DNA binding proteins of the phage λ ninR region

TROTTER, ALEXANDER,JOHN

DNA binding proteins of the phage λ *ninR* region

Alexander John Trotter B.Sc (Hons)

Van Mildert College, Durham University

Masters of Science (By Thesis) in Biological Sciences

School of Biological and Biomedical Sciences

Durham University

2015

Acknowledgements

Firstly I would like to thank my supervisor Gary for his guidance, supervision and friendship, and for giving me the opportunity to work on a project with novel research sections and so much potential for publications. I thank Phil, Fiona and Laura for the proteins, protocols and prior research they contributed to the project making my research possible. To all the wonderful members of the office and lab group of 229/231 for the friendly and cooperative work environment and for reminding me that no matter how frayed I was feeling, by comparison I was relatively sane. My parents who supported me through the time working away at the lab and for putting up with me while writing my thesis at home. And finally to my wonderful partner Libby for the countless hours spent keeping me motivated, consoled and on track with my study, without her I couldn't have gotten anywhere.

Abstract

Bacteriophages are known to disable host functions in order to impose their own enzymes to hijack the bacterium for maximal replication of the virus. One well-characterised example of this is the Red genetic recombination system of phage λ which blocks host RecBCD and replaces it with an exonuclease and synaptase to promote genomic rearrangements. Other substitute enzymes from λ include Orf, which can replace the RecA-loading activity of RecFOR, and Rap, which can substitute for the RuvC Holliday junction resolvase. This study has investigated the DNA binding activities of three gene products from the *ninR* region of phage λ that are linked to genetic recombination. These are Orf (NinB), Rap (NinG) and NinH, the latter an uncharacterised protein of 68 residues. DNA binding analysis with selected substitution mutants of the Orf protein established that the central channel and adjacent regions are critical for binding to single-stranded DNA. The predicted HNH catalytic site of the Rap DNA structure-specific endonuclease was also investigated by site-directed mutagenesis confirming that the HNH domain is essential for Holliday junction cleavage and may also participate in branched DNA binding. Finally, the NinH protein which resembles the bacterial nucleoid associated protein (NAP) Fis, was also analysed for its ability to bind double-stranded DNA. The *in vitro* data showed that NinH binds DNA and preferentially favours bent duplex substrates, consistent with its similarity to Fis. NinH could not compensate for the loss of various NAPs from *E. coli* in deletion strains, including a *fis* mutant, and generally conferred a negative effect on the growth of these strains. Thus NinH may play a role in perturbing normal bacterial replication to promote phage production. The results offer several new insights into the properties of these three DNA binding proteins and their possible roles in phage DNA replication and recombination.

Table of Contents

Tables and Figures

Chapter 1. Introduction

1.1 Bacteriophage

Phages are a group of viruses that specifically target and infect bacterial hosts; the name bacteriophage taken from Greek meaning 'bacteria eater'. Phage were first documented by Ernest Hanbury Hankin in 1896 while testing the biological properties of water said to have healing powers [1], with subsequent studies by Twort in 1915 [2] and d'Herelle in 1917 [3] confirming that bacteriophages were the agents responsible for counteracting the bacterial pathogens being treated. Since that time many diverse groups of these bacterial viruses have been studied and categorised, often in considerable detail.

1.2 Phage Genetics

Bacteriophage λ belongs to the *Siphoviridae* family, a large group of double-stranded DNA tailed phages that have been studied extensively over the last 60 years [4-8]. λ-like phages exhibit significant genetic diversity through mosaicism of their genome whereby gene segments are shuffled between phages and prophages via short regions of sequence homology at either legitimate or illegitimate sites [4, 9-12]. The capacity to acquire genes by horizontal transfer is important as it allows the phage to obtain novel functions that may confer an evolutionary advantage [13].

Diversity is also generated from mutations, with the development of virulence factors and resistance factors increasing pathogenicity to improve the chance of survival along with regulatory adaptive responses for genome flexibility. While point mutations are traditionally associated with slow micro-evolution, allowing for a genetic advantage to eventually proliferate over a fellow organism without it, small mutations in a virus can result in the rapid development of structural genes to increase pathogenicity without the acquisition of additional genes. The potential for random mutagenesis offers a strong advantage for an organism under selective pressure [14].

While recombination between sites of limited homology generates the extensive mosaicism seen in λ [4, 15], recombination generally functions to maintain genome integrity by restoring DNA breaks that might arise from a variety of sources [16, 17].

1.3 Phage λ Replication

Temperate phages, such as λ, can follow two alternative replication strategies upon entering a host. In the lysogenic phase, the phage remains dormant integrated within the bacterial genome, whereas in the lytic phase it replicates and lyses the host to release new virion copies (Figure 1.1) λ integrates at a single site on the *Escherichia coli* genome to generate a lysogen, with the embedded prophage replicating with the host chromosome [18, 19]. DNA damage or other changes in host physiology result in activation and the phage passes from the lysogenic to the lytic phase, where the host machinery is harnessed to synthesise viral DNA, RNA and proteins and ultimately release mature phage particles through a programmed lysis event (Figure 1.1).

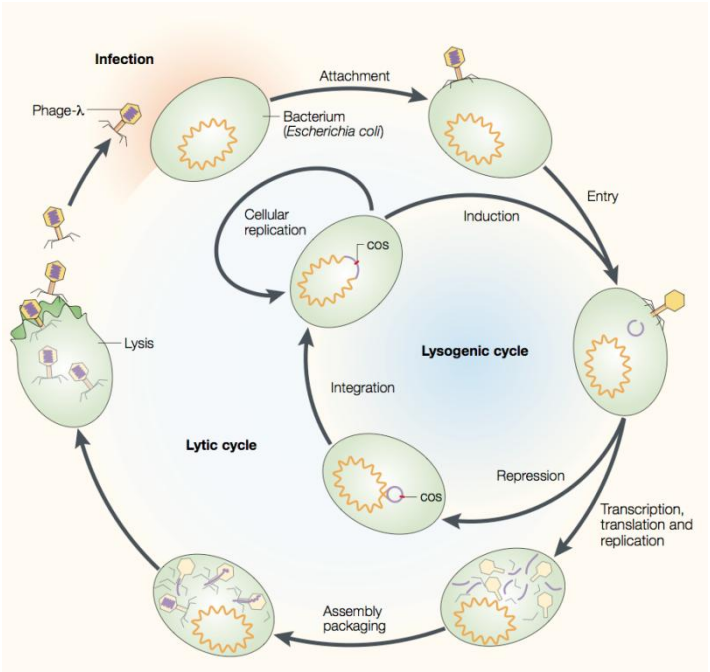

Figure 1.1. Replication cycle of a phage inside a bacterium host. The linear phage DNA can either close to form a circle at the cohesive site (cos) and integrate into the host DNA to be activated at a later time or enter rolling circle replication; immediately assembling new phage particles using the host machinery to build and package new phage. [20]

In the prophage state bacteriophages can confer useful properties upon the lysogen by expressing genes for toxins, enzymes or antibiotic resistance mechanisms [21-24]. Such events have been well documented as many virulence traits have been horizontally acquired from integrated phages. For example, the kidney-damaging Shiga toxin from *Shigella* species was picked up by a lambdoid phage and transferred to *E. coli* resulting in the evolution of the important pathogenic strain O157:H7 [25]. Subsequently other *E. coli* strains have been identified and catalogued into a Shiga toxin *E. coli* group, STEC. The potent toxins responsible for cholera and diphtheria were also delivered to their bacterial hosts by bacteriophages [26, 27]. Other non-toxin protein factors have been donated to *E. coli* by phage λ such as the *lom* and *bor* genes which code for an outer membrane protein which promotes adhesion to human host cells and the promotion of resistance to serum antibiotic activity, respectively [28, 29]. As phage λ is the best characterised phage, and its host is the model bacterium *E. coli*, λ gene functions have been dissected in detail, particularly with respect to its mechanisms of genetic recombination.

1.4 Homologous Recombination

Genetic or homologous recombination, along with replication and repair, are the processes which maintain a genome composed of DNA [16, 17]. Genetic recombination was defined by Clark [30] as being any set of pathways in which elements of nucleic acid interact with a resultant change of linkage of genes or of parts of genes. Unlike site-recombination, as in λ integration and excision, homologous recombination occurs between

relatively large sections of DNA (>100 bp) and sequence identity is used to ensure that the correct partners are involved in the exchange[31]. Recombination occurs when ssDNA is exchanged between DNA duplexes [32-34] and facilitates rearrangements that stimulate diversity or to repair damaged DNA, often resulting from problems in DNA replication at stalled or collapsed replication fork structures [16, 35].

The process of recombinational DNA repair, typically at a double-strand break. begins with either two duplexes exchanging single strands of DNA [36] or from a single unattached strand being spliced into a section of duplex DNA [34]. The strand that enters a duplex in either of these pathways can serve as a template for DNA synthesis to help replace the broken section on the damaged molecule or to restart replication. The four-stranded branched structure that arises from reciprocal strand exchange is known as a Holliday junction, named after Robin Holliday who originally proposed a model for genetic recombination in 1964 [37] (Figure 1.2). Once formed, the Holliday structure can be resolved by a group of endonucleases known as Holliday junction resolvases, which cleave opposing strands in the junction to restore separate duplexes. There are multiple pathways for formation and resolution of Holliday junctions, often dictated by the nature of the substrate or the enzymes that arrive first at the site. Phage λ makes use of two recombination systems, those of its host, *E. coli*, including a variety of Rec and Ruv proteins, and those encoded by its own genome, the Red system [38-40].

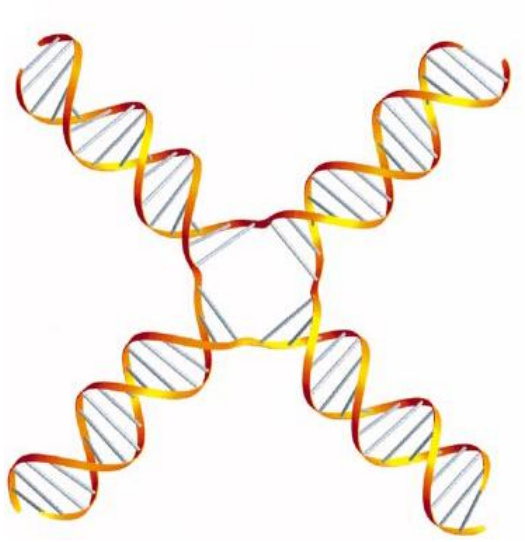

Figure 1.2. Holliday junction in 'stacked X' structure. Representational diagram of a Holliday junction held in a square planar shape as it would be found when bound by Holliday junction resolvases. The graphic shows how the exchange of single strands of DNA between two duplexes creates two new duplexes containing one strand from each, yielding heteroduplex DNA. [41]

1.5 Genetic Recombination in *E. coli*

1.5.1 RecA

In *E.coli*, homologous recombination is initiated by the action of a 38 kDa protein called RecA [34, 42-44], that is highly conserved in most cellular organisms [45]. The protein is a ssDNA-dependent ATPase [46] and assembles as a multimeric helical filament onto the DNA strand(s) that need to be paired for recombination [44, 47-52] (Figure 1.3). RecA targets single stranded DNA sections, often favouring sequences which include the

crossover hotspot instigator, Chi/χ, which contains the target sequence of 5'-GCTGGTGG-3' responsible for altering the exonuclease activity of another recombinase RecBCD [48]. RecA monomers assemble to form a nucleoprotein filament which has been demonstrated by electron microscopy to produce a right-handed helical filament so that the DNA can be used to search for a homologous duplex [53-57]. Nucleation of a DNA strand by RecA requires 4-5 protein monomers to bind initially in a reaction that is potentiated by ATP binding, which will later be needed for protein disassociation from the DNA [58]. Once the filament has nucleated on ssDNA, RecA will continue to polymerize, extending the filaments along adjacent DNA stretches [59]. Experiments using fluorescently tagged RecA showed that the filament extends along any adjacent dsDNA by 2-7 monomers [60].

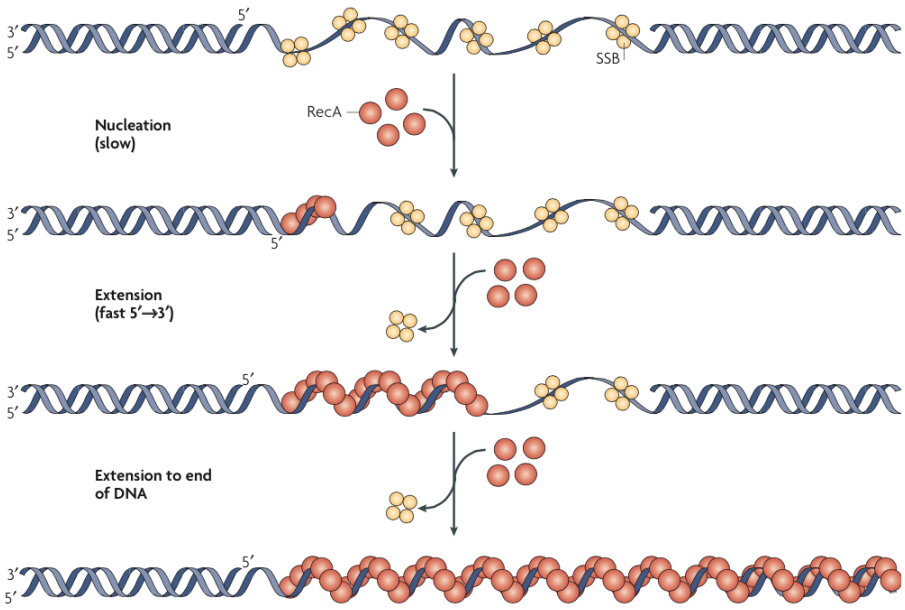

Figure 1.3. RecA assembly onto SSB bound DNA in a gapped duplex substrate. RecA monomers displace SSB in the 5' to 3' direction to form a nucleoprotein filament that extends onto the adjacent duplex in preparation for pairing with a homologous duplex to initiate recombinational exchange. [61]

Once the filament assembles a search for a homologous duplex occurs with hydrolysis of bound ATP activating the exchange of DNA strands between the homologous partners in the 5'-3' direction resulting in a 3-stranded D-loop structure that can be extended to form a Holliday junction [47, 62-64]. ATP hydrolysis also aids dissociation of RecA from the DNA following completion of strand exchange [65]. It is important that RecA can be recycled from the nucleoprotein filament to ensure recombination at other damaged sites and to prevent excessive levels of recombination, which can lead to genetic instability [59]. A number of helicases function as part of the SOS response, which is activated by the removal of the LexA repressor[66-68], which counteract overactive RecA action to remove toxic recombinase filaments [69]. One such protein is UvrD (DNA helicase II), a dimeric helicase of the helicase superfamily I [70]. UvrD is known to preferentially bind to 3' single-strand overhangs [71] and translocate in the 3'-5' direction but has also shown capability to begin strand unwinding from a nicked DNA substrate or a blunt double-strand end [72]. UvrD is capable of directly removing RecA

complexes for fork clearance in *DnaE* or *DnaN* mutants to facilitate replication restarts [73, 74]. Once the blocked fork is cleared UvrD helicase action can unwind the DNA strands promoting fork reversal [74].

1.5.2 SSB

Single-stranded DNA binding protein, SSB, is a 19 kDa tetrameric protein that protects exposed strands during DNA replication, recombination and repair [75, 76]. SSB shields ssDNA from DNA damaging agents, such as reactive oxygen species, and also protects it from exonucleases and endonucleases, which could destroy the DNA before it can be restored or lead to double-stranded breaks. SSB also limits the formation of intrastrand secondary structures, thus avoiding potential deletion and expansion of repeat sequences. Further to this structural aid, SSB also prevents interstrand re-annealing during DNA replication and similar reactions that might block recombination. Finally, SSB, via a negatively charged C-terminus, serves as a marker for a subset of replication, recombination and repair proteins which recognise and assemble at their required sites via protein-protein interactions [77-80].

While SSB promotes RecA strand exchange by eliminating secondary structures, it also tends to inhibit RecA nucleation by preventing the protein from accessing its ssDNA substrate. Hence, in order for RecA to assemble on ssDNA bound by SSB, it requires the action of other proteins to facilitate filament formation. The two sets of proteins that work to load RecA onto SSB-bound DNA are RecBCD and RecFOR [30]. Each set of proteins specialises in different recombination situations, RecBCD loads RecA at DNA ends, typically dsDNA breaks [81], while RecFOR help RecA nucleation at ssDNA gaps arising from problems during DNA replication [82].

1.5.3 RecBCD

RecBCD comprises a 330 kDa complex of three constituent proteins encoded by the *recB*, *recC* and *recD* genes [83-87]. The complex acts as a bipolar helicase and exonuclease as it degrades DNA from a double-strand break. The exonuclease activity is modified and attenuated upon encountering an 8 bp χ site (5'-GCTGGTGG-3') resulting in increased cleavage of the 5' strand but preservation of the 3' single strand [17, 88].

The individual subunits of the complex have distinct functions. RecB is a 3' to 5' helicase and exonuclease, unwinding the DNA at an end and nicking the unwound single strands [84]. RecC recognises the χ site and prevents further degradation of the 3' strand for loading of RecA and initiation of recombination. χ sites are over-represented in the *E. coli* chromosome and preferentially orientated away from the replication origin [89]. Thus RecBCD action serves to underpin DNA replication at breaks by facilitating the initiation of recombination to re-establish replication forks close to the occurrence of a break [85]. RecD is also a DNA helicase but in the 5' to 3' direction, making it the inverse counterpart to RecB [90]. However, unlike RecB, RecD requires single stranded DNA for activity whereas RecB can act at double stranded ends. RecBCD uses the ATP-driven helicase activities of both the RecB and RecD subunits as motors to allow the protein complex to translocate along the DNA from a double stranded end [87, 91, 92]. This feature ensures that the RecBCD complex will unwind dsDNA breaks into single stranded DNA, before moving off once the DNA has been secured for RecA assembly onto the recognised χ site [93]. This begins the repair of the DNA strands by

recombining them with other homologous strands resulting in the formation of a Holliday junction intermediate[94] to later be resolved by the RuvABC complex [95] (Figure 1.4).

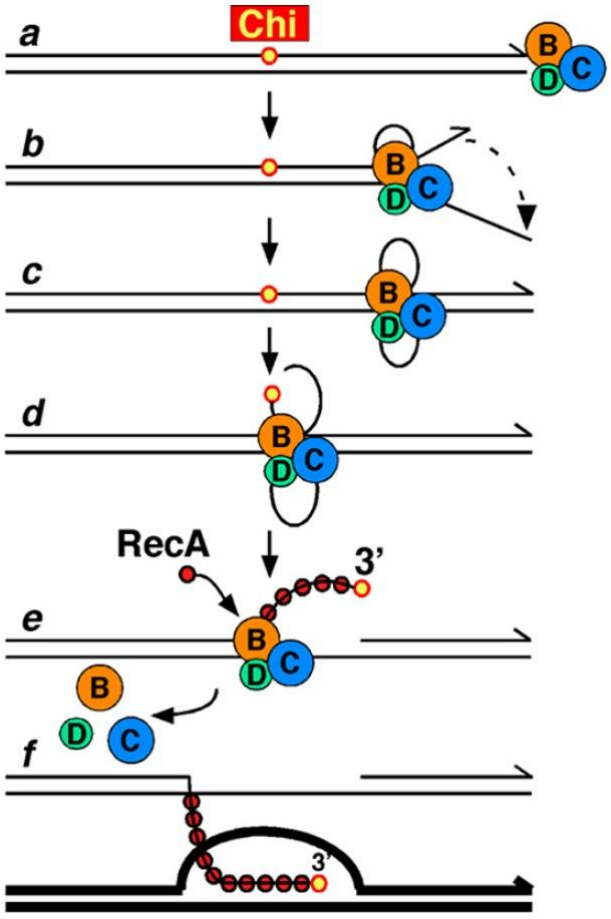

Figure 1.4. Action of RecBCD at one end of a double-stranded DNA break. The RecBCD complex recognises the DNA end (*a*) and begins unwinding and degrading the DNA while SSB binds onto the resulting single strand (*b*). Once at the χ site degradation switches from the top strand to the bottom strand (*c*) and RecBCD loads RecA onto the 3'-ended strand (*d*). RecBCD dissociates from the DNA (*e*) and the newly assembled RecA-ssDNA filament invades a homologous undamaged partner to form a 3-stranded D loop (*f*) [96].

This facilitated loading of the RecA protein onto DNA compensates for the inhibitory effect of ssDNA bound by SSB protein, allowing the RecBCD complex to control the initiation of recombination by the action of RecA[86, 97-101]. While the RecBCD complex can only repair DNA damage by double stranded breaks, the sister RecFOR pathway is less constrained. RecFOR specialises in the repair of daughter strand gaps, which can occur when replication forks stall or skip past lesions in the template [102, 103] while also being able to substitute for the action of the RecBCD pathway if needed. This latter feature is known because strains which have lost RecBCD activity due to mutations in the genes for RecB, RecC or RecD are still able to repair dsDNA breaks when the RecFOR pathway is active [104, 105].

1.5.4 RecFOR

The RecFOR complex is made of three proteins, often composed of RecFR, RecOR and RecFOR functional units, which help to load RecA monomers onto a ssDNA substrate that has been bound by SSB. The RecF subunit of RecFOR is a 40 kDa protein [106] which acts to load RecA proteins onto the single strand gaps found in damaged duplex DNA[107]. RecF binds double stranded DNA when partnered with ATP, the hydrolysis of this ATP catalyses the release of the RecF from the DNA once RecA has been loaded [108].

RecO is a 26 kDa protein[109] made up of three domains for protein-protein and protein-DNA interactions[110, 111]. The protein binds ssDNA and promotes annealing of homologous strands depending on which protein partner RecO binds [112]. When bound to SSB, RecO promotes further single strand annealing, however if RecO is bound within a RecFOR or RecFR complex the opposite effect is observed [113].

RecR is a 22 kDa protein which assembles into tetramers with a central channel wide enough to encircle dsDNA [114], however the protein is unable to bind DNA when it is not co-bound to one or other of the RecF and RecO proteins [115, 116]. When bound to RecO in a RecOR complex the complex prevents DNA from re-annealing and instead stimulates the loading of RecA onto DNA by RecF [62, 117-120]. RecR therefore regulates the amount of RecA activity through its displacement of SSB from RecO-SSB complex to change the priority of the RecFOR proteins in annealing DNA or for RecA loading [62, 112].

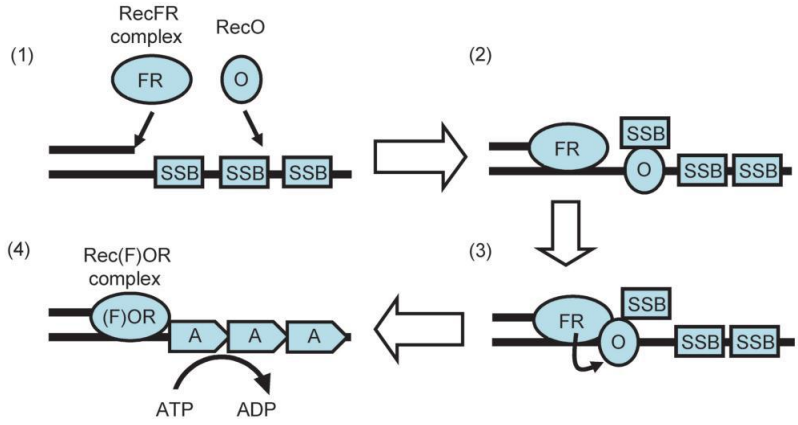

Figure 1.5. Summary of RecFOR activities. (1) RecO recognises SSB bound to ssDNA while a RecFR complex binds at the junction between single and double stranded DNA. (2) RecO binds to SSB, via the SSB C-terminus, and loads onto the ssDNA in its place. (3). RecR binds to the ssDNA via its association with RecO. (4) The now complete RecFOR complex, plus or minus RecF, facilitates nucleation of RecA onto the ssDNA substrate to initiate homologous recombination. [121]

1.5.5 RuvABC

Strand exchange promoted by RecA ultimately leads to the formation of a four-stranded Holliday junction intermediate that must be resolved to restore the duplexes and complete the recombination process. Holliday junction resolution is performed by yet another complex of three proteins known as RuvABC, which completes

the DNA recombination by cutting the DNA structure using paired incisions to yield nicked duplexes that can be sealed by DNA ligase [36]. The three proteins have been shown to be important for genome maintenance since mutation of *ruvA*, *ruvB* or *ruvC* leads to sensitivity to UV radiation, ionizing radiation and certain genotoxic antibiotics [122].

RuvA is a 22 kDa protein that forms a tetramer with grooves across one surface that can accommodate the four-stranded branched structure of the Holliday junction [123, 124] (Figure 1.6). The junction bound by RuvA is held in a square planar configuration with each arm of the junction pointing away from the crossover and in an ideal orientation for symmetrical cleavage [125]. The junction binding surface is positively charged enabling binding of the negatively charged phosphodiester backbone of DNA and hold it in place [123] (Figure 1.6). Four pairs of negatively charged residues at the centre of the tetramer determine Holliday junction specificity by preventing binding of linear duplex DNA [126].

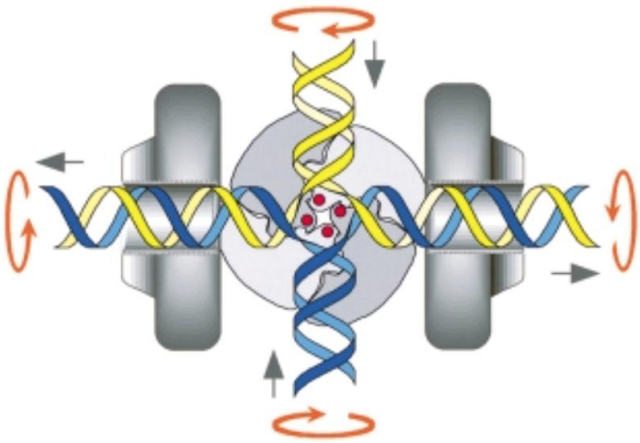

Figure 1.6. Model of Holliday junction bound by the RuvAB complex. A tetramer of RuvA at the centre holds the junctions in a square planar conformation. Negatively charged 'pins' (red circles) on each RuvA monomer prevent linear duplex binding. Two hexameric rings of RuvB assemble on opposing duplexes via protein-protein interactions with RuvA. Holliday junction branch migration is achieved by drawing the duplexes through RuvB ATP hydrolysis across the RuvA surface to generate heteroduplex DNA [127].

Once RuvA has bound a Holliday junction it acts as a target for the loading of RuvB onto the protein-DNA complex. RuvB is a 37 kDa protein with an ATP binding domain which assembles onto a Holliday junction when in the presence of both ATP and magnesium, either as one homohexameric ring or as a pair of two-fold symmetric rings [128]. Pairs of RuvB rings contact the RuvA tetramer from opposing sides which drives formation of heteroduplex DNA, via branch migration, using the RuvB ATP fuelled motor [129]. The RuvAB complex is also believed to interact directly with RecA to release it from DNA [130, 131]. With RecA removed, the Holliday junction is ready to be cleaved and the joined molecules returned to separate duplexes to complete the recombinational exchange.

RuvC is a 19 kDa endonuclease which acts as the resolving enzyme in the Holliday junction resolution process [34, 36, 95, 132, 133]. RuvC functions as a dimer allowing for symmetrical paired incisions on either side of the

junction at the site of strand crossover [123, 134, 135] resulting in two nicked duplexes, which can be repaired by DNA ligase activity to complete the recombination process [135] (Figure 1.7).

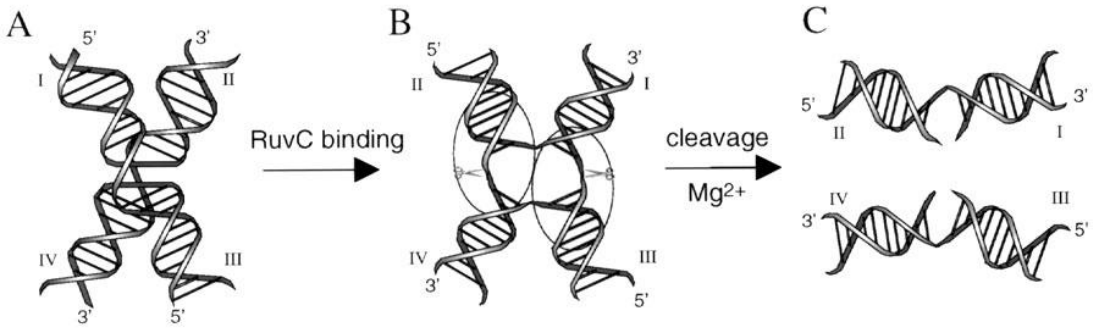

Figure 1.7. Holliday junction resolution by RuvC. (A) Holliday junction in a folded, stacked-X, confirmation prior to manipulation by RuvAB and subsequently by RuvC. (B) RuvC partially opens up the junction core and introduces symmetrical paired incisions in the presence of magnesium ions. (C) Junction resolution produces two nicked DNA duplexes that can be repaired by ligase action completing recombination. [136]

1.6 λ Red Recombination

When the phage infects a host cell it shuts off many of the host recombination pathways in order to control conditions to suit its own DNA replication and recombination strategies. The phage λ recombination pathway is known as the Red system, discovered when λ was found to be capable of undergoing recombination inside cells which were defective in the *recA*, *recB* or *recC* genes [137]. As with many of the bacterial pathways described above, the Red system is composed of three proteins coded by a cluster of genes necessary for recombination. In this case the genes all lie within the P_L operon coding for the proteins known as Redα/Exo, Redβ/Bet and Redγ/Gam [138, 139].

1.6.1 Exo

Exo is a 24 kDa exonuclease, also known as λ exonuclease, which attacks dsDNA ends, degrading one strand in the5' to 3' direction [140-146]. Exo assembles as a ring made of three monomers, forming a tapered central channel that is 30Å wide at the one end but only 15 Å wide at the other (Figure 1.8). This allows dsDNA to enter at the wide end but only single stranded to exit the narrow end of the tunnel, thus ensuring that any DNA passing through the trimer is degraded with enhanced processivity [147, 148].

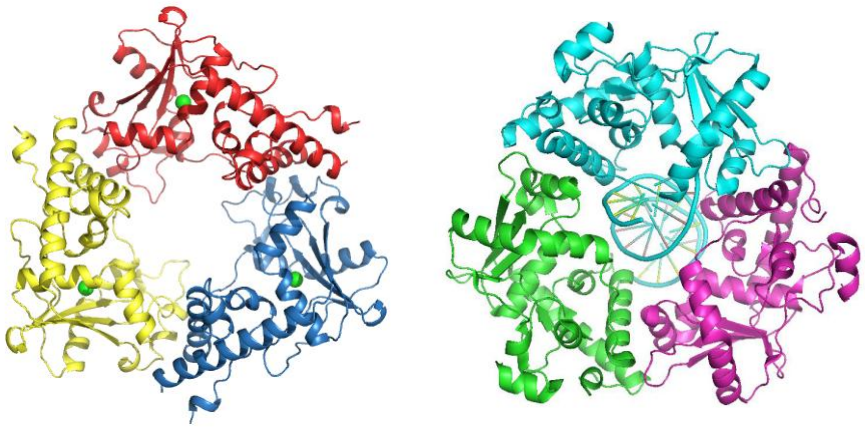

Figure 1.8. Quaternary structure of λ Exo trimer. View from the narrow end of the channel. The unbound structure (left) shows three exo proteins (red, blue and yellow) forming a trimer with the location of magnesium ions as green spheres in the active site of each subunit (PDB: 1AVQ). Bound structure (right) shows a 12 bp symmetric DNA duplex (cyan) entering an exo trimer (cyan, purple and green) at the wider end being degraded to a single strand to fit through near face narrower end (PDB: 3SLP). [147]

1.6.2 Bet

Bet is a 29 kDa protein which binds the 3'-tailed ssDNA exposed by the exonuclease action of Exo and promotes strand annealing with complementary single strands generated from other λ DNA molecules [149]. Bet associates with Exo to allow loading directly onto newly formed single strands, overcoming any inhibitory problems from host SSB assembly, and also stimulates Exo nuclease activity when together in a protein-protein complex [144, 150, 151]. Bet has been shown to assemble into different multimeric structures depending on whether DNA is present or not (Figure 1.9). In the absence of DNA Bet forms an inactive ring structure composed of 12 subunits. In the presence of ssDNA Bet assembles into a 15-18 subunit ring that is capable of strand annealing. Once complementary strands of ssDNA have been annealed to form dsDNA, Bet arranges into left-handed helical filaments, potentially as a result of ring structures merging [152]. The action of Exo and Bet resembles the recombination reactions performed by RecBCD and RecA at a DNA break, although Bet is largely restricted to strand annealing, whereas RecA predominantly performs strand invasion reactions [151, 153].

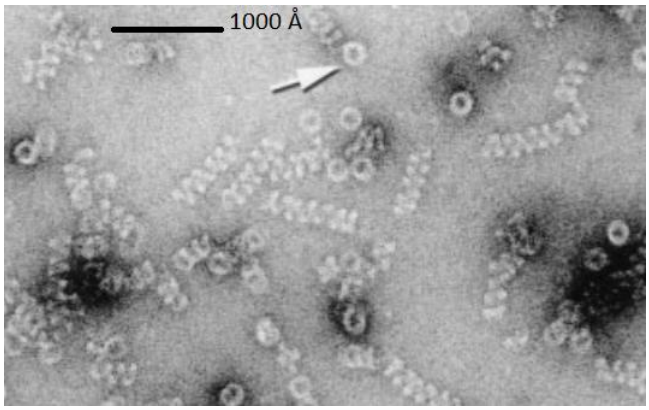

Figure 1.9. Electron micrograph of Bet rings and helices. The arrow shows a Bet ring associated with a helical tail where protein has been incubated with partial duplex DNA [152].

1.6.3 Gam

Gam acts to inhibit the exonuclease activity of RecBCD which would normally destroy the phage engaging in rolling circle replication. Gam is a 16 kDa dimer which binds the host RecBCD complex, preventing it from binding and degrading DNA, possibly as a DNA mimic [154-157]. The elimination of RecBCD means that the Red system is needed for the repair of double strand DNA break and means that the phage recombination pathway now predominates.

1.7 Orf

While Exo, Bet and Gam replace the action of RecBCD, λ also encodes a function that substitutes for the recombination pathway directed by the host RecFOR complex. For this purpose phage λ codes for a single protein called Orf [158, 159]. Genetic evidence suggested that Orf was a phage replacement protein for RecFOR alongside biochemical analysis showing that it preferentially bound to ssDNA and could also associate with the *E. coli* SSB protein [160]. Orf is a 17 kDa protein encoded by the λ gene *ninB* which is located in the *ninR* gene region along with two other DNA binding proteins which are the focus of this thesis: *ninG* and *ninH* [158, 159, 161].

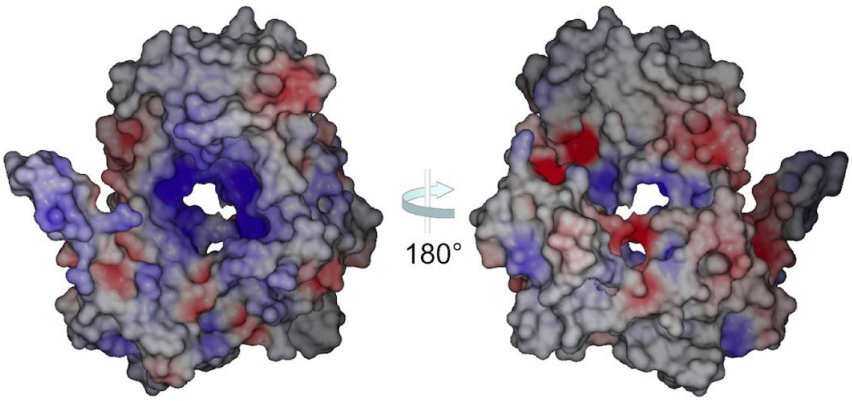

Figure 1.10. Calculated electrostatic properties of surface residues from the crystal structure of λ Orf protein. Areas in blue have a positive charge whereas those residues with a negative charge are in red. [160]

Unlike the Red system, Orf appears to serve primarily in replacing recombination for the phage itself, rather than additionally for the host bacterium [158, 162]. However, Orf is capable of aiding *E. coli* recombination if cells deficient in RecFOR carry the phage Red system proteins of Exo and Bet [163]. The crystal structure of Orf protein revealed a homodimer arranged as a toroid with a protrusion of the C-terminal helix and a cleft running perpendicular across the front of a central channel [160] (Figure 1.10). Like the Red protein Exo, the channel is wider at one side of the ring than the other, potentially allowing dsDNA to fit at one end with ssDNA passing through the dimer. However, the dimensions are considerably narrower than with Exo, making it

possible that even ssDNA could not be bound at the centre of the ring. Unlike Exo, displays preferential binding to ssDNA substrates, including DNA containing a single strand overhang and gapped duplexes[160].

Orf is able to bind dsDNA substrates, despite the central channel not being wide enough to accommodate duplex DNA, suggesting that the cleft across the face of the protein dimer is the most likely candidate for the DNA binding site. However, the central channel contains many positively charged and highly conserved residues, implying that this area could bind DNA via the negatively charged phosphodiester backbone. It has been suggested that the dimer may be able to open up its structure so that DNA, perhaps at the junction between double and single-stranded DNA, could traverse the channel when bound [164].

1.8 Rap

Another of the genes in this *ninR* region, *ninG*, specifies a Holliday junction resolvase called Rap that is a functional equivalent of the host RuvC protein [36, 165-168]. Studies on the interaction of the genes from the *ninR* region and the bacterial RecBCD dependent pathways revealed that Rap is key to the propagation of the phage since λ *rap* mutants show one hundred times less recombination than that of wild type phage [169, 170]. There is also significantly less recombination in the phage Red pathway following mutation of *rap* [161].

The *ninG* gene appears to be analogous to *rusA*, encoding another structurally distinct Holliday junction resolvase, as many lambdoid phages have either *rap* or *rusA* at the same location in their genomes [171]. RusA is highly specific for resolving the 4-stranded Holliday junction structure, despite being able to bind many different branched sequence substrates [172, 173]. Rap can resolve Holliday junctions by symmetrically paired incisions as with RusA and RuvC, however, it can also cleave 3-stranded junctions, such as D-loops, flaps and Y-junctions[167, 168]. This ability to cleave three stranded junctions may allow Rap to complete the processing of splice recombinants generated by strand annealing via the Red system [174]. Thus Rap may serve as a general debranching endonuclease as well as a Holliday junction resolvase during phage recombination.

1.9 Aims

The aims of this project were to further characterise the three proteins; Orf, Rap and NinH. Specifically, to identify how the central channel and surface cleft of the Orf protein interact with ssDNA to complete research for a publication (Appendix A). The NinH protein was investigated to establish which DNA substrates it was capable of binding and whether NinH behaved similarly to the highly homologous Fis protein, both through DNA binding studies and through in vivo tests analysing if NinH could attenuate any loss in function through mutations in Fis and similar nucleoid associated proteins. And finally work on the Rap protein was undertaken to confirm the predicted HNH catalytic domain and investigate the effect that mutations within this region would confer on the binding and cleavage of a Holliday junction structure.

