University of Glasgow

Examination of role and function of the Coiled-Coil domain of φC31 integrase

Jakub Filip Abramiuk
BSc (Hons) Genetics

Submitted to the University of Glasgow in fulfilment
of the requirements for the degree of MSc (Research)
in Molecular Genetics.

School of Molecular Biosciences
College of Medical, Veterinary and Life Sciences
University of Glasgow

September 2025

ABSTRACT:

φC31 Integrase (Int) is a recombinase enzyme used by the φC31 bacteriophage to integrate its DNA into the chromosome of a host bacterium at initiation of the lysogenic cycle. The enzyme has several useful properties, such as strong recombination site specificity, high reaction efficiency in optimal conditions and unidirectionality of the recombination, which can be reversed in the presence of a Recombination Directionality Factor (RDF) - a secondary protein factor. For those reasons, it (and other related integrases from the Large Serine Recombinase (LSR) family) have been used in biotechnological and genetic applications, such as targeted insertion/excision of large DNA fragments.

Despite this interest in the applications of the enzyme, the biochemistry underlying properties of its activity is still not fully understood and is under ongoing examination, especially with regard to directionality control. During the forward reaction φC31 Int targets and forms a synapse on a heterologous pair of sites called *attP* and *attB*, forming hybrid sites *attL* and *attR* as products of the reaction. The presence of RDF allows φC31 Int to form a synapse on *attL* and *attR* sites and inhibits synapse formation at *attP* and *attB*, resulting in a reverse reaction. Several structural studies in closely related LSRs have provided evidence of involvement of the Coiled-Coil (CC) domain (located near the C-terminus of the φC31 Int, between residues G445 and A531) in the formation of a tetrameric complex during synapse formation. This could mean that CC domain interactions are part of the control mechanism for φC31 Int reaction directionality.

This project aims to expand the knowledge on φC31 Int reaction specificity and directionality by examining several point mutations in the CC domain. Residues of interest were chosen based on sequence alignments with other LSRs, and are thought to be analogous to residues responsible for CC-CC interactions during formation of the synaptic complex in those LSRs. Information on their role in φC31 Int activity and specificity could be important for understanding the biochemistry of the enzyme, and subsequently could help unlock the ability to create designer versions with altered properties for genetic engineering and synthetic biology applications.

Mutants of L474 and Y475 residues, as well as a complete CC deletion (ΔCC) variant were produced. Other identified residues of interest included L468, L471, E472, V487, H491, F492, R493, Q496, L499, T500 and E507. *In vitro* assays for reaction efficiency (including tests with *att* sites associated with different LSRs to check for site specificity disruptions) as well as synaptic binding (including experiments with unorthodox *att* site pairings), were

conducted for the produced mutants. The alanine mutant of L474 showed some activity, albeit reduced in comparison with wild-type. Y475A and ΔCC mutants were unable to facilitate the recombination reaction under any conditions, cementing the view that the CC domain is vital during the synapsis, as well as confirming the suspicion that Y475-mediated interactions are important to CC domain function. Interestingly, the Y475H mutant showed activity in a forward reaction, which may imply that the interactions this residue takes part in are at least partially dependent on the shape and size of the tyrosine sidechain. However, it was almost completely inactive on *attL* x *attR* substrates, suggesting that its polarity might be important for interactions enabling the reverse reaction. All active mutants retained high site specificity.

ACKNOWLEDGEMENTS:

I would like to thank my supervisors, Professor Marshall Stark and Doctor Sean Colloms for their invaluable support and guidance throughout my research process, massive contributions to my understanding, skills and knowledge, as well as their sizable reserves of patience and understanding for my struggles which I am immensely grateful for and am not sure if I could finish this project without.

I would like to thank Arlene McQue for her guidance and teachings related to the numerous laboratory procedures, which enabled my work on the project and considerably expanded my laboratory skills.

I would like to thank my parents and grandparents for supporting my livelihood throughout the research period, which allowed me to dedicate my full time to the project.

I would like to thank my dear friend Lily Whyler for helping me through the process of obtaining medication which allowed me to keep my mental disabilities at bay. Without their help it may well be that the writeup process for this thesis would not have been possible.

And lastly, I would like to thank my convenor of studies, Doctor Ian Salt, for his help in handling the bureaucratic and administrative part of my postgraduate study process, as well as his frequent advice on my work and my future plans.

DECLARATION:

This thesis and any findings presented within it are results of my own work, unless explicitly stated and/or cited otherwise. This thesis or its part have not been submitted to any other degree at the University of Glasgow or at any other institution.

~ Jakub Filip Abramiuk

TABLE OF CONTENTS:

LIST OF ABBREVIATIONS:

A, Ala - alanine
AA - amino-acid
att - attachment site
bp - base pair
C, Cys - cysteine
CD - Catalytic domain
CC - Coiled-Coil domain
CTD - C-terminal domain
D, Asp - aspartic acid
Δ - deletion
ds - double-stranded
DSB - double-stranded break
DTT - Dithiothreitol
E, Glu - glutamic acid
F, Phe - phenylalanine
G, Gly - glycine
H, His - histidine
His-tag - poly-histidine affinity tag
HPLC - High-performance liquid chromatography
I, Ile - Isoleucine
IMAC - Immobilized metal affinity chromatography
Int - Integrase
K, Lys - lysine
L, Leu - leucine
LB - Lysogeny Broth
LSR - Large Serine Recombinase
M, Met - methionine
OD - optical density
ORF - open reading frame
R, Arg - arginine
RD - Recombinase domain
RDF - Recombination Directionality Factor
RE - restriction enzyme
Q, Gln - glutamine
S, Ser - serine
SDS-PAGE - sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SR - Serine recombinase
ss - single-stranded
SSR - Small Serine Recombinase
T, Thr - threonine
YR - Tyrosine recombinase
V, Val - valine

-ve - negative control
+ve - positive control
WT - wild-type
Xis - excisionase accessory factor
Y, Tyr - tyrosine
ZD - Zinc-containing domain

LIST OF FIGURES:

1. INTRODUCTION

1.1 Introduction to Site-Specific Recombinases

Site-specific recombinases are a large class of enzymes, characterised by their ability to perform excision, inversion or integration of two DNA sequences at enzyme-specific recombination sites, without causing insertion or deletion of nucleotides around the recombination site. This process is called site-specific DNA recombination, and consists of localised exchange of DNA strands (Figure 1) [Thorpe *et al.,* 2000]. Most of the known members of the superfamily are bacterial, fungal and viral enzymes [Wang *et al.*, 2011; Nern *et al.*, 2011].

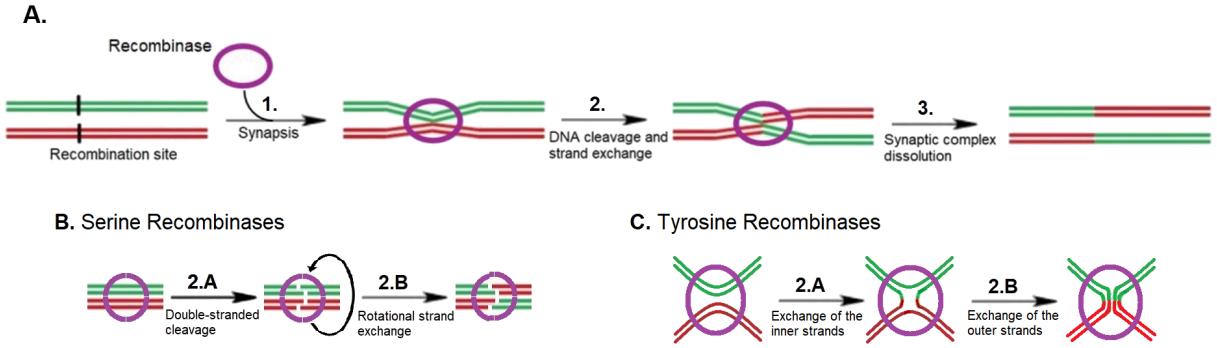


Figure 1. A - General schematic of a recombination reaction. Two double stranded DNA molecules are brought together in a synaptic complex and cleaved at the recombination sites. The resulting ends are exchanged and ligated, forming two hybrid molecules. B - Recombination mechanism as performed by serine recombinases. Both DNA strands are cleaved simultaneously, and the exchange occurs via rotational displacement facilitated by the recombinase subunits. C - Recombination mechanism as performed by the tyrosine recombinases. Inner strands are cleaved and ligated to ends of opposing inner strands, forming a Holliday junction-like intermediate. Outer strands are then cleaved and ligated in a similar manner.

