Engineering Site-Specific Recombinases for use in Synthetic Biology

Hayley Louise Macfarlane

B.Sc.

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

University of Glasgow

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

May 2017

Summary

This project examined whether it was possible to create functional hybrid serine integrases – proteins responsible for recombining DNA in a site-specific manner. Creating hybrid recognition sites, specifically engineered to be recognised by the new integrases, was examined concurrently. Ultimately, new serine integrases and recognition sites were created with the intention of increasing the repertoire of serine integrases available for use as independently functioning modules in synthetic biology assemblies. Experiments were carried out primarily on two groups of hybrid integrases – BxbI integrase and ΦC31 integrase, and the smaller recombinase Tn3 resolvase and ΦC31 integrase.

It was determined that either the BxbI integrase/ΦC31 integrase hybrids were not active on hybrid or parental recognition sites, or that the proteins themselves were not expressed at a high enough level to exhibit any activity. However, one ΦC31 integrase/BxbI integrase hybrid did exhibit activity on ΦC31 integrase recognition sites *in vivo*, though not on hybrid sites.

However, Tn3 resolvase/ΦC31 integrase hybrid proteins proved far more promising. The two hybrids exhibited recombination on sites created for them, whilst exhibiting no activity on any parental recognition sites. When both Tn3 resolvase and either hybrid integrase were present *in vitro*, recombination on combination substrate plasmids containing one copy of the Tn3 resolvase recognition site *res* site I and one copy of a hybrid recognition site was much higher than for either hybrid against hybrid sites on its own.

Additionally, throughout this investigation, it was discovered that ΦC31 integrase cleaved and recombined several sites very dissimilar to its natural *attP* and *attB* sites.

“Everything’s going so well!”

Harold Zidler; Moulin Rouge!

Acknowledgements

I originally just wanted to fill my acknowledgements section with as many stupid quotes as I could, but in the end I do have many people to thank, so the quotes will unfortunately have to be used another time.

First and foremost I would like to thank Marshall for being a wonderful supervisor and for being very enthusiastic about hybrid serine integrases even when nothing was working. To everyone else in the Hooker lab, particularly Femi, Steph and Arlene - I could not have written my thesis without all of the help and guidance you have given me over the past few years, so thank you.

Outside of the lab, I don't think I would have gotten through probably four of the most challenging years of my life without Jake, our bunny rabbits, my family and my very supportive and outrageous friends. Thank you for always being understanding if I had to cancel plans in favour of a weekend jaunt to the lab to transform some *E.coli* or for simply being there when I needed to vent about science, either with ice cream or alcohol (and sometimes both) in tow.

Thank you to the Cecilian Society for essentially being my mad Glaswegian theatre family, who provided my PhD with the musical soundtracks that it very much needed to get me through it all. I guess that's why the only quote that's made it into my thesis is ultimately one from a musical, albeit also my favourite film. Thank you for providing me with a much-needed creative outlet when lab life was getting a bit too much.

I'd also like to thank myself for finishing this without dying, so well done me.

Finally, thank you to the EPSRC for funding my research and for ultimately providing me with the opportunity to undertake this project in the first place, for which I am truly grateful.

The research reported in this thesis is my own work unless it is otherwise stated and it has not been submitted for any other degree.

Hayley Louise Macfarlane

Abbreviations

Chemicals and reagents

AcOH	Acetic acid
APS	Ammonium persulphate
ATP	Adenosine triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
EtBr	Ethidium bromide
KOAc	Potassium acetate
SDS	Sodium dodecyl sulphate
TAE	Tris acetate EDTA (electrophoresis buffer)
TEMED	N-N-N'-N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminimethane
PMSF	Phenylmethanesulfonylfluoride

Commonly used abbreviations

DNA	Deoxyribonucleic acid
HR	Homologous recombination
HDR	Homology directed repair
NHEJ	Non-homologous end joining
PAGE	Polyacrylamide gel electrophoresis
DBD	DNA binding domain
RD	Recombinase domain
TALE	Transcription activator-like effector
TALEN	TALE nuclease
TALER	TALE recombinase
UV	Ultraviolet
ZFR	Zinc finger recombinase

Units

A	Amperes	**mol**	Molar
V	Volts	**bp**	Basepairs
W	Watts	**rpm**	Revolutions per minute
L	Litres	**OD**	Optical density
m	Metres	**k**	Kilo (10^{3})
g	Grams	**c**	Centi (10^{-2})
h	Hours	**m**	Milli (10^{-3})
min	Minutes	**µ**	Micro (10^{-6})
°C	Degrees Celcius	**n**	Nano (10^{-9})
M	Moles	**p**	Pico (10^{-12})

Table of Contents

Chapter One: Introduction

1.1 DNA recombination

DNA recombination occurs as a result of molecules of DNA being broken and re-ligated together, resulting in new, recombinant DNA molecules. In the case of homologous recombination (HR), both molecules of DNA are similar in sequence to each other. However, it is also possible to recombine two dissimilar molecules of DNA via non-homologous end joining (NHEJ; Ariyoshi et al, 2000). Recombination is used as a tool in nature primarily for DNA repair, but also has an essential role in increasing genetic diversity in eukaryotic systems; recombination occurs during chromosomal crossover in meiosis.

Homologous recombination involves the formation of Holliday junction intermediates. The process requires a complex protein machinery to migrate the branches of these junctions to complete resolution of the DNA to form a recombinant molecule (Holliday, 1964; Ariyoshi et al, 2000). Life as we know it would not be possible without HR – it is critical in preserving genomic integrity and rescuing stalled replication forks which direct gene expression, amongst other essential maintenance roles (Camerini-Otero and Hsieh, 1995; San Filippo et al, 2008).

1.1.i Site-specific recombination

Recombination in a site-specific manner can be undertaken in nature via proteins known as site-specific recombinases. These proteins recognise and recombine short sections of DNA (normally between 30 and 200 bp in length), using either a serine or a tyrosine nucleophile to attack the DNA and covalently attach to it during strand exchange (Smith, 2015). Site-specific recombinases are found in prokaryotes and carry out a range of functions, including but not limited to the integration and excision of DNA, the regulation of gene expression via the controlled inversion of regulatory sequences, and catalysing the transposition of antibiotic resistance genes and gene cassettes (Smith, 2015).

The creation of a covalent DNA/protein intermediate during strand cleavage stores energy, thus absolving the reaction from requiring an external, high energy co-factor e.g. ATP (Grindley, 1994; Grindley et al, 2006).

Site-specific recombinases can catalyse the recombination of DNA in three different manners – inversion, excision and integration (Figure 1.1). However, whether inversion, excision or integration is carried out is very tightly regulated through both recombinase recognition site orientation and topological selectivity (Grindley et al, 2006). Inversion and excision can only occur between recognition sites found within the same molecule of DNA i.e. intramolecular recombination. Inversion occurs when these sites are present in indirect repeat to each other – head to head – whereas excision occurs when these sites are in direct repeat – head to tail (Figure 1.1). Conversely, integration occurs via intermolecular recombination, so long as one or both of the DNA molecules containing recombinase recognition sites is a plasmid i.e. circular in nature (Stark et al, 1992).

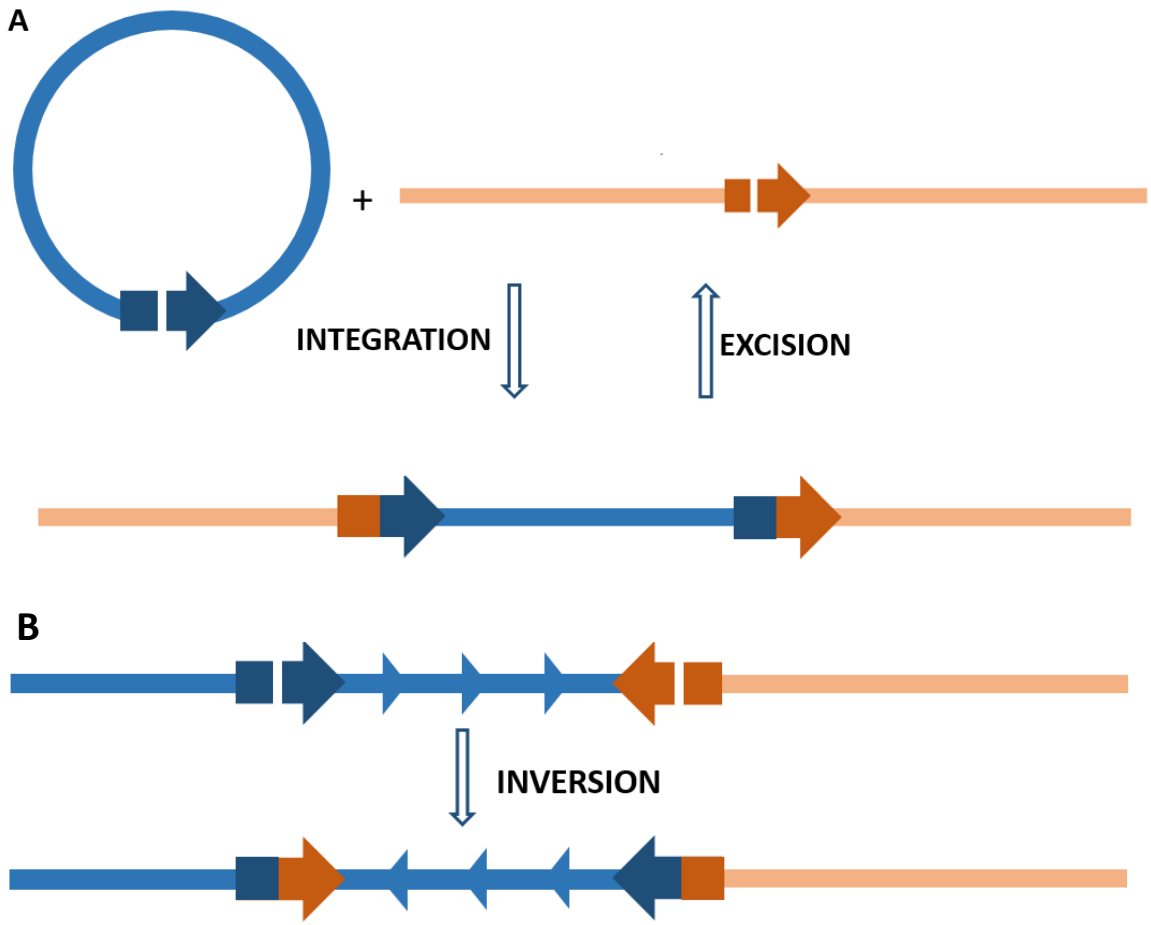

Figure 1.1: Possible results of site-specific recombination.

(A) Integration results from intermolecular recombination i.e. between recognition sites located on separate molecules of DNA, whilst excision occurs in an intramolecular fashion, between sites in a head-to-tail orientation. (B) Inversion is also intramolecular, though recognition sites must be aligned in a head-to-head fashion.

1.2 Tyrosine recombinases

This group of site-specific recombinases are defined by the presence of a tyrosine residue within their catalytic domains which is subsequently used as the nucleophile in site-specific recombination (Van Duyne, 2015).

Perhaps the most well-studied members of the tyrosine recombinase family are Flp (from *Saccharomyces cerevisiae*) and Cre (originally from the *Escherichia coli* bacteriophage P1). These recombinases, along with all other tyrosine recombinases, exchange DNA two strands at a time, thus making a Holliday junction intermediate (demonstrated in Figure 1.2). Cre has become one of the most popular site specific recombinases to use in research for genetic engineering, particularly in mammalian systems (Van Duyne, 2015).

1.3 Serine recombinases

Serine recombinases are so-called due to the residue used as the nucleophile in recombination being a serine residue as opposed to tyrosine. Unlike tyrosine recombinases, serine recombinases make a double strand break at the crossover site in each original DNA/protein duplex. The generalised process by which serine recombinases are believed to function can be seen in Figure 1.3.

For the purposes of this introduction, small serine recombinases and large serine recombinases (also called integrases) are considered separately, as their method of action is slightly different, as detailed below.

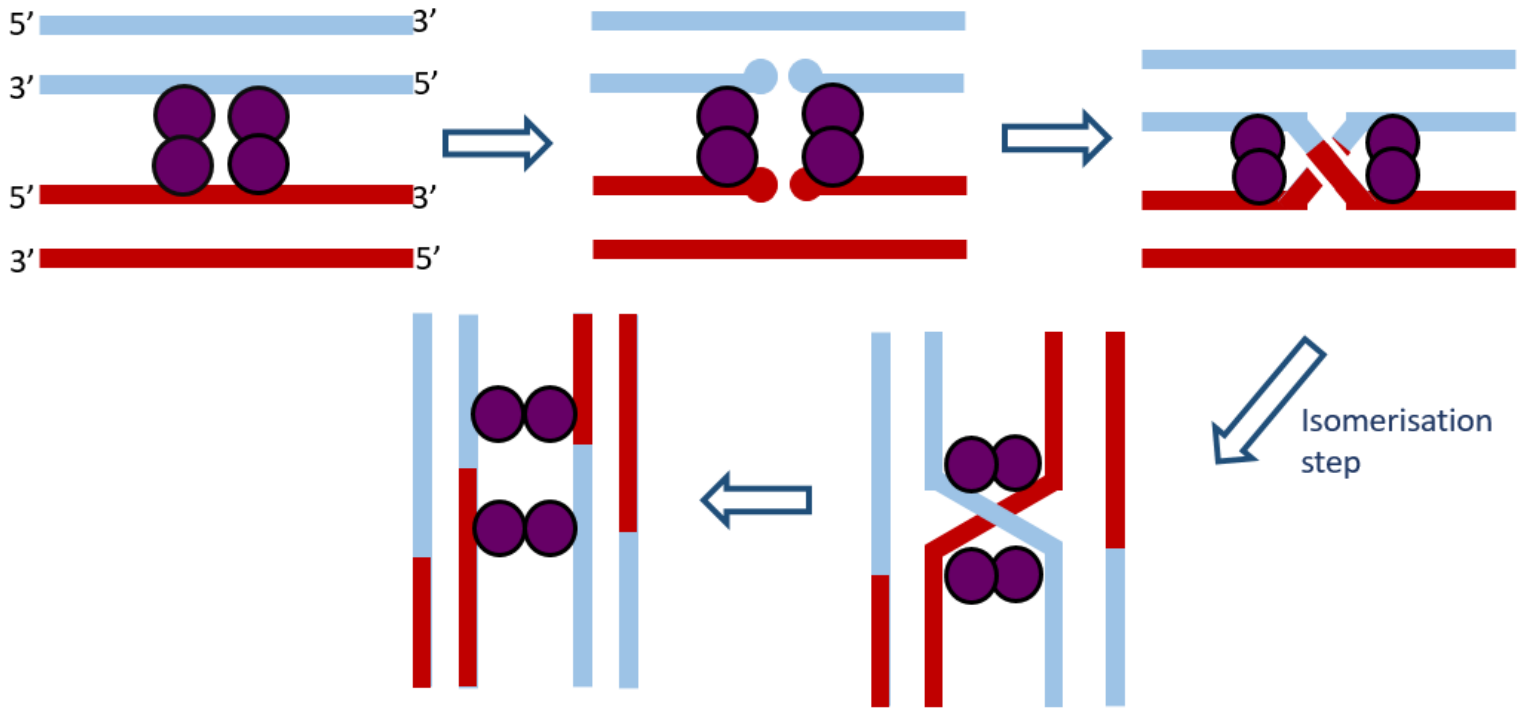

Figure 1.2: Tyrosine recombinase mechanism of action.

A tyrosine nucleophile catalyses the cleavage of one strand of DNA per duplex, which is exchanged and religated to the other to form a Holliday junction intermediate. After isomerisation of the Holliday junction, the second DNA strand from each duplex is cleaved, exchanged and subsequently religated to form two recombinant DNA complexes. Purple circles represent tyrosine recombinase monomers, whilst the DNA is represented by blue, red, or blue and red strands.

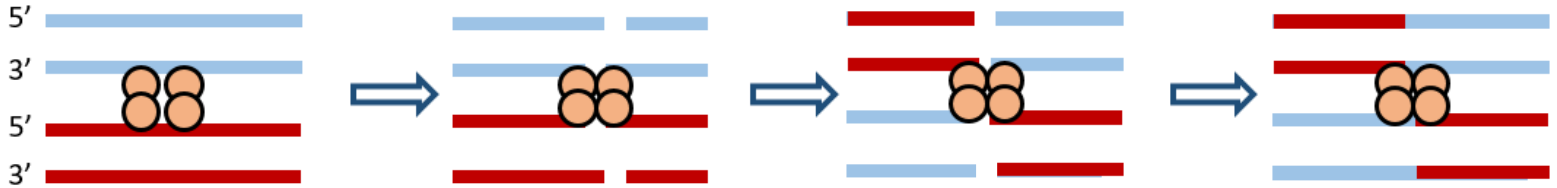

Figure 1.3: Serine recombinase mechanism of action.

Double-stranded cleavage of DNA is catalysed by a serine nucleophile, forming transient phosphoserine bonds at the 5’ ends of the DNA strands. The DNA is religated together following strand exchange, producing recombinant DNA. Peach circles represent serine recombinase monomers, whilst the DNA is represented by blue, red, or blue and red strands.

1.3.i Small serine recombinases

Tn3 resolvase is a well characterised small serine recombinase, whose function in nature is to resolve the co-integrate products formed during the replication of the Tn3 transposon during transposition (Grindley et al, 2002). Each subunit of Tn3 resolvase is made up of a catalytic N-terminal domain and a C-terminal DNA binding domain, connected via an E-helix – discussed in further detail in section 1.5 (Yang and Steitz, 1995).

Tn3 resolvase recognises and recombines a 114 bp region of DNA known as *res* site I via the process shown in Figure 1.5. This full *res* site is made up of a 28 bp length of DNA known as *res* site I, as well as the accessory sites *res* site II and *res* site III (34 bp and 25 bp long, respectively), which mediate recombination of the catalytic *res* site I (Rice, 2015). It is believed that four Tn3 resolvase catalytic subunits form a synaptic complex with two *res* site I's, whilst eight further subunits of Tn3 resolvase bind to two copies of *res* sites II and III. This results in a final synaptic complex of two Tn3 *res* sites in their entirety bound by twelve subunits of Tn3 resolvase (Rice, 2015). Figure 1.5 demonstrates resolution of DNA by Tn3 resolvase within a plasmid with two *res* sites in direct repeat. Resolution in such a fashion results in the creation of a 2-node catenane product. The dimer-dimer interactions noted to occur at *res* sites II and III promote recombination at *res* site I; they also trap three negative supercoils which consequently provide the energy required during recombination, thus allowing the reaction to occur without the need for energy-rich external co-factors such as ATP (Stark and Boocock, 1995; Burke et al 2004; Grindley et al, 2006).

Whilst some mutations within *res* site I do not result in any adverse effect on recombination, others strongly inhibit the reaction for reasons that are not yet clear (Burke et al, 2004). Conversely, some mutations within Tn3 resolvase itself allow the protein to recombine *res* site I without the accessory *res* sites II and III (Akopian et al, 2003). NM resolvase is one such hyperactive mutant which is used extensively in this project (see section 2.7 for more information).

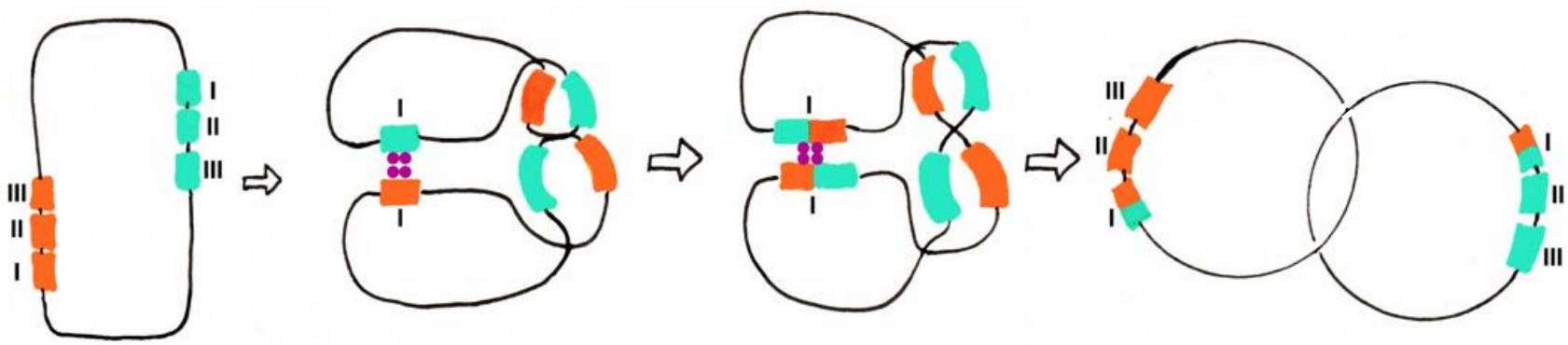

Figure 1.5: Tn3 resolvase mechanism of action.

Resolvase is illustrated as a purple circle. Only resolvase subunits operating on *res* site I are shown. Resolvase binds to each *res* sub-site (I, II and III) as a dimer. The formation of the catalytic tetramer at *res* site I is facilitated by the regulatory complexes formed at *res* sites II and III. The resultant, final product of Tn3 resolvase-mediated recombination is a 2-noded catenane.

1.3.ii Large serine integrases

Large serine recombinases – also called serine integrases - contain a catalytic domain comprising about 150 residues and a large carboxyl-terminal region (C-terminal domain or CTD) of between 300 to 550 residues (Van Duyne and Rutherford, 2013). Serine integrases act by binding as a dimer to specific sequences of DNA in the bacteriophage and in the host bacteria. These sites are known as *attP* and *attB* respectively and are regarded as being comprised of two "half sites" with a crossover dinucleotide in the middle of each full site - see Figure 1.6 (Smith et al, 2010). These sequences are short (typically around 40 bp long) and are recognised specifically by their respective integrase, which recombines these sites to form *attL* and *attR* sites (Smith et al, 2010). Unlike most site-specific recombination reactions, integrase-catalysed *attP x attB* recombination is irreversible, creating recombinant sites *attL* and *attR*. *attL x attR* recombination requires integrase plus an additional phage-encoded protein – the Recombination Directionality Factor (RDF) The generally accepted method by which serine integrases catalyse DNA integration can be seen in Figure 1.7.

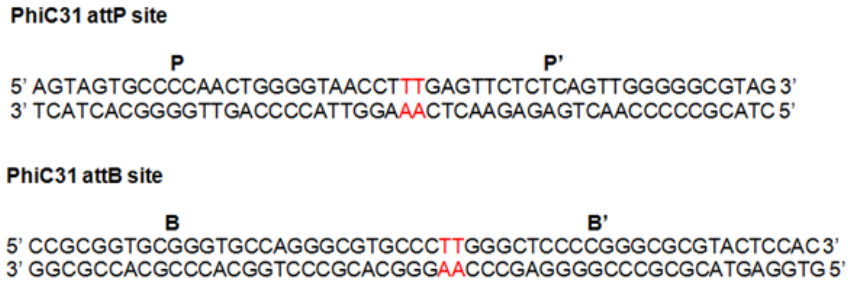

Figure 1.6: Structure of serine integrase *att* sites exemplified using ΦC31 integrase *att* sites. Both *attP* and *attB* are comprised of two half sites – P and P' and B and B' respectively. When ΦC31 integrase catalyses the integration of phage DNA into host, bacterial DNA the sites become recombined, resulting in an *attL* site (P with B') and an *attR* site (P' with B). Red nucleotides represent the crossover dinucleotide in each site (based on sequences from Olorunniji et al, 2012).

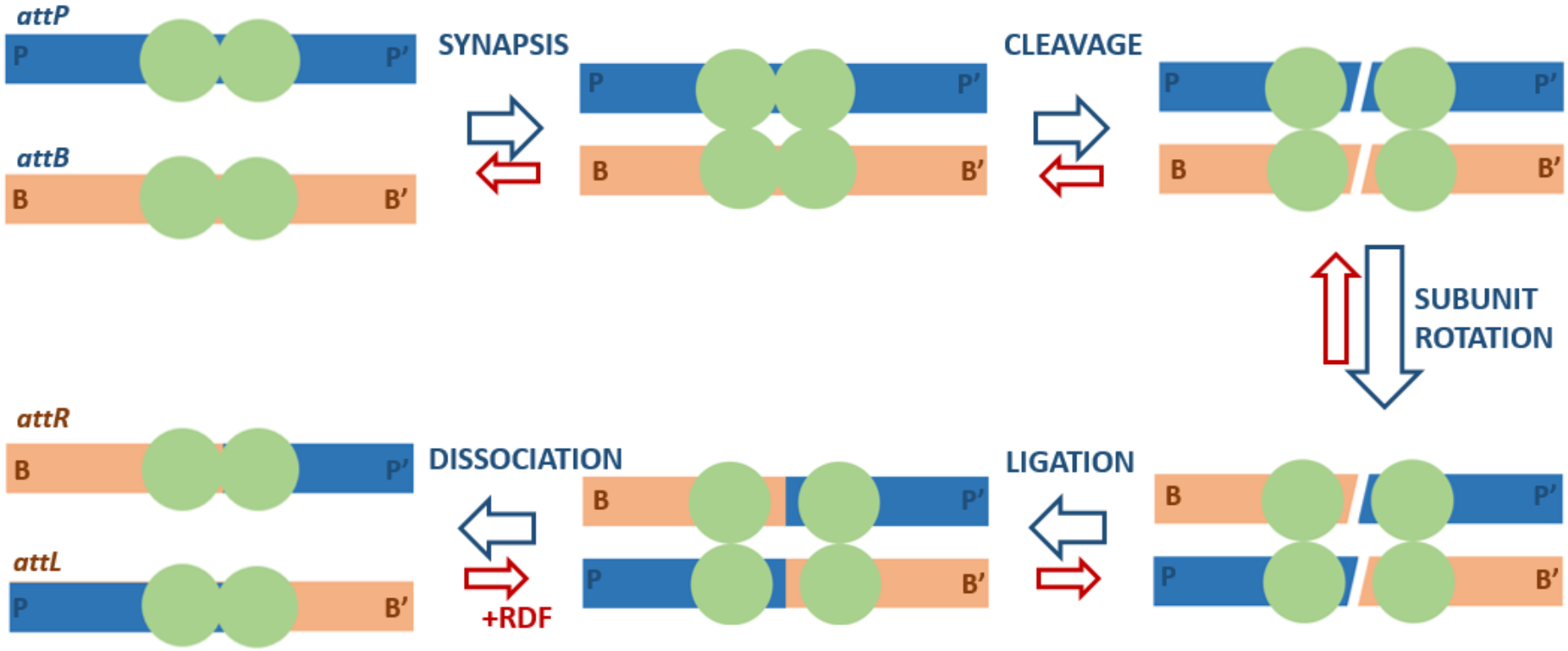

Figure 1.7: Large serine recombinase mechanism of action.

Serine integrase units are illustrated as green circles. DNA *attP* and *attB* sites are blue or orange strips respectively. Dark blue arrows denote the forward reaction between *attP* and *attB*; the red arrows denote the reverse reaction which is only possible in the presence of the recombination directionality factor (RDF). Serine integrase binds as a dimer to either *attP* or *attB,* coming together to form a synaptic tetramer. Once cleavage and strand exchange occurs, the DNA is religated together and the synaptic complex disassociates, leaving recombinant *attL* and *attR* sites bound by integrase dimers.

Serine integrases typically only recognise their own *att* sites. However, on occasion they can recombine at "pseudo" *att* sites, wherein sequences with only a few nucleotide differences from the natural *attP* or *attB* site do not prevent their respective integrase from recognising and recombining the site. Furthermore, some serine integrases can also recognise the *att* sites of other integrases; this is known as integrase cross-talk (Singh et al, 2014).

Such a case was identified in 2014 (Singh et al), whereby two integrases with 59% sequence similarity, Bxz2 and Peaches, engaged in non-reciprocal cross-talk. The two integrases had different established *attP* sites which shared less than 50% similarity but, surprisingly, identical *attB* sites (Singh et al, 2014). Upon testing their recombination efficiencies the team discovered that Bxz2 integrase recognised and bound to the Peaches *attP* site despite a lack of sequence similarity. It also recognised a hybrid *attP* site consisting of a Peaches *P* and Bxz2 *P'*, but did not recognise the reverse hybrid *attP* site (Bxz2 *P* and Peaches *P'*) (Singh et al, 2014). Conversely, Peaches did not recognise or bind the Bxz2 *attP* site at all but recognised and bound to both hybrid sites (Singh et al, 2014). The authors noted that this non-reciprocal cross-talk suggests that both integrases have varying conformational flexibility for enabling/restricting protein configurations required to bind to different *att* sites and that critical bases for *att* site recognition are perhaps positioned differently in the *att* sites looked at, and that this is likely applicable to most serine integrases (Singh et al, 2014).

Perhaps the most well-characterised, popular serine integrases for use in research are ΦC31 integrase and BxbI integrase. The *att* recognition sites for both integrases are known (Figure 1.8), as are their RDFs. The full DNA sequence and resultant amino acid sequence of both integrases is also known (seen aligned against other, major serine integrases in Figure 1.9). ΦC31 integrase and BxbI integrase are known to be orthogonal i.e. no cross-talk occurs between the two. Both integrases have been used in various cell systems, including *E.coli,* yeast, *Drosophila melanogaster*, mouse and human cells. In particular, BxbI integrase was established as the best of fifteen

candidates as the integrase of choice to integrate DNA into the human genome, with ΦC31 integrase being second (Xu et al, 2013).

Figure 1.8: ΦC31 integrase and BxbI integrase *att* sites.

Crossover dinucleotide in each site is in bold, black typeface.

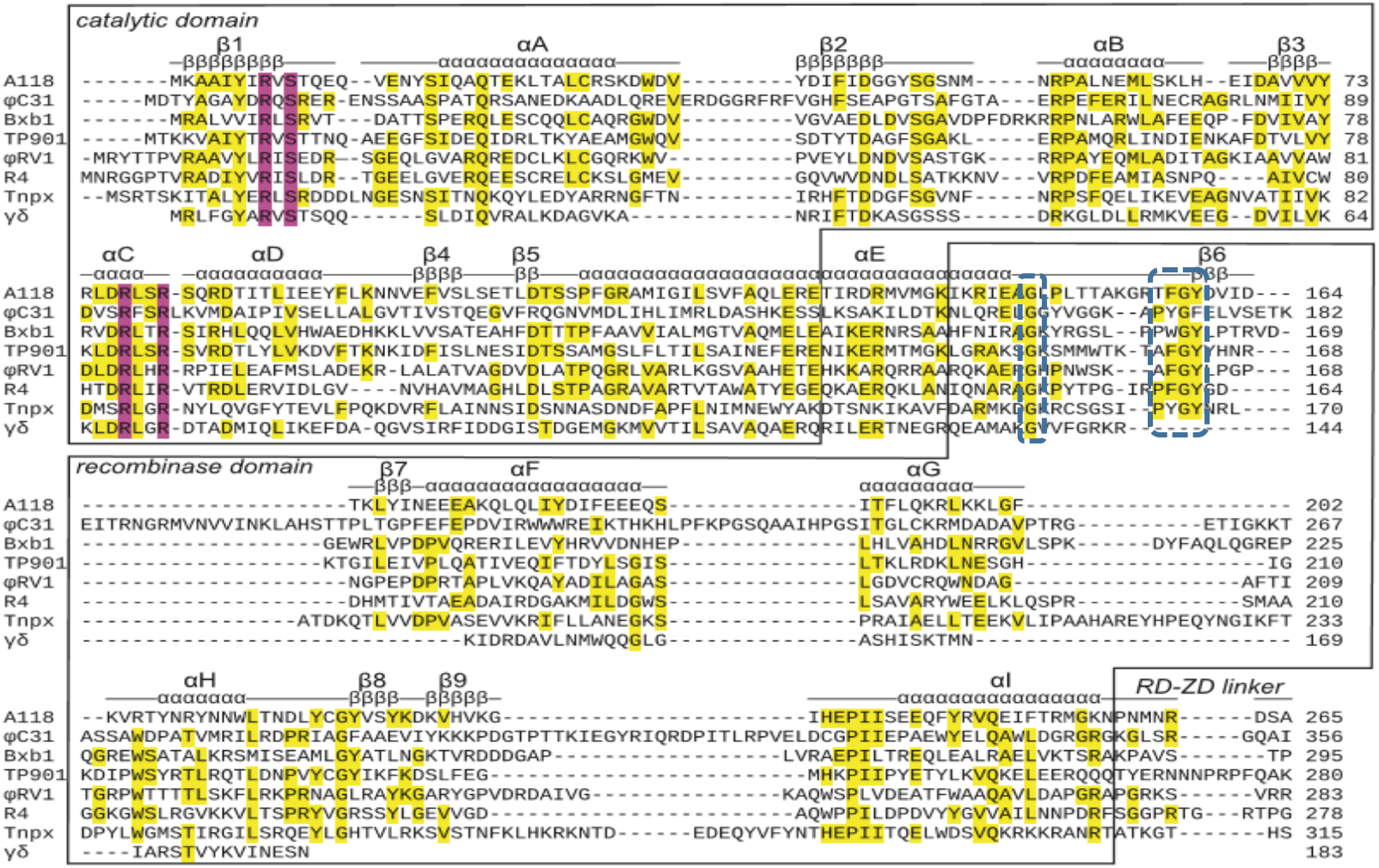

Figure 1.9: Alignment of ΦC31 integrase and Bxb1 integrase against other large serine integrases.

Yellow boxes indicate residues conserved between at least some serine integrase species. Purple boxes indicate catalytic residues and conserved between all species shown. Dashed blue boxes highlight a conserved glycine residue and a semi-conserved "piggy domain" – both found in the recombinase domain (modified from Van Duyne and Rutherford, 2013).

1.4 Crystal structure of a small serine recombinase

In 1995, Yang and Steitz successfully solved a crystal structure of a γδ resolvase dimer bound to DNA, a small serine recombinase very similar in sequence and nature to Tn3 resolvase, highlighting how the protein interacts with its cognate recognition site (Figure 1.10). It revealed that, of the 183 residues in γδ resolvase, residues 1-100 made up the N-terminal catalytic domain of the protein. Residues 101-137 formed an extended αE helix and αE helix arm and a short linker made up of residues 138-148 connected this N-terminal region of γδ resolvase to its C-terminal DNA binding domain, which consisted of residues 149-183 (Yang and Steitz, 1995).

As illustrated in Figure 1.10, the catalytic domain of γδ resolvase is constructed of four alpha helices (named αA, αB, αC and αD). A β-sheet separates αD from the other three alpha helices.

Hydrophobic residues found in the αE helix are known to form interactions between each subunit of γδ resolvase; these both stabilise the dimer structure constructed around its cognate DNA as well as position the catalytic domain of resolvase relative to the DNA. Whilst the αE helix of γδ resolvase remains structured even in the absence of DNA, this is not the case of the 'arm' region of the αE helix. This remains disordered in the absence of DNA – it is this region of γδ resolvase that is responsible for positioning of the C-terminal domain of the protein and the formation of interactions with the minor groove of its cognate DNA (Rice and Steitz, 1994).

The residues 138-148, containing the N-terminal to C-terminal linker, form interactions within the minor groove of the γδ resolvase DNA recognition site, at a conserved AT base pair which corresponds to a glycine-arginine motif (residues 141 and 142 in γδ resolvase).

Two of the three alpha helices found in the DNA binding domain – αG and αH - are responsible for forming a helix-turn-helix motif that then confers DNA

recognition to the motifs flanking the ends of each *res* site I and the regulatory sites II and III in the resolvase-DNA synaptic complex (Yang and Steitz, 1995).

Crystal structures for γδ resolvase suggest that a model known as "DNA out" fits what happens in nature most accurately with regards to the synaptic interactions of small serine recombinases and DNA at crossover sites. This model purports that resolvase subunits make up the inside of the protein-DNA complex, with DNA making up the outside (Burke et al, 2004; Li et al, 2005; Rowland et al, 2009).

Crystal structure data (alongside pre-existing biochemical and topological data) have also helped to elucidate the mechanism by which serine resolvase recombination takes place. Subunit rotation – by which the cognate DNA is cleaved by resolvase during recombination before one half of the entire DNA/protein complex is rotated 180° relative to the other half – is deemed to be the most likely mechanism. This subunit rotation occurs at the αE helix flat interface of hydrophobic residues in γδ resolvase (Grindley et al, 2006).

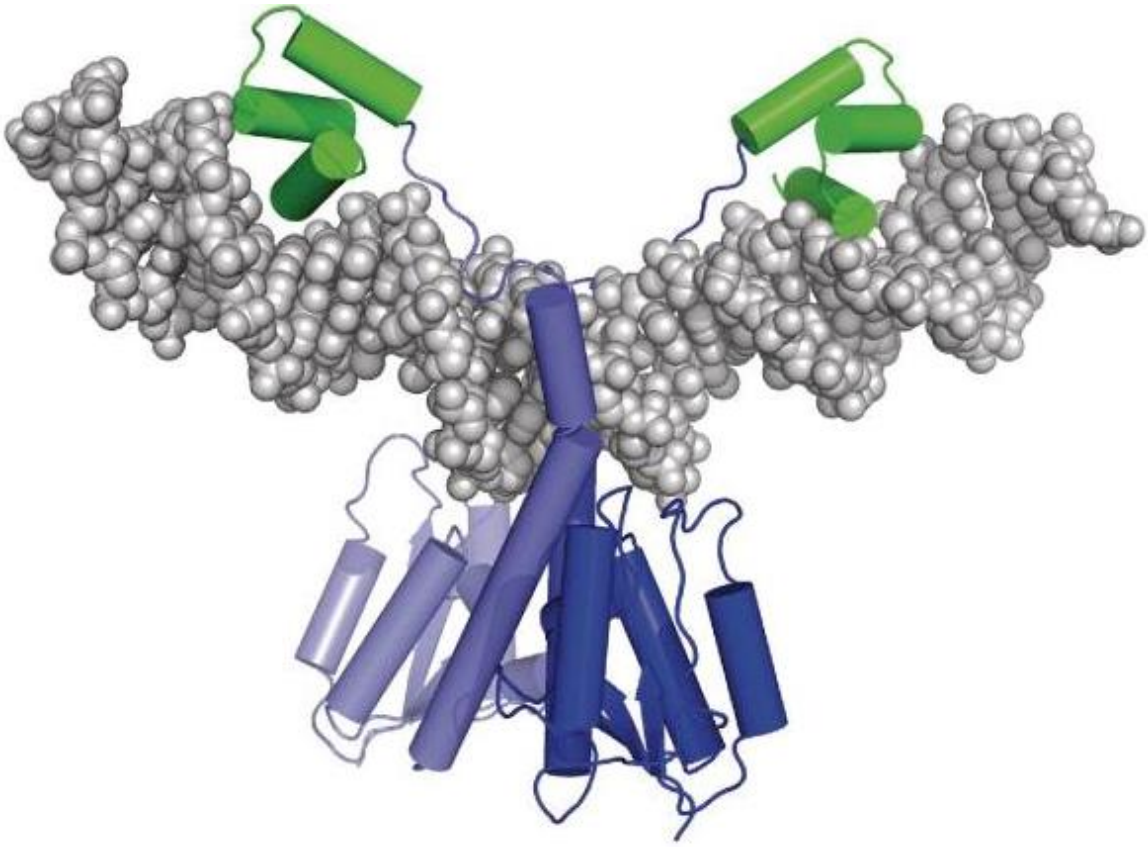

Figure 1.10: Crystal structure of γδ resolvase (1GDT) bound as a dimer to DNA.

Resolvase is represented in ribbon diagrams, with the N terminal catalytic domain and αE helix shown in dark blue whilst the C terminal DNA binding domain (DBD) is in green. The αE helix is kinked into the minor groove of the cognate DNA - represented as a space-filling model (Yang and Steitz, 1995).

1.5 Crystal structure of a large serine integrase

Until recently, any understanding of how serine integrases functioned and interacted with their recognition sites at a structural level was speculative. Whilst some serine integrases have been studied in detail via biochemical probing and quantitative analysis as well as *in vitro* investigation, it was not until 2008 that a crystal structure for a serine integrase became available. Yuan et al (2008) managed to successfully crystallise the structure of the catalytic domain for TP901-1 integrase to a resolution of 2.1Å (Figure 1.11).

The crystal structure reveals that the TP901-1 integrase catalytic domain is constructed of five α-helices, named αA-αE respectively, organised into the spatial orientation shown in Figure 1.12. By comparing their TP901-1 integrase crystal structure to that of the γδ resolvase crystal structure solved by Yang and Steitz in 1995, several things may be observed. First and foremost, superimposition of a TP901-1 integrase catalytic domain tetramer and the full γδ resolvase tetramer indicate that parts of both structures line up well against each other as shown in Figure 1.12. Conversely, other areas do not superimpose well; catalytic subdomains in the TP901-1 integrase structure which precede the αE helix differ from similar regions in the γδ resolvase structure. Based on these catalytic subdomain differences and the discovery that the catalytic serine residues within the structure are too far apart for the structure to be part of a synaptic complex, Yuan et al (2008) came to the conclusion that the catalytic domain that they had managed to crystallise is likely a unique intermediate found at another part of the recombination pathway.

However, despite what was elucidated and hypothesised from this incomplete crystal structure, a crystal structure of a full serine integrase bound to one of its *att* sites would be required to truly aid in the understanding of how serine integrases function, and to answer other questions, such as how serine integrases recognise their *att* sites (Yuan et al, 2008).

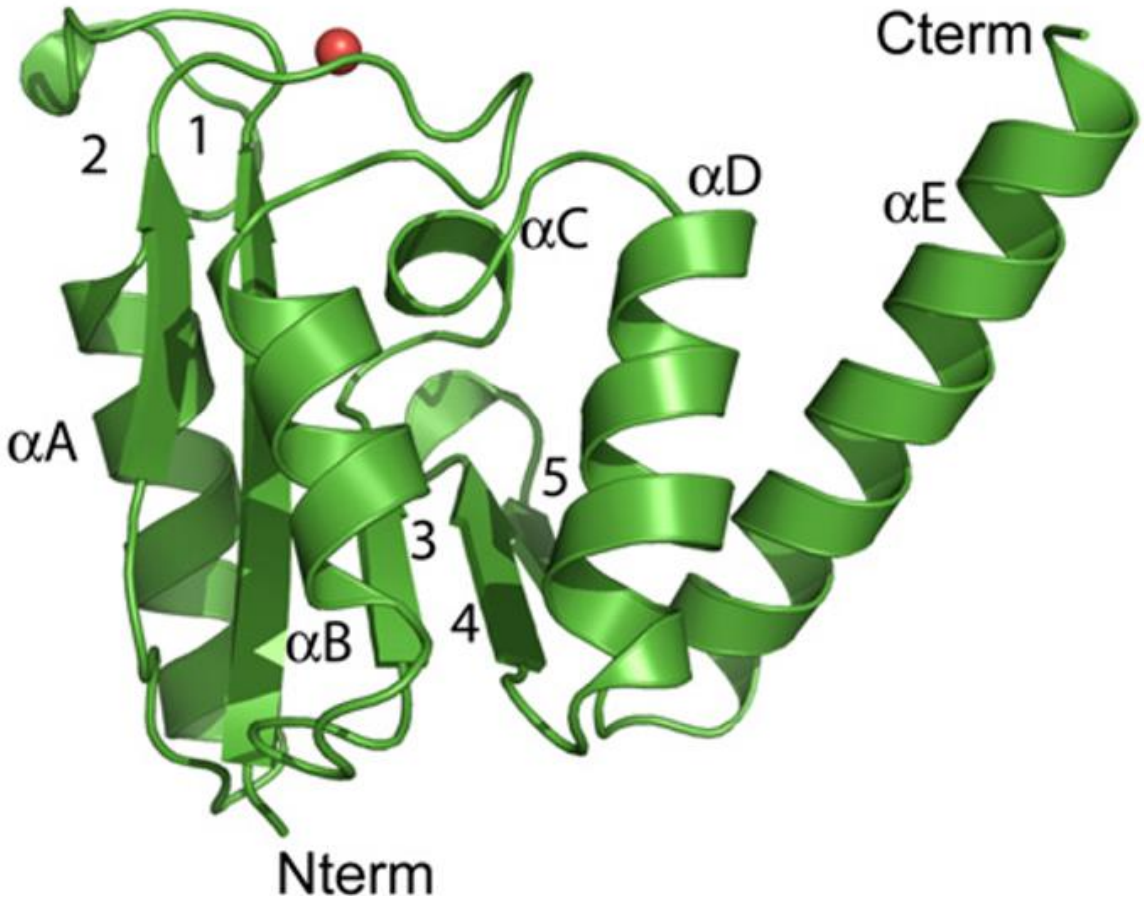

Figure 1.11: Crystal structure of TP901-1 integrase catalytic domain.

The ß-sheets are numbered 1-5, whilst the α-helices are named αA-E respectively. Cterm: C terminus; Nterm: N terminus. The catalytic serine residue is represented by a red sphere, which also highlights the location of the TP901-1 integrase active site (Yuan et al, 2008).