Chapter 2. Materials and Methods

2.1 Strains and plasmids

Table 2.1 *Escherichia coli* strains

Strain	Genotype	Source
W3110	F^- λ^- IN(*rrnD-rrnE*)1 *rph-1*	Hayashi, K et al. [175]
DH5α	F^- Φ80*lacZ*ΔM15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17*(r_k^-, m_k^+) *phoA supE44 thi-1 gyrA96 relA1* λ^-	Invitrogen
Mach1 T1R	F^- Φ80*lacZ*ΔM15 Δ*lacX74 hsdR*(r_k^-, m_k^+) Δ*recA1398 endA1 tonA*	Invitrogen
BL21-AI	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm araB*::*T7RNAP-tetA*	Invitrogen
BL21-Rosetta (DE3)	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^r)	Novagen
BW25113	*rrnB3* Δ*lacZ4787 hsdR514* Δ(*araBAD*)*567* Δ(*rhaBAD*)*568 rph-1*	Keio collection [176]
JW3229	As BW25113 but Δ*fis*::*kan*	Keio collection [176]
JW1702	As BW25113 but Δ*ihfA*::*kan*	Keio collection [176]
JW0895	As BW25113 but Δ*ihfB*::*kan*	Keio collection [176]
JW1225	As BW25113 but Δ*hns*::*kan*	Keio collection [176]
JW2644	As BW25113 but Δ*stpA*::*kan*	Keio collection [176]

Table 2.2 Plasmid constructs

Plasmid	Details	Source
pT7-7	overxpression vector, T7 promoter, Ap^r	
pET14b	overxpression vector, T7 promoter, Ap^r	
pET22b	expression vector, MCS, pelB coding sequence, T7 promoter, lacI coding sequence, pBR322 origin, Ap^r, f1 origin for ssDNA production	
pUC18	cloning vector, MCS at lacZ region	
pFC109	PCR product from λ cI857 S*am7* cut with NdeI/HindIII and inserted into pT7-7 cut with NdeI/HindIII, Ap^r	FA Curtis, unpublished
pFC110	PCR product from λ cI857 S*am7* cut with NdeI/HindIII and inserted into pT7-7 cut with NdeI/HindIII, Ap^r	FA Curtis, unpublished
pNinH-pUC18	XbaI-HindIII insert from pFC109 inserted into pUC18 cut with XbaI-HindIII	PD Townsend, unpublished
pN-His-Rap	PCR product from λ cI857 S*am7* cut with NdeI/HindIII and inserted into pET22b cut with NdeI/HindIII. An N-terminal histidine-tagged fusion. Ap^r	PD Townsend, unpublished
pC-His-Rap	PCR product from λ cI857 S*am7* cut with NdeI/XhoI and inserted into pET22b cut with NdeI/XhoI. A C-terminal histidine-tagged fusion. Ap^r	PD Townsend, unpublished

pRapC101S	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished
pRapD113A	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished
pRapH116T	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished
pRapR127A	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished
pRapN132A	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished
pRapN141D	Site-directed mutagenesis of pN-His-Rap	PD Townsend, unpublished

2.2 Cloning

2.2.1 Restriction digests

Reaction mix contained 1 µg DNA, 1-2 units of restriction endonuclease in appropriate reaction buffers. Digests were typically incubated at 37°C for 120 minutes. The following enzyme combinations were used, NdeI and BamHI in reaction buffer 2 (50 mM NaCl, 10 mM Tris-HCL pH 7.9, 10 mM $MgCl_2$, 1 mM DTT) or XbaI and HindIII in buffer 3 (100 mM NaCl, 50 mM Tris-HCL pH 7.9, 10 mM $MgCl_2$, 1 mM DTT). Enzymes and premade reaction buffers were obtained from New England BioLabs.

2.2.2 Ligation

Ligation mixes contained 8 µg insert DNA, 3 µg vector, 1 unit T4 DNA ligase in DNA ligase buffer (30 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP) and incubated at 21°C overnight. Buffers and enzymes were supplied by Promega.

2.3 Protein expression

2.3.1 Transformation

E. coli cells were grown in LB broth (casein enzymic hydrolysate 10 g/L, $NaCl_2$ 10 g/L, yeast extract 5 g/L, pH 7.5) to an A_{650nm} of 0.6, measured by Boeco S.30 spectrophotometer, then pelleted by centrifugation at 5,000 rpm for 10 min in a Boeco U-320R benchtop centrifuge. The supernatant was discarded and the bacterial pellet resuspended in 100 µl of 100 mM $CaCl_2$ per 1 ml of centrifuged culture. 0.1-0.5 µg of plasmid DNA was added to 250 µl of competent cells and incubated on ice for 20 minutes followed by 45 seconds at 42°C. Cells were returned to ice and 300 µl of LB broth added and incubated for a further 20 minutes at 37°C. 250 µl of the cell culture was spread onto LB agar supplemented with the relevant antibiotic (100 µg/ml ampicillin or 50 µg/ml kanamycin) and incubated overnight at 37°C.

2.3.2 Plasmid DNA preparation

All buffers for this procedure were supplied by QIAGEN as part of the QIAprep Spin Miniprep Kit and used according to the manufacturer's instructions. 5 ml of an overnight culture was pelleted at 5,000 rpm for 10 minutes and the supernatant discarded. The pellet was resuspended in 250 µl of buffer P1 (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNaseA)and transferred into a microcentrifuge tube with 250 µl of buffer P2 (200 mM NaOH, 1% SDS)and mixed by inverting. After 3 minutes the reaction was stopped by the addition of 350 µl of buffer N3 (4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8) and mixed by inverting. The solution was centrifuged for 10 minutes at 13,000rpm in a MSE Microcentaur and the supernatant loaded onto a QIAprep spin column and centrifuged for 1 minute at 13,000 rpm again with the flow through discarded. The column was washed with 500 µl buffer PB (5 M Gu-HCl, 30% isopropanol) and 750 µl buffer PE (10 mM Tris-HCl pH 7.5, 80% ethanol), centrifuging for 1 minute at 13,000rpm at each stage. Plasmid DNA was eluted by addition of 30 µl buffer EB (10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0), leaving to stand for 1 minute then centrifugation for 1 minute at 13,000 rpm.

2.3.3 Protein overexpression

Small scale samples were used to confirm protein overexpression from BL21-AI or BL21-Rosetta(DE3) strains transformed with plasmid constructs. 500 µl of an overnight culture was added to 4.5 ml LB broth supplemented with antibiotic and grown to A_{650nm} of 0.5. The culture was split into two and one half was induced with 1 mM IPTG (and 0.2% arabinose for BL21-AI strains) while the other half served as an uninduced control. Both cultures were incubated at 37°C with shaking at 200 rpm and 500 µl samples removed from each at 1, 2 and 3 hours. These samples were pelleted for 2 minutes at 9,000 rpm and the supernatant discarded. Cell pellets were resuspended in 50 µl Buffer A (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM DTT, 75 mM KCl) and 50 µl SDS loading dye (50 mM Tris-Cl pH 6.8, 100 mM dithiothreitol, 2% SDS, 0.1% bromophenol blue, 10% glycerol) prior to separation on 10-15% SDS-PAGE.

Once protein overexpression was confirmed, large scale cultures were prepared for protein purification. Each 1 litre flask of LB broth supplemented with antibiotic was inoculated with 20 ml of an overnight culture and grown at 37°C and shaking at 180 rpm to an A_{650nm} of 0.6. IPTG (1 mM) or arabinose (0.2%; for BL21-AI strains) and incubated for a further 3 hours under the same conditions. Cells were recovered by centrifugation at 4,000 rpm for 10 minutes in a Beckman-Coulter JLA 8.100. The supernatant was discarded and pelleted cells resuspended in 20 ml of Buffer A, or His-binding buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA) in the case of histidine-tagged proteins.

2.3.4 Protein purification

NinH, His-NinH, MBP-Orf and His-Orf proteins were purified prior to the start of the project by GJ Sharples, FA Curtis, LY Bowers. His-Rap proteins were purified during the project as described below.

2.3.4.1 Histidine-tagged fusions

The bacterial cell suspension was sonicated in an MSE Soniprep 150 for 180 seconds at medium power and then centrifuged for 20 minutes at 13,000 rpm in a Beckman-Coulter JA 25.50. The supernatant was mixed with 2 ml Ni-NTA His-affinity resin (Sigma) and incubated for 1 hour at 4°C with gentle agitation on a Stuart Scientific SRT1 roller mixer. The resin was recovered by centrifuging for 3 minutes at 3,000 rpm and loaded into a column, the supernatant was retained as the flow through containing unbound proteins. The column was washed with 10 ml His-binding buffer and eluted with a 0-250 mM gradient of imidiazole in His-binding buffer.

2.3.4.2 Heparin columns

Purified proteins from the Ni-NTA resin were dialysed overnight into buffer (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM DTT) before further purification on a 1 ml heparin-agarose (Sigma) column and eluted using a 0 M to 1 M KCl gradient in 0.1 M intervals in 1 ml fractions in Buffer A. Peak fractions from the heparin column were dialysed overnight into storage buffer (20 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM DTT, 50% glycerol) and stored in aliquots at -80°C Protein purity was confirmed by analysis on SDS-PAGE.

2.4 Gel electrophoresis

2.4.1 Agarose gel electrophoresis

Agarose gels were made with 1.5% agarose in 1x TBE buffer (90 mM Tris-Borate, 2 mM EDTA). DNA samples were mixed with 6x DNA loading dye (New England BioLabs) and loaded onto the gel at 50 V in 1x TBE buffer for 40 minutes before visualisation under UV light in a Bio-Rad GelDoc system.

2.4.2 SDS-PAGE

Protein samples for SDS-PAGE were mixed with SDS loading dye (15 µl protein sample and 5 µl dye) and denatured at 95°C for 5 minutes. 15 µl samples were separated on 15% SDS-PAGE for 90 minutes at 180 V in 1x SDS-PAGE running buffer (25 mM Tris-HCl, 200 mM glycine, 0.1% SDS). Gels were stained with Coomassie blue (40% methanol, 10% acetic acid, 0.025% Coomasie Brilliant Blue R-250) for 20-30 minutes and washed with destain (40% methanol, 10% acetic acid) until protein bands were visible.

2.5 Bacterial growth curves

Each well on a 96-well plate contained 10 µl overnight bacterial culture and 90 µl LB broth with the appropriate antibiotic. Cultures were grown with continuous shaking in a Biotek Synergy H4 plate reader at 37°C for 14 hours. Readings at A_{650nm} were taken every 15 minutes.

2.6 DNA binding and cleavage

2.6.1 ^{32}P-labelling of DNA

For DNA binding and cleavage assays, a single strand from each substrate was labelled with [γ^{32}P] ATP at the 5' end using 1-2 units of T4 polynucleotide kinase (Invitrogen). Labelled DNA was separated from unincorporated nucleotide using micro-BioSpin P30 gel columns (BioRad). Annealed DNA substrates were further purified by separation on 10% polyacrylamide gels in 90 mM Tris-borate, 2 mM EDTA. Bands were excised from the gel and transferred to microcentrifuge tubes containing sterile distilled water. Labelled substrate was recovered after incubation at 4°C overnight.

2.6.2 DNA binding assays

Protein samples in 1x GBBG (50 mM Tris-HCl pH 8.0, 5 mM EDTA, 1 mM DTT, 100 µg/ml BSA, 5% glycerol) was added to a binding mixture containing 1.5 nM ^{32}P-labelled DNA and 1x GBBG (20 µl total volume) and incubated on ice for 15 minutes. Controls without protein contained 2 µl 1x GBBG in place of the 2 µl protein sample. 12 µl of each sample was loaded on a 4% polyacrylamide gel and electroporesed at 160 V for 75 minutes in 1x LIS buffer (6.7 mM Tris-HCl pH 8.0, 3.3 mM sodium acetate, 2 mM EDTA).

2.6.3 DNA cleavage assays

Rap protein was added to a reactions (20 µl in total) containing 3 nM DNA, 1 mM $MnCl_2$ in cleavage buffer (50 mM Tris-HCl pH 8.0, 1 mM DTT, 100 µg/ml BSA) and incubated at 37°C for 30 minutes. Reactions were terminated by addition of 5 µl of 5x cleavage stop buffer (20 mM Tris-HCl pH 8.0, 0.5% SDS, 20 mM EDTA, 8 units/ml proteinase K) and incubated at 37°C for a further 10 minutes. The sample was then mixed with 4 µl of 6x loading dye and 10 µl loaded onto 10% TBE gel (90 mM Tris-borate, 2 mM EDTA) and electrophoresed for 105 minutes at 190 V.

2.6.4 Imaging gels containing ^{32}P-labelled DNA

Following electrophoresis, gels were transferred onto 3MM blotting paper and dried in a Biorad model 583 Gel Dryer for 20-40 minutes. Dried gels were placed in phosphorimaging screens and scanned on a Fuji FLA-3000 after overnight exposure. Gels were also exposed to X-ray film and exposed for varying lengths of time before developing in an Xograph Compact X4.

2.7 Data Analysis

2.7.1 ImageJ

Gels for quantitative analysis were exposed to a phosphorimaging screen for 1-18 hours and scanned using a Fuji FLA-3000 fluorescent imager with the resulting images saved in TIFF format. Images were then analysed with the image analysis software tool ImageJ (imagej.net). Lanes were selected using the 'rectangle' tool (Figure 2.3.A) and band density analysed on a grey-scale giving peaks of light intensity for each band (Figure 2.3.B). The background was removed by drawing the 'straight' tool through the baseline and peaks were separated by use of the same tool through the lowest point between peaks (both shown Figure 2.3.B). The area beneath each peak was selected using the 'wand' tool resulting in a numerical value according to the density of each band (Figure 2.3.C). Values were converted from peak area into a percentage of bound DNA in each lane derived from the total density of bands. Data are the mean of three independent experiments and the binding percentages averaged and standard deviation determined. Bound DNA is represented as a bar, with shading used to distinguish different protein-DNA complexes if more than one was evident; light grey represents the first complex formed by DNA binding while dark grey shows the formation of a second complex.

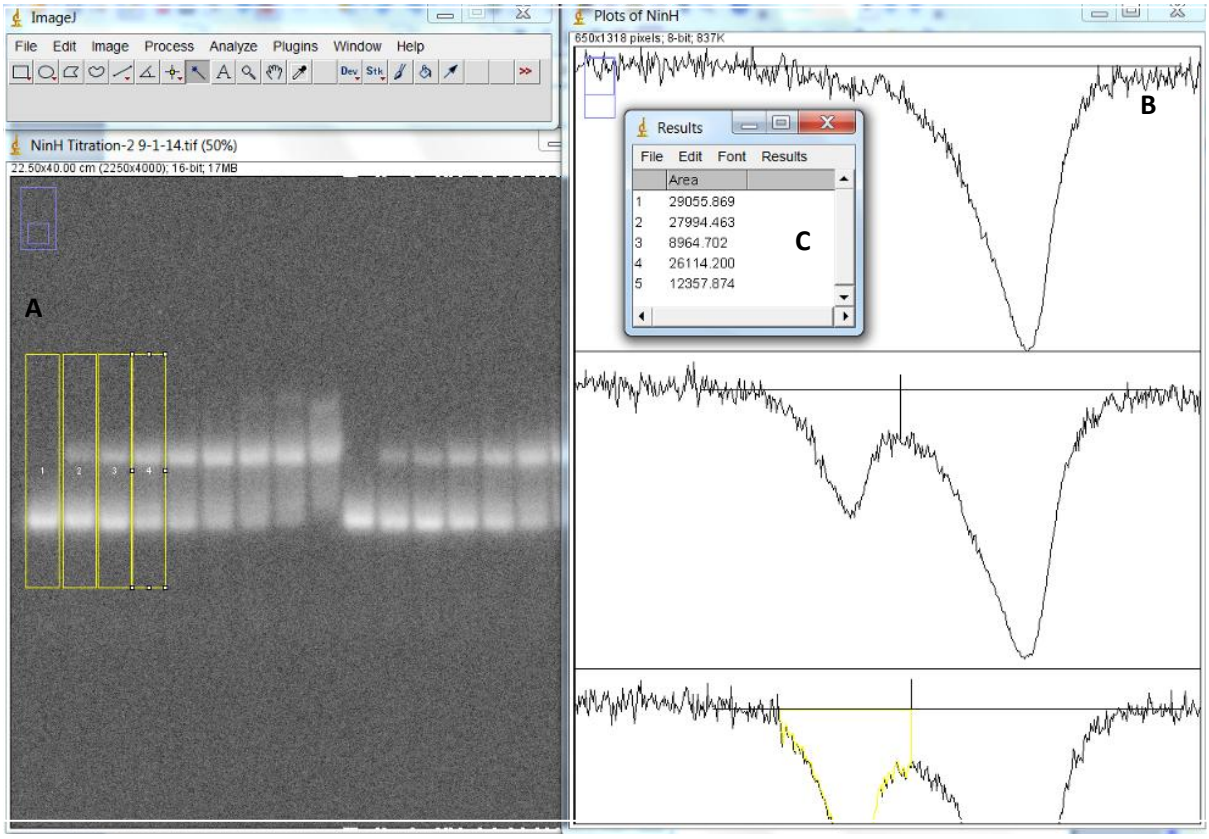

Figure 2.3 ImageJ processing of gel shift experiments using ^{32}P-radiolabelled DNA. (A) The rectangle tool (yellow) was used to select each lane. (B) The band density was plotted for each lane from left to right was represented graphically and peaks corresponding to bound and unbound DNA delineated by the line tool. (C) The area of each peak was calculated and used to determine the percentage of DNA bound from the total band density in each lane.

2.7.2 Growth curve analysis

Data from the Biotek Synergy H4 plate reader was exported from the Biotek Gen5 analysis software into Microsoft Excel. Absorbance values from wells containing media alone were subtracted from the rest of the data. Each 96-well plate contained three replicates of each bacterial strain and each set of data was averaged. Three independent plates were set up and the triplicate averages within each plate were averaged between the three experiments with a standard deviation calculated between values for each of the three plates. Absorbance values were plotted on a line graph against time of incubation.

2.7.3 Growth difference analysis

The averages of each set of three experiments were taken and standardised to starting point of the negative control of the blank vector pT7-7 for each strain. The starting absorbance of each strain with pT7-7 was subtracted from the absorbance at every time-point for both the growth curves of pT7-7 and pFC109 so that all sets of data start at zero. Next the difference in growth between bacteria containing pT7-7 and pFC109 was calculated for each strain by subtracting the value for pT7-7 from the value for pFC109 at each time-point leaving the difference in growth, positive or negative, when bacteria contain the *ninH* gene expressed from pFC109. Finally the effect on growth of pFC109 in each mutant relative to wild type was calculated by subtracting the value for the difference in growth in wild type from the difference in growth for each mutant giving a negative value for where the mutant grows less than the wild type does at that time point and positive where there is more growth. Values were plotted on a line graph showing growth +/- relative to the wild type value of constant zero.

Chapter 3. λ Orf DNA binding

Genetic studies have shown that bacteriophage λ Orf protein is involved in phage recombination [158, 162], likely in the early stages of strand exchange. Consistent with this, purified Orf binds preferentially to ssDNA over dsDNA[160], The crystal structure of the *apo* form of Orf revealed that the protein forms a ring shaped dimer although the precise DNA binding site was not immediately obvious [160, 177]. It was suggested that ssDNA could pass through the central channel of the protein although the cavity is very narrow at one end. In an effort to pinpoint the region that interacts with DNA, six substitution mutants (Figure 3.1.A) were made around the central channel of Orf, residues selected based on how well-conserved they are among Orf family proteins. DNA binding experiments were undertaken to confirm whether or not these mutant proteins would show any defect in association with different DNA substrates.

The mutant proteins had previously been purified as N-terminal MBP tagged fusions and tested for their capacity for binding ssDNA but follow up binding experiments were conducted on a 60 bp double stranded DNA substrate and a bubble structure of the same sequence featuring a twenty nucleotide unpaired region flanked by 20 bp duplexes on each side which has some features that resemble the D-loop structure that is commonly found in DNA during replication and recombination (Appendix A), although it is missing the invading strand. Mutant proteins were incubated on ice with the ^{32}P labelled DNA substrates (as described in 2.6.2) and the results of the gel shift analysis are shown in Figure 3.1.B.

3.1 Orf binding to duplex and bubble

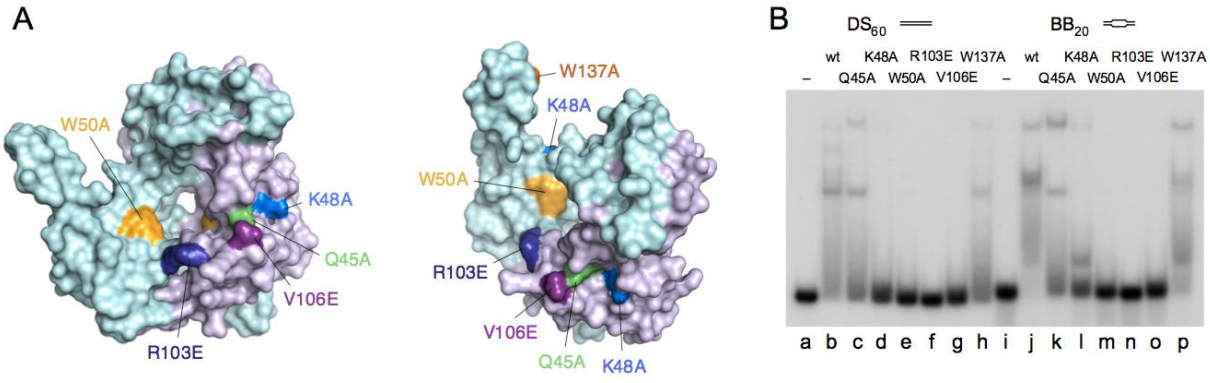

Figure 3.1. Location of Orf mutants and DNA binding analysis. (A) Location of mutated residues around central channel of Orf. (B) Orf mutant binding to double-stranded DNA (lanes a-h) and bubble DNA (i-p) in gel shift assays. Binding mixtures contained 0.15 nM ^{32}P-labelled DNA and MBP-Orf protein at 250 nM and were detected by autoradiography [178].

The DNA binding mixtures contained 0.15 nM ^{32}P-labelled DNA and MBP-Orf proteins at 250 nM. Wild-type protein at this concentration bound most of the duplex DNA (DS_{60}) although the interaction is relatively weak,

with smearing indicative of unstable protein-DNA complexes (Figure 3.1.B lane b). Three stable complexes were formed with wt MBP-Orf and likely represent assembly of more than one dimer onto the DNA. MBP-Orf wt bound all of the BB_{20} bubble DNA, forming stable complexes with similar migration although the two faster-migrating species are located close together in the band shift (Figure 3.1.B lane j). Some smearing was still evident with MBP-Orf wt consistent with unstable binding, however, the improved binding relative to DS_{60} is in keeping with the availability of ssDNA in the bubble structure.

The MBP-Orf Q45A mutant (Figure 3.1.A) showed a slight reduction in binding to the duplex compared to the wt as judged by the increase in unbound substrate (Figure 3.1.B lane c). This residue may assist in stabilising binding to DNA, although Q45A is still capable of forming complexes with a similar pattern to the wt. With bubble DNA, Q45A showed a significant reduction in binding compared to the wt (Figure 3.1.B lanes j and k) with much more of the substrate remaining unbound or smeared. The pattern of complexes also differs, with Q45A lacking the two closely-migrating complexes observed with the wt. The Q45A mutant may therefore have a more important role in stabilising contacts with ssDNA than with dsDNA, or at the junction between these in a bubble substrate.

The K48A mutant is located beyond the rim of the central channel, although it lies within a groove on one face of the protein that runs perpendicular to the cavity and could contribute to DNA binding (Figure 3.1.A). Compared to the wt, the K48A mutant has much less DNA binding activity. A faint band is visible at the top of the gel with both DS_{60} and BB_{20} substrates although most of the DNA (Figure 3.1.B lanes d and l) in a band of similar position and intensity to the no protein control, showing that the majority of DNA has not been bound by this mutant protein. There is a relatively dense band present with BB_{20} just above the position of unbound DNA suggesting that K48A does form a weak complex with bubble DNA (Figure 3.1.B lane l) although it is at a position well below the complexes observed with the wt on either duplex or bubble substrates.

The next three lanes (Figure 3.1.B lanes e-g and m-o) contain W50A, R103E and V106E which lie close together in a cleft at the entrance of the central cavity. All three of the mutants failed to form protein-DNA complexes with duplex and bubble indicating that they lack DNA binding activity.

Finally the W137A mutant showed similar binding to the wild type protein on both substrates. On dsDNA the mutant mirrors the pattern of a faint band at the top, a denser band running to the middle with a smear of unstable complexes down to the unbound substrate (Figure 3.1.B lane h). On bubble DNA a similar pattern is observed (Figure 3.1.B lane p), with the closely-migrating pair of complexes present but with more of the faster mobility complex, also seen with K48A. The presence of some unbound DNA does indicate a slight reduction in DNA binding not too dissimilar from that noted with Q45A (Figure 3.1.B compare lanes b, c, h and j, k, p)

3.2 Possible Orf clamp model for binding to ssDNA

Analysis of the Orf crystal structure revealed that hydrogen bonding is extensive in the N-terminal intertwined β-sheet region of the Orf dimer, which corresponds to the RAGNYA motif [179], whereas hydrogen bonding within the distal region of the larger C-terminal section is considerably less extensive. It is possible that stable

packing and strong hydrogen bonding within the N-terminal asymmetric region of the Orf dimer allows it to act as a hinge, permitting opening of the dimer in the C-terminal region between the antiparallel β-sheets. This would allow ssDNA access to the interior of the central channel which is highly positively charged and where several of the most conserved residues of the Orf family reside. Two approaches were devised to test whether the clamp model of OrfDNA binding was genuine. In the first, one of the two cysteines in Orf (Figure 3.2.A), Cys74 would be labelled with 1.5-IAEDANS dye to assess the open and closed states of the Orf dimer by fluorescence resonance energy transfer (FRET) with any of the nearby tryptophan residues; closing of the ring should occur in the presence of DNA and result in an increase of fluorescence. Orf contains two cysteines, so to avoid confusion by labelling both, Cys52 was replaced by theonine using site-directed mutagenesis (Figure 3.2.A). In the second approach, we attempted to force the Orf ring into a permanently closed state by adding two cysteines at His63 and Pro87 that have the potential to form a disulphide bridge (Figure 3.2.B). This should lock the ring in a closed position in the absence of a reducing agent such as dithiothreitol (DTT) but allow it to open in its presence. DNA binding should be affected under these different condition, particularly the assembly onto a gapped duplex DNA substrate.

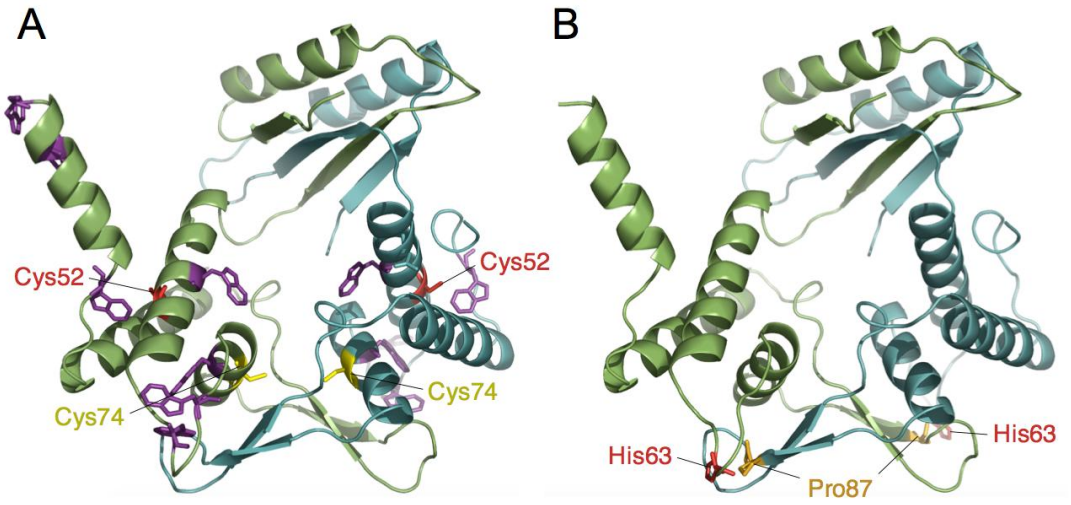

Figure 3.2. Location of residues in Orf relevant mutations used to investigate a hinge model of dimer opening. Orf subunits are coloured green and blue, tryptophans in Orf are coloured in purple. A. Cys52 was mutated to threonine. B. His63 and P87 were mutated to cysteine (PDB: 1PC6) [160].

Both C52T and H63C+P87C proteins were purified as N-terminal His-tagged fusions and binding to ssDNA evaluated in the presence or absence of DTT. The DNA binding assay was performed (as described in 2.6.2) and gel images analysed and quantified using ImageJ (following the method stated in 2.7.1).

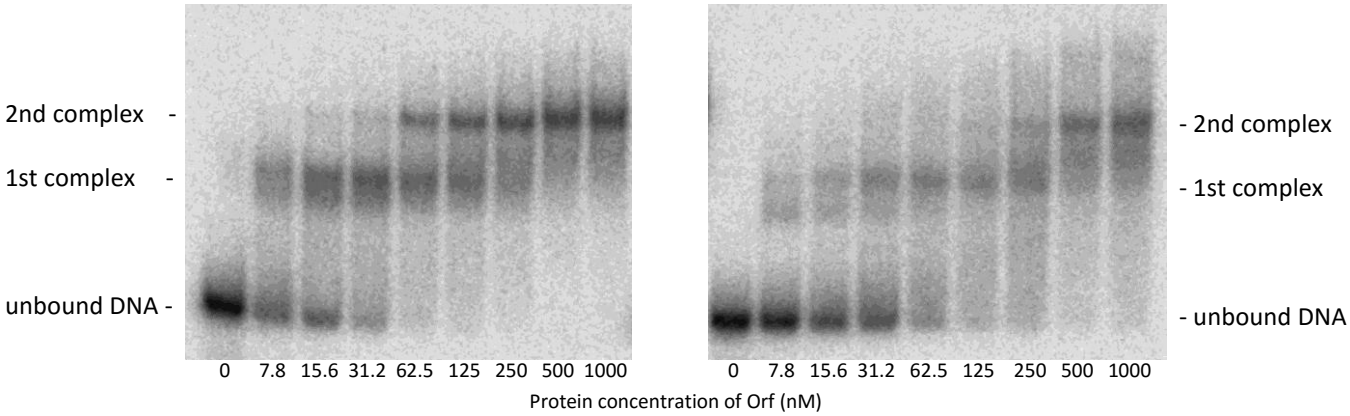

Figure 3.3. Representative Orf binding gel. Orf (1.52625-200 nM) binding to 0.15 nM ^{32}P-BB20. Complexes 1 and 2 are indicated. Gels were used to generate data for Figure 3.4 A and B. Gel A shows binding without DTT. Gel B shows binding with DTT present in the binding buffer.

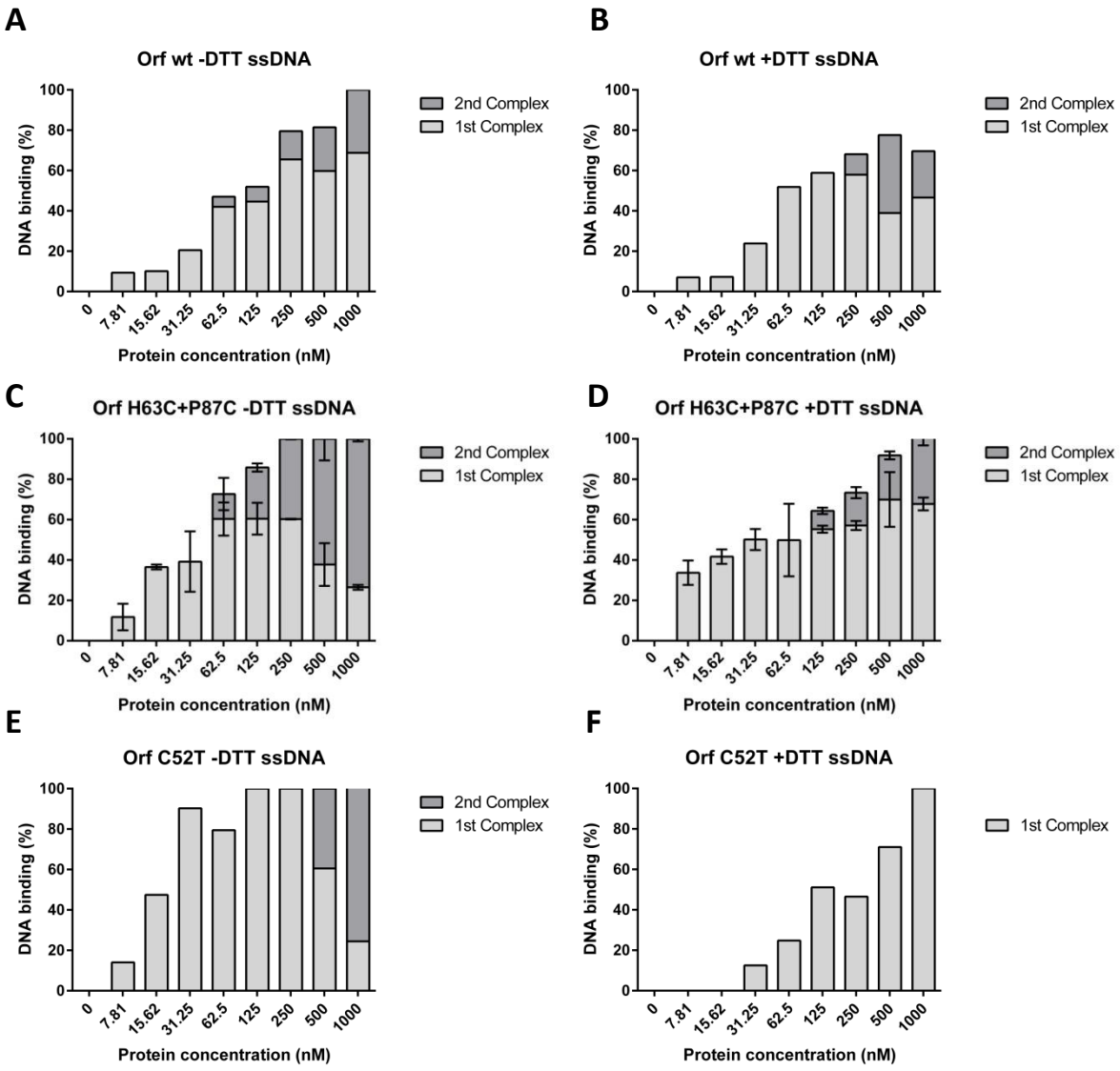

Figure 3.4. Orf mutants binding to ssDNA in gel shift assays. Binding mixtures contained 0.15 nM ^{32}P-labelled DNA (Appendix A), with and without DTT and detected by autoradiography; error bars on graphs C and D represent standard deviation of n=3

Orf wt protein without DTT shows a typical binding curve (Figure 3.4.A) with an increase in protein concentration correlating with an increase in percentage of bound DNA up to a maximum where at 1000nM all of the DNA is complexed with the protein. The first and second complex bars represent two different bands in the gel, the first where Orf has bound the DNA retarding it up the gel to one level, and the second when presumably a second Orf dimer assembles on the ssDNA structure retarding it even further up the gel to form a second band.

When DTT is introduced into the binding mixture the DNA binding activity of Orf is noticeably reduced, especially at the higher protein concentrations (Figure 3.4.B). The total amount of bound DNA is less than 80% at 1000 nM rather than the complete binding of ssDNA at this concentration in the absence of DTT (Figure 3.4.A). In addition, the appearance of a second protein-DNA complex only begins at 250nM protein in the presence of DTT, whereas 4-fold less protein is required to observe a second complex in the absence of DTT. Interactions between Orf dimers, stabilised by contacts between Cys52 and/or Cys74, may be disrupted by inclusion of DTT or there may be more indirect effects on the Orf wt structure influenced by reduction of any disulphide binds.

The H36C+P87C mutant, which adds a potential disulphide bridge by introducing another pair of cysteine residues, has a greater affinity for DNA binding than the wild type protein (Figure 3.4.C). Comparing the wild type and mutant in conditions without DTT (Figure 3.4.A and C) there was much more DNA bound at most protein concentrations by the mutant than the wild type. The mutant protein reaches 100% binding saturation at 250nM protein concentration, a 4 times lower concentration than required for wild type. Although the second complex is first formed at the same concentration by both mutants there is more of the complex visible at every concentration where it is present ranging from 2.35 to 3.47 times as dense a band in the mutant. This implies that the extra disulphide bridge possible in the mutant protein is improving the DNA binding capability of Orf, supporting the claim that the clamp action is important for protein activity.

As with the Orf wt, the addition of DTT resulted in some reduction in DNA binding with H63C+P87C (Figure 3.4.D), although binding was still better than the wt in the absence or presence of DTT. Complete binding saturation was not reached until 1000 nM of the mutant protein concentration, 4 times the concentration required to shift all of the DNA with H63C+P87C in the absence of DTT. The second complex also required more protein to form, with only 32.21% of the ssDNA bound into a second complex at the maximum protein concentration (reduced from 73.50% at 1000 nM without DTT addition). The improved binding of this mutant over the wt Orf with or without DTT, does suggest that disulphide bridges are forming and somehow allowing for a more stable association with ssDNA. The substantial improvement in formation of a second protein-DNA complex in the absence of DTT suggests that the closed ring assembly promotes more stable loading of multiple subunits, rather than the open form which should be favoured in the absence of DTT. Alternatively, the additional cysteines may simply contribute additional contacts for dimer multimeristation which are limited under reducing conditions.

The C52T mutant was engineered to remove an additional cysteine, which would hopefully not affect Orf binding to DNA. Surprisingly, the C52T mutant showed the greatest total DNA binding of any of the tested proteins in the absence of DTT, binding all of the substrate at 125 nM, half the concentrations required by

H63C+P87C and an eighth of that with Orf wt (Figure 3.4.A, C and E). However, the C52T mutant did struggle to form the second complex, despite reaching binding saturation at 125 nM a second complex is not formed until 500 nM protein, whereas the other two proteins are able to form this complex at concentrations as low as 62.5 nM. This is the first indication that formation of this complex is not correlated with maximal binding as suggested by the binding profiles of the wt and H63C+P87C.

When DTT is present in the binding mixtures (Figure 3.4.F), C52T displays rather weak binding and a second complex does not form even when all the DNA is bound at 1000 nM protein. Both wt and H63C+P87C exhibit some binding at 7.81 nM and above whereas no binding was detected with C52T until 31.25 nM. The Cys52 substitution to threonine therefore confers improved binding in the absence of DTT, suggesting that this residue normally limits the capacity to associate with ssDNA, perhaps by holding the dimer in a more closed state. In contrast, the C52T mutation reduces DNA binding when disulphide bonds are reduced in the presence of DTT as the closed state is less likely to be maintained. The reduction in formation of the second protein-ssDNA complex does show that Cys52 is important for stabilising the formation of multiple complexes on the DNA and this in turn may affect binding affinity.