Bacteriophage recombinases are site-specific, and their target DNA sequences are commonly referred to as "attachment sites" or *att*s. The sequence of those *att* sites is unique to each enzyme, allowing for a great deal of target control if the library of available recombinases is extensive enough. There are many domain types employed in the enzyme class to facilitate the recognition and binding of the target sequences, including but not limited to the Recombinase Domain in Large Serine Recombinases and Helix-turn-Helix motifs in Small Serine Recombinases [Van Duyne and Rutherford, 2013]. Some recombination sites are directional, meaning that the orientation of the site within the substrate DNA sequence can affect the result of recombination (Figure 2) [Nern *et al.*, 2011].

This allows the enzymes to perform various functions, such as excision/insertion, translocations, or fragment exchange depending on the geometry of the substrate DNA molecule(s) and the *att* sites (Figure 2). While some recombinases target a pair of identical sites, many types of bacteriophage recombinases (discussed below) have non-identical recognition sites, usually referred to as "phage attachment site" (*attP*) and "bacterial attachment site" (*attB*) and facilitate the recombination reaction between those two non-identical sites.

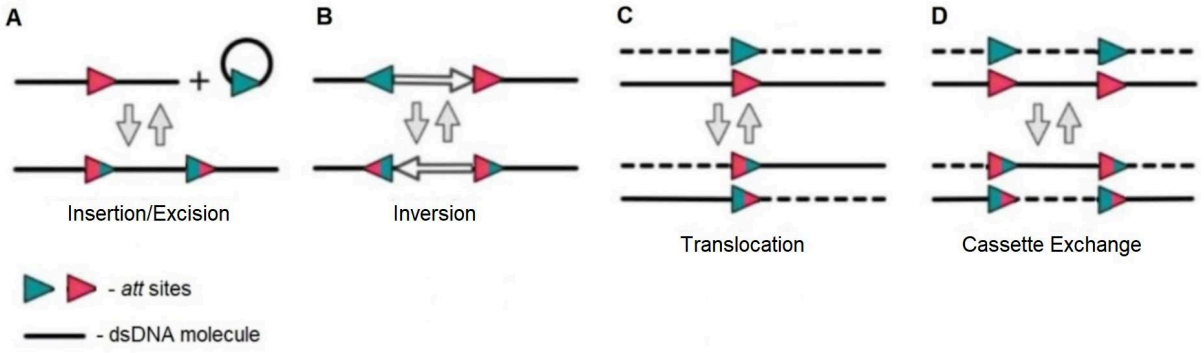


Figure 2. Common recombination reaction types facilitated by site-specific recombinases. Depending on the number and geometry of substrate DNA and the *att* recognition sites (as well as the direction of the reaction), different end results such as insertion/excision (A), direct inversion (B) and exchange of linear ends (translocation, C) or inner fragments (cassette exchange, D) can be accomplished. Figure adapted from Nern *et al.*, 2011.

Like most other DNA recombination mechanisms, recombinase systems depend on formation of a synapse - a system in which the two separate recombination sites are brought and held together by the recombinase to allow exchange of the DNA strands (Figure 1) [Thorpe *et al.,* 2000]. However, unlike some other functionally similar enzymes (such as the human Rad51 system, responsible for double-stranded break repair via homologous recombination [Bonilla *et al.*, 2020]), site-specific recombinases do not require the presence of pre-existing DSBs to facilitate recombination. Instead, after the synapse is properly formed, those recombinases create breaks at the target recombination sites themselves, cleaving the DNA using their signature catalytic domains (CDs) [Van Duyne and Rutherford, 2013; Stark, 2017]. Differences in the mechanism of this DNA cleavage are the primary factor dividing recombinases into two families. Tyrosine Recombinases (YRs) and Serine Recombinases (SRs) are distinguished from each other based on the fact that the former use a catalytic tyrosine residue to facilitate the DNA cleavage, while the latter use a catalytic serine residue to accomplish the same goal. YRs include enzymes such as Cre, FLP and λ Integrase [Wang *et al.*, 2011]. SRs include enzymes such as γδ Resolvase, A118 Integrase and φC31 Integrase [Nöllmann *et al.*, 2005; Van Duyne and Rutherford, 2013].

Because of the high site specificity, multiple geometry-dependent modes of action, relative simplicity of the recombination complexes and lack of deletion/insertion mutations caused by the recombination process, recombinases make for an attractive tool for genetic engineering [Nern *et al.*, 2011; Farrugio *et al.*, 2012]. In fact, recombinase systems were successfully used in creation of transgenic plants as early as 1991 [Dale and Ow, 1991].

1.2 Serine Recombinases

The main characteristic of Serine Recombinases which distinguishes them from other types of recombinases is their catalytic domain (CD) - typically located at the N-terminus of the protein, and always containing a catalytic serine residue, which is responsible for creating the DNA strand break necessary for strand exchange [Wang *et al.*, 2011; Van Duyne and Rutherford, 2013]. Unlike YRs, which create a single-stranded break at the recombination site followed by formation of a four-stranded intermediate akin to a Holliday junction (Figure 1 C) [Fogg *et al*., 2014], SRs native to bacteriophages create double-stranded breaks (DSBs) at both recombination sites at once. This is followed by single-step rotational exchange of the cut DNA ends [Stark, 2017]. Another difference between YRs and phage SRs is the fact that some YRs (such as λ Integrase) require the presence of accessory proteins to facilitate the recombination [Cho *et al.*, 2002]. This is especially prevalent in YRs targeting non-identical pairs of *attP* and *attB* sites. Unlike most phage SR *att* sites, YR non-identical targets tend to have high size differentials, with over 200bp differences of length between the longer *attP* site and much shorter *attB* site in case of λ Integrase. The sizable length of the *attP* sequences provides space for binding sites for accessory proteins. In contrast, phage SR systems with non-identical *att* sites, which do not use accessory proteins, tend to have much less asymmetric *attP* and *attB* sites (with average lengths of ~50bp and ~40bp respectively) [Stark, 2017].

SRs are divided into two families, Small Serine Recombinases (SSRs, also known as Resolvases/Invertases) and Large Serine Recombinases (LSRs). LSRs native to bacteriophages are also known as Serine Integrases [Thorpe *et al.*, 2000; Sarkis *et al.*, 2001; Van Duyne and Rutherford, 2013]. As the name implies, LSRs are usually significantly larger than SSRs (usually composed of over 450 AA residues, as opposed to ~200 residues of an

average SSR), and have a slightly different domain makeup and general structure (Figure 3) [Van Duyne and Rutherford, 2013].

SSRs, such as γδ Resolvase or Tn3 Resolvase, usually target a pair of identical *res* sites. Those *res* sites are conserved after the recombination [Wang *et al.*, 2011]. However, resolvases are usually only able to facilitate one specific type of reaction. For example, Tn3 Resolvase was shown to only recombine two intramolecular sites, without the ability to recombine sites on two separate plasmids [Krasnow and Cozzarelli, 1983]. This allows it to facilitate excisions, but not integrations.

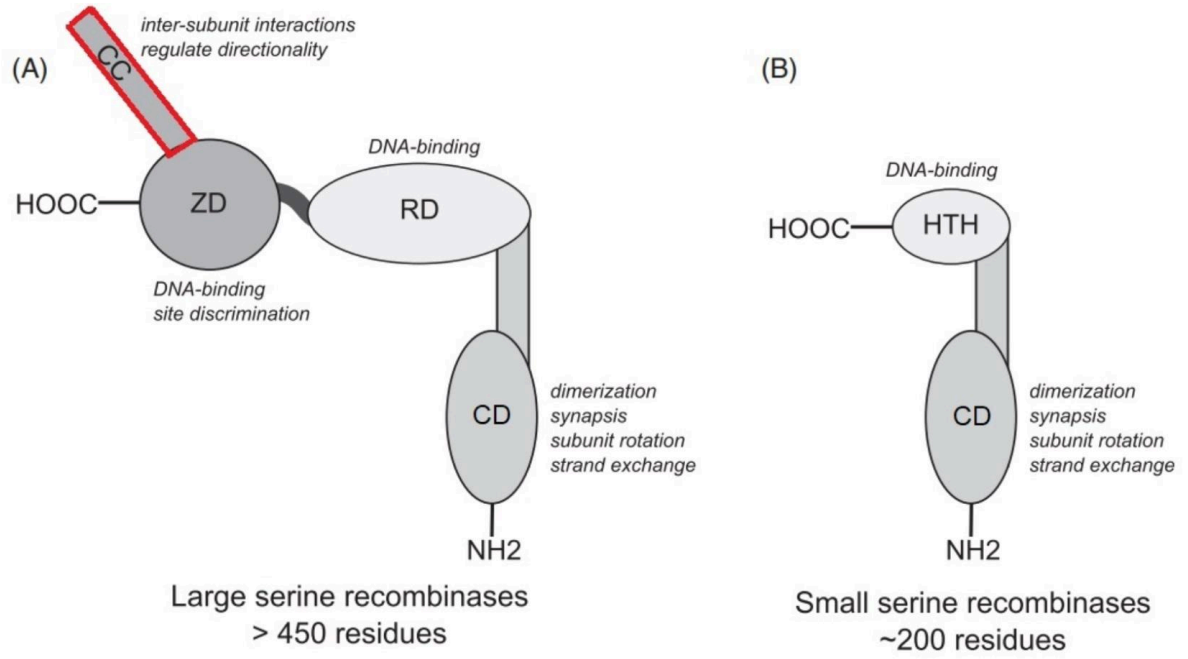


Figure 3. Structural comparison between a model Large Serine Recombinase (A) and a model Small Serine Recombinase (B). The Coiled-Coil domain of the LSR is highlighted in red. Abbreviations: CD - Catalytic domain, HTH - helix-turn-helix motif, RD - Recombinase domain, ZD - Zinc-ribbon domain, CC - Coiled-Coil domain. Figure adapted from Van Duyne and Rutherford, 2013.

