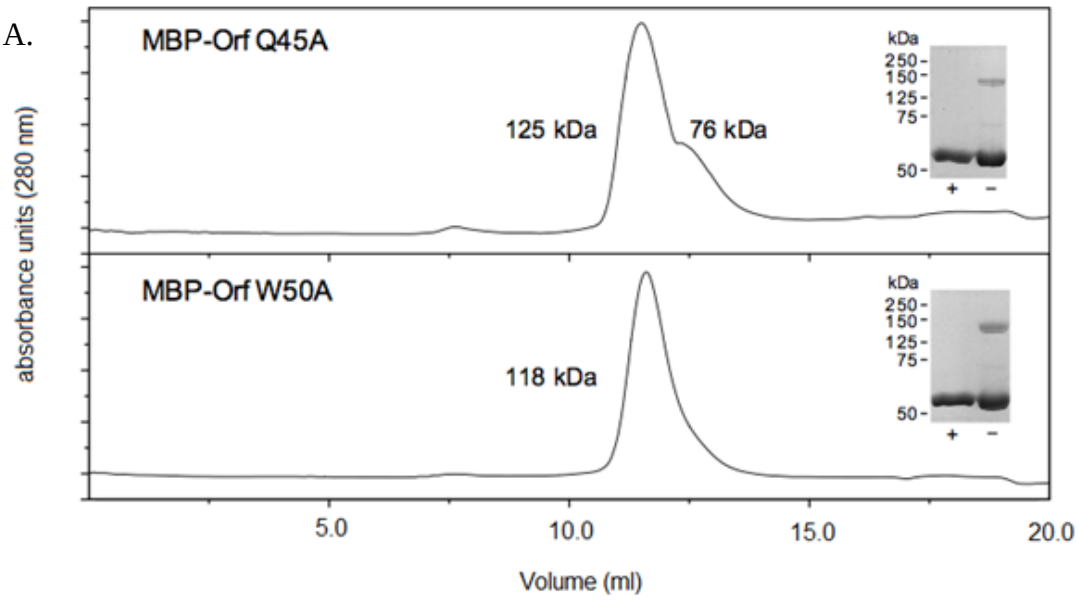
Figure 3.4 : Effect of point mutations in the Orf central channel (Q45A and W50A) on ssDNA binding . Proteins were mixed with 3 nM 32P labelled ssDNA in binding buffer. Samples were incubated on ice for 15 minutes prior to loading on 8% polyacrylamide gels in LIS buffer. A. Sample 32P gels for Wt, Orf Q45A and Orf W50A: lane 1 no protein, lane 2 500nM, lane 3 no protein, lane 4 500nM, lane 5 no protein lane 500nM. B. Quantification of ssDNA binding performed with 0, 15, 31.25, 62.5, 125, 250, 500 and 1000nM of each MBP-tagged Orf protein. Data are the mean of two independent experiments.

The binding assays reveal that the Orf Q45A and W50A mutants show significantly reduced ssDNA binding relative to the wild-type Orf protein. Maximal wild-type Orf binding occurs over 250 nM protein, whereas both mutants only shifted a similar amount of DNA at 1000 nM. The Orf Q45A mutant appears to affect DNA binding more severely than W50A, although the binding profiles are similar.

3.2.2 Quaternary structure of Orf Q45A and W50A Orf mutants in solution

It is possible that the mutations in the central cavity disrupt Orf dimer formation and this is responsible for the reduced capacity to bind ssDNA. To investigate this possibility, samples of the purified proteins were applied to a 25 ml Superose 12 gel filtration column. Orf Q45A produced a major peak consistent with dimer formation, although a smaller peak eluting later from the column suggests that a small proportion of the protein exists as a monomer. Orf W50A gave a single peak similar to the profile shown by the dimeric Orf wt (fig 3.5). Dimeric species were also detected with both proteins when samples were separated on SDS-PAGE without boiling (fig. 5A). Taken together, the results suggest that the Orf Q45A and W50A mutants do not grossly affect homodimer formation, although some monomeric species in Q45A may partly account for its defects in ssDNA binding.

A.



B.

MBP-Orf Mutant	Molecular Weight from SEC data (kDA)	Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
Q45A Peak 1	125	59531	2.1	Dimer
Q45A Peak 2	76	59531	1.28	Monomer
W50A	118	59473	1.98	Dimer

Figure 3.5: A. Size exclusion chromatography traces of Orf Q45A and W50A. 12.5% SDS-PAGE analysis of boiled (+) and unboiled (-) proteins are shown as insets. B. Table showing ratios of SEC data MW and expected MW to determine quaternary structure. Wt Orf froms dimer in solution (122)

3.2.3 Mutations in the RecA mimicking regions of Orf affect DNA binding

MBP-Orf V106E and R103E proteins carrying substitutions within the proposed RecA-interaction motif were investigated for their ability to bind 50 nucleotide ^{32}P-labelled DNA substrates. Wild-type MBP-Orf was used as a positive control and increasing amounts of protein were probed in gel shift assays (fig 3.6).

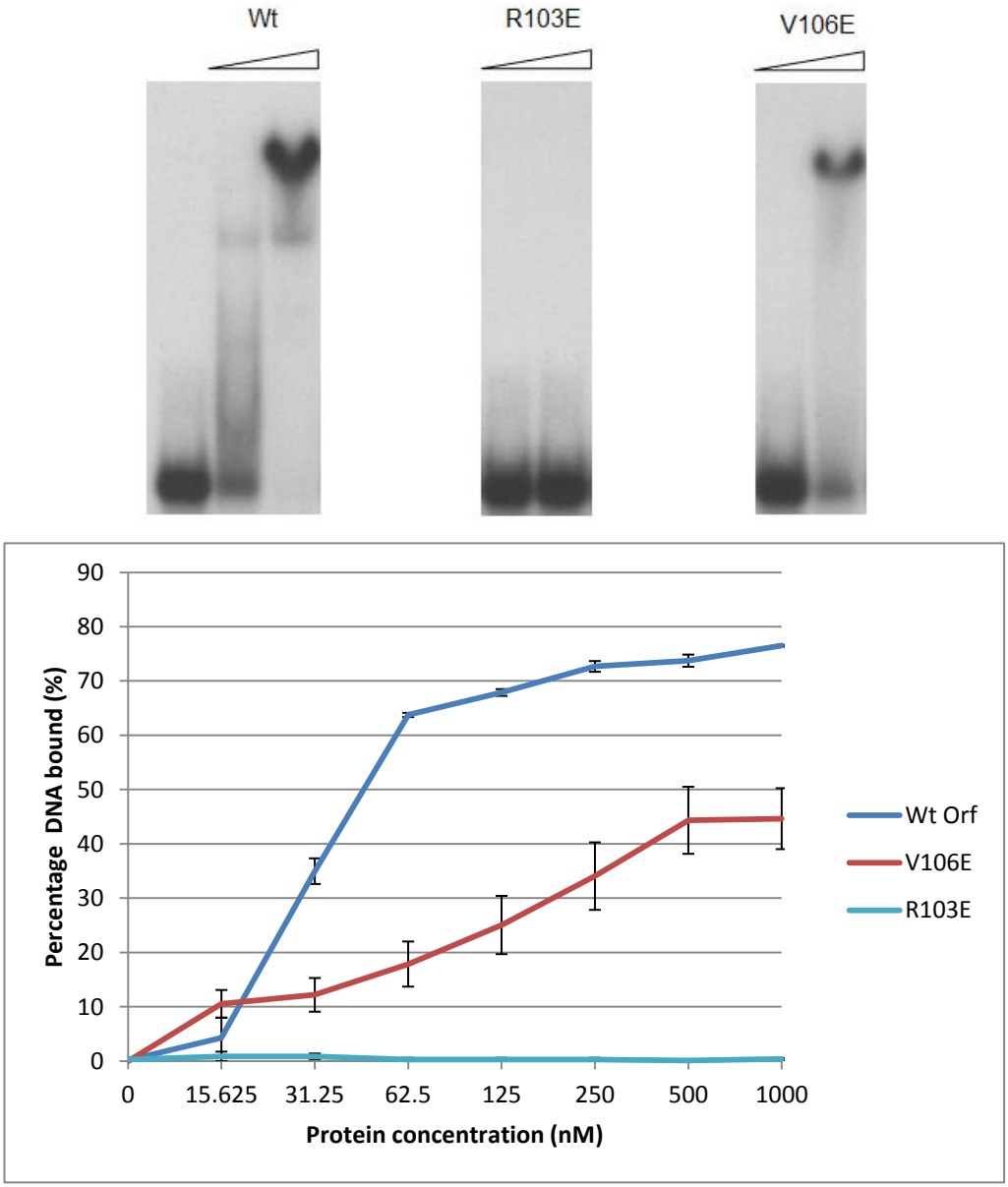


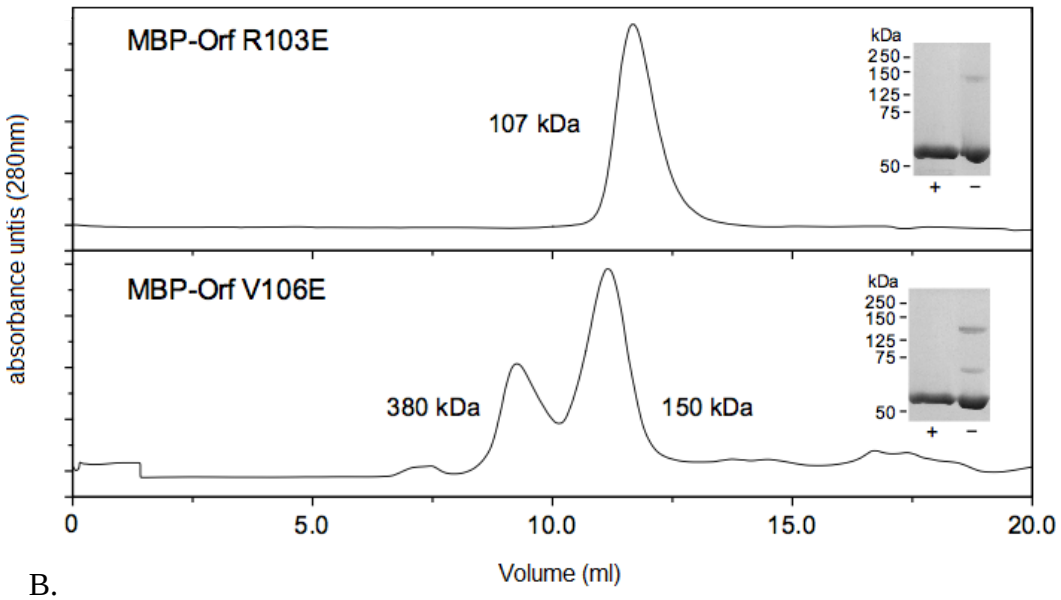
Figure 3.6: Effect of point mutations in the Orf-RecA interaction domain (R103E and V106E) on ssDNA binding. Proteins were mixed with 3 nM 32P labelled ssDNA in binding buffer. Samples were incubated on ice for 15 minutes prior to loading on 8% polyacrylamide gels in LIS buffer. A. Sample 32P gels for Wt, Orf R103E and Orf V106E: lane 1 = no protein control B. Quantification of ssDNA binding performed with 0, 15, 31.25, 62.5, 125, 250, 500 and 1000nM of each MBP-tagged Orf protein. Data are the mean of two independent experiments. Error bars represent one standard deviation

The DNA binding assays revealed that these mutants show significant defects in ssDNA binding. V106E exhibited reduced binding; where wild-type Orf bound 73& of the substrate at 250nM, V106E bound only 34% at the same concentration. While V106E retained some capacity to bind ssDNA, no DNA binding activity was detected with R103E under the conditions used.

3.2.4 Quaternary structure of Orf R103E and V106E Orf mutants in solution

To ascertain whether homodimer formation was affected in these mutants, R103E and V106E were applied to a gel filtration as described previously. While R103E provided a single peak consistent with the presence of a dimer (or potentially a trimer), V106E migrated as two distinct peaks; one indicating a dimeric molecular, the other a higher molecular weight aggregate approximately equivalent in mass to a hexamer (fig 3.7).

A.



B.

Orf Samples	Molecular Weight from SEC data (kDA)	Actual Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
R103E	107	59561.47	1.78	Dimer
V106E Peak 1	380	59618.5	6.37	Hexamer
V106E Peak 2	150	59618.5	2.51	Dimer

Figure 3.7 : A. Size exclusion chromatography trace of Orf R103E and V106E. 12.5% SDS-PAGE analysis of boiled (+) and unboiled (-) proteins are shown as insets.. B. Table showing ratios of SEC data MW and expected MW to determine quaternary structure.

3.2.5 Discussion

The data presented in this chapter indicate that the conserved residues within the Orf ring interior and those that match the oligomerisation motif in RecA protein are important for ssDNA binding. Mutation of two conserved residues within the central channel affected ssDNA binding without influencing homodimer formation. Those mutations in the proposed RecA-Orf interaction domain also affect ssDNA binding and at least one of them (V106E) also appears to induce Orf multimerisation.

Within the central channel of Orf, the Q45A mutant protein bound ssDNA with a higher affinity than W50A. Investigation of the quaternary structure revealed that while the W50A mutant formed stable dimers in solution, a small proportion of Q45A may exist as a monomer. If this is the case, this would explain, in part, the reduced capacity of Q45A to form complexes with ssDNA. In contrast, W50A appears to form a stable dimer in solution, suggesting that Trp50 is directly involved in ssDNA binding in keeping with its high conservation among Orf homologues. It also supports a model for DNA binding whereby ssDNA is threaded through the centre of the toroid or open clamp rather than passing across the shallow cleft across the protein surface.

Both mutant proteins with substitutions in the proposed RecA interaction domain showed marked defects in ssDNA binding. The significant defect in ssDNA binding by V106E at 1000nm exceeded that observed with the mutations studied in the central cavity. This could be due to altered protein folding affecting the tertiary and quaternary structure, as V106E was found to form relatively high molecular weight aggregates, as well as dimers. This may indicate a propensity for the protein to misfold, rather than a decrease in affinity for the substrate. Although the switch from the non-polar, neutral, amino acid valine to the polar and negatively charged glutamic acid and the proximity of V106E to the central channel in the Orf crystal structure may act attenuate substrate affinity, given that DNA is a negatively charged substrate.