3.3 Binding of H63C+P87C to bubble DNA

The H36C+P87C mutant and wild type Orf proteins were also tested under the same conditions for their binding to the bubble DNA structure. If the clamp-opening model for Orf is correct, assembly should depend on clamp opening as the ssDNA is flanked by duplexes and binding should be more stable as the protein once bound should be contained, preventing it from sliding off the substrate.

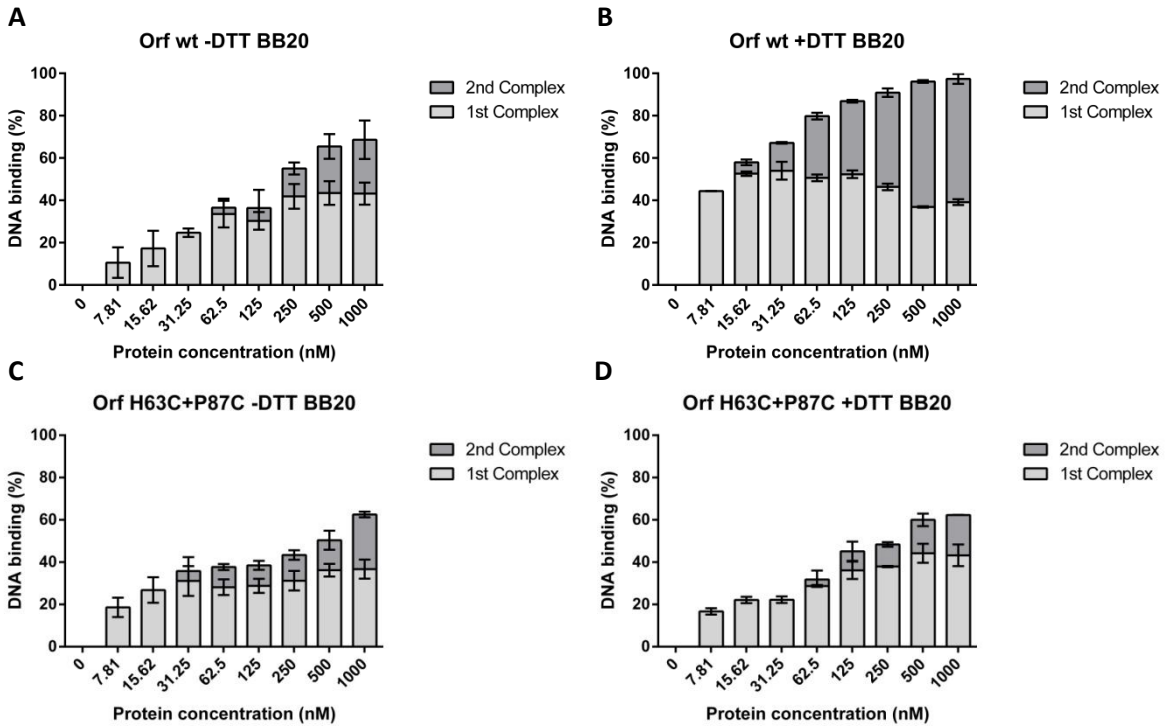

Figure3.5. Orf mutants binding bubble DNA in gel shift assays. Binding mixtures contained 0.15 nM ^{32}P-labbelled DNA (Appendix A), in the presence or absence of DTT and detected by autoradiography; error bars represent standard deviation of n=3

In the absence of DTT, wt Orf showed a similar binding isotherm as seen under the same conditions with ss-DNA (Figure 3.4.A and Figure 3.5.A), although only 68.65% of the bubble DNA was bound at 1000 nM protein, whereas all the ssDNA was bound at this protein concentration. The formation of a second complex was observed from 62.5 nM as noted with the wt on ssDNA.

In contrast to ssDNA, the introduction of DTT led to significantly improved DNA binding by wt Orf on the bubble DNA (Figure 3.5.B). The protein was able to bind well at all concentrations, starting with 44.38% of total bubble DNA bound at the lowest concentration of 7.81 nM, up from 10.57% was ssDNA at this concentration (Figure 3.4.B and Figure 3.5.B). Although the protein never bound all of the DNA at any concentration, it did reach 90% bound from 250 nM and shows a great capacity for the formation of the second complex. The second complex was formed at 15.62 nM protein concentration, earlier than any other protein tested, and on average 4.7 times as much when DTT was present than absent. The results are consistent with assembly on one or both strands and that binding is more stable than ssDNA due to the presence of flanking duplexes. Why the formation of the second complex on bubble DNA is favoured in the presence of DTT with the bubble structure rather than ssDNA is not clear.

H63C+P87C behaved very like the wt on bubble DNA in the absence of DTT, regardless of whether DTT was present or not (Figure 3.5.). None of the three conditions, wt -DTT, H63C+P87C -DTT or H63C+P87C +DTT, differed significantly in the total amount of DNA bound (Figure 3.5.). The binding observed with H63C+P87C was generally poorer than on ssDNA, possibly suggesting that restrictions in clamp opening by the introduced

cysteine pairs hampers assembly on bubble but not on ssDNA. The improved binding to the bubble noted with the wt in the presence of DTT does not occur with H63C+P87C (Figure 3.5.). H63C+P87C without DTT forms a second complex slightly earlier than both wt -DTT and H63C+P87C +DTT (Figure 3.5.C), although these difference may not be significant.

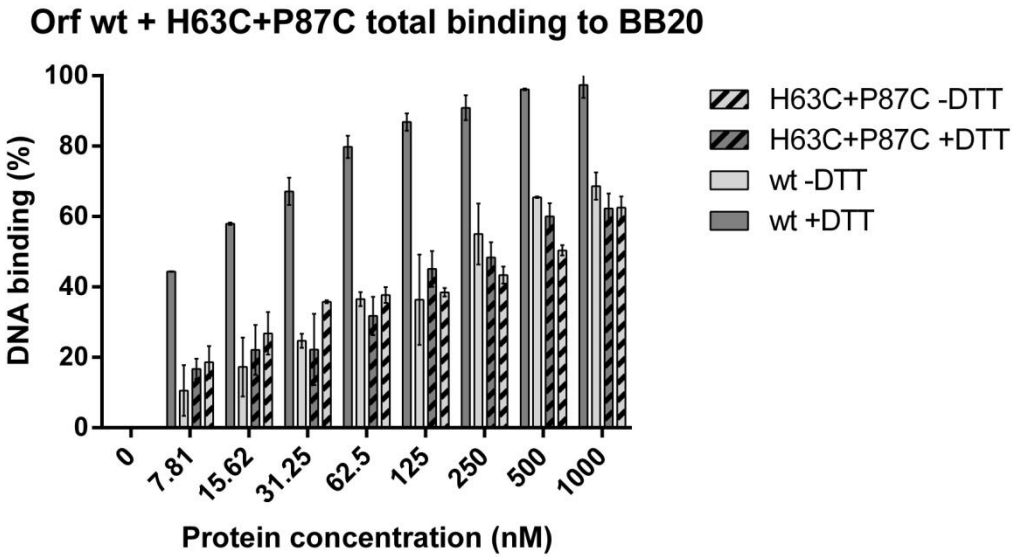

Figure 3.6. Summary of Orf wt and H63C+P87C DNA binding to bubble DNA. Binding mixtures contained 0.15 nM ^{32}P-labbelled DNA (Appendix A) with and without DTT detected by autoradiography, error bars represent standard deviation of n=3

3.4 Discussion

The significant loss of DNA binding found with the W50A mutant confirms that the central channel has at least some involvement with forming DNA-protein complexes. The tryptophan residues from both subunits of the dimer are exposed on both sides of the channel (Figure 3.1.A) suggesting that DNA must enter the channel to some degree. However, despite being close to Trp50, the Lys48 residue, which is situated outside of the channel on the outer face of the subunit, also shows a major reduction in DNA binding. Therefore the DNA must in some way associate with the outside of the protein dimer as well as the central channel, in this case DNA may be wrapped around the protein and into the cavity to reach the Trp50 and Lys48 residues on each subunit. Surprisingly the Q45A mutant does not show a similar loss in DNA binding despite the residue being situated between Trp50 and Lys48 on the rim of the channel and one of the more highly conserved residues in the Orf family. While Gln45 is non-essential for DNA binding the Val106 residue which sits next to Gln45 at the rim of the central cavity is shown to be of great importance to stabilising protein-DNA interactions as the V106E mutant shows a dramatic loss of binding with only a slight smear of unstable complexes detected. However, the V106E mutant was found to aggregate in solution suggesting that the mutation affects protein folding which could prevent assembly of stable dimers needed for DNA binding [178]. The R103E mutant, which also lies on

the lip of the channel showed a complete loss of DNA binding and as this protein was able to fold properly under normal conditions, may well be key in forming complexes with DNA targets. While Arg103 is located more along the cleft structure along the face of the dimer than the channel itself, the mutant could contribute to stabilising binding to DNA within the channel. Both the R103E and V106E mutants are more significant changes than the alanine substitutions present in the other mutants, introducing a negatively charged glutamic acid. Since DNA binding regions are often positively charged to allow association with the negatively charged phosphodiester backbone, these introduced residues could be repelling DNA from binding to cleft or within the cavity. With the exception of Q45A, the further away from the cleft at the channel entrance the mutation resides, the smaller the loss of DNA binding activity is observed. This suggests that DNA is indeed binding through the cleft as opposed to through the channel as would be expected with the obvious spatial constraints of the channel width. However, the loss of binding found with the W50A mutant suggests that DNA is associating with the channel to some extent, potentially nestled into the cavity while bound to the cleft to increase stability, further endorsing the proposal of a hinge action to allow DNA to enter the central cavity of the dimer.

The mutants made to investigate a clamp-like mechanism of Orf binding did not behave entirely as expected. DTT should have only had an effect on the H63C+P87C mutant, while the C52T mutant was expected to behave in the same manner as the wt. Overall ssDNA binding was similar between the wt and the H63C+P87C mutant, however, the addition of the extra cysteine residues did stabilise formation of a second protein-DNA complex with the mutant in either the presence of absence of DTT. This suggests that the introduced cysteines do affect assembly of multiple Orf dimers onto the same DNA strand possibly through disulphide bridge formation, without affecting initial protein-DNA interactions of single Orf dimers. The C52T mutant showed some deficiencies in DNA binding, particularly in its inability to form two complexes when DTT was present, suggesting that Cys52 is important for inter-dimer contacts on DNA. While C52T does not form second complexes as readily as the wt without DTT, when DTT is present the protein both initially struggles to bind DNA at lower concentrations and fails to form any second complex at any tested protein concentration. The loss in overall binding with DTT is not a trend shown with either the wt or H63C+P87C which may be a case of DTT interfering with overall protein folding or stability.

Contrary to the behaviour with ssDNA, DTT significantly improves the binding of wt to the bubble structure whereas the H63C+P87C mutant showed similar binding to BB_{20} irrespective of the presence of DTT. The attempt to lock the protein structure closed or open with the additional disulphide bridge does appear to eliminate the enhanced bubble DNA binding evident with the wt. While this does not follow the Orf model, which would expect the clamp mechanism to be crucial to stabilising the DNA-protein complexes, the differences in binding observed with ssDNA and bubble DNA suggests that Orf may bind differently to the two substrates. Indeed, a potential closed structure of the Orf dimer without DTT confers significantly improved binding to ssDNA with the H63C+P87C mutant, outperforming both the wt without DTT and itself with DTT. Furthermore, the highest affinity for ssDNA was observed with the C52T mutant, which if Cys52 is potentially stabilising the protein to an open clamp structure, despite not being directly involved, would compound the notion that ssDNA, unlike the larger bubble structure, needs to be clamped in place for binding.

Chapter 4. NinH

Efforts to identify the function of uncharacterized open reading frames in the λ genome revealed that the product of the *ninH* gene, located immediately downstream of *rap* (*ninG*) shares significant homology with the DNA binding region of *E. coli* Fis (Factor for Inversion Stimulation; Figure 4.2.A)[180]. The N-terminus of the 68 residue NinH protein appears to have a helix-turn-helix (Figure 4.2.B) closely related to that present at the C-terminal end of Fis. The region of homology between the two proteins encompasses the three-helix bundle that constitutes the DNA binding module of Fis (Figure 4.2.A). Remarkably, NinH possesses nearly all of the residues known to be involved in contacting DNA as determined in Fis-DNA crystal structures [181]; the single exception is Asn84, which is substituted with a serine in Fis (Figure 4.2.A and C). This residue in Fis contacts the backbone and an adenine base, but can adopt alternative conformations depending on the DNA sequence and is not essential for binding to high-affinity sites [182-185]. The bioinformatic analysis therefore indicated that NinH may be a phage counterpart to the Fis protein.

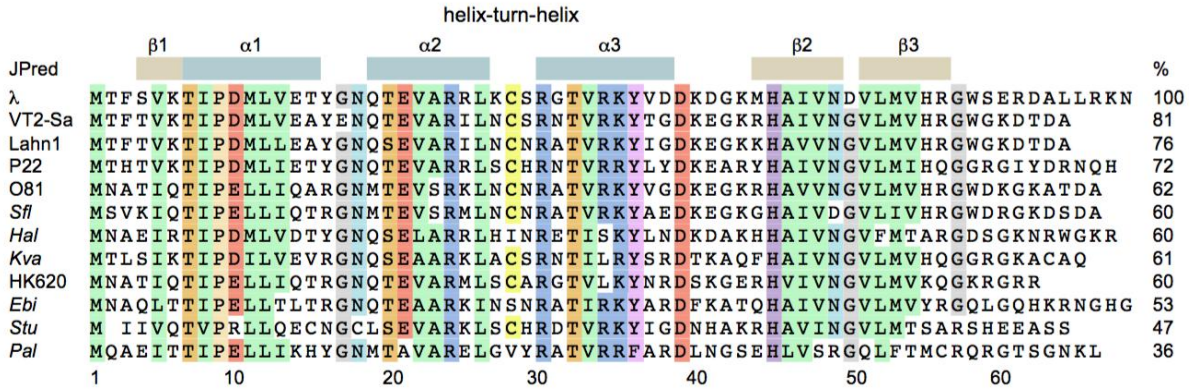

Figure 4.1. Sequence alignment of representative NinH proteins. The NinH protein from phage λ (P03771) was aligned with homologous representatives from *E. coli* O157:H7 phage VT2-Sakai (P69177; VT2-Sa), *E. coli* O84:H$^-$ phage Lahn1 (Q777W6; Lahn1), *Salmonella* phage P22 (Q38669), *E. coli* O81 ED1a (O81; B7MPX0), *Shigella flexneri* 2a prophage (A0A0C7N204), *Hafnia alvei* (*Hal*; G9Y3Q3), *Klebsiella variicola* At-22 (*Kva*; D3R9E5), enterobacteria phage HK620 (Q9AZ09), *Erwinia billingiae* Eb661 (*Ebi*; D8MV69), *Siccibacter turicensis* DSM 18703 (*Stu*: C9XYF4) and *Providencia alcalifaciens* DSM 30120 (*Pal*; B6XGT3). The percentage identity with λ NinH is shown on the right of each sequence. Conserved residues are highlighted. Secondary elements derived from the JPred4 server and the predicted helix-turn-helix motif, encompassing α2 and α3, are indicated above the alignment [186].

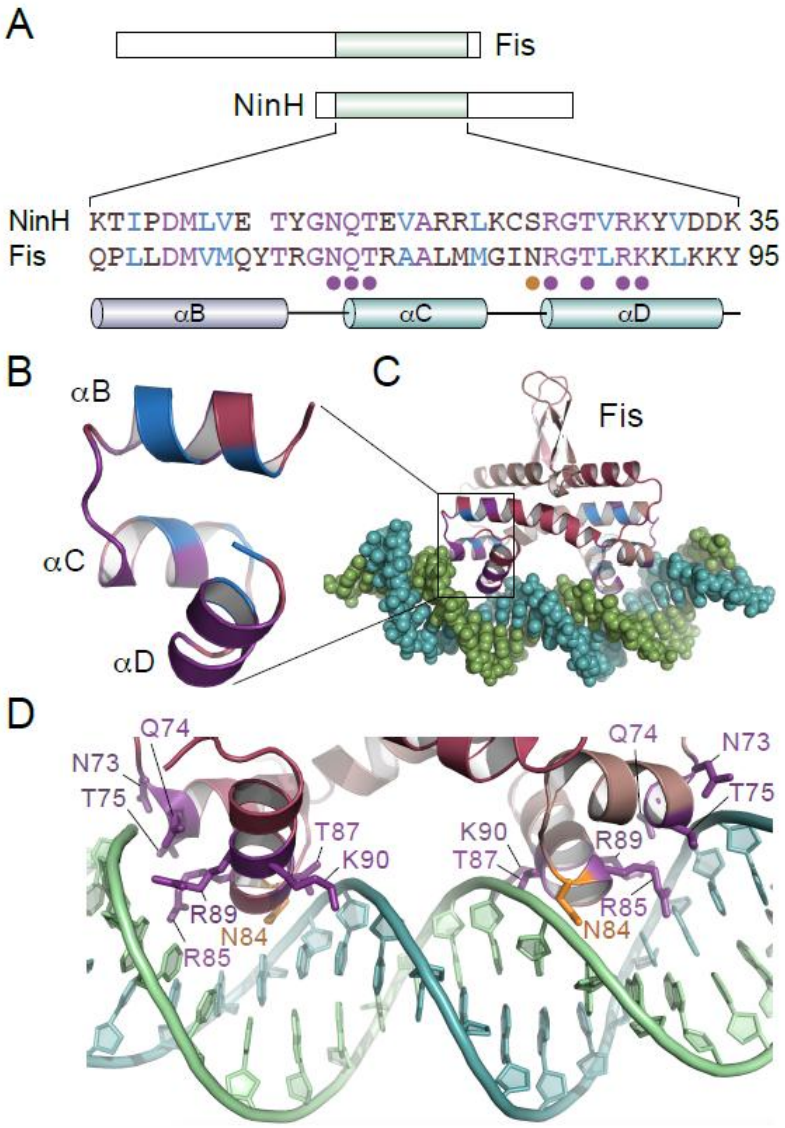

Figure 4.2. NinH homology to Fis protein. A. Sequence homology between NinH and Fis. The region of similarity spans αB and the helix-turn-helix (αC and αD) binding motif of Fis. Residues involved in DNA contacts in Fis are indicated by a filled circle below the Fis sequence. **B.** Model of NinH generated by Phyre2 based on the crystal structure of Fis. **C.** Crystal structure of the Fis-DNA complex showing the region homologous to NinH. Matching residues are coloured as in part **A**. **D. R**esidues in Fis that are known to contact DNA. With the exception of N84, all of these are present in NinH (G.J. Sharples, unpublished results).

4.1 Nucleoid-associated proteins

Fis is a member of a family of proteins called nucleoid-associated proteins (NAPs) involved in chromosome packaging within the cytoplasm of bacterial cells to form the nucleoid [187, 188]. NAPs are also responsible for unwinding the supercoiled nucleoid structure [189] so that DNA can be exposed for transcription and assembly of transcription factors within the superstructure [190-193]. NAPs influence the conformational shape of the chromosome by bending and unbending the DNA depending on the requirements of the cell [188, 190-193]. The nucleoid superstructure is rearranged in *Escherichia coli*, for example, when the cell enters the exponential

growth phase when a different set of genes needs to be upregulated[188]. Folding of the bacterial chromosome into the nucleoid structure has been studied since the 1970s, however, detailed analysis of NAP provision within an *E. coli* cell was not undertaken until 1999 [194, 195]. With the development of improved techniques for separating and identifying proteins, it has become easier to visualise the dynamics of protein action and interaction within the nucleoid [196]. Without the coordinated action of multiple NAPs within the cell, bacteria would not be able to function properly, adversely affecting DNA supercoiling and the expression of vital enzymes, including virulence factors[197] Along with Fis, three other NAPs (IHF, H-NS and StpA) are important in nucleoid formation.

Integration Host Factor (IHF) is a transcription factor that binds to a large array of DNA targets in the regulatory regions of *E. coli* [198] and is essential for both efficient site-specific integration and excision of λ [199, 200]. IHF is known to affect genome organization [201] as well as transcription [202], transposition [203] and site specific recombination [204]. IHF introduces a sharp bend (>160°) into its substrate DNA to allow other NAPs to act on the target site as co-transcription factors [205]. Protein-DNA crystal structures reveal that IHF forms a small heterodimer of highly homologous IHFα and IHFβ subunits [206].

H-NS was defined by Lammi et al. in 1984 [207]. H-NS has a preference for binding already curved DNA as well as being capable of bending DNA once it has been bound [208]. Like IHF, H-NS acts as a global regulator of transcription, affecting the expression of m**a**ny gene targets [209] in addition to maintaining genome stability [210]. In solution H-NS forms a dimer and fully coats its target DNA to repress gene activation from other transcription factors [211-213].

StpA is a paralogue of H-NS known to be capable of forming heterodimers between the two proteins [214] and in many cases share the same DNA targets in *E. coli* [215]. While deletion of the *stpA* gene has minimal affects on cell H-NS activity, the activity of StpA when H-NS has been mutated is severely reduced, suggesting that StpA mostly acts as a heterodimer with H-NS [188].

4.2 Fis

Fis is one of the NAPs that is closely associated with cell growth as its levels rise markedly as bacteria enter the exponential phase [216]. A key role of Fis is to activate Gin and Hin, a pair of site-specific recombinase proteins which act to invert sections of DNA giving it the name Factor for Inversion Stimulation, Fis [181]. The *fis* operon is autoregulated by Fis repression but stimulated by unwinding of supercoiled DNA as the cell prepares for large scale transcription in exponential growth, hence the high levels of Fis protein during cell proliferation [217]. Crystal structures of Fis protein reveal that the αC and αD helices at the C-terminus of each 98-residue subunit fold into a helix-turn-helix domain [218-221]. When bound to DNA each HTH domain is bound to the phosphodiester backbone of DNA along the major grove through nine contact points with the side chains of Asn73, Thr75 and Arg89 further anchoring the complex [181].

Once bound to DNA, Fis bends the DNA to compact and condense extended sections of the strand by forming stable DNA loop structures, folding the DNA back on itself [222]. Despite Fis having a non-discriminate

affinity for binding onto any section of DNA it is only seen to form stable complexes with regions of DNA which contain regulatory genes [191], implying Fis has a global effect on gene regulation. In particular Fis has been shown to control transcription when the cell is under exponential growth. This is in part achieved through the repressive actions of Fis on the *gyrA* and *gyrB* genes which code for the subunits of DNA gyrase [223, 224] and also of the *topA* gene which encodes topoisomerase I [225]. By regulating these genes Fis can control if and when DNA supercoiling can be undone for mRNA access for transcription and maintaining the supercoiled structure that it has folded.

4.3 NinH protein

With what we know of how Fis functions, it should be possible to evaluate whether the phage λ NinH protein behaves in a similar way, perhaps to replace Fis activity in phage transcription, genome packaging or site-specific recombination. In this chapter, the capacity of NinH to bind DNA will be investigated using a number of different DNA substrates. In addition the effect of plasmids carrying NinH will be investigated in *E. coli* mutants lacking key NAPs, namely Fis, IHF α, IHF β, H-NS and StpA.

4.4 Binding of NinH to bent DNA structures

NinH protein was examined for its ability to bind four DNA substrates of 60 nucleotides in length: single-stranded, double-stranded and two bent DNA duplexes. The bent structures were formed by addition of extra adenine residues at the centre of one strand of the duplex, causing it to bend to accommodate the extra nucleotide. BY1 contained a single additional adenine, while BY3 contained a group of three adenines (Appendix A). Adenine insertions of 1-3 nucleotides are known to introduce a kink of 50-70^{o} in duplex DNA [226, 227]. Increasing concentrations of NinH were added to these substrates on ice (as described in 2.6.2) and binding monitored by gel shift assays with results analysed using ImageJ (as described in 2.7.1).

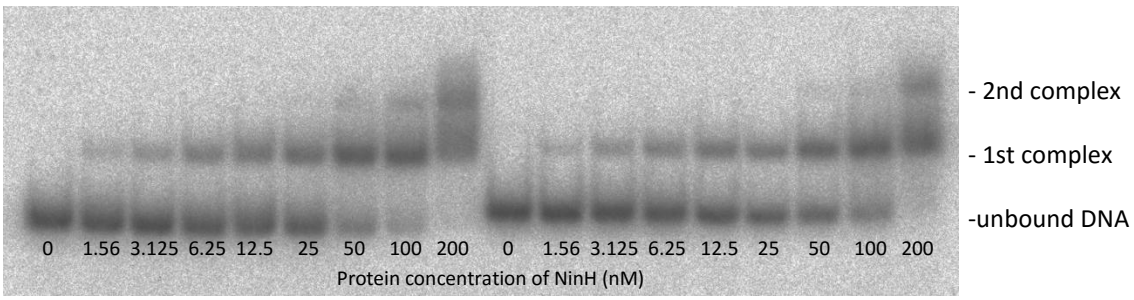

Figure 4.3. Representative NinH binding gel. NinH (1.52625-200 nM) binding to 0.15 nM ^{32}P-BY3. Complexes 1 and 2 are indicated.

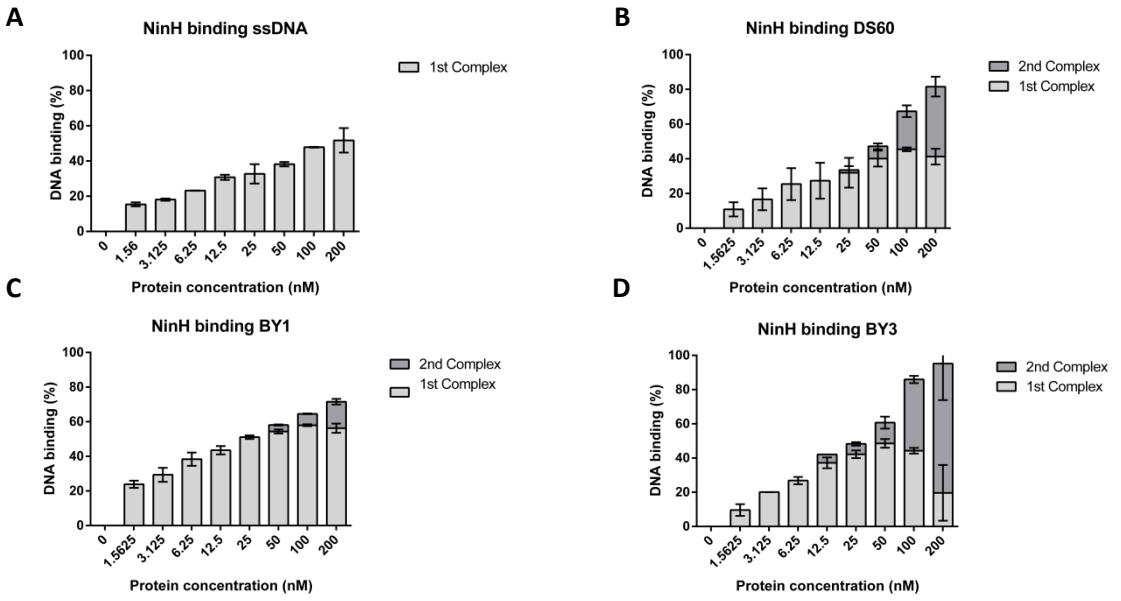

Figure 4.4. NinH binding to various DNA substrates in gel shift assays. Binding mixtures contained 0.15 nM ^{32}P-labbelled DNA (Appendix A) detected by autoradiography; error bars represent standard deviation of n=3.

NinH bound to all four of the DNA structures tested, although it did not bind fully to any of them at concentrations up to 200 nM (Figure 4.4.A-D). NinH displayed the lowest affinity for single stranded DNA, which at the highest protein concentration showed only 51.7% of the DNA bound in a single complex (Figure 4.4.A). The limited binding to single stranded DNA fits with it being a NAP-like protein which should preferentially bind dsDNA. This was confirmed with the improved binding of NinH to double stranded DNA with 81.6% of the DNA bound at 200 nM protein (Figure 4.4.B). Although there appears to be lower binding in double-stranded than single-stranded at lower concentrations the results are not statistically dissimilar between the two data sets with the error bars overlapping. While there is little difference in binding affinity at lower concentrations, the higher concentrations show drastically different results with the two substrates (Figure 4.4. A and B). This may in part reflect the emergence of a second protein-DNA complex observed from 25-200 nM NinH. The complexes probably represent the assembly of one and then two dimers of NinH protein.

Since NinH resembles NAP proteins such as Fis, we also examined its binding to DNA that already contains a bend to assess whether this would enhance binding. NinH showed a linear binding curve that appears to show a greater binding affinity than that seen with dsDNA (Figure 4.4.B and C). However, there is in fact little difference between the total DNA bound in both experiments. While the average amount of DNA bound is higher with BY1 than dsDNA up to 100 nM, it is only significantly higher between 1.5625 and 25 nM with all other lanes possessing overlapping standard deviation error bars. Binding is slightly reduced with NinH on BY1 compared to dsDNA at higher protein concentration, possibly due to its reduced ability to form a second

complex on BY3. NinH bound 1.6% of the dsDNA into a second complex at 25 nM protein whereas a second complex was not observed until 50 nM NinH with BY1. The contrast between the similarity in total binding and the difference in second complex formation between dsDNA and BY1 suggests that the bend introduced into the double helix interferes with the assembly of additional NinH protomers to the complex once the first has bound. This could be due to a reduction in the length of DNA available for binding or steric hindrance between subunits assembling on the DNA.

NinH was also tested on the BY3 substrate, containing a 3 nucleotide insertion, and showed better binding to this substrate than dsDNA or BY1 (Figure 4.4.D). NinH bound 95% of the DNA at 200 nM protein and also showed greater binding to BY3 at 12.5-100 nM than that of dsDNA (Figure 4.4.B, C and D). However, NinH bound less well to BY3 than to BY1 at lower protein concentrations. The major difference in binding may again be influenced by the ability to form two protein-DNA complexes. NinH with BY3 showed the formation of a second complex from as low a protein concentration as 12.5 nM and formed a significantly greater amount of this complex than dsDNA and BY1 at every protein concentration except for 50 nM where the standard deviations between BY3 and dsDNA overlap. Hence the bent DNA structure containing a single adenine insertion (BY1 - Appendix A) exhibits reduced second complex formation with NinH compared with linear dsDNA, whereas a DNA substrate with a more significant bend resulting from insertion of three adenines (BY3 - Appendix A) shows a much greater capacity to allow assembly of two NinH complexes.

4.5 Effects of plasmids carrying the *ninH* gene on selected NAP mutants

To investigate whether NinH could act to replace host function in *E. coli* missing various NAPs we examined the effect of introducing NinH on growth of a selection of NAP mutants obtained from the Keio collection [176]. A wild type *E.coli* strain alongside five isogenic mutant strains lacking a single NAP (IhfA, IhfB, Fis, H-NS and StpA) were transformed with pFC109 carrying the *ninH* gene in the expression vector pT7-7, although the T7 promoter is inoperative in these strains. As a negative control each strain was also transformed with pT7-7 to confirm that any effects observed were due to the presence of *ninH* and not the plasmid vector itself. Growth of the twelve transformed strains was monitored by measuring the absorbance at A_{650nm} (as described in 2.7.2).

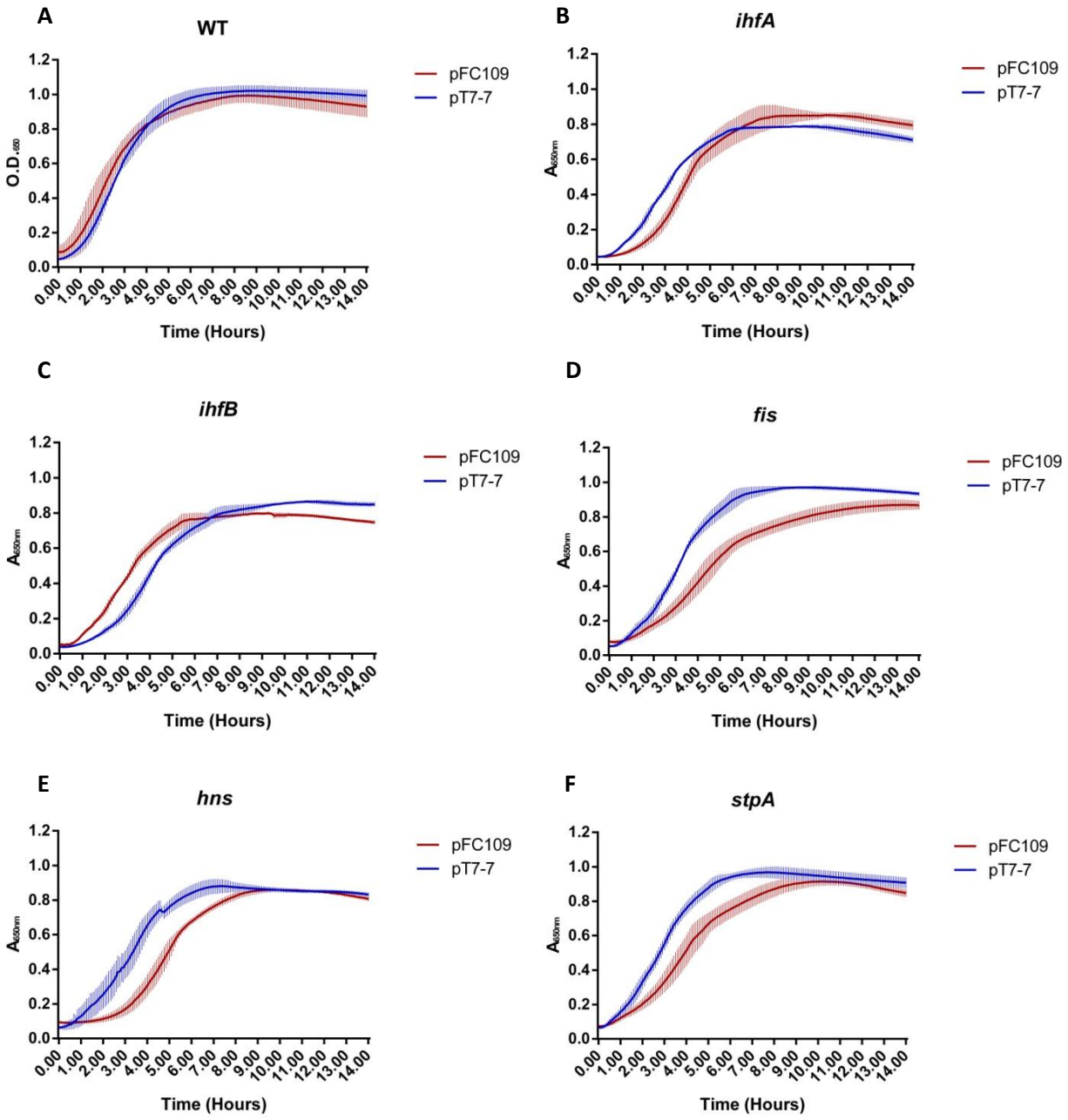

Figure 4.5. Growth of NAP mutants carrying NinH. Cultures were grown in LB media supplemented with ampicillin. Readings were taken every 5 minutes, results are an average of three independent experiments with standard deviation indicated by error bars.

The wild type strain shows that inclusion of the *ninH* gene had little impact on bacterial growth, although it did reach stationary phase slightly earlier than the control (Figure 4.5.A). Growth of the wild type strain carrying pT7-7 peaked after 9 hours 20 minutes at A_{650nm} 1.02 while the wild type carrying NinH peaked after 8 hours 40 minutes at A_{650nm} 0.99. The strain containing pFC109 started at a slightly higher optical density, however, both strains appeared to enter exponential phase at a similar time and grew at the same rate during this period. Hence it seems safe to conclude that there are no major differences in growth and that NinH has no detrimental effects in a strain that is wild-type for all of the NAPs tested here.

	pT7-7		pFC109	
	mean	S.D.	mean	S.D.
wt	56.7	10.4	51.7	5.8
ihfA	78.3	2.9	71.7	24.7
ihfB	76.7	12.6	81.7	11.5
fis	71.7	7.6	107.7	5.8
hns	87.7	7.6	80	5
stpA	75	8.7	93.3	5.8

Figure 4.6. Doubling times of NAP mutants carrying NinH. Time taken in minutes for absorbance reading to double from 0.25 to 0.5 A_{650nm}. The standard deviation (S.D.) based on three independent experiments is shown.

Analysis of strains lacking different NAPs showed that all of the mutants exhibited a reduced growth with none of them reaching as high an A_{650nm} as the wild type with or without NinH (Figure 4.5.A-F). When the cell was deficient in IhfA, one of the subunits of integration host factor, the growth curve entered stationary phase at A_{650nm} 0.78 in cells carrying the vector, well below that of the wild type and also below the 0.85 observed when the NinH plasmid was present (Figure 4.5.B). While *ihfA* carrying pT7-7 showed a peak of growth at 4 hours 30 minutes before the same strain carrying pFC109 in addition to entering exponential phase approximately an hour earlier, the doubling times between the two strains were very similar (Figure 4.6.).Contrary to the earlier exponential growth of *ihfA* with pT7-7 the opposite effect was seen with strains missing IhfA's partner subunit, IhfB. After 11 hours the *ihfB* mutant pT7-7 peaked at an A_{650nm} of 0.86, 2 hours after the strain containing the NinH gene peaks at 0.80 (Figure 4.5.C). This suggests that NinH can help promote growth when the β subunit of the host IHF NAP is deleted (*ihfB*), but has a detrimental effect when the cell is missing the α subunit (*ihfA*). However, once again the doubling times while under exponential growth remain similar between both strains.

The two strains which did show significantly different doubling times were the *fis* and *stpA* mutants for which the doubling times with pFC109 increased by an average of 36 and 18.3 minutes respectively (Figure 4.6.). Both *fis* and *stpA* mutants showed similar growth curves with pFC109 and pT7-7 entering log phase growth at the same time, however, the strains carrying the vector grew at a higher rate and reached stationary phase earlier (4 hours earlier in *fis*, 2 hours in *stpA*) at a higher maximum absorbance (Figure 4.5.D and F). These results differ from the two *ihf* mutants with expression of NinH clearly conferring a negative impact on growth rate. It is interesting that NinH, despite its homology to Fis does not appear to be able to improve the growth of the *fis* mutant, instead it results in a further reduction in growth.

The H-NS defective strain showed a result more similar to that of *ihfA* where the two growth curves of the vector and NinH give a similar growth rate but the strain carrying pFC109 shows a significantly delayed entry into exponential phase (Figure 4.5.B and E). The *hns* mutant containing the vector alone entered log growth within the first 30 minutes of incubation, whereas the culture carrying the NinH construct did not enter

logarithmic growth until after 3 hours of incubation. What separates *hns* from the other results is that the *hns* carrying pT7-7 and pFC109 reach a similar absorbance before entering stationary phase, peaking at 0.88 (pT7-7) and 0.86 (pFC109) (Figure 4.5.E). The major effect of NinH on cells deficient in H-NS is to delay entry into the exponential phase of growth.

To further analyse the effect of NinH on these mutants compared to the effect wild type, the difference in growth between the mutants containing the vector (pT7-7) or NinH (pFC109) was compared to the difference between these plasmids in the wild type as described in (link growth difference explanation in methods). This gives a graph showing the growth rate increase or decrease conferred by NinH for each mutant compared to the corresponding growth rate for that time point in wild type cells.