1.3 Large Serine Recombinases

The SSR mechanism described in Chapter 1.2 contrasts with the mechanisms of LSRs and unidirectional YRs, which always target a non-identical pair of *attP* and *attB* sites. Because of the asymmetrical nature of those sites, recombination creates hybrid *attL* and *attR* sites. Since recombinases are highly site-specific, this means that under normal conditions reaction is unidirectional. However, both YRs and LSRs show the ability to facilitate *attL* x

attR recombination in the presence of secondary protein factors, dubbed excisionase accessory factors (Xis) in YRs [Cho *et al.*, 2002] and recombination directionality factors (RDFs) in LSRs [Thorpe *et al.*, 2000; Fogg *et al.* 2018]. This is related to the biological function of those enzymes - integration of phage DNA into the bacterial genome cannot be reversed easily, as that would disrupt the lysogenic cycle. The insertion reaction resulting in hybrid *att* sites prevents the phage's own enzymes from excising the prophage from the host genome prematurely.

LSRs are more recently described cousins of Small Serine Recombinases [Thorpe *et al.,* 2000]. LSRs are generally composed of three main domains (listed from N to C terminus): a Catalytic domain (CD), which is responsible for DNA cleavage and strand exchange; a Recombinase domain (RD), which is responsible for binding to the target *att* site; and a C-terminal domain, which is usually subdivided into a Zinc-ribbon domain (ZD, responsible for *att* site recognition) and a Coiled-Coil domain (CC) (Figure 3) [Van Duyne and Rutherford, 2013; Rutherford and Van Duyne, 2014]. Both RD and ZD are thought to facilitate DNA binding via interactions within the major groove of the double helix. RD interacts with a major groove using an α-helical motif structurally similar to the SSR HTH motif, while ZD wraps around the DNA using a β-sheet hairpin [Mouw *et al.*, 2008; Van Duyne and Rutherford, 2013]. In solution, LSRs form soluble dimers, which are ready to bind to dsDNA [Van Duyne and Rutherford, 2013]. All LSRs are site-specific, with each dimer recognising and binding one of the two non-identical sites (*attP* and *attB*). After DNA is bound to the protein dimer, *attP*-bound dimer and *attB*-bound dimer form synaptic complexes via tetramerisation. This is followed by DNA cleavage, performed by the eponymous active site serine residue, then rotational exchange of DNA strands and ligation of the exchanged strands, resulting in the creation of hybrid product sites (*attL* and *attR*) [Figure 4; Van Duyne and Rutherford, 2013; Stark, 2017].

In most described LSR systems recombination is unidirectional under standard reaction conditions, and will only happen in the *attP* x *attB* -> *attL* x *attR* direction if the LSR is supplied on its own (this is dubbed the forward reaction) [McEwan *et al.*, 2009; Stark, 2017]. However, this directionality can be controlled. Most LSRs have associated recombination directionality factors (RDFs), which are unique to each enzyme. The presence of the appropriate RDF allows the enzyme to facilitate *attL* x *attR* -> *attP* x *attB* recombination (usually dubbed the reverse reaction) with a degree of efficiency that largely

shifts the equilibrium to the *attP* x *attB* products [Farrugio *et al.*, 2012; Fogg *et al.,* 2018]. It is worth noting that there are known exceptions to this general rule of strict unidirectionality - for example, φBT1 Integrase is known to facilitate reverse reaction even in absence of its RDF, albeit with significantly lesser efficiency than the forward reaction [Zhang *et al.*, 2008].

Coiled-Coil domain interactions (Figure 4) are thought to be important for activation of the synaptic tetramer, and consequently also are a vital factor in the control of directionality [Gupta *et al.*, 2017; Mandali and Johnson, 2021]. Section 1.6 provides a detailed overview of LSR CC domain studies relevant to this thesis.

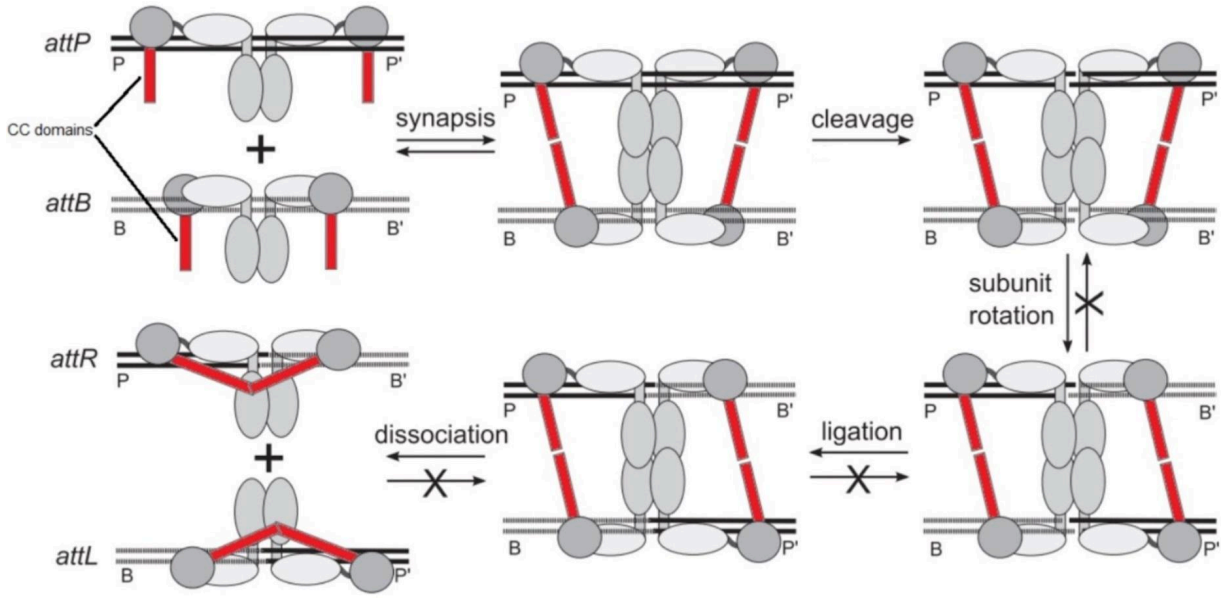


Figure 4. A representation of a model LSR-catalysed recombination reaction. First, recombinase dimers identify and bind the *attP* and *attB* sites. Then, tetramerisation occurs, resulting in a formation of the synaptic complex. Subsequently, the catalytic serine of the CD domain cleaves the DNA at the recombination point (note the overhangs created by the cleavage on both sites). Subunits on one side of the complex rotate around a central axis, swapping places of the cleaved ends (note that under standard reaction conditions this rotation is gated after a single 180^{o} turn, preventing reversion of the process [Olorunniji *et al.*, 2012]). The DNA is then relegated, forming hybrid *attL* and *attR* sites, and the synaptic complex dissociates. LSRs are unable to reverse this process (unless an RDF is present), which is indicated by the crossed arrow. Note the CC-CC interactions in the synaptic tetramer and how they change when the complex dissociates into *attL* and *attR* bound dimers (CC domains highlighted in red). Abbreviations: CC domain - Coiled-Coil domain, CD - Catalytic domain. Figure adapted from Van Duyne and Rutherford, 2013.