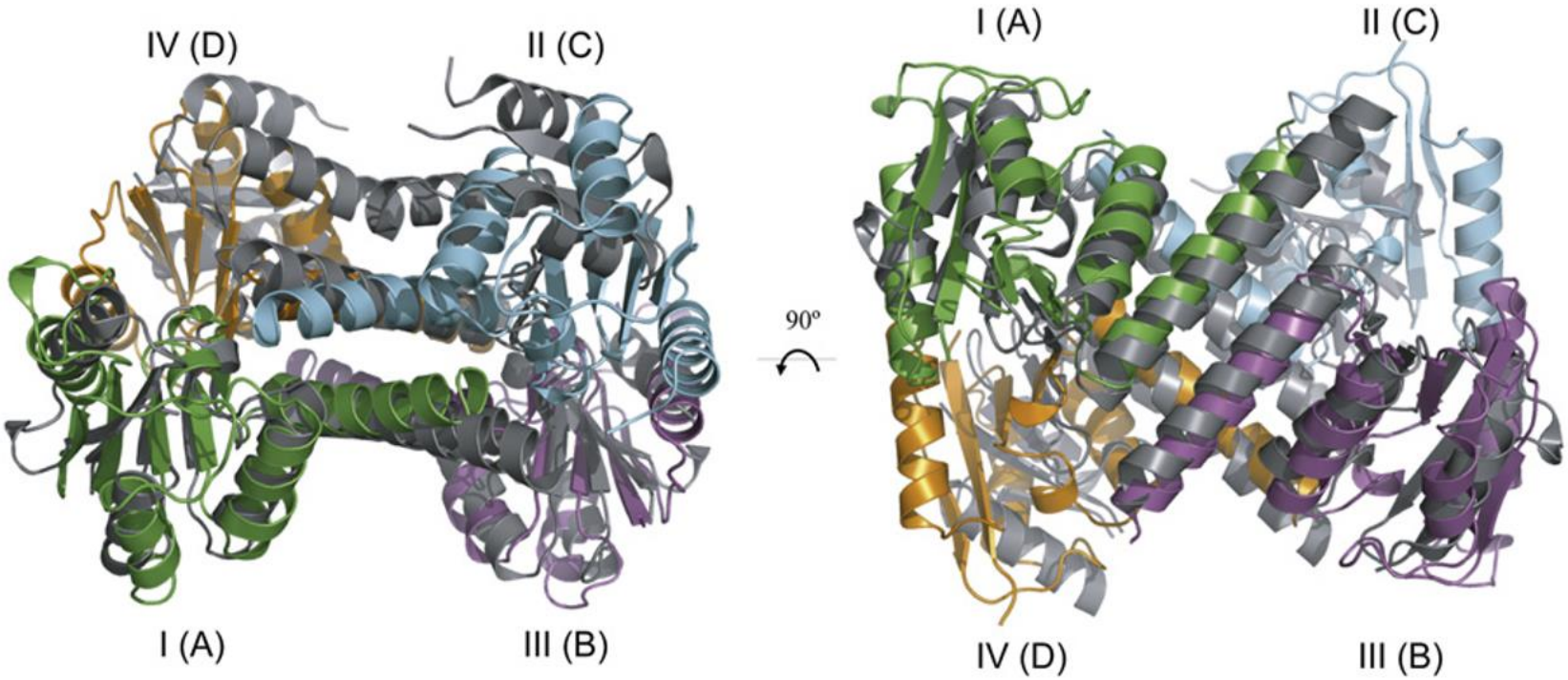

Figure 1.12: Superimposition of the crystal structures of TP901-1 integrase catalytic domain tetramer and activated γδ resolvase apo-tetramer.

Each TP901-1 integrase catalytic subunit is coloured green, light blue, purple or gold and labelled as I-IV, respectively. Resolvase is coloured grey and each subunit is labelled A-D in parentheses beside each comparable serine integrase subunit. Subunits A and B of resolvase superimpose well with the catalytic subunits I and III of TP901-1 integrase. On the other hand, subunits C and D do not superimpose well with TP901-1 integrase subunits II and IV. Two orthogonal views of the same superimposition have been included (Yuan et al, 2008).

In 2013, Rutherford et al published a crystal structure of the integrase LI, an "A118-like" serine integrase with a 2.7Å resolution (Figure 1.13). This crystal structure deals with the LI C-terminal domain bound to its *attP* half-site (Van Duyne and Rutherford, 2013).

The crystal structure reveals that the recombinase domain (RD) of LI integrase is constructed of a mixture of α-helices and ß-sheets and is connected to a zinc-nucleated domain (ZD) via an 8-residue-long linker. A coiled-coil (CC) motif, part of the ZD, stretches away from both the RD and the cognate DNA - in this instance, an *attP* half-site. The structure also contains part of the αE helix – responsible for connected the RD to the N-terminal domain (containing the catalytic domain) – which was also visualised in the Yuan et al TP901-1 crystal structure in full (2008). Rutherford et al (2013) establish that, together with what they managed to visualise of the αE helix, residues within the CTD form direct interactions along the full stretch of the *attP* half-site in both major and minor DNA grooves.

Visualisation of the four independent complexes crystallised reveals that, whilst most domains remain in similar spatial positions, the position of the CC motif is altered between the four (Figure 1.14). This has led to a hypothesis on its potential involvement in *attP* versus *attB* recognition and the prevention of the reverse reaction of *attL*/*R* back to *attB*/*P* (section 1.5.i; Rutherford et al, 2013).

Whilst Rutherford et al (2013) acknowledge that there is limited sequence similarity between the decidedly larger RD in serine integrases compared to the CTD of resolvase – made up of a much smaller helix-turn-helix (HTH) domain – they nevertheless identify a motif that appears to be structurally analogous to the resolvase HTH domain found in the very centre of the LI integrase RD. Another structure was also found to be similar to a comparable region in resolvase, this time the linker connecting the LI integrase αE helix to its ß6 sheet in the RD. This bears compounded significance when taking into account

that this region in both resolvase and LI integrase is shown to be involved in forming interactions with its cognate DNA (Yang and Steitz, 1995). This region has also been identified as an area that is likely to play an important, functional role in serine integrase-mediated recombination as mutations in this region are known to have a severe effect on the activity of integrase (Ghosh et al, 2005; Gupta et al, 2007).

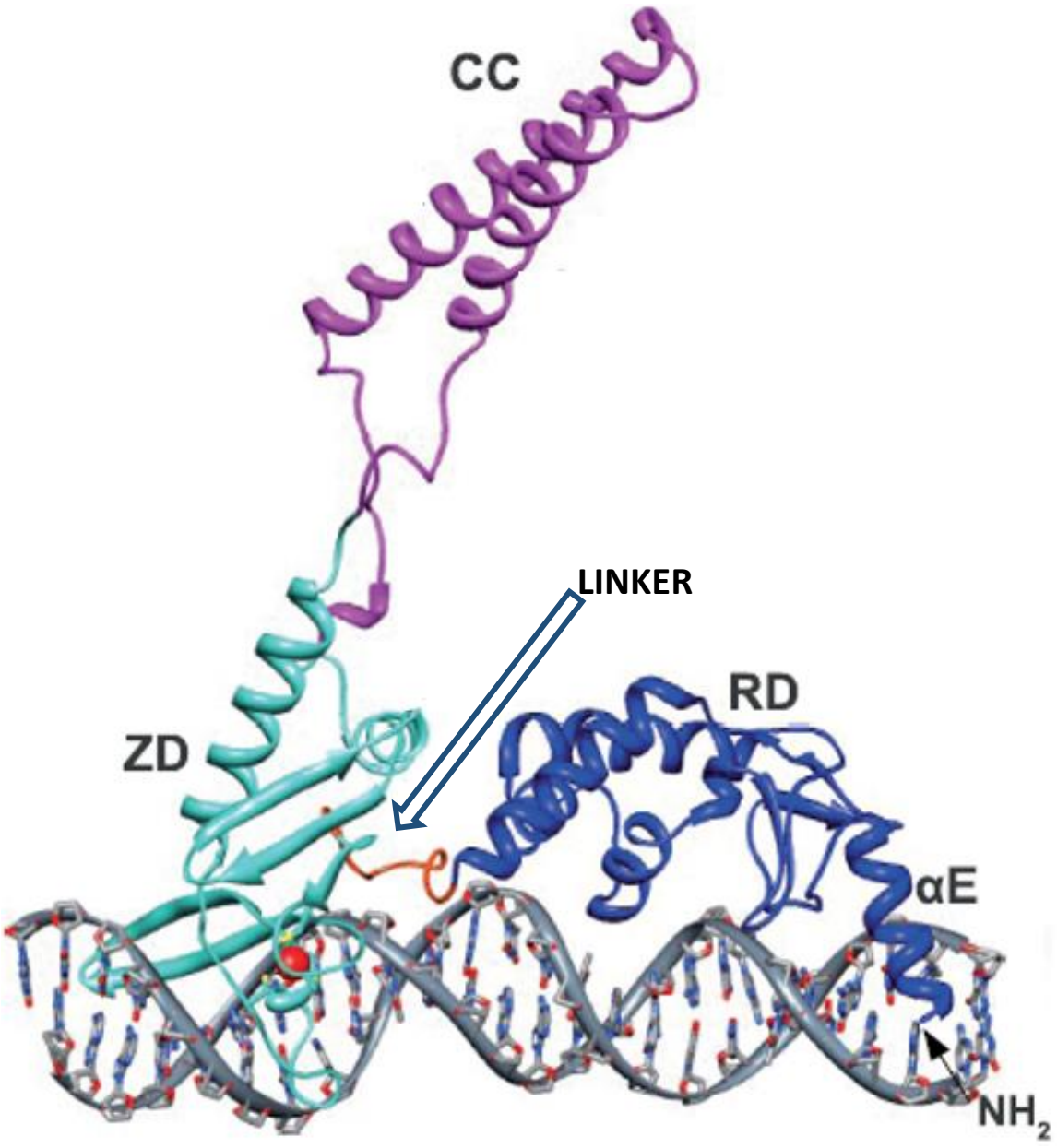

Figure 1.13: Crystal structure of LI integrase bound to its *attP* half-site. RD: recombinase domain; ZD: zinc-nucleated domain; CC: coiled-coil motif; red sphere: coordinated zinc ion. The αE helix seems to interact directly with the base sequence of the *attP* half site close to the crossover dinucleotide of what would be the full *attP* site, similarly to small serine recombinases e.g. γδ resolvase (modified from Rutherford et al, 2013).

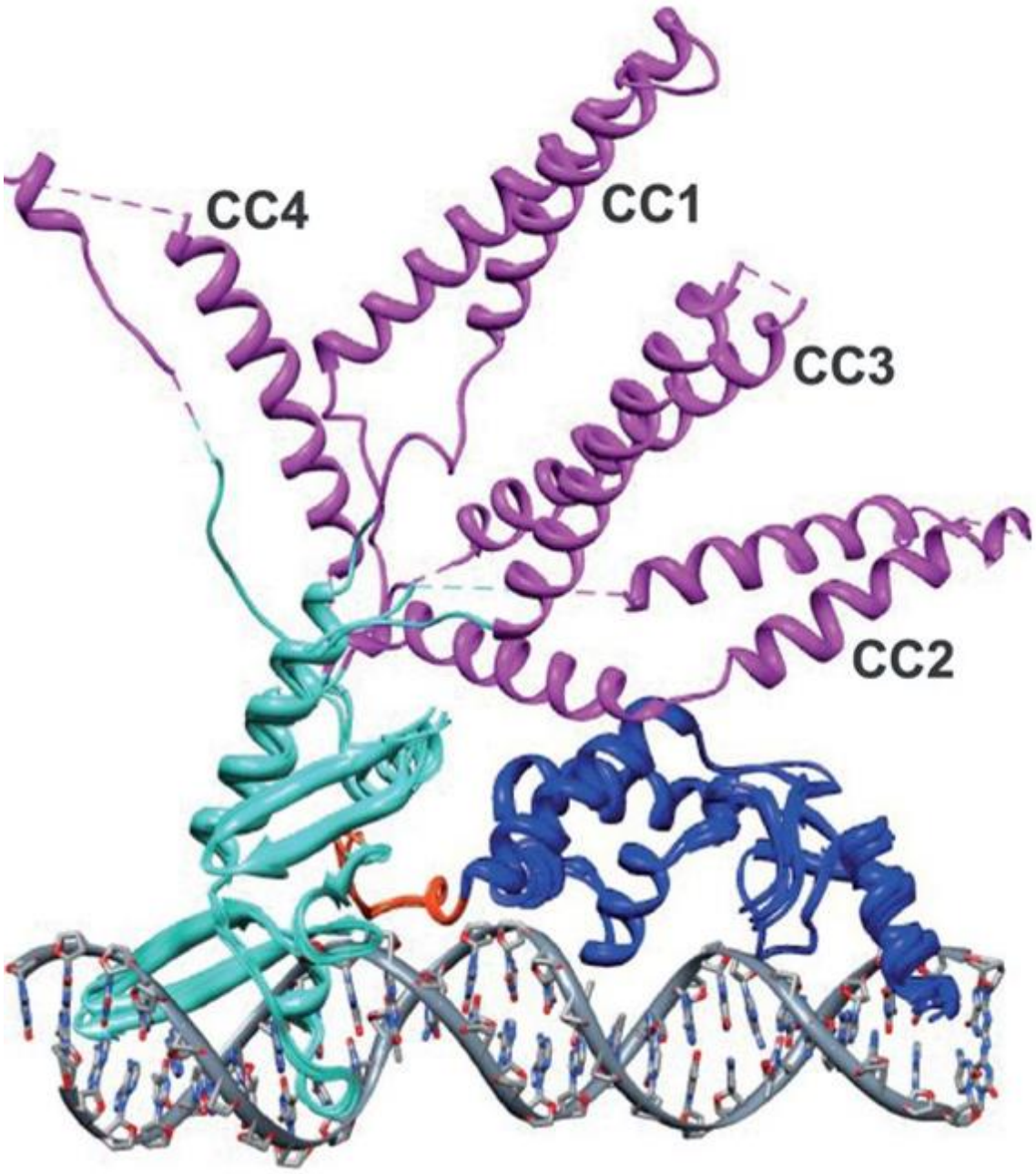

Figure 1.14: Superimposition of conformational changes of the coiled-coil domains in LI integrase.

CC1-4 are the four different conformational configurations adopted by LI integrase in each of the four crystal complexes obtained (modified from Rutherford et al, 2013).

1.5.i Hypothesised mechanism for large serine integrases

Rutherford et al (2013) hypothesise (based on the structural data obtained) that the quaternary arrangement of zinc-nucleated domains (ZD), and therefore the conformational change of the coiled-coil motifs observed in their crystal structures, may be what differentiates *attP* from *attB* recognition within the serine integrases. This would aid in the understanding of how recombination directionality is regulated at the structural level within the complex. Taking this hypothesis into account, a more comprehensive idea of how site-specific recombination occurs via large serine integrases can be seen in Figure 1.15.

However, this hypothesis is based on incomplete crystal structures. As noted by the authors, the next obvious step in the visualisation of integrase-*att* site crystal structures would be to crystallise LI integrase bound to an *attB* half-site to "complete the picture" and to help confirm their coiled-coil hypothesis (Rutherford et al, 2013). Also critical would be to successfully crystallise a complete integrase dimer bound to a full *att* site – that is, a crystal structure that includes the full catalytic domain of the protein as well as an *attP* or *attB* site (preferably both) in its entirety. A complete crystal structure would lend further credit to Van Duyne and Rutherford's hypothesis or allow one to refute it; either way, it would certainly aid in the understanding of what is truly happening during serine integrase-mediated site-specific recombination.

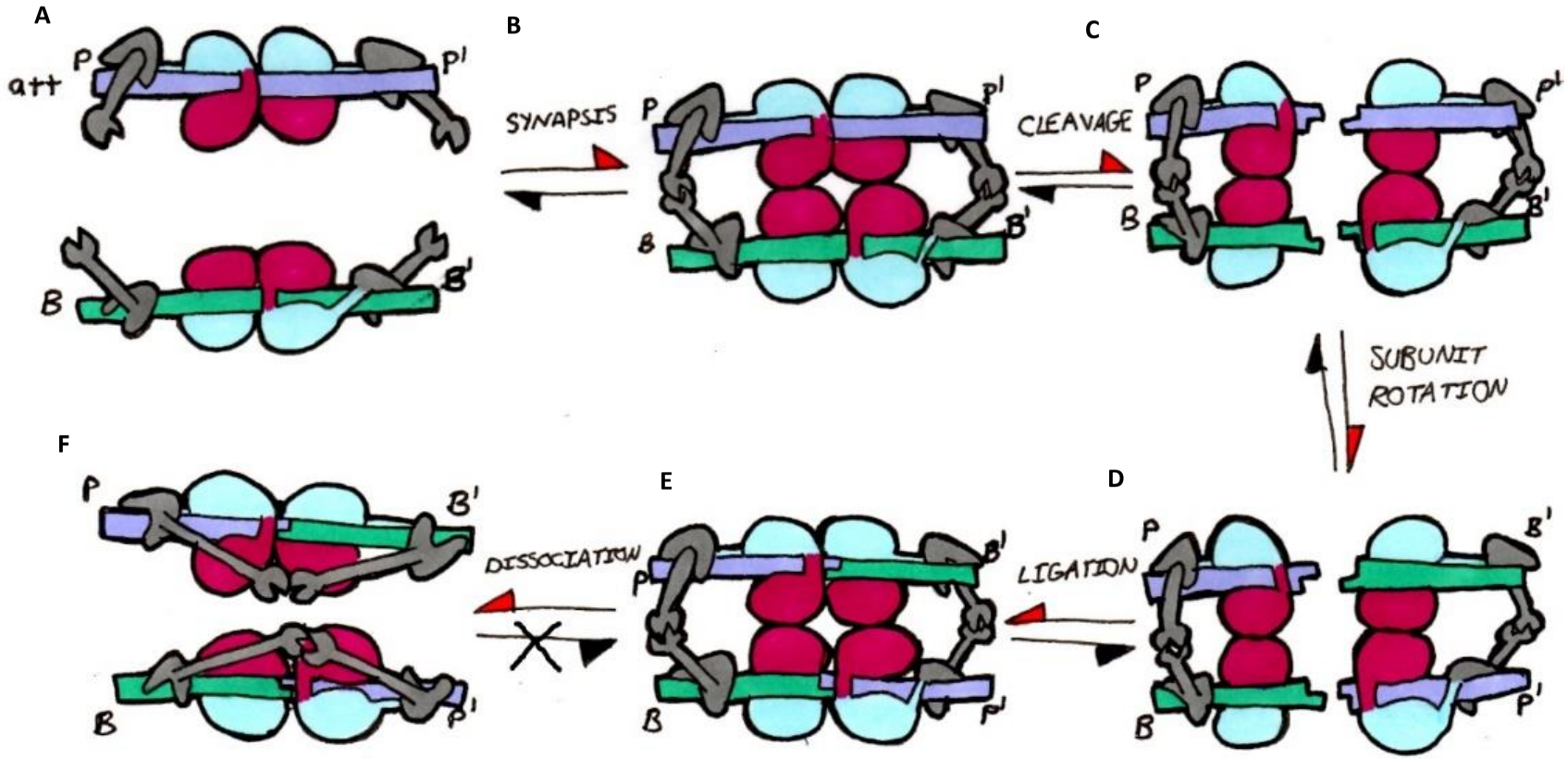

Figure 1.15: Serine integrase-catalysed site-specific recombination. Recombinase domains are shown as magenta units and catalytic domains as pale blue. DNA *attP* and *attB* sites are purple or green strips respectively. Coiled coil domains are shown in grey. Red arrows denote the forward reaction between *attP* and *attB*; the black arrows denote the reverse reaction. Grey subunits highlight the coiled coil conformational change which Rutherford and Van Duyne (2014) hypothesise is responsible for preventing the reverse reaction between *attL* and *attR* from occurring via intra-molecular interactions until the RDF is present. (A) Serine integrase binds as a dimer to its *attP* and *attB* sites, which (B) associates to form a synaptic complex. (C) 5'-phosphoserine linkages between DNA and protein subunits are formed via the coordinated cleavage of all four of the DNA strands via the catalytic domain of the integrase. 3'-overhanging dinucleotides are also formed which contain each site's crossover sequence. (D+E) The rotation of subunit pairs results in the formation of a strand exchange configuration. Providing the unpaired dinucleotides are complementary, this strand exchange conformation (not shown) will result in efficient ligation to form recombinant *attL* and *attR* sites (F). The reverse excision reaction only occurs in the presence of the RDF protein (created 2014; Rutherford and Van Duyne, 2013).

1.6 Synthetic biology

The term "synthetic biology" first appeared in 1980, when Dr Barbara Hobom used it to encompass any bacteria that had been altered using recombinant DNA technology (Hobom, 1980). Nowadays, the precise meaning of synthetic biology is debated due to the many disciplines involved in the field. However, one popular definition is "designing and constructing biological modules, biological systems, and biological machines for useful purposes" (Nakano, 2013). This definition incorporates the most prominent disciplines from biology and engineering in synthetic biology – molecular and systems biology and genetic engineering.

Because of the modular nature of many desired systems in synthetic biology, it is often a requirement for synthetically engineered biological systems that each part in a system functions independently of other modules in the same system. This allows for prediction of the final behaviour of a synthetic assembly but also enables modules from one system to be used for similar purposes in another (Benner and Sismour, 2005).

Synthetic biology has become a popular area of research in the 21st century, and has resulted in the creation of the annual iGem competition, where student-led research teams are encouraged to attempt to solve challenges found in everyday life across the globe by building genetically engineered biological systems. This has introduced the field of synthetic biology to a new generation of scientists across disciplines and will ensure the popularity of the field for years to come.

1.6.i: Uses for site-specific recombinases in synthetic biology and genome engineering

Many synthetic biologists strive to create systems which perform simple tasks, such as counting and acting as "memory" – much like computer memory - or turning a process on and off (i.e. acting as a switch). Due to the site-specific and reversible (in the presence of RDF) nature of serine integrases, these enzymes have proven themselves to be ideal for these purposes both *in vivo*

and *in vitro*. In 2012, Bonnet et al established a rewritable digital data storage system in *Escherichia coli* which was capable of maintaining passive information storage for over one hundred cell divisions by using varying ratios of BxbI integrase to its RDF to invert and restore specific DNA sequences. Additionally, this data storage was not subject to degradation despite being switched repeatedly over one hundred cell divisions (Bonnet et al, 2012).

Serine integrases have also been used as both reversible and irreversible switches, which can be tagged (for example, with green and red fluorescent protein) to demonstrate that they are on or off within a system (Figure 1.16). In a more complex setting, serine integrases can be used to target the beginning and end of genes of interest to fulfil various purposes, such as the rapid assembly of metabolic pathways containing three, four or five genes using ΦC31 integrase in *E.coli* (Colloms et al, 2013).

Because resolvase is modular in nature, its DBD and catalytic domain can by split and used separately, allowing a new DBD from a different protein to be attached to its catalytic domain in order to change the DNA recognition target for the resultant, hybrid protein (explored further in Section 1.7).

The creation of functional hybrid recombinases vastly increases the repertoire of recombinases available for use in synthetic biology and genetic engineering as well as increasing the likelihood that one might be able to create a new protein that recognises almost any desired DNA sequence, thus making recombinases all the more desirable as tailored modules in biological systems. Section 1.7 discusses in greater detail some of the functional hybrid recombinases created over the past fifteen years.

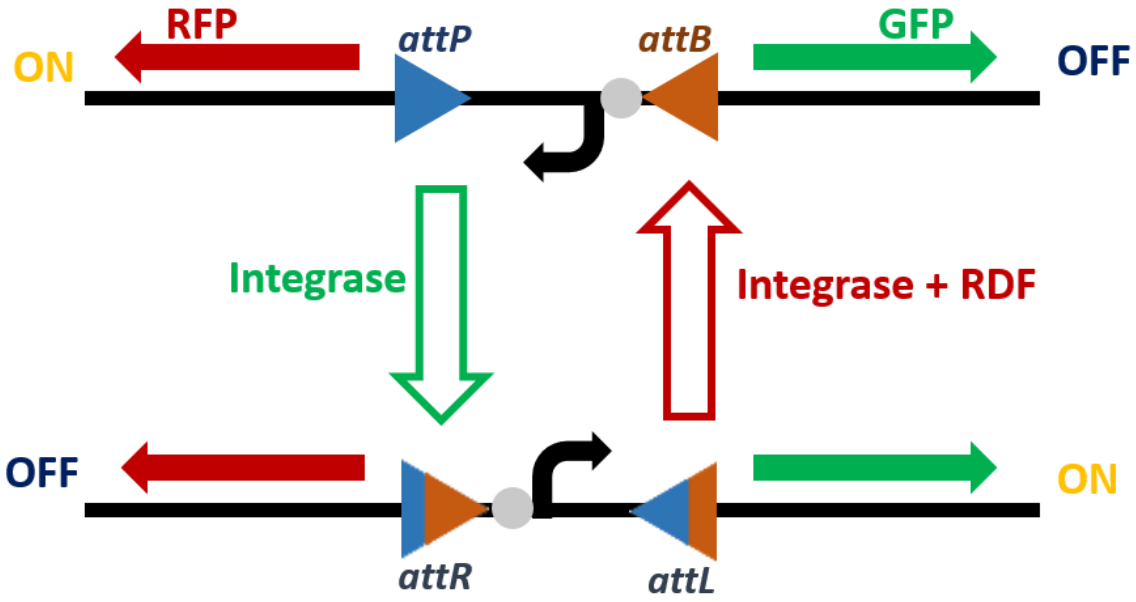

Figure 1.16: Integrase-mediated, reversible switch for *in vivo* systems in *E.coli*.

In the presence of integrase, *attP* and *attB* are recombined to form *attL* and *attR*, turning the switch on and causing the cell to glow green. In the presence of integrase and RDF, recombinant sites *attL* and *R* revert back to their original *attP* and *B* configuration, turning the switch off and causing the cell to glow red.

1.6.ii Other DNA editing/rearrangement tools useful for synthetic biology and genetic engineering

Whilst recombinases are popular and promising enzymes to use within the fields of synthetic biology and genetic engineering, other DNA modifying systems exist that are also useful in a similar context.

RNA interference

RNA interference (RNAi) is a system which has historically been subject to intensive research with regards to tailored genetic engineering. However, RNAi is often subject to off-target effects, leading to random gene down-regulation or complete silencing (Bondy-Denomy et al, 2013). Furthermore, because RNAi is ultimately an epigenetic effect, the results of RNAi on specific targets is neither permanent nor necessarily always at the same level of up- or down-regulation, leaving the system as an inefficient tool for long-term genetic engineering.

Transposons

Transposable elements are capable of moving themselves throughout or between DNA via reverse transcriptase (retrotransposons) or via transposase-mediated "cut and paste" or replicative mechanisms (DNA transposons). Transposon-derived genetic insertion vectors, such as the well-characterised DNA transposon Sleeping Beauty, have been useful systems for identifying the role of different genes (such as potential proto-oncogenes in mouse models) due to the nature of their random insertion. However, because their insertion into a genome is random it limits the use of these transposon-derived systems in genetic engineering and synthetic biology because of their potential for undesirable, off-target effects, much the same as for RNAi. Additionally, viral vectors containing strong promoters, when randomly inserted proximal to

proto-oncogenes, can lead to dysregulation and subsequent clonal expansion of mutant cells, only further consolidating the limited use for transposons in synthetic biology (Ott et al, 2006; Fischer et al, 2011; Cavazzana-Calvo et al, 2010).

Zinc finger nucleases

Zinc finger nucleases (ZFNs) are nucleases that have been engineered to target double strand breaks at specific sequences, thus initiating homology directed repair (usually homologous recombination) and non-homologous end joining (NHEJ) – both naturally occurring cellular DNA repair mechanisms (Beerli and Barbas, 2002). Structurally, ZFNs are made up of the cleavage domain of a nuclease and the DBD of specific zinc fingers e.g. Cys_2-His_2. Zinc finger domains are used because of their sequence specificity – each "finger" specifically recognises a triplet of bases in the target DNA. Because zinc finger proteins are modular, this in theory allows for different sequence specificities to be constructed simply by swapping each individual finger used in the resultant ZFN. Indeed, artificial zinc finger domains with the capability of recognising DNA sequences of varying length (between 9-18 bp) have been created, vastly increasing the usefulness of ZFNs as tools for genetic engineering (Dreier et al, 2001; Dreier et al, 2005).

Because ZFNs create double strand breaks, there is the possibility for NHEJ being employed to fix the break. NHEJ often introduces small insertions and deletions which can lead to frame-shift mutations because the system is naturally error-prone. To get around this issue, nickases – specially engineered nucleases which prevent double strand breaks and instead nick the DNA – have been employed to prevent NHEJ from occurring, causing homology directed repair to be employed instead (Gaj et al, 2012).

CRISPR/Cas

Perhaps the most well-known and well-characterised DNA modifying system from recent years is the CRISPR/Cas system. The type II prokaryotic CRISPR – short for clustered regularly interspaced short palindromic repeats – and its associated Cas proteins together form an acquired immunity system which protects the host cell from invading, foreign DNA via RNA-guided DNA cleavage through a mechanism hypothesised to be not dissimilar to that seen in RNAi (Barrangou et al, 2007). The type II CRISPR/Cas system has been organised into three distinct parts, named adaptation, biogenesis and interference. New spacers are acquired from invading DNA (from sequences known as protospacers) and incorporated into the CRISPR locus to separate direct, palindromic repeats during adaptation, whilst biogenesis deals with transcription of the new CRISPR locus to form the non-coding CRISPR RNA (crRNA) known as precursor crRNA or pre-crRNA. After pre-crRNA has been processed into short crRNAs, Cas proteins interact with crRNA to form a complex which then, during interference, provides immunity against infection by plasmids and bacteriophages from which the new spacers were acquired from in the first instance. Once the CRISPR/Cas system recognises these protospacers, it blocks replication of the foreign DNA and begins to cleave it instead (Sampson et al, 2013).

Redesigning crRNA allows for specific targeting of Cas protein cleavage – particularly that of Cas9, the most commonly utilised type II Cas protein – to desired DNA sites. This is the core of what makes the CRISPR/Cas system ideal for genetic engineering and has subsequently led to its incorporation in many laboratory settings as a commonly used research tool.

1.7 Hybrid recombinases

Several kinds of hybrid serine recombinases have been made to try and increase the repertoire of useable DNA-modifying proteins, some of which are listed in further detail here.

Zinc Finger Recombinases

Zinc finger recombinases (ZFRs) – site specific recombinases containing the catalytic domain from one of the small resolvases (for example, Tn3 or Sin resolvase) connected to the DNA binding domain of a zinc finger protein e.g. mouse transcription factor Zif268 - were pioneered by Akopian et al (2003) and expanded upon later by Prorocic et al (2011) and Proudfoot et al (2011) within the same lab. These ZFRs recognise sites known as Z-sites, which contain zinc finger binding sites flanking the crossover site sequence of the resolvase used to construct the ZFR – Figure 1.17 (Akopian et al, 2003).

Integral to the development of ZFRs has been the advent of "activated" serine recombinases – recombinases that have no need for accessory sites or accessory site proteins to recombine their main recognition site – with perhaps the most relevant activated mutants for ZFRs being Tn3 resolvase mutants. Both activated Tn3 resolvase-based ZFRs and activated Gin invertase-based ZFRs are demonstrably functional in human cell lines (Gordley et al, 2007; Gordley et al, 2009; Gaj et al, 2013).

Both Proudfoot et al (2011) and Prorocic et al (2011) created ZFRs which connected the catalytic domain of an activated Tn3 resolvase mutant to the zinc finger domain of Zif268, from an arginine triplet in Tn3 resolvase located at the end of its αE helix/beginning of its DBD (this corresponds to roughly residue 144 in γδ resolvase, which can be seen in Figure 1.9).

The zinc fingers used for ZFRs work as described in section 1.6.ii for ZFNs and, like their nuclease counterparts, ZFRs also face similar limitations. The

specificity of the central crossover sequence of the Z-site must be recognised by the resolvase catalytic domain used to construct the ZFR; because of this, a limited number of target sites can be constructed that will be recognised by ZFRs. However, in recent years directed evolution of recombinases has been employed to produce recombinases which recognise pseudo-sites i.e. sites which are similar but not identical to their original recognition sites (Sirk et al, 2014). Specifically, Proudfoot et al (2011) created a Tn3 resolvase-based ZFR which was targeted to Z-sites that contained the DNA sequence for bovine ß-casein, which could prove useful as a target for genomic engineering and integration.

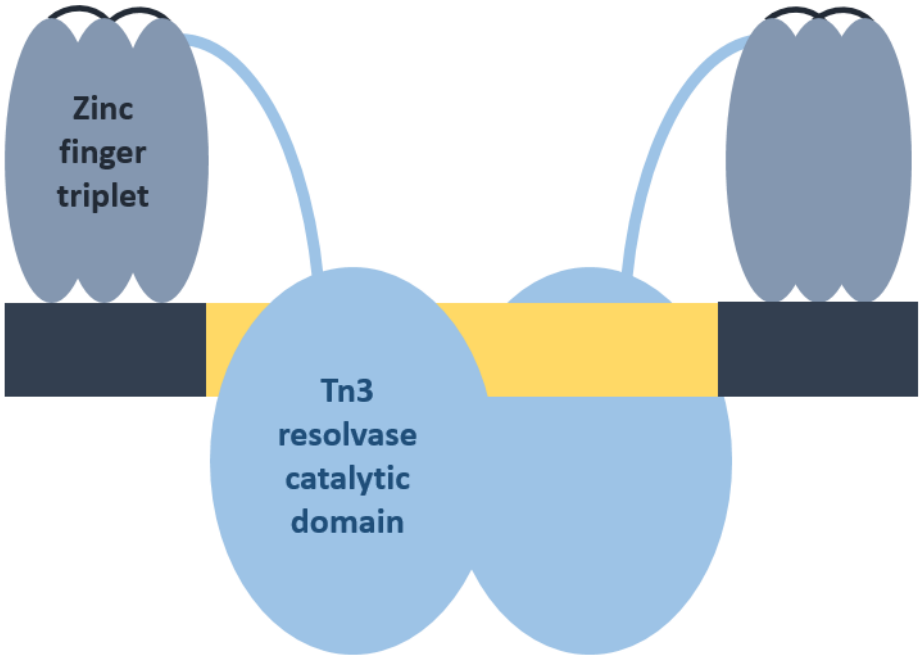

Figure 1.17: Structure of a zinc finger recombinase bound to a Z-site.

Binding sites for the zinc finger triplet are denoted in dark navy blue; the central region of Tn3 *res* site I used is denoted in pale orange. Regions of the ZFR of Tn3 resolvase origin are sky blue; those of a zinc finger origin are in navy blue. Diagrammatically, Tn3 resolvase-based ZFRs bind to DNA in much the same way as resolvase itself does (as in Figure 1.10 for γδ resolvase).

TALENS

Transcription activator like recombinases are hybrid recombinases primarily based around the DBD of TALE proteins connected to a recombinase catalytic domain (Mercer et al, 2012). TALEs – naturally occurring proteins discovered in *Xanthomonas* bacteria – are highly modifiable proteins due to the composition of their DNA binding domain. A series of 33 to 35 amino acid repeat domains within the TALE DBD recognise an individual basepair of DNA each; specificity is then determined via two hypervariable amino acids, known as repeat-variable di-residues (Joung and Sander, 2013). Repeat domains in TALE can be linked together in various orders to alter the DNA sequence recognised by each TALE protein, much like zinc fingers. However, TALE proteins do not need to undergo the stringent engineering of zinc fingers to connect these variable repeat domains, making them highly desirable tools for modular genetic engineering.

Mercer et al (2012) engineered a TALEN using the catalytic domain of a hyperactive Gin invertase mutant. Whilst they did manage to make functional TALENs, their most active variant only demonstrated approximately 7% recombination activity *in vivo*. Holt (2014) also attempted a similar TALEN construct using Tn3 resolvase activated mutants instead of Gin invertase; unfortunately, only some of the variants tested exhibited any cleavage *in vitro* and none of them appeared to exhibit any recombination activity.

With further streamlining and investigation, TALENs could prove to be more useful than ZFRs due to the TALE domain being far more easily altered to recognise different sequences. However, as of 2017, their usage remains limited compared to their ZFR counterparts.

Serine integrase hybrids in vivo

During the course of this project, hybrid integrases that displayed varying levels of activity *in vivo* on associate hybrid recognition sites were created using the serine integrases ΦC31, ΦBT1 and TG1 (Faruggio and Calos, 2014). All three integrases share a high degree of sequence similarity which subsequently made

the design process for both hybrid integrases and sites altogether much simpler than that for dissimilar proteins. In all instances, hybrids were constructed by using most of or all of the catalytic domain of one protein and the recombinase domain of the other, connecting the sequences together around the αE helix.

However, most of the constructed hybrids that proved at least partially functional in *E.coli* and/or in *HeLa* cells were those with a ΦC31 integrase catalytic domain (Faruggio and Calos, 2014), except for one ΦBT1 integrase catalytic domain/ΦC31 integrase recombinase domain hybrid. Faruggio and Calos hypothesise that a full or mostly full ΦC31 integrase catalytic domain is required to confer functionality upon the hybrid integrases, even though all three proteins are related (2014). Furthermore, they conclude that a "structurally incompatible mutation" is what renders any TG1 integrase catalytic domain/ΦC31 integrase recombinase domain hybrids inactive despite TG1 integrase being more sequentially similar to ΦC31 integrase (around 70% similar catalytic domains) than ΦBT1 integrase to ΦC31 integrase (around 40% similar catalytic domains) (Faruggio and Calos, 2014).

Faruggio and Calos also created hybrid *attP* and *attB* sites in which the central 8 or 14 bp of the site corresponded to the serine integrase used as the catalytic domain of each respective hybrid integrase i.e. if the hybrid integrase was constructed of a ΦC31 integrase catalytic domain and a TG1 recombinase domain then the central nucleotides of the associate recognition sites would be of a ΦC31 integrase *attP* or *attB* origin (2014).

However, the engineered sites were not exclusively recognised by the hybrid integrase they were designed for – various site combinations were recognised by TG1 integrase and ΦC31 integrase. For this reason, both the sites and the hybrid integrases would have limited use in a synthetic biology context, as other proteins being able to recognise their recognition sites diminishes the independent, modular nature required of serine integrases in an engineered system.

1.8 Project aims

This project aims to create hybrid serine integrases with novel DNA target sequence specificities. This would allow more widespread and complex uses for serine integrases by directly increasing the repertoire of available serine integrases that recognise varied and desirable DNA sequences for use in synthetic biology, genetic engineering and gene therapy.

This is achieved through the creation of two sets of hybrid site-specific recombinases: a set constructed between the distantly related large serine integrases ΦC31 integrase and BxbI integrase, and a set constructed between ΦC31 integrase and the smaller Tn3 resolvase. Hybrid protein construction is based on the design of successful zinc finger recombinases and hypothesised DNA-protein interactions extrapolated from various serine integrase and resolvase crystal structures.

Analogous hybrid recognition *att* sites are also engineered to test the activity of the new hybrid integrases, as well as the activity of the parental proteins on the new, hybrid *att* sites.

Chapter Two: Materials and Methods

2.1 Bacterial Strains

All *Escherichia coli* strains used in this study are derivatives of the original K-12 strain. Their names, genotypes and original sources are given in Table 2.1 below.

Bacterial strain	Genotype	Source
DS941	AB1157, *rec*F, *lacZ*ΔM15, *lacI*^{q}	Summers and Sherratt, 1988
BL21(DE3)[pLysS]	*Hsd, gal, (λcI ts 857, ind1, sam7, ini5, lac*$_{UV5}$-T7 gene-1)	Moffatt and Studier, 1986
BL21(DE3)[pLysE]	*Hsd, gal, (λcI ts 857, ind1, sam7, ini5, lac*$_{UV5}$-T7 gene-1), T7 lysozyme expressing plasmid PLysE	Moffatt and Studier, 1987
BL21(DE3)	*Hsd, gal, (λcI ts 857, ind1, sam7, ini5, lac*$_{UV5}$-T7 gene-1), T7 lysozyme expressing plasmid PLysS	Moffatt and Studier, 1987

Table 2.1: Bacterial strains used throughout this study.

2.2 Bacterial Growth Media

All liquid *E.coli* cultures were grown at 37 °C in L-Broth (10 g tryptone, 5 g yeast extract and 5 g NaCl; made up to one litre with de-ionised H_2O and

adjusted to pH 7.5 with NaOH). *E.coli* was also grown on solid media (L-broth containing 15 g/l agar).

2.3 Antibiotics

Antibiotics were all set up as stock solutions of 1000x for storage. Working concentrations of all antibiotics used in this study are shown in Table 2.2, below.

Antibiotic	Stock Concentration	Final (working) concentration
Ampicillin	100 mg/ml in H_2O	100 µg/ml
Kanamycin	50 mg/ml in H_2O	50 µg/ml
Chloramphenicol	25 mg/ml in EtOH	25 µg/ml

Table 2.2: Antibiotic stock and working concentrations.

2.4 Chemicals

Unless otherwise specified, all chemicals used throughout this study were purchased from the suppliers indicated in Table 2.3, below.

Chemical name	**Source/Supplier**
All general chemicals, organic solvents and biochemicals	Sigma/Aldrich
Media	Difco
Agarose and acrylamide	Biorad
Restriction enzymes and restriction enzyme buffers	NEB, Roche, Promega

Table 2.3: Chemical reagents and their suppliers.

2.5 Gene Synthesis

All synthetic constructs designed for use in this study were ordered from Life Technologies, GeneArt Gene Synthesis for construction.

2.6 Oligonucleotide synthesis

All oligonucleotides used throughout this study for cloning and plasmid construction were ordered from and synthesised by Eurofins Genomics (Table 2.4).

2.7 Plasmids

All plasmids were constructed through restriction digests and fragment swaps from both existing plasmids and plasmids constructed by GeneArt containing synthetic constructs. A list of all plasmids created for and used in this study can be seen in Table 2.5. The general construction methods for specific groups of plasmids with similar roles are outlined below; more specific methods e.g. restriction enzymes used, actual fragments and plasmids used can be seen in the relevant results chapter.

2.7.i High-level expression plasmids

All over-expression plasmids (for protein over-expression and purification) are based on pSA1101 (Arnold et al, 1999). pSA1101 encodes a kanamycin resistance marker and contains a ColEI origin of replication as well as an IPTG-inducible T7 promoter, which induces overexpression of any gene cloned in between the restriction sites NdeI and KpnI (in the case of pSA1101 itself, this gene encodes Tn3 resolvase).

2.7.ii Low-level expression plasmids

Low level expression plasmids, used for the *in vivo* MacConkey assay (Section 2.21), were constructed based on a pMS140 backbone (Burke et al, 2004). Like pSA1101, pMS140 contains a ColE1 origin of replication but contains an ampicillin instead of kanamycin resistance marker. A weak (unidentified) promoter is located within 400 bp upstream of the NdeI restriction site and induces expression of any gene cloned between this NdeI site and a KpnI site (pSA1101 encodes a cassetted version of the wild-type Tn3 resolvase within the same two restriction sites). This promoter induces a comparatively lower level of protein expression than the inducible, high expression vector pSA1101 and so is ideal for *in vivo* recombination assays, such as the MacConkey assay.

2.7.iii *In vitro* recombination substrate plasmids

One-site plasmids for *in vitro* recombination assays were constructed from a pMTL23 backbone – these plasmids also served as the precursor to all *in vivo* substrate plasmids. pMTL23 is a high copy number cloning vector which encodes an ampicillin resistance marker and also contains a deregulated ColEI origin of replication (Chambers et al, 1988). *Att* sites were cloned into the backbone as oligonucleotides between the restriction ends EcoRI and SacI (see Figure 3.8A or Figure 4.7A). Two-site *in vitro* substrate plasmids were then created by ligating the larger fragment from a BglII-AlwNI digested *attP* site plasmid and the smaller fragment from a BamHI-AlwNI *attB* site plasmid with the smaller, kanamycin resistance marker encoding, fragment from a BamHI digested pUC7IK plasmid (see Figure 3.8B or Figure 4.7B).

2.7.iv *in vivo* recombination substrate plasmids

Substrate plasmids used for *in vivo* recombination assays are all based on a pMS183Δ backbone; this plasmid expresses a kanamycin resistance marker and has a pSC101 low copy number origin of replication. All *in vivo* recombination substrate plasmids were constructed via two rounds of ligations, using AlwNI-NheI and AlwNI-BsrGI fragments from pMS183Δ along with a BamHI-Asp718 from a pMTL23-based single site plasmid to create a plasmid containing *attP*, and

AlwNI-XbaI and AlwNI-Asp718 fragments from pMS183Δ along with a BamHI-Asp718 from a pMTL23-based single site plasmid to create a plasmid containing *attB* (see Figure 3.9A or Figure 4.8A). Both precursor plasmids are then digested with AlwNI and NdeI, with the smaller fragment from an *attP* plasmid removed and swapped with the smaller fragment from an *attB* plasmid to create two-site *in vivo* recombination substrate plasmids (Figure 3.9B or Figure 4.8B). This cloning is made possible by the compatibility of the sticky ends made by the enzymes BsrGI and Asp718, and BglII and BamHI respectively.

Oligo name	Size (bp)	Sequence (5'-3')	Purpose
P/B attP 10nuc1	55	CAGTAGTGCCCCAACTGGGGTACGCGGTCTCATCTCTCAGTTGGGGGCGTAGGGG	To introduce ΦbΦ10 attP top strand
P/B attP 10nuc2	63	AATTCCCCTACGCCCCCAACTGAGAGATGAGACCGCGTACCCCA GTTGGGGCACTACTGAGCT	To introduce ΦbΦ10 attP bottom strand
P/B attB 10nuc1	57	CCCGCGGTGCGGGTGCCAGGGCGTGGCGGTCTCCTCCCCGGGCG CGTACTCCACCTG	To introduce ΦbΦ10 attB top strand
P/B attB 10nuc2	65	AATTCAGGTGGAGTACGCGCCCGGGGAGGAGACCGCCACGCCC TGGCACCCGCACCGCGGGAGCT	To introduce ΦbΦ10 attB bottom strand
P/B attP 12nuc1	55	CAGTAGTGCCCCAACTGGGGTCCGCGGTCTCAGCTCTCAGTTGGGGGCGTAGGGG	To introduce ΦbΦ12 attP top strand
P/B attP 12nuc2	63	AATTCCCCTACGCCCCCAACTGAGAGCTGAGACCGCGGACCCCA GTTGGGGCACTACTGAGCT	To introduce ΦbΦ12 attP bottom strand
P/B attB 12nuc1	57	CCCGCGGTGCGGGTGCCAGGGCGCGGCGGTCTCCGCCCCGGGCG CGTACTCCACCTG	To introduce ΦbΦ12 attB top strand
P/B attB 12nuc2	65	AATTCAGGTGGAGTACGCGCCCGGGGCGGAGACCGCCGCGCCC TGGCACCCGCACCGCGGGAGCT	To introduce ΦbΦ12 attB bottom strand
phiC31 top 1	61	GACCGCTCCGTATGGTTTTGAACTGGTTAGCGAAACCAAAGAAA TCACACGTAATGGTCGT	Long linker sequence part 1 between BxbI integrase and ΦC31 integrase
phiC31 bottom 1	41	CTTTGGTTTCGCTAACCAGTTCAAAACCATACGGAGCGGTC	Long linker sequence part 1 between BxbI integrase and ΦC31 integrase
phiC31 top 2	52	ATGGTTAACGT GGTTATTAACAAACTGGCACATAGCACCACACCGCTGACGT	Long linker sequence part 2 between BxbI integrase and ΦC31 integrase
phiC31 bottom 2	72	ACGTCAGCGGTGTGGTGCTATGTGCCAGTTTGTTAATAACCACGT TAACCATACGACCATTACGTGTGATTT	Long linker sequence part 2 between BxbI integrase and ΦC31 integrase

Table 2.4: Oligonucleotides used in this project.