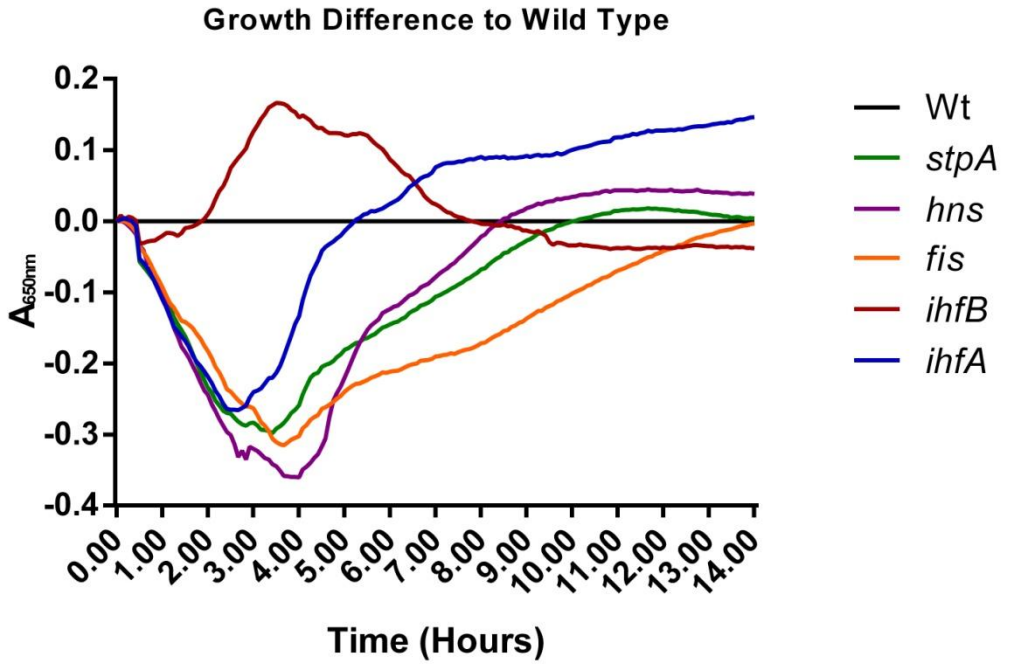

Figure 4.7. Difference in growth conferred by *ninH* plasmid in mutants compared to that in wt. The baseline value for the wild-type corresponds to the difference in absorbance between pFC109 and pT7-7 at each time point. Similarly each mutant pT7-7 was minused from mutant pFC109 and the difference in growth the NinH containing plasmid plotted against that difference in the wt (as described in 2.7.3). Readings were taken every 5 minutes.

Growth in the exponential phase of the *ihfB* is substantially improved when NinH is present compared to the difference between pT7-7 and pFC109 in the wild type (Figure 4.7). The graph also highlights the effect of NinH on delaying the entry into log phase of the other four mutant strains. Exponential growth occurred at approximately 1-8 hours which can be observed as the lines split at 30-60 minutes before rejoining close to the baseline after about 8 hours.

While the growth of *ihfA* appears slower with pFC109 than the vector control (Figure 4.5.B), the rate of log phase growth from 2.5 hours with pFC109 over pT7-7 was faster than of that in the wt (Figure 4.7). Along with the doubling time of *ihfA* being lower with pFC109 than the with the vector control (Figure 4.6) it is shown that, while entry into log growth is delayed by the plasmid for NinH, the bacteria recover their growth rate quickly, a trait not shared by all the mutants.

Indeed the *fis* mutant takes the longest time to reach wt growth levels after delayed entry to the exponential phase, only just returning to the baseline at the end of the 14 hour incubation. Although this mutant starts to return to the baseline after its peak at 3 hours 45 minutes this coincides with the peaks of *hns*, *stpA*, and on the other side of the baseline, *ihfB*. This is a result of the wild type strain coming towards the end of its log phase where growth begins slowing down around 4 hours once it has reached an A_{650nm} of 0.8 (Figure 4.5.A). As such the only mutant which shows an improved log phase with the *ninH* plasmid is *ihfA*. While *ihfB* differs from the others by entering log phase more quickly, the growth difference returns to baseline at the same time as the other mutants showing that the plasmid does not confer the enhanced growth seen with *ihfA* (Figure 4.5.B and C).

4.6 Discussion

Given its similarity to Fis, NinH bound to the different DNA substrates in the expected manner. The limited association with ssDNA was anticipated as the primary function of NAPs involves binding and manipulating dsDNA for compaction into the nucleoid. NinH displayed much better binding to dsDNA with more than one subunit assembling on linear duplex substrates. In addition, binding of NinH to bent DNA was in keeping with Fis action as a topological regulator on of supercoiled DNA [228].

Surprisingly, while DNA binding with the BY1 substrate was better than that with dsDNA, the formation of a second complex was considerably reduced on the bent DNA structure. As there was no reduction in total protein binding to DNA between these substrates, the decrease in second complex formation with BY1 implies that the single additional base in one strand of BY1 interrupts the assembly of additional NinH subunits. NinH, presumably as a dimeric species as with Fis may assemble alongside each other on linear DNA with the bend restricting binding of the second dimer. A second complex is also not formed by NinH on ssDNA which could be due to flexibility in the mobility of the backbone. Interestingly the more significant bend conferred by addition of 3 nucleotides, BY3, results in the largest amount of second complex formation, considerably more than that seen with linear duplex DNA. This fits with the similarity between NinH and Fis, since the latter would have been expected to show greater binding to DNA structures with more highly bent or supercoiled DNA. Nucleoid DNA has an irregular shape as the double helix is folded upon itself may times. Fis maintains this structure by holding the bent DNA in place, allowing protein-protein interactions to stabilise the assembly [222]. It may be that the precise position of the minor groove in each of these substrates dictates whether a second dimer can assemble [181]. The BY3 structure, having a greater bend angle than that of BY1, may be more

similar in topology to supercoiled DNA and therefore closer to the natural target of NinH explaining its high affinity for BY3. There may be certain disadvantages to Fis/NinH having string binding for a slightly bent DNA structure. Before cell division the nucleoid DNA needs to unwind from its supercoiled package in order to divide. Cells defective in the *fis* gene observed under a microscope with live/dead bacterial staining (experiment number (Data not shown) show reduced viability with elongated cells that fail to undergo normal cell division. Recent microscopy analysis has confirmed that the presence of NinH exacerbates this effect with most cells showing filamentation in *ihfB* and *fis* mutants (GJ Sharples, personal communication). The DNA superstructure needs to be switched efficiently in an analogous fashion from uncoiled to supercoiled and vice versa. Binding affinity for an intermediate step on slightly bent DNA could impede the coiling and uncoiling processes, thus interfering with replication.

In vivo, NinH interacts more with mutations in the Ihf subunits than with its homologous partner, Fis, in particular in negating the loss of the IhfB subunit. It may be that NinH can help DNA bending alongside the IhfA subunit, as this subunit can function independently of IhfB [229]. It seems unlikely that NinH and IhfA could form a functionally similar heterodimer. The growth rates do not differ significantly with NinH in either *ihfA* or *ihfB* mutants and the improvement upon *ihfB* does not reach that of wild type *E. coli*. NinH is clearly unable to support IhfB in the absence of IhfA, since the *ihfA* mutant shows reduced growth compared to the vector control. It is not clear why the presence of NinH further inhibits the growth of *ihfA* when there is no effect on the initiation of log phase in the wild type strain. NinH binding to DNA in the absence of this nucleoid protein must somehow negatively impact on proper nucleoid formation, delaying entrance into exponential growth.

It is also unclear why NinH causes further delays to cell growth rate in the *fis* and *stpA* mutants in addition to the delayed log phase entry exhibited by all of the mutants, except *ihfB*, while the plasmid has little affect on the wt. Since StpA forms a heterodimer with HN-S [214], one would have expected a similar effect on the *hns* mutant. NinH could also have an effect on global gene regulation; Fis autoregulates its own expression in order to allow DNA unwinding during cell proliferation but is switched off entirely during stationary phase [230]. The effects of NinH in causing cell filamentation do suggest that there are problems with modulating the nucleoid in preparation for DNA replication and entry into phases of rapid growth and cell division.

Chapter 5. Rap

5.1 Mutation of the HNH nuclease domain of Rap

Six point mutations were made in highly conserved residues within the predicted catalytic site of Rap protein. This region of Rap shows characteristic features of the HNH endonuclease family [231] including a cluster of cysteines involved in coordinating zinc binding and acidic and polar residues that bind magnesium or manganese for catalysis. The diagram below highlights the six residues in Rap that were targeted for mutagenesis and why these were selected based on their high degree of conservation among Rap (NinG) family proteins (Figure 5.1).

Six mutants in the *rap* gene were made in the N-His-Rap overespression construct using site-directed mutagenesis (PD Townsend, unpublished results). The mutants were selected to confirm the importance of highly conserved residues from the Pfam entry for NinG (PF05766; Figure 5.1). One of the cysteines from the zinc finger motif was replaced by a serine (C101S) to examine the importance of zinc binding for binding and cleavage of branched DNA. Conserved polar residues that may contribute to phosphodiester backbone cleavage were also targeted (D113A, H116T, R127A, N132A, N141D). Several of these residues match conserved catalytic residues found in T4 endonuclease VII [232] and *Geobacter metallireducens* Gmet_0936 [233]. The following residues were chosen to match known mutants from T4 endonuclease VII.

C101S corresponds to C23S in T4 endonuclease VII [234] with the sulphur in cysteine replaced by the oxygen group in serine. The C23S mutant is known to be defective in zinc binding and fails to bind or resolve Holliday junctions. This mutation in Rap will hopefully give us similar information indicating that the zinc finger domain plays a structural role. D113A corresponds to the H38 mutants in T4 endo VII [235]. The acidic charge of aspartic acid is replaced with the small hydrophobic group in alanine. This residue may be involved in binding magnesium or manganese for junction resolution. However in T4 endonuclease VII this residue appears to point away from the active site. However, it is likely that Rap differs here as it lacks an equivalent of D40 in T4 endo VII. In Rap sequences, residue 113 is usually either an aspartic acid or histidine (Figure 5.1). H116T matches H41T mutants of T4 endo VII [235], whereby the histidine ring is replaced with the smaller polar threonine residue. H41T mutants cannot cleave junctions but retain some DNA binding activity. This residue is conserved in all the HNH family nucleases [236]. This is likely to be an active site mutant involved in binding magnesium or manganese. N141D is the equivalent of the T4 endonuclease VII N62D mutant [232], Aspartic acid replacing asparagines would not necessarily block metal ion binding but it does prevent resolution in T4 endonuclease VII. This residue is never an aspartic acid in the HNH family [236]. This is likely to be an active site mutant involved in binding magnesium or manganese.

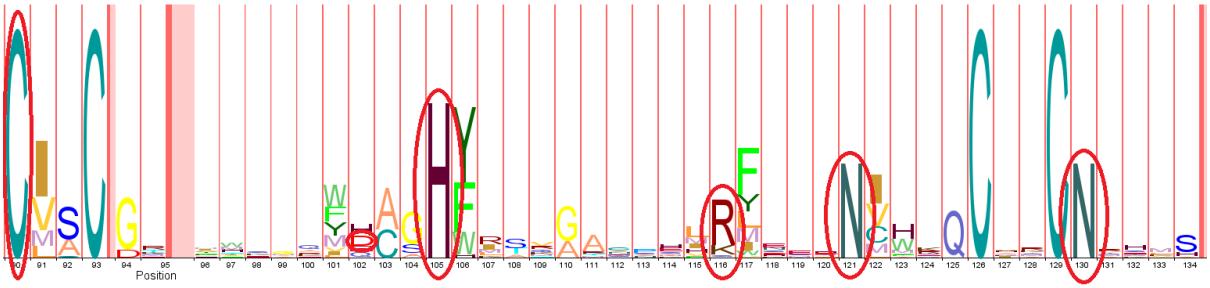

Figure 5.1. Conserved residues in Rap selected for mutagenesis. The HMM logo was taken from Pfam [237] entry PF05766 based on 590 sequences from the NinG/Rap family. The letter height denotes the extent of sequence conservation of each residue. The four cysteines comprise the zinc finger domain. The histidine, arginine, glutamate and asparagine residues may be involved in catalysis. Residues selected for mutation are circled in red.

All of the six Rap mutant proteins were expressed in *E. coli* and purified by nickel affinity and heparin-agarose chromatography (as described in 2.3.4). The purified proteins separated by SDS-PAGE are shown in Figure 5.2.

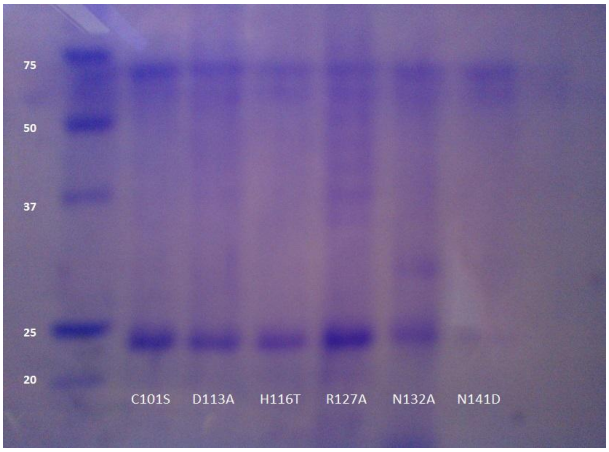

Figure 5.2. Purified Rap mutant proteins. N-His-Rap migrated as a major band at 24 kDa in a 15% SDS polyacrylamide gel. A molecular mass standard (BioRad) is shown in the first lane with band sizes indicated in kDa.

Each of the Rap mutants expressed well with the exception of N141D which expressed poorly and consequently yielded a small amount of purified protein (Figure 5.2). The apparent contaminants between 50 and 75 kDa were probably due to keratin, other contaminants will be a range of proteins from the broken up *E. coli* cells the protein was harvested from that bound to the nickel and heparin columns along with the Rap proteins.

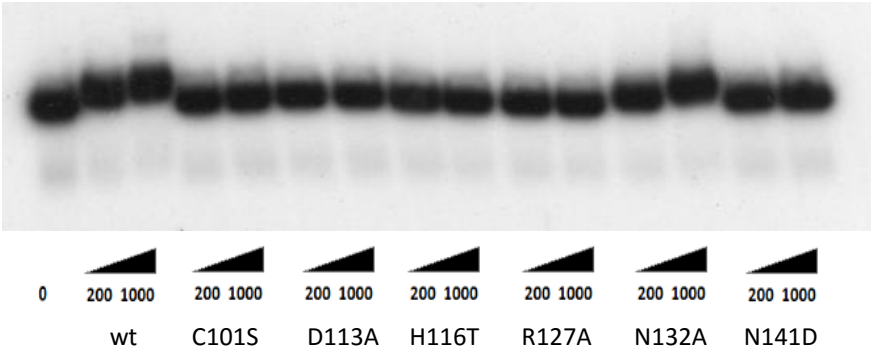

Figure 5.3. Holliday junction DNA binding of Rap mutants. The binding mixture contained 0.3 nM ^{32}P-labelled Holliday junction J11 containing an 11 bp homologous core (Appendix A) with 0, 200 or 1000 nM protein.

The N-His-Rap mutants were tested for their ability to bind a Holliday junction DNA structure, J11, alongside wild type N-His-Rap protein in a gel shift assay (as described in 2.6.2). Wild type Rap produced a small shift in the J11 DNA up the gel (Figure 5.3), although the shift is not as distinct a complex as that seen in previous studies [167, 168]. Most of the mutants showed no band shift relative to the wild type with the exception of N132A, which seemed to show a similar binding trend as the wt Rap, with 200 nM slightly retarded from the travel length of wild type and the other mutants and the 1000 nM band again slightly further up the gel than that. The results suggested that the C101S, D113A, H116T, R127A and N141D mutations abolish the capacity to associate with Holliday junctions. However, the potentially bound DNA bands travel such a similar distance that it is not possible to say with certainty that there is significant binding of the DNA, running the gel for longer may have further separated the bands to give a clearer picture of whether or not the proteins were achieving any meaningful complex formation . The experiment was repeated in a reaction buffer without EDTA and including 1 mM manganese to determine if these Rap mutants had any cleavage activity on the Holliday junction structure (Figure 5.4).

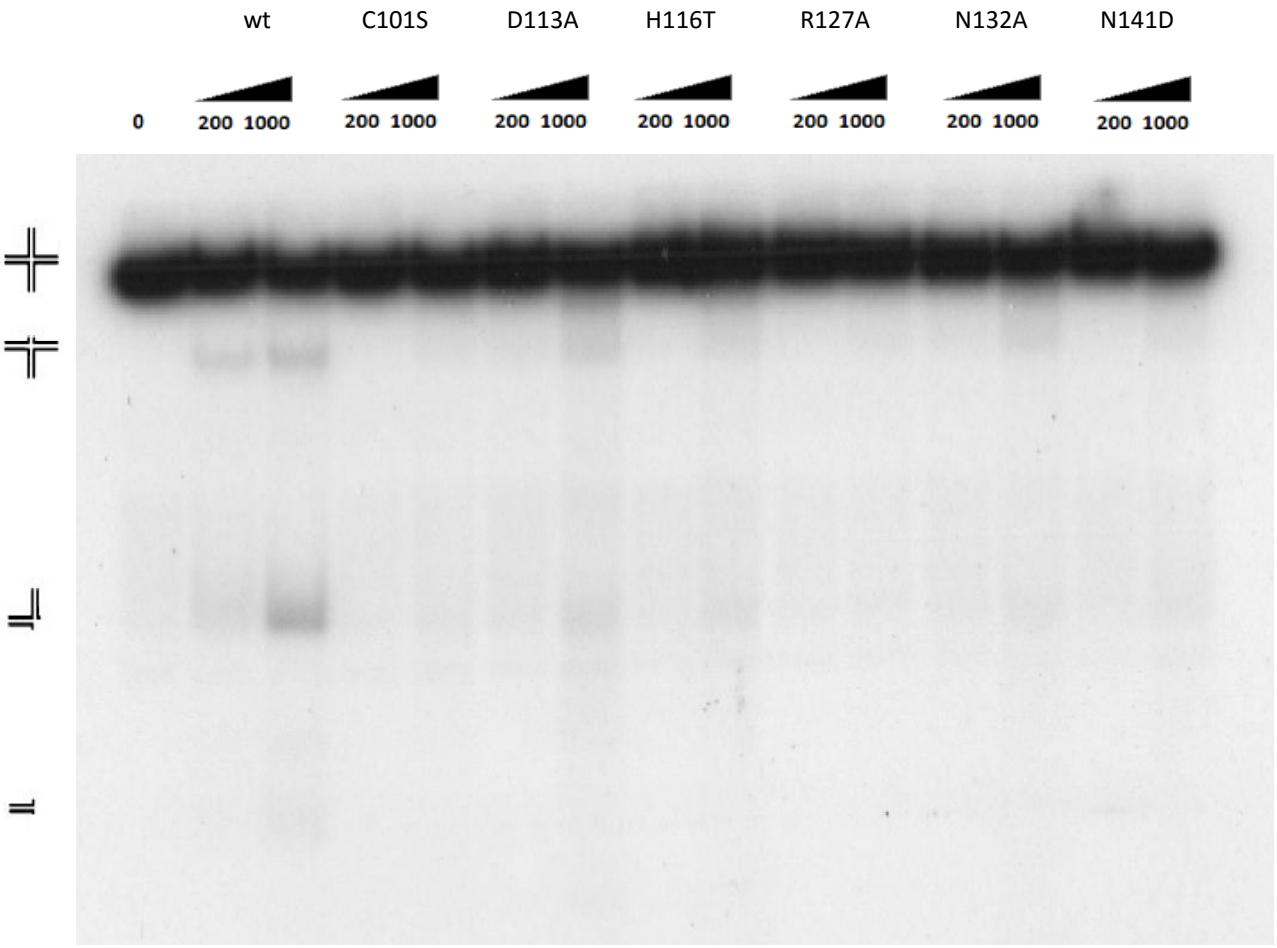

Figure 5.4. Holliday junction DNA cleavage by Rap mutants on J11. Cleavage reactions contained 0.3 nM ^{32}P-labelled J11 Holliday junction DNA (Appendix A) with 0, 200 or 1000 nM protein 1 mM $MnCl_2$.

Holliday junction cleavage by Rap wt generated three distinct products [167, 168], a major nicked duplex band (2-arm), a 3-armed product (3-arm) and a single armed product (1-arm) which is less visible as only one of the 4-arms of the junction is labelled (Figure 5.4) Cleavage by wild-type Rap was lower than expected, however, each of the mutants did exhibit some, albeit reduced, endonuclease activity. Due to the low level of endonuclease activity, the results were not quantified in the same manner as previously employed with Orf and NinH binding. For Figure 5.5 the peak areas from ImageJ (see 2.7.1) were not taken from the total DNA per well as a percentage but were left as relative peak areas; Figure 5.6. A, B and C shows the peak areas for each of the three major breakdown products.

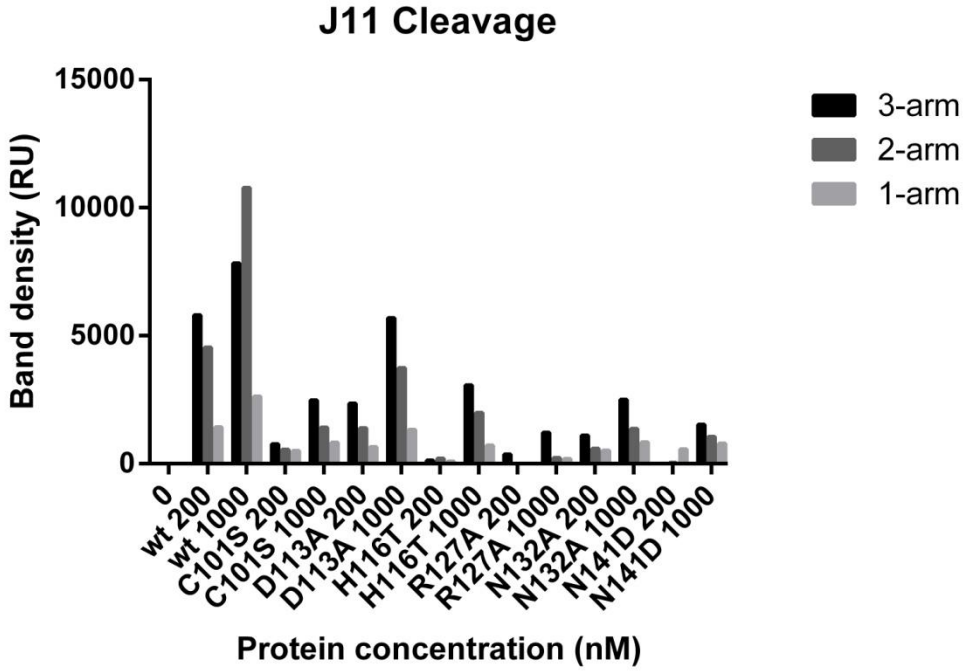

Figure 5.5. Relative density of each cleaved product of the J11 structure by Rap mutant proteins. The relative band density (RU; relative units) from ImageJ analysis of phosphorimager files sorted into the major cleavage products: 3-arm, 2-arm and 1-arm according to their migration through the polyacrylamide gel.

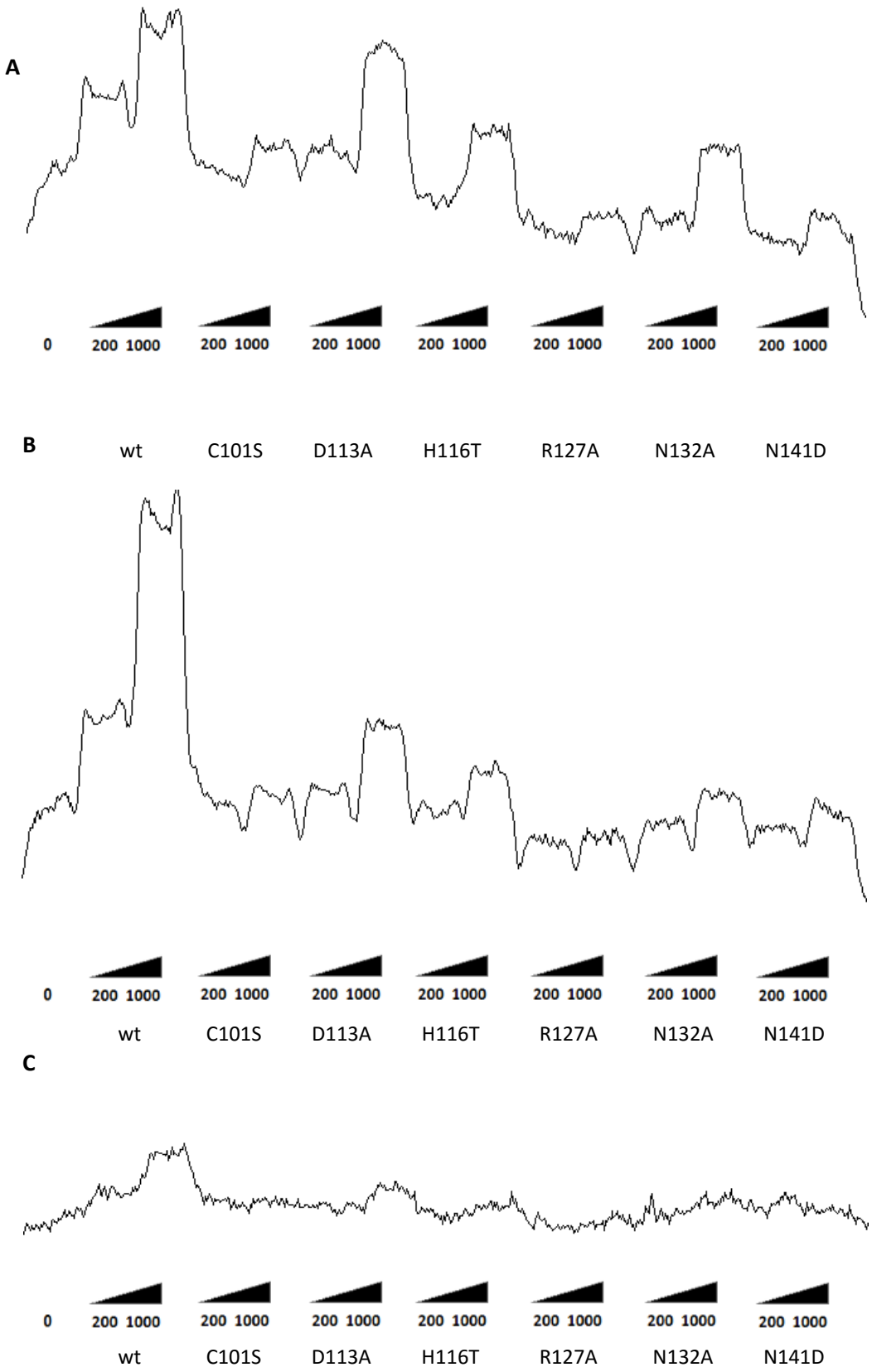

Figure 5.6. Relative density of branched cleavage products of J11 by Rap mutants. (A) Relative density of the 3-arm branched product. (B) Relative density of the 2-arm branched product. (C) Relative density of the 1-arm product. Relative units of band density analysed using ImageJ.

The wt Rap protein showed cleavage of the Holliday junction yielding all three products at both protein concentrations, with greater cleavage at the higher concentration as expected (Figure 5.5). All of the mutant proteins showed some cleavage of the J11 structure, despite most failing to show any capacity to bind the Holliday structure (Figure 5.4). D113A, H116T and N132A showed clearly higher peaks at 1000 nM compared to 200 nM consistent with a reduction in resolution activity relative to the wt. Differences between the two concentrations were less pronounced with C101S, R127A and N141D which showed considerably reduced cleavage activity at both protein concentrations suggesting these residues play an important role in catalysis.

The mutant proteins that showed higher cleavage activity maintained a trend in the ratio of cleavage products at both concentrations for example the ratio of 3-arm/2-arm/1-arm with D113A is 3.6:2.1:1 at 200 nM and 4.3:2.8:1 at 1000 nM however the wild type showed a greater increase in the 2-arm (nicked duplex) product at 1000 nM with the ratios changing from 4:3.2:1 to 3:4.1:1. This may be a result of the Rap protein acting on the 3-arm product to generate the two stranded duplex (Figure 5.6. A, B and C).

The proteins were also examined for activity on another Holliday junction structure, J12, containing a 12 bp homologous core with a different mobile sequence to that found in J11. The experiments were performed and analysed in the same way as the cleavage experiments with J11.

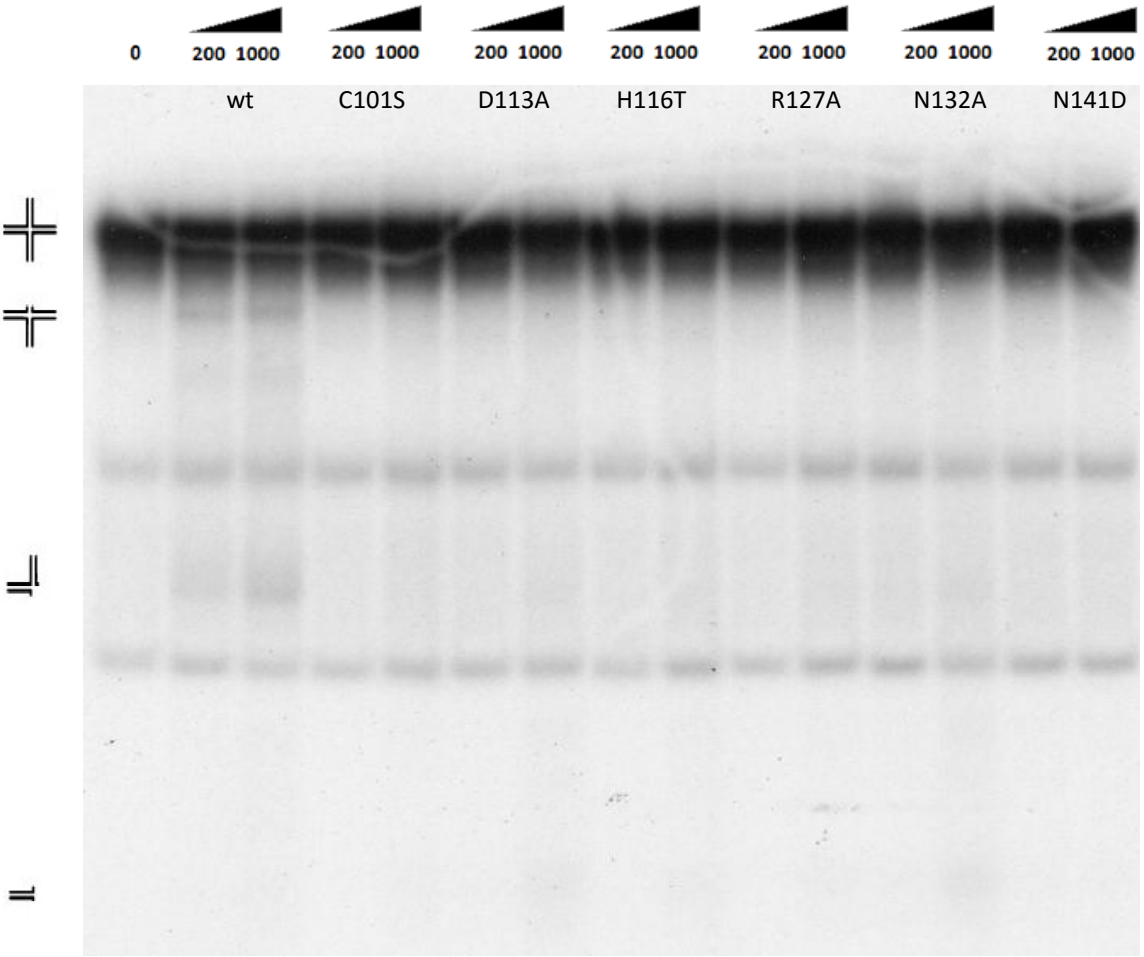

Figure 5.7. Holliday junction J12 DNA cleavage by Rap mutants. Cleavage reactions contained 0.3 nM ^{32}P-labelled J12 (Appendix A) with 0, 200 or 1000 nM protein and 1 mM $MnCl_2$.

The J12 substrate shows some instability with flayed duplex and single-stranded DNA evident in the lane without protein (Figure 5.7). The amount of protein used was the same in both experiments so this is likely a result of the structure being more vulnerable to dissociation rather than there being more contaminating nucleases present in the samples. The results were again analysed in rows across the gel rather than by lanes to avoid measuring any of these other breakdown products which are equally present in all lanes, including the control without protein, showing that they do not result from exposure to wt or mutant Rap proteins.

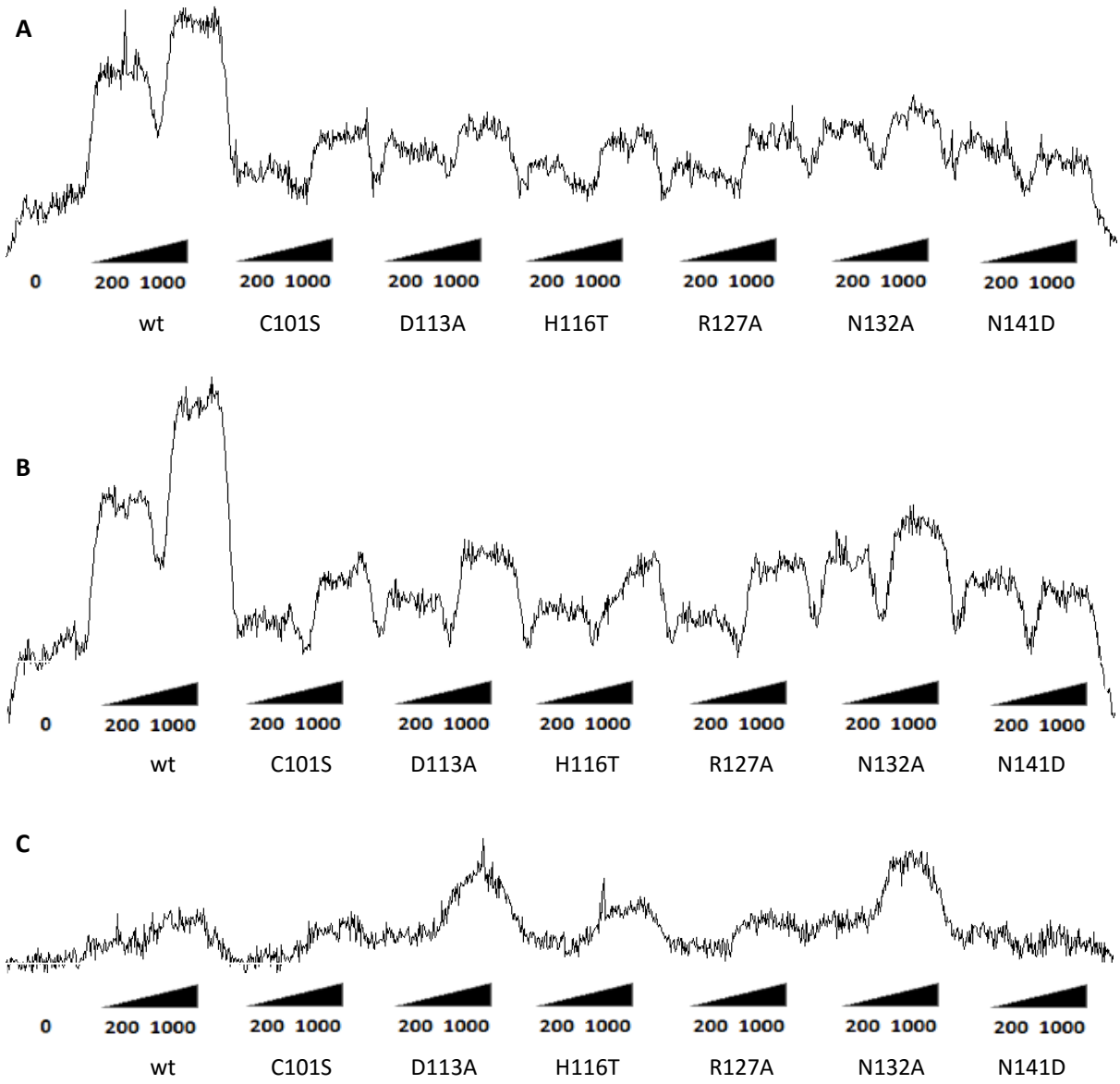

Figure 5.8. Relative density of branched cleavage products of J12 by Rap mutants. (A) Relative density of the 3-arm branched product. (B) Relative density of the 2-arm branched product. (C) Relative density of the 1-arm product. Relative units of band light density from autoradiography measured by ImageJ.

The results with J12 generally mirror those with J11 with all mutants showing reduced activity compared to Rap wt, with D113A, H116T and N132A showing weak activity and C101S, R127A and N141D showing the least. As with the junction J11, the area graphs for J12 cleavage show a similar trend of increase from 200 nM to 1000 nM protein concentration for most proteins with wild type Rap yielding the most 3-arm and 2-arm cleavage products (Figure 5.8. A and B). However, in this experiment two of the mutants, D113A and N132A show denser bands at the position of the 1-arm product than observed in the lanes containing wt Rap (Figure 5.8. C). This contrasts with the results from the J11 experiments which showed that wt Rap always yielded the most breakdown products and that the mutants tended to favour the generation of a 3-arm product (Figure 5.8. A, B and C). Here the mutants showed limited production of 3-arm product and much larger amounts of the 1-arm

product. It is possible that with J12, the 3-arm product is more prone to secondary cleavage to yield a greater quantity of single arm product.

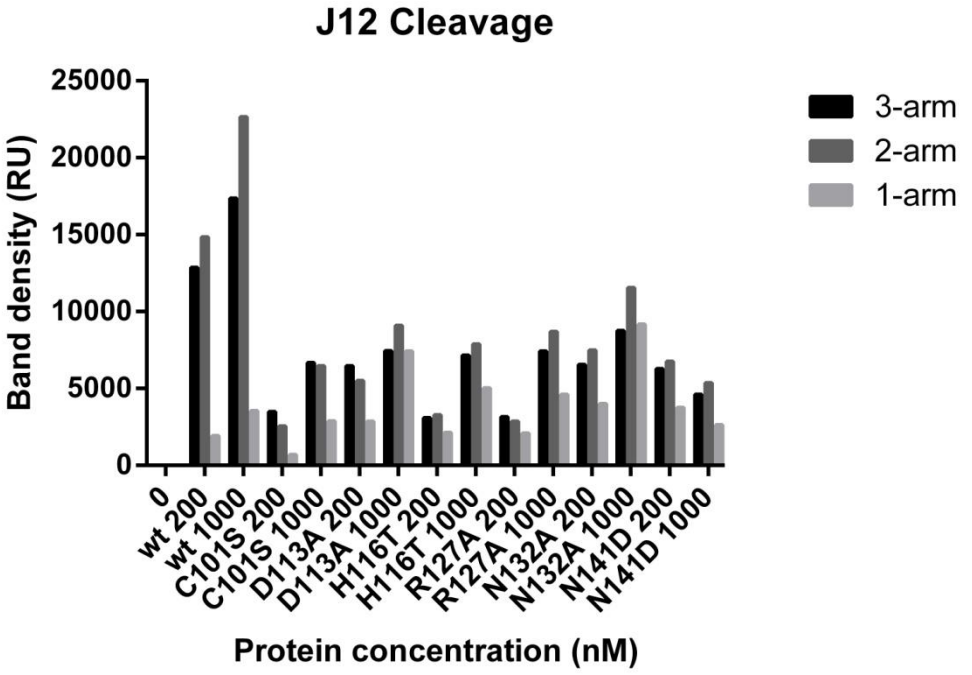

Figure 5.9. Relative densities of each cleaved product of the J12 structure by Rap mutant proteins. Relative units of band light density from autoradiography measure by ImageJ and sorted into 3-arm, 2-arm and 1-arm branched products.