1.4 Introduction to Coiled-Coil domains

CC domains are a common structural motif in proteins, consisting of a number of α-helices (usually 2, with examples comprising up to 5 helices described), wrapped around

each other in a helical pattern to form a supercoil [Mason and Arndt, 2004]. The α-helices comprising the Coiled-Coil can differ from one another in their AA sequence, but there are some conserved properties which allow the structure to maintain its shape. Most notably, residues pointing inside the supercoil are usually hydrophobic, to facilitate Van der Waals interactions between the helices, while residues pointing outwards from the CC but in close proximity to other constituent helices are often polar and facilitate ionic interactions between the constituent helices, while preventing aggregation. Outward-facing residues are also usually hydrophilic, allowing the domain to be soluble [Mason and Arndt, 2004].

Because of their relative simplicity, CC domains are known to have numerous functions in various proteins, ranging from simple structural motifs to fulfilling more complex functions such as molecular spacers, or physical receptors and signal transporters (via facilitation of conformational changes along their length) [Mason and Arndt, 2004; Truebestein *et al.*, 2016].

In LSRs it is proposed that various types of CC-CC interactions at the tips of the domain lead to tetramer activation or inhibition depending on the geometry and residues of interaction (Figure 4; Figure 6), which in turn would be dependent on the *att* sites present in the complex [Smith *et al.*, 2010; Mandali and Johnson, 2021]. The nature of those interactions is expanded upon in Chapter 1.6.

1.5 φC31 Integrase

The main focus of this study, φC31 Int, was one of the first LSRs to be described and studied [Thorpe and Smith, 1998; Smith, 2015]. It is a viral enzyme, natively expressed by a *Streptomyces* bacteriophage φC31 [Smith *et al.*, 1999]. It is responsible for integrating the phage genome into the bacterial chromosome, and thus facilitates formation of the prophage at the initiation of the lysogenic cycle [Kuhstoss, 1992; Smith *et al.*, 2010].

1.5.1 φC31 Integrase - Structure

φC31 Integrase is a large enzyme, with a length of 605 AA residues and mass of 67,124 daltons [UniProt *Q9T221*, 2025]. Its general structure follows the classical LSR domain composition presented in Figure 3; however, its sequence is significantly longer than

most other LSRs which usually have between 450 and 500 AAs. Most of the described functional elements, such as the active site of the Catalytic domain (CD) are similar in size to those smaller LSRs, with extra sequences usually modelled as loosely structured loops placed between those conserved functional elements. The vast majority of those extra residues are found in the Recombinase domain (RD), the Zinc-containing domain (ZD) and the Coiled-Coil domain (CC) [Van Duyne and Rutherford, 2013]. As with other LSRs, φC31 Int undergoes dimerisation in a solution [Van Duyne and Rutherford, 2013], and forms tetrameric synaptic complexes when presented with appropriate *attP* and *attB* sites (Figure 4).

Despite the fact that φC31 Int is both one of the first described LSRs and has received wide biotechnological usage [Kempe *et al.*, 2010; Lister *et al.*, 2011; Chaves and Calos, 2011; Farruggio *et al.*, 2012], the first major crystallography studies on LSRs have been conducted on smaller enzymes such as LI/A118 Int, as their smaller size makes the enzyme much easier to model [Rutherford and Van Duyne, 2014; Gupta *et al.*, 2017]. Access to those structures made the LI Int a more appealing candidate for future studies, as it provided additional information to base experimental design on [W.M. Stark, personal communication, 25.09.2025]. This means that structural placement of the residues unique to φC31 Int was often speculative and its general structure was mostly inferred from similarities that φC31 Int shares with other closely related LSRs. First study showcasing high resolution structures of the φC31 Int complexes has been released to review shortly after the completion date of this project, and is awaiting peer review at the time of writing [Sun *et al*., 2025].

The CC domain of φC31 Int is located near the C-terminus of the protein. It protrudes from the ZD, and as such both domains are commonly referred to in older studies collectively as the C-Terminal domain (CTD) [Van Duyne and Rutherford, 2013]. Structurally, the bulk of the φC31 Int CC domain is made from two α-helices, usually called αK and αL. In most other LSRs those two helices comprise the vast majority of the CC domain, but φC31 Int has an additional length of sequence which is modelled to be placed between the αK and αL, and which is assumed to be unstructured (Figure 5) [Gupta *et al.*, 2017]. Pre-existing sequence alignments of LI/A118 Int and φC31 Int place the φC31 Int CC domain between residues G445 and A531 [Van Duyne and Rutherford, 2013; Gupta *et al.*, 2017]. The UniProt automatic annotation algorithm identifies a Coiled-Coil structure between residues R457 and A516 [UniProt, 2025], which lies within the proposed region. The full amino-acid sequence of φC31 Int can be found in Appendix B.1.

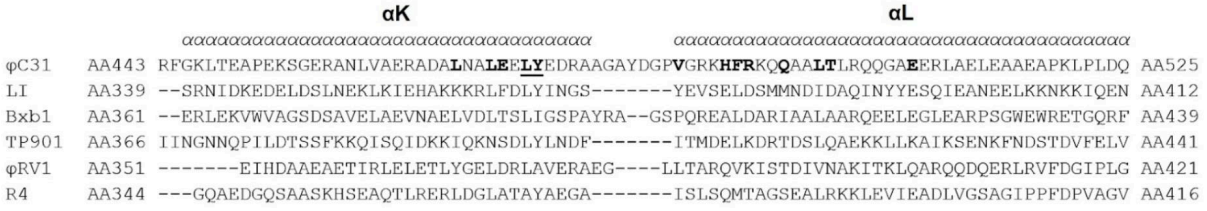


Figure 5. Amino-acid sequence alignment of the CC domains of selected LSRs (labelled with the name of the phage of origin on the left). Location of the constituent helices (αK and αL) in the LI Int is indicated above the alignment by the chain of *α* symbols. Identified residues of interest (L468, L471, E472, L474, Y475, V487, H491, F492, R493, Q496, L499, T500 and E507) are indicated by **bolding.** Residues L474 and Y475, which have been mutated for the purpose of this study, are indicated by **underscore.** The short sequence between A479 and V487 is unique to φC31 Int. Its structure is unknown, but most alignments assume it to be an unstructured region [Gupta *et al*., 2017]. Most comparisons are drawn to LI Int, as it was the LSR with most detailed crystal structures available at the time of planning this project [Rutherford and Van Duyne, 2014].

1.5.2 φC31 Integrase - Reaction Properties

Under standard conditions φC31 Int facilitates strictly unidirectional, site-specific recombination between two non-identical recognition/recombination sites *attP* and *attB* [Pokhilko *et al.*, 2016]. The sequences of the φC31 Int *attP* (39 bp) and *attB* (34 bp) sites can be found in Appendix A.1 and A.2 [Combes *et al.*, 2002]. Natively, *attP* is part of the φC31 phage genome, and is encoded immediately upstream to the φC31 Int gene [Smith *et al.*, 1999]. The *attB* site is natively encoded in the genome of *Streptomyces ambofaciens*, the natural host of the φC31 phage [Thorpe *et al.,* 2000]. The main *attB* site lies within the ORF encoding putative chromosome condensation protein, SCAC2.06c [Combes *et al.*, 2002].

φC31 Int-mediated recombination follows the general reaction trajectory shared by most LSRs and presented in Figure 4: soluble dimers bind to target sites on the dsDNA molecules and form a tetrameric synaptic complex, which is followed by double-stranded cleavage catalysed by the serine residues of the CDs, exchange of the cleaved ends (which is facilitated by rotational motion of one half of the complex), ligation of the hybrid product sites and dissociation of the complex. Such a process is called the “forward” reaction, and occurs to near-completion in the complete absence of any other protein, or chemical factors (such as free ions or ATP) [Smith, 2015; Pokhilko *et al.*, 2016]. The cleavages on top and bottom strands are conducted 2 bp apart, which results in a DSB containing 2 nt overhang sticky ends, with the sticky end created by the cleavage of *attP* being complementary to the

sticky end created by the cleavage of *attB*. Those 2 nt at which crossover between strands occurs are often called the core sequence, or central dinucleotide, and in case of φC31 Int *attP* and *attB* their sequence on the top strand is 5'-TT-3', with bottom strand being antiparallel 5'-AA-3' [Combes *et al.*, 2002]. Because of that, the ligation of the hybrid sites is only possible in one particular orientation of *attP* and *attB* in relation to each other (parallel), as if one of the sites were anti-parallel to the other, after the rotational exchange both molecules would be facing incompatible cleaved overhangs (i.e. a TT overhang would face another TT overhang) [Stark, 2017].