Significantly, R103E was unable to form any detectable interactions with ssDNA although it does form a stable dimer in solution. In the Orf crystal structure, Arg103 makes contact with D68 and E70 in the other subunit. Within the wild-type dimer, these interactions take place between the positively charged arginine and the negatively charged aspartic and glutamic acid residues. The substitution of the positive arginine for glutamic acid would severely disrupt these interactions, possibly preventing intersubunit associations in this portion of the dimer. The gel filtration data, however, suggest that dimer formation is not impaired. The defects fit with the model where an inability to close an open Orf clamp would prevent stable binding to ssDNA.

In summary, the characterisation of two conserved residues in the central channel of Orf and two residues in the predicted RecA-interaction domain have been found to have substantial effects on ssDNA binding. The evidence is consistent with the interior and key intersubunit contacts being involved in DNA binding. Two mutations (Q45A and V106E) also affect Orf dimer formation and these defects in quaternary structure may be responsible for the DNA binding deficiencies observed with these proteins. Investigation of other conserved residues within the central channel of Orf and the construction of Orf mutants unable to interact with RecA, while retaining DNA binding, should help in determining whether the RecA-interaction motif can be functionally separated from ssDNA binding.

Chapter 4

Orf-151 Binds both ssDNA and dsDNA

4.1 Introduction

The *orf151* gene resides within the *E. coli* K-12 cryptic prophage DLP12, a remnant of a previous lambda-like phage insertion (123). Orf151 constitutes one of the most distantly related members of the 124 proteins in the Orf family and initial functional analysis has confirmed that it shares DNA and SSB binding activities with its lambda counterpart (Curtis et. al, unpublished results). Although Orf151 shares only 16% overall identity with λ Orf, it is almost identical to the *orf151* gene of phage 82 (Mahdi et al JMB). The main predicted differences between λ Orf and Orf151 reside in the α3 and β3 regions and in the C-terminal α-helix (fig 4.1), the latter previously linked with both DNA recognition and structural integrity (122).

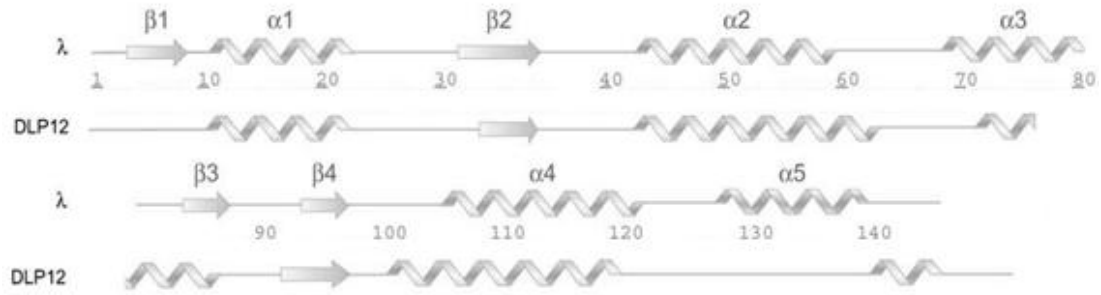


Figure 4.1: Comparison secondary structure of λ Orf determined by X-ray crystallography predicted (Jpred) secondary structure of Orf151.

An MBP-Orf151 protein was previously shown to associate with DNA and, as with λ Orf, exhibited a preference for substrates containing ssDNA (117). However, MBP-Orf151 formed complexes more readily and stably with dsDNA than λ Orf and showed reduced affinity for ssDNA when compared with λ Orf (117).

Orf151 interacted with SSB in far-western blots, ELISA and yeast-two hyrbrid assays (Curtis, F. *et al*, unpublished results). In these experiments, Orf151 actually enhanced binding to SSB relative to λ Orf; conservation of the SSB interaction, even in a distantly-related cryptic prophage Orf homologue, highlights the importance of this association for Orf family function.

4.2 Results

4.2.1 Orf151 binding ssDNA and dsDNA

The earlier experiments with MBP-Orf used protein that appeared to have been subject to some proteolysis and contamination with interacting SSB protein (Curtis, F. *et al*, unpublished results). We therefore utilised a fresh sample of MBP-Orf purified by amylose and heparin-agarose chromatography (Parkin, M., unpublished results). MBP-Orf151 was probed for its ssDNA and dsDNA binding properties on a 50 nucleotide ^{32}P-labelled ssDNA and and a complementary 50 bp ^{32}P-labelled dsDNA substrate (fig. 4.2).

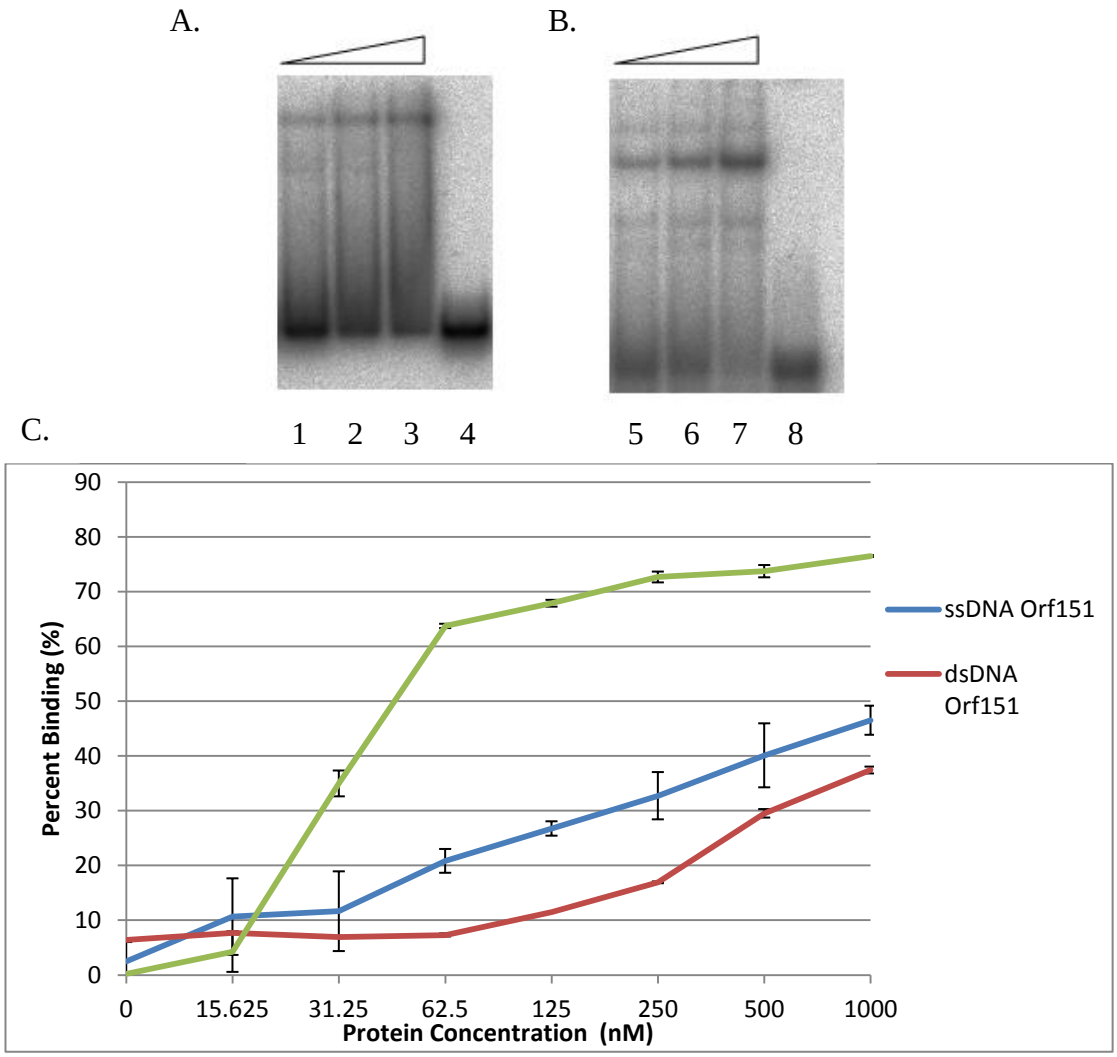
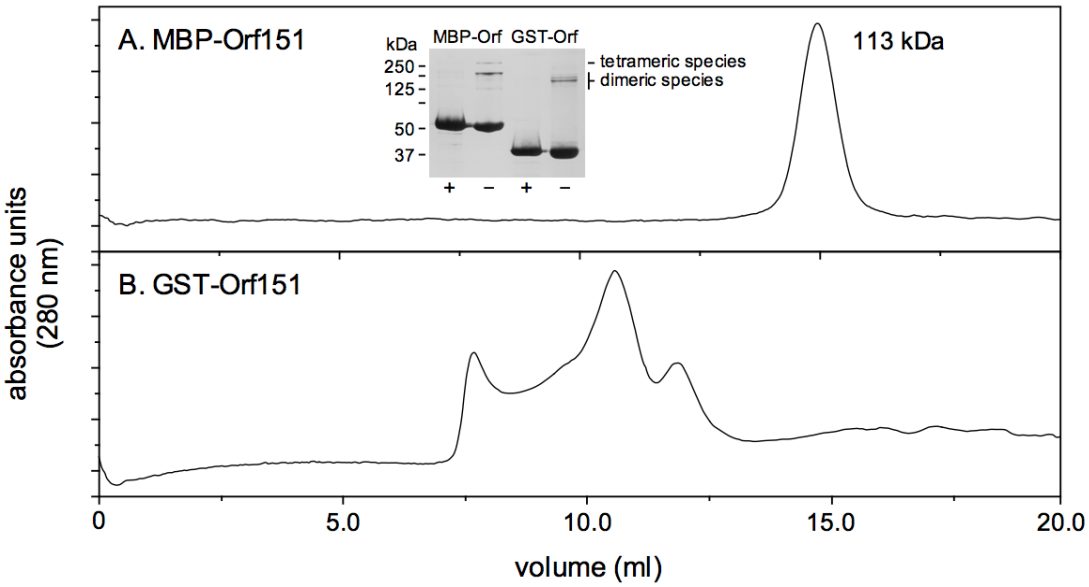


Figure 4.2 : Gel shift experiments showing MBP-Orf151 binding to dsDNA (A)and ssDNA binding (B). MBP-Orf protein was used at 250, 500, and 1000nM. Lanes 4 and 8 contained no protein. C. MBP-Orf151 binding to both ssDNA and dsDNA. MBP-Orf was used as a positive control. Data are the mean of two independent experiments. Error bars represent one standard deviation.

The DNA binding assays conducted here confirm that MBP-Orf151 has a significantly reduced capacity to bind ssDNA compared to MBP-Orf. Binding to dsDNA is less than ssDNA, though it is less discriminating than Orf, which exhibits considerable better binding to ssDNA over dsDNA of the same sequence (119).

4.2.2 The quaternary structure of Orf151

The ability of MBP-Orf151, and a previously-purified GST-Orf151 fusion, to form dimers or higher multimeric species was investigated by size exclusion chromatography. Proteins (1 mg/ml) were applied to a 25 ml Superose 12 gel filtration column and the molecular mass of peaks predicted in comparison with known molecular mass standards (fig. 4.3).



Orf151 Sample	Molecular Weight from SEC data (kDA)	Actual Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
MBP-Orf151	113	60373.51	1.87	Dimer
GST-Orf151 Peak 1	567.53888	44630.5	12.716391	Multimer
GST-Orf151 Peak 2	302.02558	44630.5	6.7672461	Multimer
GST-Orf151 Peak 3	141.67768	44630.5	3.1744588	Trimer

Figure 4.3: A. Size exclusion chromatography traces of MBP-Orf151 (A) and GST-Orf151 (B). SDS-PAGE gel analysis showing boiled (+) and unboiled (-) samples are shown in (A). C. Table showing ratios of SEC data with the experimentally-defined molecular weight and that predicted from the protein sequence to estimate quaternary structure.

The analysis reveals that MBP-Orf151 forms as a single peak consistent with formation of a homodimer. However, GST-Orf151 eluted in three, possibly four, peaks yielding estimated oligomers of 13, 7 and 3 subunits (fig. 4.3). The increased

multimerisation compared to MBP-Orf151 is most likely due to propensity of GST to dimerise and oligomerise under certain circumstances (124). Coupled with the evidence that MBP-Orf151 forms stable dimers in solution, it is likely that gel filtration yields anomalous size estimates with the fastest-eluting peaks representing MBP-Orf151 tetramers and octamers. Dimeric species were detected with both fusion proteins when samples were separated by SDS-PAGE without boiling; an apparent tetrameric species was observed with MBP-Orf (fig. 4.3A).

4.2.3 Discussion

MBP-Orf151 was investigated for its ability to bind ssDNA and dsDNA. Orf151 was found to bind both substrates with a similar binding profile, with a slight preference for ssDNA. Orf151 bound dsDNA with a higher affinity than Orf (119). There are several possible explanations for this disparity between the two Orf family proteins. Orf151, while similar to λ Orf, may have originally had a subtly different function (see Eta20, Chapter 5). Alternatively, an altered primary and secondary structure in Orf151 may have resulted from its senescence within the *E. coli* genome; with the lack of selective pressures normally found acting upon expressed genes, the gene gradually mutated and changed, resulting in a protein with severely compromised function. It is intriguing that MBP-Orf151 maintains a dsDNA binding affinity higher than that of Orf. Orf151 seems to have maintained the ability to form stable dimers in solution; it may be possible that the structure of Orf151 is more flexible due to quiescence to allow stronger binding to dsDNA. A crystal structure of Orf151 would provide a useful comparison with λ Orf and recent work has yielded crystals capable of diffraction using MBP-Orf151 (data not shown). In summary, MBP-Orf151 retains the ability to bind both DNA and SSB in keeping with its homology with the Orf family of proteins, supporting the idea that all members of this family share a similar function in phage DNA metabolism.

Chapter 5

ssDNA binding and endonuclease activities of Eta20

5.1 Introduction

Open reading frame 20 of the temperate *Staphylococcus aureus* phage φETA (phi-ETA) (125) was identified in PSI-BLAST searches as a distantly-related homologue of λ Orf in previous work (117; 119). Eta20 has an Orf-like domain with a C-terminal extension containing a zinc-finger motif related to HNH-family nucleases (126). Eta20 may therefore function to target and degrade DNA in contrast to its counterpart from λ (117). Previous work (117) confirmed that Eta20 possessed nuclease activity and that this activity resided within the C-terminal HNH-motif through characterization of an 82-residue C-terminal deletion (Eta20 ΔC82). In this Chapter, experiments to confirm these results were conducted and DNA binding and multimerisation properties were investigated.