All oligonucleotides were ordered from MWG as two sequences: top strand (denoted with either a “T” or “top” in the oligo name) and bottom strand (denoted with either a “B” or “bot” in the oligo name).

Oligo name	Size (bp)	Sequence (5'-3')	Purpose
BxbI top	51	GACCGCTCCGTGGGGTTATCTGCCGACCCGTGTTGACGGTGAATGGCGTGT	Short linker sequence between BxbI integrase and ΦC31 integrase
BxbI bottom	51	ACACGCCATTCACCGTCAACACGGGTCGGCAGATAACCCCACGGAGCGGTC	Short linker sequence between BxbI integrase and ΦC31 integrase
BpB 10 attP top	55	CTGGTTTGTCTGGTCAACCACACCTTTGAGTGTGGTGTACGGTACAAACCCGGGG	To introduce bΦb10 attP top strand
BpB 10 attP bot	65	AATTCCCCGGGTTTGTACCGTACACCACACTCAAAGGTGTGGTT GACCAGACAAACCAGAGCT	To introduce bΦb10 attP bottom strand
BpB 10 attB top	53	CTCGGCCGGCTTGTCGACGACGCCCTTGGGCGTCGTCAGGATCATCCGGGCTG	To introduce bΦb10 attB top strand
BpB 10 attB bot	64	AATTCAGGTGCCCGGATGATCCTGACGACGCCCAAGGGCGTCG TCGACAAGCCGGCCGAGAGCT	To introduce bΦb10 attB bottom strand
11 top new	54	GCCCCGTATGGTTTTCTGCCGACCCGTGTTGACGGTGAATGGCGTCTGGTTCCG	Linker sequence between ΦC31 integrase and BxbI integrase
11 bot new	55	GATCCGGAACCAGACGCCATTCACCGTCAACACGGGTCGGCAG AAAACCATACGG	Linker sequence between ΦC31 integrase and BxbI integrase
Tphi linker1 T	15	GCCCGCAGGCGTAAG	Short linker sequence between NM resolvase and ΦC31 integrase
Tphi linker1 B	14	GCCCTTACGCCTGC	Short linker sequence between NM resolvase and ΦC31 integrase
Tphi linker2 T	24	GCCCGCAGGCGTACCGTGGACAGG	Long linker sequence between NM resolvase and ΦC31 integrase
Tphi linker2 B	23	GGCCCTGTCCACGGTACGCCTGC	Long linker sequence between NM resolvase and ΦC31 integrase

Table 2.4 (continued).

Oligo name	Size (bp)	Sequence (5'-3')	Purpose
T/P attP 12 T	55	CAGTAGTGCCCCAACTGGGGTATATTATAAATTCTCTCAGTTGGGGGCGTAGGGG	To introduce ΦtΦ12 attP top strand
T/P attP 12 B	63	AATTCCCCTACGCCCCCAACTGAGAGAATTTATAATATACCCCA GTTGGGGCACTACTGAGCT	To introduce ΦtΦ12 attP bottom strand
T/P attB 12 T	57	CCCGCGGTGCGGGTGCCAGGGCGATATTATAAATTCCCCGGGC GCGTACTCCACCTG	To introduce ΦtΦ12 attB top strand
T/P attB 12 B	65	AATTCAGGTGGAGTACGCGCCCGGGGAATTTATAATATCGCCC TGGCACCCGCACCGCGGGAGCT	To introduce ΦtΦ12 attB bottom strand
T/P attP 14 T	55	CAGTAGTGCCCCAACTGGGGAATATTATAAATTATCTCAGTTGGGGGCGTAGGGG	To introduce ΦtΦ14 attP top strand
T/P attP 14 B	63	AATTCCCCTACGCCCCCAACTGAGATAATTTATAATATTCCCCA GTTGGGGCACTACTGAGCT	To introduce ΦtΦ14 attP bottom strand
T/P attB 14 T	57	CCCGCGGTGCGGGTGCCAGGGCAATATTATAAATTACCCGGGC GCGTACTCCACCTG	To introduce ΦtΦ14 attB top strand
T/P attB 14 B	65	AATTCAGGTGGAGTACGCGCCCGGGTAATTTATAATATTGCCC TGGCACCCGCACCGCGGGAGCT	To introduce ΦtΦ14 attB bottom strand

Table 2.4 (continued).

Plasmid name	Antibiotic resistance	Size (bp)	Description	Source
PHIC31_altered	Ampicillin	3204	GeneArt plasmid. Altered section of ΦC31 integrase DNA sequence (no effect on reading frame) in a pMA-T background	GeneArt
BxbI_altered	Ampicillin	2910	GeneArt plasmid. Altered section of BxbI integrase DNA sequence (no effect on reading frame) in a pMA-T background	GeneArt
pFEM14	Kanamycin	7974	ΦC31 integrase expression plasmid with inducible T7 promoter	F. Olorunniji
pFO2	Ampicillin	5469	pMS140 derivative, encoding Tn3 resolvase mutant NM (R2A, E56K, G101S, D102Y, M103I, Q105L) ORF	F. Olorunniji
pFM12	Ampicillin	2687	*in vitro* recombination substrate containing one copy of ΦC31 *attB*	F. Olorunniji
pUC71K	Kanamycin	3966	Cloning vector and source of the kanamycin resistance gene	Taylor & Rose, 1982
pMS183Δ	Kanamycin	4863	*in vivo* recombination substrate precursor plasmid, allowing cloning of *att* sites between NheI/BsrGI and XbaI/Asp718	M. Prorocic
pGD001	Kanamycin	5181	*In vivo* recombination substrate containing ΦC31 *attP* and *attB* in direct repeat flanking a *gal* K gene; pMS140 backbone	F. Olorunniji
pFM124	Kanamycin	5241	*In vivo* recombination substrate containing BxbI *attP* and *attB* in direct repeat flanking a *gal* K gene; pMS140 backbone	F. Olorunniji
pFEM32	Ampicillin	6751	ΦC31 integrase low expression plasmid with inducible T7 promoter for use *in vivo*	F. Olorunniji
pFEM35	Ampicillin	6463	BxbI integrase low expression plasmid with inducible T7 promoter for use *in vivo*	F. Olorunniji
pFM16	Kan/Amp	4201	*In vitro* recombination substrate containing ΦC31 *attP* and *attB* , flanking the Kanamycin resistance marker from pUC71K	F. Olorunniji

Table 2.5: Plasmids constructed for this project and their precursors.

Plasmid names highlighted according to plasmid type: Light blue: single site *in vitro* recombination plasmids; blue: 2-site *in vitro* recombination substrate plasmid; pale pink: precursor *in vivo* recombination substrate plasmids; pink: 2-site *in vivo* recombination substrate plasmid; lilac: pMS140 based low level expression plasmids and purple: high level expression plasmid.

Plasmid name	Antibiotic resistance	Size (bp)	Description	Source
pFM97	Kan/Amp	4201	*In vitro* recombination substrate containing BxbI *attP* and *attB*, flanking the Kanamycin resistance marker from pUC71K	F. Olorunniji
pSDC384	Kanamycin	7686	BxbI integrase expression plasmid with inducible T7 promoter	S. Colloms
pMP243	Ampicillin	4960	*In vivo* recombination substrate containing two copies of Tn3 *res* site I in direct repeat flanking a *gal* K gene; pMS140 backbone	M. Prorocic
pCO1	Ampicillin	2543	pMTL23 with a Tn3 *res* site I cloned between its EcoRI and SacI sites	C. Muir/M. Stark
pMP78	Kan/Amp	3919	*In vitro* recombination substrate containing two copies of Tn3 *res* site I, flanking the Kanamycin resistance marker from pUC71K	M. Prorocic
pFM14	Ampicillin	2687	*in vitro* recombination substrate containing one copy of ΦC31 *attB*	F. Olorunniji
pHM1	Ampicillin	2671	*in vitro* recombination substrate containing one copy of ΦbΦ10 *attP*	Section 3.3.ii
pHM1*	Kanamycin	5027	*in vivo* precursor substrate plasmid containing one copy of ΦbΦ10 *attP*	Section 3.3.ii
pHM2	Ampicillin	2673	*in vitro* recombination substrate containing one copy of ΦbΦ10 *attB*	Section 3.3.ii
pHM2*	Kanamycin	4995	*in vivo* precursor substrate plasmid containing one copy of ΦbΦ10 *attB*	Section 3.3.ii
pHM3	Ampicillin	2671	*in vitro* recombination substrate containing one copy of ΦbΦ12 *attP*	Section 3.3.ii
pHM3*	Kanamycin	5027	*in vivo* precursor substrate plasmid containing one copy of ΦbΦ12 *attP*	Section 3.3.ii

Table 2.5 (continued).

Plasmid name	Antibiotic resistance	Size (bp)	Description	Source
pHM4*	Kanamycin	4995	*in vivo* precursor substrate plasmid containing one copy of ΦbΦ12 *attB*	Section 3.3.ii
pHM7	Kanamycin	5159	*In vivo* recombination substrate containing ΦbΦ10 *attP* and *attB* in direct repeat flanking a *gal* K gene	Section 3.3.ii
pHM8	Kanamycin	5159	*In vivo* recombination substrate containing ΦbΦ12 *attP* and *attB* in direct repeat flanking a *gal* K gene	Section 3.3.ii
pHM9x	Ampicillin	3108	BΦ1 precursor expression plasmid	Section 3.3.i
pHM9	Kanamycin	7878	BΦ1 expression plasmid with inducible T7 promoter	Section 3.3.i
pHM9*	Ampicillin	6666	BΦ1 low expression plasmid with inducible T7 promoter for use *in vivo*	Section 3.3.i
pHM10x	Ampicillin	3165	BΦ2 precursor expression plasmid	Section 3.3.i
pHM10	Kanamycin	7965	BΦ2 expression plasmid with inducible T7 promoter	Section 3.3.i
pHM10*	Ampicillin	6723	BΦ2 low expression plasmid with inducible T7 promoter for use *in vivo*	Section 3.3.i
pHM11*	Ampicillin	2946	ΦB1 precursor expression plasmid	Section 3.7.i
pHM11	Kanamycin	7692	ΦB1 expression plasmid with inducible T7 promoter	Section 3.7.i
pHM12	Kan/Amp	4201	*In vitro* recombination substrate containing ΦbΦ10 *attP* and *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 3.3.ii

Table 2.5 (continued).

Plasmid name	Antibiotic resistance	Size (bp)	Description	Source
pHM15	Ampicillin	2685	*in vitro* recombination substrate containing one copy of bΦb10 *attP*	Section 3.7.ii
pHM16	Ampicillin	2686	*in vitro* recombination substrate containing one copy of bΦb10 *attB*	Section 3.7.ii
pHM23	Ampicillin	6469	ΦB1 low expression plasmid with inducible T7 promoter for use *in vivo*	Section 3.7.i
pHM28	Kanamycin	5183	*In vivo* recombination substrate containing bΦb10 *attP* and *attB* in direct repeat flanking a *gal* K gene	Section 3.7.ii
pHLM1	Ampicillin	2671	*in vitro* recombination substrate containing one copy of ΦtΦ14 *attP*	Section 3.7.ii
pHLM1*	Kanamycin	5027	*in vivo* precursor substrate plasmid containing one copy of ΦtΦ14 *attP*	Section 4.3.ii
pHLM2	Ampicillin	2673	*in vitro* recombination substrate containing one copy of ΦtΦ14 *attB*	Section 4.3.ii
pHLM2*	Kanamycin	4995	*in vivo* precursor substrate plasmid containing one copy of ΦtΦ14 *attB*	Section 4.3.ii
pHLM3	Ampicillin	2671	*in vitro* recombination substrate containing one copy of ΦtΦ12 *attP*	Section 4.3.ii
pHLM3*	Kanamycin	5027	*in vivo* precursor substrate plasmid containing one copy of ΦtΦ12 *attP*	Section 4.3.ii
pHLM4	Ampicillin	2673	*in vitro* recombination substrate containing one copy of ΦtΦ12 *attB*	Section 4.3.ii
pHLM4*	Kanamycin	4995	*in vivo* precursor substrate plasmid containing one copy of ΦtΦ12 *attB*	Section 4.3.ii

Table 2.5 (continued).

Plasmid name	Antibiotic resistance	Size (bp)	Description	Source
pHLM6	Kanamycin	7908	TΦ8 expression plasmid with inducible T7 promoter	Section 4.3.i
pHLM7	Kanamycin	5159	*In vivo* recombination substrate containing ΦtΦ14 *attP* and *attB* in direct repeat flanking a *gal* K gene	Section 4.3.ii
pHLM8	Kanamycin	5159	*In vivo* recombination substrate containing ΦtΦ12 *attP* and *attB* in direct repeat flanking a *gal* K gene	Section 4.3.ii
pHLM10	Ampicillin	6687	TΦ4 low expression plasmid with inducible T7 promoter for use *in vivo*	Section 4.3.i
pHLM11	Ampicillin	6696	TΦ8 low expression plasmid with inducible T7 promoter for use *in vivo*	Section 4.3.i
pHLM12	Kan/Amp	4201	*In vitro* recombination substrate containing ΦtΦ14 *attP* and *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 4.3.ii
pHLM13	Kan/Amp	4201	*In vitro* recombination substrate containing ΦtΦ12 *attP* and *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 4.3.ii
pHLM22	Kan/Amp	4178	*In vitro* recombination substrate containing one copy of Tn3 *res* site I and one copy of ΦtΦ14 *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 4.6.ii
pHLM24	Kan/Amp	4178	*In vitro* recombination substrate containing one copy of Tn3 *res* site I and one copy of ΦtΦ12 *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 4.6.ii
pHLM33	Kan/Amp	4201	*In vitro* recombination substrate containing one copy of ΦtΦ12 *attP* and one copy of ΦC31 *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 5.2
pHLM34	Kan/Amp	4201	*In vitro* recombination substrate containing one copy of ΦC31 *att P* and one copy of ΦtΦ12 *attB*, flanking the Kanamycin resistance marker from pUC71K	Section 5.2

Table 2.5 (continued).

2.8 Competent cell procedure

Two kinds of competent cells were used throughout this study; where standard DNA transformation was required, chemically competent cells were used, and where high transformation efficiency was required, electro-competent cells were used. The methods by which the competent cells were created are detailed below.

2.8.i Chemically competent *E.coli* cells

A 200 µl sample from the desired *E.coli* overnight culture was used to inoculate 10 ml of L-broth, which was then grown in a New Brunswick Scientific Excella E24 Incubator Shaker at 250 rpm, set to a temperature of 37 °C. Once the culture had grown to mid-log phase (OD^{600} of 0.4-0.6), the culture was divided into four aliquots of 1.2 ml and chilled on ice for twenty minutes. Cells were then harvested at 4 °C by centrifugation at 4 °C for sixty seconds at 10000 rpm, after which the supernatant was removed. The cell pellet was re-suspended in 1.2 ml of ice-cold 50 mM $CaCl_2$, after which centrifugation was repeated as before. Following the second centrifugation step, the supernatant was removed and the cell pellet was re-suspended in 200 µl ice-cold 50 mM $CaCl_2$. Cells prepared by this method were then stored on ice in a 4 °C cold room, remaining competent for up to 48 hours.

2.8.ii Electrocompetent *E.coli* cells

Electrocompetent cells were prepared by inoculating 500 ml of L-broth with 2 ml of an overnight *E.coli* culture. This culture was then grown at 37 °C with shaking at 250 rpm (New Brunswick Scientific Excella E24 Incubator Shaker) to an OD600 of 0.4-0.6 (mid-log phase). The culture was subsequently split into two 250 ml aliquots and chilled on ice for twenty minutes. Cells were harvested by centrifugation in a pre-cooled Beckman Coulter JA-14 rotor at 4 °C for 15 minutes at 4500 rpm. The supernatant was removed; the cell pellet was then re-suspended in 500 ml ice-cold, filtered 10% glycerol and centrifugation was repeated. The resulting cell pellet was re-suspended in 250 ml ice cold 10% glycerol and the centrifugation was repeated. The resultant cell pellet was

then re-suspended in 20 ml ice-cold 10% glycerol, and the suspension was transferred to a 30 ml polypropylene tube and centrifuged (Beckman coulter JA-20 rotor) at 4 °C for 15 minutes at 4500 rpm. Lastly, the pellet was re-suspended in 2 ml ice-cold 10% glycerol. The cells were then divided into 50 µl aliquots and stored at -70 °C. Cells prepared in this way can be stored for up to six months.

2.9 Transformation of DNA

2.9.i Using chemically competent cells

All chemically competent cell transformations were performed using 100 µl of chemically competent cells mixed with 0.01-0.1 µg of plasmid DNA. If the plasmid DNA was the result of a ligation, the mixture was then incubated on ice for 20 minutes; if the plasmid DNA was supercoiled, this step was skipped. The cells were then heat-shocked at 42 °C in a water bath for 2 minutes, followed by a 5 minute incubation on ice. 1 ml of L-broth was subsequently added the cells, which were then incubated at 37 °C for 60-90 minutes to allow for expression of the plasmid. 100 µl of this liquid culture were spread on selective L-agar plates (again dependent on the antibiotic resistance marker present on the DNA) and incubated at 37 °C for sixteen hours.

2.9.ii Using electrocompetent cells

0.01 µg of DNA was added to a 50 µl aliquot of electrocompetent cells and transferred to an ice-cold electroporation cuvette after a five minute incubation step on ice. An electrical pulse was delivered by a Biorad Micropulser; 500 µl L-broth was added immediately following this. The culture was then transferred to 15 ml Corning tubes and incubated at 37 °C with shaking at 250 rpm (New Brunswick Scientific Excella E24 Incubator Shaker) for 60-90 minutes (dependent on the antibiotic resistance marker present on the plasmid DNA). After this recovery stage, 100 µl of the culture was spread on selective agar plates and incubated for sixteen hours at 37 °C.

2.10 Plasmid DNA preparation

Plasmid DNA for use in all experiments was prepared on a small scale using a Qiagen plasmid mini-prep kit, which uses a silica-gel membrane in order to bind DNA. All mini-preps were carried out from cell pellets created by spinning down 4.5 ml of bacterial culture via centrifugation.

2.11 Ethanol precipitation of DNA

In some instances, it was necessary to purify ligations as well as Qiagen mini-prepped plasmid DNA further; this was performed via ethanol precipitation. A solution of ammonium acetate was added to give a final concentration of 0.5 M, after which a 2.5x sample volume of 100% ethanol was then added. This mix was then incubated on ice for a minimum of 20 minutes (or up to sixteen hours), followed by centrifugation at 4 °C for 60 minutes at 13000 rpm. The supernatant was removed and the pellet was washed with 100 µl 80% ethanol. Centrifugation was repeated at 4 °C at 13000 rpm for 30 minutes; the resultant supernatant was removed and the final pellet was then re-suspended in 10 µl ddH_2O.

2.12 DNA restriction enzyme digests

Restriction enzymes from various suppliers (see Table 2.3) were used to digest DNA, using 2-10 units of enzyme per µg of DNA and the recommended buffer system for each enzyme. Single restriction enzyme digests were typically incubated at 37 °C for one hour, whereas double restriction enzyme digests were incubated at 37 °C for two hours. After restriction digest but prior to loading on an agarose gel, samples were mixed with SDS loading buffer (Section 2.15) in a 1:5 ratio in order to stop the digest reaction by denaturation of the restriction enzyme.

2.13 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate restriction digest fragments. Biorad ultrapure agarose was dissolved in 100 ml TAE buffer (40 mM Tris/acetate; 1 mM EDTA; pH8.5) by heating in a microwave oven for a final weight/volume of agarose of between 1-1.2%. The final percentage used was dependent on the intended use of the gel – if DNA bands were to be visualised only, then an agarose gel of 1.2% w/v was made; if DNA bands were intended to be cut from the gel for use in ligations or to be transformed, 1% w/v agarose gels were typically set up. Once the agarose had dissolved in TAE buffer, the solution was cooled and subsequently poured onto a gel plate and fitted with a well-forming comb. Once set, this comb was removed and the gel was immersed in TAE buffer. Samples were then loaded onto the gel and the gel was run at room temperature for various times and different voltages dependent on the experiment; typically, gels were run at 125 V for 60-90 minutes. Gels were then stained and visualised following the protocol in Section 2.16.

2.14 DNA marker ladder

An NEB 1kb or 1kb^{+} DNA fragment size ladder was used as and where appropriate for DNA fragment size estimation on agarose gels.

2.15 Sample loading buffers for agarose gel electrophoresis

SDS loading buffer (1% SDS, 50% glycerol, 0.01% Bromophenol blue) was added to samples to be loaded onto agarose gels for electrophoresis in a 1:5 ratio of loading buffer to sample. In the case of samples from *in vitro* cleavage and recombination reactions (Section 2.24), the same ratio of SDS-K loading buffer was used instead (1% SDS, 50% glycerol, 0.01% Bromophenol blue, 1 mg/ml protease K). Protease K digests any protein present, allowing for appropriate visualisation of DNA bands on the agarose gel.

2.16 DNA visualisation via ethidium bromide staining

DNA bands on agarose gels were visualised by staining with a 0.6 µg/ml ethidium bromide (EtBr) solution (15 mg/ml EtBr stock, added to 1x TAE gel running buffer once the gel had finished running) for 30-60 minutes. Gels stained in this way were de-stained by soaking in de-ionised water for a further 30-60 minutes to remove background ethidium bromide fluorescence. EtBr-stained bands were visualised on a long wavelength (365 nm) UV transilluminator for preparative gels where bands were to be extracted for DNA purification purposes, or on a Biorad short wavelength (254 nm) UV transilluminator for visualisation and photographing of gels.

2.17 DNA gel extraction

DNA was extracted from agarose gels by cutting out the required DNA band from the gel (visualised by 350 nm transillumination of an EtBr-stained gel) using a scalpel, and then recovering the DNA either by filtering through a 0.45 µM Costar column by centrifugation for 1 minute at 10,000 rpm or using a Qiagen gel extraction kit.

2.18 Annealing of oligonucleotides

Oligonucleotides obtained from MWG Eurofins Operon were re-suspended in the volume of TE buffer (10 mM Tris/HCl, 0.1 mM EDTA) specified by the provided oligonucleotide synthesis sheet, resulting in a final DNA concentration of 100 µM. These single-stranded oligonucleotides were then annealed by mixing the top and bottom strands in a 1:1 ratio with 100 mM NaCl and 1x TE buffer, heating to 90 °C for ten minutes and then leaving to cool back down to room temperature before any subsequent usage.

2.19 Ligation of DNA restriction digest fragments

DNA fragments were ligated in Invitrogen ligation buffer (50 mM Tris/HCl [pH 7.6], 10 mM MgCl2, 1 mM Dithiothreitol [DTT], 5% w/v polyethylene glycol-8000), using 1 unit of T4 DNA ligase in a final volume of 20 µl. The ratio of vector to insert was typically 1:3. Ligation reactions were carried out at room temperature for sixteen hours and used to transform electro-competent *E.coli* cells (Section 2.9.ii).

2.20 Sequencing plasmid DNA

Sequencing of DNA samples was carried out by MWG Eurofins Genomics (Ebersberg, Germany). Plasmid DNA was prepared as recommended by MWG. For high copy number expression vectors, the MWG T7 sequencing primer was routinely used and for the pMTL23-type substrate vectors, the MWG M13-rev (-49) sequencing primer was typically used.

2.21 *In vivo* recombination assay

E.coli cells (DS941 strain, which is *galk*-) containing an *in vivo* recombination substrate plasmid (Section 2.7.iv) were made competent (Section 2.8), using the chemically competent cell method and then transformed (Section 2.8.1) with a low-level recombinase expression plasmid (Section 2.7.ii). Following transformation, cells were plated out on MacConkey indicator agar plates (Difco) containing kanamycin (to select for the substrate plasmid), ampicillin (to select for the recombinase expression plasmid) and 2% w/v galactose, and incubated for sixteen hours at 37 °C. Plates were photographed (Nikon Coolpix p510). A sample of cells was then scraped off the plates and grown for sixteen hours in L-broth. Plasmid DNA isolated from pelleted cells was then analysed by agarose gel electrophoresis following EtBr staining of the DNA. Alternatively, as for Section 3.4, all cells were scraped off the plates and grown for sixteen hours in L-broth then mini-prepped as described in Section 2.10. These samples were then analysed by agarose gel electrophoresis, followed by EtBR staining of DNA. The samples were also used to transform chemically competent cells

which were then plated on MacConkey agar plates containing ampicillin (selecting for the substrate plasmid only). These plates were then photographed (Nikon Coolpix p510). Both the agarose gel and new MacConkey plates were used as a means of observing very low levels of recombination that may not have been detectable on the original MacConkey plates.

2.22 Overexpression and purification of his-tagged proteins

Overexpression of proteins

Expression plasmids with T7 inducible promoters and polyhistidine tags (Section 3.2, table 2.5) were induced with isopropylthio-ß-galactoside (IPTG; final concentration of 1 mM) in the E.coli strain BL21 (DE3) [PlysS] after the cells had grown to an OD_{600} of between 0.5-0.6; 400 ml cultures were left for sixteen hours at 20 °C with shaking at 200 rpm once IPTG had been added. These cultures were then centrifuged for ten minutes at 9000 rpm (Beckman JA-14 rotor) to pellet the cells and remove the supernatant.

Purification of proteins

Cell pellets obtained from 400 ml of culture were resuspended in lysis buffer (3.8 mM di-sodium phosphate [Na_2HPO_4], 16.2 mM monosodium phosphate [NaH_2PO_4], 1 M NaCl, 1 mM DTT, 50 mM imidazole and 1 mM PMSF) and were harvested and broken by sonication at 40% amplitude (Sonics & Materials, VibraCell™) by three, twenty second sonication pulses of cells, with five minute steps on ice when the sample was stirred between each sonication. 30 µl of 100 mM PMSF was added immediately after the first sonication. Samples were then homogenised on ice for ten minutes before being centrifuged at 18500 rpm (Beckman JA-20 rotor) for thirty minutes at 4 °C.

Samples prepared as above were then subject to HPLC gradient fractionation using the ÄKTA purifier with a 1 ml HisTrap HP column attached. Samples were added to nickel affinity HisTrap columns and washed with 20 ml buffer A (3.8 mM di-sodium phosphate [Na_2HPO_4], 16.2 mM monosodium phosphate

[NaH_2PO_4], 1 M NaCl, 1 mM DTT and 50 mM imidazole), followed by a gradient of buffer A to buffer B (3.8 mM di-sodium phosphate [Na_2HPO_4], 16.2 mM monosodium phosphate [NaH_2PO_4], 1 M NaCl, 1 mM DTT and 500 mM imidazole) of 100% buffer B over fifty minutes, with fractions collected from the beginning of the gradient. Fractions eluted when a peak for an absorbance reading of 280 nm appeared on the ÄKTA purifier multi-wavelength UV monitor were retained; all other fractions were discarded.

Chosen fractions were further purified and concentrated via dialysis, using dialysis tubing. Tubes were immersed in glycerol dialysis buffer (25 nM Tris.HCl pH 7.5, 50% glycerol, 1 M NaCl and 1 mM DTT) for sixteen hours on ice in a 4 °C cold room. Purified protein was then harvested and transferred to Nunc tubes to be stored at -20 °C.

2.23 Analysis of protein via Laemmli gel SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE; Laemmli, 1970) was then used to determine which eluted fractions contained the purest protein samples. SDS-PAGE was also used to determine if the desired protein was expressed in induced versus uninduced cell cultures containing high-expression plasmids.

Samples were run through two gel layers of different ionic strength and acrylamide percentage. The first gel – the stacking gel - was typically 5% acrylamide (5% acrylamide/bisacrylamide (37.5:1), 125 mM Tris/HCl (pH 6.8), 0.1% SDS, 0.1% APS and 0.2% TEMED), whilst the second gel – the resolving gel - was typically 10% (10% acrylamide/bisacrylamide (37.5:1), 375 mM Tris/HCl (pH 8.8), 0.1% SDS, 0.1% APS and 0.2% TEMED). The resolving gel was poured between clamped, sealed glass plates and covered with 1 ml 100% isopropanol to exclude oxygen to aid polymerisation. When set, the isopropanol was washed off, the stacking gel solution was poured in and a well-forming comb was inserted. After approximately thirty minutes of polymerisation, the comb was removed and the gel was inserted into a vertical electrophoresis kit. The gels

were run in Laemmli electrophoresis buffer (25 mM Tris base, 250 mM glycine and 0.1% SDS).

Prior to loading of protein samples, 2x Laemmli loading buffer (LLB; 20% v/v glycerol, w/v 4% SDS, 100 mM Tris/HCl [pH 6.8], 200 mM DTT and 0.2% Bromophenol blue) was added at a ratio of 1:1 buffer to sample. For non-purified samples i.e. induced cell culture, 200 µl of cell culture was centrifuged for two minutes at 13000 rpm and the supernatant removed. The resultant pellet was then resuspended in 50 µl 10 mM Tris.HCl (pH 6.8) and then 50 µl 2xLLB was added. All samples were heated to 95 °C for 5 minutes to denature the proteins. Samples were then cooled and centrifuged at 13000 rpm for five minutes. 20 µl of each sample was then loaded onto the gel and run at 30-40 mA (200 V) for 2.5 - 3 hours. The gels were then stained in Coomassie stain (2g Coomassie Brilliant Blue G, 450 ml 100% methanol and 100 ml 100% glacial acetic acid and ddH_2O to 1 litre) for 45 minutes and destained in Coomassie blue destain (50 ml 100% methanol, 70 ml 100% glacial acetic acid and ddH_2O to 1 L) for up to sixteen hours. Gels were then placed between two transparent, acetate sheets and scanned into a computer to obtain digital images of 300 dpi quality.

2.24 *In vitro* assays

2.24.i Recombination assay

For *in vitro* recombination assays, samples were prepared as follows: 2 µl of 4 µM protein, 8 µl of 125 µg/ml substrate plasmid, 2 µl ddH_2O and 28 µl buffer B105 (10 mM Tris.HCl pH 7.5, 0.1 mM EDTA pH 8.0, 5 mM spermidine and 0.1 mg/ml BSA). For control samples, integrase dilution buffer (IDB; 25 mM Tris.HCl pH 7.5, 1M NaCl, 50% glycerol and 1mM DTT) was used in place of protein. In all instances, protein was added to each sample last.

Samples were incubated at 30 °C for sixteen hours (unless otherwise specified), after which the reaction was stopped by heating samples at 80 °C for 10 minutes. Samples were then typically digested with the restriction enzyme NruI

for one hour (Section 2.12), then treated with 7.5 µl of SKE loading buffer (Section 2.15) and incubated at 37 °C for 15 – 30 minutes before being loaded onto a 1.2% agarose gel (Section 2.13), which was typically run for two hours at 100 V. If supercoiled samples were to be viewed on the agarose gel, the digest with NruI was omitted.

2.24.ii 40% ethylene glycol cleavage assay

In vitro cleavage assay samples were set up as follows: 2 µl of 4 µM protein, 4 µl of 125 µg/ml substrate plasmid and 14 µl of IRB3 (10 mM Tris.HCl pH 7.5, 0.1 mM EDTA pH 8.0, 5 mM spermidine, 0.1 mg/ml BSA and 40% v/v 100% ethylene glycol). In all instances, protein was added to each sample last.

Unless otherwise specified, samples were incubated at 30 °C for sixteen hours, after which the reaction was stopped by the addition of 5 µl SKE buffer (Section 2.15). Samples were then loaded onto a 1.2% agarose gel (Section 2.13) which was typically run for two hours at 100 V.

Chapter Three: BxbI integrase/ΦC31 integrase hybrids

3.1 Introduction

As detailed in Section 1.4.ii, serine integrases are a group of enzymes which recombine short sections of target DNA in a site-specific manner. The way in which serine integrases promote recombination makes them ideal candidates for use in genome engineering, synthetic biology and in gene therapy. For example, serine integrases can be implemented as independently-functioning switches in genetic circuits (Bonnet, et al, 2012) and targeted towards sequences at the beginning and end of a gene of interest for various purposes, such as metabolic gene assembly (Colloms et al, 2013) or the creation of a hybrid genome constructed from the genomes of different organisms. Subsequently, scientists are looking to maximise their control over serine integrases and their recognition sites due to their modular and editable nature.

Perhaps the most straightforward method of expanding the repertoire of serine integrases available for use by scientists is to look for new ones in nature. However, there are several issues with this method: one must identify the *att* sites recognised by the new integrase; likewise, in order to be able to reverse the reaction, the companion protein recombination directionality factor (RDF) must also be identified; the integrase may not recombine its cognate sites with a desirable efficiency and, perhaps most importantly, integrase cross-talk may be occurring – see Section 1.3.ii.

Non-reciprocal cross-talk reduces the potential for such serine integrases to be used in synthetic biology (and other fields in general) where independent, non-cross-talking integrases are preferred for the making of genetic circuits.

Another method of obtaining new integrases that gets around the above listed problems is to engineer novel integrases. One approach is to create hybrids between pre-existing, well established large serine integrases which might recognise and recombine *att* sites which are also hybrids of the natural sites for the original, parental integrases; this is the focus of this chapter.

As established in Section 1.3.ii, ΦC31 integrase and BxbI integrase are both well-characterised serine integrases often used in research; subsequently they were the two integrases chosen to create new, hybrid integrases.

ΦC31 integrase and BxbI integrase share very limited sequence similarity apart from in a few, very specific regions that appear to be conserved across species (see Figure 1.9). This lack of similarity is, in effect, a double-edged sword when it comes to creating hybrid integrases between the two – neither integrase recognises the other's *att* sites, limiting the possibility of reciprocal or non-reciprocal cross-talk occurring in a hybrid of the two, but also increases the probability that the resultant hybrid has issues with protein folding or limited activity.

There are several issues to consider when selecting how to create a hybrid integrase - how much of a domain from one parental integrase should be used at the expense of the other parental integrase; whether parts of regions from one integrase that do not align with the other should be removed or kept in (as is the case with ΦC31 integrase, which is larger than BxbI integrase as well as many other serine integrases – L1 integrase, TP901 integrase, ΦRV1 integrase, ΦBt1 integrase and TG1 integrase for example - and consequently has parts of its sequence that do not correspond to any of BxbI integrase) and, relevant to creating recognition sites, how much from one integrase's *att* site that should be taken and put together with the other. On a technical level, one is also somewhat limited by how many restriction endonuclease sites are present in useful locations within the genomes of both ΦC31 integrase and BxbI integrase.

Part of the basis for expecting a hybrid serine integrase to work is the previous, successful creation of ZFRs, as detailed in Section 1.7. Furthermore, one year into this project, Faruggio and Calos (2014) published a paper detailing the creation of functional serine integrase chimeras which showed activity *in vivo* in *E.coli*. Of the many fusions they constructed, the ones which were successful all contained a ΦC31 integrase catalytic domain and at least part of its αE helix and recombinase domain. However, Farruggio and Calos' hybrids were constructed using integrases which are far more sequentially similar to ΦC31

integrase than BxbI integrase is and so this may be at least part of the reason for why their hybrids are functional – see Section 1.7 for more detail (2014).

3.2 First hybrid integrases and site designs

The serine integrases used to construct all hybrid integrases in this chapter were ΦC31 integrase and BxbI integrase. Consequently, the hybrid *att* sites created were based upon the native *att* sites of both parental proteins.

3.2.i Hybrid integrase design

The rationale behind the domain composition of the hybrid integrases was based in part on the successful ZFRs (detailed in Section 1.7) and in part on the DNA-protein interactions perceived to occur in the LI integrase crystal structures solved by Rutherford et al (2013; detailed in Section 1.5). A closer look at this crystal structure suggests that many direct interactions between LI integrase and its *attP* half-site are centred at the αE helix of the integrase (which connects the catalytic domain to the recombinase domain) and the beginning of the recombinase domain (Figure 3.1).

Taking functional ZFRs and this crystal structure into account, it was decided that the first hybrid integrases designed would consist of a BxbI integrase catalytic domain and αE helix, and a ΦC31 integrase recombinase domain, since this seemed the most likely construct to function. This construct is portrayed pictorially in Figure 3.2.

The point at which the domains from both parent integrases would be connected was decided by aligning the DNA and amino acid sequence of ΦC31 integrase and BxbI integrase within the region of interest, looking for conserved areas that were likely to perform a similar function (see Figure 1.9 for full alignment of both integrases against other serine integrases). Additionally, areas that were vastly different between the two integrases were also of interest. Consequently, two different linkers were designed to connect the two parts of the hybrid integrase (described further in Section 3.3.i).

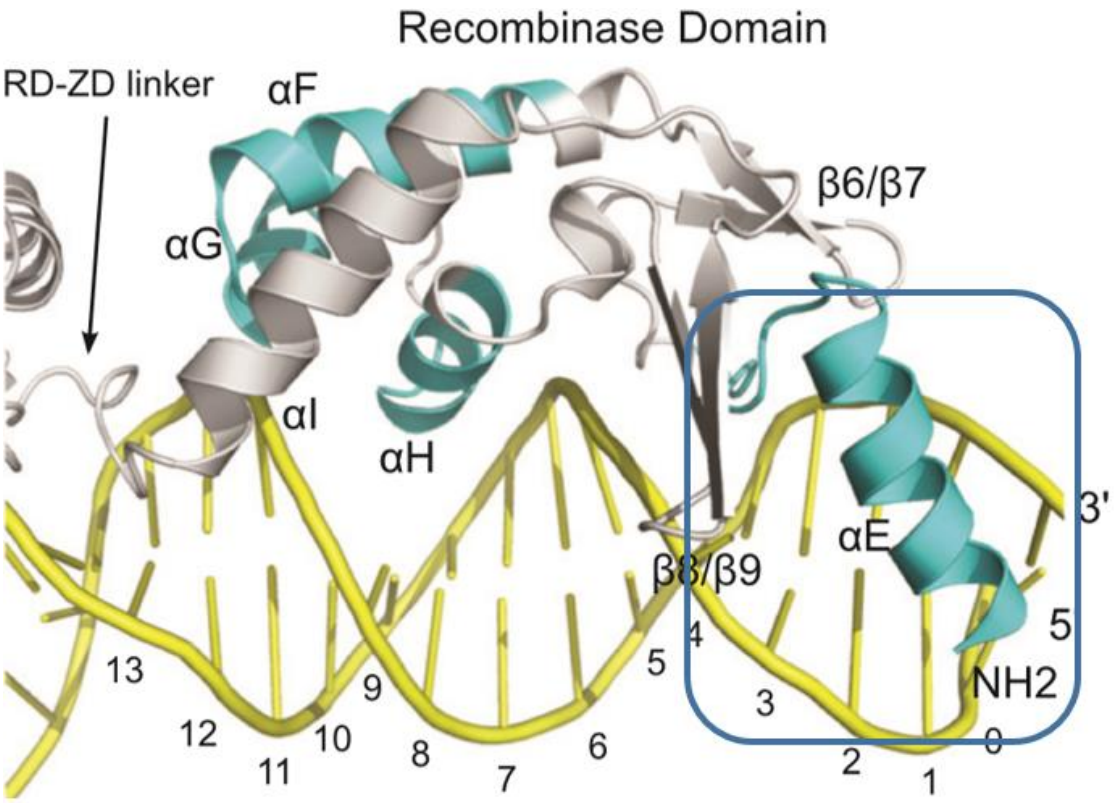

Figure 3.1: A closer look at A118 integrase-*attP* half site interactions.

The direct interactions of interest that A118 integrase has with its *attP* half-site involve the αE helix of the integrase as well as the beginning of its recombinase domain and take place across the first five nucleotides of the site (outlined in the blue box). If this were extrapolated to include a full- rather than half-site, then the direct interactions A118 integrase has with its *attP* recognition site fall across the central ten nucleotides of the site (modified from Van Duyne and Rutherford, 2013).

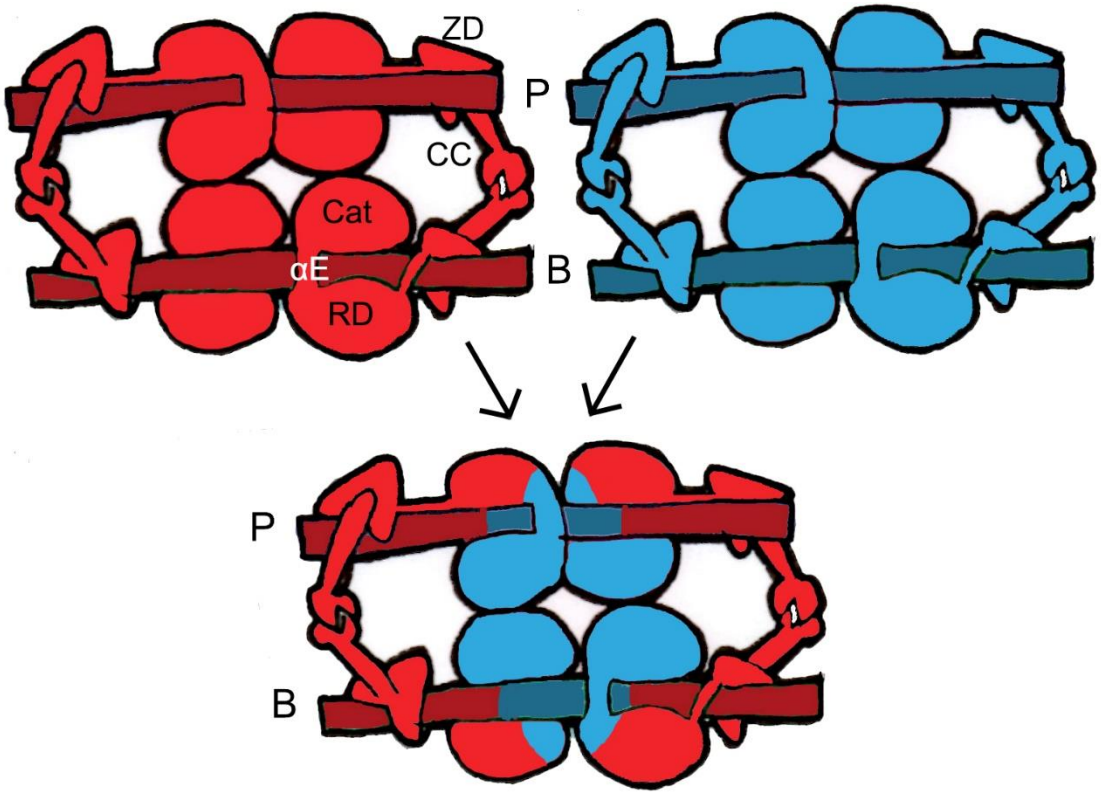

Figure 3.2: Design constructs for hybrid integrases.

(A) Design construct of hybrid integrases based on Rutherford et al (2013) integrase-*att* site interaction model. Blue denotes a BxbI integrase origin and red denotes a ΦC31 integrase origin. Cat: catalytic domain; RD: recombinase domain; αE: αE helix; ZD: zinc ribbon domain; CC: coiled coil domain; P: *attP* site; B: *attB* site.

3.2.ii Hybrid site design

The various regions of conserved nucleotides found between *attP* and *attB* sites for both ΦC31 integrase and BxbI integrase were taken into account when deciding upon the design of the hybrid sites (Figure 3.3; full *att* site sequences can be found in Figure 3.4B). It was decided that two sets of hybrid *att* sites would be created – one set with the central 10 bp and one set with the central 12 bp coming from the BxbI integrase *attP*/*B* sites respectively. The rest of the site's basepairs would originate from the ΦC31 integrase *attP*/*B*. Altering either 10 or 12 bp in the middle of the site is based on the direct DNA-protein interactions observed for the LI integrase crystal structure (Figure 3.1). Since this structure leads us to suggest that the central 10 bp (5 bp from the centre of the *attP* site on either side) are subject to direct DNA-protein interactions by LI integrase, whether the sixth basepair from the centre of the site would also be critical for identification of the site by its respective hybrid integrase or either parent integrase is of interest. This is demonstrated pictorially in Figure 3.4A, with the resultant full hybrid *att* site sequences designed shown in Figure 3.4C.