As a result of completion of the forward recombination, the former *attP* and *attB* sites form two recombinant sites, *attL* and *attR* [Smith, 2015]. One of the advantages of the φC31 Int system is that the reaction is independent of the conformation of the target DNA - being able to recombine both supercoiled, non-supercoiled and linear DNA; as well as working both for reactions of *att* sites located on two separate DNA molecules and reactions with both *att* sites located on the same DNA molecule [Stark, 2017].

φC31 Int dimers have shown the ability to bind to all 4 of the relevant *att* sites (*P, B, L* and *R*). However, tetramerisation and thus proper synapse formation will only occur between *attP* and *attB* sites in the absence of the RDF [Thorpe *et al.,* 2000]. In the presence of RDF, φC31 Int will begin to mediate a "reverse" reaction, which entails recombination of an *attL* site with an *attR* site, following the same general steps as the forward reaction and resulting in recreation of *attP* and *attB* sites. Most notably, the presence of RDF is the only requirement for the reverse reaction to not only occur, but also become more energetically favourable - the reverse reaction still does not require any other external molecules such as ATP [Thorpe *et al.*, 2000; Pokhilko *et al.*, 2016].

In vitro, if recombination occurs between two *att* sites located on the same product plasmid, the two resulting product plasmids might form in various different topologies, including, but not limited to, interlinked catenanes. This is caused by the rotational mechanism of strand exchange facilitated by the integrase, and can be manipulated to deliberately create chains of interlinked plasmids or intricate DNA "knots" with precisely defined geometry [Mouw *et al.*, 2008; Olorunniji *et al.*, 2012].

1.5.3 φC31 Integrase - Recombination Directionality Factor

Each LSR has a unique Recombination Directionality Factor (RDF). These RDFs are accessory proteins, which, via interactions with the integrase, allow it to perform the reverse recombination, (ie *attL* x *attR* -> *attP* x *attB*) [Fogg *et al.,* 2018]. The φC31 Int RDF is called gp3. It is 244 AA residues long and has a mass of 27,328 daltons [UniProt *Q9T216*, 2025]. Interestingly, gp3 appears to be unrelated to some other described LSR RDFs such as the A118 Int RDF, despite the sequence similarity shared between the Integrases themselves [Fogg *et al.,* 2018]. This could imply that the unidirectional LSR mechanism evolved first, and various forms of RDFs evolved afterwards, providing additional versatility to a preexisting system. Recent studies provide evidence that most LSR RDFs could have evolved from other functional phage proteins which have mutated to create a compatible pair with an LSR, with examples such as Bxb1 Int RDF showing remarkable similarity to a 3'-5' editing exonuclease [Alsaleh *et al.*, 2025].

gp3 binding to φC31 Int enables the reverse reaction. Its binding site location was identified near the C-terminus of the enzyme, specifically at the base of the CC domain and in the linker connecting the CC domain with the ZD [Fogg *et al.*, 2018]. That region was identified in the same study as a "hinge", mobility of which is vital for allowing various CC-CC interactions, which are necessary for the activation of the synaptic complex and recombination catalysis. It is therefore hypothesised that gp3 binding might alter the orientation of the CC domains, enabling the interactions normally unique to the *attP* x *attB* synaptic complex geometry to happen in the *attL* x *attR* synapse [Mandali and Johnson, 2021].

1.5.4 φC31 Integrase - Practical Applications

As a site-specific recombinase, φC31 Int is an attractive tool for DNA modification [Nern *et al.*, 2011; Farrugio *et al.*, 2012]. While the oldest and simplest application of φC31 Int is site-specific insertion of large heterologous fragments [Stark, 2017], the enzyme can be used to facilitate various other genetic modifications, such as site-specific excisions or inversions of gene fragments (as shown in Figure 2) [Nern *et al.,* 2011] or even wide scale genomic rearrangements if coupled with other recombinase systems [Stark, 2017]. Because of the independence of φC31 Int activity from DNA conformation, it could be possible to use

it to transform a target genome with massive constructs containing several functional genes, as long as appropriate *att* sites are present.

It has been shown that, because of those properties, φC31 Int has the potential to become a tool for gene therapy. Its ability to predictably and reliably insert an entire plasmid into a chromosome provides a potential way to safely and permanently provide cells of a patient with a functioning copy of the gene. This would offset conditions caused by loss-of-function mutations [Chaves and Calos, 2011].

φC31 Int has found some more novel uses as well. It has been used in creation of biological logic gates and computing systems [Fogg *et al.*, 2014; Siuti *et al.*, 2014; Zhao *et al.*, 2019], as well as editable systems for storage of digital data within DNA [Bonnet *et al.*, 2012] and intracellular systems for archiving chemical and biochemical events in the living cell [Roquet *et al.*, 2016]. Those computational uses are enabled by possibility to flank promoters with *att* sites, allowing those promoters to be switched on and off by inverting their orientation (see Figure 2 B). The unidirectional nature of LSRs means that the orientation of the promoters (and thus the position of the molecular switch) can be tightly controlled, as reversing the process requires expression of the RDF alongside the Integrase [Fogg *et al.*, 2014].

1.6 Previous Research on LSR Coiled-Coil domains

Studies of CD deletion in φC31 Int show that the Int dimers are able to form synaptic tetramers in absence of CD. The same study found that inducing mutations into the CC severely disrupted the synapsis, suggesting that CC interactions are important to the stability of the synaptic complex [McEwan *et al.*, 2009].

Structural properties and interactions of CC domains in the LI Int were investigated, alongside providing updated, more accurate sequence alignments between various members of the serine integrase family (including φC31 Int) [Gupta *et al.*, 2017]. LI Int is a smaller protein than φC31 Int and thus was picked as a comparatively easier target for structural analysis. The paper provides structural, as well as genetic evidence (via targeted point-mutation) that isolated CC domains of LI Int are capable of interaction with each other *in vitro*. Several residues in the region between AA365 and AA380 of LI Int were identified

and determined to be crucial to this interaction. This region corresponds to the CC domain tip - a region between AA471 and AA493 of φC31 Int (Figure 5). A significant portion of those residues of interest are conserved between multiple members of the LSR family, providing further supporting evidence of the significance of the discovered CC-CC interaction, as well as a potential reference point for similar future examination in the other integrases. However, as the structural analysis of those interactions was conducted on the isolated CC domains rather than complete protein, it did not factor in potential interactions of CC domains with other Integrase domains, nor how the rest of the complex would affect the CC-CC interactions. The same study also presents strong evidence for the domain's involvement in control of directionality, namely its role in inhibition of the reverse reaction [Gupta *et al.*, 2017].

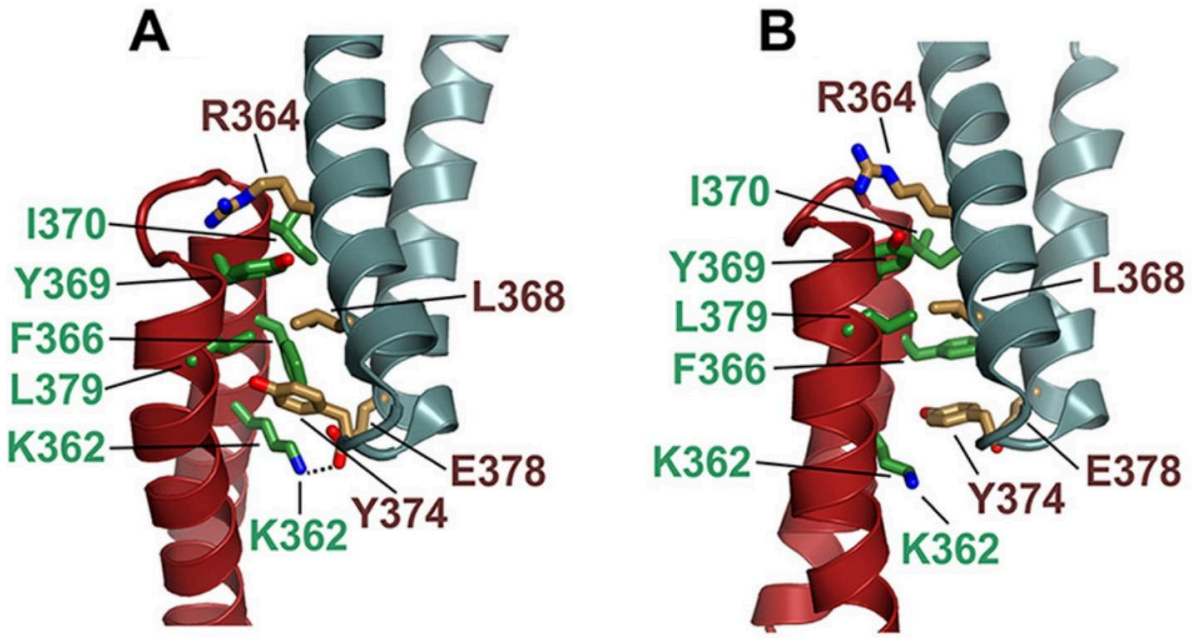


Figure 6. Structure of the CC domain ends and proposed interfaces of CC-CC interactions in LI Int. A and B represent two different modes of interaction which have been found in the crystal structure of the isolated CCs [Gupta *et al.*, 2017]. Note that the geometry of the interface and nature of the interactions between particular residues is shifted between structure A and structure B. In a WT Integrase this difference in CC-CC interaction could depend on the geometry of the complex, and therefore on the *att* site pair bound in the synapse. It is hypothesised that binding of the RDF at the base of the CC can facilitate a change of the geometry of the interaction, consequently shifting from inhibition to promotion of the recombination. Figure adapted from Mandali and Johnson, 2021.