5.2 MBP-Eta20 exhibits endonuclease activity lacking in Eta20 carrying an 82-residue C-terminal deletion

To probe the endonuclease function of Eta20 and Eta20Δ82, protein samples were incubated with ϕX174 virion DNA and the samples visualized by agarose gel electrophoresis (fig. 5.1).

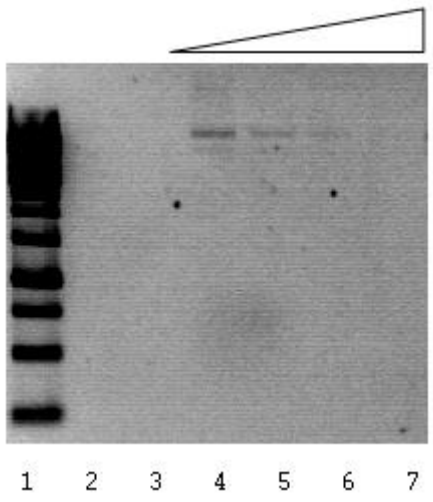


Figure 5.1: Endonuclease activity of MBP-Eta20 on circular ϕX174 ssDNA. Lane 1 1,000 kb ladder; lane 3 No DNA substrate, 500nM ETA-20; lane 4. 62.5 nM protein, lane 5. 125 nM protein, lane 6. 250 lane 7. 500 nM MBP-ETA20. Eta20Δ82 did not exhibit any nuclease activity (data not shown)

5.3 ETA20 binding to single-stranded DNA

DNA gel shift assays with MBP-Eta20 and MBP-Eta20 ΔC82 were undertaken to investigate the effect of the removal of the C-terminal HNH-nuclease domain on ssDNA binding (fig. 5.2).

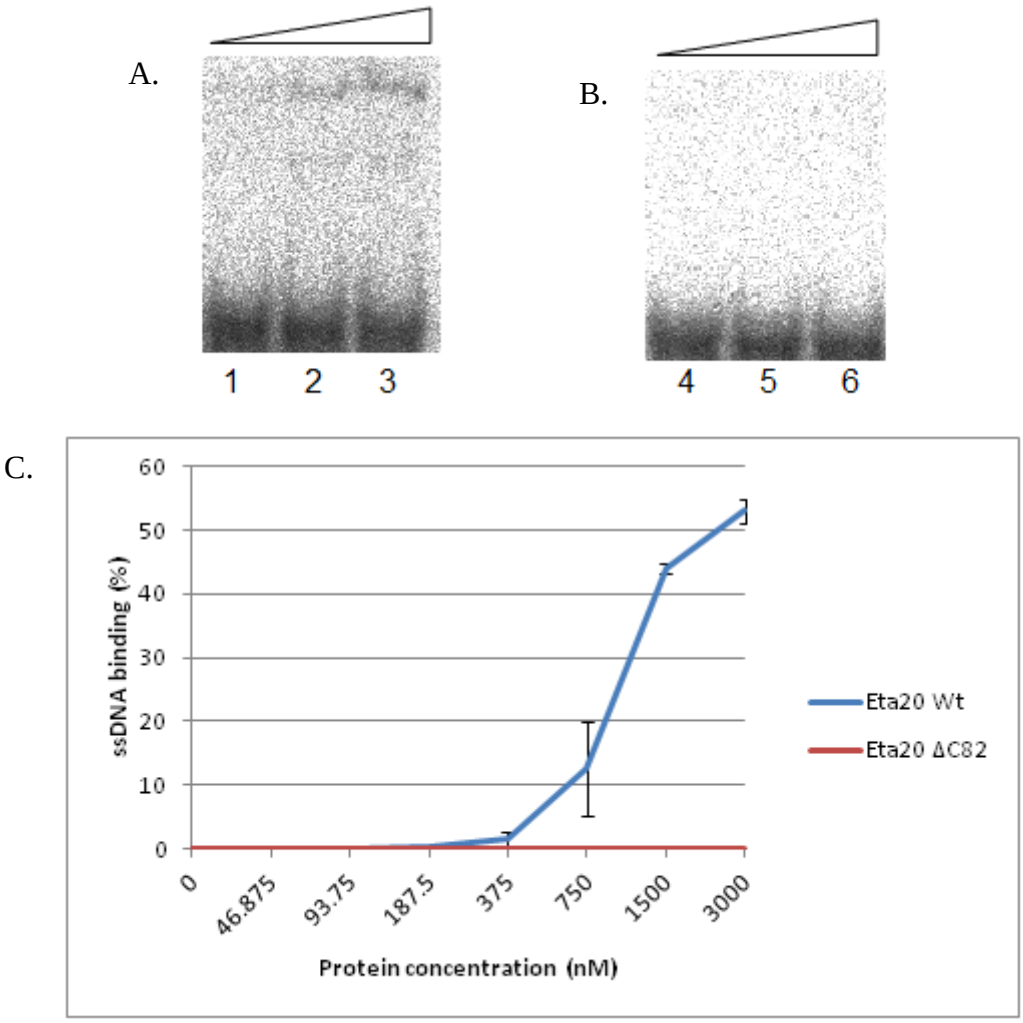
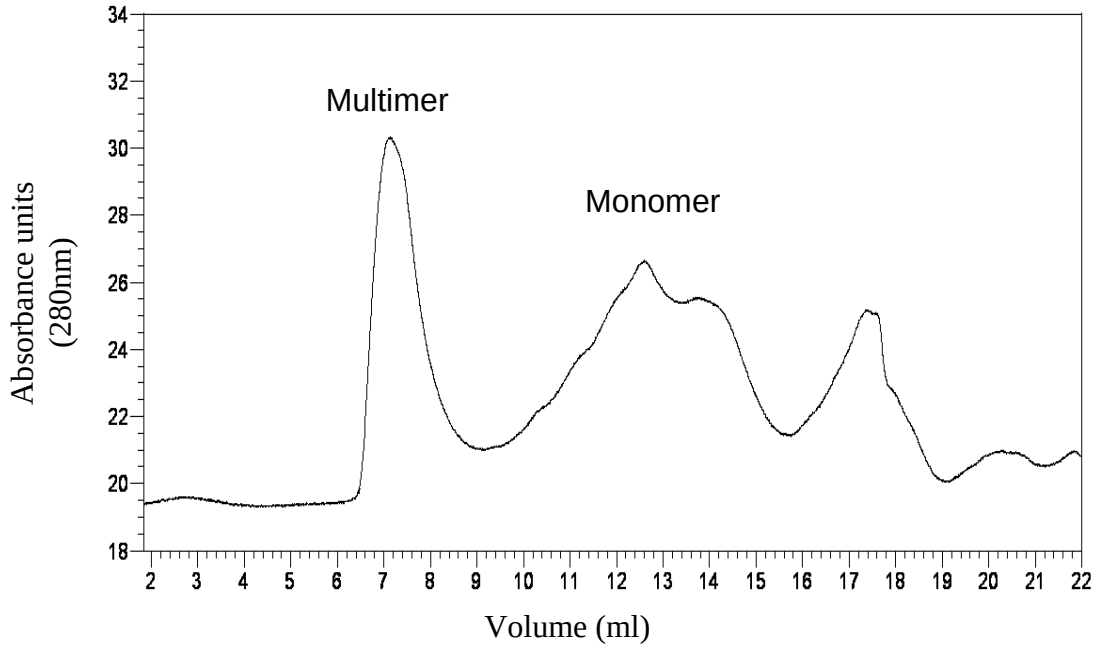


Figure 5.2: ssDNA binding assays of MBP- Eta20 (A) and MBP-Eta20 ΔC82 (B) containing 125nM, 250 nM and 500nM protein. (C.)Gel shift experiments showing MBP- Eta20 (A) and MBP-Eta20 ΔC82 binding to ssDNA. MBP-Eta20 and MBP-Eta20 ΔC82 protein was used at 250, 500, and 1000nM. C. MBP-Orf151 binding to both ssDNA and dsDNA. MBP-Orf was used as a positive control. Data are the mean of two independent experiments. Error bars represent one standard deviation.

Under the conditions used, MBP-Eta20 was found to bind ssDNA, but only at high concentrations (>1500nM). In contrast, Eta20ΔC82 was unable to bind ssDNA even at high concentrations up to 3000 nM. The ssDNA binding data support the Eta20 being a member of the Orf family. However, this region alone is insufficient for DNA binding lacks any ssDNA binding abilities.

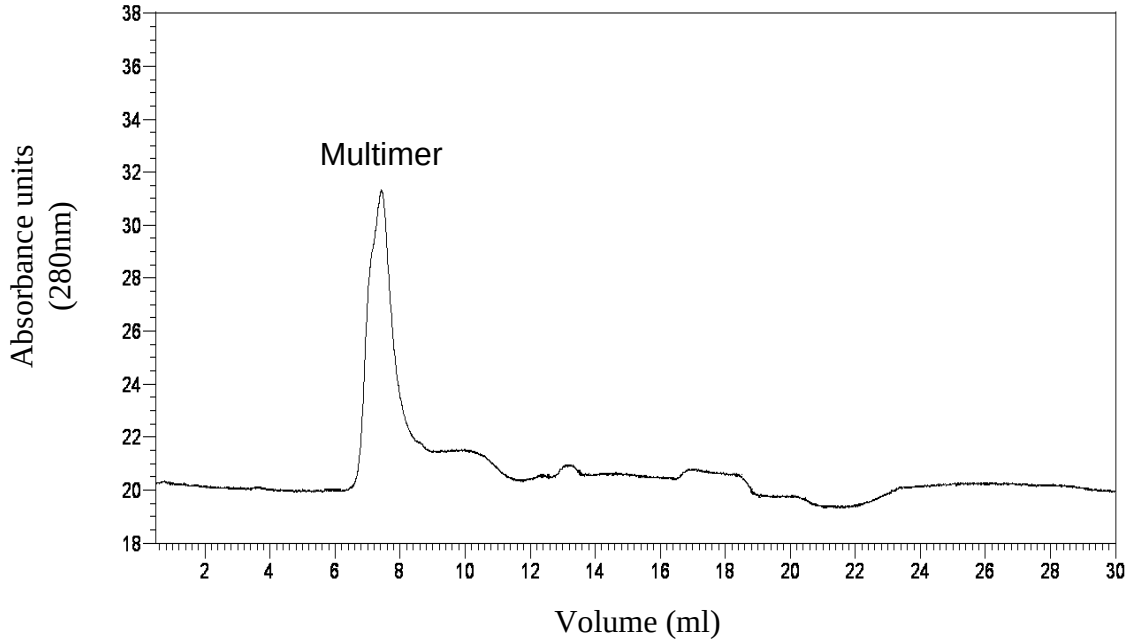
5.4 The quaternary structure of Eta20 and Eta20ΔC82

Gel filtration was utilized to investigate whether MBP-Eta20 and MBP-Eta20 ΔC82 form homodimers as with MBP-Orf and MBP-Orf151 using a 25 ml superose 12 column (fig. 5.3 and 5.4).



Samples	Molecular Weight from SEC data (kDA)	Actual Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
Eta20 Peak 1	2661.814311	76902.16	34.6	Multimer
Eta20 Peak 2	85.53422733	76902.16	1.1	Monomer

Figure 5.3: Size exclusion chromatography data for Eta20 at 1mg/ml in buffer A and 150 mM KCL. Table showing ratios of SEC data with the experimentally-defined molecular weight and that predicted from the protein sequence to estimate quaternary structure.



Samples	Molecular Weight from SEC data (kDA)	Actual Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
Eta20 ΔC82	2273.472903	67261	33.8	Multimer

Figure 5.4: Size exclusion chromatography trace for Eta20ΔC82 at 1mg/ml in buffer A and 150 mM KCL. Table showing ratios of SEC data with the experimentally-defined molecular weight and that predicted from the protein sequence to estimate quaternary structure.

The gel filtration data show that both MBP-Eta20 proteins form multimeric aggregates in solution. MBP-Eta20 also exists as a monomeric species, not seen with the mutant version of the protein, MBP-Eta20 ΔC82.

5.5 Discussion

Nuclease activity, ssDNA binding and quaternary structure were examined for MBP-Eta20 and an MBP-Eta20 ΔC82 deletion lacking the predicted HNH nuclease domain. Eta20 was found, under these conditions, to bind ssDNA. This DNA binding activity depended on the presence of the C-terminal nuclease motif, since no ssDNA binding was detected with MBP-Eta20 ΔC82. The results indicate that, although there is a degree of similarity between Orf and Eta20 in primary structure, Eta20 does not bind ssDNA in the same manner; instead Eta20 needs the nuclease domain extension for stable DNA binding. The endonuclease activity of MBP-Eta20 was confirmed a Volume (ml) in the deletion mutant. If Eta20 does share an evolutionary ancestor with Orf then its current function is significantly different. Further study is required to define the precise properties of the Eta20 nuclease activity.

Investigation of the quaternary structure of both MBP-Eta proteins by gel filtration showed that wild type Eta20 formed both monomeric and multimeric species in solution, while Eta20 formed only multimers. The absence of the nuclease domain may drastically affect protein folding and increase the propensity of the rest of the protein to multimerise. Subsequent circular dichroism (CD) experiments have revealed that both proteins show considerable aggregation and unfolding (Sanderson, J and Sharples, G., unpublished results); thus only a small proportion of the proteins may be properly folded, explaining the high concentration required to detect DNA binding and nuclease activities. The results from this chapter, therefore, needed to be interpreted cautiously.

Chapter 6

NinH

6.1 Introduction

NinH is a short, 68 residue, polypeptide located in the *ninR* region of the phage λ genome, just downstream of *rap* (*ninG*) that specifies a Holliday junction rsolvase (ref). It is found in numerous enterobacterial phages, although its function is currently undefined.