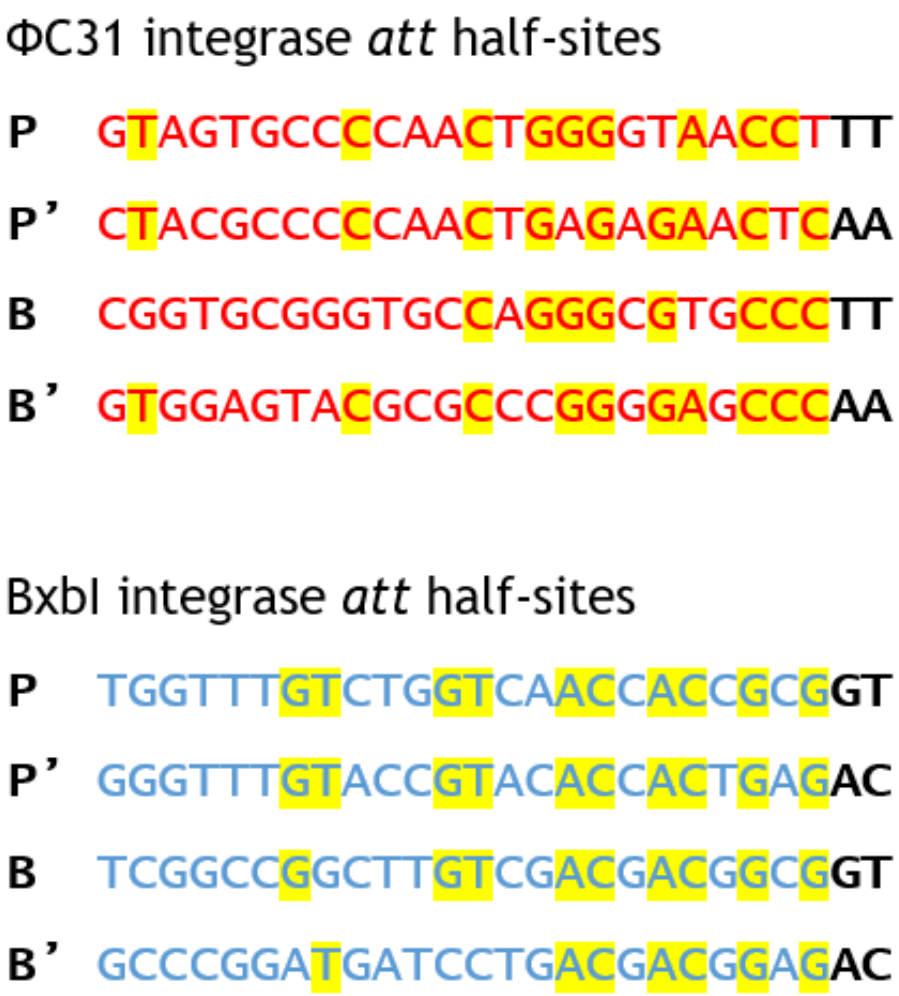

Figure 3.3: ΦC31 integrase and BxbI integrase *att* half-sites aligned, highlighting conserved regions.

Conserved nucleotides in at least three of four positions between aligned *att* half-sites are highlighted in yellow. The crossover dinucleotide for each site is shown in black, boldface type (modified from Van Duyne and Rutherford, 2013).

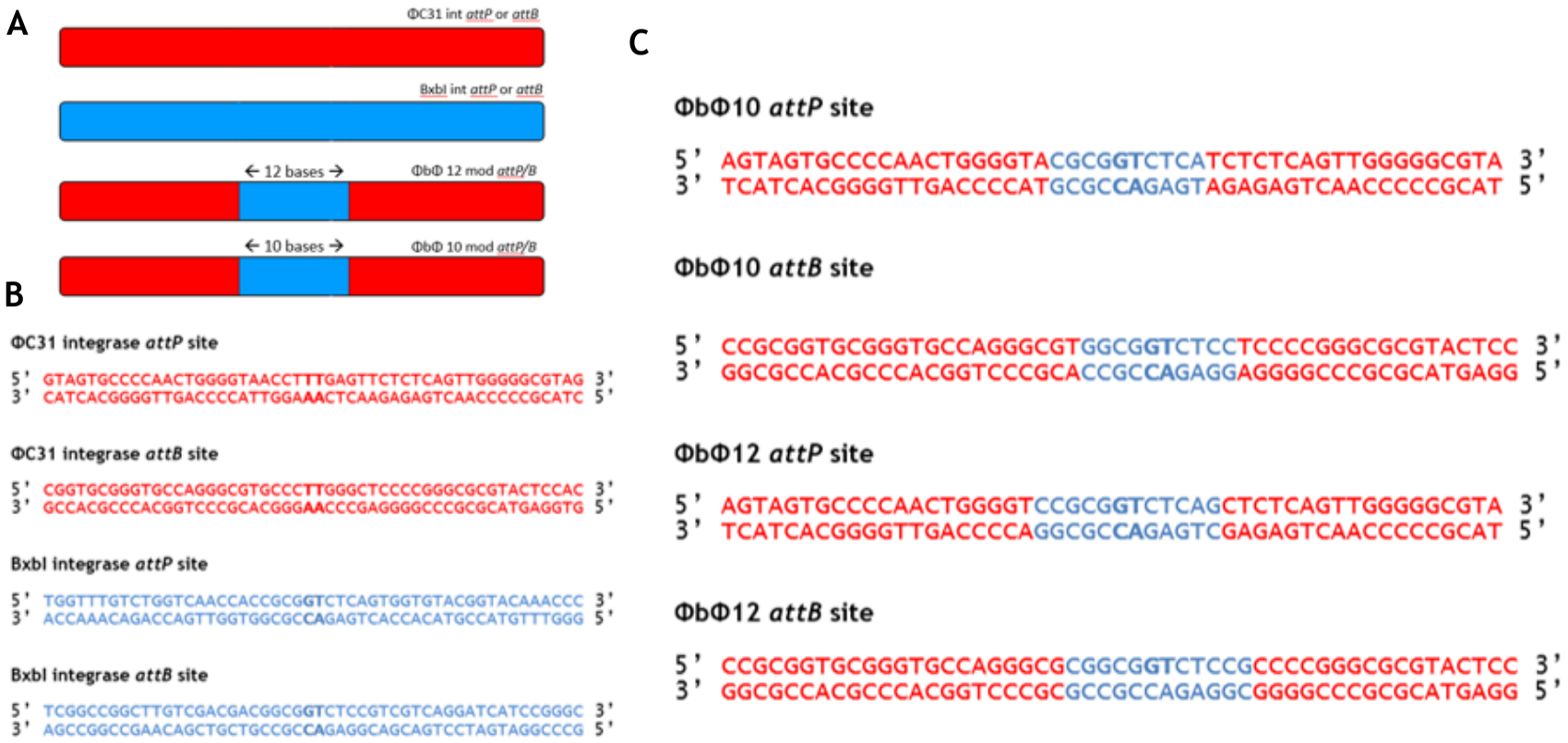

Figure 3.4: Recognition *att* sites for parental and hybrid integrase.

(A) Pictorial representation of hybrid *att* site design; (B) exact *attP* and *attB* recognition site sequences for ΦC31 integrase and BxbI integrase and (C) exact *attP* and *attB* recognition site sequences for the hybrid integrases BΦ1 and BΦ2. In all instances, red denotes a ΦC31 integrase origin and blue denotes a BxbI integrase origin. The crossover dinucleotide in each site is in bold typeface. ΦbΦ10 *attP* indicates a recognition site in which the central ten nucleotides come from the BxbI integrase *attP* site, whilst the nucleotides flanking this region come from the ΦC31 integrase *attP* site. This naming scheme remains consistent throughout all hybrid sites.

3.3 Construction of hybrid integrases and sites

3.3.i Hybrid integrase construction

In order to allow greater flexibility in choosing where exactly to connect BxbI integrase domains to ΦC31 integrase domains, restriction enzyme sites were introduced to the coding DNA sequences of both parent integrases across their catalytic domains, αE helix and recombinase domains which do not alter the translated amino acid sequence. These altered regions of both integrases (spanning 824 bp/274 residues and 536 bp/178 residues for ΦC31 integrase and BxbI integrase respectively) were synthesised by GeneArt within plasmids with a pMA-T vector backbone (see Table 2.5 for detailed plasmid information).

As mentioned in Section 3.2.i, two oligonucleotide linkers were designed to connect the gene fragment of BxbI integrase to the gene fragment of ΦC31 integrase, from the restriction enzyme site AccI located at the beginning of the BxbI integrase recombinase domain to the site RsrII located further into the recombinase domain of ΦC31 integrase (exact locations shown and the regions of interest from both integrase shown aligned in Figure 3.5**)**. One of these linkers was designed to follow the amino acid sequence of BxbI integrase from AccI to where its sequence would end relative to the RsrII site located in ΦC31 integrase; conversely, the second linker was designed to follow the amino acid sequence of ΦC31 integrase from where its sequence would start relative to the AccI site in BxbI integrase to the site RsrII (Figure 3.5). As a result, the second linker is longer than the first as ΦC31 integrase is a larger protein than BxbI integrase, and consequently when aligned the relative region from AccI to RsrII in ΦC31 integrase is one such area with no sequence similarity to BxbI integrase. All linkers were designed as oligonucleotides and were synthesised by Eurofins Genomics. Because the second linker was so long, it was synthesised as two separate pairs of oligonucleotides (see list of oligonucleotides in Table 2.4).

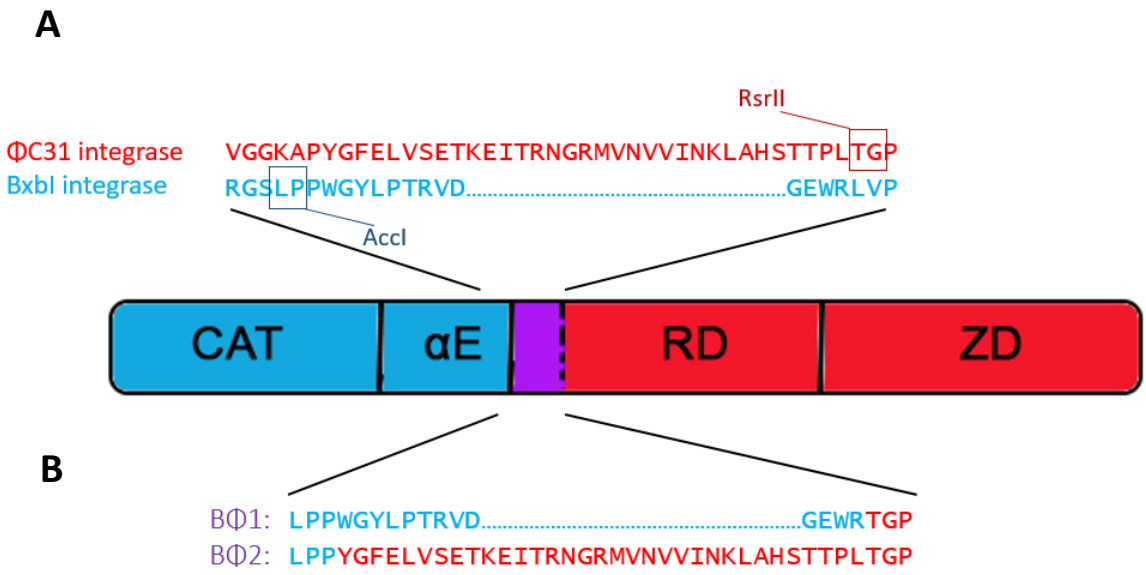

Figure 3.5: (A) Regions of interest in ΦC31 integrase and BxbI integrase amino acid sequences aligned and (B) amino acid sequence for both hybrid integrase linkers.

In all instances, ΦC31 integrase is denoted in red, BxbI integrase in blue and the hybrid integrase linkers in purple. In (A), the amino acid sequence of ΦC31 integrase from the end of its αE helix to the beginning of its recombinase domain is aligned against the equivalent region in BxbI integrase, with the location of the restriction enzymes AccI and RsrII indicated. Ellipses are used to denote no equivalent amino acids in the BxbI integrase sequence to align to ΦC31 integrase. In (B), the amino acid sequences for the two hybrid integrase linkers BΦ1 and BΦ2 are shown. A pictorial diagram of the hybrid integrase construction is included to represent where these linkers connect BxbI integrase domains to those of ΦC31 integrase relative to the aligned amino acid sequences for both parental proteins shown in (A). Cat: catalytic domain; αE: αE helix; RD: recombinase domain; ZD: zinc ribbon domain.

The GeneArt constructed plasmid BxbI_altered was digested with the restriction enzymes DrdI and AccI; likewise the GeneArt constructed plasmid PHIC31_altered was digested with the restriction enzymes DrdI and RsrII. The resultant, smaller fragment from PHIC31_altered was ligated with larger fragment from BxbI_altered and either the BΦ1 or BΦ2 linker oligonucleotides (Table 2.4). This created the *in vitro* precursor plasmids pHM9x (BΦ1) and pHM10x (BΦ2; Figure 3.6A).

Once pHM9x and pHM10x were created, both plasmids were cut with the restriction enzymes NdeI and BamHI and fragment swapped into the T7 inducible, high copy number expression plasmid pFEM14 (Table 2.5). pFEM14 contains the full length reading frame of ΦC31 integrase and so the deletion of the smaller NdeI-BamHI fragment from this plasmid and subsequent replacement with the equivalent fragment from pHM9x or pHM10x successfully completes the full reading frame of both hybrids in a plasmid backbone suitable for protein overexpression and purification (Figure 3.6B). These plasmids were named pHM9 and pHM10, respectively.

For the creation of *in vivo* expression plasmids for both hybrid integrases, pHM9 and pHM10 were cut with the restriction enzymes NdeI and KpnI and fragment swapped into the T7 inducible, low copy number expression plasmid pFO2 (Table 2.1), also cut with the same enzymes. This resulted in the *in vivo* expression plasmids pHM9* and pHM10* (Figure 3.7).

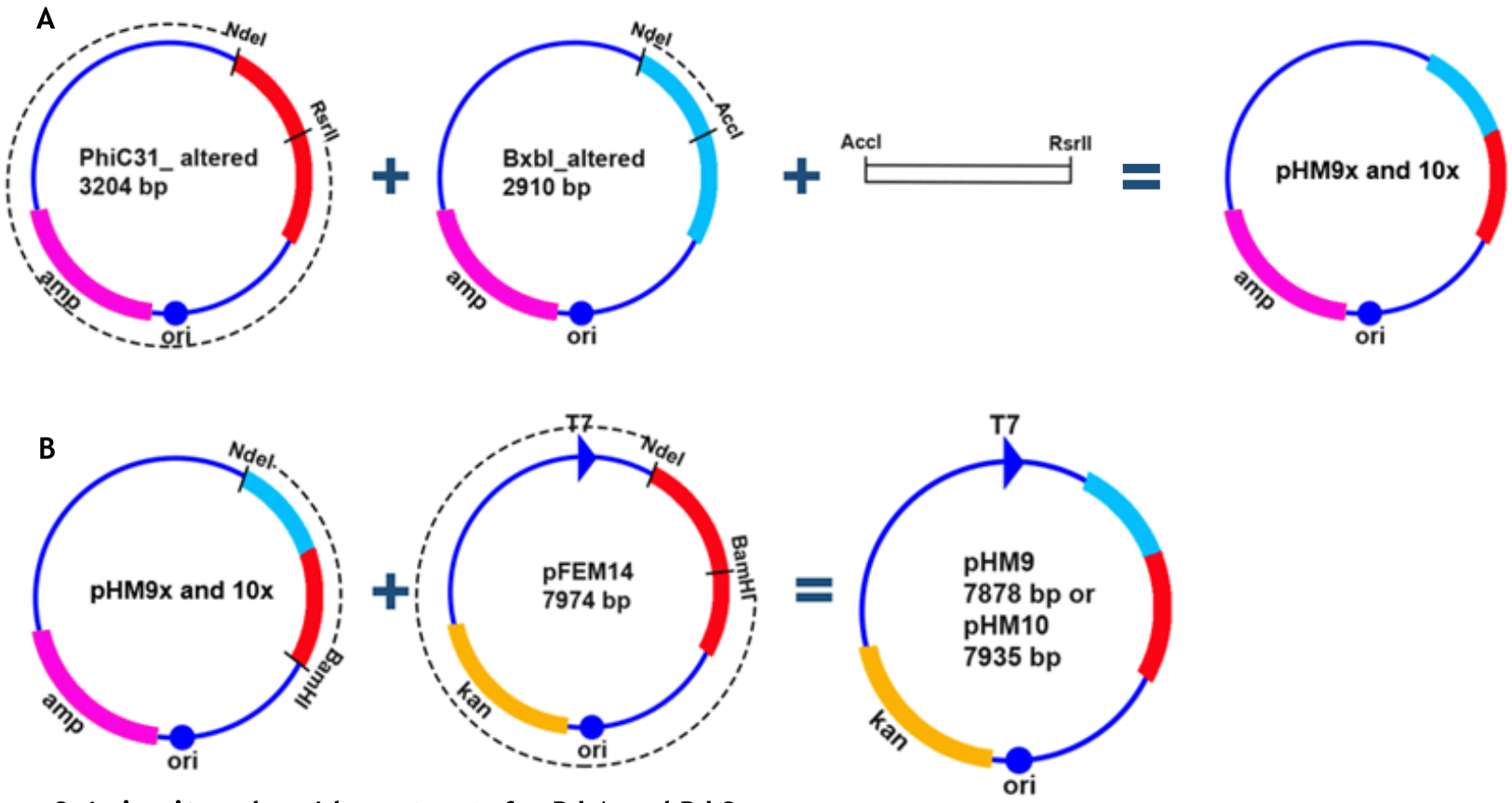

Figure 3.6: *in vitro* plasmid constructs for BΦ1 and BΦ2.

(A) The smaller fragment from an NdeI and AccI digest performed on the plasmid Bxbl_altered, the larger fragment from an NdeI and RsrII digest performed on the plasmid PhiC31_altered and an oligonucleotide linker with AccI and RsrII ends were ligated to create the *in vitro* precursor expression plasmids pHM9x and pHM10x. (B) The smaller fragment from an NdeI and BamHI digest performed on pHM9x or pHM10x and the larger fragment from the same digest performed on pFEM14 were ligated together to create the *in vitro* expression plasmids pHM9 (BΦ1) and pHM10 (BΦ2), respectively. In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

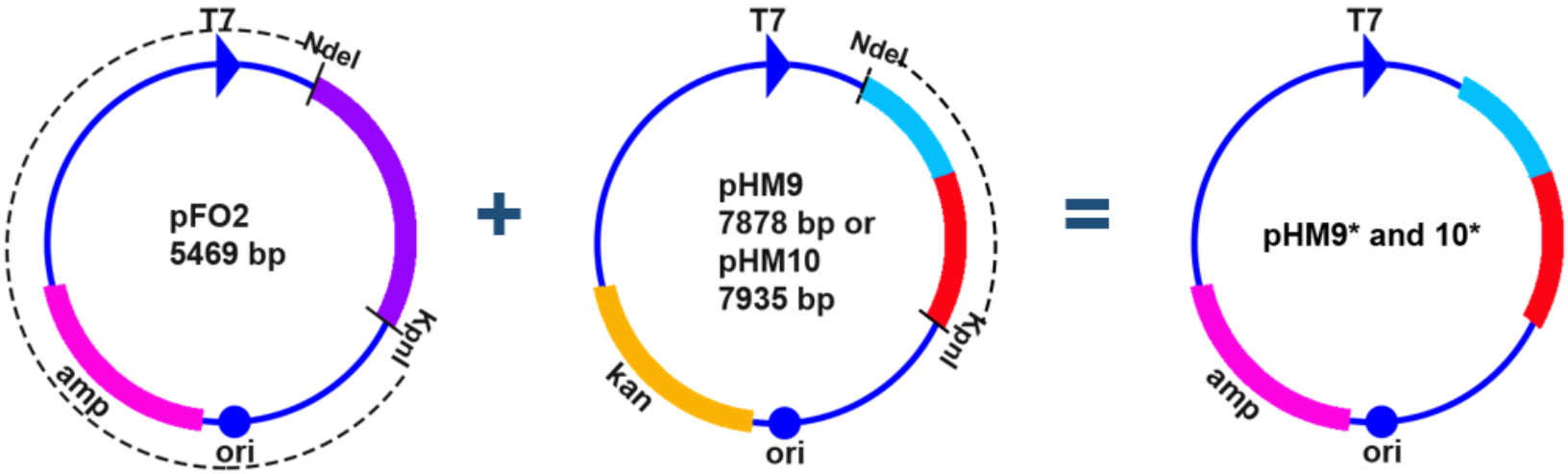

Figure 3.7: Construction of *in vivo* expression hybrid integrase plasmids.

The larger fragment from an NdeI and KpnI digest on pFO2 and the smaller fragment from an NdeI to KpnI digest on either pHM9 or pHM10 were ligated together to create the *in vivo* expression plasmids pHM9* and pHM10*. In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

3.3.ii Hybrid site construction

The oligonucleotide sequences for the hybrid recognition sites ΦbΦ12 *attP* and *attB* and ΦbΦ10 *attP* and *attB* can be seen in Table 2.4. All *in vitro* substrate plasmids were created via two rounds of cloning – the plasmid pFM12 was cut with the restriction enzymes SacI and EcoRI and the substrate oligonucleotide in question was then ligated in to create pHM1 (ΦbΦ10 *attP*), pHM2 (ΦbΦ10 *attB*), pHLM3 (ΦbΦ12 *attP*) and pHLM4 (ΦbΦ12 *attB*) (Figure 3.8A**)**. After these one site substrate plasmids were created, two site substrate plasmids were created by digesting pHM1 and pHM3 with AlwNI and BglII, pHM2 and pHM4 with AlwNI and BamHI and pUC7IK with BamHI only. The required fragments were ligated together to create the two site plasmids pHM12 (ΦbΦ10 *attP* and *attB*) and pHM13 (ΦbΦ12 *attP* and *attB*) (Figure 3.8B). These were subsequently digested with the restriction enzyme NruI and run on a 1.2% agarose gel to ensure that the orientation of the *attP* and *attB* sites would result in the most informative NruI digest – this is because the pUC71K fragment can be ligated in to the new plasmid either way around due to its identical BamHI ends (Figure 3.9).

For *in vivo* substrate plasmids, two rounds of cloning were again required. The plasmid pMS183Δ was digested separately with either AlwNI and BsrGI or AlwNI and NheI to insert a new *attP* site, and either AlwNI and XbaI or AlwNI and Asp718 to insert a new *attB* sites, whilst the single site *in vitro* plasmids pHM1, pHM2, pHM3 and pHM4 were digested with Asp718 and XbaI. The smaller fragment from the *in vitro* substrate plasmids was ligated together with either set of fragments from pMS183Δ depending on whether the site in question was *attP* or *B* (Figure 3.10A). The resultant plasmids were then digested with AlwNI and NdeI and fragment swapped together to create two site *in vivo* substrate plasmids named pHM7 (ΦbΦ10 *attP* and *attB*) and pHM8 (ΦbΦ12 *attP* and *attB*) (Figure 3.10B).

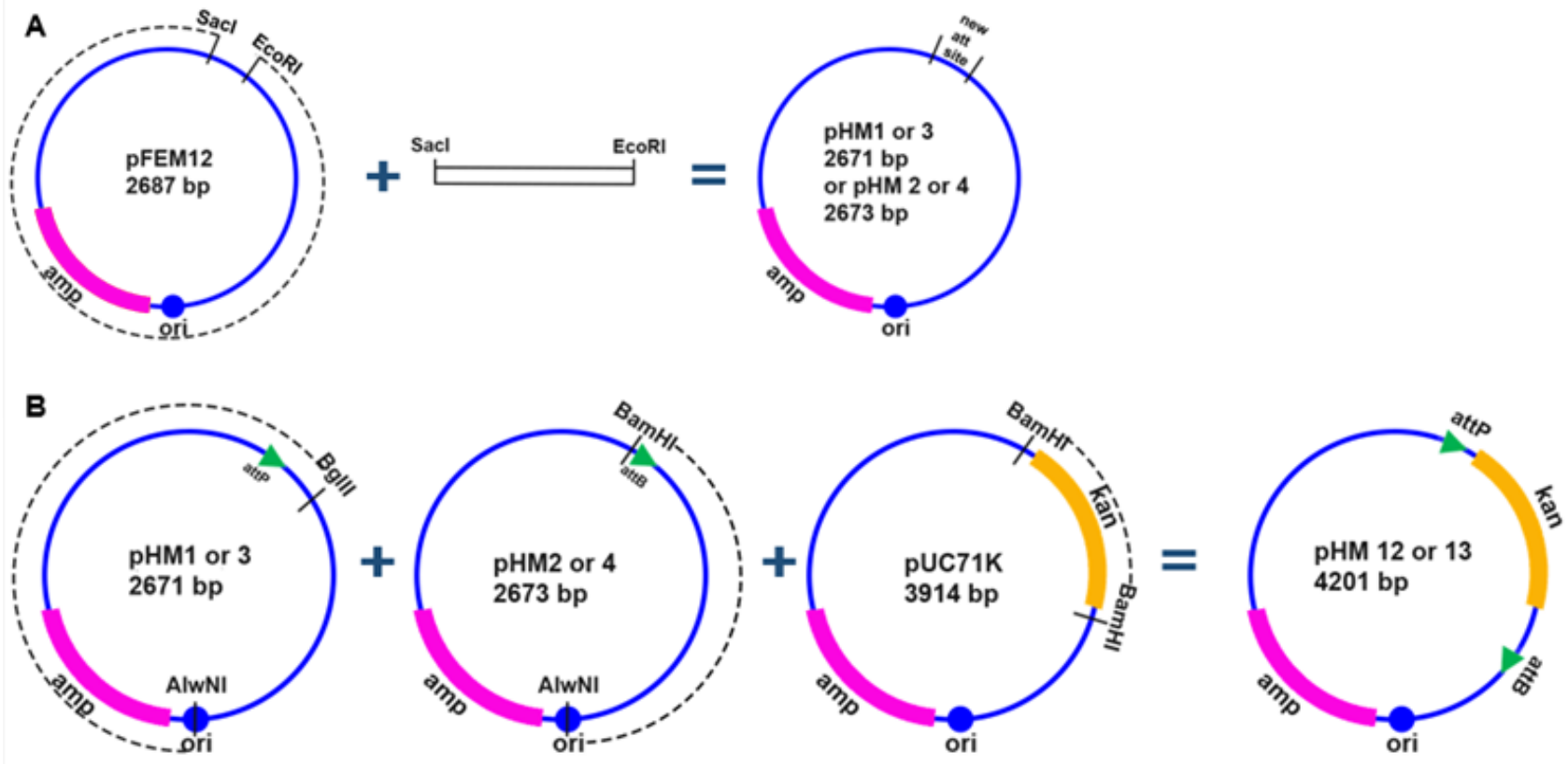

Figure 3.8: Construction of *in vitro* precursor and two-site substrate plasmids.

(A) The plasmid pFM12 was cut with SacI and EcoRI and an oligonucleotide of a hybrid *attP* or *B* site with the same restriction site ends was ligated in. The resultant precursor substrate plasmid was named either pHM1, pHM2, pHM3 or pHM4 depending on the *att* site within it (see main text). (B) pUC71K was digested with BamHI and the smaller fragment was ligated together with the larger fragment of an AlwNI and BglII digested *attP*-containing plasmid (e.g. pHLM1) and the smaller fragment from an AlwNI and BamHI digested *attB*-containing plasmid (e.g. pHLM2). The resultant *in vitro* two-site substrate plasmids were named pHM12 or pHM13. In all instances, the fragment taken to create the final plasmids is denoted with a dashed line.

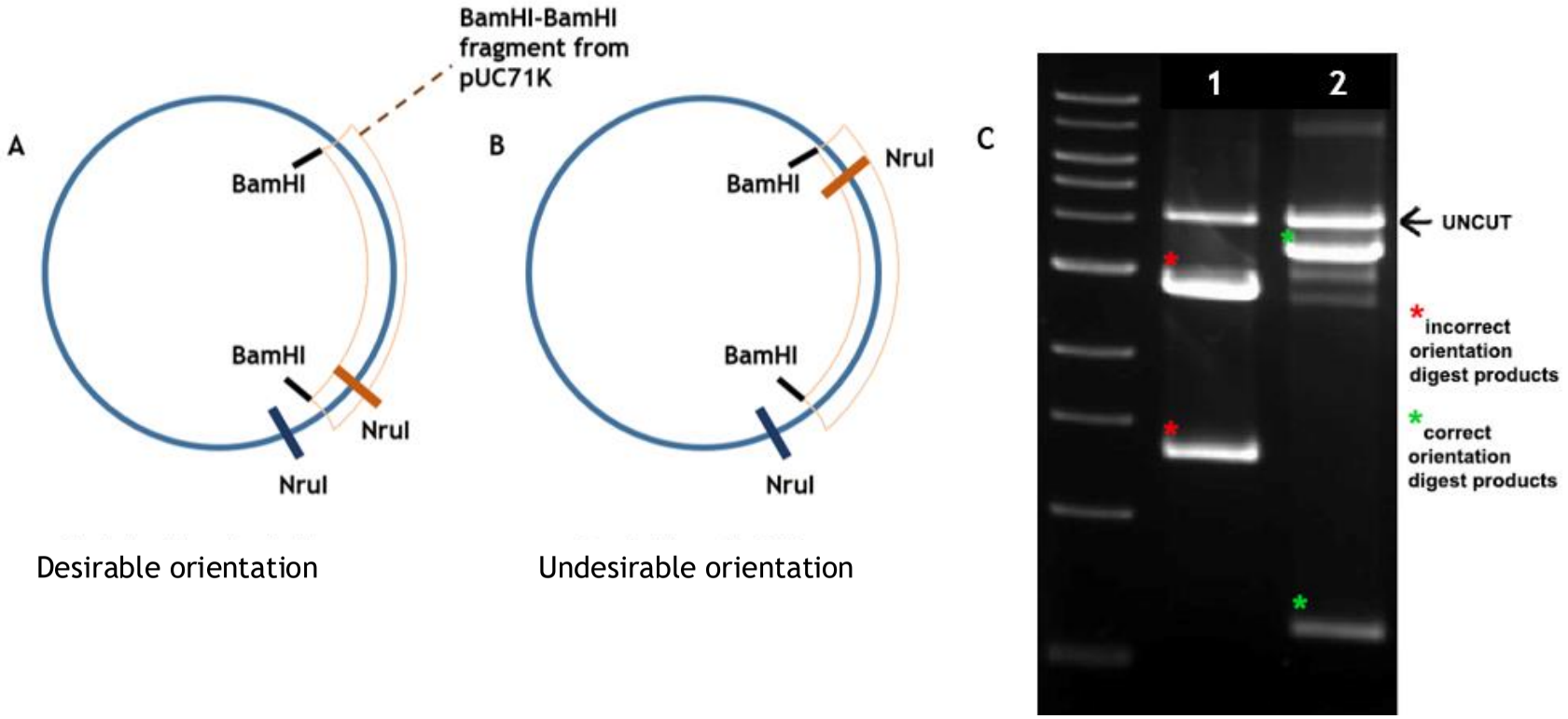

Figure 3.9: Possible orientations for *in vitro* two site substrate plasmids once cut with NruI.

The double BamHI-ended pUC71K fragment used in the construction of such plasmids can ligate in either way, resulting in (A) two-site substrate plasmids with a desirable orientation for NruI digests and (B) two-site substrate plasmids with an undesirable orientation for NruI digests. (C) NruI digest products on two different two-site substrate plasmids: (1) undesirable orientation, where the digest products are approximately 1.3kb and 2.8kb in size and (2) desirable orientation of two site plasmids, where the digest products are approximately 550 bp and 3.4kb in size. 1kb ladder is used for size comparison.

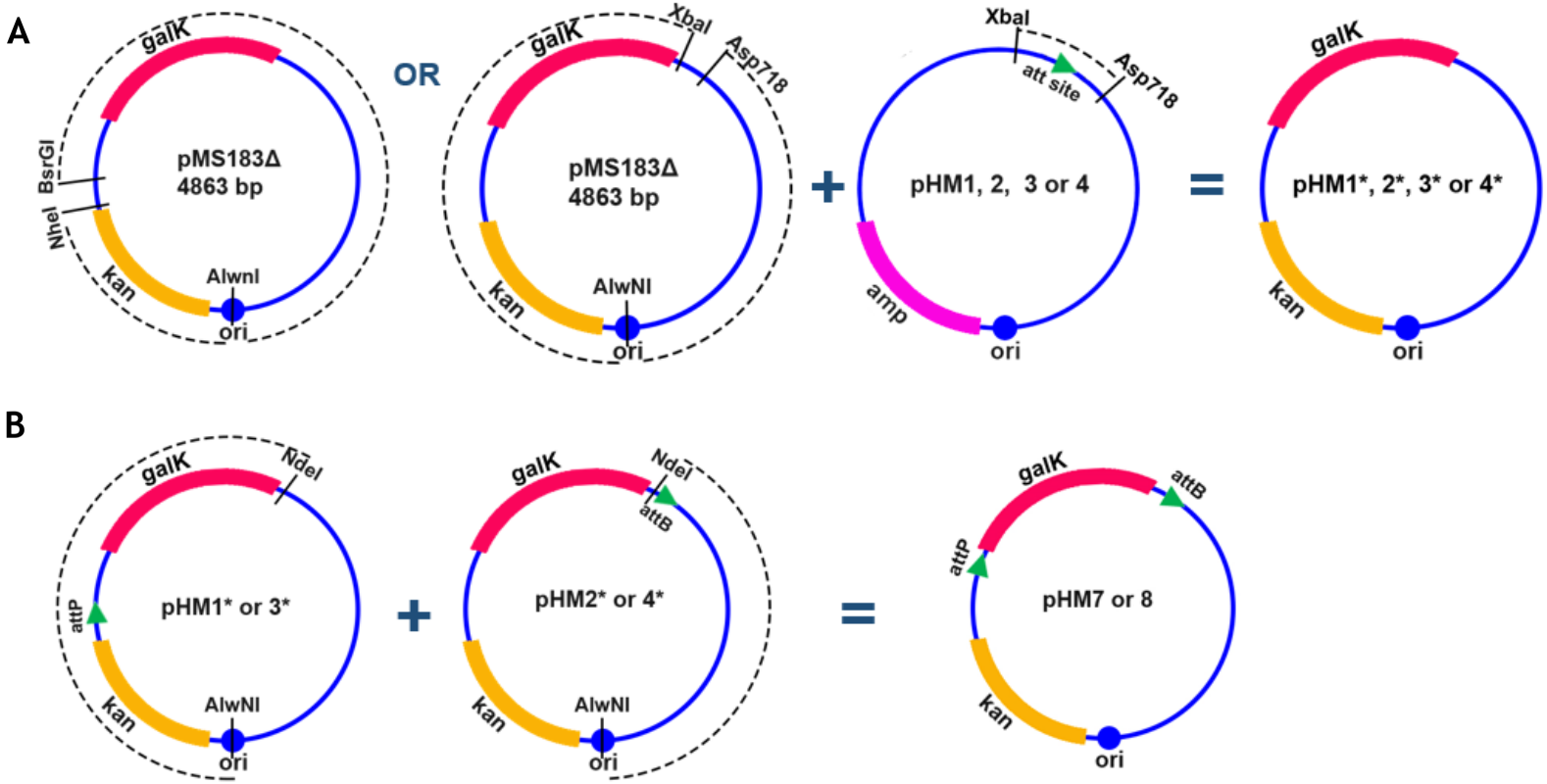

Figure 3.10: Construction of *in vivo* two-site substrate plasmids

(A) The two larger fragments indicated with a dashed line in the left hand pMS183Δ plasmid along with the smaller fragment from an Asp718 and XbaI digested one-site *att* site *in vitro* plasmid (e.g. pHM1) were ligated together to create the *in vivo* plasmids pHLM1* and pHLM3*. Similarly, the two larger fragments indicated with a dashed line in the right hand pMS183Δ plasmid along with the smaller fragment from an Asp718 and XbaI digested one-site *att* site *in vitro* plasmid were ligated together to create the *in vivo* plasmids pHLM2* and pHLM3*. (B) The larger fragment from an AlwNI and NdeI digest of *attP*-containing plasmids pHM1* or pHM3* and the smaller fragment taken from the same digest for *attB*-containing plasmids pHM2* or pHM4* were ligated together to form the two site *in vivo* plasmids pHM7 and pHM8, respectively. In all instances, the fragments used to create the final plasmids are denoted with a dashed line.

3.4 Hybrid activity *in vivo*

Both BΦ1 and BΦ2 (as well as ΦC31 integrase and BxbI integrase) were tested for their recombination activity *in vivo* using a MacConkey assay (as described in Section 2.21) against the two-site hybrid substrate plasmids pHM7 and pHM8.

The MacConkey assay (Blake, 1993) can be used to screen for recombination activity *in vivo* (explained fully in Figure 3.11). Resolution of a substrate plasmid containing two recombination sites by its respective integrase or resolvase (e.g. ΦC31 integrase or Tn3 resolvase) produces two circular DNA molecules - one containing the origin of replication and the other containing the *galK* gene. The latter cannot be maintained in the cell as it has no origin of replication and so is lost in subsequent cell divisions. MacConkey agar has no carbon source other than the galactose added to it and contains the pH indicator 2-methyl-3-amino-6dimethylaminophenazine, which is yellow at a pH greater than 8.0 and red at a pH less than 6.8. Successful resolution of the substrate plasmid forces the *E.coli* to metabolise amino acids in the agar as they are unable to metabolise galactose, resulting in an increase in pH and production of white colonies. Unresolved substrate containing the *galK* gene allows the *E.coli* to metabolise galactose, lowering the pH and producing red colonies.

Their activity on the native ΦC31 integrase *att* sites in plasmid pGD001 and the native BxbI integrase *att* sites in plasmid pFM124 was also tested. However, all colonies other than positive controls were red in colour, indicating no recombination (data not shown). In order to see if any resolution was taking place at a low level, each MacConkey plate was scraped of all colonies and grown for sixteen hours in L-broth at 37 °C. These samples were then miniprepped (Section 2.10 and 2.21) and run on a 1.2% agarose gel to test if any observable recombination had taken place. They were also transformed into electro-competent DS941 and subsequently plated out on another round of MacConkey plates (see Section 2.21 for full protocol). Both the gel and new MacConkey plates demonstrated that ΦC31 integrase does exhibit some activity

against pHM7 (ΦbΦ10) but not against pHM13 (ΦbΦ12) (Figure 3.12 and Figure 3.13). Colony counts for each of these plates can be seen in Table 3.1.

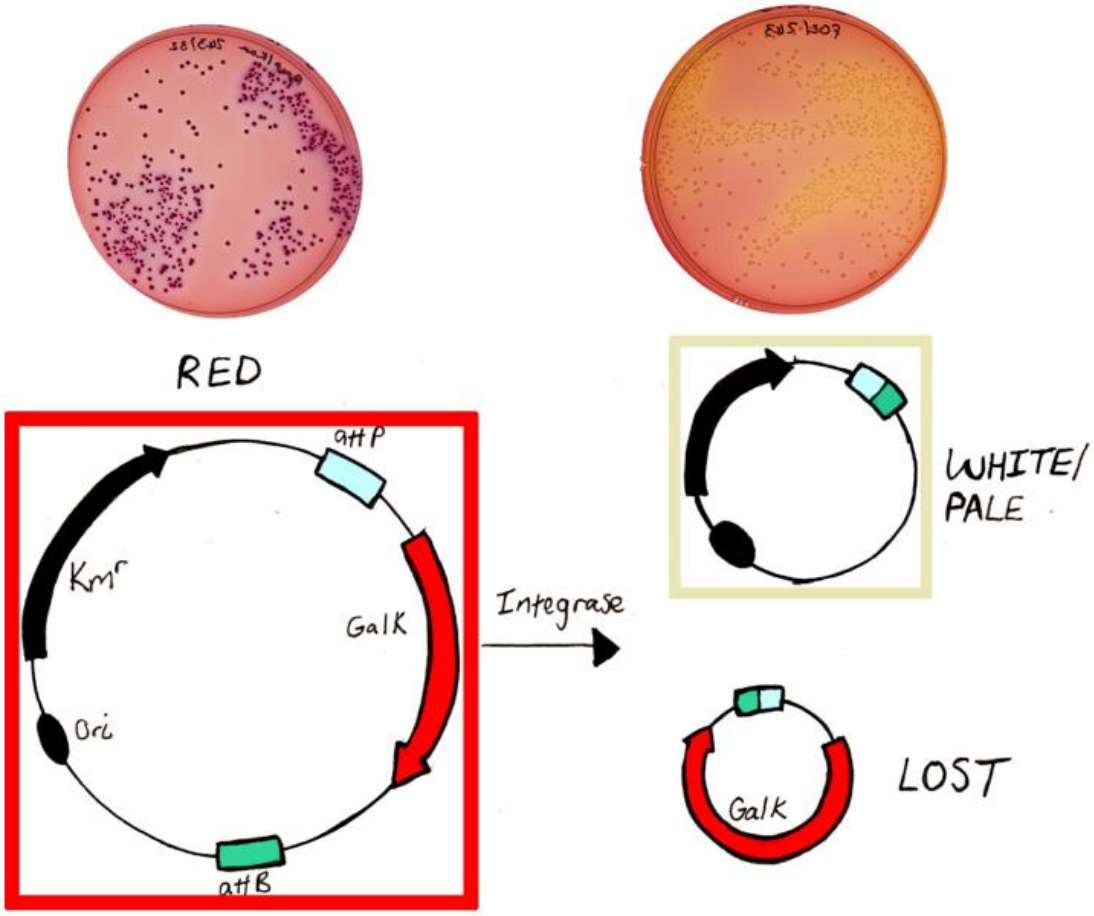

Figure 3.11: MacConkey assay.

GalK-expressing *att* site plasmids are transformed into DS941, a *recF*$^{-}$, *galK*$^{-}$, F^{-} strain. Colonies of DS941 containing *galK*-expressing plasmids are red on MacConkey agar plates which contain galactose, as the presence of *galK* allows DS941 to use galactose as a food source. Resolution of these plasmids by integrase separates the *galK* gene from the origin of replication. Plasmids which do not contain the origin of replication are subsequently lost. The resultant colonies are *galK*$^{-}$ and pale yellow ('white') on MacConkey plates, as they cannot use galactose as a food source. Colonies containing unresolved plasmids remain red on MacConkey plates.

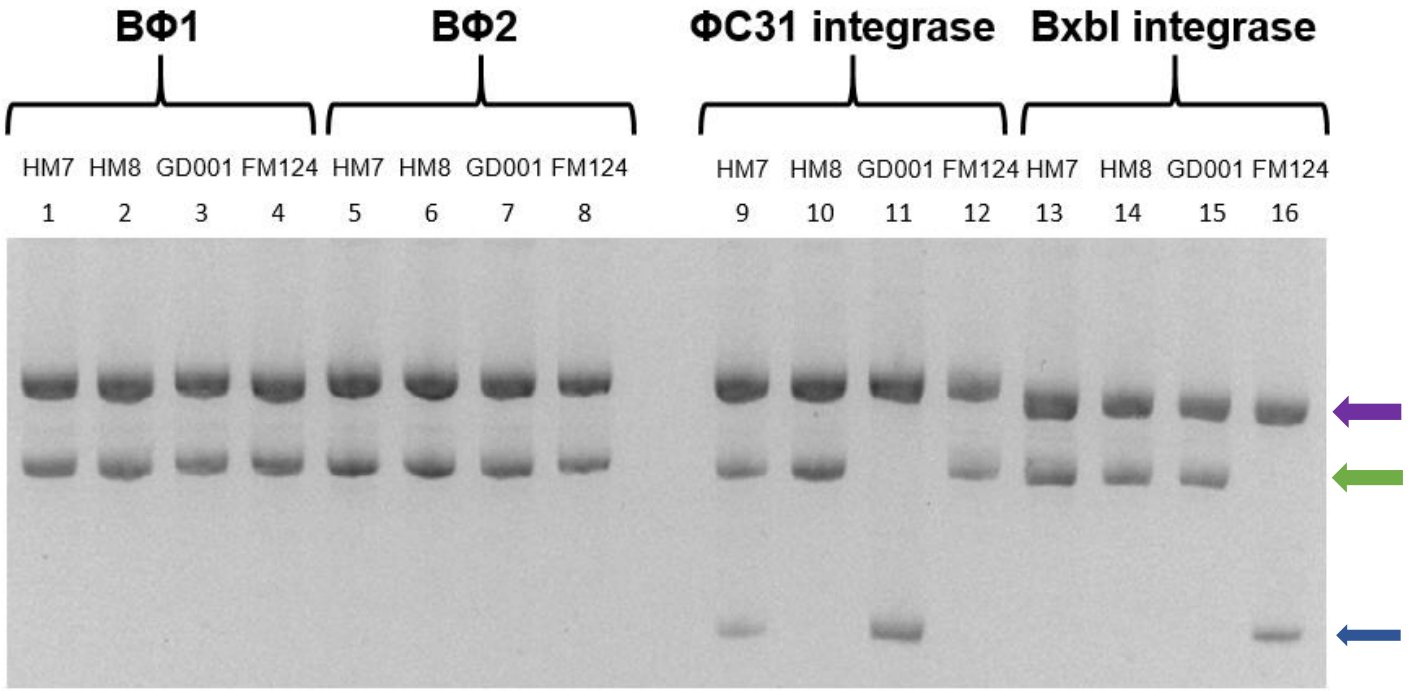

Figure 3.12: MacConkey assay overnight cultures recombination results.