An independent study [Mandali and Johnson, 2021] performed on A118 Int provides further evidence for the CC domain's involvement in directionality control. It studied residues at the tips of the CC domains (including those identified in the previously described experiment) as responsible for CC interactions in vitro. This experimental data was obtained

using alanine scanning of amino acids in the region AA362-AA383. It shows that residues responsible for interactions between coil-coiled domains are different depending on whether the synaptic complex is formed on *attP* x *attB* (substrate of the forward reaction) or attL x attR (substrate of the reverse reaction). This suggests that directionality is caused by two mutually exclusive types of CC domain interactions (Figure 6) [Mandali and Johnson, 2021].

Historically one of the hypothesised mechanisms for determination of LSR reaction directionality was formation of conformationally different complexes depending on the DNA binding sites [Smith *et al.*, 2010], which was proposed to explain how LSRs are able to have high binding affinity for all four DNA sites (*attP*, *attB*, *attL* and *attR*) but are able to recombine only *attP* x *attB* and *attL* x *attR* pairings, while seemingly ignoring all other possible pairings of the four sites. Those results support such a hypothesis, as different complex conformations would expose different parts of opposing CC domains to each other, allowing different types of interactions depending on the DNA binding sites. Those different interactions could in turn promote or inhibit the formation of the synapse and recombination [Mandali and Johnson, 2021]. Additionally, since the region of the CC domain in which both types of interaction were identified corresponds to the region identified previously [Gupta e*t al.*, 2017], it provides evidence that observed interactions were not just an incidental product of the purified domains, but are present in the complete protein as well.

The same study also describes how the deletion of CC domain inhibits synapse formation, and conducted experiments attempting to use designed gain-of-function mutants in other domains to facilitate recombination in the complete absence of CC domains. While some of the mutants form synaptic complexes successfully despite the CC deletion, no attempt yielded a successful recombination reaction, further cementing the importance of CC interactions for the recombination process [Mandali and Johnson, 2021].

A different study on φC31 Int used data obtained from observation of the recombination reactions to create a mathematical model of the reaction's kinetics and thermodynamics, and thus to predict the biochemistry of the reaction [Pokhilko *et al.*, 2016]. The resulting model predicts the formation of stable intermediate DNA-integrase complexes, which would be structurally different depending on the presence or absence of the RDF. It also presents experimental data disproving some of the previously made hypotheses on the mechanism of the directionality, such as inactivation of integrase mid-reaction, which is

important, as it allows narrowing down the search area for future research. While the paper admits to omission of some less kinetically significant steps for simplicity, and provides only a theoretical model of the hypothetical reaction, some of its predictions have already been supported by the cumulative results of the follow-up studies. One of its well supported predictions is the existence of two types of structurally distinct and thermodynamically stable synapses, depending on the presence or absence of the RDF. Studies supporting the hypothesis that RDF binding changes the structure of the φC31 Int complexes have been published [Fogg *et al.*, 2018]. Additionally, evidence exists that the structure of the A118 Int synaptic complex changes depending on the DNA sites bound in the synapse [Mandali and Johnson, 2021]. If the mode of stabilisation of those intermediate complexes is thermodynamical, as Pokhilko's model predicts, the CC domain interactions corresponding to those described in related integrases [Gupta *et al.*, 2017; Mandali and Johnson, 2021] might be responsible for making particular complexes favourable. It has been demonstrated that the CC domains are essential for integrase dimer stabilisation [Gupta *et al.*, 2017], making them one of the potential factors determining which of the hypothetical intermediates would form in the most energetically favourable way.

1.7 Aims of the Project

My research project aims to gather biochemical data on the role of particular AA residues within the CC domain of φC31 Int, using designed point mutations of particular residues to examine their importance to the reaction and its directionality. Using LSR family sequence alignments from Gupta *et al.*, 2017, I have identified the φC31 Int residues corresponding to the previously described CC domain interactions in LI and A118 integrases. Using those alignments I have selected L468, L471, E472, L474, Y475, V487, H491, F492, R493, Q496, L499, T500, E507 as residues of interest. Those particular residues were selected as they were either corresponding to LI/A118 Int residues responsible for CC-CC interactions identified by Gupta *et al.*, 2017 and Mandali and Johnson, 2021, or the LSR sequence alignments implied that they were heavily conserved within the family. I aimed to induce point mutations in some of those residues, in order to examine if their importance is conserved between the proteins and if so, how each of the residues affects integrase dimer and synapsis formation, recombination efficiency and specificity of directionality of the reaction. In the end, the investigation of one L474 mutant and two Y475 mutants, as well as of the complete CC deletion was conducted. Experiments included a study of reaction

efficiency, including a survey for anomalies related to directionality control, binding assays for study of synaptic complex formation, including unorthodox pairings of *att* sites, as well as a specificity study aiming to identify if any of the mutants is able to recombine a pair of *att* sites natively recognised by different LSRs.

The intended outcome is to add further detail to our understanding of the role of the CC domains and their interactions in the φC31 Int mechanism, as well to use the data to assess how closely those interactions and mechanisms are conserved in relation to the better structurally understood LI/A118 integrases.

2. MATERIALS AND METHODS

2.1 Bacterial strains

All bacteria used for cloning, transformation, plasmid amplification and protein expression were strains of *Escherichia coli*. The following strains of *E. coli* were used during the course of this project:

- **DH5** [*F-, φ80, Δ(lacZYA-argF), U169, hsdR17, (rk-, mk+), phoA, supE44, thi-1, gyrA96, relA1, λ-*]
- **DH5α** [*F-, φ80, lacZΔM15, Δ(lacZYA-argF), U169, recA1, endA1, hsdR17, (rk-, mk+), phoA, supE44, thi-1, gyrA96, relA1, λ-*]
- **BL21(DE3)pLysS** [*F-, ompT, hsdSB (rB-, mB-), dcm, gal, λ(DE3), pLysS, Cm^r*]

2.2 Growth media

All growth media were provided by the University of Glasgow College of Medical, Veterinary and Life Sciences prep room. The following liquid growth media and agar plates were used during the course of this project:

- L Broth
    - 5 g/L yeast extract powder
    - 10 g/L tryptone
    - 10 g/L NaCl
    - made up in water and autoclave sterilised at 121 ^oC for 20 min
- 2xYT Broth
    - 10 g/L yeast extract powder
    - 16 g/L tryptone
    - 5 g/L NaCl
    - adjusted to pH 7.0 using NaOH solution
    - made up in water and autoclave sterilised at 121 ^oC for 20 min
- L-Agar
    - As L Broth, with addition of 15 g/L of agar powder before autoclave sterilisation

2.3 Chemicals

The following chemicals and chemical solutions were used during the duration of this project. All stock solutions were made up with double-distilled water unless otherwise specified.

- Agarose powder
- Ammonium acetate (2 M solution)
- Ammonium persulphate (10% solution)
- Ampicillin (25 mg/mL solution)
- Calcium chloride ($CaCl_2$; 50 mM solution)
- Chloramphenicol (25 mg/mL solution in 100% Ethanol)
- DTT (dithiothreitol) (20 mM)
- Ethanol
- Ethidium bromide (10 mg/mL solution)
- Glycerol
- IPTG (isopropyl-β-D-thiogalactoside) (100 mM solution)
- Kanamycin (25 mg/mL solution)
- Magnesium chloride ($MgCl_2$; 100 mM solution)
- Phenylmethylsulfonyl fluoride (PMSF) (100 mM in 100% Ethanol; made up freshly before each use)

2.4 Enzymes

Following enzymes were used during the duration of this project. Unless otherwise stated, all enzymes were obtained from commercially available New England Biolabs catalogue.