6.2 Bioinformatic analysis of NinH

A number of bioinformatic tools were used in the initial characterization of NinH. A search was conducted using the protein Basic Local Alignment Search Tool (BLASTp) from the National Centre for Biotechnology Information (http://blast.ncbi.nlm.nih.gov) using the non-redundant protein sequence database. Closely-related homologues from several enterobacterial phages, including P22, ES18 and ST104 were identified. Subsequent alignment of these sequences using the ClustalW2 multiple alignment tool (127) (http://www.ebi.ac.uk/Tools/msa/clustalw2/) illustrates the conservation between these phage proteins (Figure 6.1). No proteins from sources other than pahge or prophage were found.

```
Lambda    MTFSVKTIPDMLVE AYGNQTEVARRLKCSRGTVRKYVDDKDGKMHAIVNDVLMVHRGWSERDALLRKN
phi-4795  MTFTVKTIPDMLLE AYGNQTEVARILNCNRATVRKYIGDKEGKKHAVVNGVLMVHRGWGK
Sakai-VT  MTFTVKTIPDMLVE AYENQTEVARILNCSRNTVRKYTGDKEGKRHAIVNGVLMVHRGWGK
H19B      MTFTVKTIPDMLLE AYGNQSEVARILNCNRATVRKYIGDKEGKKHAVVNGVLMVHRGWGK
Lahn1     MTFTVKTIPDMLLE AYGNQSEVARILNCNRATVRKYIGDKEGKKHAVVNGVLMVHRGWGK
1639      MTFTVKTIPDMLVE AYENQTEVARILNCSRNTVRKYTGDKEGKRHAIVNGVLMVHRGWGK
ST64T     MTHTVKTIPDMLIE TYGNQTEVARRLSCHRNTVRRYLYDKEARYHAIVNGVLMIHQG
P22       MTHTVKTIPDMLIE TYGNQTEVARRLSCHRNTVRRYLYDKEARYHAIVNGVLMIHQG
PS34      MTHTVKTIPDMLIE TYGNQTEVARRLSCHRNTVRRYLYDKEARYHAIVNGVLMIHQG
ST104     MTHTIKTIPDMLIE TYGNQTEVARRLSCHRNTVRRYLYDKEARYHAIVNGVLMIHQG
Sfl       MSVKIQTIPELLIQ TRGNMTEVSRMLNCNRATVRKYAEDKEGKGHAIVDGVLMVHRGW
HK620     MNATIQTIPELLIQ TRGNQTEVARMLSCARGTVLKYNRDSKGERHVIVNGVLMVKQG
Kva       MTLSIKTIPDILVE VRGNQSEAARKLACSRNTILRYSRDTKAQFHAIVNGVLMVHQGGRGKACAQ
```

Figure 6.1: Alignment of λ NinH with homologues from phage and prophages. Sfl, *Shigella flexneri*, etc

NinH was then analysed with Position-Specific Iterated BLAST (PSI-BLAST) in an attempt to detect more distantly related homologues. This variation of the BLAST tool allows the top matches from an initial BLAST search to be probed again in an iterative manner to identify related proteins with lower similarity scores. Identifying such proteins, that have themself been previously characterized, may shed light on the function of NinH. PSI-BLAST was run with altered search parameters to return the maximum number of aligned sequences (20,000)

and the BLOSUM-80 scoring matrix, rather than the standard BLOSUM-62, as this substitution matrix is recommended for query lengths between 50 and 85 residues (128). NinH entered four iterations before no further new sequences were identified. Among the sequences recovered, a significant majority were transcriptional regulators or other DNA binding factors. These included FIS (factor for inversion stimulation) and NtrC (Nitrogen regulatory protein C) family transcription regulators. The homologous regions revealed that NinH possesses a potential helix-turn-helix motif.

6.2.1 NinH helix-turn-helix Motif

The helix-turn-helix (HTH) is a sub-structure that, at its simplest, constitutes 20-25 amino acids that comprises a two or three turn α-helix, followed by a four-residue turn, flanked by another α-helix of four turns (129). The helix-turn-helix motif was first identified within transcription regulatory proteins in model organisms, such as the catabolite gene activator protein (CAP) from *E. coli* (130) and the Cro repressor from bacteriophage λ (131). Often, transcription factors that include this motif are dimeric, allowing binding of appropriately-orientated inverted repeat nucleotide sequences. The final helix has been found to be generally more important for DNA recognition, as it is this that forms the major point of contact with the major groove of bound DNA, and is therefore termed the 'recognition helix' (129; 132). Helix-turn-helix sub-domains are found in both prokaryotic and eukaryotic proteins, with several variations based around the same molecular theme (129). Additionally, although the motif does make specific and non-specific contacts with DNA, it can also act as an anchor to align residues upstream or downstream of the HTH-motif that are also able to form specific or non-specific interactions with DNA (129).

To investigate the possibility that NinH contains a HTH-like domain, the sequence was compared with a selection of known HTH-forming protein sequences using the Dotlet flash program (http://myhits.isb-sib.ch/cgi-bin/dotlet) (133) with sample sequences taken from Brennan and Matthews (1989) (132) (data not shown). Results varied, with some of the sequences returning one or two possible HTH motifs within NinH, while others indicated there may be more (approximating the amino acid regions 16-38, 35-50 and 42-52). To visualize this more readily, sequences containing the HTH motif were entered into the Kalign program (134) (http://www.ebi.ac.uk/Tools/msa/kalign/) alongside NinH (figure 6.2).

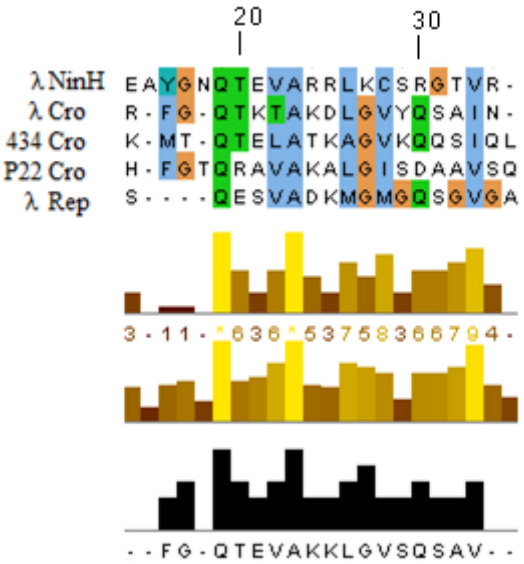


Figure 6.2: Multiple alignment between the predicted NinH HTH region (residues 40-60) with known HTH containing phage proteins.

This analysis flagged up a region extending from residue 16 to 33 in NinH matching the HTH-containing portions of several phage regulatory proteins. In order to give a more definitive answer, the primary sequence of NinH was subjected to analysis with a helix-turn-helix prediction program available from Institut de Biologie et Chemie des Proteines, Lyon, France (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html) (135) using a method updated by Dodd and Egan (1990) (136). The resulting analysis predicted, with 100% probability, that NinH contained a helix-turn-helix motif spanning 22 amino acids, from residue 17-38 (figure 6.3).

```
          10          20          30          40          50          60
          |           |           |           |           |           |
MTFSVKTIPD MLVEAYGNQT EVARRLKCSR GTVRKYVDDK DGKMHAIVND VLMVHRGWSE RDALLRKN
```

Figure 6.3: Results of HTH analysis on NinH showing the region (red) predicted to contain a HTH motif.

This area corresponds precisely to the portion aligned with the HTH-containing phage proteins and also extends the region further by several residues.

To provide a clearer picture of the potential structure of NinH, as well as support the identification of the predicted HTH motif, the primary sequence of NinH was analysed with Jpred3 secondary structure prediction software (fig. 6.4). With a high degree of predicted

accuracy, Jpred assigned the sequence highlighted by the HTH-prediction software as two α-helices connected by a short linker.

```
: 1---------11--------21--------31--------41--------51--------61------ :
: MTFSVKTIPDMLVEAYGNQTEVARRLKCSRGTVRKYVDDKDGKMHAIVNDVLMVHRGWSERDALLRKN :
: -----HHHHHHHHHH---HHHHHHHHH--HHHHHHHH------EEEEE--EEEEEEE------EE--- :
```

Figure 6.4: Jpred3 output for NinH with the probable HTH motif highlighted in cyan

Finally, to unify the findings from Jpred3 and the HTH prediction software, NinH was subjected to analysis with the tertiary structure prediction program Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (137). Both 'normal' and 'intensive' modelling modes were used. Under 'normal' modelling conditions, Phyre2 modelled 47% of NinH with 92% confidence based on a major template containing a 3-helical DNA/RNA binding fold belonging to the Fis family with 31% identity and 92% confidence. The major area of shared identity included the previously identified HTH motif. The 3D model generated consisted of the first three helices arranged in a non-parallel, twisted conformation as with other proteins containing the HTH domain (fig. 6.5).

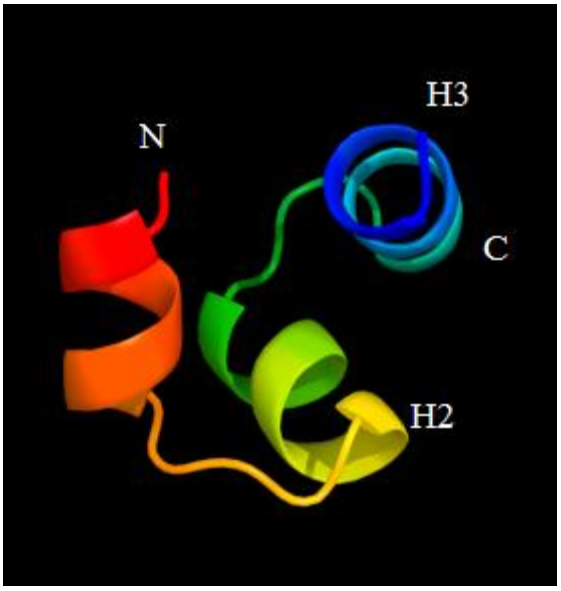


Figure 6.5: Jmol representation of NinH including the suspected HTH motif (H2= first helix, T= turn, H3= recognition helix)

Under 'intensive' modelling conditions, Phyre2 made extended predictions concerning the NinH N- and C-termini, although with a marked increase in uncertainty outside the area modelled under normal conditions and with minor changes to the HTH-region. A longer α-helix was modelled, using structural information from the TetR transcriptional regulator from

E. coli and the Brk repressor from *Drosophila melanogastor* that both contained an extended region of similarity (138). Again, Phyre2 was able to predict 47% of the structure with >90% confidence. Only the extreme termini contained regions with low prediction confidence (figure 6.6).

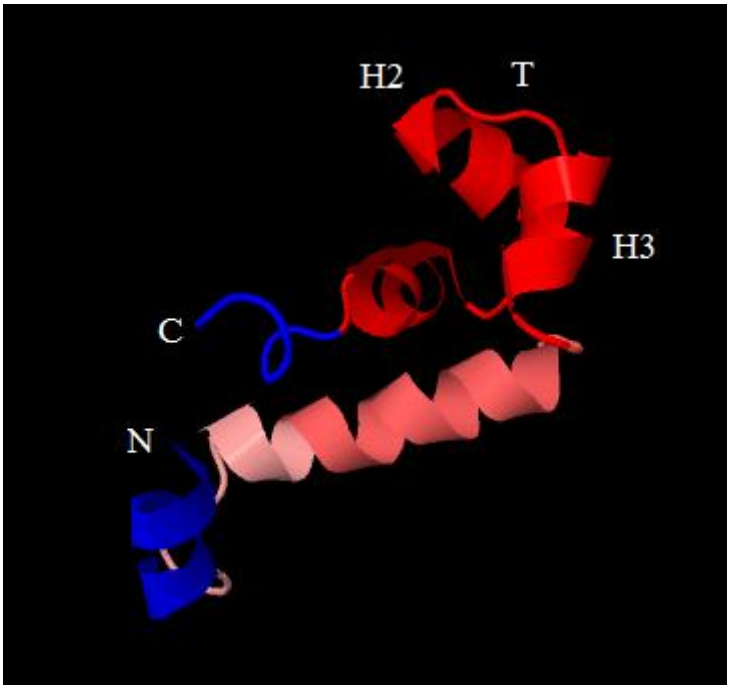


Figure 6.6: Extended model of NinH displaying prediction confidence (Red > Blue) based on *C. glutamicum* XxxX.

From the predicted tertiary structure of NinH, the placement of the HTH is appropriate for a genuine DNA interaction motif. In conjunction with the consensus obtained from both Phyre2 and Jpred3, the analysis fits with the designation of NinH as a small, DNA binding protein containing a HTH-motif.

6.2.2 Investigation of NinH-NtrC family homology

The PSI-BLAST yielded multiple NtrC protein matches, suggesting that a search of the literature and further bioinformatics was warranted. Nitrogen regulatory protein C (NtrC) is encoded by the *glnG* gene. NtrC regulates genes involved in nitrogen assimilation under conditions of limited nitrogen availability (139). NtrC appears to be involved in regulating approximately 2% of the entire *E. coli* genome, with the vast majority of the operons affected being associated with transport (139). Under low nitrogen conditions, *E. coli* activates transporters that scavenge nitrogenous compounds from the environment or periplasm; NtrC is part of a two-component system that regulates this response. The second component is

NtrB, a membrane-associated protein kinase that activates NtrC through autophosphorylation, after NtrB is autophosporylated during nitrogen and carbon-limited growth. However, given the bioinformatic findings described above, it may simply be the case that, as a DNA binding protein, NtrC shares homology with NinH solely with the HTH or a HTH-like domain, although it is possible that they evolved from a common ancestral motif.

The highest scoring NtrC sequence found by PSI-BLAST belonged to bacteria from *Magnetococcus* species. Using the LALIGN multiple subsegment alignment tool with default search parameters (http://www.ch.embnet.org/software/LALIGN_form.html) (140) set with the Blosum45 matrix, a homologous segment with an overlap of 32 amino acids was found with an identity of 37.5%. The area of homology spanned amino acids 8-39 of NinH and amino acids 468-499 on NtrC. NinH and an area containing the highlighted portion of NtrC was entered into KALIGN with the output shown in figure 6.7.