Lanes 11 and 16 are positive resolution controls. Resolved substrate plasmids observed in lanes 9, 11 and 16 are denoted with a blue arrow; unresolved substrate plasmids are denoted with a green arrow and recombinase expression plasmids are denoted with a purple arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

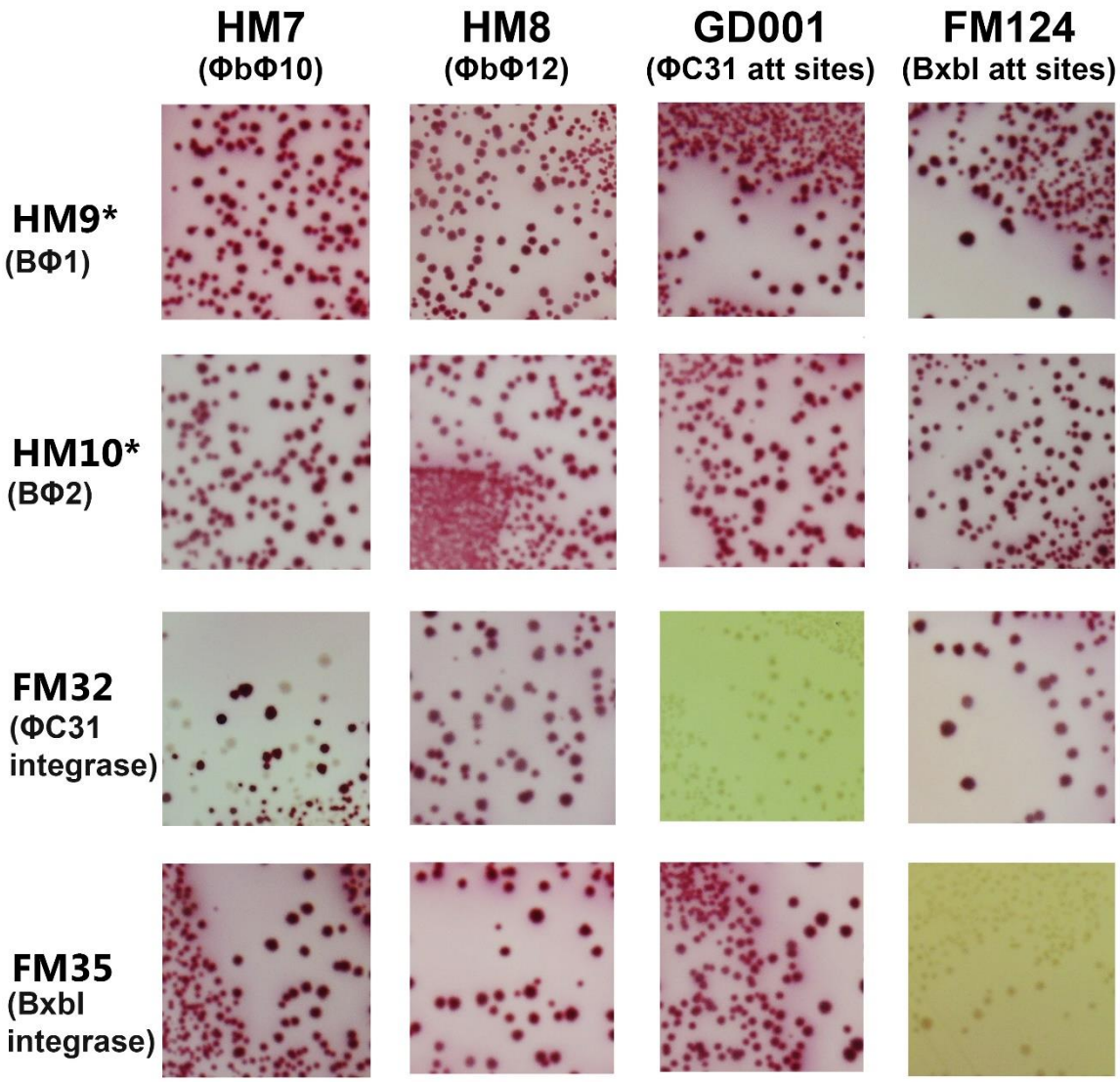

Figure 3.13: Second round MacConkey plate results for BΦ1 and BΦ2.

Red/pink colonies denote no recombination whilst white/cream colonies denote recombination. A plate containing a combination of both red and white colonies (such as pFEM32 [ΦC31 integrase] with pHM7 [ΦbΦ10 *attP/B*]) denotes incomplete resolution of the substrate-containing plasmids.

Name	Site name	No. of red colonies	No. of white colonies	% resolved
HM7	HM12	2449	0	0
	HM13	2301	0	0
	GD001	2133	0	0
	FM124	1780	0	0
HM8	HM12	2550	0	0
	HM13	2178	0	0
	GD001	2003	0	0
	FM124	1998	0	0
FEM32	HM12	1965	640	24.57
	HM13	2300	0	0
	GD001	0	2145	100
	FM124	2263	0	0
FEM35	HM12	2577	0	0
	HM13	2305	0	0
	GD001	1847	0	0
	FM124	0	1805	100

Table 3.1: Colony counts for second round MacConkey plates for BΦ1 and BΦ2.

24.57% resolution of pHM12 (ΦbΦ10 *attP*/*B*) is a result of allowing pFEM32 (ΦC31 integrase) more time in which to allow recombination to occur.

3.5 Hybrid site recognition *in vitro*

Following on from the *in vivo* results, whether either parent integrase would be able to recognise and recombine either of the hybrid two-site plasmids pHM12 (ΦbΦ10 *attP* and *attB*) or pHM13 (ΦbΦ12 *attP* and *attB*) as pure protein *in vitro* was tested using a recombination assay (described in Section 2.24.i). Both pFM16 (ΦC31 *attP* and *attB*) and pFM97 (BxbI *attP* and *attB*) were used as positive controls for ΦC31 integrase and BxbI integrase, respectively. Unfortunately, neither ΦC31 integrase nor BxbI integrase recombined pHM12 or pHM13 (Figure 3.14). Consequently, the more sensitive ethylene glycol cleavage assay (described in Section 2.24.ii) was performed to determine if either parent integrase could bind and cleave the hybrid sites. The plasmid pFM16 was used as a positive control. In sixteen hour reactions, ΦC31 integrase cleaved both sets of hybrid sites (Figure 3.15). In a time course cleavage assay, it took two hours before cleavage by ΦC31 integrase could be observed on pHM12 and four hours for pHM13.

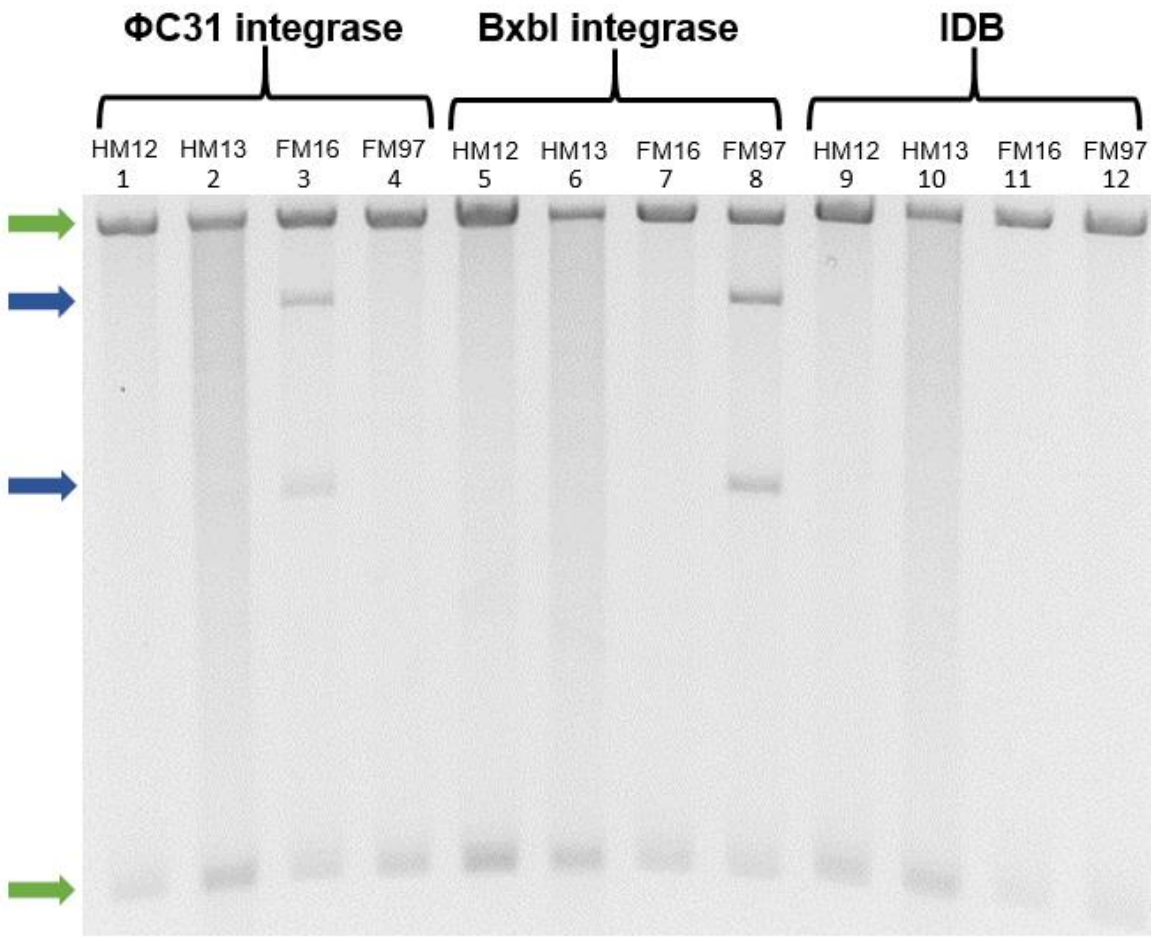

Figure 3.14: Recombination assay results for ΦC31 integrase and Bxbl integrase on hybrid substrate plasmids pHM12 and pHM13.

IDB (integrase dilution buffer) is used in place of protein as a negative control in lanes 9, 10, 11 and 12. Positive controls are in lanes 3 and 8. Recombination products identified in lanes 3 and 8 are marked with blue arrows. Unrecombined products have been identified with green arrows. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

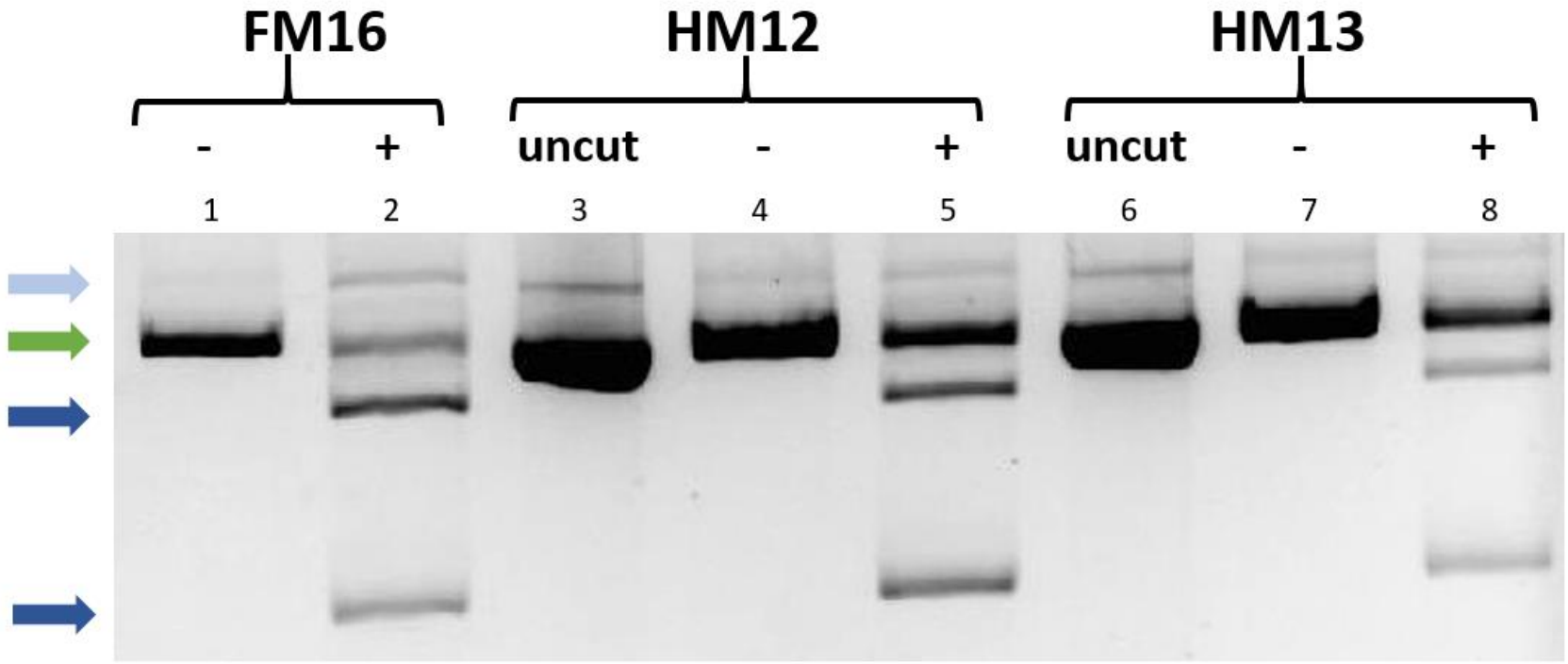

Figure 3.15: Ethylene glycol cleavage assay results for ΦC31 integrase on hybrid substrate plasmids pHM12 and pHM13.

IDB used as a negative control in lanes 1, 4 and 7. Positive control is in lane 2. Double-site cleavage products identified in lanes 2, 5 and 8 are marked with dark blue arrows. Single-site cleavage products identified with a pale blue arrow. A green arrow identifies non-cleaved substrate plasmid. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

3.6 Second round hybrid integrase and site designs

3.6.i Hybrid integrase design

One year into this project, Farruggio and Calos (2014) published a paper detailing the creation of functional serine chimeras, most of which were constructed using a ΦC31 integrase catalytic domain (see Section 1.7). Taking this into account, a hybrid integrase was subsequently constructed as a "reverse integrase" of BΦ1 and BΦ2 i.e. a ΦC31 integrase catalytic domain and a BxbI integrase recombinase domain, named ΦB1 (Figure 3.16). A short linker was created to join the two integrases together (Figure 3.16; Table 2.4).

3.6.ii Hybrid site design

One set of hybrid sites was created for ΦB1 following the same *att* site design as described in Section 3.2.ii but in reverse –the central 10 bp coming from the ΦC31 integrase *attP* and *attB* sites and the outermost basepairs coming from BxbI integrase *attP* and *attB* sites. These sites can be seen in Figure 3.17.

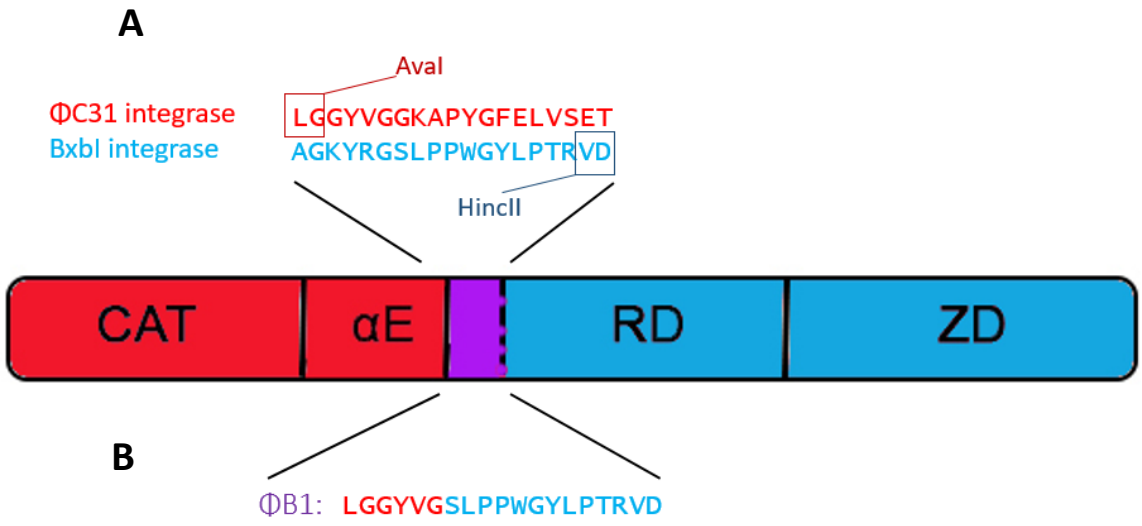

Figure 3.16: (A) Regions of interest in ΦC31 integrase and BxbI integrase amino acid sequences aligned and (B) amino acid sequence for hybrid integrase linker.

In all instances, ΦC31 integrase is denoted in red, BxbI integrase in blue and the hybrid integrase linkers in purple. In (A), the amino acid sequence of ΦC31 integrase from the end of its αE helix to the beginning of its recombinase domain is aligned against the equivalent region in BxbI integrase, with the location of the restriction enzymes AvaI and HincII indicated. In (B), the amino acid sequence for the hybrid integrase linker ΦB1 is shown. A pictorial diagram of the hybrid integrase construction is included to represent where these linkers connect ΦC31 integrase domains to those of BxbI integrase relative to the aligned amino acid sequences for both parental proteins shown in (A). Cat: catalytic domain; αE: αE helix; RD: recombinase domain; ZD: zinc ribbon domain.

bΦb10 *attP* site

5' TGGTTTGTCTGGTCAACCACACCTTTGAGTGTGGTGTACGGTACAAACCC 3'
3' ACCAAACAGACCAGTTGGTGTGGAAACTCACACCACATGCCATGTTTGGG 5'

bΦb10 *attB* site

5' TCGGCCGGCTTGTCGACGACGCCCTTGGGCGTCGTCAGGATCATCCGGGC 3'
3' AGCCGGCCGAACAGCTGCTGCGGGAACCCGCAGCAGTCCTAGTAGGCCCG 5'

Figure 3.17: Recognition *att* sites for the hybrid integrase ΦB1.

Exact *attP* and *attB* recognition site sequences for the hybrid integrases ΦB1. Please refer to Figure 3.4 for full ΦC31 integrase and BxbI integrase *att* site sequences. Red denotes a ΦC31 integrase origin and blue denotes a BxbI integrase origin. The crossover dinucleotide in each site is in bold typeface. bΦb10 *attP* indicates a recognition site in which the central ten nucleotides come from the ΦC31 integrase *attP* site, whilst the nucleotides flanking this region come from the BxbI integrase *attP* site. This naming scheme remains consistent throughout all hybrid sites.

3.7 Second hybrid integrase and site construction

3.7.i Hybrid integrase construction

The GeneArt plasmid BxbI_altered was digested with NdeI and HincII, whilst the GeneArt plasmid PHIC31_altered was digested with NdeI and AvaI. The larger fragment from BxbI_altered was ligated with the smaller fragment from PHIC31_altered as well as the ΦB1 linker, creating the *in vitro* precursor plasmid pHM11*. This plasmid was then fragment-swapped with the *in vitro* expression plasmid pSDC384 using the restriction enzymes NdeI and BamHI to create the *in vitro,* high expression plasmid pHM11 (Figure 3.18). To create an *in vivo*, low-expression plasmid, pHM11 was fragment swapped with pFO2 using the restriction enzymes NdeI and KpnI, resulting in pHM23 (Figure 3.19).

3.7.ii Hybrid site construction

The oligonucleotide sequences for the hybrid recognition sites bΦb10 *attP* and *attB* can be seen in Table 2.4. Substrate plasmids were constructed exactly as described in Section 3.3.ii, creating the precursor, one-site substrate plasmids pHM15 (bΦb10 *attP*) and pHM16 (bΦb10 *attB*) and the *in vivo* two-site substrate plasmid pHM28 (bΦb10 *attP*/*B*).

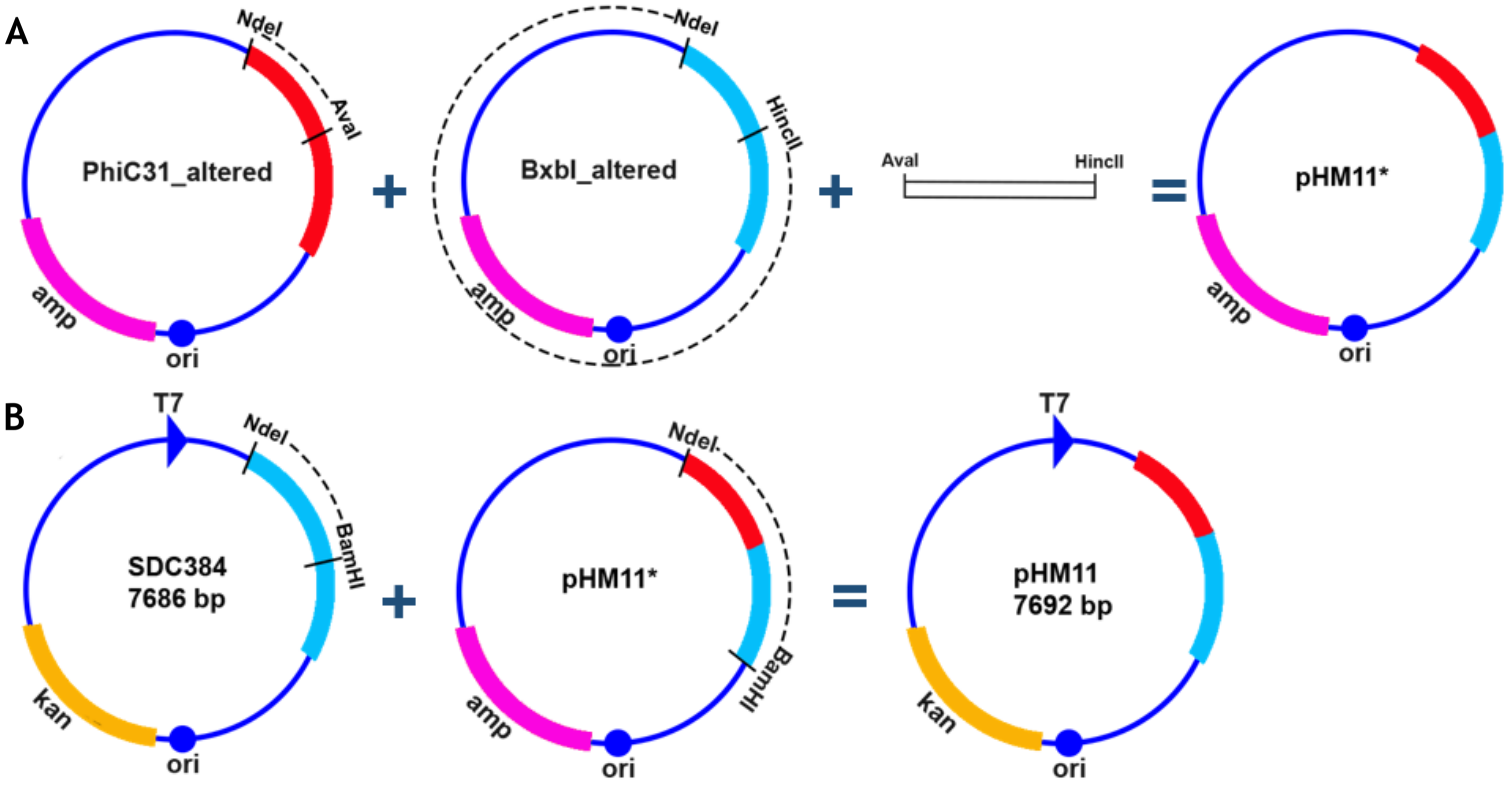

Figure 3.18: *in vitro* plasmid constructs for ΦB1.

(A) The larger fragment from an NdeI and HincII digest performed on the plasmid BxbI_altered, the smaller fragment from an NdeI and AvaI digest performed on the plasmid PhiC31_altered and an oligonucleotide linker with HincII and AvaI ends were ligated to create the *in vitro* precursor expression plasmid pHM11x. (B) The smaller fragment from an NdeI and BamHI digest performed on pHM11x and the larger fragment from the same digest performed on pSDC384 were ligated together to create the *in vitro* expression plasmid pHM11. In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

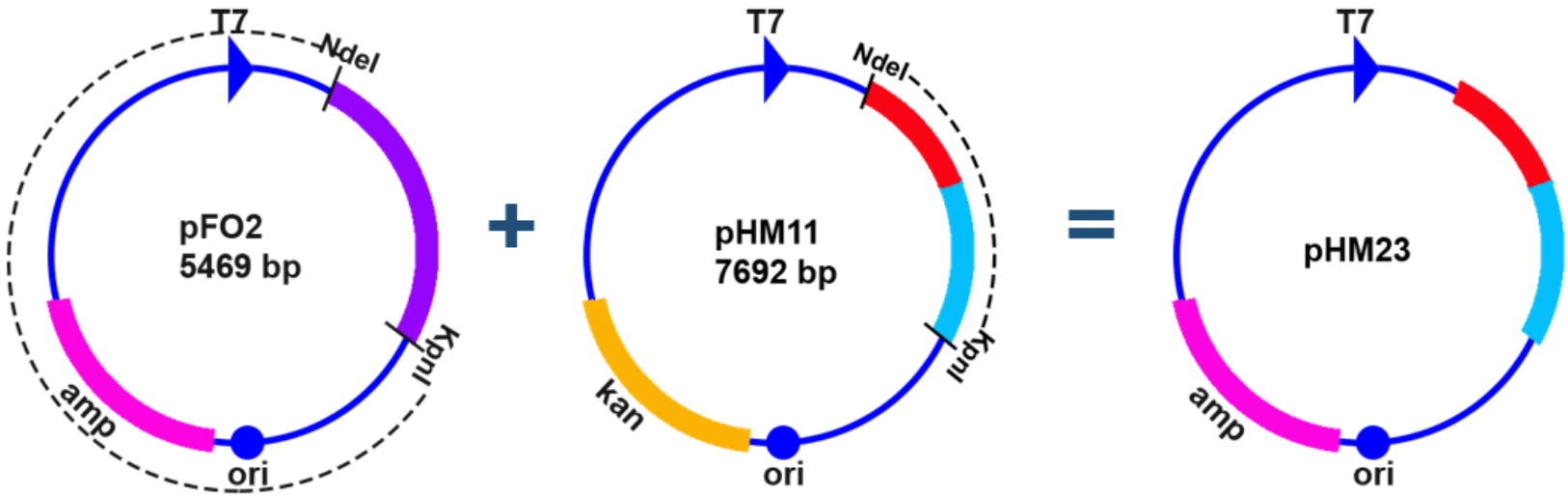

Figure 3.19: Construction of an *in vivo* expression hybrid integrase plasmid.

The larger fragment from an NdeI and KpnI digest on pFO2 and the smaller fragment from an NdeI to KpnI digest on pHM11 were ligated together to create the *in vivo* expression plasmid pHM23 In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

3.8 Hybrid activity *in vivo*

A MacConkey assay was performed to test the recombination activity of pHM23 (ΦB1), pFEM32 (ΦC31 integrase) and pFEM35 (BxbI integrase) on the hybrid two-site plasmid pHM28 (bΦb10 *attP* and *attB*). The plasmids pGD001 (ΦC31 *attP* and *attB*) and pFM35 (BxbI *attP* and *attB*) were used as positive controls for ΦC31 integrase and BxbI integrase, respectively. The results indicate that ΦB1 does not recognise pHM28, but does resolve pGD001 at an efficiency of around 33% (Figure 3.20 and Table 3.2).

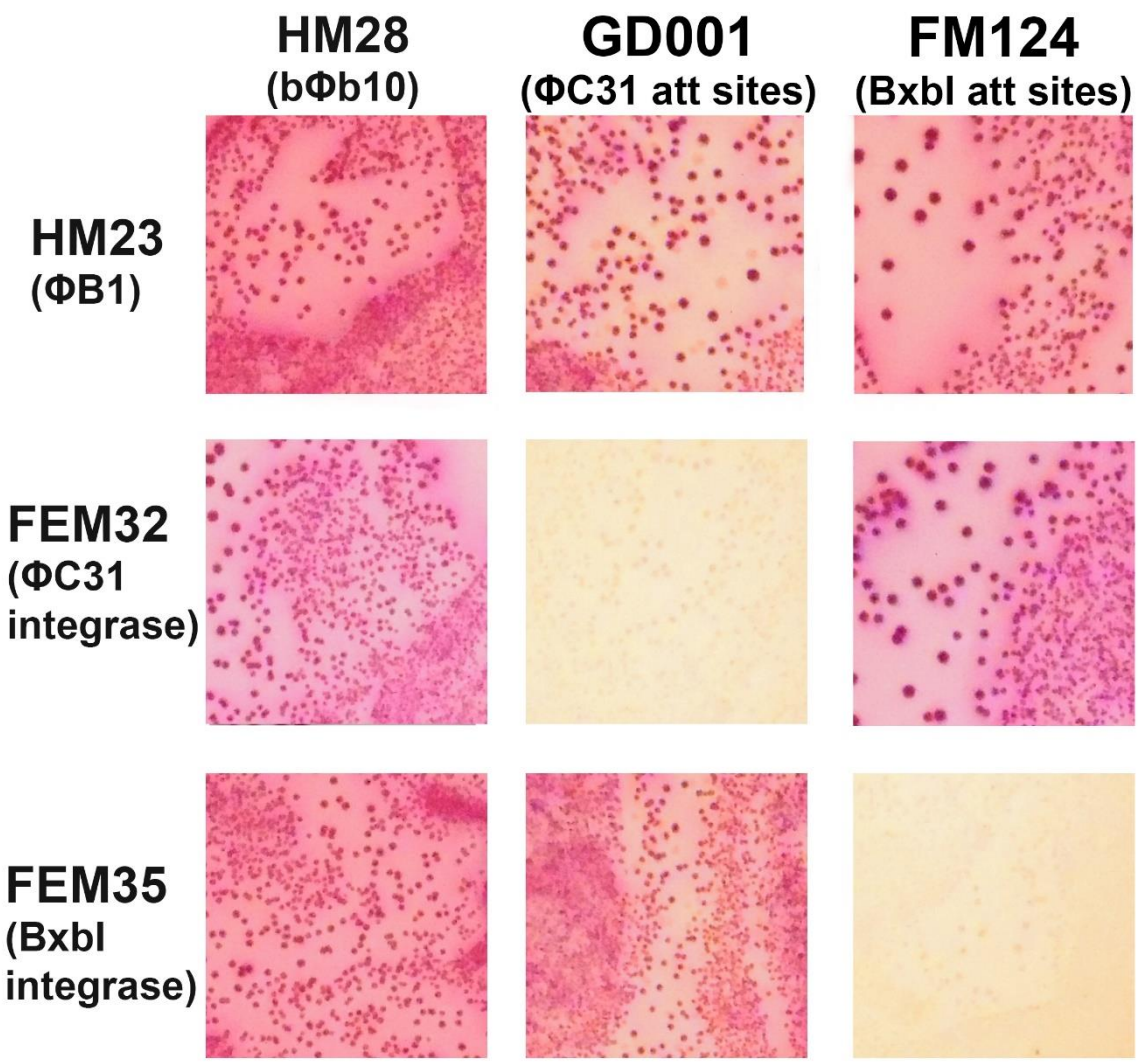

Figure 3.20: MacConkey assay results for ΦB1.

Red colonies denote no recombination whilst white colonies denote recombination. A plate containing a combination of both red and white colonies (such as pHM23 [ΦB1] with pHM28 [bΦb10 *attP*/*B*]) denotes incomplete resolution of the substrate-containing plasmids.

Name	Site name	No. of red colonies	No. of white colonies	% resolved
HM23	HM28	2006	0	0
	GD001	1762	986	35.8
	FM124	1982	0	0
FEM32	HM28	2664	0	0
	GD001	0	2753	100
	FM124	1786	0	0
FEM35	HM28	2132	0	0
	GD001	1985	0	0
	FM124	0	2104	100

Table 3.2: Colony counts for first round MacConkey plates for ΦB1 plus controls.

Results indicate 35.8% resolution of pGD001 (ΦC31 integrase *attP/B* recognition sites) by pHM23 (ΦB1), but no recognition of the hybrid two-site plasmid pHM28 by any integrase tested.

3.9 Hybrid integrase overexpression attempts

Many attempts were made to over express and subsequently purify the hybrid integrases BΦ1, BΦ2 and ΦB1 (Section 2.22) but, unfortunately, all attempts proved unsuccessful. Only one small-scale induction in the cell line BL21(DE3)[plysS] for BΦ1 indicated any discernible level of usable protein overexpression (Figure 3.21). However, all subsequent inductions failed to show even this small level of over-expression.

Inductions were attempted for all three hybrid integrases under many variable conditions: altered temperature (20 °C or 37 °C); time courses (one hour through to sixteen hours); induction with IPTG at varying OD_{600} values (from 0.4 through to 0.7, with 0.55-0.6 being the expected optimal OD_{600}); altering final concentration of IPTG in the induced sample; changing the location of the His-tag attached to the hybrid integrase from the C-terminus to the N-terminus, and changing the cell line to BL21(DE3)[plysE]and BL21(DE3) (see Table 2.1). Protein overexpression was observed when using BL21(DE3) for BΦ1 and ΦB1 but not BΦ2 (Figure 3.22), though protein obtained from this cell line is often unusable for purification.

Any attempts at protein purification from large-scale over-expression procedures (Section 2.22) resulted in protein products sticking to the His-column that were far too small to be any of the desired hybrid integrases.

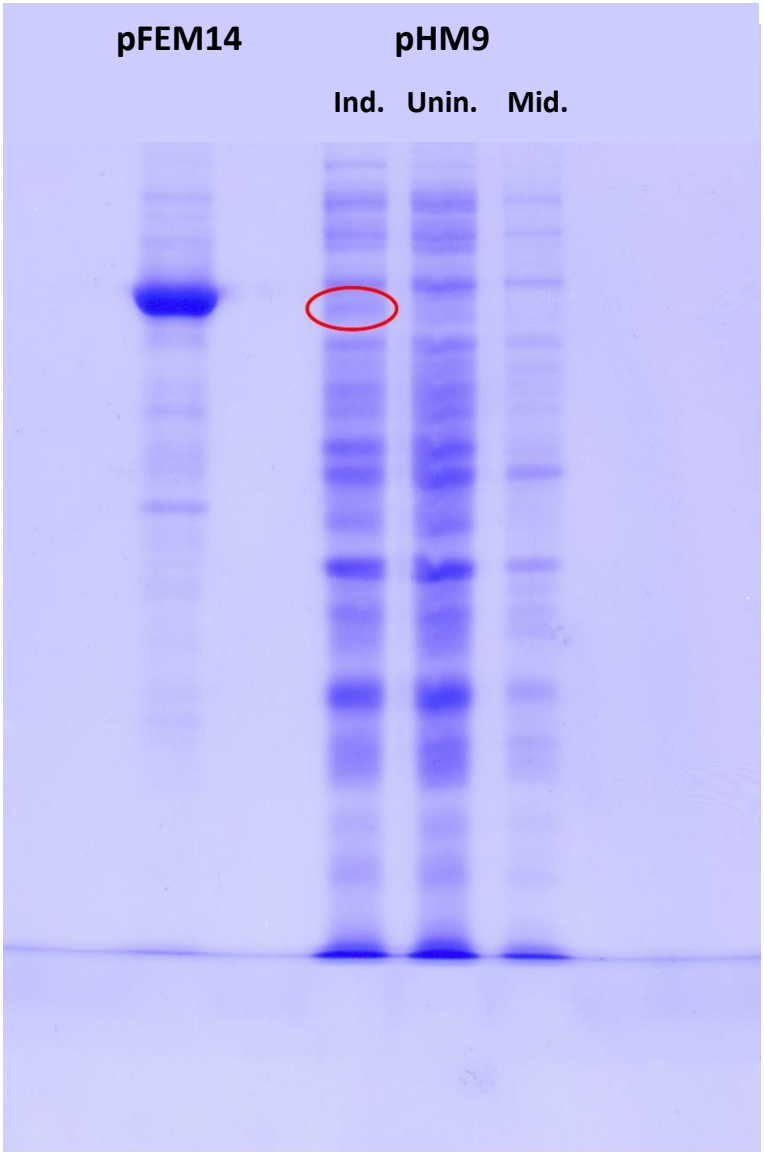

Figure 3.21: Attempted protein over-expression of pHM9 (BΦ1) in cell line BL21(DE3)[plysS].

The plasmid pFEM14 (ΦC31 integrase used as a size marker of 67kDa; expected size of hybrid integrase is slightly smaller than ΦC31 integrase (65kDa) and the band assumed to be the hybrid integrase protein is circled in red in the induced sample for BΦ1. Ind: induced with IPTG and left for 16 hours at 20 °C; unind: not induced with IPTG and left for 16 hours at 20 °C; mid: mid-log sample grown to OD_{600} of 0.6 with no IPTG added. Samples run on Laemmli gel SDS-PAGE and stained with Coomassie brilliant blue.

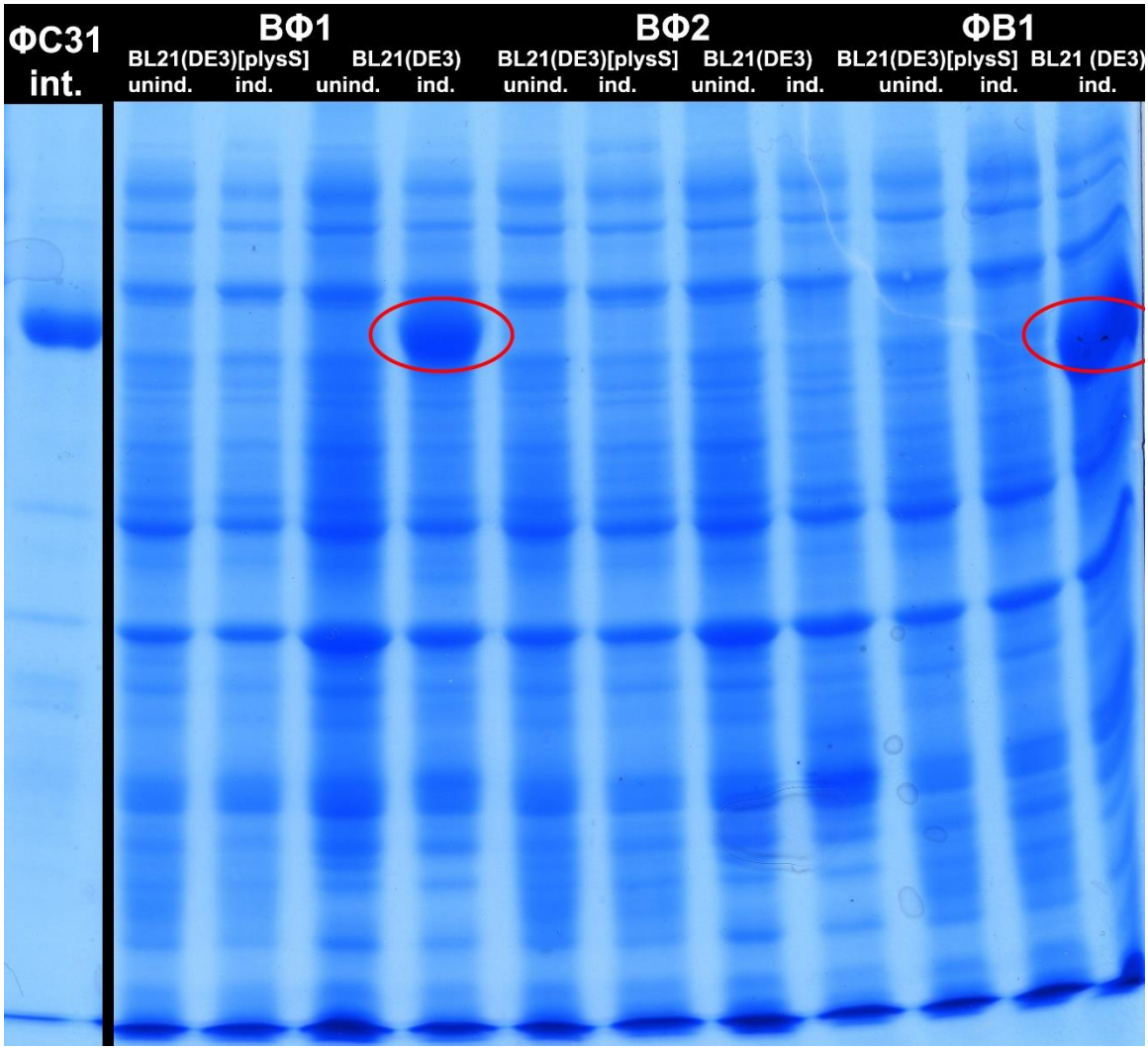

Figure 3.22: Induction attempts for BΦ1, BΦ2 and ΦB1 in the cell strains BL21(DE3)[plysS] and BL21(DE3).

ΦC31 integrase used as a size control of around 67kDa – BΦ1 and BΦ2 should be around 65kDa whilst ΦB1 should be around 58kDa. In both of the induced BL21(DE3) samples for BΦ1 and ΦB1, bands of approximately the right size are indicated (circled in red). Ind: induced with IPTG and left for 16 hours at 20 °C; unind: not induced with IPTG and left for 16 hours at 20 °C. Samples run on Laemmli gel SDS-PAGE and stained with Coomassie brilliant blue.

3.10 Discussion

In this chapter, two hybrid serine integrases were tested for their activity *in vitro* and *in vivo* on a variety of recognition sites, including sites created specifically for them. These new integrases consisted of an N-terminal BxbI integrase catalytic domain, a C-terminal ΦC31 integrase recombinase domain and a variable length linker of either BxbI integrase or ΦC31 integrase identity to connect the two domains together; the resultant hybrid integrases were named BΦ1 and BΦ2 to denote the short linker (BΦ1) and the longer linker (BΦ2) used to create them. Both sets of new recognition sites were based on ΦC31 *att* sites, with either the central 12 or 10 bp in the site replaced with the central 12 or 10 bp from the equivalent BxbI *att* site, resulting in sites named ΦbΦ12 *attP* and *attB* and ΦbΦ10 *attP* and *attB*. A hybrid integrase consisting of an N-terminal ΦC31 integrase catalytic domain and a BxbI integrase C-terminal recombinase domain was also constructed, named ΦB1. A new pair of *att* sites, named bΦb10 *attP* and *attB*, were also constructed, based on BxbI *att* sites but with the central 10 bp replaced with those of ΦC31 *att* sites.

Both hybrid integrases BΦ1 and BΦ2 were tested for their ability to carry out recombination *in vivo* on parental *att* sites and hybrid *att* sites. Hybrid *att* sites were also tested to see whether either parent integrase - ΦC31 integrase or BxbI integrase – could carry out recombination and/or cleavage both *in vivo* and *in vitro* on them. Based on the experimental results obtained, both BΦ1 and BΦ2 either failed to carry out any recombination on the *att* sites tested, or were not expressed at a high enough level to exhibit any activity on these *att* sites. Whilst ΦB1 demonstrated partial resolution of parental ΦC31 integrase *att* sites, failure to successfully overexpress and purify the integrase means that the activity of ΦB1 *in vitro* is, currently, unknown.

Hybrid integrase overexpression and purification

There are two reasons most likely to result in all three hybrid integrases failing to overexpress even when many different induction parameters were tested.

One possible reason is that the resultant proteins are insoluble and so were not present in the solution purified. Another reason is that both parental integrases are large and share very limited sequence similarity to each other, and are therefore likely to fold differently. It may be that attaching regions from one of these integrases to the other produces a protein unable to fold properly, resulting in protein misfolding. Protein misfolding is often toxic to the cell meant to produce it, as a build-up of misfolded protein usually triggers the stress responses put in place by the cell to protect against such proteins, but this in turn induces certain cell death pathways because of the prolonged stress put on the cell (Rao and Bredesen, 2004).

This hypothesis is backed up by the results of the inductions in different BL21 cell lines. Both BL21(DE3)[pLysS] and [pLysE] constitutively express the T7 lysosyme, which reduces the basal expression of the target gene (in this case, BΦ1, BΦ2 or ΦB1) via the inhibition of T7 RNA polymerase (Moffatt and Studier, 1987). The pLysE plasmid expresses higher levels of T7 lysosyme than pLysS, making it ideal to use when the protein intended to be expressed may be toxic, as it further limits the amount of protein overexpressed. However, if the protein product is indeed toxic, it could potentially kill the cells expressing it or become rapidly degraded, which may explain in part why little discernible expression of any of the hybrid integrases was observed. However, the strain BL21(DE3) does not contain a plasmid which expresses T7 lysosyme and, as a result, much higher levels of protein expression can be reached, particularly for toxic and misfolded proteins which are otherwise inhibited. This was the case for both BΦ1 and ΦB1, which successfully overexpressed in BL21(DE3), but not BΦ2. It is possible that BΦ2 failed to overexpress due to experimental error, but it is also possible that the longer, ΦC31 integrase-based linker in BΦ2 makes protein folding so difficult that BΦ2 is simply not produced at all.

If it is true that expression of any of the hybrid integrases is toxic to the cell, it follows that failure to observe even low levels of resolution occurring within the MacConkey assay may be due to the cell repressing expression of the hybrid integrase in order to prevent cell death. Consequently, there would have been little point in attempting to purify either BΦ1 or ΦB1 from a BL21(DE3)-based

overexpression to test the proteins *in vitro*, as any activity then observed would not be reproducible within an *in vivo* environment, thus reducing the possible uses of both hybrid integrases as tools for genetic engineering, gene therapy or biological computing.

Activity of ΦC31 integrase and BxbI integrase on hybrid att *sites*

The results obtained for BxbI integrase with both sets of hybrid *att* sites ΦbΦ10 and 12 suggest that having the central 10 or 12 bp following the BxbI integrase *att* site sequence is not enough to allow this integrase to recognise, cleave or recombine the sites. It is very likely that there are regions further out from the central crossover dinucleotide in the *att* sites essential for recognition by their cognate integrase. This would explain why BxbI integrase does not recognise the hybrid sites but ΦC31 integrase does, as the basepairs further out from the centre of the hybrid sites follow the sequence of the ΦC31 *att* sites. Indeed, there are conserved motifs outwith the central 10 or 12 bp between the natural *attP* and *attB* sites for each integrase which may be the regions responsible for integrase recognition, allowing the integrase to perform the initial cleavage of the sites required prior to recombination (Figure 3.4).