The majority of the mutant proteins showed reduced amounts of the 3-arm product with J12 relative to J11, with only C101S, D113A and R127A showing evidence of higher levels of 3-arm product (Figure 5.9). This may be due to the different core nucleotide sequences between the two junctions and their impact on the preferential resolution of the junction. Rap shows a limited preference for resolution between 5'-GC-3' dinucleotides, of which there are two in the homologous core of J11 but only one in the J12 [166]. It may be that these different mobile core sequences affect the outcome of cleavage meaning that 3-arm and 1-arm products are less likely to be formed.

There was also a much higher ratio of 1-arm product to 3-arm than noted in the J11 experiment. For example the ratio of 3-arm:2-arm:1-arm with D113A on J11 was 3.6:2.1:1 at 200 nM protein and 4.3:2.8:1 at 1000 nM but with J12 the ratios were 2.3:1.9:1 and 1:1.2:1 respectively. The results could be due to a contaminating nuclease activity, however, this is unlikely since a higher level of 1-arm product was not noted with J11. Hence it is more likely that these mutants are more prone to cleave 3-arm and 2-arm reaction products are the initial cleavage has been made. In contrast, wt Rap showed a low proportion of 1-arm product on J12, with a ration of 6.7:7.7:1 at 200 nM protein and 4.9:6.4:1 at 1000 nM.

5.2 MBP-Rap deletions

Two deletion mutants of Rap (Figure 5.13) tagged at the N-terminus with maltose-binding protein (MBP) that had previously been purified (F.A. Curtis, unpublished results) were also tested for binding and cleavage of the J11 Holliday structure (Appendix A). Samples of purified MBP-Rap wt, NΔ80 (lacking 80 residues from the N-terminus, including a possible zinc binding domain) and CΔ104 (lacking 104 residues from the C-terminus, including the HNH domain) were analysed by SDS-PAGE to evaluate purity. All of the proteins showed a high level of purity suitable for biochemical analysis and the deletion derivatives showed bands of appropriately reduced molecular mass. Rap CΔ104-2 was selected for testing alongside the MBP-Rap wt and NΔ80 samples.

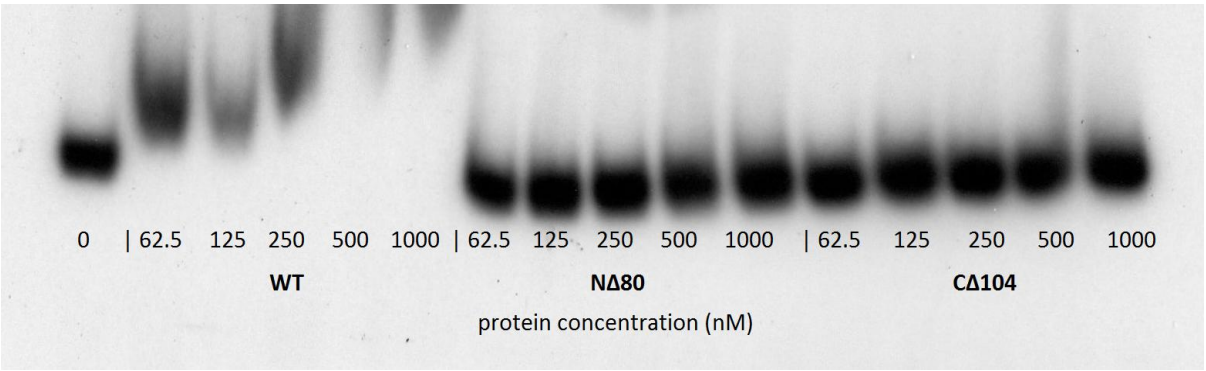

Figure 5.10. Holliday junction DNA binding of MBP-Rap deletion mutants in a gel shift assay. The binding mixture contained 0.3 nM ^{32}P-labelled J11 (Appendix A) and protein at the indicated concentration.

The Rap proteins were mixed with the Holliday junction J11 on ice to asses DNA binding under the same conditions as the His-Rap mutants. Unfortunately there was a problem with drying of the gel which distorted the results and there was insufficient time to repeat the experiment. However, the results are still interpretable. MBP-Rap wt bound the junction with high affinity shifting all of the DNA even at the lowest concentration used (62.5 nM). In contrast, neither of the two deletion mutants showed any binding to the Holliday junction even at 1000 nM protein (Figure 5.10). This result confirms that the 80 residues at the N-terminus and 104 at the C-terminus are critical for Rap binding to Holliday junction DNA. However, many of the His-Rap substitution mutants showed no Holliday junction binding but were still capable of some cleavage activity. Hence the NΔ80 and CΔ104 deletion mutants were assayed for resolution activity of the J11 structure (Figure 5.11).

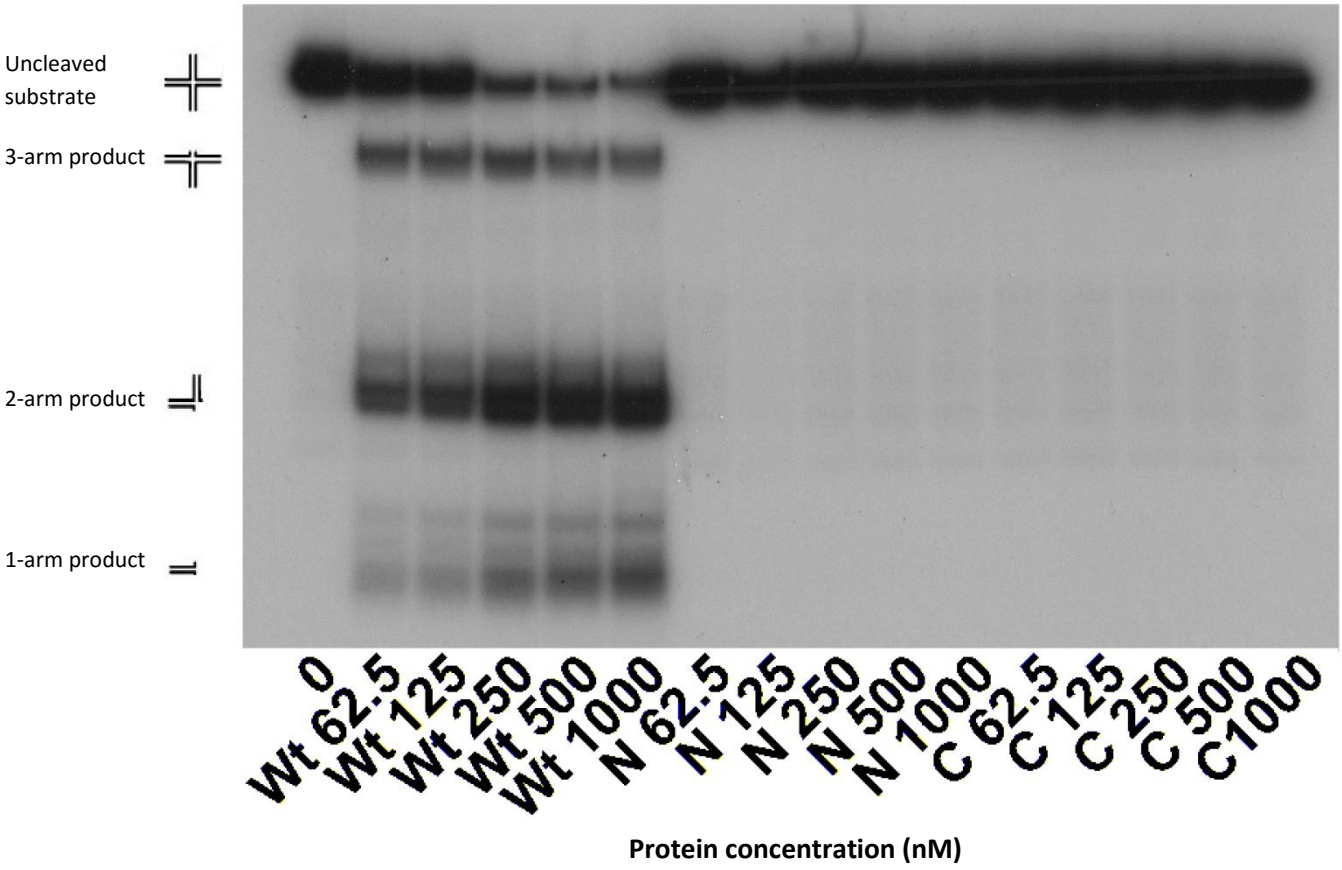

Figure 5.11. Holliday Junction cleavage by Rap deletion mutants. Cleavage reactions contained 0.3 nM ^{32}P-labelled Holliday junction J11 (Appendix A) and Rap proteins as indicated. N: NΔ80 and C: CΔ104.

The results show that MBP-Rap shows high activity on J11 to generate 3-,2- and 1-arm products, considerably more than His-Rap wt under the same experimental conditions (compare Figures 5.4 and 5.11). The amounts of each product were analysed using the approach taken with the binding gels involving NinH and Orf.

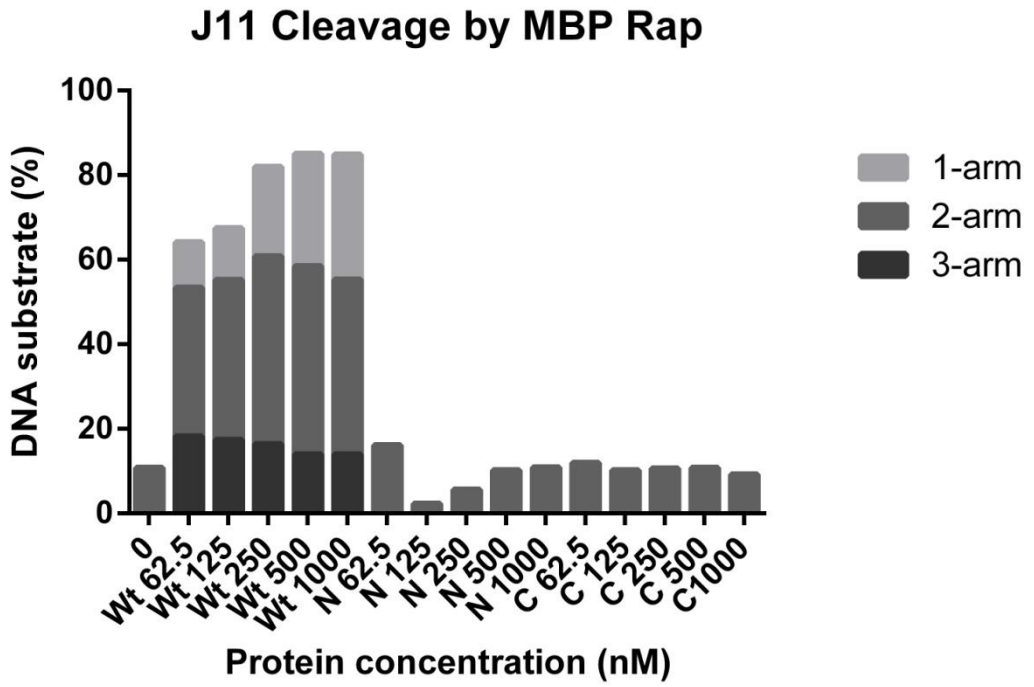

Figure 5.12. Percentage Holliday junction DNA cleavage by MBP-Rap mutants. Cleavage reactions contained 0.3 nM ^{32}P-labelled J11 Holliday junction DNA (Appendix A) and 0.1 mM $MnCl_2$. Relative amounts of 3-arm, 2-arm and 1-arm products were quantified using ImageJ.

MBP-Rap wt cleaved the majority of the junction DNA substrate with the lowest protein concentration (62.5 nM) yielding 64.1% of the DNA in cleaved products while the highest concentration (1000 nM) left only 15.2% of the J11 substrate intact (Figure 5.12). This activity was much greater than that seen with His-Rap under at the same experimental conditions and protein concentrations. Although the published results do show an improved activity for MBP-Rap (Rap JMB paper), the differences in the differentially tagged proteins was not as extreme as seen here. Thus the results obtained with the His-Rap proteins need to be interpreted with caution.

Consistent with their inability to bind Holliday Junction DNA (Figure 5.10), the two deletion mutants showed no significant cleavage activity on J11. Although there are traces of a 2-arm band in the lanes containing NΔ80 and CΔ104, there is no more present than in the control without Rap protein. The band at this position is probably due to instability in the substrate producing a 2-stranded fork that just happens to coincide with the 2-arm resolution product generated by MBP-Rap.

5.3 Discussion

A number of conserved residues in the proposed HNH nuclease domain of Rap were targeted for mutagenesis and six substitution mutants were purified. All of these mutants show reduced Holliday junction cleavage activity confirming that this region is likely to be the catalytic site of Rap. His-Rap wt protein showed preferential production of a 2-arm product from J11, which in most cases is likely to be a nicked duplex consistent with symmetrical resolution of the Holliday junction [166] (Figure 5.5). In contrast, the mutant Rap derivates on J11 tended to generate more 3-arm products perhaps indicating that they show an increased uncoupling of junction resolution. While the Rap mutants did show more cleavage resulting in a 2-arm product on the J12 junction (Figure 5.9), this still did not reach the same levels achieved with the wild type. It is possible that Rap preferentially targets 3-arm products to generate more of the 2-arm product, however, this seems unlikely given its structure specificity [166, 168]. In reactions with the His Rap proteins there was still sufficient 4-stranded Holliday junction DNA available for resolution in preference to 3-arm products, unless this junction was in a state that limited cleavage by folding or at sequences that were not suitable for resolution (Figures 5.4 and 5.7). The increased production of a 1-arm product with some of the mutants on J12 does fit with an increase in multiple, uncoupled incisions at the branch point rather than the paired symmetrical incisions that should be favoured by wt Rap. Of the six mutants tested the D113A and N132A mutants retain the highest junction binding and cleavage activity suggesting these residues are less important for Rap function (Figure 5.13). D113A cleaves J11 best whereas N132A performs best on J12, suggesting that there may be some effect on sequence selectivity in each of these mutants. The other four mutants show significant defects in Holliday junction resolution suggesting that the zinc finger module is essential for catalysis as judged by mutation of Cys101 and that His116, Arg127 and Asn141 are also critical for endonuclease function (Figure 5.13). His116 and Asn141 both have equivalents in T4 endonuclease VII (Figure 5.13 B and C) and could fulfil a similar role in coordinating magnesium or manganese binding for phosphodiester bond hydrolysis [232, 234, 235].

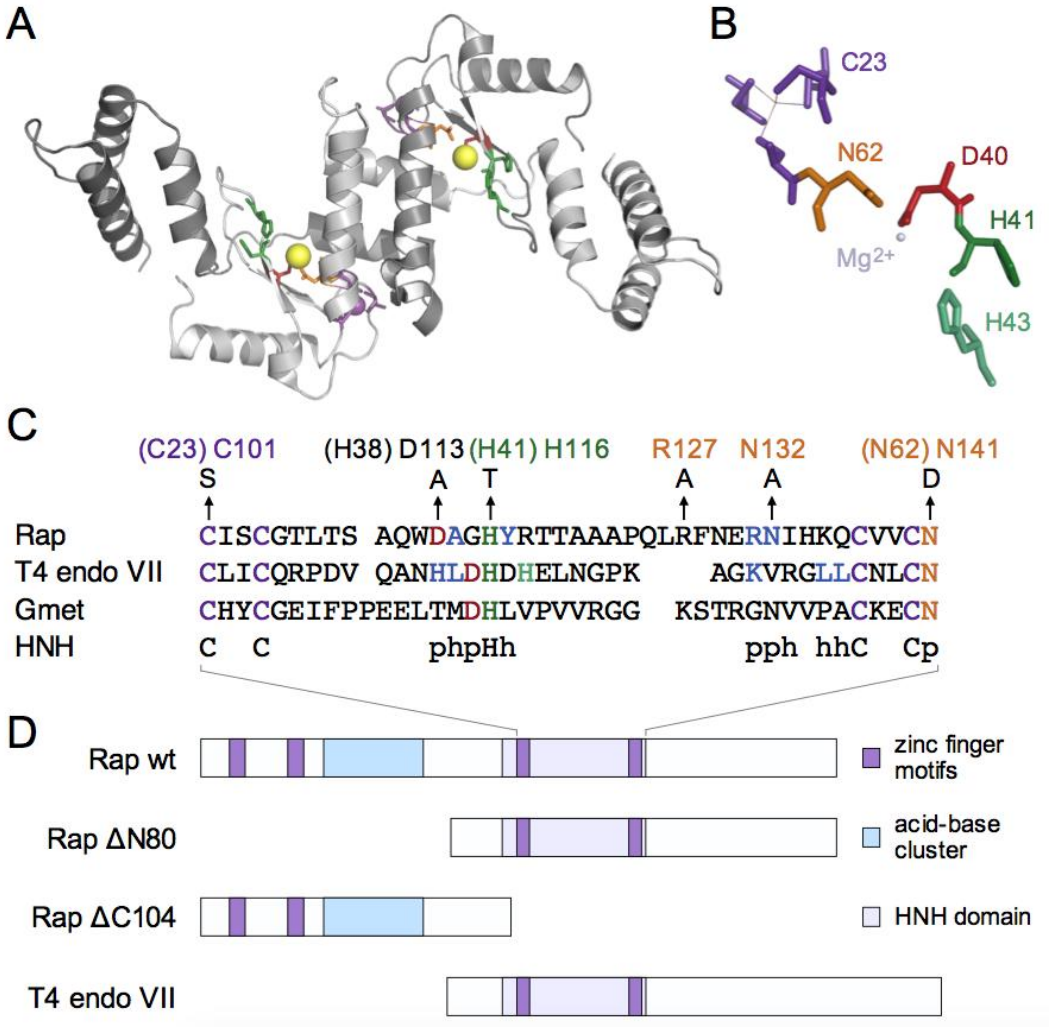

Figure 5.13. Structural features of Rap and T4 endonuclease VII. A. T4 endonuclease VII structure from PDB entry 2QNC [238]. The two subunits are shown in different shades of grey. Residues involved in the HNH endonuclease domain are highlighted. Zinc and magnesium atoms are shown as spheres in purple and yellow, respectively. B. Residues in T4 endonuclease VII that are likely to be involved in catalysis. They are coloured as in (A) and (C). C. Alignment of HNH endonuclease motifs of Rap, T4 endonuclease VII [232] and *Geobacter metallireducens* Gmet_0936 [233]. Conserved residues are highlighted as in (A) and (B). The HNH consensus is indicated below the aligned residues. p=polar and h = hydrophobic. Residues selected for mutagenesis are shown above with the corresponding position in T4 endo VII given in parenthesis. D. Diagram illustrating the location of conserved features in Rap and the position of the two Rap deletions.

The MBP-Rap deletion mutants essentially removed either half of the protein with CΔ104 retaining the N-terminal portion containing a zinc binding module and region rich in positively and negatively charged amino acids but lacking the HNH motif along with the remainder of the C-terminus. NΔ80 was essentially the opposite, retaining the HNH potion with the C-terminus (Figure 5.13). Previous experiments [167] indicated that truncated versions of Rap carrying C-terminal deletions could still bind to Holliday junctions.

The results here reveal that neither of these deletion mutants were functional in binding or cleaving model Holliday junctions. Thus Rap requires both the N (zinc finger) and C terminal (HNH motif which incorporates a zinc finger) for branched DNA recognition and resolution.

Chapter 6. Conclusions and Future Work

Genetic recombination in lambdoid bacteriophages is important to preserve genomic integrity by offering a repair mechanism for double-stranded DNA breaks. A by-product of this repair pathway is the provision for exchanges at short regions of sequence homology resulting in the generation of significant genomic diversity with the capacity to gain and lose functions within the constraints of viability. In this study we have characterised three DNA binding proteins from the *ninR* region of phage λ that are involved in genetic recombination or other aspects of phage DNA metabolism. Specifically we have focused on the DNA binding properties of NinB (Orf), NinG (Rap) and NinH, the latter being previously uncharacterised.

Experiments with a selection of Orf mutant proteins carrying single amino acid substitutions revealed that the central channel of the Orf dimer ring is important for the formation of DNA-protein complexes. The conservation of residues in this region and its overall positive charge fits with DNA being threaded through the toroidal structure. However, residues located away from the central cavity also contributed to stable DNA binding indicating that there may be wrapping of ssDNA around the dimer, perhaps helping to destabilise interactions between SSB and ssDNA to facilitate loading of a recombinase [164, 177, 178]. The possibility that ssDNA can enter the narrow channel in Orf implies that the dimer must open by some sort of hinge mechanism since Orf can bind to gapped duplex substrates that should prevent loading due to the larger dimensions of the flanking duplexes. It is possible that there are two modes by which Orf associates with ssDNA, one involving traversal of the core channel, the other involving wrapping around the dimer, possibly with involvement of the C-terminal helix [177, 178]. DNA binding experiments with mutant proteins aimed at evaluating the clamp model were partly inconclusive. Further analysis of DNA binding would be informative and analysis of opened and closed Orf dimers by FRET would merit further investigation.

The *in vitro* studies with NinH revealed that it behaved very much like the bacterial Fis protein as predicted by bioinformatics analysis. NinH displayed a preference for dsDNA over ssDNA along with a greater affinity for dsDNA containing a bend. This was particularly evident in the formation of two complexes with DNA containing 3 nucleotides on one strand of the substrate compared with bent DNA with only a single additional nucleotide. The *in vivo* studies showed that NinH could not functionally replace the growth defect of a *fis* mutant; in fact it conferred a significant defect on growth and appeared to delay entry into the exponential phase. Similar negative effects were observed with several other mutants defective in NAPs but did not majorly affect wild-type strains. Thus NinH may interfere with nucleoid structure or global gene expression, which may promote phage lambda replication. Given that binding sites for Fis and IHF are found within the origin of replication (*oriC*) in *E.coli*, it would be useful to determine if NinH shows any preferential binding to such sequences. It would also be worthwhile to evaluate the effects of NinH on nucleoid formation and replication initiation. Microscopy could be used to evaluate any lethal effects of NinH expression in NAP mutants (with DAPI used to stain the chromosomal DNA) and live-dead staining to monitor the viability and cell wall integrity. Exposure to UV radiation or other genotoxic agents could be tested to see if they exacerbate the effects of NinH. The replication of lambda should also be investigated as Fis is known to influence the integration and excision reactions by binding *attP* [239, 240]. A preliminary NMR structure of NinH has been determined (C Redfield, personal communication) which fits with the structural predictions described here. A refined version of this or a crystal structure would be insightful in understanding how DNA binding and bending are mediated.

Site directed mutagenesis of the predicted HNH endonuclease motif of Rap confirmed that the highly conserved residues selected for mutation did appear to be important for binding to Holliday junction DNA. All of the mutants also showed varying defects in Holliday junction cleavage confirming that this region is likely to constitute the catalytic site. Several of the Rap mutants also appear to show a reduction in the ability to produce symmetrically-related paired incisions. This may be a consequence of the uncoupling of endonuclease activity between individual Rap subunits to give a lop-sided cut. This could be due to a reduction in Rap dimer formation, although it is possible that Rap functions as a monomer as dimer formation has never been conclusively demonstrated [166-168]. Asp113 appears to be the least important of the residues studied as D113A retained more cleavage activity than the other five mutants testes; this observation fits with Asp113 being the least conserved of the selected residues. The studies with deletion derivatives of Rap confirmed that Holliday junction binding and cleavage by Rap requires both the predicted zinc finger domain in the N-terminus and the HNH endonuclease domain. The Rap mutants behaved largely as expected in the assays performed although the activities were considerably lower than anticipated. The binding and cleavage assays should therefore be repeated. It would also be good to study any effects of the mutant proteins on the sequence specificity of junction resolution using denaturing gel electrophoresis. A major goal would be to determine the crystal structure of the Rap protein which proved to be beyond the scope of this present study.

Appendix A. Oligonucleotides

Oligo	Nucleotide sequence (5'-3')
o-SS60	TTTGGTCTAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCTGAATCAGGTTATTA
o-DS60	TAATAACCTGATTCAGCGAAACCAATCCGCGGCATTTAGTAGCGGTAAAGTTAGACCAAA
o-BB20	TAATAACCTGATTCAGCGAA**TGACCGATAACGTCCACTTG**AGCGGTAAAGTTAGACCAAA
o-BY1	TAATAACCTGATTCAGCGAAACCAATCC**A**GCGGCATTTAGTAGCGGTAAAGTTAGACCAAA
o-BY3	TAATAACCTGATTCAGCGAAACCAATCC**AAA**GCGGCATTTAGTAGCGGTAAAGTTAGACCAAA
J11-1	GGCGACGTGATCACCAGAT**GATTGCTAGGC**ATGCTTTCCGCAAGAGAAGC
J11-2	GGCTTCTCTTGCGGAAAGCAT**GCCTAGCAATC**CTGTCAGCTGCATGGAAC
J11-3	GGTTCCATGCAGCTGACAG**GATTGCTAGGC**TCAAGGCGAACTGCTAACGG
J11-4	ACCGTTAGCAGTTCGCCTTGA**GCCTAGCAATC**ATCTGGTGATCACGTCGC
J12-1	GACGCTGCCGAATTCTGGC**TTGCTAGGACAT**CTTTGCCCACGTTGACCC
J12-2	TGGGTCAACGTGGGCAAAG**ATGTCCTAGCAA**TGTAATCGTCTATGACGTT
J12-3	CAACGTCATAGACGATTACA**TTGCTAGGACAT**GCTGTCTAGAGACTATCGA
J12-4	ATCGATAGTCTCTAGACAGC**ATGTCCTAGCAA**GCCAGAATTCGGCAGCGT

A.A. Oligonucleotides used to generate DNA substrates. Nucleotides highlighted show where sequences differ from o-SS60 to generate the bubble and bend structures when annealed or in the case of the J11 and J12 Holliday junction arms denote the 11 or 12 bp mobile core.

Appendix B. Publications

OPEN ACCESS Freely available online

PLOS ONE

Phage Orf Family Recombinases: Conservation of Activities and Involvement of the Central Channel in DNA Binding

Fiona A. Curtis[1], Ali D. Malay[2], Alexander J. Trotter[1], Lindsay A. Wilson[1], Michael M. H. Barradell-Black[1], Laura Y. Bowers[1], Patricia Reed[1], Christopher R. T. Hillyar[1], Robert P. Yeo[1], John M. Sanderson[3], Jonathan G. Heddle[2], Gary J. Sharples[1]*

1 School of Biological and Biomedical Sciences, Biophysical Sciences Institute, Durham University, Durham, United Kingdom, 2 Heddle Initiative Research Unit, RIKEN, Wako, Saitama, Japan, 3 Department of Chemistry, Biophysical Sciences Institute, Durham University, Durham, United Kingdom

Abstract

Genetic and biochemical evidence suggests that λ Orf is a recombination mediator, promoting nucleation of either bacterial RecA or phage Redβ recombinases onto single-stranded DNA (ssDNA) bound by SSB protein. We have identified a diverse family of Orf proteins that includes representatives implicated in DNA base flipping and those fused to an HNH endonuclease domain. To confirm a functional relationship with the Orf family, a distantly-related homolog, YbcN, from *Escherichia coli* cryptic prophage DLP12 was purified and characterized. As with its λ relative, YbcN showed a preference for binding ssDNA over duplex. Neither Orf nor YbcN displayed a significant preference for duplex DNA containing mismatches or 1-3 nucleotide bulges. YbcN also bound *E. coli* SSB, although unlike Orf, it failed to associate with an SSB mutant lacking the flexible C-terminal tail involved in coordinating heterologous protein-protein interactions. Residues conserved in the Orf family that flank the central cavity in the λ Orf crystal structure were targeted for mutagenesis to help determine the mode of DNA binding. Several of these mutant proteins showed significant defects in DNA binding consistent with the central aperture being important for substrate recognition. The widespread conservation of Orf-like proteins highlights the importance of targeting SSB coated ssDNA during lambdoid phage recombination.

Citation: Curtis FA, Malay AD, Trotter AJ, Wilson LA, Barradell-Black MMH, et al. (2014) Phage Orf Family Recombinases: Conservation of Activities and Involvement of the Central Channel in DNA Binding. PLoS ONE 9(8): e102454. doi:10.1371/journal.pone.0102454

Editor: Steven J. Sandler, University of Massachusetts, United States of America

Received May 1, 2014; **Accepted** June 17, 2014; **Published** August 1, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Biotechnology and Biological Sciences Research Council (www.bbsrc.ac.uk/) grants BB/C514523/1 and BB/F020503/1 to GS. JGH and ADM were funded by RIKEN Initiative Research Funding awarded to JGH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: gary.sharples@durham.ac.uk

Introduction

Recombination in bacteriophages salvages genomes for packaging by restoring damaged or broken molecules via exonuclease processing and annealing. Illegitimate exchanges promote rapid evolution as new gene combinations or acquisitions can be generated during joint formation. In phage λ the Red system is responsible for exchanges at DNA ends [1]. Redγ protein serves to inhibit the *Escherichia coli* RecBCD exonuclease, ensuring that rolling circle replication can proceed [2,3]. λ exonuclease (Redα) degrades the DNA duplex to generate 3′ ssDNA tails, which are bound by Redβ protein, a strand annealing protein that searches for homologous ssDNA sequences [1,4]. The combined action of Redα and Redβ would be expected to favor splice-type recombinants, although there is evidence to indicate that annealing events occur regularly in the context of exposed ssDNA at a replication fork [5–7]. Thus, DNA synthesis primed by the 3′ annealed strand or by template switching can provide a means of generating intact genomes suitable for incorporation into capsids.

Phage λ encodes an ancillary recombinase called Orf, which was identified by its ability to complement the recombination deficiencies apparent in *recF*, *recO* and *recR* mutants of *E. coli* [8,9]. The products of *recF*, *recO* and *recR* [10] function predominantly at ssDNA gaps and aid assembly of the primary bacterial recombinase, RecA, helping to overcome the inhibitory effects of ssDNA binding protein (SSB). Further *in vivo* studies revealed that the susceptibility of *recFOR* mutants to ultraviolet light could also be overcome by Orf [11]. Orf is a 17 kDa protein that behaves as a dimer in solution and interacts with *E. coli* SSB protein. It binds equally well to ssDNA, gapped duplex and 5′ and 3′ flap DNA and less well to fully duplex substrates [12]. The Orf crystal structure revealed an asymmetrical toroid with a central tunnel flanked by positively charged residues [12,13]. Although the channel is too narrow to accommodate dsDNA, it could potentially encompass ssDNA. One side of the central cavity comprises a RAGNYA motif found in a wide range of DNA binding proteins [14]. DNA binding could also occur along a shallow groove across the protein surface since Orf bound equally well to a ssDNA flanked by duplexes, which would prohibit

threading of ssDNA through the central hole [12]. Residues in the flanking C-terminal helices also contribute to stabilizing the association with DNA [13].

In this study we describe a family of phage and prophage Orf proteins exhibiting considerable sequence diversity but sharing common ancestry within a core domain. To substantiate their functional relatedness we purified one of the most dissimilar members, YbcN, from the *E. coli* cryptic prophage DLP12 and compared its properties with its counterpart from phage λ. YbcN was previously identified as a T:G mismatch and base-flipping enzyme [15], an activity that is not easily reconciled with a role in suppressing the genetic recombination and UV light repair defects of *recF*, *recO* and *recR* mutants.

YbcN behaved remarkably like λ Orf in its preference for binding ssDNA and in its association with *E. coli* SSB protein, supporting a functional relationship between the two phage proteins. However, some differences were noted in their interactions with DNA and SSB. YbcN bound much less well than Orf to ssDNA, although both proteins did show improved binding to bubble DNA structures. In addition, YbcN showed a specific interaction with the C-terminus of SSB, a feature not shared by λ Orf. Neither Orf nor YbcN exhibited an enhanced affinity for DNA containing a single G:G mismatch over fully duplex DNA, although Orf showed a slight preference for a T:G mispair. A number of mutants in λ Orf were generated to investigate the importance of the central cavity in DNA binding. Several of these mutants exhibited significant defects in DNA binding consistent with ssDNA passing through the interior of the Orf ring. These results, in combination with the ability of Orf to recognize bubble structures, support a clamp model for Orf assembly. The conservation of ssDNA and SSB binding activities in diverse members of the Orf family underscores the importance of facilitated loading of a partner recombinase for the successful initiation of phage recombination.

Results

The phage Orf family

A position-specific iterative BLAST search performed with the λ Orf protein identified homologs exclusively from phage genomes, predominantly Myoviridae, Siphoviridae and Podoviridae, or associated with prophage-like elements (Pfam05772 [16]). The initial BLASTP search uncovered multiple closely related sequences with an expectation value of less than 5×10^{-5}. All of these proteins displayed greater than 30% identity with Orf. At the first PSI-BLAST iteration additional sequences with an expectation value of less than 4×10^{-9} were identified, including a highly diverged Orf-like protein (YbcN) from the DLP12 cryptic prophage of *E. coli* belonging to the PRK09741 conserved domain [17], currently classified as distinct from the PF05772 NinB (λ Orf) family. This gene matches almost exactly (97% identity) the *ybcN* (*orf151*) gene of phage 82 [18], yet shares only 16% overall identity with Orf. Further searches initiated with DLP12 YbcN recovered typical Orf proteins including those from Aaϕ23, Stx2 and λ at the second PSI-BLAST iteration confirming the relationship. YbcN showed significant similarity to several prophage-associated proteins including *E. coli* O157:H7 (CP-933X; 93% identity), *Shigella sonnei* (81% identity), *Photorhabdus luminescens* (41% identity) and *Neisseria gonorrhoeae* (27% identity). The Phyre[2] protein fold recognition tool was used to compare the predicted secondary structures of DLP12, *P. luminescens* and *N. gonorrhoeae* YbcN proteins with known three-dimensional structures in the structural classification of proteins (SCOP) database. In all three cases, the λ Orf fold was returned with >95% estimated precision. Representative Orf family proteins identified in this study varied in length from 124-181 residues; a list, including accession numbers, is available in Table S1.

A phylogram was prepared using the PAUP program to visualize similarity between these selected Orf family representatives (Figure S1). In this comparison there appear to be at least four discrete Orf groups occurring in phages from diverse species, although most belong to the β and γ proteobacteria. As might be expected, examples from closely related species tend to cluster together. For example, one of the groups contains highly related Orf proteins displaying 91-99% identity and is principally derived from *E. coli* O157:H7 or O84:H$^-$ resident prophages (Figure S1, shaded in blue). However, this is not always the case as other clusters incorporate representatives from *E. coli* H phage HK620, *Actinobacillus actinomycetemcomitans* phage Aaϕ23 and *Haemophilus influenzae* prophages. Indeed some orthologs, mainly from prophages residing in dissimilar bacterial species, do not fit unambiguously into any of the above groups. These discrepancies illustrate the difficulty in unraveling phage and prophage lineages due to high rates of horizontal gene transfer and mutation.

The *Listeria* proteins belong to the Orf family but are distantly related, sharing only 18% identity with λ Orf (Figure S1 and Table S1). Both *L. innocua* Orf-like proteins belong to yet another distinct conserved domain DUF968 (Pfam06147), which includes homologous polypeptides from *Staphylococcus*, *Lactobacillus* and *Enterococcus* phages and prophages. The N-terminal domain of these latter proteins share 23–24% identity with λ Orf consistent with a distant relationship with the Orf family. However, members of the DUF968 family possess an additional C-terminal extension, absent from the majority of Orf homologs, which incorporates a zinc finger HNH endonuclease motif [19]. These phages may therefore encode a novel single-strand specific deoxyribonuclease arising from fusion of an Orf DNA binding domain to an HNH nuclease domain. It is interesting to note that the DUF968 domain is found fused to a number of recombinases, including the strand annealing proteins, Erf and Rad52, and the Holliday junction resolvase, RusA [17]. The HNH region alone from this family is also found joined to the mismatch recognition factor, MutS.

Orf family gene organization among lambdoid phages

The λ *orf* gene lies in the *ninR* region between replication gene P and the Q antiterminator (Figure 1A). In addition to Orf, this section encodes two other genes associated with DNA metabolism: Rap (NinG), a DNA branch-specific endonuclease [20] and NinH, a dsDNA binding protein (G.J. Sharples, unpublished results). Where possible to determine based on homology of the surrounding genes, *orf* resides in a similar context in many phages and prophages, sandwiched between replication and lysis functions (Figure 1A). Significant shuffling of the genes in this region is evident with acquisition of unrelated genes, many of unknown function, and elimination of others. Despite these rearrangements, there are notable consistencies in gene order. For instance, preservation of synteny is particularly evident in the occurrence of the short *ninD*, *ninE* and *ninF* genes (Figure 1A). In every case *orf* lies upstream of *rap* or *rusA* genes, which specify structurally distinct Holliday junction DNA resolvases [21]. The genetic organization of the *orf*-like gene in DLP12 (*ybcN*) located just three genes 5' of *rusA* adds support to the notion that it is a genuine ortholog of λ Orf (Figure 1A).