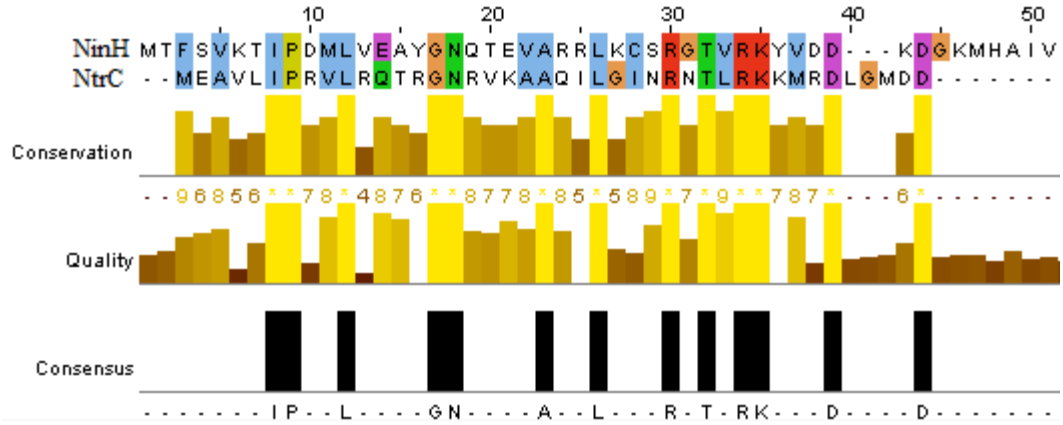


Figure 6.7: KALIGN search result in ClustalW format, showing the alignment of NinH and NtrC from *E. coli*

As before, this alignment is centred on the area highlighted in earlier bioinformatic searches as likely to contain a HTH motif. A search of the current literature revealed that only the structure of the N-terminal receiver domain of NtrC has been solved in both inactivated and activated forms (141; 142). The area of alignment with NinH lies at the far C-terminal end of the primary sequence; an area yet to structurally resolved. Directed by the aligment produced by Kalign, the Magnetococcal NtrC was also subjected to a HTH-motif search (135). This highlighted a sequence starting at residue 477 and returning a HTH-probability of 100% (see figure 6.8). This motif was found to be common to every NtrC homolog checked in the PSI-BLAST generated list, all with 100% probabilities of containing a HTH motif.

```
  470                              500
  |                                |
RVLRQTRGNRVKAAQILGINRNTLRKKMRDLGMDD
```

Figure 6.8: Result of HTH prediction software on Magnetococcal NtrC protein. The highlighted area is predicted to contain a HTH motif.

These results are consistent with a short, common, structural motif within NtrC homologues that is also shared by NinH.

6.2.3 Investigation of NinH-Fis homology

The PSI-BLAST searches highlighted a possible relationship between NinH and the Factor for Inversion Stimulation (Fis) protein. The specific Fis proteins identified originated from *Chromobacterium violaceum, Candidatus glomeribacter gigasporarum*, *Burkholderia* species and *Haemophilus haemolyticus*. Fis is a small, bacterial, nucleoid-associated protein that defines one of the four major families of prokaryotic histone-like proteins that share structural resemblances and functional similarities (143). The diverse functions of Fis include DNA bending and packaging (144), regulation of site-specific recombination events (145) and transcriptional regulation of a large number of different genes during various stages of growth (146); among the latter are rRNA and tRNA genes (147) and those involved in biofilm formation (148), virulence (149), phage integration and excision reactions, as well as autoregulatory functions (150). Fis was originally discovered due to its requirement in site-specific recombination involving the Hin and Gin recombinases that function by catalyzing the inversion of segments of DNA, from where it draws its full name (151). Fis is also responsible for interacting with the λ excisionase (Xis) to catalyse the extraction of the lambda prophage from the host chromosome (152).

Fis in *E. coli* is a homodimer of 98 residues, with two α-helices from each monomer forming a tightly-interacting globular domain with a pair of helix-turn-helix motifs protruding outwards along the same edge (153) (figure 6.9). The 24 residues present at the N terminus of each subunit are highly disordered in the absence of DNA (153). Mutational analysis and Fis-DNA co-crystallization has revealed that when bound to DNA, the N-termini form β-hairpin arms that can associate with other Fis-DNA complexes (151). The removal of these β-hairpin arms prevents Fis-activation of DNA invertases (150).

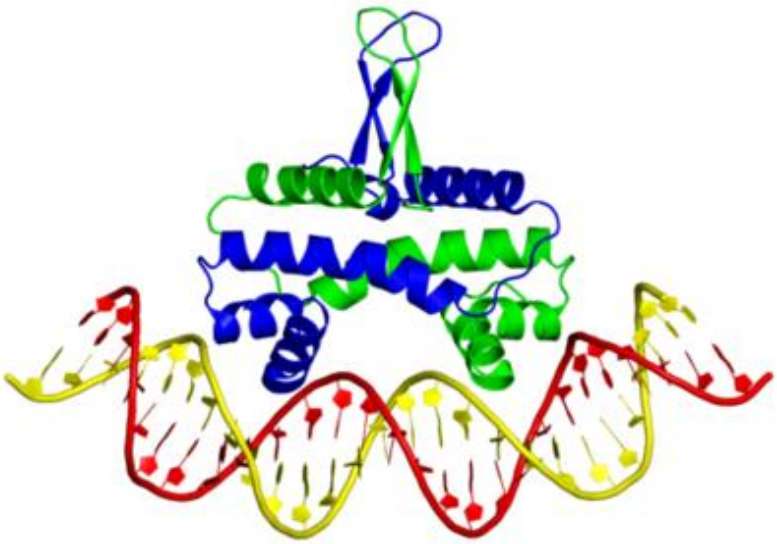


Figure 6.9: 3-dimensional structure of *E. coli* Fis showing monomer A (blue) and B (green) bound to double-stranded DNA (red and yellow) (154).

Fis is one of the most abundant DNA-binding proteins found in *E. coli* during early exponential growth, reaching in excess of 50,000 copies per cell just before the commencement of cell division, while only present at less than 100 copies per cell during stationary phase (155). Fis condenses DNA by binding 21 base-pair segments *via* its nonspecific DNA-binding and bending activities (156). However, Fis is also capable of forming stable complexes with specific DNA sequences found in regulatory areas of the genome, as expected given its numerous DNA interactions (157). Sequences of 15 base pairs have been found to specify high-affinity Fis binding sites, although they are poorly related at the sequence level (151). Evidence suggests that Fis initially targets DNA regions based upon the width of the minor DNA groove, followed by induced fit binding to the major groove following DNA binding; Fis maintains its structural integrity and pulls the DNA over its two HTH motifs, causing the DNA to bend by ~65° (151). Each HTH is held in place by multiple side-chain contacts. Of particular importance are the side chains of residues asparagine-73, threonine-75, arginine-89, as well as the amides from glutamine-74 and threonine-75, which act to pull the phosphate backbone towards the HTH domains (151). Mutational studies have

shown that, of these residues, the polar side chain of Asn73 plays an especially vital role in binding and bending DNA (151).

Three adjacent residues, asparagine-84, threonine-87 and lysine-90, have been found to stabilize the recognition helix. Of these particular residues, the side chain of Thr87 seems to be the most important (158), while Lys90 appears to play a significant role in non-specific interactions (156). Asn84 and Arg85 are the only residues in the recognition helix so far that have been found to form vital contacts with bases in the major groove; the loss of Arg85 leading to complete loss of all Fis-DNA interactions (158; 156; 151)

The highest scoring Fis homolog from the PSI-BLAST NinH searches belonged to *Chromobacterium violaceum*. Lalign (140) analysis highlighted three areas of alignment: 38.7% alignment over 31 amino acids from NinH (residues 11-41) and Fis (residues 46-76) (figure 6.10), 25% identity over 24 amino acids on NinH (residues 30-53) and Fis (residues 12-35). The final section of overlap was relatively small: 6 amino acids sharing 21.4% similarity from residue 50-63 in NinH and residue 44-57 in Fis. NinH was also compared to *E. coli* Fis using the same approach. This also yielded three sections of overlap with higher degrees of similarity: 52.6% identity over 19 amino acids 17-35 and 72-90, 28.6% over residues 45-65 and 60-80, and finally, 30% identity over residues 12-31 and 53-72. Only the first alignment yielded two discernibly similar regions, that of the HTH motif.

```
          20        30
NinH   GNQTEVARRLKCSRGTVRK
       ::::::  .  .:::.::
Fis    GNQTRAALMMGINRGTLRK
              80        90
```

Figure 6.10: NinH and *E. coli* Fis alignment of HTH domain highlighted by Lalign with an overall identity of 52.6%. Residues in red indicate vital or important residues in Fis, all bar Asn84 are conserved in NinH. The area of no homology correlates with the predicted 'turn' sub-domain of the HTH motif.

In order to quantify the conservation of these residues, NinH and *E.coli* Fis were compared using the T-COFFEE multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/tcoffee/) (159) (figure 6.11)

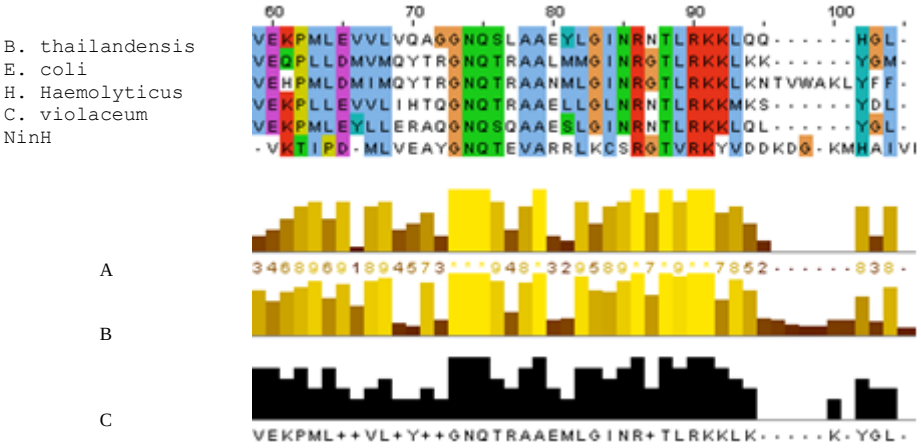


Figure 6.11: Multiple alignment of lambda NinH and Fis homologues: A. Conservation Score, B. Quality Score, C. Consensus Score

From this analysis, it is clear that the HTH motif in NinH shares not only the gross architecture found in HTH motifs, but also all the important functional residues found in the HTH motif found in Fis; even the neutral polar side chain of the asparagine at position 84 in Fis is substituted by the biochemically similar serine-30 in NinH. In terms of secondary structure, the HTH motifs are situated at the far ends of each molecule; NinH towards the N terminus, Fis towards the C-terminus. As the N-terminal end of Fis is involves other functions, such as the Fis-dimer and invertase interface, other functions may reside in the C-terminal region of NinH.

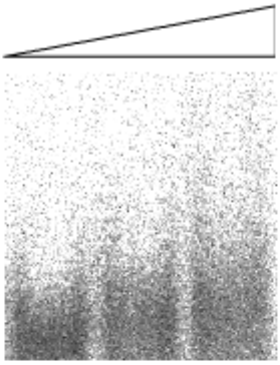
6.2.4 Conclusion

The function of NinH is still yet to be fully defined, but this work indicates that NinH may be a Fis-like DNA binding protein. Given the high degree of similarity between the HTH-motif of Fis and that of NinH, it may bind in a similar manner and have a similar or a related function. NinH could play a role in λ gene regulation, or it may have a role in site-specific recombination. It may have multiple roles in a multitude of processes, much like Fis. Evidence that NinH does not dimerise may indicate a divergent function as dimerisation is important for Fis function, although this needs to be investigated (see next section).

6.3 NinH binding to ssDNA

Both wild-type and histidine-tagged NinH proteins had been purified previously (Curtis, F. and Bowers, L., unpublished results). DNA binding experiments utilising these proteins and a 50 nt ^{32}P-labelled ssDNA substrate showed that NinH is capable of binding single-stranded DNA (figure xxxx).

A.



B.

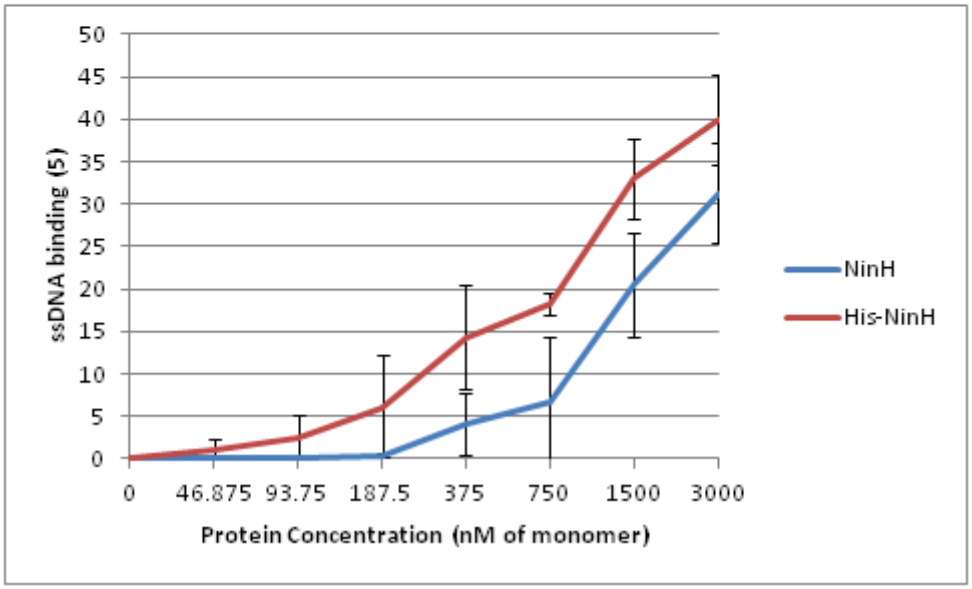
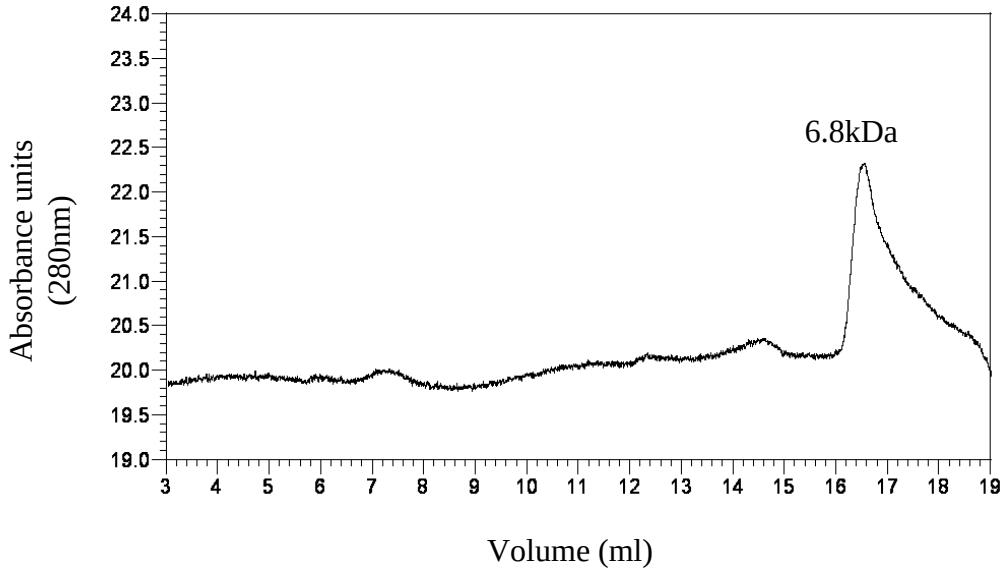


Figure : NinH and His-NinH binding to ssDNA Proteins were mixed with 3 nM ^{32}P labelled ssDNA in binding buffer. Samples were incubated on ice for 15 minutes prior to loading on 8% polyacrylamide gels in LIS buffer. A. NinH binding to ssDNA. B. Quantification of ssDNA binding performed with 0, 47, 94, 188, 375, 750, 1500 and 3000nM of NinH or His-NinH proteisn. Data are the mean of two independent experiments. Error bars represent one standard deviation.