Results for ΦC31 integrase with both hybrid sets of sites help to narrow down the nucleotides in an *att* site which may be essential for integrase cleavage of its own sites. Since ΦC31 integrase cleaved ΦbΦ10 much faster than it did ΦbΦ12 *in vitro,* and only ΦbΦ10 was resolved *in vivo,* the sixth basepair from the centre of each *att* site (a C for BxbI *att* sites; an A or a T for ΦC31 *att* sites) may be involved in *att* site cleavage and, at least in the case of ΦC31 integrase, essential for efficient site cleavage. It is not surprising that ΦC31 integrase did not recombine either set of hybrid sites *in vitro,* as the central basepairs within the hybrid sites hypothesised to interact with the αE helix and beginning of the recombinase domain of an integrase are taken from BxbI *att* sites. Furthermore, there are very few basepair positions within ΦC31 natural *attP* and *attB* sites which remain the same in BxbI integrase sites – only 2 bp (one being the T of the central dinucleotide) in *attP* between species and 4 bp

(one being the T of the central dinucleotide) in *attB* (Figure 3.4). This decreases the likelihood that ΦC31 integrase would have been able to successfully recombine either hybrid *att* sites ΦbΦ10 or 12 *attP* and *attB* to any significant level as the sequences between both parent integrase recognition sites are so dissimilar.

The observation that neither ΦC31 integrase nor BxbI integrase resolved bΦb10 *in vivo* lends further credence that, without the necessary conserved regions of the *att* sites out with the central 10 bp, recombination by either parent integrase is impossible. It also shows that BxbI integrase is not capable of the small amount of resolution demonstrated by ΦC31 integrase on the opposite set of hybrid *att* sites ΦbΦ10, implying that recognition of outlying conserved basepairs in the BxbI integrase-like regions of bΦb10 *attP* and *attB* is not sufficient to cleave the site and even partially resolve it, as was the case for ΦC31 integrase. This possibly suggests that BxbI integrase may be more site specific than ΦC31 integrase is, even though the current consensus on ΦC31 integrase based on existing research is that it is unable to recognise its own *att* sites when even single nucleotide changes have been made (Smith et al, 2004; discussed further in Chapter 5). This also defends the stance of Singh et al (2014), who concluded that regions essential for site recognition likely differed between integrase species (see Section 1.3.ii).

Activity of ΦB1 in vivo

The partial resolution observed for ΦB1 on ΦC31 natural *attP* and *attB* sites is more promising. It demonstrates that ΦB1 is likely to be less toxic to the cell than either BΦ1 or BΦ1, providing evidence that it may be possible to successfully overexpress and purify the protein for *in vitro* work in the future. It also suggests that a combination of domains from both parent integrases more in line with the successful chimeric constructs of Faruggio and Calos (2014) – that is, an N-terminal ΦC31 integrase catalytic domain and a C-terminal recombinase domain from another integrase (in this case BxbI integrase) – is more likely to function than the other way around (see Section

1.7). However, this may only be true in the case of two large serine integrases with a high degree of similarity.

However, the observation that ΦB1 is capable of some activity *in vivo* demonstrates that a hybrid integrase need not be constructed of domains from very similar integrases, as is the case for the chimeras that Faruggio and Calos (2014) constructed between ΦC31 integrase and its two related proteins, ΦBT1 integrase and TG1 integrase.

If work was to be continued on large serine hybrid integrases, exploring the ability of ΦB1 to recognise and recombine further combinations of *att* sites as well as continuing attempts to purify the protein for *in vitro* experiments should be the first priority. From this, decisions can be made as to the viability of this method for creating integrases for use in synthetic biology, genetic engineering and gene therapy since, as it stands, using ΦC31 integrase and BxbI integrase to create hybrid integrases has been notoriously difficult. It may be that altering the large serine integrases used to engineer the hybrids would make constructing functional hybrid integrases much easier, or using a combination of large serine integrases and the smaller resolvases instead, as discussed in Chapter 4.

Chapter Four: Tn3 resolvase/ΦC31 integrase hybrids

4.1 Introduction

As demonstrated by the zinc finger recombinases - in which the catalytic domain of the small serine recombinase Tn3 resolvase was attached to the DNA binding domain of the zinc finger protein Zif268 (described further in Section 1.7) - the modular, independent nature of the serine recombinases allows for the swapping of domains from one species to another to create new and functional proteins. This chapter describes research aiming to create two, new hybrid integrases using the catalytic domain from Tn3 resolvase and the recombinase domain from ΦC31 integrase, in a manner similar to the creation of the integrases described in Chapter 3.

4.2 Hybrid integrase and site design

For the purposes of this project, the catalytic domain of the resolvase mutant NM was used (in place of wild type Tn3 resolvase) due to its hyperactive nature (see Section 2.7). ΦC31 integrase protein remained identical to that used in Chapter 3.

The rationale behind hybrid integrase and site design remained much the same as for the hybrids created and discussed in Chapter 3.

4.2.i Hybrid integrase design

Two hybrids were designed with a similar domain composition as those seen in Chapter 3, with an NM resolvase catalytic domain and αE helix in place of BxbI integrase and a ΦC31 integrase recombinase domain (shown pictorially in Figure 4.1). As in Chapter 3, the difference between the two hybrids is the length of the linker joining the two parental proteins together (discussed further in Section 4.3 below).

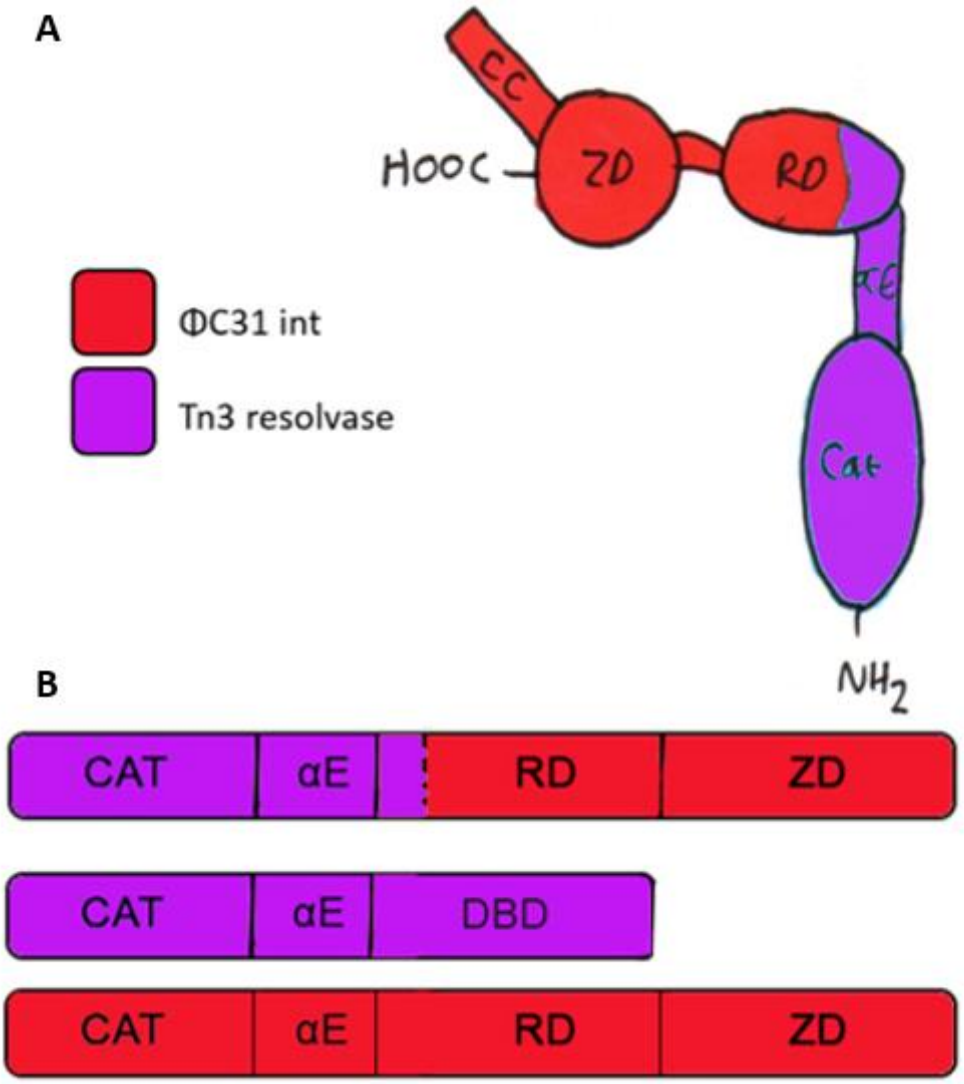

Figure 4.1: Pictorial representation of Tn3 (NM) resolvase/ΦC31 integrase hybrids.

(A) Basic structural diagram of the new hybrid proteins, denoting Tn3 resolvase origin in purple and ΦC31 integrase in red. (B) Linear diagrams of the basic hybrid construct and its parental proteins, demonstrating where Tn3 resolvase and ΦC31 integrase will be joined together.

4.2.ii Hybrid site design

Hybrid sites were created following much the same rationale as in Chapter 3 – a central group of basepairs from the ΦC31 integrase *att* sites was removed and replaced with the central basepairs from Tn3 *res* site I.

Additionally, two further observations were taken into account. Bearing in mind that the ultimate goal is to create sites which are only recognised by the hybrid integrases specifically created for them, the more rapid *in vitro* cleavage and *in vivo* recognition observed for ΦbΦ10 *att* sites over ΦbΦ12 *att* sites by ΦC31 integrase (Section 3.5) is a result we would hope to avoid for the new sites, so taking more than ten central basepairs from Tn3 *res* site I seemed logical to help prevent ΦC31 integrase itself from recombining the hybrid sites. Secondly, zinc finger recombinase recognition sites containing only the sixteen central basepairs from Tn3 *res* site I are sufficiently recognised by the zinc finger recombinases created for them (Prorocic et al, 2011; Proudfoot et al, 2011). With this in mind, it was decided to create two sets of *att* sites – one set which contained the central 12 bp from Tn3 *res* site I and one set which includes 14 bp (shown in Figure 4.2).

A comparison between the central 16 bp of ΦC31 integrase *attP* and *attB* and Tn3 *res* site I (shown in Figure 4.3) demonstrates that 8 out of 16 bp and their relative locations are identical in *res* site I and the *attP* site. Whilst this level of sequential similarity could potentially prove to be problematic with regards to cross-talk (see Section 1.3.ii), in contrast, only 2 out of 16 bp and their relative locations are identical in *res* site I and the *attB* site, which may limit the ability of ΦC31 integrase to recognise the new hybrid sites.

Sites were designed as oligonucleotides and were synthesised by Eurofins Genomics.

ΦC31 integrase *attP* site

5’ GTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAG 3’
3’ CATCACGGGGTTGACCCCATTGGAAACTCAAGAGAGTCAACCCCCGCATC 5’

ΦC31 integrase *attB* site

5’ CGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCAC 3’
3’ GCCACGCCCACGGTCCCGCACGGGAACCCGAGGGGCCCGCGCATGAGGTG 5’

16 central Tn3 *res* site I nucleotides

5’ AAA TAT TAT AAA TTA T 3’
3’ TTT ATA ATA TTT AAT A 5’

ΦtΦ12 (purple) or 14 (purple+orange) *attP* site:

5’ AGT AGT GCC CCA ACT GGG GTA TAT TAT AAA TTC TCT CAG TTG GGG GCG TAG 3’
3’ TCA TCA CGG GGT TGA CCC CAT ATA ATA TTT AAG AGA GTC AAC CCC CGC ATC 5’

ΦtΦ12 (purple) or 14 (purple+green) *attB* site:

5’ CCG CGG TGC GGG TGC CAG GGC GAT ATT ATA AAT TCC CCG GGC GCG TAC TCC AC 3’
3’ GGC GCC ACG CCC ACG GTC CCG CTA TAA TAT TTA AGG GGC CCG CGC ATG AGG TG 5’

Figure 4.2: Exact *attP* and *attB* recognition site sequences for ΦC31 integrase, 16 bp of interest from Tn3 *res* site I and the new sequences for the hybrid *att* sites ΦtΦ14 and ΦtΦ12.

In all instances, red denotes a ΦC31 origin and purple denotes a Tn3 origin. The crossover dinucleotide in each site is in bold typeface. ΦtΦ12 *attP* indicates a recognition site in which the central 12 bp come from Tn3 *res* site I, whilst the basepairs flanking this region come from the ΦC31 integrase *attP* site. This naming scheme remains consistent throughout all modified sites.

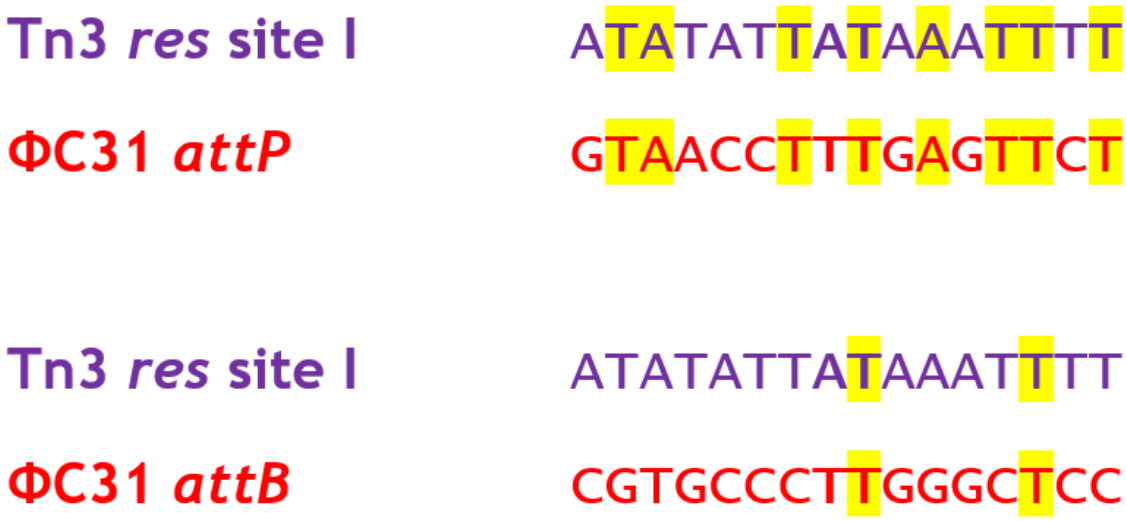

Figure 4.3: Comparison between the central 16 bp of Tn3 *res* site I and the central 16 bp of ΦC31 integrase *attP* and *attB* sites (top strands shown only).

Identical basepairs are highlighted in yellow. The crossover dinucleotide in each site is shown in boldface type.

4.3 Construction of hybrid integrases and sites

4.3.i Hybrid integrase construction

As mentioned in Section 4.2.i, two oligonucleotide linkers were designed to connect NM resolvase to ΦC31 integrase. Both linkers follow the amino acid sequence of Tn3 resolvase from residue 146 to either residue 149 or residue 152 (Figure 4.4). The DNA sequences for both are shown in Table 2.4 (Section 2.6). The corresponding amino acid sequence for each as well as each linker's relative location in the resultant hybrid integrase are shown in Figure 4.4. Both linkers begin with an EagI restriction site and end with a DraII site. The linkers, and the subsequent new hybrid integrases created using them, were named TΦ4 and TΦ8, with the number denoting the number of amino acids which are derived from Tn3 resolvase in the linker.

These linkers were designed with the construction of functional zinc finger recombinases specifically in mind – several ZFRs have been created which connect NM resolvase between residues 146 to 152 to Zif268 (Prorocic et al, 2011; Proudfoot et al, 2011). Both linkers were designed as oligonucleotides and synthesised by Eurofins Genomics.

For the creation of *in vitro* expression plasmids for both hybrids, the larger fragment from the NM resolvase-containing plasmid pFO2 digested with the restriction enzymes NdeI and EagI, the smaller fragment from the T7 inducible, high copy number ΦC31 integrase-containing plasmid pFEM14 digested with NdeI and DraII and either the TΦ4 or TΦ8 linker were ligated together to form pHLM5 and 6, respectively (Figure 4.5).

In vivo expression plasmids were created by digesting both pHLM5 and 6 and the low copy number expression plasmid pFO2 with NdeI and KpnI, ligating the larger fragment from pFO2 and the smaller fragment from either pHLM5 or 6 to create plasmids pHM10 (TΦ4) and 11 (TΦ8) (Figure 4.6).

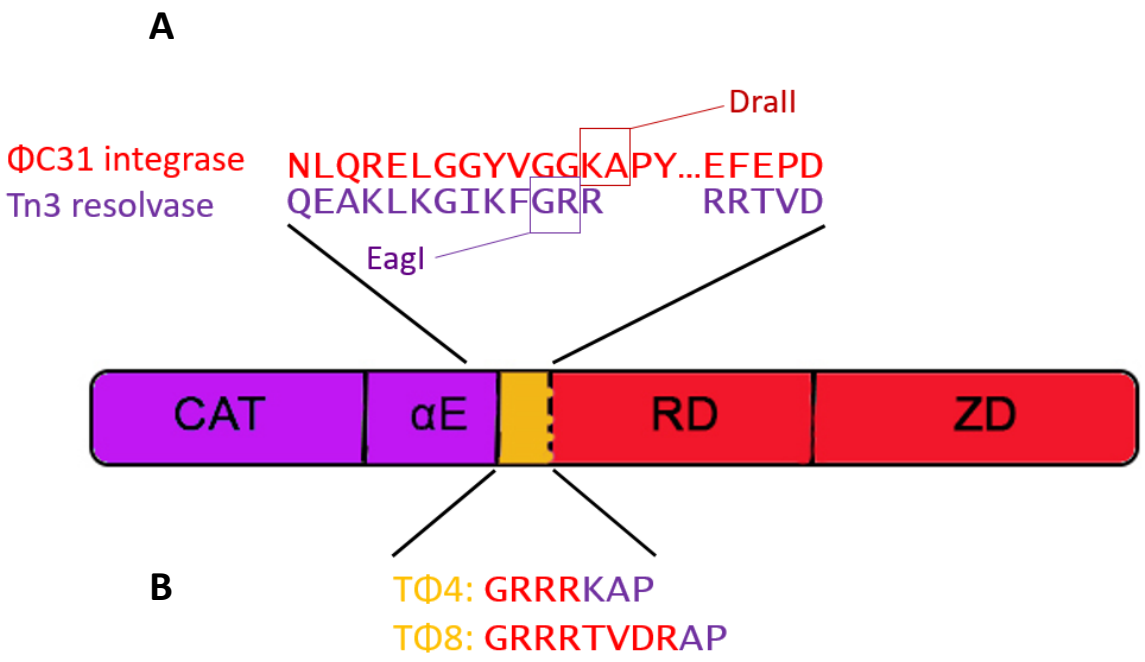

Figure 4.4: (A) Regions of interest in ΦC31 integrase and Tn3 resolvase amino acid sequences aligned and (B) amino acid sequence for both hybrid integrase linkers.

In all instances, ΦC31 integrase is denoted in red, Tn3 resolvase in purple and the hybrid integrase linkers in orange. In (A), the amino acid sequence of ΦC31 integrase from the end of its αE helix to the beginning of its recombinase domain is aligned against the equivalent region in Tn3 resolvase, with the location of the restriction enzymes EagI and DraII indicated. Ellipses are used to denote the skipping of amino acids in the ΦC31 integrase sequence. In (B), the amino acid sequences for the two hybrid integrase linkers TΦ4 and TΦ8 are shown. A pictorial diagram of the hybrid integrase construction is included to represent where these linkers connect Tn3 resolvase domains to those of ΦC31 integrase relative to the aligned amino acid sequences for both parental proteins shown in (A). Cat: catalytic domain; αE: αE helix; RD: recombinase domain; ZD: zinc ribbon domain.

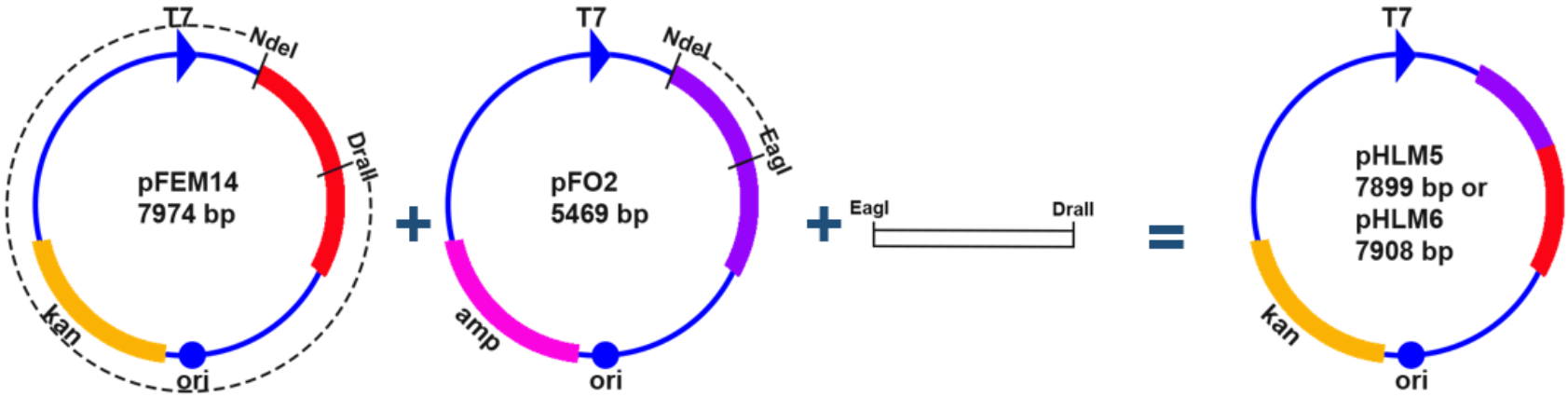

Figure 4.5: Construction of *in vitro* expression hybrid integrase plasmids.

The smaller fragment from NdeI to DraII was removed from pFEM14 and replaced with the smaller fragment from NdeI to EagI from pFO2 as well as either of two EagI-DraII oligonucleotide linkers to create the plasmids pHLM5 and pHLM6. In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

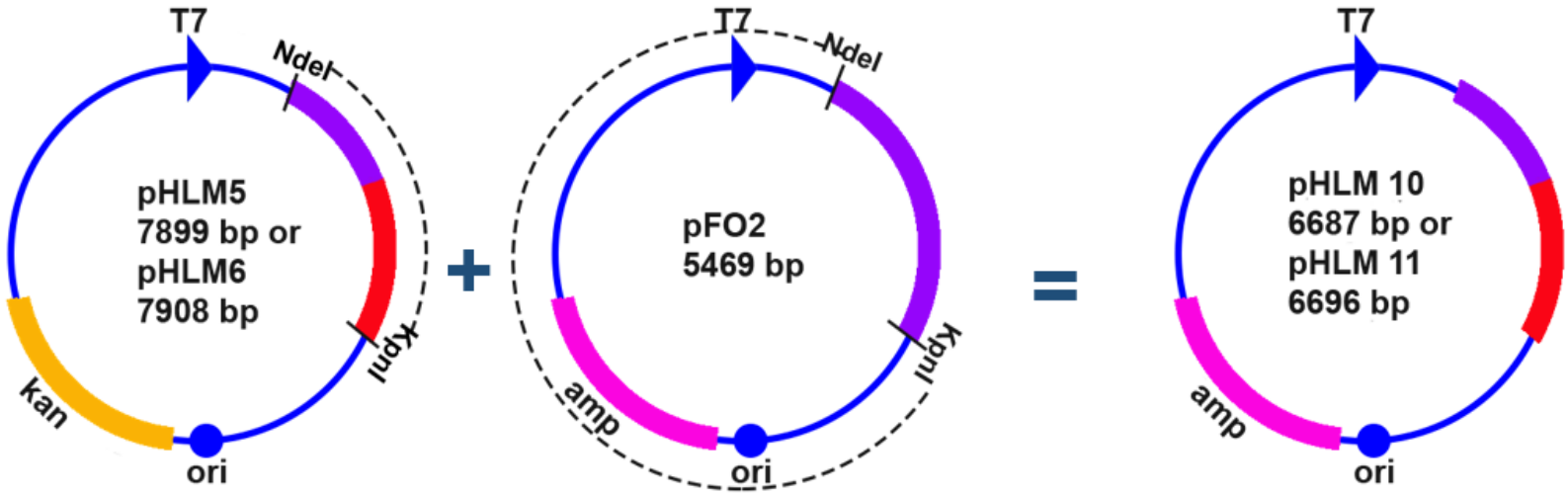

Figure 4.6: Construction of *in vivo* expression hybrid integrase plasmids.

The smaller fragment from NdeI to KpnI was removed from pFO2 and replaced with the smaller fragment from NdeI to KpnI from either pHLM5 or pHLM6 to create the plasmids pHLM10 and pHLM11. In all instances, the fragment taken to create the final plasmid is denoted by a dashed line.

4.3.ii Hybrid site construction

The oligonucleotide sequences for the hybrid recognition sites ΦtΦ12 *attP* and *attB* and ΦtΦ14 *attP* and *attB* can be seen in Table 2.4. *In vitro* and *in vivo* substrate plasmids were constructed as described in Section 3.3.ii, creating the *in vitro* precursor, one-site plasmids pHLM1 (ΦtΦ14 *attP*), pHLM2 (ΦtΦ14 *attB*), pHLM3 (ΦtΦ12 *attP*) and pHLM4 (ΦtΦ14 *attB*); the *in vitro* two-site plasmids pHLM12 (ΦtΦ14 *attP*/*B*) and pHLM13 (ΦtΦ12 *attP* and *attB*) and the *in vivo* two-site plasmids pHLM7 (ΦtΦ14 *attP*/*B*) and pHLM8 (ΦtΦ12 *attP* and *attB*) (Figures 4.7 and 4.8). All two-site plasmids were constructed to give the most informative NruI restriction digest (see Figure 3.9).

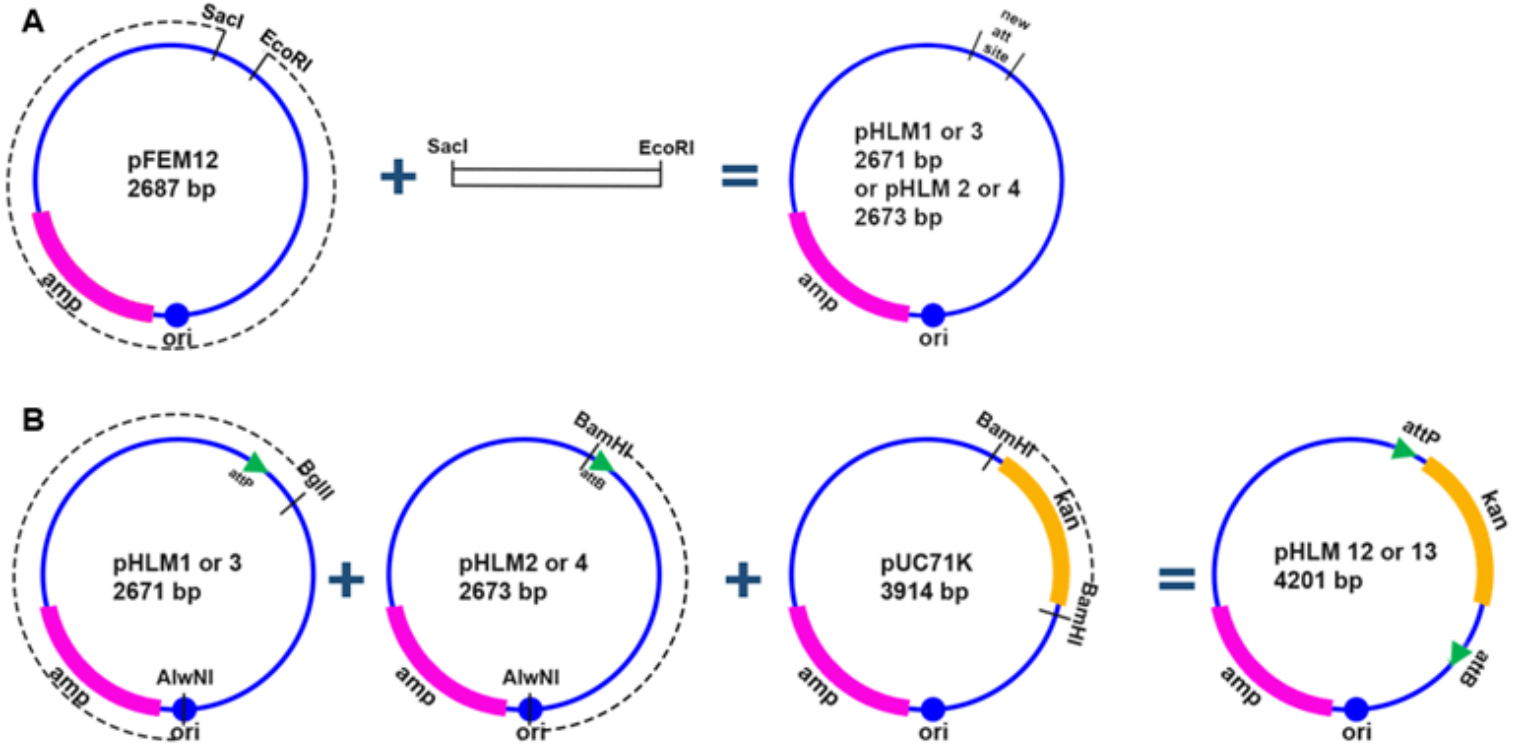

Figure 4.7: Construction of *in vitro* precursor and two-site substrate plasmids.

(A) The plasmid pFM12 was cut with SacI and EcoRI and an oligonucleotide of a hybrid *attP* or *B* site with the same restriction site ends was ligated in. The resultant precursor substrate plasmid was named either pHLM1, pHLM2, pHLM3 or pHLM4 depending on the *att* site within it (see main text). (B) pUC71K was digested with BamHI and the smaller fragment was ligated together with the larger fragment of an AlwNI and BglII digested *attP*-containing plasmid (e.g. pHLM1) and the smaller fragment from an AlwNI and BamHI digested *attB*-containing plasmid (e.g. pHLM2). The resultant *in vitro* two-site substrate plasmids were named pHLM12 or pHLM13. In all instances, the fragment taken to create the final plasmids is denoted with a dashed line.

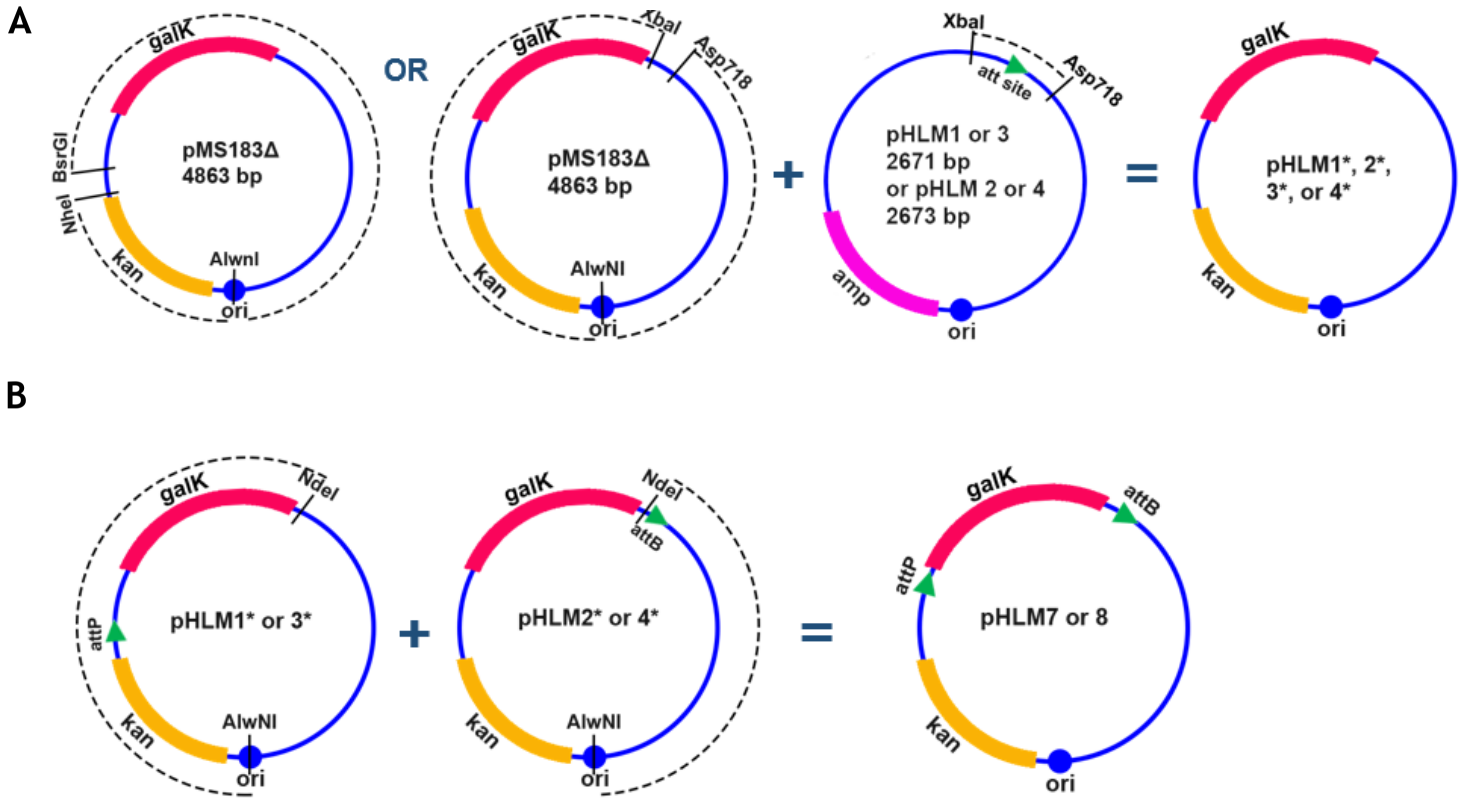

Figure 4.8: Construction of *in vivo* one-site and two-site substrate plasmids

(A) The two larger fragments indicated with a dashed line in the left hand pMS183Δ plasmid along with the smaller fragment from an Asp718 and XbaI digested one-site *att* site *in vitro* plasmid (e.g. pHLM1) were ligated together to create the *in vivo* plasmids pHLM1* and pHLM3*. Similarly, the two larger fragments indicated with a dashed line in the right hand pMS183Δ plasmid along with the smaller fragment from an Asp718 and XbaI digested one-site *att* site *in vitro* plasmid were ligated together to create the *in vivo* plasmids pHLM2* and pHLM3*. (B) The larger fragment from an AlwNI and NdeI digest of *attP*-containing plasmids pHLM1* or pHLM3* and the smaller fragment taken from the same digest for *attB*-containing plasmids pHLM2* or pHLM4* were ligated together to form the two site *in vivo* plasmids pHLM7 and pHLM8, respectively. In all instances, the fragments used to create the final plasmids are denoted with a dashed line.

4.4 Hybrid integrase activity *in vivo*

The hybrid recombinases TΦ4 and TΦ8 (as well as their parental proteins NM resolvase and ΦC31 integrase) were tested for their activity on various recognition sites *in vivo* using a MacConkey assay (Section 2.21). The results indicate a low level of activity for TΦ4 and TΦ8 on both sets of hybrid sites - ΦtΦ14 *attP* and *attB* and ΦtΦ12 *attP* and *attB* – but no activity on either ΦC31 integrase or Tn3 resolvase natural recognition sites (Figure 4.9; colony counts in Table 4.1). Unexpectedly, ΦC31 integrase resolved both ΦtΦ14 *attP* and *attB* and ΦtΦ12 *attP* and *attB* as well as it resolved its cognate sites (Figure 4.9); this was confirmed by running select colonies on a 1.2% agarose gel (Figure 4.10**)** and sending the recombinant gel products to Eurofins Genomics for sequencing. This is explored further in Section 4.7 below.

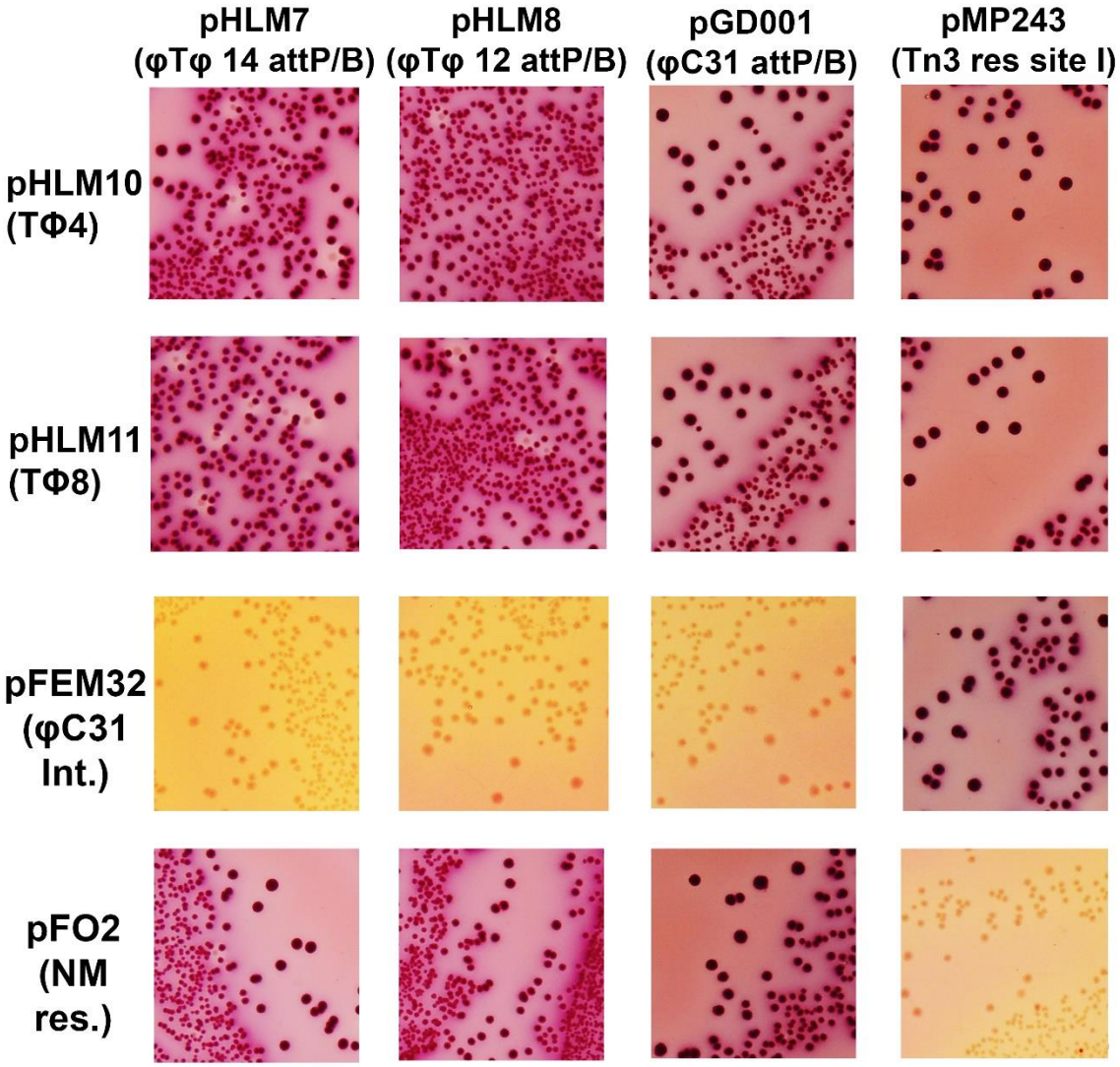

Figure 4.9: MacConkey results for hybrid integrases TΦ4 and TΦ8.

White colonies denote site resolution whilst red colonies denote no resolution. A mix of both coloured colonies denotes partial resolution. Both pFEM32 (ΦC31 integrase) and pFO2 (NM resolvase) fully resolve their cognate site plasmids (pGD001 and pMP243 respectively) whilst pFEM32 also fully resolves both hybrid plasmids pHLM7 (ΦTΦ14 *att* sites) and pHLM8 (ΦtΦ12 *att* sites). Hybrid integrase plasmids pHLM10 and pHLM11 (TΦ4 and TΦ8 respectively) both resolve a low percentage of pHLM10 and pHLM11.

Name	Site name	No. of red colonies	No. of white colonies	% resolved
TΦ4	ΦtΦ 14 *attP/B*	504	12	2.38
	ΦtΦ 12 *attP/B*	1204	8	0.67
	ΦC31 *attP/B*	1762	0	0
	Tn3 *res* site I	2001	0	0
TΦ8	ΦtΦ 14 *attP/B*	456	8	1.75
	ΦtΦ 12 *attP/B*	1004	7	0.8
	ΦC31 *attP/B*	1844	0	0
	Tn3 *res* site I	1982	0	0
ΦC31 int.	ΦtΦ 14 *attP/B*	0	2034	100
	ΦtΦ 12 *attP/B*	0	1956	100
	ΦC31 *attP/B*	0	2753	100
	Tn3 *res* site I	2263	0	0
Tn3 res.	ΦtΦ 14 *attP/B*	2132	0	0
	ΦtΦ 12 *attP/B*	1566	0	0
	ΦC31 *attP/B*	2030	0	0
	Tn3 *res* site I	0	2104	100

Table 4.1: Colony counts for MacConkey Assay.

Activity of each integrase on respective *att* sites is measured as a percentage of resolved (white) colonies against total number of observed (red and white) colonies.

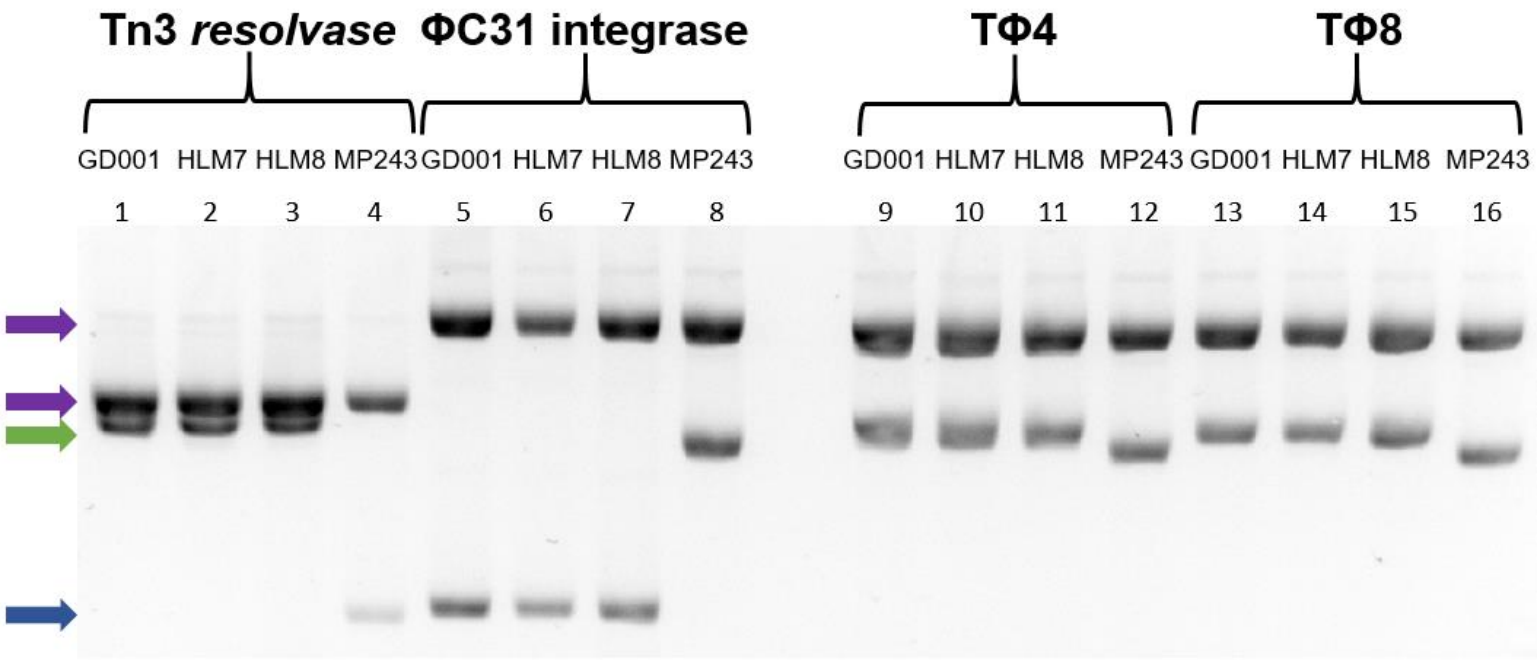

Figure 4.10: MacConkey assay resolution products.

Lanes 4 and 5 are positive resolution controls. Resolved substrate plasmids seen in lanes 4, 5, 6 and 7 are denoted with a blue arrow; unresolved substrate plasmids are denoted with a green arrow and recombinase expression plasmids are denoted with a purple arrow. The level of resolution observed for the hybrid integrases TΦ4 and TΦ8 on hybrid substrate plasmids pHLM7 and pHLM8 in the MacConkey assay is too small to be observed on this gel. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

4.5 Hybrid overexpression and activity *in vitro*

The hybrid integrase expression plasmids HLM5 and HLM6 were transformed into the cell line BL21(DE3)pLysS (using the chemically competent transformation protocol detailed in Section 2.8.i) for induction and subsequent over-expression of TΦ4 and TΦ8. Following a small-scale induction of 20 ml of bacterial culture, the samples were visualised by SDS-PAGE to confirm the presence of hybrid protein.

Once the small-scale induction was shown to be successful, large-scale inductions of 400 ml of culture for each plasmid were performed in order to purify the hybrid proteins. IPTG was added to a final concentration of 1 mM to each culture once they had reached an OD_{600} of 0.6 (determined by spectrophotometry), after which the inductions were left for sixteen hours at 4 °C. Following this, TΦ4 and TΦ8, which both contain a C-terminal His_6 tag were purified (see Section 2.22). The resultant fractions were then dialysed to further concentrate them (Section 2.22). These fractions were then analysed using SDS-PAGE and the fractions containing hybrid protein were then used for subsequent *in vitro* experiments (Figure 4.11).