The gene arrangement is radically different among the DUF968 (Pfam06147 [17]) Orf-like proteins from *Listeria* and *Staphylococcus* in keeping with phage sequences from different evolutionary origins. However, homologs of *dnaB*, *dnaC* or *ssb* are located close

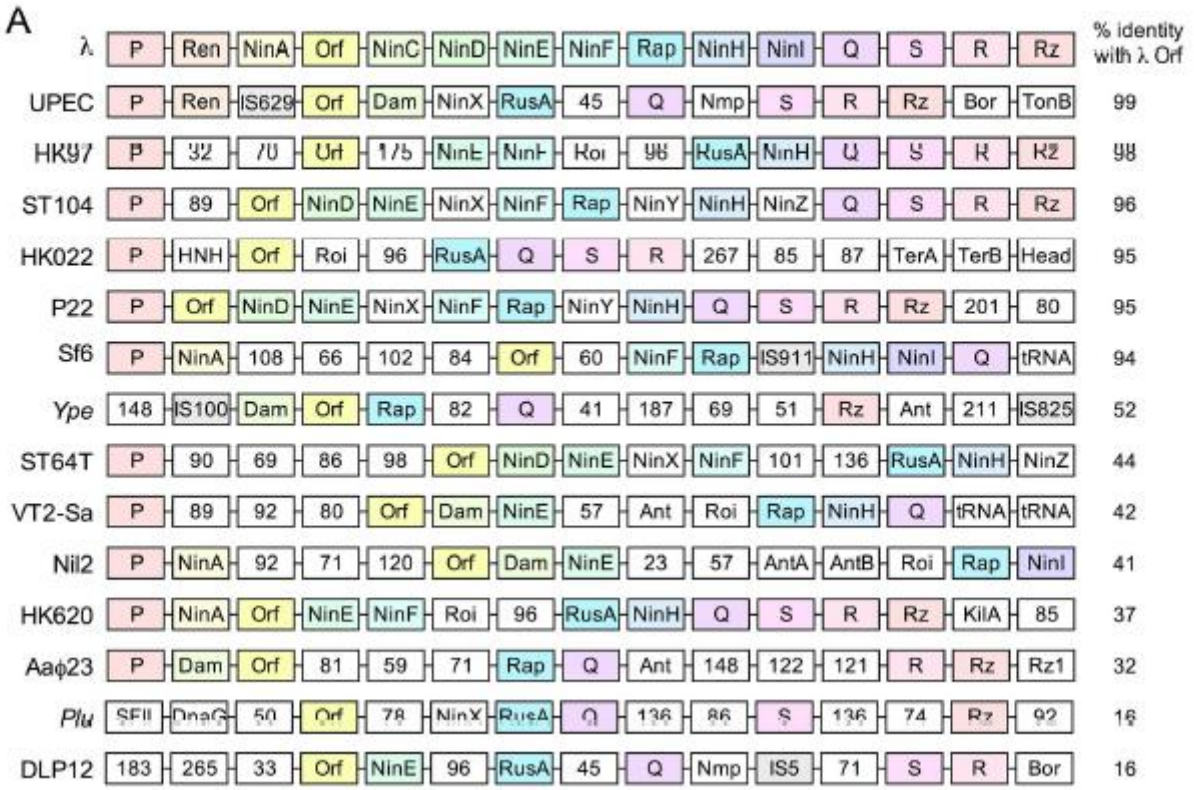

Figure 1. Conservation of genomic organization and protein architecture among selected members of the Orf family. (**A**) Conserved gene order of *orf* relative to the replication and lysis genes of lambdoid phages and prophages. Genes are shown as boxes and colored according to their order in the λ *ninR* region; in almost all cases they are transcribed from left to right. Orthologous genes are highlighted by name and color. Those that failed to match any of the λ genes in this region are shown as white boxes (the residue length of unannotated open-reading frames is indicated). Additional abbreviations refer to gene products that match the HNH nuclease domain family (HNH), DNA/RNA helicase superfamily II (SFII) or phage head proteins (head). The percentage identity to λ Orf is shown for each Orf homolog and listed in order of similarity. Many of the prophages are reservoirs for insertion sequences (shaded in grey); note that additional genes encoded by these elements are not shown. The putative 64-residue polypeptide between Q and S in λ was also omitted. (**B**) Conserved primary structure of λ Orf, DLP12 YbcN and φETA Orf20 proteins. λ Orf

is aligned with homologous representatives from *Y. pestis* (*Ype*), *E. coli* O157:H7 prophage Stx2-converting phage VT2-Sa and *E. coli* H phage HK620. DLP12 YbcN is aligned with a prophage homolog from *P. luminescens* (*Plu*). *Staphylococcus aureus* ϕETA Orf20 is aligned with a homolog from *Enterococcus faecalis* (*Efa*). Conserved residues within each subset are highlighted in blue (basic), red (acidic) and lilac (others); those occurring in both Orf/YbcN and Orf/Orf20 alignments are marked with an asterisk. Secondary elements from the crystal structure of λ Orf [12] are indicated above the alignment, as are the site-directed mutants Q45A, K48A, W50A, R103E, V106E and W137A. A pair of cysteines from the HNH motif are highlighted in orange.
doi:10.1371/journal.pone.0102454.g001

to the predicted *orf* gene in many of these sequences, linking these regions with functionality in DNA metabolism.

Sequence similarity between λ Orf and DLP12 YbcN

Secondary structure predictions and modeling using Jpred and Phyre[2] with DLP12 YbcN (and representatives of the DUF968 family) indicated a very similar overall architecture to that found with λ Orf (Figure S2). The main differences appear to be in the vicinity of α3 and β3 and in the C-terminal α-helix (Figure 1B). Any alterations in this first region could potentially affect dimerisation, although there are unlikely to be gross distortions in the overall fold. By examining Orf family sequence alignments alongside the crystal structure of λ Orf [12], functionally important residues in Orf and YbcN homologs were identified (Figure 1B). Nine hydrophobic residues from two discrete segments of Orf, namely β1-α1-β2 (comprising the RAGNYA motif [14]) and α3-β3-β4, contribute to the dimer interface. Four of these residues are conserved in YbcN homologs, while the remainder could be replaced by adjacent functionally equivalent amino acids (Figure 1B). Interestingly, the most highly conserved residues cluster around α-helix 2 and flank the positively charged channel that traverses the dimer [12]. Equivalents of λ Orf Arg41, Gln45, Asn46 and Arg48 are present in the majority of Orf sequences (Figure 1B, Figure S2 and data not shown). Arg41 from subunit B lies on the rim of the dimer cavity and could potentially contact the ssDNA backbone. In contrast, Arg41 in subunit A is hidden at the base of the cavity. Gln45 and Asn46 residues are symmetrically orientated and project from opposite sides towards the central channel. Asn46, in combination with Trp50 (which occurs as Trp or His in almost all Orf sequences), could interact with nucleotide bases if ssDNA can be induced to enter the ring interior. Finally, Arg48 is located at the rim of the aperture in each subunit, on opposite faces of the dimer, and could also make contacts with the C-terminal α-helix.

Purification of DLP12 MBP-YbcN

To resolve whether members of the PRK09741 (YbcN) group are genuine functional equivalents of λ Orf, we set out to purify YbcN from *E. coli* DLP12 and validate its properties *in vitro*. Plasmids carrying the *ybcN* gene were constructed in order to overexpress wild-type and an N-His_6 fusion of YbcN protein. High levels of expression were achieved, however, all of the YbcN recovered precipitated with the cellular debris upon lysis (data not shown). In an attempt to recover soluble protein we designed a construct that fuses the *E. coli* maltose binding protein (MBP) to the N-terminus of YbcN. Large quantities of soluble YbcN were recovered and the protein purified to near homogeneity by amylose and heparin chromatography. A GST-YbcN fusion was also purified at the same time using glutathione sepharose, heparin agarose and Q sepharose chromatography steps. MBP-YbcN and GST-YbcN proteins were analyzed in parallel with purified MBP-Orf and GST-Orf fusions from bacteriophage λ [13].

YbcN binds preferentially to ssDNA

λ Orf protein favors binding to ssDNA over dsDNA in keeping with its predicted role in the initial steps of recombinational exchange [12]. To confirm that YbcN is functionally analogous to Orf, we investigated the ability of MBP-YbcN to bind short ssDNA (SS_{50}) and dsDNA (DS_{50}) substrates in an electrophoretic mobility shift assay carried out in the presence of EDTA to exclude metal ions. Similar to MBP-Orf (Figure 2A, lanes b-e), MBP-YbcN formed a single complex with the 50 nucleotide (nt) ssDNA substrate (Figure 2A, lanes l-o). In both cases smearing of the substrate is indicative of unstable protein-ssDNA complexes that tend to dissociate during electrophoresis. MBP-YbcN exhibited a lower affinity for ssDNA relative to MBP-Orf protein (Figure 2B). Replacing EDTA with 1 mM $MgCl_2$, in both binding reaction and gel, significantly reduced binding of MBP-YbcN and MBP-Orf to SS_{50} (Figure 2B). MBP-Orf was much more sensitive than MBP-YbcN to inclusion of Mg^{2+} ions. No further reduction in ssDNA binding was observed when 10 mM $MgCl_2$ was used (data not shown). The results are consistent with a counterion effect and protein-DNA interactions occurring through electrostatic contacts with the DNA backbone. Neither MBP-YbcN nor MBP-Orf bound well to 50 bp dsDNA, although both proteins did manage to form a stable complex with a small proportion of the substrate with smearing again consistent with a weak association (Figure 2A, lanes g-j and q-t).

To provide more quantitative data on the relative affinities of YbcN and Orf for ssDNA, we employed fluorescence anisotropy using a fluorescein-labeled SS_{60} substrate. This method relies on the fact that protein assembly onto DNA slows the spin of the fluorescently-labeled molecule in polarized light resulting in an increase in observed anisotropy (r_{obs}). Studies are typically conducted with DNA substrates of <40 bp in length but larger substrates of 70 bp also yield reliable data [22]. As in the gel shift assays, MBP-YbcN bound much less well to ssDNA than MBP-Orf, with a 21-fold difference in their estimated dissociation constants (Table 1).

Orf and YbcN binding to mismatch, bubble and bent DNA

Since YbcN has been implicated in binding preferentially to DNA containing a T:G mismatch or abasic site [15], we compared the binding properties of MBP-YbcN and MBP-Orf on 60 bp DNA substrates with 1, 5, 13 and 20 nt unpaired regions, utilizing 60 nt ssDNA as a control. The larger bubble structures resemble structures more likely to be encountered during recombination rather than DNA mismatch recognition and repair. MBP-Orf formed two complexes with SS_{60} (Figure 2C, lane b), suggesting that the additional 10 nt permits assembly of two dimers of Orf compared to the 50 nt substrate (Figure 2A, lanes b-e). As noted before (Figure 2A), binding to dsDNA was less efficient and the complexes formed were prone to dissociation (Figure 2C, lane e). Inclusion of mismatches in the DNA affected the pattern of complexes formed, with BB_5, BB_{13} and BB_{20} showing improved binding and formation of a more stable, distinct complex (Figure 2C, lanes k, n and q); the complex formed with these substrates is actually composed of two closely-migrating species, possibly indicating two alternative conformations of Orf bound to bubble DNA. Binding of MBP-Orf to the single G:G mismatch

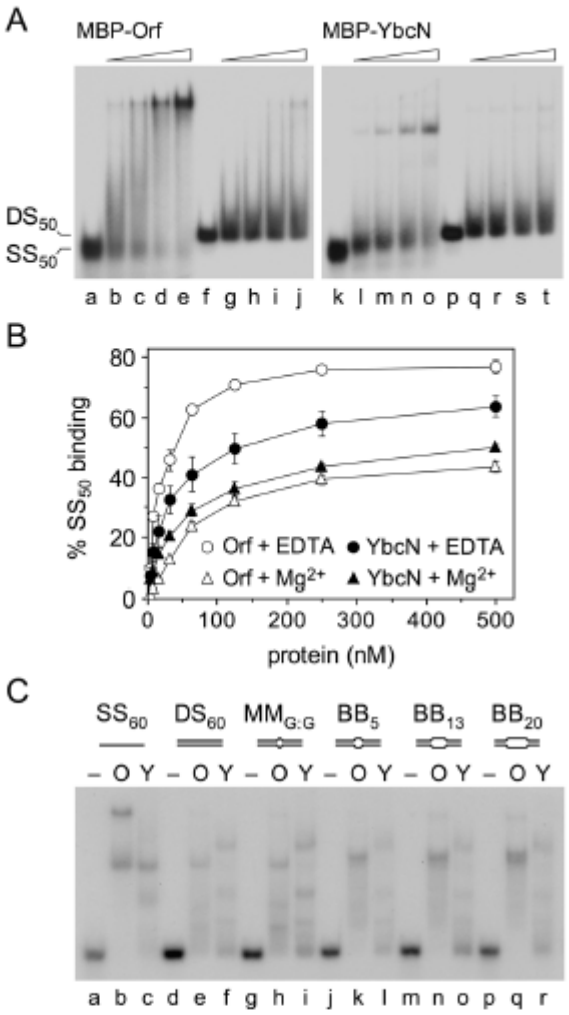

Figure 2. Comparison of MBP-YbcN and MBP-Orf binding to DNA. (**A**) Binding to ssDNA and dsDNA. Gel mobility shift assays contained 0, 62.5, 125, 250 and 500 nM MBP-Orf or MBP-YbcN proteins, 5 mM EDTA and either 0.3 nM of ^{32}P-labelled 50 nt (SS_{50}) ssDNA (lanes a-e and k-o) or 50 bp dsDNA (lanes f-j and p-t). (**B**) MBP-YbcN and MBP-Orf binding to ssDNA. Binding reactions contained 0.3 nM SS_{50} with 0, 7.81, 15.625, 31.25, 62.5, 125, 250 and 500 nM protein and either 5 mM EDTA or 1 mM $MgCl_2$. Data are the mean and standard deviation of two independent experiments. (**C**) Binding to mismatch and bubble DNA. Gel mobility shift assays contained 250 nM MBP-Orf (O) or MBP-YbcN (Y) proteins, 5 mM EDTA and 0.15 nM of ^{32}P-labelled 60 nt (SS_{60}) ssDNA (lanes a-c), 60 bp (DS_{60}) dsDNA (lanes d-f), 1 bp ($MM_{G:G}$) mismatch (lanes g-i), 5 bp (BB_5) bubble (lanes j-l), 13 bp (BB_{13}) bubble (lanes m-o) and 20 bp (BB_{20}) bubble (lanes p-r).
doi:10.1371/journal.pone.0102454.g002

(Figure 2C, lane h) did not differ significantly from that seen with the fully paired substrate (Figure 2C, lane e).

MBP-YbcN formed two complexes on SS_{60} (Figure 2C, lane c) and three with DS_{60}, although with a slightly altered mobility (Figure 2C, lane f). Surprisingly, as the length of unpaired DNA increased from 1-20 nt, there was no marked improvement in DNA binding or stability of complex formation (Figure 2C, lanes i, l, o and r). A 2.8-fold preference for BB_{20} over SS_{60} was apparent with MBP-YbcN in fluorescence anisotropy assays (Figure S3; Table 1). It is possible that the flanking duplexes in the bubble substrate help stabilize Orf, but not YbcN, complex formation by preventing protein dimers from sliding off the ends.

The absence of any preference for $MM_{G:G}$ (a G:G mispair) relative to DS_{60} (compare Figure 2C, lanes f and i) could be because YbcN has a high degree of specificity for single T:G base pair mismatches [15]. To assess this possibility, an $MM_{T:G}$ substrate with a T:G rather than a G:G mismatch was employed in fluorescence anisotropy assays. MBP-YbcN exhibited no significant differences in binding to $MM_{T:G}$ or a fully duplex control (Figure S4A). MBP-Orf showed some preference for the DNA containing a T:G mispair suggesting some specificity for this type of replication error, although the data sets varied, consistent with a relatively poor association with duplex DNA (Figure S4A). Binding of MBP-YbcN and MBP-Orf was also tested on substrates containing centrally located insertions of 1, 2 or 3 adenines in one of the strands to generate bent DNA (Figure S4B). Adenine bulges of 1-3 nt are known to introduce a kink of 50-70° in duplex DNA [23,24]. These substrates resemble mismatches by reducing the thermodynamic stability of the helix and inducing localized distortion of the DNA backbone. Neither protein showed any preference in binding these three substrates relative to DS_{60} (Figure S4B). Hence under the conditions employed here, neither YbcN nor Orf appear to favor bent DNA or duplexes where mispaired or bulged nucleotides might encourage the formation of extrahelical bases.

YbcN binds SSB

Previous studies [12,13] revealed an interaction between λ Orf and *E. coli* SSB consistent with a role at an early stage in phage recombination. Far western blotting experiments were conducted as in these earlier studies to determine whether an association with SSB is conserved in YbcN. As with λ MBP-Orf, an interaction between MBP-YbcN and SSB bound to the membrane was detected with anti-MBP antibodies (Figure 3A, lanes a-c), suggesting that contacts between these proteins are preserved and are therefore functionally relevant. In fact, YbcN yielded a stronger signal than Orf [13] indicating that it has a higher affinity for SSB. No bands were detected using MBP alone to probe membrane-bound SSB protein showing that YbcN is necessary to detect binding to SSB. The association with SSB was further validated using a purified GST-YbcN fusion in far western blots probed with antibodies specific for the GST tag. The results mirrored those obtained with the MBP-tagged YbcN protein (Figure 3A, lanes d-f).

A GST-YbcN interaction with SSB, similar to a positive control containing GST captured on the plate, was also detected in an enzyme-linked immunosorbent assay (Figure 3B). GST-Orf failed to produce a signal suggesting that some of the SSB epitopes normally recognized by Orf may be occluded by the manner of SSB attachment to the surface of the plate. Addition of a 51-nt ssDNA at the outset or upon addition of GST-Orf or GST-YbcN did not significantly alter the results obtained (data not shown). In addition, YbcN-SSB complex formation was probed using the yeast two-hybrid system. A clear interaction was detected between YbcN and SSB (Figure 3E). λ Orf and *E. coli* SSB also associated in equivalent assays but much less strongly [13].

The C-terminus of SSB is required for YbcN binding

Ten residues at the carboxy-terminus of *E. coli* SSB (168-PMDFDDDIPF-177) comprise a negatively-charged tail that functions as an assembly point for multiple proteins involved in DNA replication, recombination and repair [25]. λ Orf has previously been shown to bind SSB in far western assays in the

Table 1. MBP-YbcN and MBP-Orf mutant dissociation constants.

protein	SS_{60}	BB_{20}
YbcN	4602±275.1	1619±3.2
Orf wt	223±72.6	173±20.9
Q45A	272±20.4	264±15.2
K48A	622±0.7	551±6.4
W50A	3535±287.8	1606±418.2
R103E	–	–
V106E	522±12.4	540±73.2
W137A	628±99.9	285±1.9

Fluorescence anisotropy assays were performed in 100 mM Tris.HCl pH 7.0 with 10 nM fluorescein-labeled DNA. The apparent K_D (nM) was calculated for the protein as a dimer. Standard deviations of the duplicate data sets are shown.
doi:10.1371/journal.pone.0102454.t001

presence or absence of this C-terminal region [13]. Similar experiments were carried out with MBP-YbcN using SSB113, which carries a single amino acid substitution in this region (P176S) and is known to confer a temperature-sensitive phenotype [26], and SSBΔC10, which lacks the last 10 residues entirely. Although SSB113 and SSBΔC10 proteins transferred less well to the membrane than wt SSB (Figure 3C, row ii, lanes a-i), it does appear that the MBP-YbcN interaction with the SSB C-terminal mutants is reduced relative to the wt, especially with SSB113 (Figure 3, row i, lanes a-i). Residual binding could potentially be due to traces of wt SSB in the SSB113 and SSBΔC10 samples as both mutant proteins were, by necessity, purified from strains carrying an intact *ssb* gene.

The possible reduced association between YbcN and the two mutant SSB proteins was investigated further using the enzyme-linked immunosorbent assay with SSB bound to the plate and anti-GST antibody used to detect any interaction with GST-YbcN (Figure 3D). In this assay, no binding between YbcN and

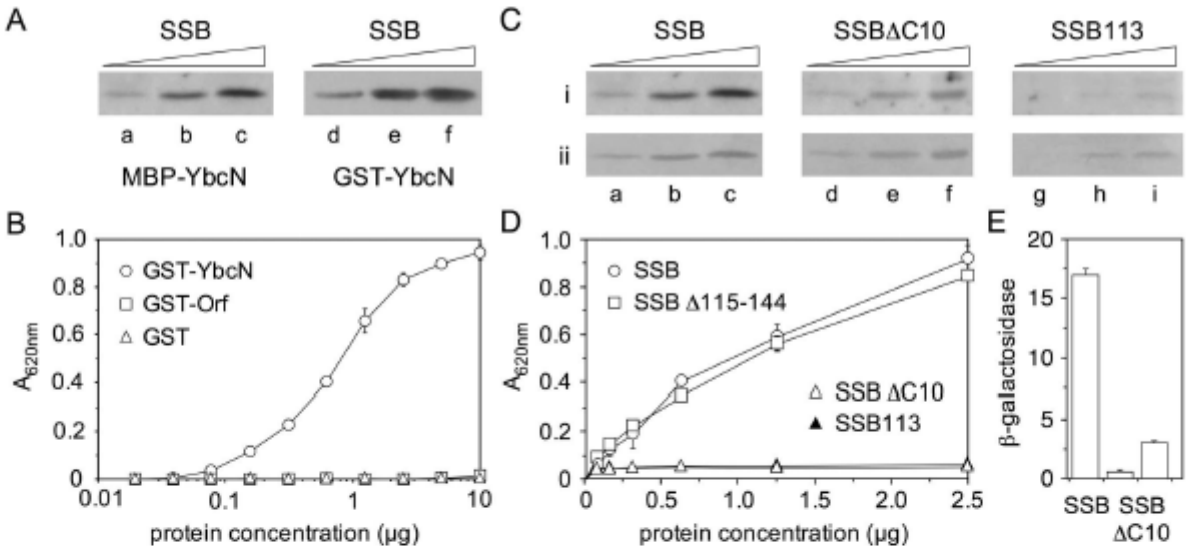

Figure 3. YbcN binding to SSB. (**A**) Interaction of YbcN with SSB in far western assays. SSB protein (0.5, 1.9, 3.8 μg) separated on 15% SDS-PAGE was blotted and probed with 20 μg MBP-YbcN (lanes a-c) or GST-YbcN (lanes d-f). Interactions were detected with antibodies specific for either the MBP or GST domains. (**B**) Interaction of GST-YbcN with SSB as detected by ELISA. SSB (0.5 μg) bound to Immulon-1 microtitre plates was exposed either to GST-YbcN, GST-Orf or GST proteins (0.0195-10 μg). Positive interactions were detected by addition of anti-GST antibody-HRP conjugate. Data are the mean and standard deviation of four independent experiments. (**C**) YbcN binding to SSBΔC10 and SSB113 proteins in far western assays. 0.5, 1.9, 3.8 μg SSB (lanes a-c), SSBΔC10 (lanes d-f) and SSB113 (lanes g-i) separated on 15% SDS-PAGE were blotted and probed with 20 μg MBP-YbcN. Interactions were detected with antibodies specific for MBP (row i). Row ii shows the efficiency of SSB, SSBΔC10 and SSB113 transfer onto blotted membranes as revealed by staining with amido black. Panel i (lanes a-c) is reproduced from (A) to facilitate comparisons with MBP-YbcN binding to wild-type SSB. (**D**) Interaction of GST-YbcN with SSB, SSBΔC10, SSB113 and SSBΔ115-144 as detected by ELISA. SSB wild-type and mutant proteins (0.5 μg) bound to microtitre plates were exposed to GST-YbcN (0.078-2.5 μg) and probed as in (B). Data are the mean and standard deviation of two independent experiments. (**E**) Yeast two-hybrid analysis of YbcN interactions with SSB and SSBΔC10. Experiments were performed using constructs fused to the GAL4 DNA-binding or activating domains. The β-galactosidase activities were the mean and standard deviation of three independent experiments. The vertical bars represent the β-galactosidase activity (Miller units) and values for YbcN-SSBΔC10 are given for pGBKT7-53 and pGADT7-T constructs in both orientations.
doi:10.1371/journal.pone.0102454.g003

SSBΔC10 or SSB113 was detected, whereas an interaction was readily observed with wt SSB. Another SSB deletion mutant, Δ115-144 [27], behaved like wt SSB, suggesting that this portion of SSB is dispensable for YbcN-SSB binding (Figure 3D). Anti-SSB antibodies were used to confirm that wt SSB, SSB113, SSBΔC10 and SSBΔ115-144 proteins all bound similarly well to the microtiter plate (data not shown). In addition, yeast two-hybrid analysis with the SSBΔC10 mutant showed a significant reduction in binding to YbcN (Figure 3E). Taken together, the results are consistent with the last 10 residues of SSB being necessary for a stable YbcN interaction. This contrasts with experiments showing that λ Orf retains the capacity to associate with SSBΔC10 and SSB113 mutants [13].

Secondary and quaternary structure of MBP-YbcN

CD spectroscopy was used to provide information on the secondary structure and folding of MBP-YbcN and GST-YbcN proteins. MBP-YbcN yielded a typical spectrum for a globular protein and appears to be folded correctly (Figure S5A). The predicted proportion of helices, sheet, turn and coil are similar to those observed with MBP-Orf, in keeping with a similar overall fold (Table S2). In contrast, GST-YbcN appears rather more disordered (Figure S5A) and secondary structure predictions do not differ significantly from GST protein alone (Table S2). However, the protein does appear to be functional in binding to SSB (Figure 3) and ssDNA (Figure S5A, lanes j-l).

Size exclusion chromatography was employed to investigate the oligomeric state of MBP-YbcN and GST-YbcN proteins. MBP-YbcN eluted with a relative molecular mass of 113 kDa in agreement with the 120 kDa predicted for a homodimer (Figure S5B). The dimeric species is relatively stable and can be detected as an ~170 kDa species in SDS-polyacrylamide gels (Figure S5B). Hence both YbcN and λ Orf [13] exist as a homodimer in solution. Gel filtration analysis of GST-YbcN, however, identified three major peaks that could correspond to tetramer, octamer and a larger multimeric species (Figure S5B). Stable tetrameric species of 140-150 kDa (a monomer of GST-YbcN is expected to yield a 44.5 kDa protein) were observed in SDS polyacrylamide gels (Figure S5B). Since GST is known to dimerize [28], higher molecular mass complexes consisting of dimer-dimer combinations are not unexpected.

MBP-Orf substitution mutants flanking the central cavity

A number of site-directed mutants were constructed in λ Orf to help identify residues critical for interaction with ssDNA (Figure 4). Alanine substitution mutants were introduced in an aromatic residue projecting into the central cavity (W50A), as well as two nearby polar residues at the channel entrance (Q45A, K48A). These residues match several that are conserved within the Orf family (Figure 1B and Figure S2). More radical glutamic acid substitutions (R103E, V106E) were generated further from the cavity entrance that could also potentially participate in stabilizing contacts with DNA (Figure 4). The latter two mutations were also chosen as they lie in a motif (102-SRMRVGE-108) that resembles a region at the subunit-subunit interface in the *E. coli* RecA filament (25-SIMRLGE-31) [29,30] and could potentially mimic a RecA monomer to nucleate polymerization as a recombination mediator. A related motif has been identified in BRCA2 that promotes assembly of Rad51, the human RecA ortholog, at ssDNA gaps [29,31]. An additional alanine substitution was introduced in the C-terminal helix (W137A) as it was anticipated that it would behave similarly to mutants obtained previously in this region [13], which are known to impair DNA binding to linear ssDNA and dsDNA. All mutant proteins were purified as fusions

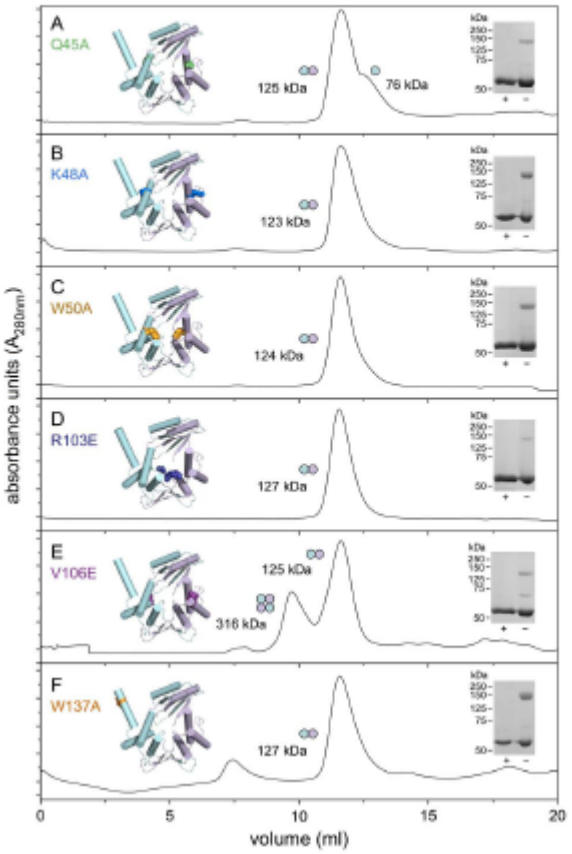

Figure 4. Size-exclusion chromatography of MBP-Orf mutant proteins. Proteins (1 mg/ml) were applied to a 24 ml Superose 12 HR 10/30 column. In each panel, the crystal structure of Orf protein shows the location of the relevant substitution mutant. Chain A is colored cyan and Chain B, lilac; the projecting C-terminal helix is absent from the latter. Predicted molecular weights for MBP-Orf are 59.6 kDa (monomer) and 119.2 kDa (dimer). Oligomeric states are depicted with a circle representing a single subunit and placed adjacent to the corresponding peak. Boiled (+) and unboiled (–) samples of each purified protein were separated on 12.5% SDS-PAGE and stained with Coomassie blue.
doi:10.1371/journal.pone.0102454.g004

with maltose binding protein to improve Orf solubility, simplify the recovery of milligram quantities of pure protein and facilitate far-western assays for protein-protein interactions.

Quaternary and secondary structure of MBP-Orf mutants and effect on SSB binding

To confirm that the six mutations did not substantially affect dimerization or folding, size exclusion chromatography, static light scattering experiments and circular dichroism (CD) spectroscopy were performed. The quaternary structure of each mutant was assessed by size exclusion chromatography. The results reveal that the mutant proteins tend to elute as a single major peak of 123-127 kDa (Figure 4) in keeping with the 120 kDa predicted for an MBP-Orf homodimer. Dimeric species of 135-140 kDa resistant to denaturation were also observed on SDS-PAGE when purified samples were separated in the absence of a reducing agent and without boiling (Figure 4). The Q45A mutant showed an additional peak in gel filtration that corresponded to the size of

a monomer (Figure 4A) that may indicate that this residue normally contributes to dimerization, although this may not be a direct effect since it is located far from the subunit interface. Gel filtration analysis on V106E revealed two major peaks with the larger species consistent with formation of a pair of Orf dimers (Figure 4E). However, when samples were reexamined at pH 7, instead of pH 8, a single dimeric peak at 125 kDa was observed at both 0.5 mg/ml or 1 mg/ml protein (data not shown). Thus electrostatic interactions may be important in the formation of multimeric complexes in this mutant.

Each of the mutants was also examined by static light scattering, which determines the particle size in solution and provides an estimate of the oligomeric status of a protein sample. MBP-Orf yielded a 6.1 nm particle radius and similar sizes of 5.2-5.9 nm were obtained with Q45A, K48A, W50A, R103E and W137A (Figure 5A and Table S3). These correspond to an average molecular mass of 188 kDa, about 1.6 times that of an MBP-Orf dimer, although estimates are based on the assumption that a particle is spherical. The smallest dimension of 5.2 nm was obtained with R103E potentially indicating some conformational change affecting particle shape (Table S3). The V106E mutant differed from the other mutants (Figure 5A) in producing a large particle size with a radius of 114.9 nm (Table S3). It also showed a high percentage polydispersity of 55.8% suggestive of protein aggregation. Under these conditions, MBP also produced an apparent multimer of 17.9 nm (Table S3), which is in accord with previous gel filtration data indicating it exists as a large protein complex [13].

The mutant MBP-Orf proteins were also studied by CD spectroscopy to assess secondary structure and folding properties. All six mutants appear to be folded correctly (Figure 5B) and secondary structure predictions suggest that the R103E, V106E and W137A do not differ significantly from wt MBP-Orf (Table S2). The three mutants clustered around the central channel on α2 of the Orf dimer, Q45A, K48A and W50A, do however show differences in the proportion of predicted α-helix, β-sheet and turn (Table S2). K48A showed the most dissimilar CD spectrum in the far-UV region (Figure 5B) with a significantly increased percentage of α-helix and a corresponding decrease in both β-sheet and turn (Table S2). Hence, although these three proteins are folded and dimerize normally, the alanine substitutions may affect the local conformation of α2 and adjacent secondary structures.

Taken together the data do not reveal any substantial changes in mutant protein folding or homodimer formation, although subtle changes in secondary structure were observed with Q45A, K48A and W50A. V106E has a propensity to form multimers as noted in gel filtration and static light scattering experiments. In the case of the latter mutant it is possible that the presence of an additional negative charge on the surface encourages protein-protein interactions between Orf dimers that are generally positively charged over their protein surface.

Far-western blotting experiments were performed to determine whether any of the Orf mutants affected the association with *E. coli* SSB. None of the mutants showed any reduction in interaction with SSB (Figure S6), suggesting that contacts between the two proteins must be located elsewhere.

Effect of MBP-Orf mutations on DNA binding

The impact of the six MBP-Orf substitution mutants (Figure 6A) on ssDNA and bubble DNA binding was examined in electrophoretic mobility shift and fluorescence anisotropy assays. Binding to a 50 nt radiolabeled linear ssDNA was examined first (Figure 6B) in the absence of metal ions. MBP-Orf wt forms a single major complex with the ssDNA (Figure 6B, lanes a-b), with

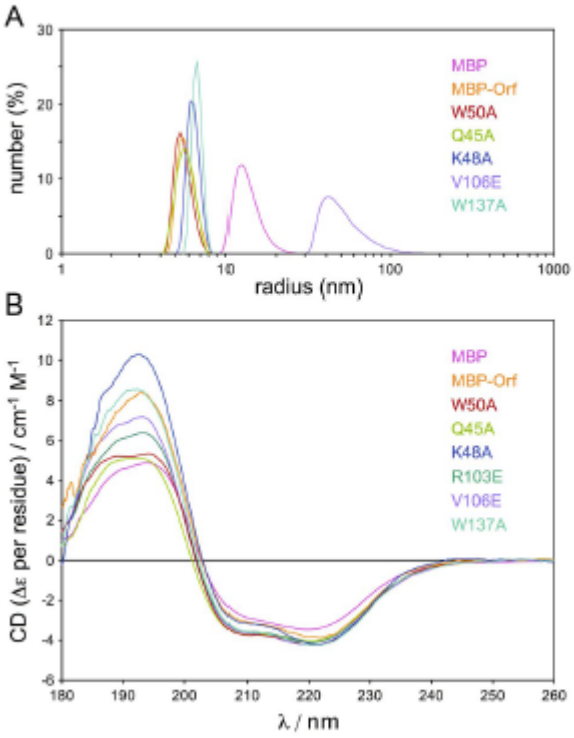

Figure 5. CD spectra and SLS analysis of MBP-Orf mutant proteins. (**A**) SLS analysis of selected MBP-Orf mutant proteins. Proteins at 0.2 mg/ml in 100 mM Tris-HCl pH7 at 25°C were analyzed in a Zetasizer μV. (**B**) CD analysis of MBP-Orf mutant proteins. CD spectra (180-260 nm) were obtained for MBP-Orf proteins in ultrapure water at 20°C and analysis performed using CDSSTR, Contin/LL and Selcon3 programs from the CDPro suite [54].
doi:10.1371/journal.pone.0102454.g005

smearing consistent with an unstable interaction with the substrate as noted previously [13]. The Q45A mutant showed a minor defect in binding as evidenced by the increased presence of unbound substrate (Figure 6B, lane d) relative to the wt. W137A also showed a slight reduction in ssDNA binding, similar to that seen with a W141F mutant [13] and consistent with the α5 helix contributing to DNA binding stability. More severe defects in ssDNA binding were evident with K48A and V106E, while W50A bound only weakly (Figure 6B). No DNA binding was detected with the R103E mutant (Figure 6B, lanes j-k).

Gel shift experiments were also performed on dsDNA (DS_{60}) and 20 nt bubble (BB_{20}) substrates (Figure 6C). The results generally mirrored those with SS_{50} (Figure 6B), although they highlighted more severe DNA binding defects with several of the mutants. While wt MBP-Orf binds relatively poorly to DS_{60} (Figure 6C, lane b), K48A, W50A, R103E and V106E displayed little or no association with this substrate (Figure 6C, lanes d-g). Similarly poor binding was observed with the bubble DNA, although K48A did bind a small proportion of BB_{20} but only as a much faster-migrating complex (Figure 6C, lane l). Q45A and W137A showed less extreme defects on DS_{60} (Figure 6C, lanes c and h). However, the stability of complexes detected on BB_{20} was reduced slightly with W137A (Figure 6C, lane p) and significantly with Q45A (Figure 6C, lane k) compared to the wt. Interestingly, Q45A showed an inability to form the two closely-migrating

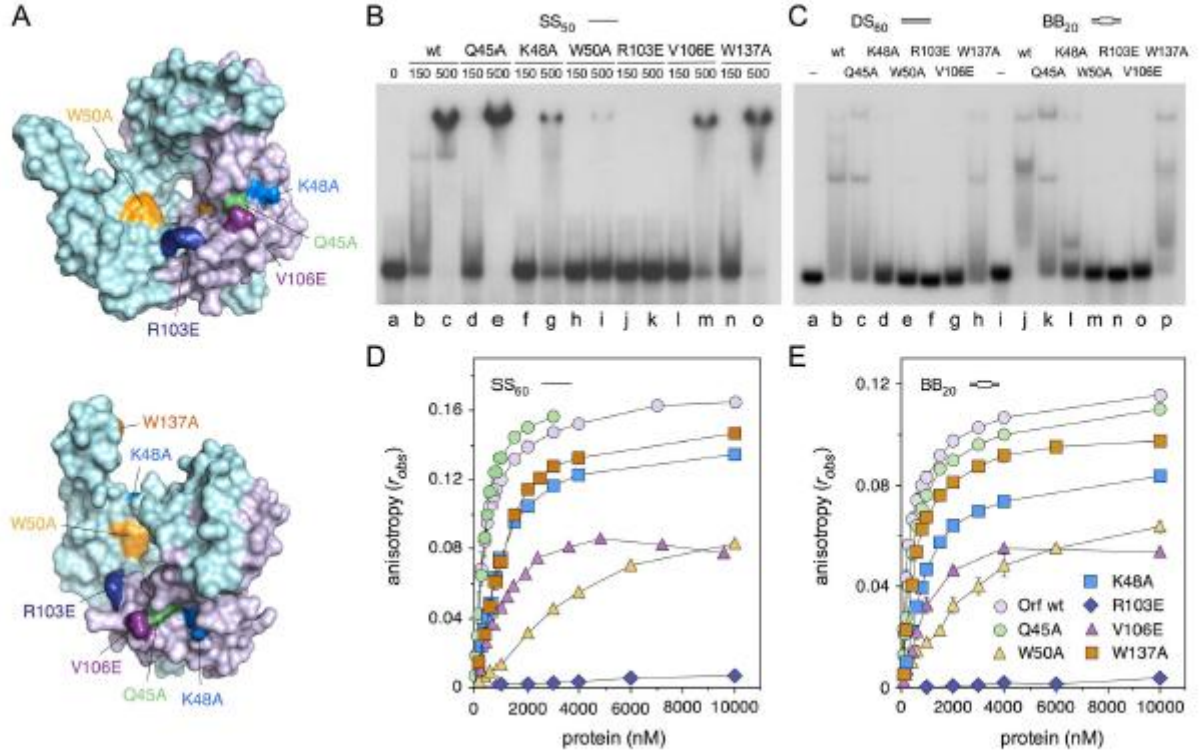

Figure 6. DNA binding by MBP-Orf mutant proteins. (**A**) Crystal structure of Orf protein (1PC6) highlighting the location of site-directed mutations. Chain A is colored cyan and Chain B, lilac. The projecting C-terminal helix is absent from the latter. (**B**) MBP-Orf mutant binding to ssDNA. Gel mobility shift assay contained 0.3 nM ^{32}P-labelled ssDNA (SS_{60}) and protein at 150 and 500 nM as indicated. Samples were separated on 4% polyacrylamide gel electrophoresis and visualized by autoradiography. (**C**) MBP-Orf mutant binding to double-stranded and bubble DNA. The gel shift assay contained 0.15 nM ^{32}P-labelled dsDNA (DS_{60}; lanes a-h) or bubble DNA (BB_{20}; lanes I-p) and 250 nM MBP-Orf protein. Samples were separated by 4% polyacrylamide gel electrophoresis and visualized by autoradiography. (**D** and **E**) MBP-Orf mutant binding to 10 nM fluorescein-labeled ssDNA (D) and BB_{20} DNA (E) as determined by fluorescence anisotropy. Binding isotherms are colored as in (A). Data are the mean and standard deviation of two independent experiments.
doi:10.1371/journal.pone.0102454.g006

complexes seen with the wt and W137A on BB_{20} (Figure 6C, compare lanes j, k and p).