The results show that NinH is binds ssDNA, although relatively high concentrations of protein are required. His-NinH appears to have a higher affinity than the purified wild-type NinH. This may be due to a stabilising effect from the tag that aids protein folding.

6.4 The quaternary structure of NinH

In order to investigate the quaternary structure of NinH, NinH was applied to a 25 ml Superose 12 gel filtration column as before.



Samples	Molecular Weight from SEC data (kDA)	Actual Molecular Weight (g mol^{-1})	Ratio of Sec Data : Calculated MW	Aggregation
NinH	6.8	7881.1	0.87	monomer

Figure : Size exclusion chromatography data for NinH.

NinH yielded a single peak consistent with the formation of a monomeric species in solution and does not dimerise like Fis. As Fis has a function dependent on its ability to dimerise, NinH may have a distinct function.

6.5 Discussion

NinH was predicted by bioinformatic analyses to be a HTH-type protein. It displayed limited ssDNA binding properties and formed a monomer in solution. Given that it shares key residues found in Fis but does not dimerise, it likely has distinct DNA binding functions, although it may interact with DNA in a similar manner to Fis but without dimerising. Alternatively, NinH may dimerise upon association with DNA. Like Fis, it may preferentially bind dsDNA, rather than ssDNA, and may interact with other proteins. Further work is

needed to determine the affinity of NinH for dsDNA and its role in lambda gene regulation and/or DNA metabolism.

Chapter 7

Conclusions and future directions

The study of lambda recombination proteins has helped to show that bacteriophages are capable of conducting recombination with a considerably smaller repertoire of less complex proteins than that possessed by their prokaryotic hosts. This may not be too surprising since the problems encountered by a phage are smaller in number and they can always harness host recombinases as required. In this project, Orf family proteins and mutant derivatives were probed for their ability to bind DNA and whether they formed oligomers in solution. A further phage λ protein, NinH, was characterised, largely by undertaking a detailed bioinformatic analysis.

Experiments on the phage λ Orf recombinase, revealed that conserved residues in Orf affect both ssDNA binding and homodimerisation. Mutants (W50A and Q45A) in conserved residues in the central channel of Orf displayed significant defects in DNA binding. Trp50 could potentially stack with bases to help bind ssDNA. The Q45A mutant also showed reduced DNA binding and this residue could make contacts with the nucleotide bases, or alternatively with the backbone. In the case of the latter, some of the defects may in part be due to reduced homodimer stability. The fact that these mutations show defects in DNA binding supports a model where ssDNA is accommodated through the centre of the Orf ring. Since Orf can bind a gapped duplex substrate (Maxwell et al, 2005), and the centre of the Orf ring is too narrow to encircle dsDNA, the data supports a model whereby the Orf protein opens, like a clamp, to bind ssDNA. This model also fits with the majority of conserved residues lining the walls of the Orf aperture. Further investigation into the other conserved residues in this region, (e.g. R41 and N46), should shed further light on the elements critical for Orf binding to ssDNA.

Mutations in the predicted RecA oligomerisation recognition domain have also been found to be important for DNA binding. This may indicate that the Orf-RecA interaction is coupled to ssDNA binding, perhaps undergoing a conformational change that enable Orf to act as a nucleation site for RecA polymerisation. Any such changes in Orf may also facilitae displacement of SSB from the ssDNA substrate. The Orf R103E mutant was completely defective in binding ssDNA, yet retained the ability to form dimers in solution. Arg103

makes contacts with residues in the other Orf subunit so may be important in stabilising dimer formation. Since dimer formation was not affected by this mutation it may be that closing of the Orf ring to clamp onto ssDNA is prevented. It is also possible that this residue contacts ssDNA directly and is important for forming stable DNA complexes. The Orf V106E mutant also showed a defect ssDNA binding, although less severe than in R103E, it was a more impaired in binding than W50A and Q45A. Given the radical nature of the changes (inserting a negatively charged glutamic acid), it may simply be that the mutations were too extreme to allow a proper evaluation on Orf function. While R103E formed dimers in solution, V106E seemed prone to multimerisation, meaning that its reduction in ssDNA binding may be due to impaired folding or improper aggregation, rather than decreased ssDNA binding affinity. Investigation of alternative, less extreme, substitution mutants may prove useful. Further, more quantitative, analysis of DNA binding should also be conducted. This could include isothermal titration calorimetry and fluorescence anisotropy. Both mutants should be tested to see if an interaction with RecA is blocked. Affinity pull-down assays, far-western blotting, size-exclusion chromatography or yeast two-hybrid assays are all approaches that could be employed.

Two highly diverged Orf homologues, Orf151 and Eta20 were analysed as aprt of this study. A distantly-related homologue of Orf (Orf151) from an *E. coli* prophage showed remarkably similar properties to its λ counterpart. It differed in showing a slightly higher affinity for dsDNA but otherwise behaved as a homodimer. Given the distant relationship to λ Orf and the ease with which milligram quantities of this protein can be recovered, obtaining a crystal structure of this protein would be extremely useful. It is possible that an ‘open’ clamp conformation could be obtained to help explain how Orf binds DNA. Eta20, an Orf homologue harbouring a C-terminal HNH nuclease domain, was found to be functionally distinct from Orf. It is possible that Eta20 acts in a solely degradative capacity; in the evolutionary past it may have targeted degradation at ssDNA gaps, however the results presented here suggest that it may function to degrade bacterial chromosomal DNA non-specifically. The ssDNA binding activity of Eta20 depended on the presence of the HNH nuclease extension; hence the Eta20 ΔC82 mutant protein failed to bind ssDNA. Eta20 did not form a homodimeric complex in the manner of Orf, instead the majority of the protein was in monomeric or multimeric form. It may be that expression in *E. coli*, rather than its normal host *Staphylococcus aureus*, means that proper folding of Eta20 is not achieved.

Further work on the Eta20 DNA binding and nuclease activities using appropriate ssDNA and dsDNA substrates would be valuable in defining the possible roles that this phage Orf variant fulfils *in vivo*.

Preliminary investigation of the λ NinH protein by a bioinformatic approach highlighted several interesting aspects of this short polypeptide located in the *orf* and *rap* operon. As a small, helix-turn-helix protein, it shares a conserved motif with the *E. coli* Fis DNA binding protein. While Fis exists as a dimer, NinH was found to be monomeric in solution, although it could potentially dimerise when interacting with DNA. NinH was only tested for ssDNA binding, which it was able to bind to with relatively weak affinity compared to many of the proteins tested. Further investigation of NinH on additional DNA substrates, including dsDNA, should help clarify its function and substrate preferences. Its small size may permit structural determination using nuclear magnetic resonance methodologies. More direct approaches to examine its impact on site-directed recombination, homologous recombination and phage gene regulation could also be undertaken.

Bibliography

1. *Cryptic prophages help bacteria cope with adverse environments.* **Wang, X., et al.** 147, 2010 21-December, Nat. Comm., Vol. 1, pp. 1-9.
2. *Bacteriophages: Evolution of the majority.* **Hendrix, R.W.** 2002 Vol. 61, pp. 471-480.
3. *Bacteriophage Therapy.* **Sulakvelidze, A. Alavidze, Z., and Morris Jr., J.G.** 3, 2001, Antimicrob.Agents Chem., Vol. 45, pp. 649-659.
4. *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion.* **Brussow H., Canchaya C., and Hardt WD.** 3, 2004 September, Microbiology and Molecular Biology Reviews, Vol. 68, pp. 560-602.
5. *Prophages and bacterial genomics: what have we learned?* **Casjens, S.** 2, 2003, Mol. Microbiol., Vol. 49, pp. 277-300.
6. *The lysis-lysogeny descision of phage lambda: explicit programming and responsiveness.* **Herskowitz, I. and D., Hagen.** 1980, Ann. Rev. Gen, Vol. 14, pp. 399-445.
7. **Luria, S.E., et al.** *General Virology.*: John Wiley & Sons, 1978. 0-471-55640-8.
8. *Phage resistance in lactic acid bacteria.* **Sanders, M.E.** 1988 24-11, Biochimie, Vol. 70, pp. 411-421.
9. *Isolation of the Bacteriophage lambda receptor.* **SCHWARTZ, M. RANDALL-HAZELBAUER L. and.** 3, 1973 December, J. Bacteriol., Vol. 116, pp. 1436-1446.
10. *Nucleotide sequence of bacteriophage lambda DNA.* **Sanger F., Coulson A.R., Hong G.F., Hill D.F. and Petersen G.B.** Cambridge, 1982, J. Mol. Biol., Vol. 162, pp. 729-773.
11. *Comparative molecular biology of lambdoid phages.* **Campbell, A.** 1994, Ann. Rev. Microbiol, Vol. 48, pp. 193-222.
12. *Evidence that the normal route of replication-allowed Red-mediated recombination involves double-chain ends.* **Thaler, D.S., Stahl, M.M. and Stahl F.W.** 10, 1987 , EMBO Journal, Vol. 6, pp. 3171-3176.
13. *DNA struture specificity of Rap endonuclease.* **Sharples G.J., Corbett L.M., McGlynn P.** 21, Oxford : s.n., 1999 , Nucleic Acids Research , Vol. 27.
14. *Toward a metabolic interpretation of genetic recombination of E. coli and its phages.* **Clark, A.J.** 1971 , Annu. Rev. Microbiol., Vol. 25, pp. 437-64.
15. *The X philes: structure-specific endonucleases that resolve Holiday Junctions.* **Sharples, Gary J.** 4, 2001 , Molecular Microbiology, Vol. 39, pp. 823-834.
16. *Biochemistry of Homologous Recombination in Escherichia coli.* **Kowalczykowski S.C., Dixon D.A., Eggleston A.K.,Scoit D. Lauder, Rehrauer W.M.** 3, 1994 September, Microbiological Reviews, Vol. 58, pp. 401-465.
17. *Rescue of arrested replication forks by homologous recombination.* **Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M, Bidnenko V.** 15, 2001 July, Proc Natl Acad Sci U S A., Vol. 98, pp. 8181-8188.
18. *The Junction Resolving Enzymes.* **Lilley D.M.J., White M.F.** 2001 June, Nature Reviews Molecular Cell Biology, Vol. 2, pp. 433-443.
19. *Recombinational Repair of DNA Damage in Escherichia coli and Bacteriophage Lambda.* **Kuzminov, A.** 4, Eugene : s.n., 1999 December, MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Vol. 63, pp. 751-813.
20. *The importance of repairing stalled replication forks.* **Cox M.M., Goodman M.F., Kreuzer K.N., Sherratt D.J., Sandlerk S.J. & Marians K.J.** 2000 2-March, Nature, Vol. 404, pp. 37-41.

21. *Motoring along with the bacterial RecA protein.* **Cox, M.M.** 2007 February , Nature Reviews: Molecular Cell Biology, Vol. 8, pp. 127-138.
22. *Bacterial endosymbionts in animals.* **Baumann, Nancy A Moran and Paul.** 2000 , Current Opinion in Mcrobiology, Vol. 3, pp. 270-275.
23. *Pleiotropic effect of the rec A gene of Escherichia coli: uncoupling of cell division from deoxyribonucleic acid replication.* **M., Inouye.** 2, May 1971, J Bacteriol., Vol. 106, pp. 539-42.
24. *Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli.* . **GC., Walker.** 1, Mar 1984, Microbiol Rev., Vol. 48, pp. 60-93.
25. *Deletions generated by the transposon Tn10 in the srl recA region of the Escherichia coli K-12 chromosome.* **Csonka LN, Clark AJ.** 2, Oct 1979 , Genetics., Vol. 93, pp. 321-43.
26. *In vitro selection of preferred DNA pairing sequences by the Escherichia coli RecA protein.* **Tracy RB, Kowalczykowski SC.** (5, Aug 1, 1996, Genes Dev., Vol. 10, pp. 1890-903.
27. *Recombination Activities of E. coli RecA Protein.* **Radding, C.M.:.** 1981 July, Cell, Vol. 25, pp. 3-4.
28. *Regulation of bacterial RecA protein function.* . **Cox MM.** 1, 2007 January-Februaury, Crit Rev Biochem Mol Biol., Vol. 42, pp. 41-63.
29. *C-terminal deletions of the Escherichia coli RecA protein. Characterization of in vivo and in vitro effects.* **Lusetti SL, Wood EA, Fleming CD, Modica MJ, Korth J, Abbott L, Dwyer DW, Roca AI, Inman RB, Cox MM.,.** 18, 2003 2-May, J Biol Chem. , Vol. 278, pp. 16372-80.
30. *The C terminus of the Escherichia coli RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein.* . **Eggler AL, Lusetti SL, Cox MM.,.** 18, 2003 2-May, J Biol Chem. , Vol. 278, pp. 16389-96.
31. *The UvrD helicase and its modulation by the mismatch repair protein MutL.* . **Matson SW, Robertson AB.,.** 15, 2006 , Nucleic Acids Res. , Vol. 34, pp. 4089-97.
32. **Cox, MM.** The Bacterial RecA Protein: Structure, Function, and Regulation. [ed.] A Aguilera and R Rothstein. *Molecular Genetics of Recombination.* s.l. : Springer Berlin / Heidelberg, 2007, Vol. 17, pp. 53-94.
33. *Replication restart in UV-irradiated Escherichia coli involving pols II, III, V, PriA, RecA and RecFOR proteins.* . **Rangarajan S, Woodgate R, Goodman MF.,.** 3, 2002 February, Mol Microbiol. , Vol. 43, pp. 617-28.
34. *The DinI protein stabilizes RecA protein filaments.* . **Lusetti SL, Voloshin ON, Inman RB, Camerini-Otero RD, Cox MM.** 29, 2004 16-July, J Biol Chem. , Vol. 279, pp. 30037-46.
35. *Escherichia coli RecX inhibits RecA recombinase and coprotease activities in vitro and in vivo.* **Stohl EA, Brockman JP, Burkle KL, Morimatsu K, Kowalczykowski SC, Seifert HS.,.** 4, 2003 24-January, J Biol Chem., Vol. 278, pp. 2278-85.
36. *A RecA filament capping mechanism for RecX protein.* . **Drees JC, Lusetti SL, Chitteni-Pattu S, Inman RB, Cox MM.,.** 5, 2004 10-September, Mol Cell. , Vol. 15, pp. 789-98.
37. *An SOS inhibitor that binds to free RecA protein: the PsiB protein.* **Petrova V, Chitteni-Pattu S, Drees JC, Inman RB, Cox MM.** 1, 2009 9-October, Mol Cell. , Vol. 36, pp. 121-30.
38. *A DNA-Unwinding Protein Isolated from Escherichia coli: Its Interaction with DNA and with DNA Polymerases.* **Sigal, N., et al.** 12, 1972 December, Proc. Nat. Acad. Sci. USA, Vol. 69, pp. 3537-3541.
39. *Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7.* **Hollis, T., et al.** 17, 2001 14-August, PNAS, Vol. 98, pp. 9557–9562.