- Acc65I
- AlwNI
- BamHI
- BssHII
- EcoRI
- gp3 (φC31 Int RDF) - produced on site by Arelene McQue using protocol 2.18
- Klenow polymerase

- KpnI
- NdeI
- NheI
- NruI
- SacII
- SpeI
- StuI
- T4 DNA Ligase
- Wild-type φC31 Integrase - produced on site by Arelene McQue using protocol 2.18
- XbaI

2.5 Buffers

The following buffers were used during the course of this project. Unless otherwise stated, water used in creation of all buffers was double-distilled.

- 4x Assembly buffer (used in Section 2.21)
    - 40 mM Tris-HCl, pH 7.5
    - 20 mM Spermidine 3HCl
    - 0.4 mM Ethylenediaminetetraacetic acid (EDTA)
    - 0.4 mg/mL Bovine serum albumin
- 5x Binding Buffer + $MgCl_2$ (used in Section 2.22)
    - 100 mM Tris-HCl, pH 7.5
    - 50% Glycerol, filter sterilised
    - 100 µg/mL Poly(deoxyinosinic-deoxycytidylic) acid
    - 0.5 mM DDT
    - 0.002% Bromophenol Blue
    - 10 mM Magnesium Chloride ($MgCl_2$), filter sterilised
    - made up in water, stored at -20 ºC

- Coomassie blue staining solution (used in Section 2.19)
    - 0.2% Coomassie Blue (Serva Blau)
    - 7.5% Acetic acid
    - 50% Ethanol
    - made up in water, stored at 4 ^{o}C
- 10x CutSmart™ buffer (commercially available from New England Biolabs) (used in Section 2.14)
- 10 mM dNTP mixture (used in Section 2.16)
    - 10 mM dATP
    - 10 mM dCTP
    - 10 mM dGTP
    - 10 mM dTTP
    - made up in water
- 1x Pellet Wash Buffer (PWB) (used in Section 2.18)
    - 20 mM Tris-HCl, pH 7.57
    - 10 mM Magnesium chloride ($MgCl_2$)
    - made up in water
- 6% Polyacrylamide gel (40 mL per one gel) (used in Section 2.22)
    - 4 mL of 10x TB Synapsis Buffer
    - 8 mL of 50% Glycerol, filter sterilised
    - 6 mL of 40% Polyacrylamide solution (Acrylamide/Bisacrylamide, 37:1 (2.6% crosslinker))
    - 480 μL of 10% Ammonium persulphate
    - 20 μL of Tetramethylethylenediamine (TEMED)
    - 21.5 mL of ddH_2O
- 1x Protein Purification Buffer A (Buffer A) (used in Section 2.18)
    - 20 mM Sodium phosphate (NaH_2PO_4; pH 7.4)
    - 1 M Sodium chloride (NaCl)
    - 1 mM DTT
    - 50 mM Imidazole
    - made up in water, filter sterilised

- 1x Protein Purification Buffer B (Buffer B) (used in Section 2.18)
    - 20 mM Sodium phosphate (NaH_2PO_4; pH 7.4)
    - 1 M Sodium chloride (NaCl)
    - 1 mM DTT
    - 500 mM Imidazole
    - made up in water, filter sterilised
- 1x Recombinase dialysis/dilution buffer (RDB) (used in Section 2.18)
    - 25 mM Tris-HCl, pH 7.5
    - 1 M Sodium chloride (NaCl)
    - 1 mM DTT
    - 50% glycerol
    - made up in water, filter sterilised
- 1x SDS-PAGE Loading Buffer (used in Section 2.19)
    - 5x Laemmli Loading Buffer
        - 470 mM Tris-HCl (pH 6.8)
        - 11.3% Sodium dodecyl sulphate
        - 62.5% Glycerol
        - 7.5% β-mercaptoethanol
        - 0.04% Bromophenol blue
    - 10 mM DDT
- 5x T4 DNA Ligation buffer (commercially available from New England Biolabs) (used in Section 2.17)
- 1x TAE (used in Section 2.12)
    - 40 mM Tris base
    - 20 mM Acetic acid
    - 1 mM EDTA
    - made up in deionized water as 50x concentration and diluted as needed
- 10x TB Synapsis buffer (used in Section 2.22)
    - 890 mM Tris base
    - 890 mM Boric acid ($B(OH)_3$)
    - made up in water, filter sterilised

2.6 Bacterial cultures

Plate cultures were grown on L-Agar plates poured in single-use plastic petri dishes at ~25 mL of agar per plate. Antibiotic selection was performed using either Ampicillin (at 100 μg/mL final concentration), Kanamycin (at 50 μg/mL final concentration) or Chloramphenicol (at 25 μg/mL final concentration) depending on the plasmid being selected for, and were always incubated overnight at 37 ºC in a static incubator.

Overnight cultures were grown in 6 mL of L Broth in reusable aerated glass flasks. Antibiotic selection was performed using either Ampicillin (at 100 μg/mL final concentration), Kanamycin (at 50 μg/mL final concentration) or Chloramphenicol (at 100 μg/mL final concentration) depending on the plasmid being selected for, and were always incubated overnight at 37 ºC and 250 rpm in a shaker incubator.

2.7 Electrocompetent cell preparation

20 mL of L Broth was inoculated with a single colony of DH5 or DH5α *E. coli* and grown overnight at 37 ºC and 250 rpm. After the overnight incubation of the initial 20 mL culture was completed, two growth flasks with 200 mL of L Broth were inoculated using 2 mL of the culture. Both flasks were grown at 37 ºC and 225 rpm until OD_{600} reached 0.5-0.6 (measurements taken using Thermoscienfic Bio Spectrophotometer), which took about 2.5 h of growth. Growth flasks were then cooled on ice for 20 min, after which the cultures were transferred to appropriate centrifugation vessels. Cultures were centrifuged at 4000 rpm for 15 min at 4 ºC using a pre-chilled rotor. Supernatant was discarded, and pellets were gently resuspended with 200 mL of ice-cold filter sterilised 10% (v/v) Glycerol per 200mL of original culture. Resuspended cells were centrifuged at 4000 rpm for 15 min at 4ºC. Supernatant was discarded, and pellets were gently resuspended with 100 mL of ice-cold filter sterilised 10% (v/v) Glycerol each. Resuspended cells were centrifuged at 4000 rpm for 15 min at 4ºC. Supernatant was discarded, and pellets were gently resuspended with 10 mL of ice-cold filter sterilised 10% (v/v) Glycerol each, transferred to pre-chilled 30 mL centrifuge tubes and centrifuged at 4000 rpm for 15 min at 4 ºC using a pre-chilled rotor. Cells were gently resuspended in 1 mL of ice cold-10% Glycerol per pellet and separated into 50 μL aliquotes, stored at -80 ºC.

2.8 Chemically competent cell preparation

5 mL L Broth culture with 25 μg/mL Chloramphenicol was inoculated with 100 μL BL21(DE3) pLysS *E. coli* glycerol stock and incubated at 37 ºC and 250 rpm for 1 h 55 min. 1 mL of liquid culture was aliquoted into 4 pre-chilled 1.5 mL Eppendorf tubes, and centrifuged in the cold room (4 ºC) for 2 min at 6000 rpm. Supernatant was discarded, and the cells were resuspended in 1 mL ice-cold 50 mM $CaCl_2$. Resuspended cells were centrifuged in the cold room (4 ºC) for 2 min at 6000 rpm. Supernatant was discarded, and the cells were resuspended in 100 μL ice-cold 50 mM $CaCl_2$. Ready competent cells were kept on ice in the cold room (4 ºC) for 1 h prior to use.

2.9 Electrocompetent cell transformation

Electrocompetent cell stock was defrosted and kept on ice. 1 μL of DNA samples was added to fresh Eppendorf tubes kept on ice. 30 μL of electrocompetent cell stock was added to each tube, and incubated on ice for 1 min. The transformation mixes were transferred into corresponding pre-chilled 1 mL electroporation cuvettes, while ensuring that liquid was touching both walls of the cuvette. Cuvette was inserted into an electroporation machine and subjected to a 2.5 V 6.10 ms pulse. Instantly after, 1 mL of L Broth was added to the cuvette. Mixture was transferred to a fresh, labelled Eppendorf tube and incubated for 1.5 h in a shaker incubator at 37 ºC and 250 rpm.

2.10 Chemically competent cell transformation

1 μL of DNA samples was added to fresh Eppendorf tubes and incubated on ice for 2 min. 50 μL of Chemically competent cell stock was added to each tube, aspirated and incubated on ice for 2 min. Cells were then heat shocked in a 37 ºC water bath for 5 min and immediately returned to ice to incubate for 2 min. 500 μL of L Broth was added after the incubation was completed, and the resulting mix was incubated in a shaker incubator at 37 ºC and 250 rpm for 1 h.