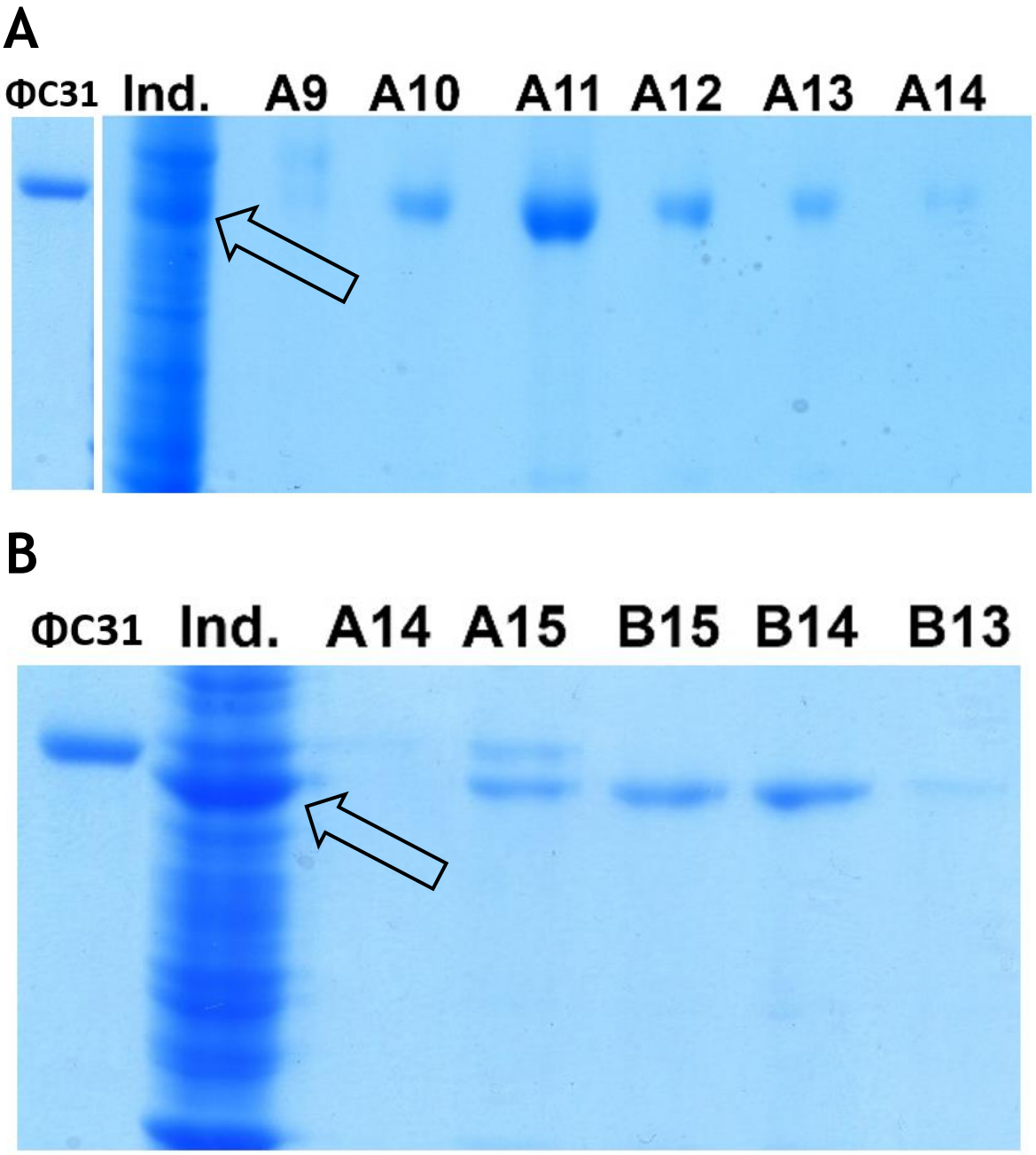

Figure 4.11: Purified fractions for the hybrid integrases (A) TΦ4 and (B) TΦ8.

ΦC31 integrase is used as a size control of ~69 kDa; both hybrid integrases are ~65 kDa. "Ind." is the unpurified, IPTG induced sample to demonstrate correct protein size (indicated with an arrow in both cases). Column headings are the names of the different fractions purified. For TΦ4, A10, A11 and A12 were chosen to conduct *in vitro* experiments with; for TΦ8, B15 and B14 were chosen.

The activity levels of TΦ4 and TΦ8 on hybrid site combinations ΦtΦ14 *attP* and *attB* (plasmid pHLM12) and ΦtΦ12 *attP* and *attB* (pHLM13) as well as ΦC31 integrase *att* sites (pFM16) was tested using a recombination assay (described in Section 2.24.i). ΦC31 integrase was used as a positive control on its own recognition sites. The assay was left to run for 16 hours. As before in Chapter 3, recombination assay samples were digested with the restriction enzyme NruI to create linearised products to be run on a 1.2% agarose gel. Neither hybrid integrase recombined ΦC31 integrase cognate sites; however, TΦ4 and TΦ8 exhibited very slight activity on pHLM12 and pHLM13, with TΦ4 exhibiting more activity on pHLM13 than TΦ8 (Figure 4.12).

To complement the recombination assay, an ethylene glycol cleavage assay was undertaken (described in Section 2.24.ii). ΦC31 integrase was used as a positive control on its own recognition sites. Reactions were left for sixteen hours before being run on a 1.2% agarose gel. Neither hybrid integrase exhibited any cleavage activity on ΦC31 integrase *att* sites. Both pHLM12 and pHLM13 were subject to partial cleavage by TΦ4 and TΦ8, again with TΦ4 being the more active of the two proteins (Figure 4.13).

pHLM12 and pHLM13 were also tested in a recombination assay with NM resolvase but no activity was observed on either substrate plasmid.

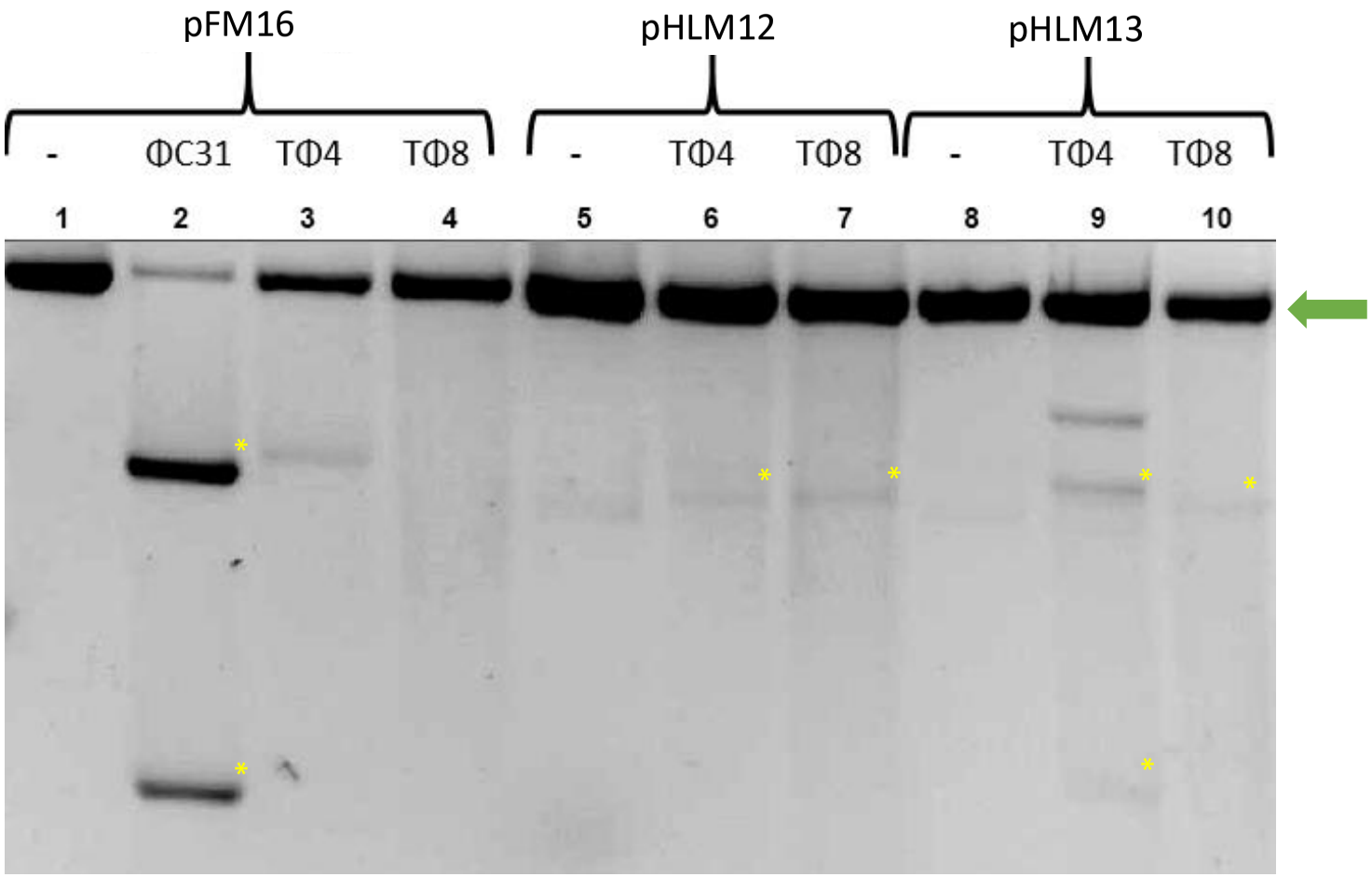

Figure 4.12: Recombination assay results for TΦ4 and TΦ8.

IDB was used as a negative control in lanes 1, 5 and 8. A ΦC31 integrase positive control is in lane 2. Recombination products identified in lanes 2, 6, 7, 9 and 10 are marked with a yellow asterisk. Unrecombined products are indicated by a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

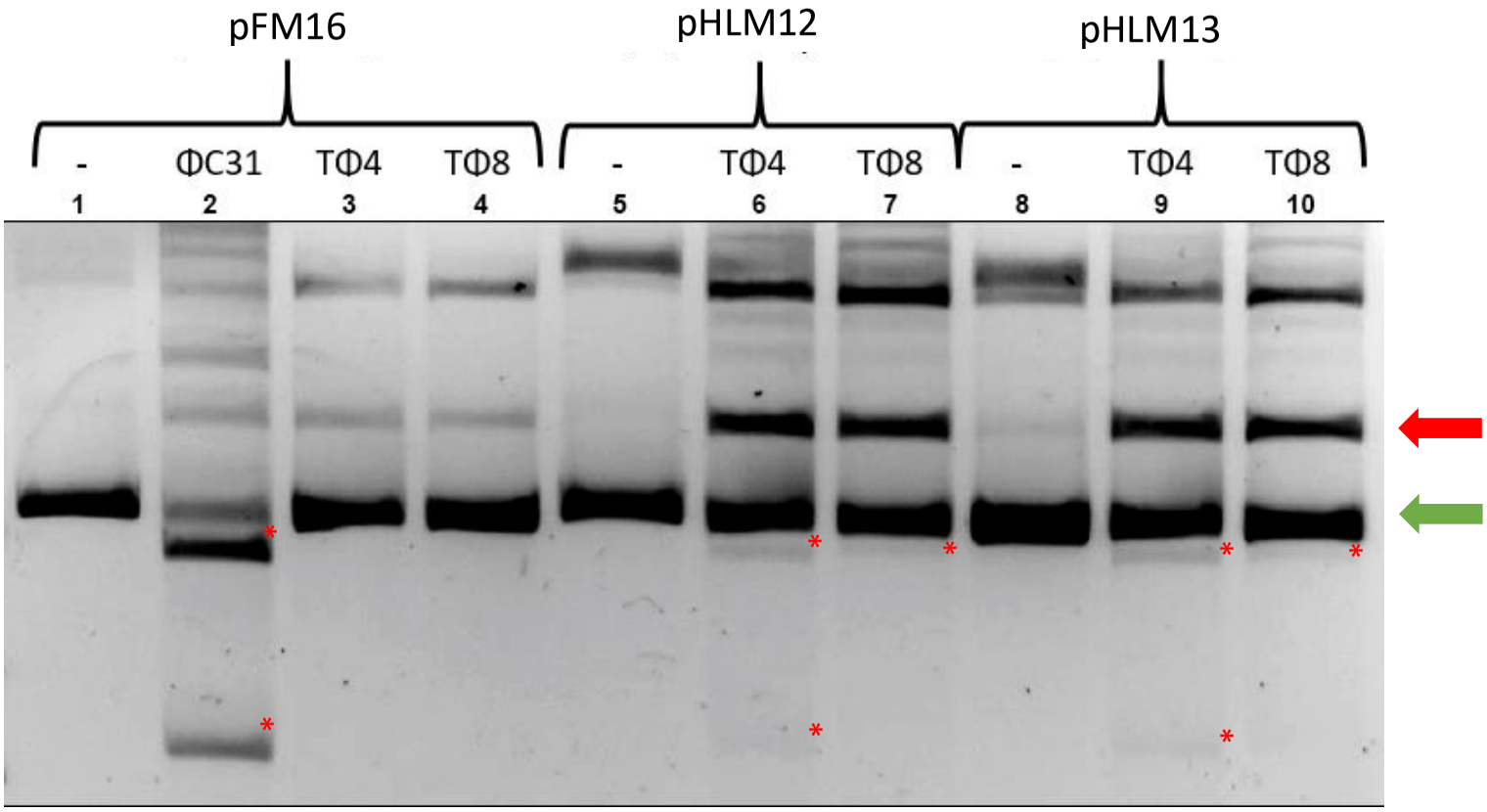

Figure 4.13: Ethylene glycol cleavage assay results for TΦ4 and TΦ8.

IDB was used as a negative control in lanes 1, 5 and 8. A ΦC31 integrase positive control is in lane 2. Two-site cleavage products identified in lanes 2, 6, 7, 9 and 10 are marked with a red asterisk. A red arrow marks sing-site cleavage products. A green arrow identifies non-cleaved plasmid. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

4.6 Hybrid recombinase and Tn3 resolvase activity *in vitro*

Following the results obtained in Section 4.5, several experiments were conducted to explore the ability of the hybrid integrases TΦ4 and TΦ8 to interact with one of their parental proteins, NM resolvase, and whether the resultant interaction would be capable of exhibiting any activity on two-site plasmids containing one hybrid recognition site e.g. ΦtΦ14 *attB* and one copy Tn3 *res* site I. These experiments were based on the hypothesis that NM resolvase would be capable of forming a synaptic complex with either hybrid integrase, wherein NM resolvase would be responsible for recombining Tn3 *res* site I, and then either TΦ4 or TΦ8 would be responsible for recombining a hybrid *attB* site.

4.6.i Hybrid site construction

New *in vitro* two-site Tn3 *res* site I/hybrid *att* site plasmids were created following the same cloning protocol described in Section 3.3.ii, using fragments from pHLM2 and pHLM4 (see Figure 4.7), the single site Tn3 *res* site I plasmid pCO1 and pUC71K. This resulted in pHLM22 (Tn3 *res* site I/ΦtΦ14 *attB*) and pHLM24 (Tn3 *res* site I/ΦtΦ12 *attB*) (Figure 4.14). As before, all two-site, combination plasmids were constructed to give the most informative NruI restriction digest (Figure 3.9).

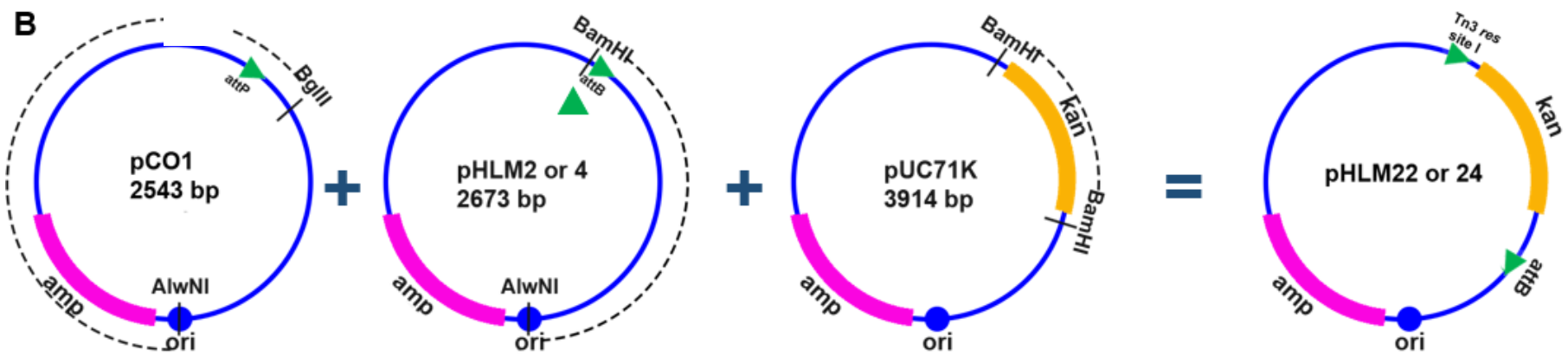

Figure 4.14: Construction of *in vitro* two-site substrate plasmids.

The plasmid pUC71K was digested with BamHI and the smaller fragment taken to be ligated together with the larger fragment of an AlwNI and BglII digested pCO1 (Tn3 *res* site I-containing plasmid) and the smaller fragment from an AlwNI and BamHI digested *attB*-containing plasmid pHLM2 or pHLM4. The resultant *in vitro* two-site substrate plasmids were named pHLM22 and pHLM24. In all instances, the fragment taken to create the final plasmids is denoted with a dashed line.

4.6.ii Activity *in vitro*

To investigate the capability of NM resolvase to interact with either of the hybrid integrases TΦ4 or TΦ8, a recombination assay (Section 2.24.i) was undertaken with the plasmids pHLM22 (Tn3 *res* site I/ΦTΦ14 *attB*) and pHLM24 (Tn3 *res* site I/ΦtΦ12 *attB*). NM resolvase acting on the Tn3 *res* site-containing plasmid pMP78 was used as a positive control. As before in Chapter 3 and Section 4.5, recombination assay samples were digested with the restriction enzyme NruI to create linearised products to be run on a 1.2% agarose gel.

The results suggested that neither TΦ4 nor TΦ8 on their own are capable of recombining pHLM22 or pHLM24, nor could NM resolvase (Figure 4.15). However, NM resolvase was able to recombine copies of Tn3 *res* site I on separate plasmids to form intermolecular rather than intramolecular recombination products, which run at double the size of standard, intramolecular recombination products (peach arrows in Figure 4.15). NM resolvase is also noted to do this when presented with pMP78, which contains two copies of Tn3 *res* site I. In this instance, NM resolvase recombines the two sites found on pMP78 as well as recombining sites on separate plasmids, resulting in the banding pattern seen in lane 2 in Figure 4.15.

From this, it was observed that NM resolvase was capable of interacting with TΦ4 and TΦ8, which then successfully recombine Tn3 *res* site I with either ΦtΦ14 *attB* or ΦtΦ12 *attB* (blue arrows in Figure 4.15). However, NM resolvase also continued to form intermolecular recombination products between copies of Tn3 *res* site I on separate copies of the plasmids pHLM22 and pHLM24 being tested (Figure 4.15).

To identify whether there was an ideal ratio of NM resolvase to hybrid integrase that would result in maximum intramolecular recombination and minimal intermolecular Tn3 *res* site I recombination, a recombination assay was set up, using increasing concentrations of hybrid integrase to NM resolvase. The reaction was left for sixteen hours, as before. However, the results showed that intermolecular recombination occurs regardless of the ratio of NM resolvase to hybrid protein (Figures 4.16 and 4.17). A ratio of 1:1 to 1:2 of NM resolvase to

hybrid protein resulted in optimal intramolecular recombination (Figures 4.16 and Figure 4.17).

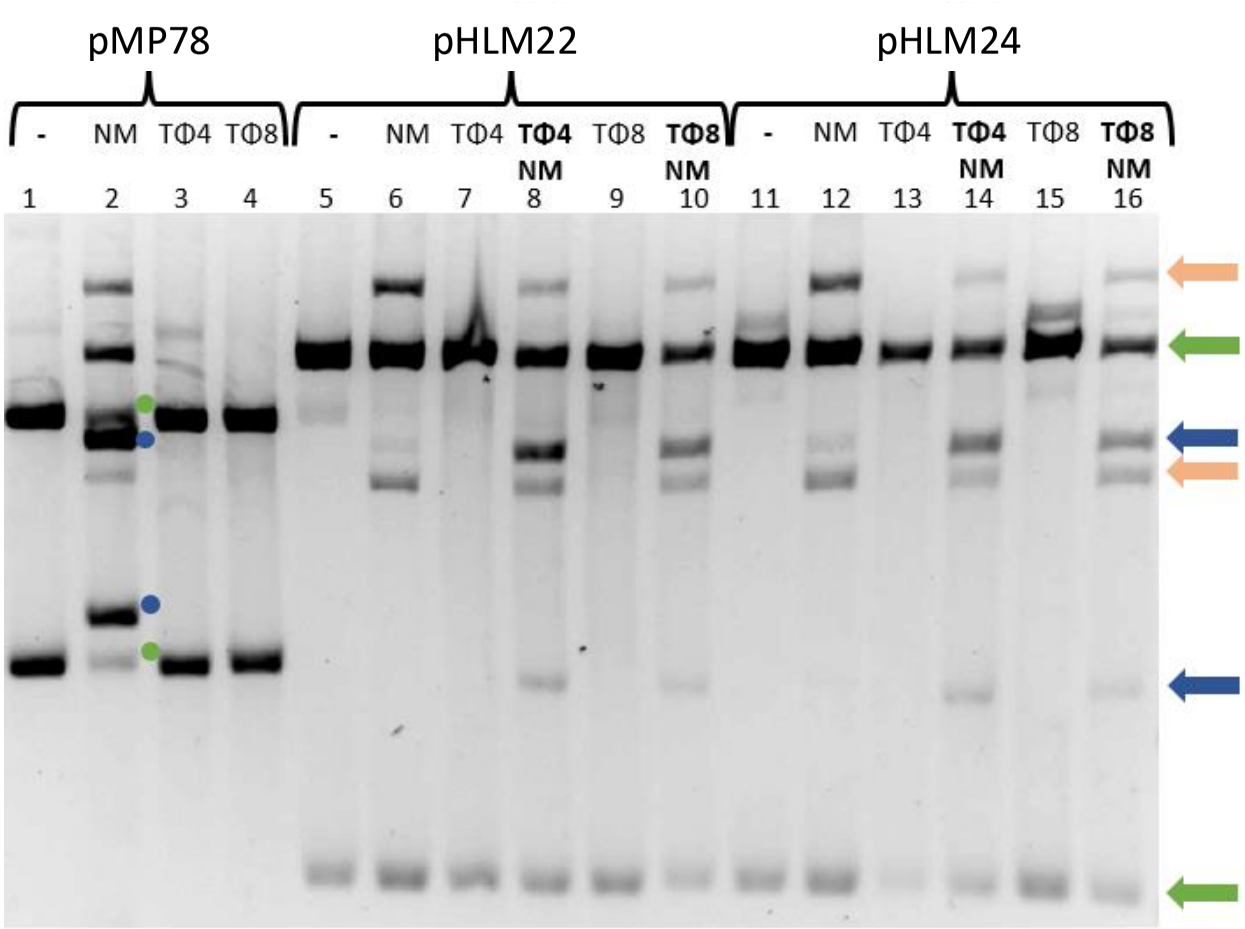

Figure 4.15: Recombination assay results for NM resolvase, TΦ4, TΦ8 and NM/hybrid protein interactions.

Blue arrows and dots denote the correct size for intramolecular recombination products whilst peach arrows denote intermolecular recombination products created through the recombination of two copies of Tn3 *res* site I on separate plasmids. Green arrows and dots denote non-recombined digest products. IDB was used as a negative control in lanes 1, 5 and 11 whilst lane 2 is a positive recombination control (NM resolvase acting on the Tn3 *res* site I two-site plasmid pMP78). 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

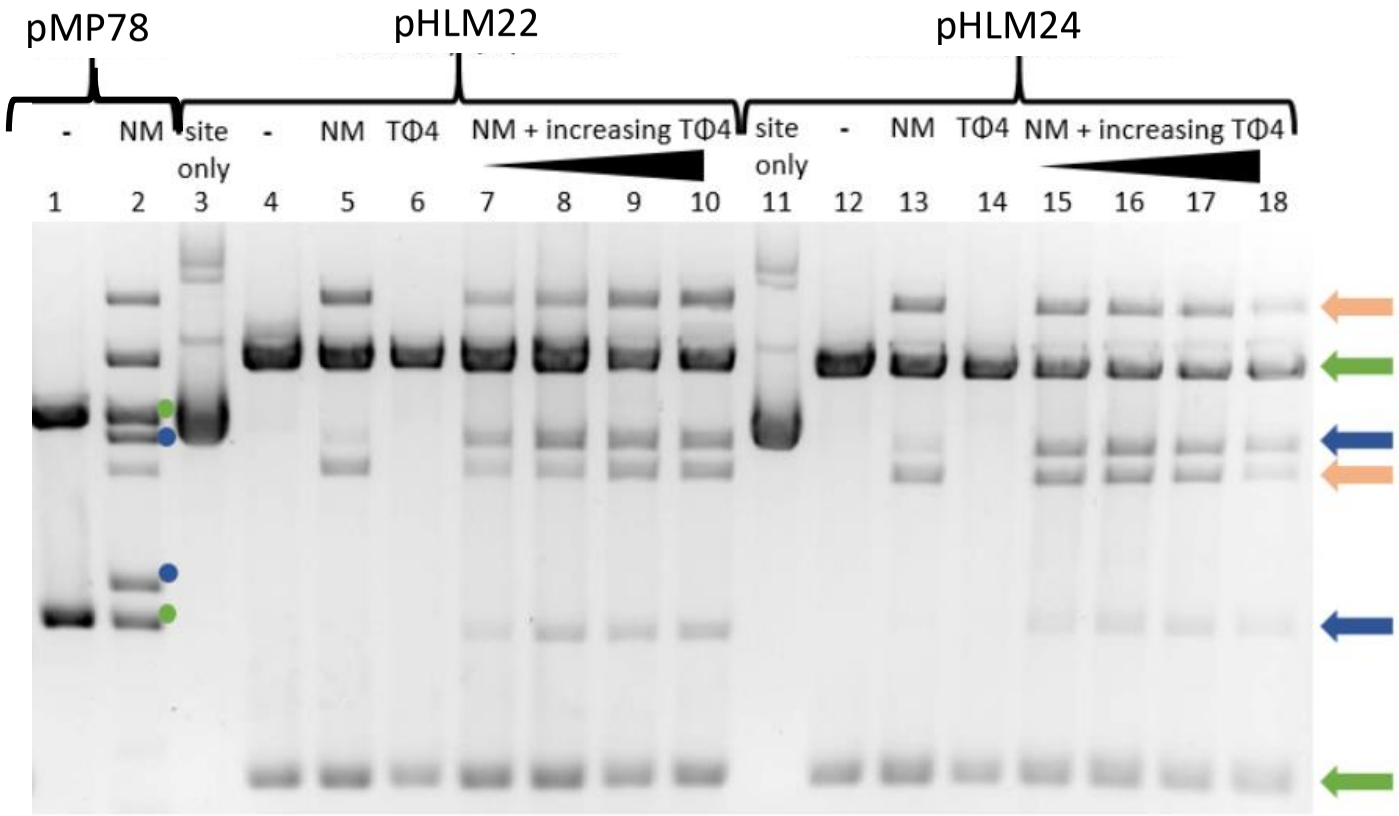

Figure 4.16: Recombination assay results for NM resolvase, TΦ4 and NM/hybrid protein interactions.

Blue arrows and dots denote the correct size for intramolecular recombination products whilst peach arrows denote intermolecular recombination products created through the recombination of two copies Tn3 *res* site I on separate plasmids. Green arrows and dots denote non-recombined digest products. NM resolvase to TΦ4 ratios are 1:1, 1:2, 1:4 and 1:8 in lanes 7, 8, 9 & 10 and lanes 15, 16, 17 & 18 respectively. IDB was used as a negative control in lanes 1, 4 and 12 whilst lane 2 is a positive recombination control (NM resolvase acting on the Tn3 *res* site I two-site plasmid pMP78). 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

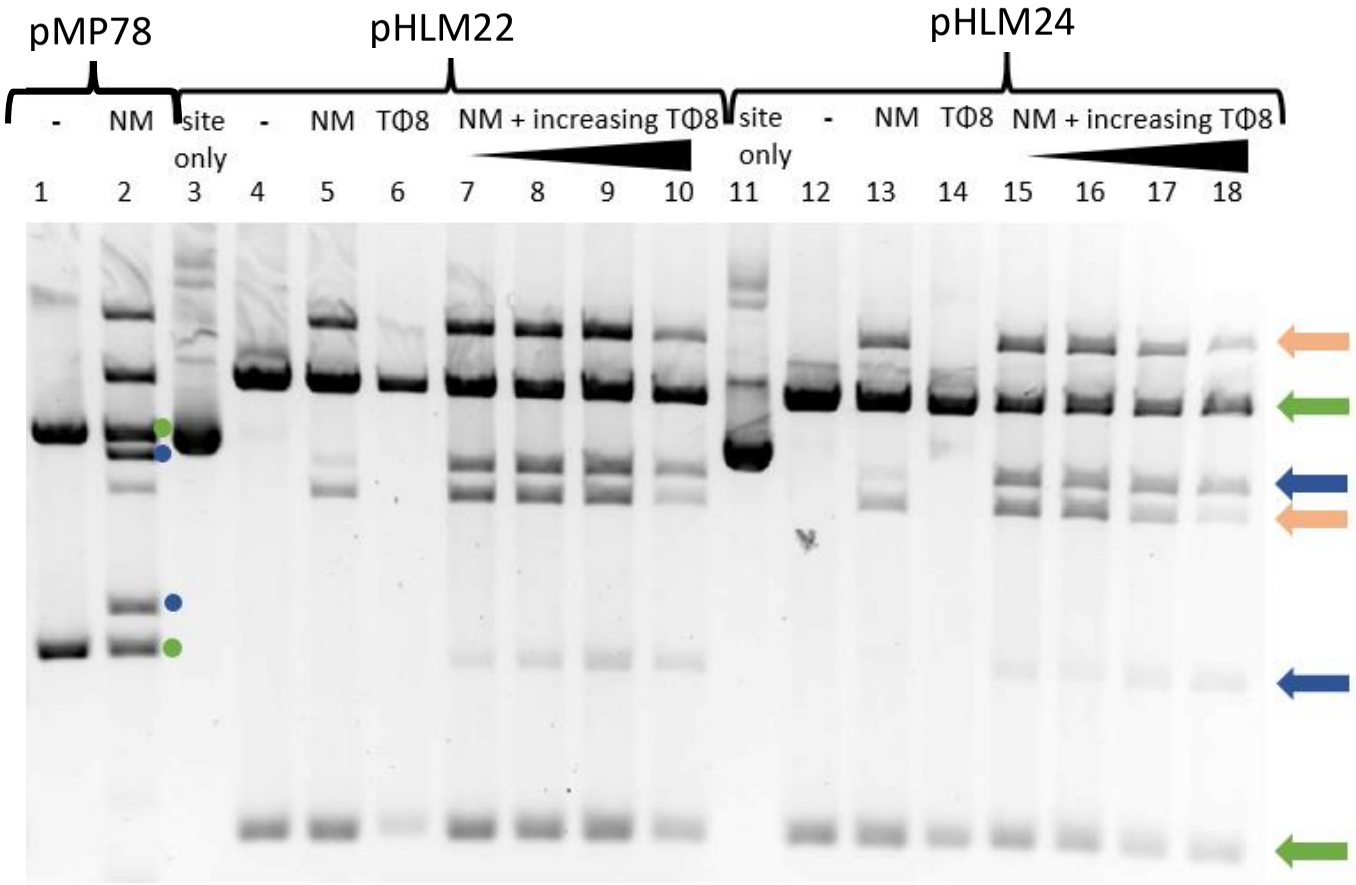

Figure 4.16: Recombination assay results for NM resolvase, TΦ8 and of NM/hybrid protein interactions.

Blue arrows and dots denote the correct size for intramolecular recombination products whilst peach arrows denote intermolecular recombination products created through the recombination of two copies Tn3 *res* site I on separate plasmids. Green arrows and dots denote non-recombined digest products. NM resolvase to TΦ8 ratios are 1:1, 1:2, 1:4 and 1:8 in lanes 7, 8, 9 & 10 and lanes 15, 16, 17 & 18 respectively. IDB was used as a negative control in lanes 1, 4 and 12 whilst lane 2 is a positive recombination control (NM resolvase acting on the Tn3 *res* site I two-site plasmid pMP78). 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

4.7 Discussion

In this chapter, two hybrid serine integrases were tested for their activity *in vitro* and *in vivo* on a variety of recognition sites, including sites created specifically for them. These new integrases consisted of an N-terminal Tn3 resolvase catalytic domain, a C-terminal ΦC31 integrase recombinase domain and a variable length oligonucleotide linker to connect the two domains together; they were named TΦ4 and TΦ8 to denote the short linker (TΦ4) and the longer linker (TΦ8) used to create them. Both sets of new recognition sites were based on ΦC31 integrase *att* sites, with either the central 14 or 12 bp in the site replaced with the central 14 or 12 bp from Tn3 *res* site I, resulting in sites named ΦtΦ14 *attP* and *attB* and ΦtΦ12 *attP* and *attB*. TΦ4 and TΦ8 were also tested for their ability to interact with the hyperactive Tn3 resolvase mutant, NM resolvase, to recombine different *att* site combinations *in vitro*.

Hybrid integrase overexpression and purification

TΦ4 and TΦ8 were both easily induced and overexpressed in the cell line BL21(DE3)[pLysS] and subsequently His-tag column purified; that either hybrid integrase was overexpressed to a level sufficient for protein purification is a success compared to the hybrid integrases created in Chapter 3, none of which were successfully purified. It demonstrates that a protein structure consisting of domains from Tn3 resolvase and ΦC31 integrase does not have the same hypothesised solubility and/or misfolding issues that a hybrid protein consisting of BxbI integrase and ΦC31 integrase domains has (see Sections 3.9 and 3.10). This may be due to the smaller size of Tn3 resolvase compared to BxbI integrase allowing for fewer opportunities for protein misfolding. Alternatively, it may be that the Tn3 resolvase protein folds in a manner far more similar to ΦC31 integrase than BxbI integrase, meaning that a hybrid between the two proteins is more likely to fold in the same way. Furthermore, inductions for TΦ4 and TΦ8 worked as well as for control inductions used to overexpress ΦC31 integrase, indicating that both hybrid proteins are no more toxic to the cell line than non-hybrid proteins. It is also possible that TΦ4 and TΦ8 may simply be

naturally more soluble than BΦ1, BΦ2 and ΦB1, thus allowing them to be purified following the protocols outlined in Section 2.22.

Activity of the hybrid integrases TΦ4 and TΦ8

Both the *in vivo* MacConkey assay and the *in vitro* recombination assay revealed the ability of the hybrid integrases TΦ4 and TΦ8 to recombine the hybrid *att* sites ΦtΦ14 and 12 *attP* and *attB*, though the overall percentage of substrate plasmids recombined was low. The extent of cleavage *in vitro* on both sets of hybrid sites by TΦ4 and TΦ8 was greater than the extent of recombination.

Despite the low activity demonstrated by TΦ4 and TΦ8 on hybrid sites, the results are encouraging; neither hybrid protein cleaved or recombined the cognate sites of the parental integrases Tn3 resolvase and ΦC31 integrase but they did cleave and recombine the sites created for them, thus providing proof of concept that functional *in vitro* hybrid serine integrases are possible. From these results, it is clear that using an N-terminal NM resolvase and C-terminal ΦC31 integrase construct produces viable hybrid integrases, that cleave and recombine *att* sites wherein the central basepairs follow the DNA sequence of the N-terminal protein's natural recognition site i.e. Tn3 *res* site I, and the rest of the basepairs in the site follow the DNA sequence of the C-terminal protein's recognition site i.e. ΦC31 *attP* and *attB*. This provides credence to the hypothesis made based on the LI integrase crystal structure: that it is likely the region around the αE helix is responsible for cleavage and recombination of a serine integrase *att* site, and that many (but not all) of the basepairs in the *att* site critical for cleavage and recombination lie within the centre 14 bp of the site itself.

It may well be that altering the ratio of Tn3 resolvase to ΦC31 integrase in a hybrid integrase construct will increase the level of activity exhibited on hybrid sites. Increasing the number of basepairs taken from Tn3 *res* site I to be put into the hybrid *att* sites may also help with increasing the proficiency with

which a Tn3 resolvase/ΦC31 integrase hybrid protein can cleave and recombine these sites.

Activity of Tn3 (NM) resolvase as a dimer with TΦ4 and TΦ8

NM resolvase had no observable activity on either hybrid site ΦtΦ14 or 12 *attP* and *attB* on its own *in vitro* or *in vivo*, which was to be expected; this is likely due to the protein lacking the ability to bind to the hybrid sites, since the hybrid sites were created with a ΦC31 integrase recombinase domain in mind, rather than the DNA-binding domain of Tn3 resolvase.

However, NM resolvase was capable of significant recombination *in vitro* on pHLM22 (Tn3 *res* site I/ΦtΦ14 *attB*) and pHLM24 (Tn3 *res* site I/ΦtΦ12 *attB*) substrate plasmids when either TΦ4 or TΦ8 were also present. The most probable explanation for this is that NM resolvase forms a dimer with itself and binds to the Tn3 *res* site I found in the substrate plasmid, whilst TΦ4/8 does the same but binds to the ΦtΦ14 or 12 *attB* site in the plasmid. Both dimers would then come together to form a synaptic complex and follow through with site-specific recombination as described in the mechanism for large serine integrase recombination in Figures 1.7 and 1.16 (Figure 4.17). This would also explain how intermolecular recombination was observed for *in vitro* recombination reactions; the NM resolvase dimers would be able to form synaptic complexes with each other across two substrate plasmids.

It is clear from the recombination assay results that, regardless of how NM resolvase and either TΦ4 or TΦ8 interact with each other, the presence of NM resolvase facilitates the activity of the two hybrid integrases on the recognition sites tested. However, the intermolecular recombination observed is problematic; it limits the usage of such a protein and recognition site combination in applicable contexts, since the system is neither independently-functioning nor controllable because no way to prevent recombination between plasmids has been identified thus far. Indeed, rather than attempting to prevent this from occurring, time would be better spent trying to alter the

hybrid integrase domain composition to make the new proteins more active on hybrid sites by themselves, without the presence of NM resolvase.

It would be useful to perform an *in vivo* MacConkey assay using the same protein and substrate combinations and subsequently sequencing any resolution products observed. This would provide information regarding how effectively or "properly" the Tn3 *res* site I/ΦtΦ14 *attB* or Tn3 *res* site I/ΦtΦ12 *attB* site combinations were resolved, as well as highlighting the ratio of Tn3 *res* site I intermolecularly-recombined products to Tn3 *res* site I/ ΦTΦ14 or 12 *attB* site intramolecularly-recombined products present by viewing the products on an agarose gel, as described in Section 4.4.

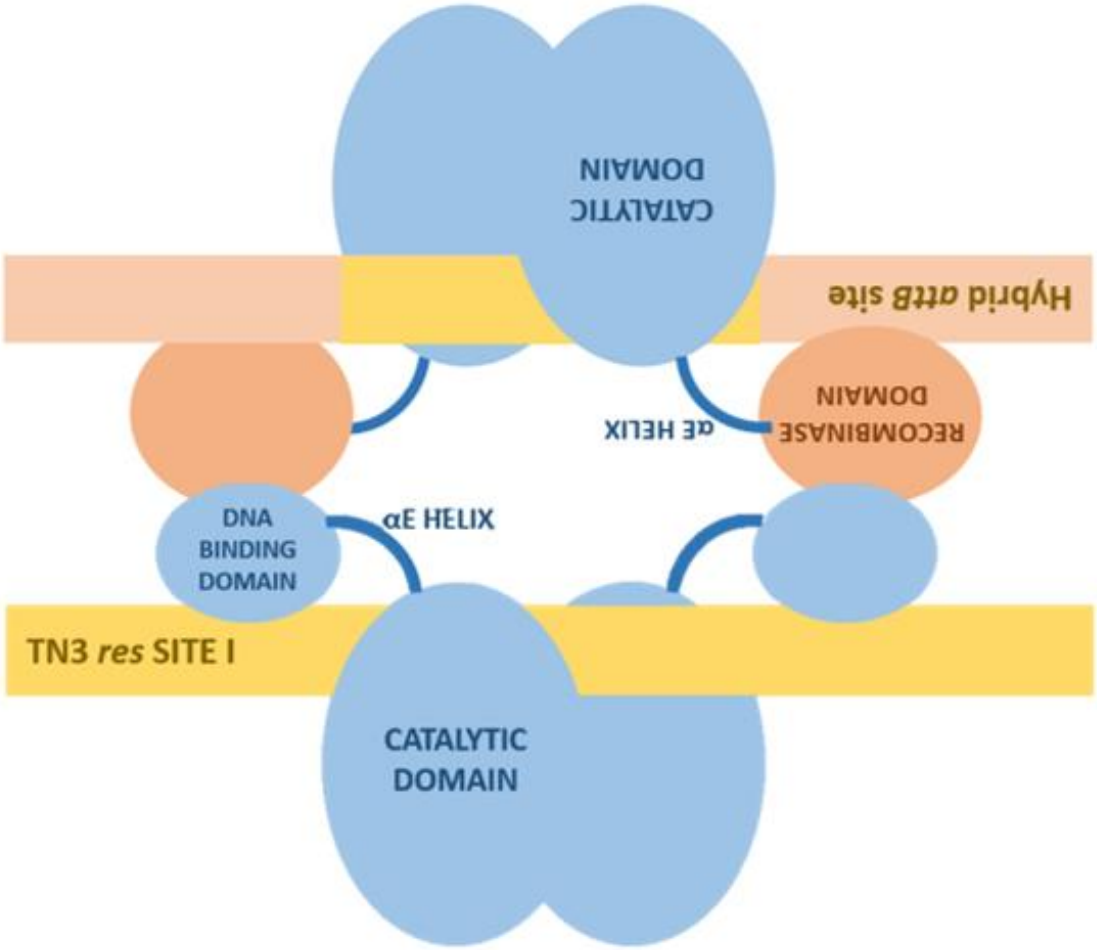

Figure 4.17: NM resolvase/hybrid integrase proposed synaptic complex.

NM resolvase is shown as a blue dimer bound to a pale orange Tn3 *res* site I, whilst TΦ4/TΦ8 is shown as a blue and pale red dimer bound to a peach and pale orange *attB* site. The two dimers come together to form a synaptic complex but are responsible for cleaving and recombining their own recognition site only, following a similar site-specific recombination pathway to other serine integrases.

Chapter Five: ΦC31 integrase recognition site variability

5.1 Introduction

Based on some of the results obtained for hybrid *att* sites throughout this project, further investigation was undertaken into the ability of ΦC31 integrase to recognise different *att* site combinations.

Because ΦC31 integrase fully resolved both hybrid plasmids pHLM7 (ΦtΦ14 *attP*/*B*) and pHLM8 (ΦtΦ12 *attP*/*B*) *in vivo*, an investigation into which *att* sites ΦC31 integrase would recognise was undertaken. As well as testing both sets of hybrid sites (pHLM12 and pHLM13) *in vitro* with ΦC31 integrase, two combination *att* site plasmids were also tested - ΦtΦ12 *attP*/ΦC31 *attB* and ΦC31 attP/ΦtΦ12 *attB*.

5.2 Substrate plasmid construction

Two-site *in vitro* substrate plasmids were created exactly as described in Section 3.3.ii, creating pHLM33 (ΦtΦ12 *attP*/ΦC31 *attB*) from the ΦtΦ12 *attP*-containing plasmid pHLM3, the ΦC31 *attB*-containing plasmid pFM12 and pUC71K, and pHLM34 (ΦC31 attP/ΦtΦ12 *attB*) from the ΦtΦ12 *attB*-containing plasmid pHLM4, the ΦC31 *attP*-containing plasmid pFM14 and pUC71K (Figure 5.1). As before, the two, new plasmids were constructed to give the most informative NruI digest (Figure 3.9).

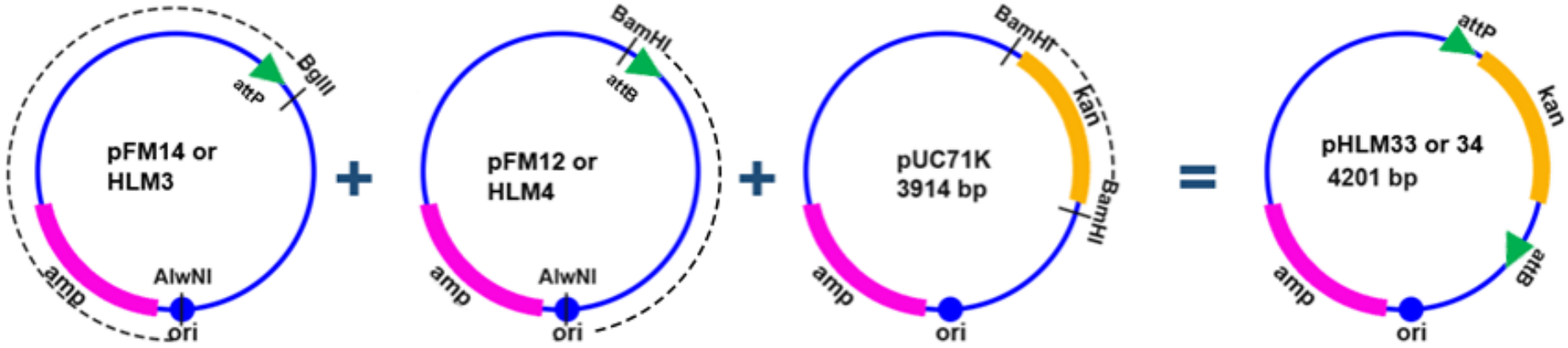

Figure 5.1: Construction of *in vitro* two-site substrate plasmids.