40. *Escherichia coli single-strand binding protein organizes single-stranded DNA in nucleosome-like units.* **Chrysogelos, S. and J., Griffith.** 1982 October, Proc. NatL Acad. Sci. USA, Vol. 79, pp. 5803-5807.
41. *Structure of the DNA binding domain of E. coli SSB bound to ssDNA.* **Raghunathan, S., et al.** 8, 2000 August, Nature Structural Biology, Vol. 7, pp. 648-652.
42. *Homologous recombination-mediated double-strand break repair.* **Wyman, C., Ristic, D. and R., Kanaar.** 2004 24-April, DNA Repair, Vol. 3, pp. 827-833.
43. *Initiation of genetic recombination and recombination-dependent replication .* **Kowalczykowski, S.C.:.** 2000 April, TIBS, Vol. 25, pp. 156-165.
44. *RecBcD: The supercar of DNA Repair.* **Wigley, D.B.** 2007 16-November, Cell, Vol. 131, pp. 651-653.
45. *Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme.* **Arnold DA, Kowalczykowski SC.** 16, Apr 21, 2000 , J Biol Chem. , Vol. 275, pp. 12261-5.
46. *RecBCD enzyme overproduction impairs DNA repair and homologous recombination in Escherichia coli.* **Dermić D, Halupecki E, Zahradka D, Petranović M.,.** 3, 2005 April, Res. Microbiol. , Vol. 156, pp. 304-311.
47. *RecBCD enzyme is a bipolar DNA helicase.* **Dillingham MS, Spies M, Kowalczykowski SC.,.** 6942, 2003 19-June, Nature, Vol. 423, pp. 893-897.
48. *RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity.* **Taylor AF, Smith GR.,.** 6942, 2003 19-June, Nature, Vol. 423, pp. 889-893.
49. *Functions of the ATP hydrolysis subunits (RecB and RecD) in the nuclease reactions catalyzed by the RecBCD enzyme from Escherichia coli.* **Chen HW, Randle DE, Gabbidon M, Julin DA.** 1, 1998 25-April, J Mol Biol., Vol. 278, pp. 89-104.
50. *RecBCD enzyme switches lead motor subunits in response to chi recognition.* **Spies M, Amitani I, Baskin RJ, Kowalczykowski SC.,.** 4, 2007 16-November, Cell. 2007 Nov 16;131(4):694-705., Vol. 131, pp. 694-705.
51. *Recombination proteins and rescue of arrested replication forks. .* **Michel B, Boubakri H, Baharoglu Z, LeMasson M, Lestini R.** 7, Jul 1, 2007 , DNA Repair (Amst). , Vol. 6, pp. 967-80.
52. *Reconstitution of initial steps of dsDNA break repair by the RecF pathway of E. coli.* **N., Handa, et al.** 2009 , Genes & Development, Vol. 23, pp. 1234-1245.
53. *Is RecF a DNA replication protein?* **Kogoma T.** 8, 1997 15-April, Proc Natl Acad Sci USA, Vol. 94, pp. 3483-4.
54. *The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins. .* **Inoue J, Honda M, Ikawa S, Shibata T, Mikawa T.,.** 1, 2008 January, Nucleic Acids Res. , Vol. 36, pp. 94-109.
55. *Crystal structure and DNA-binding analysis of RecO from Deinococcus radiodurans.* **Leiros I, Timmins J, Hall DR, McSweeney S.,.** 5, 2005 9-March, EMBO J., Vol. 24, pp. 906-18.
56. *RecFOR and RecOR as distinct RecA loading pathways. .* **Sakai A, Cox MM.,.** 5, 2009 30-January, J Biol Chem., Vol. 284, pp. 3264-72.
57. *The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins. .* **Inoue J, Honda M, Ikawa S, Shibata T, Mikawa T.,.** 1, 2008 January, Nucleic Acids Res. , Vol. 36, pp. 94-109.
58. *The RecF protein antagonizes RecX function via direct interaction. .* **Lusetti SL, Hobbs MD, Stohl EA, Chitteni-Pattu S, Inman RB, Seifert HS, Cox MM.,.** 1, 2006 6-January, Mol Cell. , Vol. 21, pp. 41-50.
59. *Evidence for ATP binding and double-stranded DNA binding by Escherichia coli RecF protein. .* **Madiraju MV, Clark AJ.** 23, 1992 December, J Bacteriol. , Vol. 174, pp. 7705-10.

60. *Structural conservation of RecF and Rad50: implications for DNA recognition and RecF function.* . **Koroleva O, Makharashvili N, Courcelle CT, Courcelle J, Korolev S.,.** 3, 2007 7-February, EMBO J. , Vol. 26, pp. 867-77.
61. *RecR-mediated modulation of RecF dimer specificity for single- and double-stranded DNA.* . **Makharashvili N, Mi T, Koroleva O, Korolev S.,.** 3, 2009 16-January, J Biol Chem. , Vol. 284, pp. 1425-34.
62. *RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair.* **Morimatsu K, Kowalczykowski SC.** 5, May 2003, Mol Cell., Vol. 11, pp. 1337-47.
63. *Identification of the RecR Toprim domain as the binding site for both RecF and RecO. A role of RecR in RecFOR assembly at double-stranded DNA-single-stranded DNA junctions.* **Honda M, Inoue J, Yoshimasu M, Ito Y, Shibata T, Mikawa T.** 27, Jul 7, 2006, J Biol Chem. , Vol. 281, pp. 18549-59.
64. *Mechanism of RecO recruitment to DNA by single-stranded DNA binding protein.* **Ryzhikov M, Koroleva O, Postnov D, Tran A, Korolev S.** 2011 18-April, Nucleic Acids Res.
65. *Postreplication repair mechanisms in the presence of DNA adducts in Escherichia coli.* . **Bichara M, Meier M, Wagner J, Cordonnier A, Lambert IB.,.** 3, 2011 May-June, Mutat Res. , Vol. 727, pp. 104-22.
66. *Isolation and characterization of an Escherichia coli ruv mutant which forms nonseptate filaments after low doses of ultraviolet light irradiation.* **Otsuji N, Iyehara H, Hideshima Y.** 2, 1974 February, J Bacteriol., Vol. 117, pp. 337-44.
67. *Processing of recombination intermediates by the RuvABC proteins.* **West, S.C.** 1997 , Annu. Rev. Genet., Vol. 31, pp. 213–44.
68. *Role of RuvA in branch migration reactions catalyzed by the RuvA and RuvB proteins of Escherichia coli.* **Mitchell AH, West SC.** 32, Aug 9, 1996 , J Biol Chem., Vol. 271, pp. 19497-502.
69. *Purification and properties of the RuvA and RuvB proteins of Escherichia Coli.* **Tsaneva, I.R., et al.** 1992 11-May, Mol Gen Gene, Vol. 235, pp. 1-10.
70. *Escherichia coli RuvA and RuvB proteins specefically interact with Holliday junctions and promote branch migration.* **Iwasaki, H., et al.** 1992 27-August, Genes and Development, Vol. 6, pp. 2214-2220.
71. *Structure of a multisubunit complex that promotes DNA branch migration.* **Parsons, C.A., et al.** 1995 , Nature, Vol. 374, pp. 375–378.
72. *The kinetics of spontaneous DNA branch migration.* **Panyutin, I.G. and Hsieh, P.** 1994 , Proc. Natl. Acad. Sci. USA, Vol. 91, pp. 2021–2025.
73. *Interactions Between RuvA and RuvC at Holliday Junctions: Inhibition of Junction Cleavage and Formation of a RuvA-RuvC-DNA Complex.* **Whitby, M.C., et al.** 1996 , J. Mol. Biol., Vol. 264, pp. 878-890.
74. *Formation Of A Stable Ruva Double Tetramer Is Required For Efficient Branch Migration In Vitro And For Replication Fork Reversal In Vivo.* **Bradley, A.S., et al.** 2011 24-June, J. Biol. Chem., Vol. 25, pp. 22372-83.
75. *Crystal structure of the Holliday junction DNA in complex with a single RuvA tetramer.* **Ariyoshi, M., et al.** 15, 200 9-April, PNAS, Vol. 97, pp. 8257–8262.
76. *Distantly related sequences in the a- and f-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold.* **Walker, J.E., et al.** 8, 1982 5-July, The EMBO Joumal, Vol. 1, pp. 945-951.
77. *The Escherichia coli RuvB branch migration protein forms double hexameric rings around DNA.* **Stasiak, A., et al.** 1994 August, Proc. Natl. Acad. Sci. USA, Vol. 91, pp. 7618-7622.

78. *The Processing of Recombination Intermediates: Mechanistic Insights from Studies of Bacterial Proteins.* **West, S.C.** 1994 14-January, Cell,, Vol. 76, pp. 9-15.
79. *Branch Migration of Holliday Junctions Promoted by the Escherichia coli RuvA and RuvB Proteins.* **Muller, B., Tsaneva, I.R. and S.C., West.** 23, 1993 15-August, J. Biol. Chem., Vol. 268, pp. 17185-17189 .
80. *RuvB Protein-Mediated ATP Hydrolysis: Functional Asymmetry in the RuvB hexamer.* **Marrione, P.E. and M.M., Cox.** 1995 , Biochemistry, Vol. 34, pp. 9809-9818.
81. *Crystal structure of the Holliday junction migration motor protein RuvB from Thermus thermophilus HB8.* **Yamada, K., et al.** 4, 2001 , PNAS, Vol. 98, pp. 1442-1447.
82. *Dissociation of RecA filaments from duplex DNA by the RuvA and RuvB DNA repair proteins.* **Adams, D.E., Tsaneva, I.R. and West, S.C.** 21, 1994 11-October , PNAS , Vol. 91, pp. 9901-9905.
83. *In vitro reconstitution of the late steps of genetic recombination in E. coli.* **Eggleston, A.K., Mitchell, A.H. and West, S.C.** 1997 , Cell , Vol. 89, pp. 607–617.
84. *Crystal structure of the RuvA-RuvB complex: a structural basis for the Holliday junction migrating motor machinery.* **Yamada, K, et al.** 3, 2002 September, Mol. Cell, Vol. 10, pp. 671-681.
85. *The RuvC protein dimer resolves Holliday junctions by a dual incision mechanism that involves base-specific contacts.* **Shah, R., Cosstick, R. and West, S.C.** 6, 1997 , The EMBO Journal, Vol. 16, pp. 1464–1472.
86. *Formation and resolution of recombination intermediates by E. coliRecA and RuvC proteins.* **Dunderdale, H.J., et al.** 1991 26-December, Nature, Vol. 354, pp. 506-510.
87. *Holliday Junction Processing in Bacteria: Insights from the Evolutionary Conservation of RuvABC, RecG, and RusA.* **Sharples, GJ, Ingleston, SM and Lloyd, RG.** 18, 1999 September, Journal of Bacteriology, Vol. 181, pp. 5543-5550.
88. *Resolution of Holliday Junctions by RuvC Reslvase: Cleavage Specificity and DNA Distortion.* **Bennett, R.J., Dunderdaie, H.J. and West, S.C.** 1993 24-Sepetember , Cell, Vol. 74, pp. 1021-1031.
89. *Cloning, Overexpression, Purification, and Characterization of the Escherichia coli RuvC Holliday Junction Resolvase.* **Dunderdale, H.J., et al.** 7, 1994 18-February, THE JOURNAL OF BIOLOGICAL CHEMISTRY, Vol. 269, pp. 5187-5194.
90. *The X philes: structure-specific endonucleases that resolve Holliday junctions.* **Sharples, G.J.:.** 4, Nottingham : s.n., 2001 , Molecular Microbiology, Vol. 39, pp. 823-834.
91. *Analysis of conserved basic residues associated with DNA binding (Arg69) and catalysis (Lys76) by the RusA Holliday junction resolvase. .* **Bolt, E.L., Sharples, G.J. and Lloyd, R.G.** 2000 , J Mol Biol, Vol. 304, pp. 165-176.
92. *Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase.* **Saito, A, et al.** 1995. , Proc. Natl. Acad. Sci. USA, Vol. 92, pp. 7470–7474.
93. *Structure and mechanism of the RuvB Holliday junction branch migration motor.* **Putnam CD, Clancy SB, Tsuruta H, Gonzalez S, Wetmur JG, Tainer JA.** 2, 2001 10-August, J Mol Biol., Vol. 311, pp. 297-310.
94. *Formation of a Stable RuvA Protein Double Tetramer Is Required for Efficient Branch Migration in Vitro and for Replication Fork Reversal in Vivo.* **Bradley, AS, et al.** 25, 2011 25-June, J Biol Chem., Vol. 286, pp. 22372-22383.
95. *Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of holliday junction resolution.* **van Gool AJ, Hajibagheri NM, Stasiak A, West SC.** 14, 1999 15-July, Genes Dev., Vol. 13, pp. 1861-70.