Fluorescence anisotropy was again used to give quantitative information on binding to SS_{60} and BB_{20} substrates. Binding isotherms are presented (Figure 6D-E) alongside dissociation constants derived from the data (Table 1). The results match well those obtained in gel shift assays (Figure 6C) with gradually decreasing DNA binding activity in the order wt, Q45A, W137A, K48A, V106E, W50A and R103E. However, differences were noted within and between the two substrates that were particularly evident in the estimated K_D for each mutant protein (Table 1). In all cases, Q45A showed only a slight defect in DNA binding relative to wt. Similarly, R103E was the most severely deficient in DNA binding, followed by W50A. The remaining three mutants varied in order depending on the substrate. V106E bound better than K48A and W137A to SS_{60}, whereas W137A bound more tightly than V106E and K48A to BB_{20} (Table 1). W50A also showed a preference for binding to BB_{20} over SS_{60}. The observed anisotropy gradually decreased in the V106E mutant at concentrations above 4 μM with both SS_{60} and BB_{20} (Figure 6D-E). This could be due to a contaminant nuclease destroying the substrate, but it is more likely that V106E aggregates as the amount of protein increases resulting in a reduction in protein available for DNA binding. This conclusion fits with the data in Figures 4E and 5A showing that V106E has a tendency to form higher molecular mass complexes.

Discussion

Phage λ Orf substitutes for the *E. coli* RecFOR proteins in genetic exchange reactions, possibly by directing assembly of either bacterial RecA or phage β recombinases onto ssDNA coated with SSB [8,11]. Consistent with this role, purified Orf interacts with both *E. coli* SSB and ssDNA substrates [12,13]. In this study we identified a family of Orf recombinases from diverse lambdoid phages sharing similarity in primary sequence, predicted secondary structure and genomic context within a conserved module located between replication and lysis genes. The authenticity of the proposed family relationship was investigated by studying the biochemical properties of YbcN, one of the most distantly-related members of the Orf family. YbcN is encoded by the *E. coli* cryptic lambdoid prophage DLP12, is 97% identical to a protein from phage 82 [18] and has previously been shown to bind preferentially to abasic sites and T:G mismatches, with the capacity to flip mispaired bases out of the double helix [15]. However, an involvement of YbcN in mismatch recognition does not seemingly fit with an activity responsible for suppressing the recombination and UV light repair deficiencies of *recFOR*

mutants. The specific DNA binding properties of YbcN were therefore examined in direct comparison with λ Orf.

YbcN shared with Orf a preference for binding to short 50 or 60 nt ssDNA substrates rather than duplexes of the same length and sequence. However, YbcN bound much less well to ssDNA than Orf in the absence of metal ions, whereas their binding profiles were more similar in the presence of Mg^{2+} ions (conditions more closely resembling those found in the cell). No difference in binding was detected between a fully complementary dsDNA and one containing a single T:G or G:G mismatch with YbcN protein, although Orf did show a slight preference for the T:G mispair. YbcN showed no improved binding to related substrates containing a bubble of 5, 13 or 20 nt in gel shift assays, although a slight preference for 20 nt bubble DNA was evident in fluorescence anisotropy assays. In contrast, Orf bound much more tightly to these bubble DNA structures in keeping with its ability to load on ssDNA. It is possible that YbcN has a larger footprint than Orf and therefore requires a longer stretch of ssDNA to assemble in a stable complex. Alternatively, it may be that a free end is necessary for loading of YbcN, and that the flanking duplexes in these bubble substrates prevented assembly. The DNA binding properties of YbcN, and the observation that it forms homodimers, are in accord with a recombinase that is homologous to λ Orf and are consistent with a role at the early stages of recombination, binding to ssDNA exposed by phage exonuclease or targeting replication forks or bubbles in the template.

A high-throughput screen of *E. coli* protein-protein interactions had previously indicated an association between YbcN and SSB [32]. Far western blotting, ELISA and yeast two-hybrid experiments confirmed this interaction, with YbcN showing a higher binding affinity for its bacterial partner than Orf. These results confirm that a distantly-related Orf family member has the capacity to bind bacterial SSB and supports the notion that they form a conserved, functional complex. Multiple members of the replication, recombination and repair apparatus have been shown to target the negatively-charged C-terminus of SSB [25], presumably to promote delivery at sites requiring their action. Previous studies had revealed that the terminal 10 residues of SSB were dispensable for λ Orf binding [13]. In contrast, YbcN showed significantly reduced binding to SSB mutant proteins lacking this region (ΔC10) or carrying an SSB113 mutation, known to disrupt interactions with partner proteins [33,34]. Some residual binding of YbcN to SSB was detected by far western blotting and yeast two-hybrid assays, suggesting there may be other portions of the flexible C-terminus of SSB (residues 113-177) or elsewhere that are more important for the Orf-SSB association. It may be that the region on SSB targeted by Orf contributes to remodeling of the SSB-ssDNA complex that allows loading of recombinase onto the exposed ssDNA. It is noteworthy that while both YbcN and Orf interact specifically with *E. coli* SSB, they appear to do so by recognizing disparate features on their bacterial partner.

Although there are differences in YbcN and Orf binding to DNA, their preference for single-stranded DNA does support the prediction that they belong to a conserved family with a comparable role in genetic recombination. It is difficult to reconcile any specificity for abasic sites and mismatches [15] with an activity that suppresses defects associated with mutation of the *recF*, *recO* or *recR* genes, unless the Orf domain has been co-opted to generate a base-flipping enzyme in YbcN. Neither Orf nor YbcN share any homology with enzymes that might incise DNA at a mismatch (e.g. a DNA glycosylase or apurinic/apyrimidinic endonuclease) or to modify any flipped out bases (e.g. a methyltransferase). Moreover, under the assay conditions used here, no specificity for G:G or T:G mismatches was detected with YbcN protein. One possibility is that Orf and YbcN can in certain circumstances capture and stabilize exposed DNA bases as part of their normal function in recognizing ssDNA for assembly of partner recombinases. Such an activity may ensure that a sufficient amount of ssDNA is released from SSB-ssDNA complexes to permit nucleation of a partner recombinase. RecOR are known to accelerate nucleation of RecA on SSB-coated ssDNA, apparently by trapping transient states where an SSB tetramer slides or unwraps from the DNA strand [35]. Recognition of SSB by YbcN and Orf, as with RecO [36,37], would allow targeting of appropriate ssDNA substrates even if the protein-protein interactions do not directly contribute to SSB dissociation [36].

The prophage encoded YbcN could contribute directly to homologous recombination reactions in *E. coli*, although experiments with *rusA* [18], which lies downstream of *ybcN* in DLP12, suggest that insertion mutations that supply an active promoter upstream of this operon are needed for adequate gene expression at this locus. However, the presence of homology with other lambdoid phages in this region does mean that *ybcN* can serve as a repository of alternative phage genetic material that can be picked up by incoming phage, thus contributing to the ongoing evolution of mosaic genomes [38]. The presence of homologs of the defective prophage DLP12 *ybcN* gene in functional phage genomes, including 82 and φ80, suggests that it does specify an active Orf-like protein.

The crystal structure of Orf [12] suggested two possible ssDNA binding modes, either negotiating the narrow (8 Å) central channel or bound within a shallow groove running perpendicular to this cavity. Binding to a gapped duplex substrate, and the bubble DNA substrates used in this study, implied that the latter interaction was the genuine one [13]. However, the experiments did not exclude the possibility that the dimeric ring opens to permit ssDNA access. Electrostatic potential calculations performed on the Orf crystal structure support this possibility as the interior of the toroid is highly positively charged [12]. Furthermore, several of the most highly conserved residues in the Orf family line the walls of this cavity. The identification of a new DNA binding fold in the N-terminal portion of the Orf structure also fits with this idea. This domain, termed RAGNYA [14], comprises two α-helices (α1 from each monomer in Orf) and four β-strands (β1 and β2 from each Orf subunit) with the predicted DNA binding surface across the face of the sheet region and within the central cavity. It may be that additional contacts involve the C-terminal helix, which protrudes from either side of the dimeric ring (Figure 4) and has previously been shown to participate in DNA binding [13].

Mutants in Orf at several conserved sites within or at the entrance to the central channel, alongside a substitution in the C-terminal helix, were studied to help clarify their contribution to ssDNA recognition. The W50A mutant protein exhibited a significant defect in DNA binding providing the first clear evidence that the central cavity is important for stabilizing contacts with ssDNA. Trp50 projects from both sides into the space within the dimer interior and could stabilize binding to nucleotide bases as they enter the channel. Lys48 is located nearby on the cavity rim and a K48A mutation also conferred a significant defect in DNA binding. This residue may make contact with α-helix 5 and either directly or indirectly affect ssDNA that may traverse the groove between this helix and the main body of the dimer, potentially allowing ssDNA to wrap around the toroid. Surprisingly, the Q45A mutant displayed only a minor defect in DNA binding activity compared to the wt, despite being highly conserved among the Orf family. It may play a more subtle role as

the formation of protein:DNA complexes in gel shift assays showed some variation from the wt on both bubble and dsDNA structures.

Mutation of Trp137 to alanine resulted in a modest reduction in DNA binding, comparable to other substitution mutants in the Orf C-terminal helix (R132A, R140A and W141F) [13]. This entire helix is flexible and could adopt a variety of conformations to clamp onto DNA or associate with the RAGNYA domain as suggested previously [13]. The two remaining mutants, R130E and V106E, are located close to the cavity entrance. V106E showed a significant defect in DNA binding, while R103E was completely defective in binding any of the DNA substrates tested. The V106E mutant had a tendency to aggregate and thus it may serve in a structural role, stabilizing intrasubunit interactions between helices α2 and α4. Addition of a negatively-charged residue close to the rim of the channel could also disrupt electrostatic interactions between Orf and the DNA backbone. The R103E mutant could behave in a similar fashion as it lies at each end of the central channel and could be critical for stabilizing contacts with ssDNA. Arg103 in one of the Orf monomers forms a salt bridge with both Asp68 and Glu70 from the other subunit, however, the R103E mutant does appear to fold and dimerize normally suggesting that it does not significantly disrupt the dimer interface. If the Orf protein can open as a clamp, the R103E substitution could potentially affect opening and closing of the ring preventing its association with DNA. Arg103 and Val106 lie in a short motif that resembles part of the subunit interface in RecA [29,30] and this region protrudes on both sides of the Orf cavity, ideally situated to aid nucleation of RecA onto ssDNA. However, as yet, we have been unable to confirm a physical interaction between Orf and RecA.

The assembly of Orf on ssDNA *in vivo* could facilitate loading of either the bacterial RecA strand exchange protein or phage β strand annealing protein to stimulate recombination [1]. Assuming that Orf behaves similarly to *recF*, *recO* and *recR*, the genes it substitutes for *in vivo* [8,11], it could help nucleate RecA at replication forks, D-loops (similar to bubble structures) or at gaps resected in duplex DNA. This scenario would be analogous to bacteriophage T4, whereby the UvsY mediator promotes assembly of the UvsX recombinase (a RecA homolog) onto ssDNA, overcoming the block presented by the phage Gp32 ssDNA binding protein [39]. RecFOR are known to help RecA filaments assemble at dsDNA-ssDNA junctions [40–42], destabilizing SSB-ssDNA complexes that normally prevent access to the template [35]. The fact that RecFOR promote RecA filament assembly at gaps with either DNA or RNA at the 5′ terminus suggests that the stalled replication fork may well be a target for their mediator function [43]. Facilitated loading of RecA in this context could sanction lesion bypass and replication restart downstream of the block. Restoration pathways that depend on fork regression have been suggested [44,45] and there is evidence that RecFOR play an important role [46] that can function independently of RuvAB-mediated Holliday junction processing [47]. If Orf could stimulate a similar recovery pathway, even by simply safeguarding the fork, this might explain why it can also partially substitute for the activities of RuvABC in Holliday junction branch migration and resolution [11].

Orf suppresses the recombination and DNA repair defects of *recFOR* mutants even in the absence of Redαβ [8,11]. However, recent experiments examining recombinational exchanges between partially homologous partners on phages and prophages, suggest that Orf is required primarily for exchanges promoted by the Red system rather than by RecA [38]. It is possible that RecFOR normally prevail over Orf in mediating the loading of RecA onto ssDNA bound by SSB, whereas the bacterial activities are less active in situations involving the Red system (e.g. at DNA ends). Depending on the context and availability of recombination enzymes, it may be that Orf aids nucleation of either phage Redβ or bacterial RecA to promote recombinational exchange. Both RecA (strand invasion) and Red (annealing) pathways [48] can promote the recovery of viable phage genomes, although the latter may lead to more rearrangements at sites of limited sequence homology, thus fostering phage genetic diversity and mosaicism [38].

Materials and Methods

Sequence analysis

Proteins related to Orf were identified using the TBLASTN and PSI-BLAST algorithms [49] on the non-redundant database at the National Center for Biotechnology Information (NIH, Bethesda, MD). An inclusion threshold of 0.05 was applied in iterative searches to restrict those matches selected in the position-specific weight matrix (PSSM). Sequences were aligned using ClustalW with a Gap Creation Penalty of 10 and Gap Extension Penalty of 1 [50]. Phylogenetic relationships were estimated by the neighbor-hood-joining method incorporated within the PAUP package [51]. The Jnet protein structure prediction tool on the Jpred server was used to estimate secondary structure features of *E. coli* DLP12 YbcN and *S. aureus* ϕETA Orf20 [52]. The Phyre^2 protein fold recognition server is an improved version of 3D-PSSM [53].

MBP and GST fusion constructs

The *orf*-like gene (*orf151* or *ybcN*) from *E. coli* prophage DLP12 [18] was amplified from AB1157 genomic DNA using Pfx DNA polymerase and oligonucleotides 5′-CGGAGGGAATTCATGAACCTCTCAC-3′ and 5′-GACGATTTGGATCCCTGTAGATGTG-3′. The PCR product was digested with EcoRI and BamHI (underlined) and inserted into pMALc2 to give pLB101. This construct expresses *E. coli* MBP (maltose binding protein) fused to the N-terminus of YbcN. An N-terminal GST (glutathione S-transferase) fusion was constructed by transferring the insert from pLB101, using EcoRI and SalI sites, into pGEX-6P-1 (GE Healthcare) to generate pLB108 (GST-YbcN).

The QuickChange (Stratagene) site-directed mutagenesis method was used to generate Q45A (pFC256), K48A (pFC262), W50A (pFC259), R103E (pFC265), V106E (pFC267) and W137A (pEH101) mutant derivatives in pMALc2. The integrity of all cloned genes was confirmed by DNA sequencing.

Proteins

GST, GST-Orf, MBP, MBP-Orf, SSB, SSB113 and SSBΔC10 proteins were purified as described [13]. Purified SSBΔ115-144 protein [27] was kindly provided by Robert Lloyd, University of Nottingham. MBP-YbcN was purified from 1 l of DH5α or BL21 (DE3) Rosetta carrying pLB101 grown in LB broth supplemented with 10 mM glucose and ampicillin at 150 μg/ml and 37 μg/ml chloramphenicol as required. Expression was induced with 0.5 mM IPTG when the culture reached an A_{650nm} of 0.5 and incubation continued for 3 hours at 37°C. Harvested cells were resuspended in column buffer (20 mM Tris-HCl pH7.4, 200 mM NaCl, 1 mM EDTA) and lysed by sonication. The cleared lysate was applied to a 4 ml amylose column (New England Biolabs) in column buffer and bound proteins eluted in the presence of 10 mM maltose. Fractions containing MBP-YbcN were collected and dialyzed in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM DTT, 200 mM KCl, 50% (v/v) glycerol. Aliquots were stored at -80°C and a total of 70 mg of highly pure MBP-YbcN was

recovered at a concentration of 17.5 mg/ml. Additional stocks of MBP-YbcN were purified by both amylose and heparin agarose chromatography, with the majority of protein eluting from the latter matrix at 0.45 M NaCl.

GST-YbcN was purified from 6 l of BL21 Codon Plus cells carrying pLB108. Cells were grown at 25°C in LB broth containing ampicillin at 150 µg/ml and protein expression induced as before. Harvested cells were resuspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 pH 7.3) and lysed in a cell disruptor at 15-25 psi. The supernatant was applied to a 4 ml glutathione sepharose 4B column (GE Healthcare) equilibrated in the same buffer. Fusion proteins were eluted in 50 mM Tris-HCl, 10 mM reduced glutathione pH 8.0 and further purified by passage through heparin agarose and Q sepharose chromatography, collecting the flow-though at each stage. A total of 32 mg GST-YbcN (16.2 mg/ml) was recovered.

MBP-Orf mutant proteins were purified from 500 ml of BL21-AI carrying the appropriate pMALc2 construct as described [13]. MBP-Orf Q45A, W50A, R103E and V106E were further purified by heparin-agarose chromatography after the amylose affinity purification step.

Protein concentrations were determined in a NanoDrop 2000 (Thermo Scientific) micro-volume spectrophotometer and by a modified Bradford assay (BioRad) using bovine serum albumin as a standard; amounts of Orf proteins are expressed as moles of dimer. Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase and Pfx DNA polymerase were obtained from Invitrogen.

DNA substrates

Oligonucleotides used in this study are listed in Table S4. The 50 nt ssDNA substrate (SS_{50}) consisted of o-SS_{50}, which was annealed to its complement (o-DS_{50}) to give a 50 bp dsDNA duplex (DS_{50}). The 60 nt ssDNA substrate (SS_{60}) consisted of o-SS_{60}, which was annealed to its complement (o-DS_{60}) to give a fully complementary 60 bp dsDNA duplex (DS_{60}). Centrally located mismatch (MM) or bubble (BB) structures of 1, 5, 13 and 20 nt ($MM_{G:G}$, $MM_{T:G}$, BB_5, BB_{13} and BB_{20}) were made by annealing o-SS_{60} with o-$MM_{G:G}$, o-$MM_{T:G}$, o-BB_5, o-BB_{13} and o-BB_{20}, respectively. Bent DNA substrates containing 1 (BT_1), 2 (BT_2) or 3 (BT_3) adenine insertions were made by annealing o-SS_{60} with o-BT_1, o-BT_2 and o-BT_3, respectively. In gel shift assays, o-SS_{50} or o-SS_{60} were labeled with [γ^{32}P] ATP at the 5′ end using T4 polynucleotide kinase. Labeled DNA was separated from unincorporated nucleotide using MicroBioSpin columns (BioRad). Annealed substrates were further purified by separation on 10% polyacrylamide gels in 90 mM Tris-borate, 2 mM EDTA. A 60 nt fluorescein 3′-end-labeled ssDNA equivalent in sequence to o-SS_{60} was used in fluorescence anisotropy experiments, alongside DS_{60}, $MM_{T:G}$ and BB_{20} substrates generated by annealing the fluorescein-labeled strand to appropriate partner oligonucleodides.

DNA binding assays

Band shift assays (20 µl) with 50 and 60 nt ^{32}P-labelled DNA substrates were conducted in 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 1 mM dithiothreitol, 5% glycerol, 100 µg/ml BSA. Samples were incubated on ice for 15 min prior to separation on 4% PAGE in 6.7 mM Tris-HCl pH 8.0, 3.3 mM sodium acetate, 2 mM EDTA at 160 V. EDTA was replaced with 1 mM $MgCl_2$ in both binding mixtures and gels to assess protein-DNA interactions in the presence of Mg^{2+} ions. Gels were dried on filter paper and analyzed by autoradiography and phosphorimaging. Data were analyzed using ImageQuant and ImageJ software.

DNA-protein interactions measured by fluorescence anisotropy utilized 3′-fluorescein-labeled derivatives of SS_{60}, DS_{60}, $MM_{T:G}$ and BB_{20} in 100 mM Tris-HCl pH 7.0. Protein was titrated into 1 ml of 10 nM DNA in a 1 ml two-way quartz cuvette (10 mm path length). Changes in anisotropy were monitored on a modified Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) fitted with polarizing filters with excitation at 494 nm and emission at 521 nm. Five replicates were taken with an averaging time of 20 sec. Data were analyzed using DynaFit 3 software to estimate the K_D for MBP-YbcN, MBP-Orf wt and mutant proteins on SS_{60} and BB_{20}.

Size exclusion chromatography

The apparent molecular mass of proteins was estimated using an AKTA FPLC system with a 24 ml Superose 12 HR 10/30 column (GE Healthcare). Proteins (1 mg/ml) were applied to the gel filtration column in 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM DTT, 150 mM KCl at a flow rate of 0.5 ml/min. Molecular mass standards (BioRad) contained thyroglobulin (670 kDa), γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin B12 (1.35 kDa).

Far western blotting

Purified SSB wt and mutant proteins were separated on 15% SDS-PAGE and transferred to a PVDF membrane by electroblotting in 48 mM Tris, 39 mM glycine, 1.3 mM SDS. A prestained protein molecular weight standard (BioRad) served as a size marker. Blots were probed with YbcN or Orf fused to MBP or GST and interactions detected with monoclonal anti-MBP or anti-GST antibodies and mouse IgG peroxidase conjugate (Sigma). Chemiluminescence was observed by exposure to X-ray film following treatment with ECL reagents (GE Healthcare). Appropriate controls, either positive (appropriately tagged YbcN or Orf bound to the membrane) or negative (omitting tagged YbcN/Orf or including purified MBP or GST), were analyzed in parallel. Protein transfer to membranes was monitored by staining with 1% amido black (Sigma) in 10% acetic acid.

ELISA

SSB wt and mutant proteins in PBS were applied to a flat bottom Immulon 2HB microtitre plate (Thermo Scientific) and incubated for 16 h at 37°C to allow binding. Wells were washed in PBS containing 0.3% Tween-20 (Sigma) and incubated for 1 hr at 37°C in this buffer containing 5% milk powder. After washing, GST-Orf or GST-YbcN proteins were added and incubated for 1 hr at 37°C. Wells were washed again prior to addition of HRP-anti-GST antibody (Sigma) at a 1:1000 dilution. Tetramethylbenzidine (Sigma) was added and the intensity of the color reaction measured at A_{620nm} in an Anthos HTII plate reader.

Static light scattering

Proteins at 0.2 mg/ml (50 µl) in 100 mM Tris-HCl pH 7 at 25°C were analyzed in a Zetasizer µV (Malvern Instruments Ltd). Thirteen measurements at 10 sec intervals were recorded and the data set repeated in triplicate for each sample.

Circular dichroism (CD) spectroscopy

Proteins were dialyzed overnight in Slide-A-Lyzer Mini Dialysis Units (Pierce) in ultrapure water at 4°C. Dialyzed samples were centrifuged at 15 000 rpm for 30 minutes to remove any precipitated protein. Protein concentrations were determined by

measurement of the absorbance at 280 nm. Far-UV CD spectra and the corresponding blanks were recorded on a Jasco J-810 Spectropolarimeter in a 1 ml cuvette (0.2 cm path length). Eight accumulations recorded at a rate of 10 nm/min, a pitch of 0.5 nm, a bandwidth of 1 nm and a response time of 2 s, were averaged. After subtraction of the appropriate blank, adaptive smoothing was carried out within the Jasco Spectra Analysis program (version 1.50, Jasco, Great Dunmow, UK). Smoothed data in the range 185-240 nm were analyzed for protein secondary structure using the CDSSTR, SELCON3, and CONTIN/LL programs in the CDPro package [54] with the SP37 reference set of soluble proteins [55]. Secondary structure content of known crystal structures was determined by DSSP [56] and accessed through the sequence retrieval system (SRS) at the European Bioinformatics Institute (http://srs.ebi.ac.uk).

Supporting Information

Figure S1 Phylogram of representative Orf family proteins. The tree was obtained by the neighborhood joining method using *Staphylococcus aureus* phage 187 and *Listeria innocua* sequences lin1258 and lin1740 sequences as the outgroup. *S. typhimurium* phage ST64T and *E. coli* O157:H7 phage 933W were excluded from the phylogenetic tree because they carry incomplete *orf* genes. Only branches with significant values (>70%) are shown. Abbreviations are: *Eco*, *E. coli*; *Sso*, *Shigella sonnei*; *Plu*, *Photorhabdus luminescens*; *Sen*, *Salmonella enterica*; *Bbr*, *Bordetella bronchiseptica*; *Nar*, *Novosphingobium aromaticivorans*; *Mma*, *Magnetospirillum magnetotacticum*; *Psy*, *Pseudomonas syringae*; UPEC, Uropathogenic *E. coli*; *Sty*, *Salmonella enterica* serovar Typhimurium; *Sfl*, *Shigella flexneri*; *Ype*, *Yersinia pestis*; *Ahy*, *Aeromonas hydrophila*; *Hin*, *Haemophilus influenzae*; *Aac*, *Actinobacillus actinomycetemcomitans*; *Ngo*, *Neisseria gonorrhoeae*; *Bvi*, *Burkholderia vietnamiensis*; *Lin*, *Listeria innocua*; *Sau*, *Staphylococcus aureus*. Four groups of closely-related homologs are highlighted in color. The true phylogeny is difficult to ascertain due to the likelihood of gene transfer between phage genomes. Full designations of sources are given in Table S1.
(TIFF)

Figure S2 Structural similarity between λ Orf, DLP12 YbcN and ϕETA Orf20 proteins. Structural models of *E. coli* DLP12 YbcN (residues 16-147), belonging to the PRK09741 domain, and *S. aureus* ϕETA Orf20 (residues 26-130), a member of the DUF968 domain, were based on the λ Orf crystal structure (1PO6) using Phyre[2]. Only subunit A is shown, with the RAGNYA domain colored in yellow. Conserved and potential functionally-equivalent residues located close to the central channel of the dimer are highlighted.
(TIFF)

Figure S3 YbcN binding to single-stranded and bubble DNA. Comparison of MBP-YbcN binding to 10 nM fluorescein-labeled SS_{60} and BB_{20} as determined by fluorescence anisotropy. Data are the mean and standard deviation of two independent experiments.
(TIFF)

Figure S4 YbcN and Orf binding to mismatch and bent DNA. (**A**) Comparison of MBP-YbcN and MBP-Orf binding to 10 nM fluorescein-labeled $MM_{G:G}$ and $MM_{T:G}$ DNA as determined by fluorescence anisotropy. Data are the mean and standard deviation of two independent experiments. (**B**) Comparison of MBP-YbcN and MBP-Orf binding to bent DNA. Gel mobility shift assays contained 125 nM MBP-Orf (O) or MBP-YbcN (Y) proteins, 5 mM EDTA and 0.15 nM of ^{32}P-labelled 60 nt (SS_{60}) ssDNA (lanes a-c), 60 bp (DS_{60}) dsDNA (lanes d-f), 1 nt (BT_1) insertion (lanes g-i), 2 nt (BT_2) insertion (lanes j-l) and 3 nt (BT_3) insertion (lanes m-o).
(TIFF)

Figure S5 Analysis of DLP12 MBP-YbcN and GST-YbcN proteins. (**A**) CD analysis of YbcN proteins and binding to ssDNA. CD spectra (180–260 nm) were obtained for MBP, GST, MBP-YbcN and GST-YbcN proteins in ultrapure water at 20°C. Gel shift assays contained 0.3 nM ^{32}P-labelled ssDNA (SS_{50}), 5 mM EDTA and 62.5, 125 and 250 nM MBP-Orf (lanes b-d) and 250, 500 and 1000 nM MBP-YbcN (lanes f-h) and GST-YbcN (lanes j-l). (**B**) Size-exclusion chromatography of DLP12 YbcN. MBP-YbcN and GST-YbcN proteins (1 mg/ml) were applied to a 24 ml Superose 6HR 10/30 column in 20 mM Tris-HCl pH8, 1 mM EDTA, 0.5 mM DTT, 250 mM KCl. The predicted molecular weights for each protein monomer are 60.4 kDa for MBP-YbcN and 44.6 kDa for GST-YbcN. Oligomeric states are depicted with a circle representing a single subunit and placed adjacent to the corresponding peak. Boiled (+) and unboiled (–) samples of each purified protein were separated on 12.5% SDS-PAGE and stained with Coomassie blue.
(TIFF)

Figure S6 MBP-Orf mutant protein binding to SSB in far western assays. SSB protein (5 µg) separated on 15% SDS-PAGE was blotted and probed with 30 µg MBP-Orf mutant proteins. MBP-Orf-SSB interactions were detected with antibodies specific for the MBP domain.
(TIFF)

Table S1 Representative Orf family proteins.
(PDF)

Table S2 Secondary structure predictions for YbcN and Orf mutant proteins from CD data.
(PDF)

Table S3 Static light scattering analysis of MBP-Orf mutants.
(PDF)

Table S4 Oligonucleotides used to generate DNA substrates.
(PDF)

Acknowledgments

We thank Robert Lloyd for SSBΔ115-144 protein, Peter McGlynn for anti-SSB antibody, Elinor Hanlon for the W137A mutant construct and Michael Parkin, Sarah Francis and Emma Rochester for additional protein purification.

Author Contributions

Conceived and designed the experiments: FC LW RY JS GS. Performed the experiments: FC AM AT MB-B LW LB PR CH RY GS. Analyzed the data: FC AM LW RY JS JH GS. Contributed to the writing of the manuscript: GS JH.

References

1. Poteete AR (2001) What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol Lett 201: 9–14.
2. Murphy KC (2007) The λ Gam protein inhibits RecBCD binding to dsDNA ends. J Mol Biol 371: 19–24.
3. Court R, Cook N, Saikrishnan K, Wigley D (2007) The crystal structure of λ-Gam protein suggests a model for RecBCD inhibition. J Mol Biol 371: 25–33.

4. Court DL, Sawitzke JA, Thomason LC (2002) Genetic engineering using homologous recombination. Annu Rev Genet 36: 361–388.
5. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci USA 98: 6742–6746.
6. Poteete AR (2008) Involvement of DNA replication in phage λ Red-mediated homologous recombination. Mol Microbiol 68: 66–74.
7. Li XT, Thomason LC, Sawitzke JA, Costantino N, Court DL (2013) Bacterial DNA polymerases participate in oligonucleotide recombination. Mol Microbiol 88: 906–920.
8. Sawitzke JA, Stahl FW (1992) Phage λ has an analog of *Escherichia coli recO*, *recR* and *recF* genes. Genetics 130: 7–16.
9. Sawitzke JA, Stahl FW (1997) Roles for λ Orf and *Escherichia coli* RecO, RecR and RecF in lambda recombination. Genetics 147: 357–369.
10. Cox MM (2007) Regulation of bacterial RecA protein function. Crit Rev Biochem Mol Biol 42: 41–63.
11. Poteete AR (2004) Modulation of DNA repair and recombination by the bacteriophage λ Orf function in *Escherichia coli* K-12. J Bacteriol 186: 2699–2707.
12. Maxwell KL, Reed P, Zhang RG, Beasley S, Walmsley AR, et al. (2005) Functional similarities between phage λ Orf and *Escherichia coli* RecFOR in initiation of genetic exchange. Proc Natl Acad Sci USA 102: 11260–11265.
13. Curtis FA, Reed P, Wilson LA, Bowers LY, Yeo RP, et al. (2011) The C-terminus of the phage λ Orf recombinase is involved in DNA binding. J Mol Recognit 24: 333–340.
14. Balaji S, Aravind L (2007) The RAGNYA fold: a novel fold with multiple topological variants found in functionally diverse nucleic acid, nucleotide and peptide-binding proteins. Nucl Acids Res 35: 5658–5671.
15. Chen CS, Korobkova E, Chen H, Zhu J, Jian X, et al. (2008) A proteome chip approach reveals new DNA damage recognition activities in *Escherichia coli*. Nat Methods 5: 69–74.
16. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucl Acids Res 40: D290–301.
17. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, et al. (2013) CDD: conserved domains and protein three-dimensional structure. Nucl Acids Res 41: D348–352.
18. Mahdi AA, Sharples GJ, Mandal TN, Lloyd RG (1996) Holliday junction resolvases encoded by homologous *ruvA* genes in *Escherichia coli* K-12 and phage 82. J Mol Biol 257: 561–573.
19. Shub DA, Goodrich-Blair H, Eddy SR (1994) Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. Trends Biochem Sci 19: 402–404.
20. Sharples GJ, Curtis FA, McGlynn P, Bolt EL (2004) Holliday junction binding and resolution by the Rap structure-specific endonuclease of phage λ. J Mol Biol 340: 739–751.
21. Sharples GJ, Bolt EL, Lloyd RG (2002) RusA proteins from the extreme thermophile *Aquifex aeolicus* and lactococcal phage r1t resolve Holliday junctions. Mol Microbiol 44: 549–559.
22. Heyduk T, Ma Y, Tang H, Ebright RH (1996) Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. Methods Enzymol 274: 492–503.
23. Gohlke C, Murchie AI, Lilley DM, Clegg RM (1994) Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer. Proc Natl Acad Sci USA 91: 11660–11664.
24. Lilley DM (1995) Kinking of DNA and RNA by base bulges. Proc Natl Acad Sci USA 92: 7140–7142.
25. Shereda RD, Kozlov AG, Lohman TM, Cox MM, Keck JL (2008) SSB as an organizer/mobilizer of genome maintenance complexes. Crit Rev Biochem Mol Biol 43: 289–318.
26. Chase JW, L'Italien JJ, Murphy JB, Spicer EK, Williams KR (1984) Characterization of the *Escherichia coli* SSB-113 mutant single-stranded DNA-binding protein: cloning of the gene, DNA and protein sequence analysis, high pressure liquid chromatography peptide mapping, and DNA-binding studies. J Biol Chem 259: 805–814.
27. Moore T, McGlynn P, Ngo HP, Sharples GJ, Lloyd RG (2003) The RdgC protein of *Escherichia coli* binds DNA and counters a toxic effect of RecFOR in strains lacking the replication restart protein PriA. EMBO J 22: 735–745.
28. Fabrini R, De Luca A, Stella L, Mei G, Orioni B, et al. (2009) Monomer-dimer equilibrium in glutathione transferases: a critical re-examination. Biochemistry 48: 10473–10482.
29. Pellegrini L, Yu DS, Lo T, Anand S, Lee M, et al. (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. Nature 420: 287–293.
30. Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature 453: 489–484.
31. Jensen RB, Carreira A, Kowalczykowski SC (2010) Purified human BRCA2 stimulates RAD51-mediated recombination. Nature 467: 678–683.
32. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, et al. (2006) Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. Genome Res 16: 686–691.
33. Cadman CJ, McGlynn P (2004) PriA helicase and SSB interact physically and functionally. Nucl Acids Res 32: 6378–6387.
34. Yuzhakov A, Kelman Z, O'Donnell M (1999) Trading places on DNA – a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. Cell 96: 153–163.
35. Bell JC, Plank JL, Dombrowski CC, Kowalczykowski SC (2012) Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. Nature 491: 274–278.
36. Umezu K, Kolodner RD (1994) Protein interactions in genetic recombination in *Escherichia coli*: interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. J Biol Chem 269: 30005–300013.
37. Kantake N, Madiraju MV, Sugiyama T, Kowalczykowski SC (2002) *Escherichia coli* RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: a common step in genetic recombination. Proc Natl Acad Sci USA 99: 15327–15332.
38. De Paepe M, Hutinet G, Son O, Amarir-Bouhram J, Schbath S, et al. (2014) Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases. PLoS Genet 10: e1004181.
39. Liu J, Morrical SW (2010) Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. Virol J 7: 357.
40. Morimatsu K, Kowalczykowski SC (2003) RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell 11: 1337–1347.
41. Sakai A, Cox MM (2009) RecFOR and RecOR as distinct RecA loading pathways. J Biol Chem 284: 3264–3272.
42. Handa N, Morimatsu K, Lovett ST, Kowalczykowski SC (2009) Reconstitution of initial steps of dsDNA break repair by the RecF pathway of *E. coli*. Genes Dev 23: 1234–1245.
43. Morimatsu K, Wu Y, Kowalczykowski SC (2012) RecFOR proteins target RecA protein to a DNA gap with either DNA or RNA at the 5' terminus: implication for repair of stalled replication forks. J Biol Chem 287: 35621–35630.
44. Higgins NP, Kato K, Strauss B (1976) A model for replication repair in mammalian cells. J Mol Biol 101: 417–425.
45. Robu ME, Inman RB, Cox MM (2001) RecA protein promotes the regression of stalled replication forks *in vitro*. Proc Natl Acad Sci USA 98: 8211–8218.
46. Courcelle J, Donaldson JR, Chow KH, Courcelle CT (2003) DNA damage-induced replication fork regression and processing in *Escherichia coli*. Science 299: 1064–1067.
47. Donaldson JR, Courcelle CT, Courcelle J (2004) RuvAB and RecG are not essential for the recovery of DNA synthesis following UV-induced DNA damage in *Escherichia coli*. Genetics 166: 1631–1640.
48. Stahl MM, Thomason L, Poteete AR, Tarkowski T, Kuzminov A, et al. (1997) Annealing vs. invasion in phage lambda recombination. Genetics 147: 961–977.
49. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res 25: 3389–3402.
50. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 31: 3497–3500.
51. Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4 Sinauer Associates, Sunderland, Massachusetts.
52. Cuff JA, Barton GJ (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. Proteins 34: 508–519.
53. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4: 363–371.
54. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287: 252–260.
55. Johnson WC (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins 35: 307–312.
56. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.