The plasmid pUC71K was digested with BamHI and the smaller fragment taken to be ligated together with the larger fragment of an AlwNI and BglII digested *attP*-containing plasmid pFM14 or pHLM3 and the smaller fragment from an AlwNI and BamHI digested *attB*-containing plasmid pFM12 pHLM4. The resultant *in vitro* two-site substrate plasmids were named pHLM33 and pHLM34. In all instances, the fragment taken to create the final plasmids is denoted with a dashed line.

5.3 ΦC31 integrase activity *in vitro* and *in vivo*

To test the activity of ΦC31 integrase on pHLM12 (ΦtΦ14 *attP*/*B*) and pHLM13 (ΦtΦ12 *attP*/*B*), a recombination assay was undertaken (Section 2.25.i). Reactions were left for sixteen hours, after which reactions were digested with NruI and run on a 1.2% agarose gel. The ΦC31 integrase *attP*/*B*-containing plasmid pFM16 was used as a positive control. The results show that ΦC31 integrase exhibits a low level of activity on both sets of hybrid sites, though not as strongly as might have been expected given the *in vivo* MacConkey assay results (Figure 5.2).

The two-site substrate plasmids were also treated with ΦC31 integrase in a cleavage assay (Section 2.25.ii). The results show that ΦC31 integrase promotes cleavage of both substrates. Cleavage of these sites was comparatively favourable with the level of cleavage exhibited on natural ΦC31 integrase *att* sites (Figure 5.3).

To complement these results, further recombination and ethylene glycol cleavage assays were undertaken in which the two-site substrate plasmids pHLM12 and pHLM13 were replaced with their *in vivo* two-site substrate plasmid counterparts, pHLM7 and pHLM8 respectively (Table 2.5 and Section 4.3.ii). The assays were performed as described previously (Section 2.25). Because pHLM7 and pHLM8 also contain NruI restriction sites in similar positions to pHLM12 and pHLM13, it is possible to directly compare recombination and cleavage products between the two. The results from these experiments suggest that ΦC31 integrase is capable of cleavage and recombination of both plasmids, though not as well as was demonstrated for the *in vitro* substrate plasmids pHLM12 and pHLM13 (Figure 5.4 and Figure 5.5).

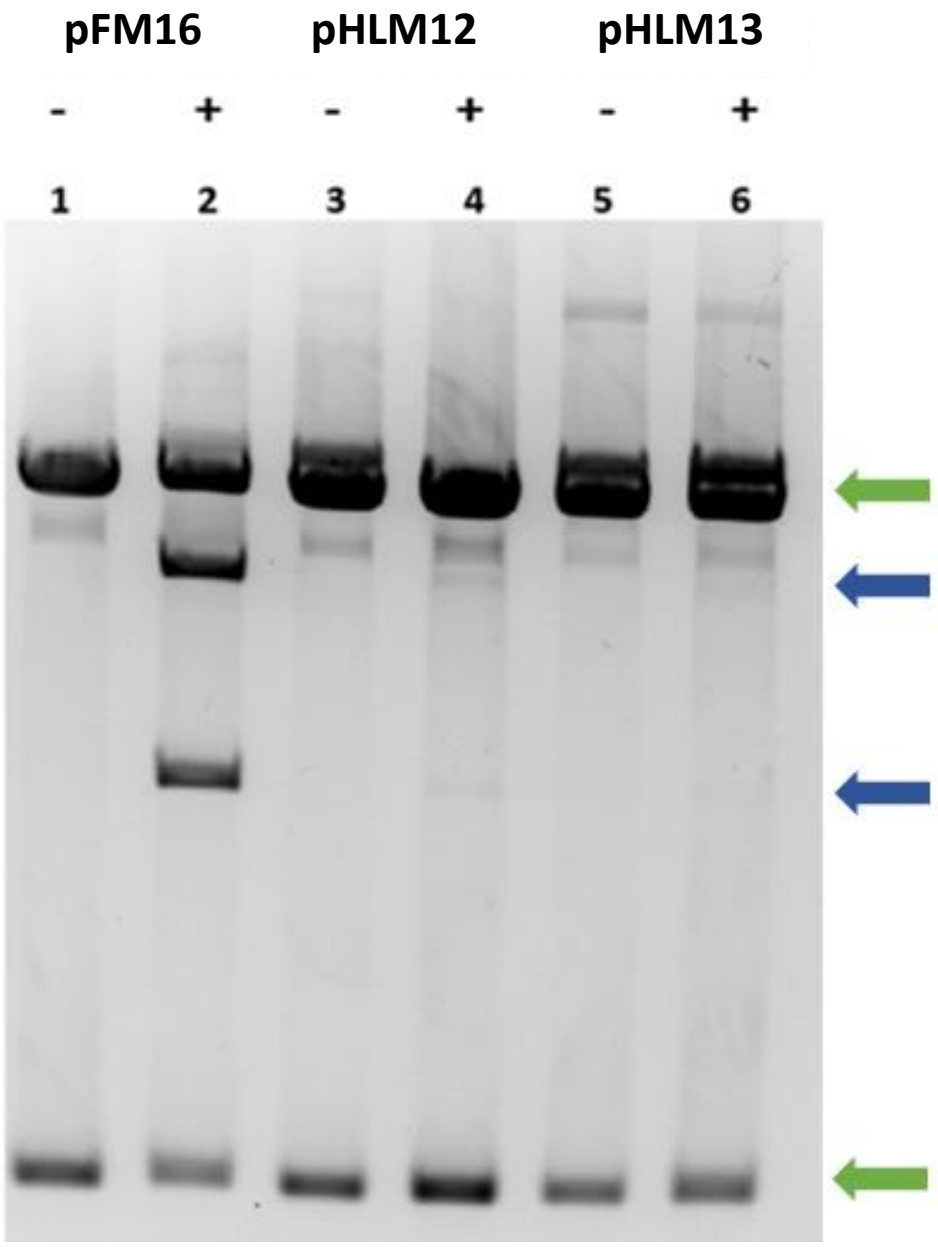

Figure 5.2: Recombination assay results for ΦC31 integrase on hybrid two-site *in vitro* substrate plasmids.

IDB used as a negative control in lanes 1, 3 and 6. ΦC31 integrase positive control is in lane 2. Recombination products identified in lanes 2, 4 and 6 are marked with a blue arrow. Unrecombined products are marked by a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

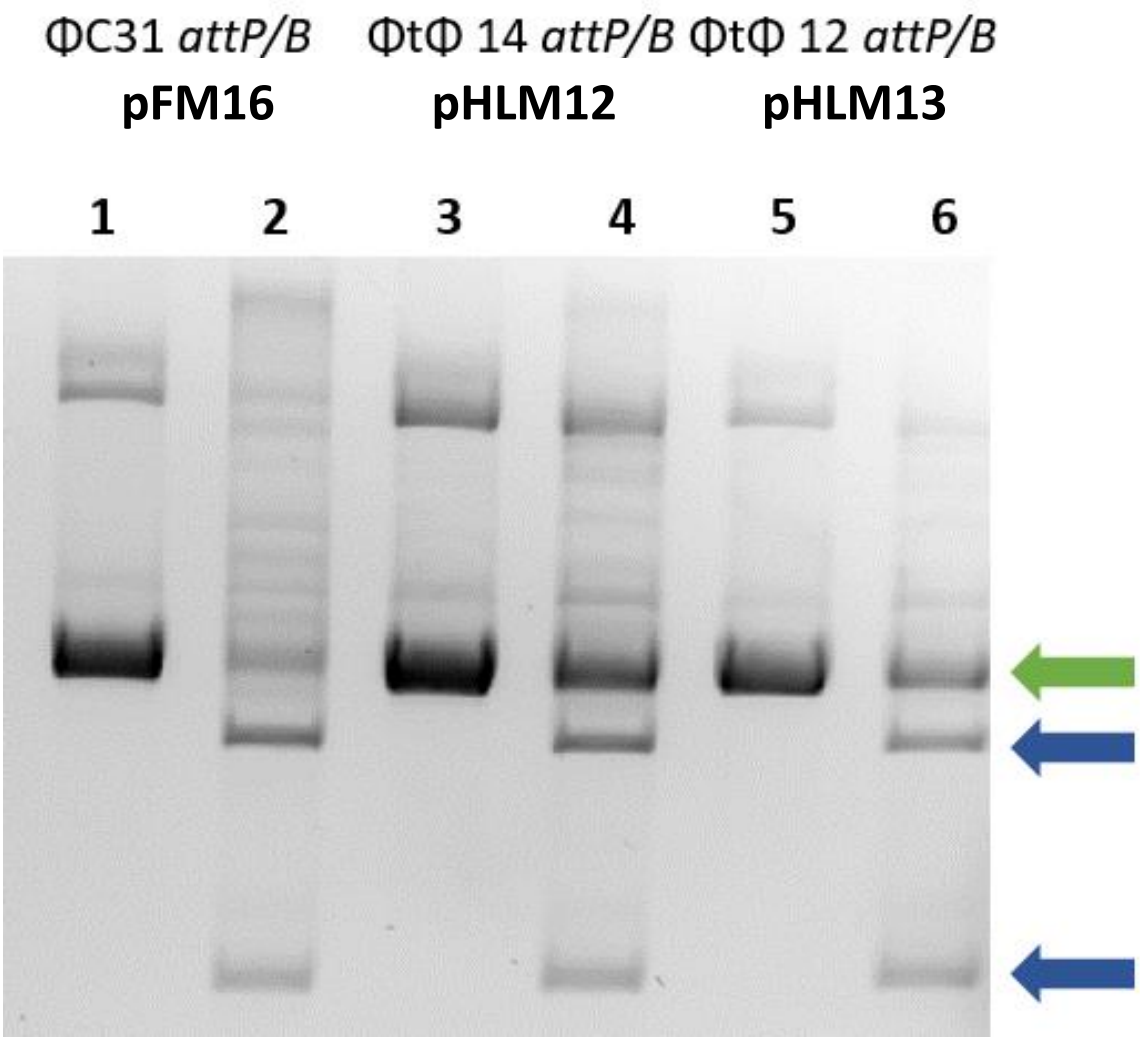

Figure 5.3: Ethylene glycol cleavage results for ΦC31 integrase on hybrid two-site *in vitro* substrate plasmids.

IDB used as a negative control in lanes 1, 3 and 6. ΦC31 integrase positive control is in lane 2. Cleavage products identified in lanes 2, 4 and 6 are marked with a blue arrow. Uncleaved products are marked by a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

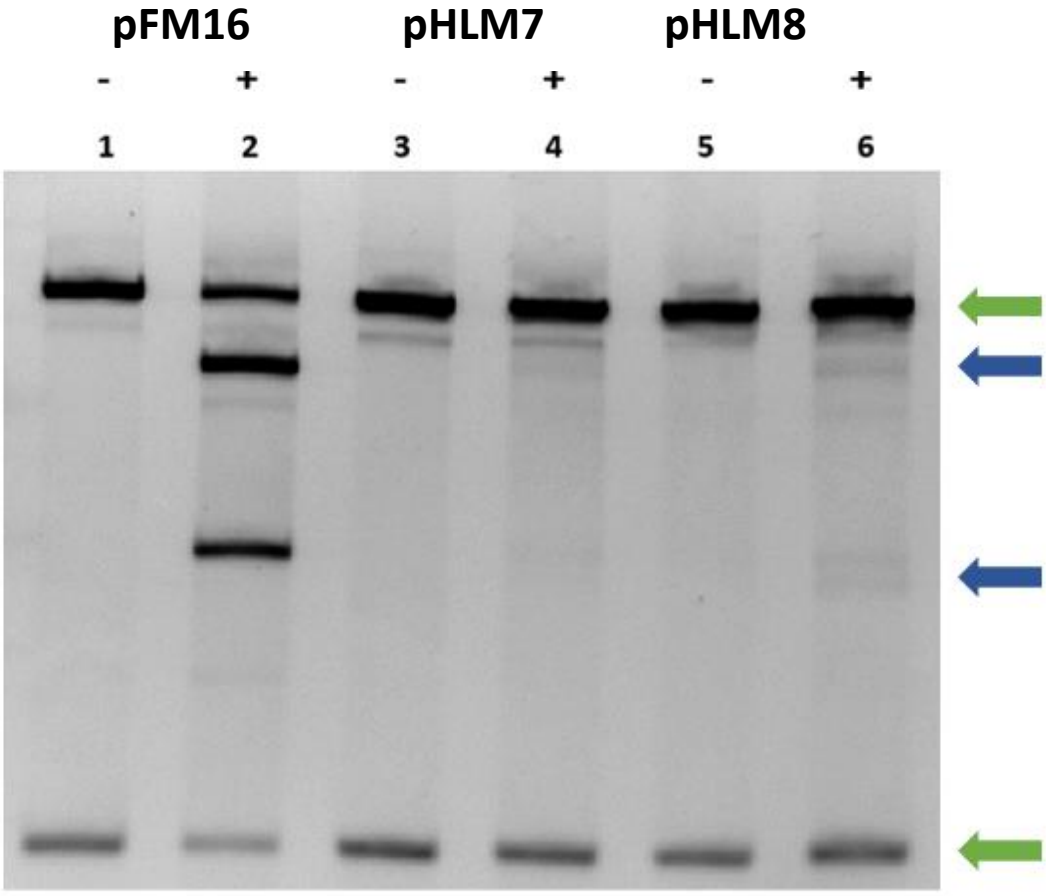

Figure 5.4: Recombination assay results for ΦC31 integrase on hybrid two-site *in vivo* substrate plasmids.

IDB used as a negative control in lanes 1, 3 and 6. ΦC31 integrase positive control is lane 2. Recombination products identified in lanes 2, 4 and 6 are marked with a blue arrow. Unrecombined products are marked by a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

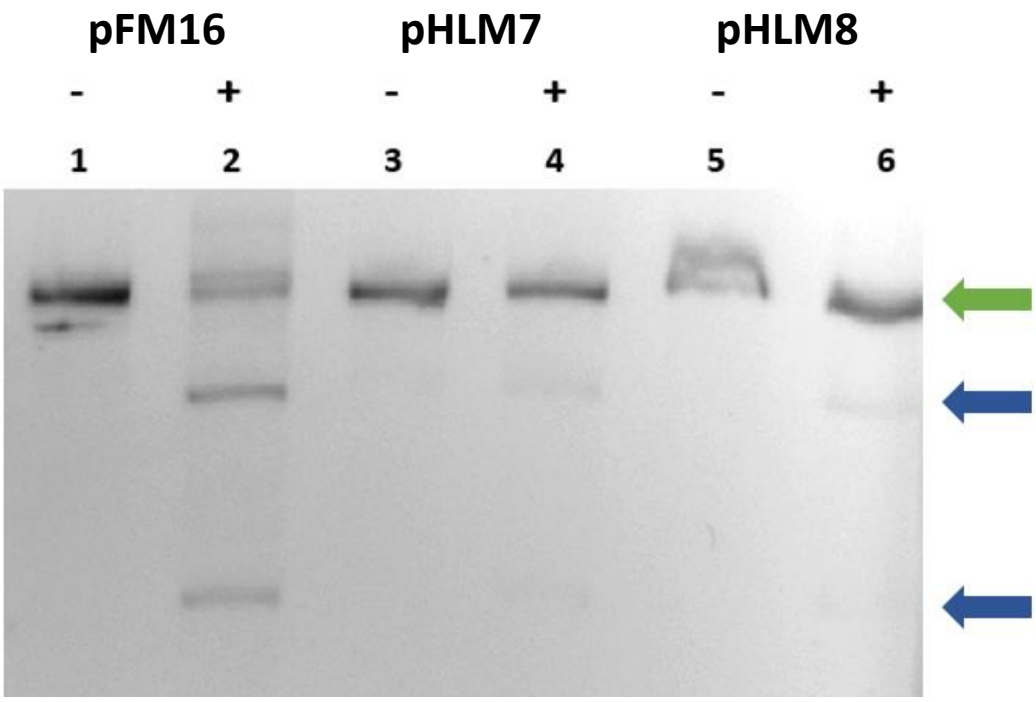

Figure 5.5: Ethylene glycol cleavage results for ΦC31 integrase on hybrid two-site *in vivo* substrate plasmids.

IDB used as a negative control in lanes 1, 3 and 6. ΦC31 integrase positive control is in lane 2. Cleavage products identified in lanes 2, 4 and 6 are marked with a blue arrow. Uncleaved products are marked by a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

To test whether ΦC31 integrase was capable of activity on a combination of its own *att* sites and hybrid *att* sites, a recombination assay and an ethylene glycol cleavage assay (Section 2.25) were subsequently performed using the plasmids pHLM33 (ΦtΦ12 *attP*/ΦC31 attB) and pHLM34 (ΦC31 attP/ΦtΦ12 *attB*). The results of the recombination assay showed no recombination of either combination substrate plasmid even after sixteen hours; this was to be expected, given that the two sites within the plasmids tested do not share the same crossover dinucleotide sequence.

However, cleavage assay results demonstrated cleavage of both plasmids after sixteen hours. A time course was then performed following the same ethylene glycol cleavage protocols to determine at which point ΦC31 integrase cleaves either two-site substrate plasmid. The results from this experiment suggest that, whilst a low level of cleavage is observed after one and four hours, most of the site cleavage occurs between four and sixteen hours (Figure 5.6).

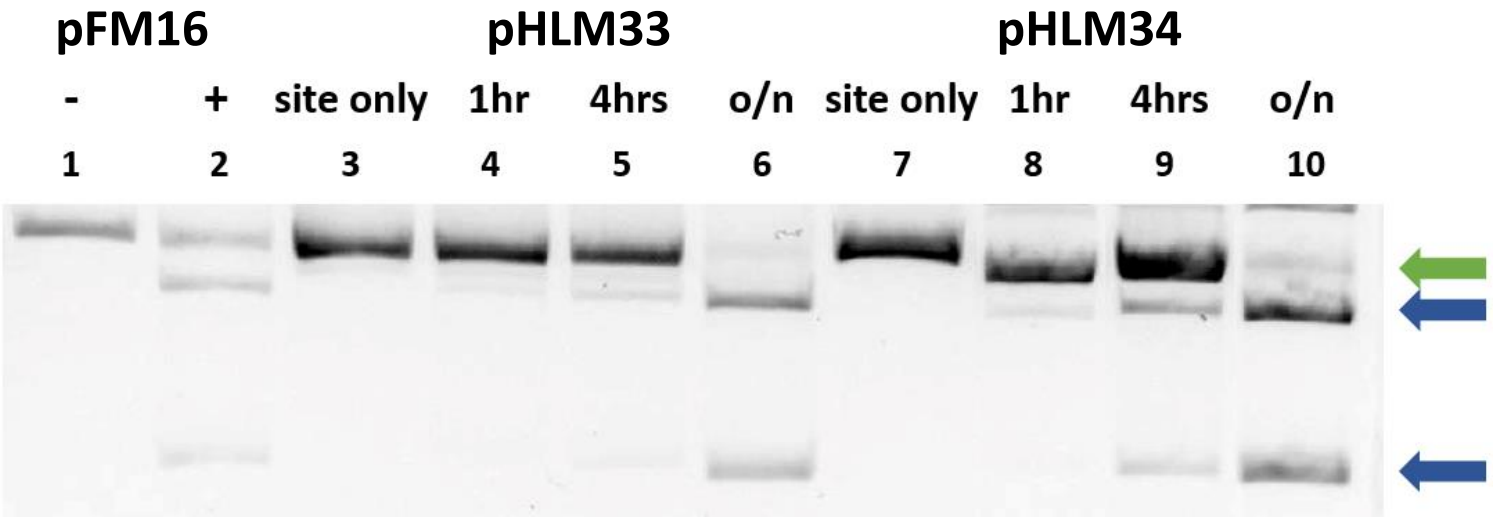

Figure 5.6: Ethylene glycol cleavage results for ΦC31 integrase on two-site substrate plasmids.

IDB used as a negative control in lane 1. ΦC31 integrase positive control is in lane 2. Cleavage products identified in lanes 2, 4, 5, 6, 8, 9 and 10 are marked with a blue arrow. Uncleaved products are marked with a green arrow. 1.2% agarose gel run at 125 V for 90 minutes, followed by 30 minutes of staining in ethidium bromide and 30 minutes of destaining in deionised water. Gel was visualised using a short-wavelength UV transilluminator.

5.4 Discussion

In this chapter, the ability of ΦC31 integrase to recombine and cleave different *att* sites *in vitro* was explored. Both recombination and ethylene glycol cleavage assays were employed to investigate the activity of the protein.

It was observed that ΦC31 integrase was capable of partial recombination of both hybrid *att* sites ΦtΦ14 and ΦtΦ12 *in vitro* and complete resolution of both hybrid *att* sites *in vivo*. These results were surprising; since the hybrid sites contained large differences in DNA sequence from the native ΦC31 integrase *att* sites, such efficient recombination, particularly *in vivo*, was not expected. ΦC31 integrase also cleaved both hybrid sites efficiently *in vitro*. Given that it was already established in Chapter 3 that ΦC31 integrase was capable of cleaving the hybrid sites ΦbΦ12 and ΦbΦ10 reasonably well – and that all of the hybrid sites tested in this project used by and large the same regions of the ΦC31 *attP* and *attB* sites sequences in their construction - this result was less surprising than the recombination results observed.

The results suggest that cleavage of recognition sites by ΦC31 integrase is not strictly related to the final formation of recombinant *attL* and *attR* sites and that cleavage of *att* sites by ΦC31 integrase is not always dependent on the crossover dinucleotide in each *att* site being TT (it is AT in ΦtΦ14 and 12, and GT in ΦbΦ12 and 10). Furthermore, the cleavage results observed *in vitro* for two-site substrate plasmids containing ΦtΦ12 *attP*/ΦC31 *attB* or ΦC31 *attP*/ΦtΦ12 *attB* indicate that the crossover dinucleotide need not even match for cleavage to occur.

This is directly contradictory to much of the current literature surrounding ΦC31 integrase; Smith et al (2004) observed that the insertion or deletion of even single basepairs into the ΦC31 *attB* site terminated the recombination process somewhere between synapsis and cleavage – only a very small amount of cleaved *attB* and the partner site *attP* was seen. The hypothesis behind this result was that even single basepair changes prevent recombination and cleavage from occurring due to spacing requirements between sequences in *attB* required to be recognised by ΦC31 integrase that subsequently correctly

positions the catalytic serine residue with the scissile phosphate (Smith et al, 2004).

Furthermore, Gupta et al (2007) investigated the effect that specific basepair changes in the ΦC31 *attB* site would have in order to create functional pseudo-*attB* sites. They discovered that, whilst some alterations to the site caused little to no change in site cleavage and recombination by ΦC31 integrase, changing positions 2, 15 or 16 on either side of the centre of the crossover dinucleotide critically interrupted cleavage and recombination. Whilst positions 15 and 16 have remained identical to ΦC31 *attB* for all hybrid sites explored in this project, position 2 on both sides of the crossover dinucleotide has been different for all hybrids tested (discussed further in Chapter 6). Specifically, Gupta et al (2007) noted that changing position 2 in *attB* prevented ΦC31 integrase from cleaving the site altogether, halting the DNA-protein complex at synapse formation. This was clearly not the case for ΦC31 integrase interacting with ΦtΦ14 and 12 *attP* and *attB*, which the protein very clearly cleaved and partially recombined, nor for ΦbΦ10 and 12 *attP* and *attB*, which ΦC31 integrase cleaved and partially recombined *in vivo* (ΦbΦ10) or cleaved only (ΦbΦ12). Since this is so contradictory to the findings of Gupta et al (2007), further experimentation into the ability of ΦC31 integrase to recombine *att* sites with altered central basepairs to those of its natural sites would be useful to help identify just how site-specific the integrase actually is.

It is possible, as for the resolution observed for the MacConkey assay performed in Chapter 3 on ΦC31 integrase/BxbI integrase hybrid sites, that cleavage on its own is sufficient to induce resolution of substrate plasmids *in vivo*, in that the sites could be cut by the protein present and then manage to religate together without the protein having to appropriately mediate recombination. However, sequencing of resolution products from the *in vivo* MacConkey assay carried out for ΦC31 integrase on ΦtΦ14 and 12 *attP* and *attB* confirmed that both sets of hybrid sites had been recombined correctly to form *attL* and *attR* products, implying that recombination was carried out efficiently and mediated by ΦC31 integrase (data not shown).

This could be explored further by investigating the ability of ΦC31 integrase to perform the reverse recombination reaction of *attLxattR* back to *attP* and *attB* in the presence of its recombination directionality factor protein.

Taken together, the results for ΦC31 integrase in this chapter, as well as in Chapters 3 and 4, suggest that the protein is far less site-specific than previously thought, particularly *in vivo*.

Chapter Six: General discussion and conclusions

This project dealt with two primary goals – the construction of a functional hybrid serine integrase, and the construction of recognition *att* sites specifically designed for said hybrid integrase. Additionally, based on the results obtained for these *att* sites, experiments looking into the nature of ΦC31 integrase and its ability to recognise and recombine different *att* sites were also undertaken.

Functional hybrid integrase structure and activity

Chapter 3 discussed the creation and attempted purification of hybrid integrases based on either an N-terminal BxbI integrase catalytic domain and a C-terminal ΦC31 integrase recombinase domain construction (named BΦ1 and BΦ2) or the opposite domain structure of a ΦC31 integrase N-terminus and a BxbI integrase C-terminus (named ΦB1). New *att* sites were also created wherein either the central 10 or 12 bp from ΦC31 *att* sites were replaced with those from BxbI *att* sites, or the central 10 bp from BxbI *att* sites were replaced with those from ΦC31 *att* sites. Whilst BΦ1 and BΦ2 showed no demonstrable activity *in vivo* and failed to overexpress to a level suitable for purification and subsequent *in vitro* experiments, ΦB1 was capable of resolving approximately one-third of total ΦC31 *att* sites *in vivo,* though purification of this protein was also unsuccessful.

The results for hybrid integrases created in Chapter 3 suggest that there may be issues with protein solubility or protein misfolding, as discussed in Section 3.10. This leads to the possibility that using the domains of dissimilar large serine integrases creates an inviable hybrid integrase. However, since ΦB1 was capable of some activity *in vivo*, it is also possible that the specific combination of N-terminal BxbI integrase/C-terminal ΦC31 integrase is ultimately what results in inviable hybrid integrases. Consequently, adjusting protein

purification protocols further in order to purify ΦB1 may yet yield some *in vitro* results for this integrase.

The hybrid integrases discussed in Chapter 4 proved altogether more promising. Both of the N-terminal Tn3/NM resolvase catalytic domain and C-terminal ΦC31 integrase recombinase domain hybrid integrases (named TΦ4 and TΦ8) demonstrated a low level of activity *in vivo* and *in vitro* on the hybrid sites ΦtΦ14 and 12 *attP* and *attB*, wherein the central 14 or 12 bp from ΦC31 *att* sites were replaced with those from Tn3 *res* site I. Furthermore, the activity of either hybrid integrase with NM resolvase on combination plasmids containing one copy of Tn3 *res* site I and one copy of either ΦtΦ14 or 12 *attB* was stronger than either hybrid integrase on its own on either ΦtΦ14 or 12 *attP* and *attB*.

Whilst the level of activity exhibited by both TΦ4 and TΦ8 *in vitro* and *in vivo* was relatively low, the very observation that they are active at all – and only on sites specifically created for them to recognise – is proof of concept that a functional hybrid serine integrase is indeed possible. If it were to be the case that only hybrid integrases created using Tn3/NM resolvase and ΦC31 integrase were functional, it would limit the number of useful proteins and sites that could be engineered, as the innermost basepairs of the hybrid *att* sites would always have to derive from Tn3 *res* site I. However, this seems unlikely.

It is far more plausible that using domains from the smaller resolvases together with domains from the large serine integrases can create further hybrid integrases. The best way to investigate this would be to construct hybrid integrases using a similar domain composition as described in Chapter 4 but altering the proteins used to construct them e.g. replace ΦC31 integrase with TG1 integrase or LI integrase, or Tn3 resolvase with γδ resolvase. One could also create hybrid integrases following the domain construction of Chapter 3 i.e. use two large serine integrases to test whether any combinations outwith N-terminus BxbI integrase/C-terminus ΦC31 integrase are functional.

Hybrid att *site recognition*

ΦC31 integrase was capable of a low level of resolution on the hybrid sites ΦbΦ10 *attP* and *attB in vivo* and managed to cleave both ΦbΦ10 and ΦbΦ12 *attP* and *attB in vitro* when left for at least four hours. ΦC31 integrase also completely resolved both ΦtΦ14 and 12 *attP* and *attB in vivo,* as well as cleaving them both and partially recombining them *in vitro*.

Since, in the instance of the hybrid sites ΦtΦ14 and 12 *attP* and *attB*, only half of the central 12 or 14 bp for *attP* and two of the central 12 or 14 bp for *attB* are identical to ΦC31 *attP* and *attB* respectively (Figure 6.1), one would reasonably expect cleavage between the two sites to be difficult. One possibility is that the basepairs that remain the same in the centre of the site are far more essential for cleavage by ΦC31 integrase than those that have been changed. The creation of further hybrid *att* sites with directed, single-basepair changes would be the most straightforward way of exploring this hypothesis, and would help in identifying which basepairs within the centre of an *attP* or *attB* are critical for cleavage by ΦC31 integrase.

More generally, since NM/Tn3 resolvase failed to recognise either ΦtΦ14 or 12 *attP* and *attB* on its own but both hybrid integrases TΦ4 and TΦ8 did, this suggests that the recognition machinery for choosing an *att* site to bind to, cleave and recombine does not lie solely within the αE helix and beginning of the recombinase domain as originally hypothesised. The *in vivo* and *in vitro* results obtained for BxbI integrase with both sets of hybrid sites ΦbΦ10 and 12 *attP* and *attB* also support this claim: they suggest that having the central 10 or 12 bp following the BxbI *att* site sequence is not enough to allow this integrase to recognise, cleave or recombine the sites.

This leads to the hypothesis that there are regions further out from the central basepairs that were explored in this project essential for recognition by their cognate integrase. This would explain why BxbI integrase does not recognise ΦbΦ10 and 12 *attP* and *attB* but ΦC31 integrase cleaves them both, as the

basepairs further out from the centre of the hybrid sites follow the sequence of ΦC31 *att* sites. Indeed, there are conserved motifs outwith the central 10 or 12 bp between the *attP* and *attB* sites for each integrase which may be the regions critically responsible for complete integrase recognition alongside the central nucleotides (see Figure 3.3).

As discussed in Section 3.10, the results for ΦC31 integrase with ΦbΦ10 and 12 *attP* and *attB* help to narrow down the nucleotides in an *att* site which may be essential for integrase to be able to cleave the site. Since ΦC31 integrase cleaved ΦbΦ10 *attP* and *attB* much faster than it did ΦbΦ12 *attP* and *attB in vitro,* and only ΦbΦ10 *attP* and *attB* was resolved *in vivo,* the sixth nucleotide from the centre of each *att* site is likely involved in site cleavage and, at least in the case of ΦC31 integrase, essential for improved site cleavage.

The results for ΦC31 integrase completely resolving both ΦtΦ14 and 12 *attP* and *attB in vivo* as well as cleaving them both and partially recombining them *in vitro* only adds further credence to this. As mentioned previously, when the central 12 bp of Tn3 *res* site I are compared to the same region of ΦC31 *attP* and *attB* sites, one can see that 6 of the 12 bp are conserved between the two in the *attP* site, including the sixth basepair on either side of the centre of the site (Figure 6.1). However, only 2 of 12 bp are conserved between this region in Tn3 *res* site I and the ΦC31 *attB* site, which would perhaps explain how ΦC31 integrase fails to completely recombine the hybrid sites *in vitro.* In contrast, the central 12 bp taken from BxbI *att* sites to create sites ΦbΦ10 and 12 *attP* and *attB* only contain 2 of 12 and 4 of 12 conserved basepairs compared to ΦC31 *attP* and *attB* sites respectively. Taking these results together, the ability of ΦC31 integrase to cleave and recombine different hybrid *att* sites can be seen to increase when more basepairs in the centre of the hybrid sites matches the basepairs originally in that position in ΦC31 natural *att* sites, though some of these positions hold more sway over this activity e.g. position 6 from the centre of the crossover dinucleotide.

To test the above hypotheses in the future, new *att* sites should be created to test whether increasing or decreasing the number of basepairs swapped from

the centre of the site can alter the efficiency of cleavage and recombination by the hybrid integrase created for them. Furthermore, if new hybrid integrases were to be created using different serine recombinase domains, complementary *att* sites should also be created for them following a similar rationale.

Additionally, as touched upon in Section 5.4, given that ΦC31 integrase completely resolved both ΦtΦ14 and 12 *attP* and *attB in vivo,* it would be interesting to investigate whether ΦC31 integrase is capable of recombining the resultant *attL* and *attR* hybrid sites in the presence of its RDF, the protein gp3, back to *attP* and *attB.* If possible, this would provide information on how efficiently and "correctly" ΦC31 integrase manages to cleave and recombine hybrid *attLxattR* back to *attP* and *attB.*

Concluding remarks

The aim of this project was to demonstrate that functional hybrid serine integrases, along with complementary hybrid *att* sites, were possible. Whilst the activity of TΦ4 and TΦ8 on hybrid sites was low, they nevertheless demonstrated that it is indeed possible to engineer hybrid serine integrases to recognise specific sites designed with the hybrid integrase in mind, both *in vivo* and *in vitro*. Going forward, the priority should be to optimise the design of future hybrid integrases to maximise the activity they exhibit on engineered *att* sites. Work should also be undertaken to try and minimise the ability of pre-existing integrases to cleave and recombine newly engineered *att* sites so that such sites are recombined only by the hybrid integrases designed for them. This would help a great deal in solidifying the place of serine integrases as a modifiable, modular tool to use in independently-functioning circuits in synthetic biology, as well as in genetic engineering and *in vivo* gene therapy.

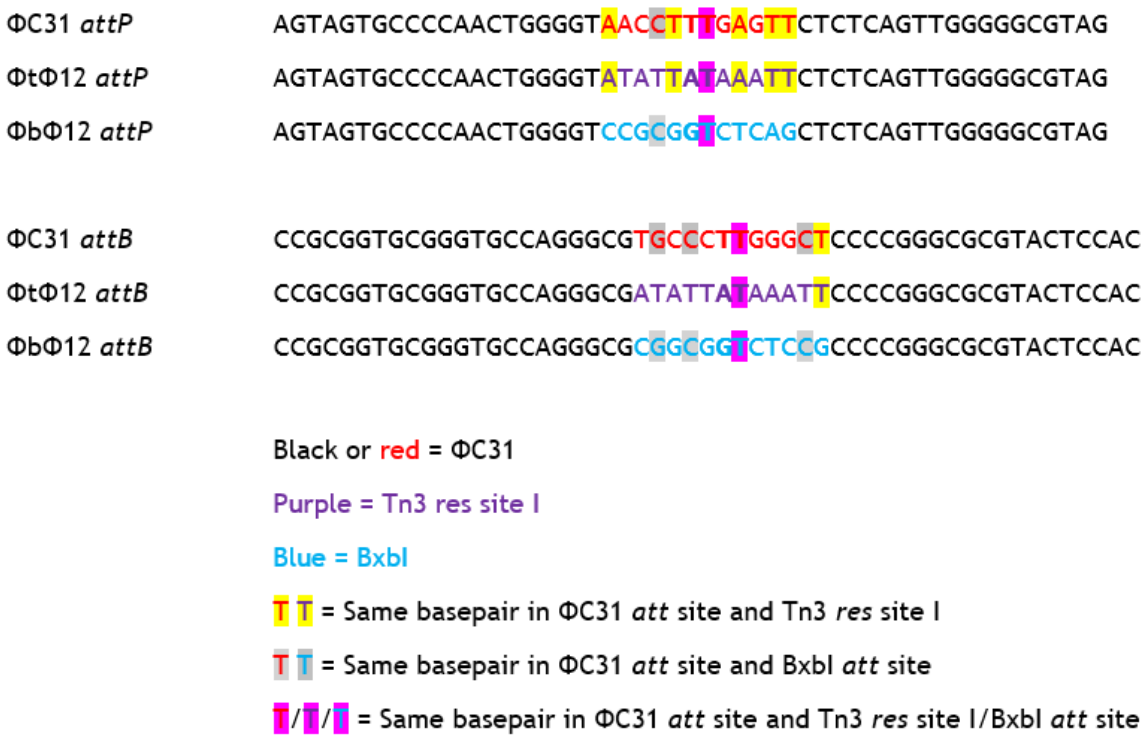

Figure 6.1: Conserved basepair comparison between ΦC31 natural *att* sites and select ΦC31/Tn3 and ΦC31/BxbI hybrid *att* sites.

In all instances, only the top strand (5' to 3') is shown.

References

Akopian, A., He, J., Boocock, M. R., & Stark, W. M. (2003). Chimeric recombinases with designed DNA sequence recognition.' *Proceedings of the National Academy of Sciences,* 100(15) 8688-8691.

Arnold, P. H., Blake, D. G., Grindley, N. D., Boocock, M. R., & Stark, W. M. (1999). Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *The EMBO journal,* 18(5), 1407-1414.

Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H. and Morikawa, K. (2000). Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer. *Proc Natl Acad Sci USA*, 97(15), 8257-62.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819), 1709-1712.

Beerli, R. R., & Barbas, C. F. (2002). Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology*, 20(2), 135-141.

Benner, S. A. and Sismour, A. M. (2005). Synthetic Biology. *Nature Reviews*, 6, pp.533-543.

Blake, D.G. (1993): 'Binding of Tn3 resolvase to its recombination site.' PhD thesis, University of Glasgow.

Bondy-Denomy, J., Pawluk, A., Maxwell K. L. and Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature,* 493, pp.429-432.

Bonnet, J., Subsoontorn, P. and Endy, D. (2012). Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, *109*(23), 8884-8889.

Burke, M. E., Arnold, P. H., He, J., Wenwieser, S. V., Rowland, S. J., Boocock, M. R. and Stark, W. M. (2004). Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol Microbiol*, 51(4), 937-48.

Camerini-Otero, R. D. and Hsieh, P. (1995). Homologous recombination proteins in prokaryotes and eukaryotes. *Annu Rev Genet*, 29, 509-52.

Cavazzana-Calvo, M., Payen, E., Negre, O., Wang, G., Hehir, K., Fusil, F., ... & Leboulch, P. (2010). Transfusion independence and hmga2 activation after gene therapy of human [bgr]-thalassaemia. *Nature*, 467(7313), 318-322.

Chambers, S. P., Prior, S. E., Earstow, D. A. and Minton, N. P. (1988). pMTL nick-cloning vectors. I. improved pUC polylinker regions to facilitate the use of Sonicated DNA for nucleotide sequencing. *Gene 68*, 139-149.

Colloms, S. D., Merrick, C. A., Olorunniji, F. J., Stark, W. M., Smith, M. C., Osbourn, A. and Rosser, S. J. (2014). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic acids research*, *42*(4), e23-e23.

Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., & Barbas, C. F. (2001). Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. *Journal of Biological Chemistry*, 276(31), 29466-29478.

Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., ... & Barbas, C. F. (2005). Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. *Journal of Biological Chemistry*, 280(42), 35588-35597.

Farruggio, A. P. and Calos, M. P. (2014). Serine integrase chimeras with activity in E. coli and HeLa cells. *Biology open*, BIO20148748.

Fischer, A., Hacein-Bey-Abina, S. & Cavazzana-Calvo, M. (2011). Gene therapy for primary adaptive immune deficiencies. *Journal of Allergy and Clinical Immunology*, 127(6), 1356-1359.

Gaj, T., Guo, J., Kato, Y., Sirk, S. J., & Barbas III, C. F. (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nature methods,*9(8), 805-807.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*,31(7), 397-405.

Ghosh, P., Pannunzio, N. R., & Hatfull, G. F. (2005). Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *Journal of molecular biology*, *349*(2), 331-348.

Gordley, R. M., Smith, J. D., Gräslund, T. & Barbas III, C. F. (2007): 'Evolution of programmable zinc finger-recombinases with activity in human cells.' *Journal of molecular biology*, 367(3), 802-813.

Gordley, R. M., Gersbach, C. A., & Barbas, C. F. (2009) Synthesis of programmable integrases. *Proceedings of the National Academy of Sciences*,106(13), 5053-5058.

Grindley, N.D.F., (1994). Resolvase-mediated site-specific recombination. Nucl. *Acids Mol. Biol*. 8, pp. 236–267.

Grindley N. D. F., Craig N.L, Craigie R, Geller M, Lambowitz AM (2002). The movement of Tn3-like elements: 'Transposition and cointegrate resolution'. *Mobile DNA II*. Washington (D. C.): ASM Press. pp. 272–302.

Grindley, N. D., Whiteson, K. L. and Rice, P. A. (2006). Mechanisms of sitespecific recombination. *Annu Rev Biochem*, 75, 567-605.

Gupta, M., Till, R. and Smith, M. C. M. (2007). Sequences in *attB* that affect the ability of ΦC31 integrase to synapse and to activate DNA cleavage. *Nucleic Acids Research,* 35(10), pp.3407-3419.

Hobom, B. (1980). Surgery of genes-at the doorstep of synthetic biology. *Medizinische Klinik*, *75*(24), 14-21.

Holliday, R. (1964): A mechanism for gene conversion in fungi. *Genet Res*, 89(5-6), 285-307.

Holt, S. E. (2014). *Target site DNA recognition by Tn3 and Sin resolvases* (Doctoral dissertation, University of Glasgow).

Joung, J. K. and Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature reviews,* 14, pp.49-55.

Laemmli , U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685

Li, W., Kamtekar, S., Xiong, Y., Sarkis, G. J., Grindley, N. D. and Steitz, T. A. (2005). Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs. *Science,* 309(5738), 1210-5.

Mercer, A. C., Gaj, T., Fuller, R. P., & Barbas, C. F. (2012). Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic acids research,* 40(21), 11163-11172.

Moffatt, B. A. and Studier, F. W. (1987). T7 Lysozyme Inhibits Transcription by T7 RNA Polymerase. *Cell,* 49, pp.221-227.

Nakano, T., Eckford, A. W., & Haraguchi, T. (2013). *Molecular communication*. Cambridge University Press.

Olorunniji, F.J., Buck, D.E., Colloms, S.D., McEwan, A.R., Smith, M.C.M., Stark, W.M. and Rosser, S.J. (2012). Gated rotation mechanism of site-specific recombination by ΦC31 integrase. *PNAS,* 109(48), pp.19661-19666.

Ott, M. G., Schmidt, M., Schwarzwaelder, K., Stein, S., Siler, U., Koehl, U. & Grez, M. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nature medicine,* 12(4), 401-409.

Prorocic, M. M., Wenlong, D., Olorunniji, F. J., Akopian, A., Schloetel, J. G., Hannigan, A. and Stark, W. M. (2011). Zinc-finger recombinase activities in vitro. *Nucleic acids research*, gkr652.

Proudfoot, C., McPherson, A. L., Kolb, A. F. and Stark, W. M. (2011). Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS One, 6*(4), e19537.

Rae, R. V. and Bredesen, D. E. (2004). Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr Opin Cell Biol,* 16(6), pp.653-662.

Rice, P. A. (2015). Serine Resolvases. *Microbiology Spectrum,* pp.1-15.

Rice, P. A., & Steitz, T. A. (1994). Model for a DNA-mediated synaptic complex suggested by crystal packing of gamma delta resolvase subunits. *The EMBO journal*, 13(7), 1514.

Rowland, S. J., Boocock, M. R., McPherson, A. L., Mouw, K. W., Rice, P. A. and Stark, W. M. (2009). Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome. *Mol Microbiol,* 74(2), 282-98.

Rutherford, K., Yuan, P., Perry, K., Sharp, R., & Van Duyne, G. D. (2013). Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic acids research*, *41*(17), 8341-8356

Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L., & Weiss, D. S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature,* 497(7448), 254-257.

San Filippo, J. Sung, P. and Klein. H (2008). Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry* Vol. 77: 229-257

Singh, S., Rockenbach, K., Dedrick, R. M., VanDemark, A. P., & Hatfull, G. F. (2014). Cross-talk between diverse serine integrases. *Journal of molecular biology*, *426*(2), 318-331.

Sirk, S. J., Gaj, T., Jonsson, A., Mercer, A. C., & Barbas, C. F. (2014). Expanding the zinc-finger recombinase repertoire: directed evolution and

mutational analysis of serine recombinase specificity determinants. *Nucleic acids research*, 42(7), 4755-4766.

Smith, C. M. (2015). Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiology Spectrum,* pp.1-19

Smith, M.C.A., Till, R., Brady, K., Soultanas, P., Thorpe. H. and Smith, M.C.M (2004). Synapsis and DNA cleavage in ΦC31 integrase-mediated site-specific recombination. *Nucleic Acids Research,* 32(8), pp.2607-2617.

Smith, M.C.M., Brown, W.R.A, McEwan, A.R. and Rowley, P.A. (2010). Site-specific recombination by φC31 integrase and other large serine recombinases. *Biochemical Society Transactions,* 38(2), pp.388-393.

Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992). Catalysis by sitespecific recombinases. *Trends Genet,* 8(12), 432-9.

Stark, W. M. and Boocock, M. R. (1995). Gatecrashers at the catalytic party. *Trends Genet,* 11(4), 121-3.

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, *189*(1), 113-130.

Summers, D. K., & Sherratt, D. J. (1988). Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. *The EMBO journal*, 7(3), 851.

Van Duyne, G. D. (2015). Cre Recombinase. *Microbiology Spectrum,* pp.1-19.

Van Duyne, G. D. and Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol,* 48(5), 476-491.

Yang, W. and Steitz, T. A. (1995). Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell,* 82 (2), 193-207.

Yuan, P., Gupta, K. and Van Duyne, D.G. (2008). Tetrameric Structure of a Serine Integrase Catalytic Domain. *Structure*, 16, pp.1275-1286.