96. *What makes the bacteriophage V Red system useful for genetic engineering: molecular mechanism and biological function.* **Poteete, A.R.** 2001 , FEMS Microbiology Letters , Vol. 201, pp. 9-14.
97. *Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli.* **KC., Murphy.** 8, 1998, J Bacteriol., Vol. 180, pp. 2063-71.
98. *Lambda Gam protein inhibits the helicase and Chi-stimulated recombination acrivities of Escherichia coli RecBCD enzyme.* **Murphy, K. C.** 1991 , Journal of Bacteriology , Vol. 173, pp. 5808-5821.
99. *The Crystal Structure of λ-Gam Protein Suggests a Model for RecBCD Inhibition.* **Court, R., et al.** 2007 , Journal of Molecular Biology , Vol. 371, pp. 25-33.
100. *The gamma protein specified by bacteriophage lambda. Structure and inhibitory activity for the recBC enzyme of Escherichia coli.* **Karu, A. E., et al.** 1975 , Journal of Biological Chemistry , Vol. 250, pp. 7377-7387.
101. *The Beta Protein of Phage Lambda Promotes Strand Exchange.* **Li, Z., et al.** 1998 , J. Mol. Biol., Vol. 276, pp. 733-744.
102. *An Exonuclease Induced by Bacteriophage Lambda.* **Little, J.W.** 4, 1967 25-February, The Journal of Biological Chemistry, Vol. 242, pp. 679-686.
103. *Toroidal Structure of Lambda-Exonuclease.* **Kovall, R. and Matthews, B.W.** 1997 19-September, Science, Vol. 277, pp. 1824-1826.
104. *Classification and evolutionary history of the single-stranded annealing proteins, RecT, RedBeta, ERF and RAD52.* **Iyer, L. M., Koonin, E. V. and L., Aravind.** 8, 2002 21-March, BMC Genomica, Vol. 3.
105. *Rings and filaments of b protein from bacteriophage lambda suggest asuperfamily of recombination proteins.* **Passy, S.I., et al.** 1999 April, Proc. Natl. Acad. Sci. USA Biochemistry, Vol. 96, pp. 4279-4284.
106. *RecE/RecT and Red alpha/Red beta initiate double-stranded break repair by specifically interacting with their respective partners.* **Muyrers, J.P.P., et al.** 2000 2-June, Genes & Development , Vol. 14, pp. 1971-1982.
107. *Efficient RecABC-dependent, homologous recombination between coliphage lambda and plasmids requires a phage ninR region gene.* **Hollifield, W.C., Kaplan, E.N. and Huang, H.V.** 1987 , Molecular Genetics and Genomics, Vol. 210, pp. 248-255.
108. *Phage Lambda Has an Analog of Escherichia coli rec0, recR and recF Genes.* **Sawitzke, J.A. and Stahl, F.W.** 1992 January, Genetics, Vol. 130, pp. 7-16.
109. *Modulation of DNA repair and recombination by the bacteriophage lambda orf function in Escherichia coli K12 coli.* **Potetee, A.R.** 9, 2004 May, Journal of Bacteriology, Vol. 186, pp. 2699-2707.
110. *Modulation of DNA repair and recombination by the bacteriophage lambda Orf function in Escherichia coli K-12.* **AR., Poteete.** 9, 2004 May, J Bacteriol. , Vol. 186, pp. 2699-707.
111. *Functional similarities between phage lambda Orf and Escherichia coli RecFOR in initiation of genetic exchange.* **Maxwell KL, Reed P, Zhang RG, Beasley S, Walmsley AR, Curtis FA, Joachimiak A, Edwards AM, Sharples GJ.** 32, 2005 9-August , Proc Natl Acad Sci U S A. , Vol. 102, pp. 11260-5.
112. *Diverse members of the phage Orf recombination mediator family share similar DNA and SSB binding properties.* **Curtis, F.A, et al.** Unpublished.
113. **Sharples, G.**, *personal communication*.
114. *Strand invasion promoted by recombination protein beta of coliphage lambda. .* **Rybalchenko N, Golub EI, Bi B, Radding CM.** 49, 2004 7-December, Proc Natl Acad Sci U S A. , Vol. 101, pp. 17056-60.

115. **Bowers, L.Y.** *Orf protein modulates phage and bacterial pathways of genetic recombination.* s.l. : University of Durham, 2008.
116. *The protein interaction map of bacteriophage lambda.* . **Rajagopala SV, Casjens S, Uetz P.** 213 , Sep 26, 2011 , BMC Microbiol. , Vol. 11.
117. **Reed, P.** *Function of bacteriophage Orf recombinases in genetic exchange.* s.l. : University of Durham, 2006.
118. *BRCA2 BRC motifs bind RAD51-DNA filaments.* **Galkin VE, Esashi F, Yu X, Yang S, West SC, Egelman EH.** 24, Jun 14, 2005, Proc Natl Acad Sci U S A., Vol. 102, pp. 8537-42.
119. *The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA.* . **Thorslund T, McIlwraith MJ, Compton SA, Lekomtsev S, Petronczki M, Griffith JD, West SC.** 10, Oct 2010 , Nat Struct Mol Biol. , Vol. 17, pp. 1263-5.
120. *The C-terminus of the phage λ Orf recombinase is involved in DNA binding.* . **Curtis FA, Reed P, Wilson LA, Bowers LY, Yeo RP, Sanderson JM, Walmsley AR, Sharples GJ.** 2, Mar-Apr 2011 , J Mol Recognit. , Vol. 24, pp. 333-40.
121. *Prophage genomics.* **Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H.** 2, 2003 Jun, Microbiol Mol Biol Rev. , Vol. 67, pp. 238-76. Erratum in: Microbiol Mol Biol Rev. 2003 Sep;67(3):473.
122. *Improved isolation of proteins tagged with glutathione S-transferase.* **Vinckier NK, Chworos A, Parsons SM.** 2, Feb 2011 , Protein Expr Purif. , Vol. 75, pp. 161-4.
123. *Phage conversion of exfoliative toxin A production in Staphylococcus aureus.* . **Yamaguchi T, Hayashi T, Takami H, Nakasone K, Ohnishi M, Nakayama K, Yamada S, Komatsuzawa H, Sugai M.** 4, 2000 Nov, Mol Microbiol. , Vol. 38, pp. 694-705.
124. *HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily.* . **Mehta P, Katta K, Krishnaswamy S.** 1, 2004 Jan, Protein Sci., Vol. 13, pp. 295-300.
125. *ClustalW and ClustalX version 2.* **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG.** 21, 2007 , Bioinformatics 2007 , Vol. 23, pp. 2947-2948.
126. BLAST Substitution Matrices. *NCBI.* [Online] [Cited: 2011 10-October.] http://www.ncbi.nlm.nih.gov/blast/html/sub_matrix.html.
127. *The winged-helix DNA-binding motif: another helix-turn-helix takeoff.* . **Brennan, RG.** 5, 1993 Sep, Cell, Vol. 74, pp. 773-6.
128. *Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA.* . **McKay DB, Steitz TA.** 5809, 1981 30-Apr, Nature, Vol. 290, pp. 744-9.
129. *Structure of the cro repressor from bacteriophage lambda and its interaction with DNA.* . **Anderson WF, Ohlendorf DH, Takeda Y, Matthews BW.** 5809, 1981 3o-Apr, Nature. , Vol. 290, pp. 754-8.
130. *The helix-turn-helix DNA binding motif.* . **Brennan RG, Matthews BW.** 4, 1989 5-Feb, J Biol Chem. , Vol. 264, pp. 1903-6.
131. *Dotlet: diagonal plots in a web browser.* . **Junier, Thomas and Pagni, Marco.** 2, 2000 Feb, Bioinformatics. , Vol. 16, pp. 178-9.
132. *Kalign--an accurate and fast multiple sequence alignment algorithm.* . **Lassmann T, Sonnhammer EL.** 6 , 2005 Dec , BMC Bioinformatics. , Vol. 12, p. 298.
133. *Network Protein Sequence Analysis* . **Combet C., Blanchet C., Geourjon C. and Deléage G.** 3, 2000 March, TIBS, Vol. 25, pp. 147-150.
134. *Improved detection of helix-turn-helix DNA-binding motifs in protein sequences.* . **Dodd IB, Egan JB.** 17, 1990 11-Sep, Nucleic Acids Res. , Vol. 18, pp. 5019-26.

135. *Protein structure prediction on the web: a case study using the Phyre server.* **Kelley, LA and Sternberg, MJE.** 2009 , Nature Protocols , Vol. 4, pp. 363-371.
136. Protein: Transcriptional regulator Cgl1640/Cg1846 from Corynebacterium glutamicum. *SCOPE.* [Online] [Cited: 2011 09-November.] http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.b.j.b.bc.dd.html.
137. *Nitrogen regulatory protein C-controlled genes of Escherichia coli: scavenging as a defense against nitrogen limitation.* **Zimmer DP, Soupene E, Lee HL, Wendisch VF, Khodursky AB, Peter BJ, Bender RA, Kustu S.** 26, 2000 19-Dec, Proc Natl Acad Sci U S A., Vol. 97, pp. 14674-9.
138. *A time-efficient, linear-space local similarity algorithm.* **Xiaoqiu Huang, Webb Miller,.** 3, 1991 September , Advances in Applied Mathematics, Vol. 12, pp. 337-357.
139. *Three-dimensional solution structure of the N-terminal receiver domain of NTRC.* **Volkman BF, Nohaile MJ, Amy NK, Kustu S, Wemmer DE.** 4, 1995 31-Jan, Biochemistry, Vol. 34, pp. 1413-24.
140. *High-resolution solution structure of the beryllofluoride-activated NtrC receiver domain.* . **Hastings CA, Lee SY, Cho HS, Yan D, Kustu S, Wemmer DE.** 30, 2003 5-Aug, Biochemistry. , Vol. 42, pp. 9081-90.
141. *Histone-Like Proteins of Bacteria (Review).* **Anuchin, A. M., et al.** 6, 2011 , Applied Biochemistry and Microbiology, Vol. 47, pp. 580-585.
142. *Fis modulates growth phase-dependent topological transitions of DNA in Escherichia coli.* . **Schneider, R., Travers, A. and Muskhelishvili, G.** 1997 , Mol. Microbiol. , Vol. 26, pp. 519–530.
143. *The Fis protein: it's not just for DNA inversion anymore.* . **Finkel, S. E. and Johnson, R. C.** Mol. Microbiol. , Vol. 6, pp. 3257–3265.
144. *Global effects of Fis on Escherichia coli gene expression during different stages of growth.* . **Bradley, M. D., et al.** 2007 , Microbiology, Vol. 153 , pp. 2922–2940.
145. *FIS-dependent trans-activation of tRNA and rRNA operons of Escherichia coli.* . **Bosch, L., et al.** 1990 , Biochim. Biophys. Acta., Vol. 1050, pp. 293–301.
146. *Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli.* **Sheikh, J., Hicks, S., Dall'Agnol, M., Phillips, A. D. & Nataro, J. P.** 2001 , Mol. Microbiol., Vol. 41, pp. 983–997.
147. *Role of the nucleoid-associated protein Fis in the regulation of virulence properties of enteropathogenic Escherichia coli.* **Goldberg, M. D., Johnson, M., Hinton, J. C. & Williams, P. H.** 2001 , Mol. Microbiol., Vol. 41, pp. 549–559.
148. *The N-terminal part of the E.coli DNA binding protein FIS is essential for stimulating site-specific DNA inversion but is not required for specific DNA binding.* **Koch C, Ninnemann O, Fuss H, Kahmann R.** 21, 1991 11-Nov, Nucleic Acids Res. , Vol. 19, pp. 5915-22.
149. *The shape of the DNA minor groove directs binding by the DNA-bending protein Fis.* **Stella S, Cascio D, Johnson RC.** 8, 2010 15-Apr, Genes Dev. , Vol. 24, pp. 814-26.
150. *Efficient excision of phage lambda from the Escherichia coli chromosome requires the Fis protein.* **Ball CA, Johnson RC.** 13, 1991 Jul, J Bacteriol., Vol. 173, pp. 4027-31.
151. *Three-dimensional structure of the E. coli DNA-binding protein FIS.* . **Kostrewa D, Granzin J, Koch C, Choe HW, Raghunathan S, Wolf W, Labahn J, Kahmann R, Saenger W.** 349 (6305), 1991 Jan, Nature. , Vol. 10, pp. 178-80.
152. The Crystal Structure of E. coli Fis. *European Bioinformatics Institute* . [Online] [Cited: 2011 14-November.] http://www.ebi.ac.uk/pdbe-srv/view/entry/1etq/quaternary.html.
153. *Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli.* . **Ball CA, Osuna R, Ferguson KC, Johnson RC.** 24, 1992 Dec, J Bacteriol. , Vol. 174, pp. 8043-56.

154. *Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis.* **Skoko D, Yoo D, Bai H, Schnurr B, Yan J, McLeod SM, Marko JF, Johnson RC.** 4 , 2006 8-Dec , J Mol Biol. , Vol. 364, pp. 777-98 .
155. *Bacterial chromatin.* **Travers, A. and Muskhelishvili, G.** 5, 2005 October , Current Opinion in Genetics & Development, Vol. 15 , pp. 507-514.
156. *Common and variable contributions of Fis residues to high-affinity binding at different DNA sequences.* **Feldman-Cohen LS, Shao Y, Meinhold D, Miller C, Colón W, Osuna R.** 6, 2006 Mar, J Bacteriol. , Vol. 188, pp. 2081-95 .
157. *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* **Notredame C, Higgins DG, Heringa J.** 1, 2000 8-Sep , J Mol Biol. , Vol. 302, pp. 205-17.
158. *Crystal structure of the homo-tetrameric DNA binding domain of Escherichia coli single-stranded DNA-binding protein determined by multiwavelength x-ray diffraction on the selenomethionyl protein at 2.9-A resolution.* **Raghunathan S, Ricard CS, Lohman TM, Waksman G.** 13, 1997 24-June, Proc Natl Acad Sci U S A., Vol. 94, pp. 6652-7.
159. *Integration requires a specific interaction of the donor DNA terminal 5'-cytosine with glutamine 148 of the HIV-1 integrase flexible loop.* **Johnson AA, Santos W, Pais GC, Marchand C, Amin R, Burke TR Jr, Verdine G, Pommier Y.** 1, 2006 6-January, J Biol Chem., Vol. 281, pp. 461-467.