Bibliography

1. Sulakvelidze, A., Z. Alavidze, and J.G. Morris, Jr., *Bacteriophage therapy.* Antimicrob Agents Chemother, 2001. **45**(3): p. 649-59.
2. Twort, F.W., *An investigation on the nature of ultra-microscopic viruses by Twort FW, L.R.C.P. Lond., M.R.C.S. (From the Laboratories of the Brown Institution, London).* Bacteriophage, 2011. **1**(3): p. 127-129.
3. d'Herelle, M.F., *On an invisible microbe antagonistic to dysentery bacilli. Note by M. F. d'Herelle, presented by M. Roux. Comptes Rendus Academie des Sciences 1917; 165:373–5.* Bacteriophage, 2011. **1**(1): p. 3-5.
4. Juhala, R.J., et al., *Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages.* J Mol Biol, 2000. **299**(1): p. 27-51.
5. Neely, M.N. and D.I. Friedman, *Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release.* Mol Microbiol, 1998. **28**(6): p. 1255-67.
6. Miyamoto, H., et al., *Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions.* DNA Res, 1999. **6**(4): p. 235-40.
7. Plunkett, G., 3rd, et al., *Sequence of Shiga toxin 2 phage 933W from Escherichia coli O157:H7: Shiga toxin as a phage late-gene product.* J Bacteriol, 1999. **181**(6): p. 1767-78.
8. Vander Byl, C. and A.M. Kropinski, *Sequence of the genome of Salmonella bacteriophage P22.* J Bacteriol, 2000. **182**(22): p. 6472-81.
9. Hendrix, R.W., et al., *Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 2192-7.
10. Highton, P.J., Y. Chang, and R.J. Myers, *Evidence for the exchange of segments between genomes during the evolution of lambdoid bacteriophages.* Mol Microbiol, 1990. **4**(8): p. 1329-40.
11. Hendrix, R.W., *Bacteriophages: evolution of the majority.* Theor Popul Biol, 2002. **61**(4): p. 471-80.
12. Campbell, A., *Comparative molecular biology of lambdoid phages.* Annu Rev Microbiol, 1994. **48**: p. 193-222.
13. Gyles, C. and P. Boerlin, *Horizontally transferred genetic elements and their role in pathogenesis of bacterial disease.* Vet Pathol, 2014. **51**(2): p. 328-40.
14. Ziebuhr, W., et al., *Evolution of bacterial pathogenesis.* Cell Mol Life Sci, 1999. **56**(9-10): p. 719-28.
15. Szabo, M., et al., *Importance of illegitimate recombination and transposition in IS30-associated excision events.* Plasmid, 1999. **42**(3): p. 192-209.
16. Kreuzer, K.N., *Interplay between DNA replication and recombination in prokaryotes.* Annu Rev Microbiol, 2005. **59**: p. 43-67.
17. Kowalczykowski, S.C., *Initiation of genetic recombination and recombination-dependent replication.* Trends Biochem Sci, 2000. **25**(4): p. 156-65.
18. Serra-Moreno, R., et al., *Use of the lambda Red recombinase system to produce recombinant prophages carrying antibiotic resistance genes.* BMC Mol Biol, 2006. **7**: p. 31.
19. Brussow, H., C. Canchaya, and W.D. Hardt, *Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion.* Microbiol Mol Biol Rev, 2004. **68**(3): p. 560-602, table of contents.
20. Campbell, A., *The future of bacteriophage biology.* Nat Rev Genet, 2003. **4**(6): p. 471-477.
21. Waldor, M.K., *Bacteriophage biology and bacterial virulence.* Trends Microbiol, 1998. **6**(8): p. 295-7.
22. Miao, E.A. and S.I. Miller, *Bacteriophages in the evolution of pathogen-host interactions.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9452-4.

23. Wagner, P.L. and M.K. Waldor, *Bacteriophage control of bacterial virulence.* Infect Immun, 2002. **70**(8): p. 3985-93.
24. Boyd, E.F. and H. Brussow, *Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved.* Trends Microbiol, 2002. **10**(11): p. 521-9.
25. Perna, N.T., et al., *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.* Nature, 2001. **409**(6819): p. 529-33.
26. Waldor, M.K. and J.J. Mekalanos, *Lysogenic conversion by a filamentous phage encoding cholera toxin.* Science, 1996. **272**(5270): p. 1910-4.
27. Buck, G.A., et al., *DNA relationships among some tox-bearing corynebacteriophages.* Infect Immun, 1985. **49**(3): p. 679-84.
28. Barondess, J.J. and J. Beckwith, *A bacterial virulence determinant encoded by lysogenic coliphage lambda.* Nature, 1990. **346**(6287): p. 871-4.
29. Barondess, J.J. and J. Beckwith, *bor gene of phage lambda, involved in serum resistance, encodes a widely conserved outer membrane lipoprotein.* J Bacteriol, 1995. **177**(5): p. 1247-53.
30. Clark, A.J., *Toward a metabolic interpretation of genetic recombination of E. coli and its phages.* Annu Rev Microbiol, 1971. **25**: p. 437-64.
31. Weinstock, G.M., *Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed.* 1996: p. 1034.
32. Court, D.L., J.A. Sawitzke, and L.C. Thomason, *Genetic engineering using homologous recombination.* Annu Rev Genet, 2002. **36**: p. 361-88.
33. Cox, M.M., et al., *The importance of repairing stalled replication forks.* Nature, 2000. **404**(6773): p. 37-41.
34. Kowalczykowski, S.C., et al., *Biochemistry of homologous recombination in Escherichia coli.* Microbiol Rev, 1994. **58**(3): p. 401-65.
35. Cox, M.M., *A broadening view of recombinational DNA repair in bacteria.* Genes Cells, 1998. **3**(2): p. 65-78.
36. Sharples, G.J., *The X philes: structure-specific endonucleases that resolve Holliday junctions.* Mol Microbiol, 2001. **39**(4): p. 823-34.
37. Holliday, R., *A mechanism for gene conversion in fungi.* Genetics, 1964.
38. Lam, S.T., et al., *Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity.* Genetics, 1974. **77**(3): p. 425-33.
39. McMilin, K.D., M.M. Stahl, and F.W. Stahl, *Rec-mediated recombinational hot spot activity in bacteriophage lambda. I. Hot spot activity associated with spi-deletions and bio substitutions.* Genetics, 1974. **77**(3): p. 409-23.
40. Stahl, F.W., et al., *The distribution of crossovers along unreplicated lambda bacteriophage chromosomes.* Genetics, 1974. **77**(3): p. 395-408.
41. Gopaul, D.N., F. Guo, and G.D. Van Duyne, *Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination.* The EMBO Journal, 1998. **17**(14): p. 4175-4187.
42. Sheridan, S.D., et al., *A comparative analysis of Dmc1 and Rad51 nucleoprotein filaments.* Nucleic Acids Res, 2008. **36**(12): p. 4057-66.
43. Cox, M.M., *Regulation of bacterial RecA protein function.* Crit Rev Biochem Mol Biol, 2007. **42**(1): p. 41-63.
44. Muniyappa, K., et al., *Mechanism of the concerted action of recA protein and helix-destabilizing proteins in homologous recombination.* Proc Natl Acad Sci U S A, 1984. **81**(9): p. 2757-61.
45. Moran, N.A. and P. Baumann, *Bacterial endosymbionts in animals.* Curr Opin Microbiol, 2000. **3**(3): p. 270-5.
46. Cox, M.M., *Motoring along with the bacterial RecA protein.* Nat Rev Mol Cell Biol, 2007. **8**(2): p. 127-38.

47. Takahashi, M., et al., *Calorimetric analysis of binding of two consecutive DNA strands to RecA protein illuminates mechanism for recognition of homology.* J Mol Biol, 2007. **365**(3): p. 603-11.
48. Rajan, R., J.W. Wisler, and C.E. Bell, *Probing the DNA sequence specificity of Escherichia coli RECA protein.* Nucleic Acids Res, 2006. **34**(8): p. 2463-71.
49. Mazin, A.V. and S.C. Kowalczykowski, *A novel property of the RecA nucleoprotein filament: activation of double- stranded DNA for strand exchange in trans.* Genes Dev, 1999. **13**(15): p. 2005-16.
50. Shibata, T., et al., *Homologous pairing in genetic recombination: formation of D loops by combined action of recA protein and a helix-destabilizing protein.* Proc Natl Acad Sci U S A, 1980. **77**(5): p. 2606-10.
51. Shibata, T., et al., *Homologous pairing in genetic recombination: complexes of recA protein and DNA.* Proc Natl Acad Sci U S A, 1979. **76**(10): p. 5100-4.
52. Shibata, T., et al., *Purified Escherichia coli recA protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments.* Proc Natl Acad Sci U S A, 1979. **76**(4): p. 1638-42.
53. Biet, E., J. Sun, and M. Dutreix, *Conserved sequence preference in DNA binding among recombination proteins: an effect of ssDNA secondary structure.* Nucleic Acids Res, 1999. **27**(2): p. 596-600.
54. Dutreix, M., *(GT)n repetitive tracts affect several stages of RecA-promoted recombination.* J Mol Biol, 1997. **273**(1): p. 105-13.
55. Tracy, R.B., J.K. Baumohl, and S.C. Kowalczykowski, *The preference for GT-rich DNA by the yeast Rad51 protein defines a set of universal pairing sequences.* Genes Dev, 1997. **11**(24): p. 3423-31.
56. Tracy, R.B. and S.C. Kowalczykowski, *In vitro selection of preferred DNA pairing sequences by the Escherichia coli RecA protein.* Genes Dev, 1996. **10**(15): p. 1890-903.
57. Yu, X., et al., *What is the structure of the RecA-DNA filament?* Curr Protein Pept Sci, 2004. **5**(2): p. 73-9.
58. van der Heijden, T., et al., *Real-time assembly and disassembly of human RAD51 filaments on individual DNA molecules.* Nucleic Acids Res, 2007. **35**(17): p. 5646-57.
59. Holthausen, J.T., C. Wyman, and R. Kanaar, *Regulation of DNA strand exchange in homologous recombination.* DNA Repair (Amst), 2010. **9**(12): p. 1264-72.
60. Galletto, R., et al., *Direct observation of individual RecA filaments assembling on single DNA molecules.* Nature, 2006. **443**(7113): p. 875-8.
61. Cox, M.M., *Motoring along with the bacterial RecA protein.* Nat Rev Mol Cell Biol, 2007. **8**(2): p. 127-138.
62. Umezu, K. and R.D. Kolodner, *Protein interactions in genetic recombination in Escherichia coli. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein.* J Biol Chem, 1994. **269**(47): p. 30005-13.
63. Gupta, R.C., et al., *Polarity of DNA strand exchange promoted by recombination proteins of the RecA family.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9843-8.
64. DasGupta, C. and C.M. Radding, *Polar branch migration promoted by recA protein: effect of mismatched base pairs.* Proc Natl Acad Sci U S A, 1982. **79**(3): p. 762-6.
65. Modesti, M., et al., *Fluorescent human RAD51 reveals multiple nucleation sites and filament segments tightly associated along a single DNA molecule.* Structure, 2007. **15**(5): p. 599-609.
66. Giese, K.C., C.B. Michalowski, and J.W. Little, *RecA-dependent cleavage of LexA dimers.* J Mol Biol, 2008. **377**(1): p. 148-61.
67. Little, J.W., *Autodigestion of lexA and phage lambda repressors.* Proc Natl Acad Sci U S A, 1984. **81**(5): p. 1375-9.
68. Little, J.W., *Mechanism of specific LexA cleavage: autodigestion and the role of RecA coprotease.* Biochimie, 1991. **73**(4): p. 411-21.

69. Lestini, R. and B. Michel, *UvrD controls the access of recombination proteins to blocked replication forks.* EMBO J, 2007. **26**(16): p. 3804-14.
70. Gorbalenya, A.E.a.K., E.V., *Helicases: amino acid sequence compaisons and structure-function relationships.* Current Opinion in Structural Biology, 1993: p. 3:419-429.
71. Matson, S.W., *Escherichia coli helicase II (urvD gene product) translocates unidirectionally in a 3' to 5' direction.* J Biol Chem, 1986. **261**(22): p. 10169-75.
72. Runyon, G.T., D.G. Bear, and T.M. Lohman, *Escherichia coli helicase II (UvrD) protein initiates DNA unwinding at nicks and blunt ends.* Proc Natl Acad Sci U S A, 1990. **87**(16): p. 6383-7.
73. Flores, M.J., V. Bidnenko, and B. Michel, *The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants.* EMBO Rep, 2004. **5**(10): p. 983-8.
74. Flores, M.J., N. Sanchez, and B. Michel, *A fork-clearing role for UvrD.* Mol Microbiol, 2005. **57**(6): p. 1664-75.
75. Sigal, N., et al., *A DNA-unwinding protein isolated from Escherichia coli: its interaction with DNA and with DNA polymerases.* Proc Natl Acad Sci U S A, 1972. **69**(12): p. 3537-41.
76. Raghunathan, S., et al., *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-Å resolution.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(13): p. 6652-6657.
77. Chrysogelos, S. and J. Griffith, *Escherichia coli single-strand binding protein organizes single-stranded DNA in nucleosome-like units.* Proc Natl Acad Sci U S A, 1982. **79**(19): p. 5803-7.
78. Raghunathan, S., et al., *Structure of the DNA binding domain of E. coli SSB bound to ssDNA.* Nat Struct Biol, 2000. **7**(8): p. 648-52.
79. Wyman, C., D. Ristic, and R. Kanaar, *Homologous recombination-mediated double-strand break repair.* DNA Repair (Amst), 2004. **3**(8-9): p. 827-33.
80. Lohman, T.M. and M.E. Ferrari, *Escherichia Coli Single-Stranded DNA-Binding Protein: Multiple DNA-Binding Modes and Cooperativities.* Annual Review of Biochemistry, 1994. **63**(1): p. 527-570.
81. Wigley, D.B., *Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB.* Nat Rev Microbiol, 2013. **11**(1): p. 9-13.
82. Sakai, A. and M.M. Cox, *RecFOR and RecOR as distinct RecA loading pathways.* J Biol Chem, 2009. **284**(5): p. 3264-72.
83. Sutherland, J.H., et al., *SOS induction by stabilized topoisomerase IA cleavage complex occurs via the RecBCD pathway.* J Bacteriol, 2008. **190**(9): p. 3399-403.
84. Wigley, D.B., *RecBCD: the supercar of DNA repair.* Cell, 2007. **131**(4): p. 651-3.
85. Singleton, M.R., et al., *Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks.* Nature, 2004. **432**(7014): p. 187-93.
86. Arnold, D.A. and S.C. Kowalczykowski, *Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme.* J Biol Chem, 2000. **275**(16): p. 12261-5.
87. Roman, L.J., A.K. Eggleston, and S.C. Kowalczykowski, *Processivity of the DNA helicase activity of Escherichia coli recBCD enzyme.* J Biol Chem, 1992. **267**(6): p. 4207-14.
88. Bianco, P.R. and S.C. Kowalczykowski, *The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5'-GCTGGTGG-3'.* Proc Natl Acad Sci U S A, 1997. **94**(13): p. 6706-11.
89. Dillingham, M.S., M. Spies, and S.C. Kowalczykowski, *RecBCD enzyme is a bipolar DNA helicase.* Nature, 2003. **423**(6942): p. 893-7.
90. Taylor, A.F. and G.R. Smith, *RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity.* Nature, 2003. **423**(6942): p. 889-93.
91. Handa, N., et al., *Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after chi recognition.* Mol Cell, 2005. **17**(5): p. 745-50.
92. Korangy, F. and D.A. Julin, *Efficiency of ATP hydrolysis and DNA unwinding by the RecBC enzyme from Escherichia coli.* Biochemistry, 1994. **33**(32): p. 9552-60.

93. Dixon, D.A. and S.C. Kowalczykowski, *The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme.* Cell, 1993. **73**(1): p. 87-96.
94. Lusetti, S.L. and M.M. Cox, *The bacterial RecA protein and the recombinational DNA repair of stalled replication forks.* Annu Rev Biochem, 2002. **71**: p. 71-100.
95. Sharples, G.J., S.M. Ingleston, and R.G. Lloyd, *Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA.* J Bacteriol, 1999. **181**(18): p. 5543-50.
96. Smith, G.R., *How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view.* Microbiol Mol Biol Rev, 2012. **76**(2): p. 217-28.
97. Spies, M., et al., *RecBCD enzyme switches lead motor subunits in response to chi recognition.* Cell, 2007. **131**(4): p. 694-705.
98. Joo, C., et al., *Real-time observation of RecA filament dynamics with single monomer resolution.* Cell, 2006. **126**(3): p. 515-27.
99. Spies, M. and S.C. Kowalczykowski, *The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins.* Mol Cell, 2006. **21**(4): p. 573-80.
100. Anderson, D.G., J.J. Churchill, and S.C. Kowalczykowski, *Chi-activated RecBCD enzyme possesses 5'-->3' nucleolytic activity, but RecBC enzyme does not: evidence suggesting that the alteration induced by Chi is not simply ejection of the RecD subunit.* Genes Cells, 1997. **2**(2): p. 117-28.
101. Kowalczykowski, S.C. and R.A. Krupp, *Effects of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA.* J Mol Biol, 1987. **193**(1): p. 97-113.
102. Cox, M.M., *Recombinational DNA repair of damaged replication forks in Escherichia coli: questions.* Annu Rev Genet, 2001. **35**: p. 53-82.
103. Kuzminov, A., *Recombinational repair of DNA damage in Escherichia coli and bacteriophage lambda.* Microbiol Mol Biol Rev, 1999. **63**(4): p. 751-813, table of contents.
104. Lloyd, R.G. and C. Buckman, *Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12.* J Bacteriol, 1985. **164**(2): p. 836-44.
105. Kushner, S.R., et al., *Genetic recombination in Escherichia coli: the role of exonuclease I.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 824-7.
106. Horii, Z. and A.J. Clark, *Genetic analysis of the recF pathway to genetic recombination in Escherichia coli K12: isolation and characterization of mutants.* J Mol Biol, 1973. **80**(2): p. 327-44.
107. Morimatsu, K. and S.C. Kowalczykowski, *RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair.* Mol Cell, 2003. **11**(5): p. 1337-47.
108. Madiraju, M.V. and A.J. Clark, *Evidence for ATP binding and double-stranded DNA binding by Escherichia coli RecF protein.* J Bacteriol, 1992. **174**(23): p. 7705-10.
109. Kolodner, R., R.A. Fishel, and M. Howard, *Genetic recombination of bacterial plasmid DNA: effect of RecF pathway mutations on plasmid recombination in Escherichia coli.* J Bacteriol, 1985. **163**(3): p. 1060-6.
110. Leiros, I., et al., *Crystal structure and DNA-binding analysis of RecO from Deinococcus radiodurans.* EMBO J, 2005. **24**(5): p. 906-18.
111. Makharashvili, N., et al., *A novel structure of DNA repair protein RecO from Deinococcus radiodurans.* Structure, 2004. **12**(10): p. 1881-9.
112. Kantake, N., et al., *Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination.* Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15327-32.

113. Koroleva, O., et al., *Structural conservation of RecF and Rad50: implications for DNA recognition and RecF function.* EMBO J, 2007. **26**(3): p. 867-77.
114. Mahdi, A.A. and R.G. Lloyd, *The recR locus of Escherichia coli K-12: molecular cloning, DNA sequencing and identification of the gene product.* Nucleic Acids Res, 1989. **17**(17): p. 6781-94.
115. Lee, B.I., et al., *Ring-shaped architecture of RecR: implications for its role in homologous recombinational DNA repair.* EMBO J, 2004. **23**(10): p. 2029-38.
116. Webb, B.L., M.M. Cox, and R.B. Inman, *An interaction between the Escherichia coli RecF and RecR proteins dependent on ATP and double-stranded DNA.* J Biol Chem, 1995. **270**(52): p. 31397-404.
117. Inoue, J., et al., *The process of displacing the single-stranded DNA-binding protein from single-stranded DNA by RecO and RecR proteins.* Nucleic Acids Res, 2008. **36**(1): p. 94-109.
118. Bork, J.M., M.M. Cox, and R.B. Inman, *The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA.* EMBO J, 2001. **20**(24): p. 7313-22.
119. Shan, Q., et al., *RecA protein filaments: end-dependent dissociation from ssDNA and stabilization by RecO and RecR proteins.* J Mol Biol, 1997. **265**(5): p. 519-40.
120. Umezu, K., N.W. Chi, and R.D. Kolodner, *Biochemical interaction of the Escherichia coli RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein.* Proc Natl Acad Sci U S A, 1993. **90**(9): p. 3875-9.
121. Lusetti, S.L., et al., *The RecF protein antagonizes RecX function via direct interaction.* Mol Cell, 2006. **21**(1): p. 41-50.
122. Mitchell, A.H. and S.C. West, *Role of RuvA in branch migration reactions catalyzed by the RuvA and RuvB proteins of Escherichia coli.* J Biol Chem, 1996. **271**(32): p. 19497-502.
123. Tsaneva, I.R., et al., *Purification and properties of the RuvA and RuvB proteins of Escherichia coli.* Mol Gen Genet, 1992. **235**(1): p. 1-10.
124. Parsons, C.A., et al., *Structure of a multisubunit complex that promotes DNA branch migration.* Nature, 1995. **374**(6520): p. 375-8.
125. Whitby, M.C., et al., *Interactions between RuvA and RuvC at Holliday junctions: inhibition of junction cleavage and formation of a RuvA-RuvC-DNA complex.* J Mol Biol, 1996. **264**(5): p. 878-90.
126. Stasiak, A., et al., *The Escherichia coli RuvB branch migration protein forms double hexameric rings around DNA.* Proc Natl Acad Sci U S A, 1994. **91**(16): p. 7618-22.
127. Ingleston, S.M., G.J. Sharples, and R.G. Lloyd, *The acidic pin of RuvA modulates Holliday junction binding and processing by the RuvABC resolvasome*. Vol. 19. 2000. 6266-6274.
128. Muller, B., I.R. Tsaneva, and S.C. West, *Branch migration of Holliday junctions promoted by the Escherichia coli RuvA and RuvB proteins. I. Comparison of RuvAB- and RuvB-mediated reactions.* J Biol Chem, 1993. **268**(23): p. 17179-84.
129. Marrione, P.E. and M.M. Cox, *RuvB protein-mediated ATP hydrolysis: functional asymmetry in the RuvB hexamer.* Biochemistry, 1995. **34**(30): p. 9809-18.
130. Yamada, K., et al., *Crystal structure of the RuvA-RuvB complex: a structural basis for the Holliday junction migrating motor machinery.* Mol Cell, 2002. **10**(3): p. 671-81.
131. Shah, R., R. Cosstick, and S.C. West, *The RuvC protein dimer resolves Holliday junctions by a dual incision mechanism that involves base-specific contacts.* EMBO J, 1997. **16**(6): p. 1464-72.
132. Bennett, R.J., H.J. Dunderdale, and S.C. West, *Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion.* Cell, 1993. **74**(6): p. 1021-31.
133. Dunderdale, H.J., et al., *Cloning, overexpression, purification, and characterization of the Escherichia coli RuvC Holliday junction resolvase.* J Biol Chem, 1994. **269**(7): p. 5187-94.
134. Lilley, D.M. and M.F. White, *The junction-resolving enzymes.* Nat Rev Mol Cell Biol, 2001. **2**(6): p. 433-43.

135. Saito, A., et al., *Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase.* Proc Natl Acad Sci U S A, 1995. **92**(16): p. 7470-4.
136. Shah, R., R. Cosstick, and S.C. West, *The RuvC protein dimer resolves Holliday junctions by a dual incision mechanism that involves base-specific contacts*. Vol. 16. 1997. 1464-1472.
137. Clark, A.J., *Recombination deficient mutants of E. coli and other bacteria.* Annu Rev Genet, 1973. **7**: p. 67-86.
138. Poteete, A.R., *What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function.* FEMS Microbiol Lett, 2001. **201**(1): p. 9-14.
139. Stahl, F.W., *Recombination in phage lambda: one geneticist's historical perspective.* Gene, 1998. **223**(1-2): p. 95-102.
140. Perkins, T.T., et al., *Sequence-dependent pausing of single lambda exonuclease molecules.* Science, 2003. **301**(5641): p. 1914-8.
141. Subramanian, K., et al., *The enzymatic basis of processivity in lambda exonuclease.* Nucleic Acids Res, 2003. **31**(6): p. 1585-96.
142. Matsuura, S., et al., *Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field.* Nucleic Acids Res, 2001. **29**(16): p. E79.
143. Mitsis, P.G. and J.G. Kwagh, *Characterization of the interaction of lambda exonuclease with the ends of DNA.* Nucleic Acids Res, 1999. **27**(15): p. 3057-63.
144. Cassuto, E. and C.M. Radding, *Mechanism for the action of lambda exonuclease in genetic recombination.* Nat New Biol, 1971. **229**(1): p. 13-6.
145. Little, J.W., *An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction.* J Biol Chem, 1967. **242**(4): p. 679-86.
146. Radding, C.M. and D.C. Shreffler, *Regulation of lambda exonuclease. II. Joint regulation of exonuclease and a new lambda antigen.* J Mol Biol, 1966. **18**(2): p. 251-61.
147. Kovall, R. and B.W. Matthews, *Toroidal structure of lambda-exonuclease.* Science, 1997. **277**(5333): p. 1824-7.
148. Kovall, R.A. and B.W. Matthews, *Structural, functional, and evolutionary relationships between lambda-exonuclease and the type II restriction endonucleases.* Proc Natl Acad Sci U S A, 1998. **95**(14): p. 7893-7.
149. Kmiec, E. and W.K. Holloman, *Beta protein of bacteriophage lambda promotes renaturation of DNA.* J Biol Chem, 1981. **256**(24): p. 12636-9.
150. Poteete, A.R., *Involvement of DNA replication in phage lambda Red-mediated homologous recombination.* Mol Microbiol, 2008. **68**(1): p. 66-74.
151. Muniyappa, K. and C.M. Radding, *The homologous recombination system of phage lambda. Pairing activities of beta protein.* J Biol Chem, 1986. **261**(16): p. 7472-8.
152. Passy, S.I., et al., *Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins.* Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4279-84.
153. Li, Z., et al., *The beta protein of phage lambda promotes strand exchange.* J Mol Biol, 1998. **276**(4): p. 733-44.
154. Court, D.L., A.B. Oppenheim, and S.L. Adhya, *A new look at bacteriophage lambda genetic networks.* J Bacteriol, 2007. **189**(2): p. 298-304.
155. Murphy, K.C., *Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme.* J Bacteriol, 1991. **173**(18): p. 5808-21.
156. Murphy, K.C., *The lambda Gam protein inhibits RecBCD binding to dsDNA ends.* J Mol Biol, 2007. **371**(1): p. 19-24.
157. Karu, A.E., et al., *The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of Escherichia coli.* J Biol Chem, 1975. **250**(18): p. 7377-87.
158. Sawitzke, J.A. and F.W. Stahl, *Phage lambda has an analog of Escherichia coli recO, recR and recF genes.* Genetics, 1992. **130**(1): p. 7-16.

159. Sawitzke, J.A. and F.W. Stahl, *The phage lambda orf gene encodes a trans-acting factor that suppresses Escherichia coli recO, recR, and recF mutations for recombination of lambda but not of E. coli.* J Bacteriol, 1994. **176**(21): p. 6730-7.
160. Maxwell, K.L., et al., *Functional similarities between phage lambda Orf and Escherichia coli RecFOR in initiation of genetic exchange.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11260-5.
161. Tarkowski, T.A., et al., *Gene products encoded in the ninR region of phage lambda participate in Red-mediated recombination.* Genes Cells, 2002. **7**(4): p. 351-63.
162. Sawitzke, J.A. and F.W. Stahl, *Roles for lambda Orf and Escherichia coli RecO, RecR and RecF in lambda recombination.* Genetics, 1997. **147**(2): p. 357-69.
163. Poteete, A.R., *Modulation of DNA repair and recombination by the bacteriophage lambda Orf function in Escherichia coli K-12.* J Bacteriol, 2004. **186**(9): p. 2699-707.
164. Reed, P., *Function of bacteriophage Orf recombinases in genetic exchange.* PhD thesis. University of Durham., 2006.
165. Poteete, A.R., A.C. Fenton, and H.R. Wang, *Recombination-promoting activity of the bacteriophage lambda Rap protein in Escherichia coli K-12.* J Bacteriol, 2002. **184**(16): p. 4626-9.
166. Sharples, G.J., et al., *Holliday junction binding and resolution by the Rap structure-specific endonuclease of phage lambda.* J Mol Biol, 2004. **340**(4): p. 739-51.
167. Sharples, G.J., L.M. Corbett, and I.R. Graham, *lambda Rap protein is a structure-specific endonuclease involved in phage recombination.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13507-12.
168. Sharples, G.J., L.M. Corbett, and P. McGlynn, *DNA structure specificity of Rap endonuclease.* Nucleic Acids Res, 1999. **27**(21): p. 4121-7.
169. Hollifield, W.C., Kaplan, E.N. and Huang, H.V., *Molecular Genetics and Genomics.* 1987: p. p. 248-255.
170. Shen, P. and H.V. Huang, *Homologous recombination in Escherichia coli: dependence on substrate length and homology.* Genetics, 1986. **112**(3): p. 441-57.
171. Mahdi, A.A., et al., *Holliday junction resolvases encoded by homologous rusA genes in Escherichia coli K-12 and phage 82.* J Mol Biol, 1996. **257**(3): p. 561-73.
172. Chan, S.N., et al., *Sequence specificity and biochemical characterization of the RusA Holliday junction resolvase of Escherichia coli.* J Biol Chem, 1997. **272**(23): p. 14873-82.
173. Giraud-Panis, M.J. and D.M. Lilley, *Structural recognition and distortion by the DNA junction-resolving enzyme RusA.* J Mol Biol, 1998. **278**(1): p. 117-33.
174. Stahl, F.W., Shurvington, C.E, Thamoason, L.C., Hill, S. and Stahl, M.M, *Genetics.* 1995: p. p. 1107-1121.
175. Hayashi, K., et al., *Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110.* Mol Syst Biol, 2006. **2**: p. 2006 0007.
176. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol Syst Biol, 2006. **2**: p. 2006 0008.
177. Curtis, F.A., et al., *The C-terminus of the phage lambda Orf recombinase is involved in DNA binding.* J Mol Recognit, 2011. **24**(2): p. 333-40.
178. Curtis, F.A., et al., *Phage ORF family recombinases: conservation of activities and involvement of the central channel in DNA binding.* PLoS One, 2014. **9**(8): p. e102454.
179. Balaji, S. and L. Aravind, *The RAGNYA fold: a novel fold with multiple topological variants found in functionally diverse nucleic acid, nucleotide and peptide-binding proteins.* Nucleic Acids Res, 2007. **35**(17): p. 5658-71.
180. Kroger, M. and G. Hobom, *A chain of interlinked genes in the ninR region of bacteriophage lambda.* Gene, 1982. **20**(1): p. 25-38.
181. Stella, S., D. Cascio, and R.C. Johnson, *The shape of the DNA minor groove directs binding by the DNA-bending protein Fis.* Genes Dev, 2010. **24**(8): p. 814-26.

182. Koch, C., et al., *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.* Nucleic Acids Res, 1991. **19**(21): p. 5915-22.
183. Feldman-Cohen, L.S., et al., *Common and variable contributions of Fis residues to high-affinity binding at different DNA sequences.* J Bacteriol, 2006. **188**(6): p. 2081-95.
184. Pan, C.Q., et al., *Structure of the Escherichia coli Fis-DNA complex probed by protein conjugated with 1,10-phenanthroline copper(I) complex.* Proc Natl Acad Sci U S A, 1994. **91**(5): p. 1721-5.
185. Pan, C.Q., et al., *Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts.* J Mol Biol, 1996. **264**(4): p. 675-95.
186. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Research, 2015. **43**(Web Server issue): p. W389-W394.
187. de Vries, R., *DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins.* Biochimie, 2010. **92**(12): p. 1715-21.
188. Browning, D.F., D.C. Grainger, and S.J. Busby, *Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression.* Curr Opin Microbiol, 2010. **13**(6): p. 773-80.
189. Postow, L., et al., *Topological domain structure of the Escherichia coli chromosome.* Genes Dev, 2004. **18**(14): p. 1766-79.
190. Dillon, S.C. and C.J. Dorman, *Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.* Nat Rev Microbiol, 2010. **8**(3): p. 185-95.
191. Dorman, C.J., *Nucleoid-associated proteins and bacterial physiology.* Adv Appl Microbiol, 2009. **67**: p. 47-64.
192. Hardy, C.D. and N.R. Cozzarelli, *A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure.* Mol Microbiol, 2005. **57**(6): p. 1636-52.
193. Luijsterburg, M.S., et al., *The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes.* Crit Rev Biochem Mol Biol, 2008. **43**(6): p. 393-418.
194. Ali Azam, T., et al., *Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid.* J Bacteriol, 1999. **181**(20): p. 6361-70.
195. Azam, T.A. and A. Ishihama, *Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity.* J Biol Chem, 1999. **274**(46): p. 33105-13.
196. Grainger, D.C. and S.J. Busby, *Global regulators of transcription in Escherichia coli: mechanisms of action and methods for study.* Adv Appl Microbiol, 2008. **65**: p. 93-113.
197. Dorman, C.J., *DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria.* Infect Immun, 1991. **59**(3): p. 745-9.
198. Grainger, D.C., et al., *Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome.* Nucleic Acids Res, 2006. **34**(16): p. 4642-52.
199. Miller, H.I., et al., *Site-specific recombination of bacteriophage lambda: the role of host gene products.* Cold Spring Harb Symp Quant Biol, 1979. **43 Pt 2**: p. 1121-6.
200. Miller, H.I. and D.I. Friedman, *An E. coli gene product required for lambda site-specific recombination.* Cell, 1980. **20**(3): p. 711-9.
201. Pettijohn, D.E., *The nucleoid. In Escherichia coli and Salmonella: Cellular and Molecular Biology.* 1996: p. 158-166.
202. Goosen, N. and P. van de Putte, *The regulation of transcription initiation by integration host factor.* Mol Microbiol, 1995. **16**(1): p. 1-7.
203. Crellin, P., S. Sewitz, and R. Chalmers, *DNA looping and catalysis; the IHF-folded arm of Tn10 promotes conformational changes and hairpin resolution.* Mol Cell, 2004. **13**(4): p. 537-47.

204. Esposito, D., J.S. Thrower, and J.J. Scocca, *Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation.* Nucleic Acids Res, 2001. **29**(19): p. 3955-64.
205. Corcoran, C.P. and C.J. Dorman, *DNA relaxation-dependent phase biasing of the fim genetic switch in Escherichia coli depends on the interplay of H-NS, IHF and LRP.* Mol Microbiol, 2009. **74**(5): p. 1071-82.
206. Mangan, M.W., et al., *The integration host factor (IHF) integrates stationary-phase and virulence gene expression in Salmonella enterica serovar Typhimurium.* Mol Microbiol, 2006. **59**(6): p. 1831-47.
207. Lammi, M., et al., *Proteins from the prokaryotic nucleoid: biochemical and 1H NMR studies on three bacterial histone-like proteins.* Adv Exp Med Biol, 1984. **179**: p. 467-77.
208. Rimsky, S., *Structure of the histone-like protein H-NS and its role in regulation and genome superstructure.* Curr Opin Microbiol, 2004. **7**(2): p. 109-14.
209. Hommais, F., et al., *Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS.* Mol Microbiol, 2001. **40**(1): p. 20-36.
210. Atlung, T. and F.G. Hansen, *Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in Escherichia coli.* J Bacteriol, 2002. **184**(7): p. 1843-50.
211. Rimsky, S., et al., *A molecular mechanism for the repression of transcription by the H-NS protein.* Mol Microbiol, 2001. **42**(5): p. 1311-23.
212. Lucht, J.M., et al., *Interactions of the nucleoid-associated DNA-binding protein H-NS with the regulatory region of the osmotically controlled proU operon of Escherichia coli.* J Biol Chem, 1994. **269**(9): p. 6578-8.
213. Badaut, C., et al., *The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA.* J Biol Chem, 2002. **277**(44): p. 41657-66.
214. Leonard, P.G., et al., *Investigation of the self-association and hetero-association interactions of H-NS and StpA from Enterobacteria.* Mol Microbiol, 2009. **73**(2): p. 165-79.
215. Uyar, E., et al., *Differential binding profiles of StpA in wild-type and h-ns mutant cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis.* J Bacteriol, 2009. **191**(7): p. 2388-91.
216. Ball, C.A., et al., *Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli.* J Bacteriol, 1992. **174**(24): p. 8043-56.
217. Dorman, C.J., *Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription.* Biochem Soc Trans, 2013. **41**(2): p. 542-7.
218. Cheng, Y.S., et al., *Structural analysis of the transcriptional activation region on Fis: crystal structures of six Fis mutants with different activation properties.* J Mol Biol, 2000. **302**(5): p. 1139-51.
219. Kostrewa, D., et al., *Three-dimensional structure of the E. coli DNA-binding protein FIS.* Nature, 1991. **349**(6305): p. 178-80.
220. Safo, M.K., et al., *The transactivation region of the fis protein that controls site-specific DNA inversion contains extended mobile beta-hairpin arms.* EMBO J, 1997. **16**(22): p. 6860-73.
221. Yuan, H.S., et al., *The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding.* Proc Natl Acad Sci U S A, 1991. **88**(21): p. 9558-62.
222. Skoko, D., et al., *Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis.* J Mol Biol, 2006. **364**(4): p. 777-98.
223. Keane, O.M. and C.J. Dorman, *The gyr genes of Salmonella enterica serovar Typhimurium are repressed by the factor for inversion stimulation, Fis.* Mol Genet Genomics, 2003. **270**(1): p. 56-65.
224. Schneider, R., et al., *A DNA architectural protein couples cellular physiology and DNA topology in Escherichia coli.* Mol Microbiol, 1999. **34**(5): p. 953-64.

225. Weinstein-Fischer, D. and S. Altuvia, *Differential regulation of Escherichia coli topoisomerase I by Fis.* Mol Microbiol, 2007. **63**(4): p. 1131-44.
226. Gohlke, C., et al., *Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence resonance energy transfer.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11660-4.
227. Lilley, D.M., *Kinking of DNA and RNA by base bulges.* Proc Natl Acad Sci U S A, 1995. **92**(16): p. 7140-2.
228. Cameron, A.D., D.M. Stoebel, and C.J. Dorman, *DNA supercoiling is differentially regulated by environmental factors and FIS in Escherichia coli and Salmonella enterica.* Mol Microbiol, 2011. **80**(1): p. 85-101.
229. Zablewska, B. and J. Kur, *Mutations in HU and IHF affect bacteriophage T4 growth: HimD subunits of IHF appear to function as homodimers.* Gene, 1995. **160**(1): p. 131-2.
230. Ninnemann, O., C. Koch, and R. Kahmann, *The E.coli fis promoter is subject to stringent control and autoregulation.* EMBO J, 1992. **11**(3): p. 1075-83.
231. Guhan, N. and K. Muniyappa, *Structural and functional characteristics of homing endonucleases.* Crit Rev Biochem Mol Biol, 2003. **38**(3): p. 199-248.
232. Raaijmakers, H., et al., *X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture.* The EMBO Journal, 1999. **18**(6): p. 1447-1458.
233. Xu, S.-y., et al., *Structure Determination and Biochemical Characterization of a Putative HNH Endonuclease from <italic>Geobacter metallireducens</italic> GS-15.* PLoS ONE, 2013. **8**(9): p. e72114.
234. Giraud-Panis, M.J., D.R. Duckett, and D.M. Lilley, *The modular character of a DNA junction-resolving enzyme: a zinc-binding motif in bacteriophage T4 endonuclease VII.* J Mol Biol, 1995. **252**(5): p. 596-610.
235. Giraud-Panis, M.J. and D.M. Lilley, *T4 endonuclease VII. Importance of a histidine-aspartate cluster within the zinc-binding domain.* J Biol Chem, 1996. **271**(51): p. 33148-55.
236. Aravind, L., K.S. Makarova, and E.V. Koonin, *SURVEY AND SUMMARY: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories.* Nucleic Acids Res, 2000. **28**(18): p. 3417-32.
237. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic Acids Res, 2014. **42**(Database issue): p. D222-30.
238. Biertumpfel, C., W. Yang, and D. Suck, *Crystal structure of T4 endonuclease VII resolving a Holliday junction.* Nature, 2007. **449**(7162): p. 616-620.
239. Esposito, D. and G.F. Gerard, *The Escherichia coli Fis protein stimulates bacteriophage lambda integrative recombination in vitro.* J Bacteriol, 2003. **185**(10): p. 3076-80.
240. Ball, C.A. and R.C. Johnson, *Multiple effects of Fis on integration and the control of lysogeny in phage lambda.* J Bacteriol, 1991. **173**(13): p. 4032-8.