2.11 Plasmid DNA isolation

Plasmid DNA was isolated from overnight cell cultures using a commercially available Qiagen QIAprep Spin Miniprep Kit, according to the manufacturer's instructions. The resulting DNA was quantified using Nanodrop (Thermofisher).

2.12 Agarose gel electrophoresis

Agarose gels were prepared by dissolving Agarose powder in 1xTAE buffer. A standard gel concentration used was 1.2%, with a standard gel volume of 100 mL.

1/4th of the sample volume of loading dye was added to all the samples prior to loading the gels.

Gels were run using Horizon gel boxes, submerged in 1xTAE buffer, at 100 V. Standard run time varied between 1 h 45 min and 2 h.

2.13 Agarose gel staining and imaging

Freshly run Agarose gel was placed on a glass plate inside a 1 L plastic tray and the tray was filled with the 1xTAE buffer used during electrophoresis. The tray was transferred to a fume hood to add 40 μL of Ethidium bromide, and placed on a shaker platform. Shaker was set to shake for 30 min in slow, circular motion. After the shaking was completed, the Ethidium bromide solution was discarded and the gel was washed with distilled water. The gel was then placed in a gel imaging system to take pictures. BioRad Gel Doc™ XR+ was used to take all agarose gel images for this project.

2.14 Restriction digest

Standard restriction digest reaction mix was as follows:

- 5 units of enzyme per 10 μL final volume (5 units per enzyme if multiple enzymes used)
- 1/10th of the final volume of 10x CutSmart™ buffer (substituted by final concentration of 10 mM Magnesium chloride in case of NruI digest of Integrase reaction assay samples)

- Approximately 0.5 μg of plasmid DNA, or 10 μL of Integrase reaction assay samples
- Fill to final volume with ddH_2O

Reactions were run in 37 ºC water baths for 1 h or 2 h depending on enzymes used. Since most reactions were immediately run on an Agarose gel after completion (either for DNA gel purification or for data production), no heat inactivation of restriction enzymes was necessary unless otherwise advised in other listed protocols.

2.15 DNA gel purification

Bands of interest were excised from Agarose gels under long-wavelength UV transilluminator using fresh scalpel blades and placed in fresh Eppendorf tubes. Gels were not imaged prior to purification to avoid any mutations caused by short-wavelength UV of the imaging system.

Purification was conducted using commercially available Qiagen QIAquick Gel Extraction Kit, according to the manufacturer's instructions.

2.16 Klenow polymerase reaction

If plasmid DNA was digested in preparation for blunt end ligation (eg. a portion of the plasmid needed to be excised), sticky ends were filled in before the gel purification using the Klenow polymerase reaction. Restriction enzymes were deactivated by incubation at 80 ºC for 20 min. Samples were then cooled at room temperature for 5 min, and dNTP mixture was added, to the final concentration of 125 μM. Afterwards, Klenow polymerase was added(1 unit per 20 μL reaction volume), and the reactions were incubated at 37 ºC for 15 min. Enzyme was deactivated by incubation at 70 ºC for 5 min, and the mix was cooled to room temperature before using in DNA ligation reactions.

2.17 DNA Ligation

Both sticky end ligation and blunt end ligation were conducted using T4 DNA Ligase in a following reaction mix:

- Up to 8 µL of DNA mix consisting of 3:1 molar proportion of insert to vector (or 16 µL of Klenow polymerase reaction product if re-circularising a plasmid by blunt end ligation)
- 2 µL of 5x T4 Ligation buffer (doubled for blunt end ligation)
- 1 µL of T4 Ligase (doubled for blunt end ligation)
- Fill to 10 µL with ddH_2O if needed (20 µL for blunt end ligation)

2.18 Enzyme production and purification

The φC31 Int production and purification protocol was adapted from Olorunniji et. al, 2017, pages 306 to 312. Purification of produced protein was conducted using nickel column affinity chromatography on an HPLC machine. The adapted protocol used here is as follows:

1. Introduce the expression plasmid into *E. coli* strain BL21 (DE3) pLysS via chemically competent cell transformation (see 2.10). Spread transformation reactions on L-Agar plates containing 25 µg/mL Chloramphenicol and 50 µg/mL Kanamycin. Incubate overnight at 37 °C.
2. Use a fresh single transformant colony to inoculate a starter culture in 2xYT broth (10 mL) supplemented with 50 µg/mL Kanamycin and 25 µg/mL Chloramphenicol. Grow overnight (~16 h) at 37 °C with shaking (225 rpm).
3. While cells are growing overnight, pre-warm 400 mL 2xYT broth to 37 °C by incubating overnight in sterile bottles.
4. The next day, pour the pre-warmed 2YT broth into sterile 2 L conical flasks with gas-permeable (cotton wool or similar) plugs. Add 0.4 mL each of Kanamycin and Chloramphenicol stock solutions (see Section 2.3). Inoculate with the expression strain by adding 4 mL of the fresh overnight culture prepared in Step 3.
5. Grow the culture at 37 °C in a suitable shaking incubator. Shake continuously (250 rpm), and monitor OD_{600} every 20-30 min until it reaches 0.5−0.6. This typically takes 2-3 h.
6. Take 800 µL samples of the cultures for protein gel analysis by SDS-PAGE as a pre-induction control.
7. Cool the main culture rapidly to below 20 °C by placing it in an ice bath. Induce protein expression by adding IPTG stock solution to the culture, to a final concentration of 0.5 mM IPTG.

8. Grow for a further 16 h in a refrigerated shaking incubator at 20 °C, with continuous shaking at 250 rpm.
9. Take 800 µL samples of the cultures for protein gel analysis by SDS-PAGE. Measure OD_{600} of taken samples before proceeding. If OD_{600} exceeds 0.8, dilute 5-fold using 2xYT Broth.
10. Pellet 200 µL of samples from steps 6 and 9 by centrifuging, and freeze the pellets. When ready to perform SDS-PAGE analysis, resuspend those pellets in 100 µL of SDS-PAGE loading buffer, boil at 95 ^{o}C for 5 min, and proceed as per Chapter 2.19.
11. Split the main culture into two separate aliquots of 200 mL. Pre-weigh 200 mL centrifuge tubes before the transfer of the liquid culture to allow the wet weight of the harvested cell pellet to be determined after centrifugation. Centrifuge the culture at 9800 g for 10 min at 4 °C.
12. Discard the supernatant and resuspend the pellets in 100 mL Pellet Wash Buffer (PWB). Use a clean glass rod or gently pipette up and down to completely suspend the cell pellet. Keep the sample on ice as much as possible.
13. Centrifuge at 9800 g for 10 min at 4 °C to re-pellet the cells. Discard the supernatant.
14. Weigh the cell pellet. Pellets can be frozen at -20 °C or -70 °C for future use.

Steps 15-19 are performed on ice.

15. If the pellet was frozen, thaw it in the centrifuge tube at room temperature, and transfer to an ice bucket as soon as the pellet is completely thawed.
16. To proceed with protein extraction and purification from the cell pellet, add 25 mL ice-cold Buffer A to the cell pellet per each 2-3 g of the pellet. Mix thoroughly but gently to suspend the cells into a smooth homogenous paste without lumps. This can be done by pipetting up and down while disrupting the cell pellet using a 10 mL pipette. Alternatively, use a clean glass rod to stir. Avoid introducing bubbles into the suspension.
17. Transfer the cell suspension to a clean pre-cooled container suitable for sonication The container must be narrow enough so that the cell suspension is more than 3 cm deep, for efficient sonication.
18. Immerse the sonicator probe in the cell suspension to a depth of at least 1 cm and sonicate in 3 short pulses of 20 seconds at 30% amplitude, until the cells are completely lysed. Keep the sample on ice throughout and cool for 2 min between each pulse. Add PMSF to a final concentration of 1 mM immediately after the first

sonication step. Ensure that the sonicator does not introduce excessive air into the lysate.

19. Transfer the sample to a 50mL centrifuge tube and centrifuge at 48000 g for 30 min at 4 °C. Collect the supernatant carefully, avoiding the pellet, and transfer to a suitable container. Store on ice until ready to load onto the nickel affinity column.
20. Ni^{2+} affinity column chromatography is used to purify the produced protein from the crude *E. coli* extracts on a standard HPLC system for protein liquid chromatography following the manufacturer's instruction manual. Place one pump inlet in a bottle containing 500 mL of Buffer A and the other pump inlet in a bottle containing 500 mL of Buffer B. The alternative inlet to the first pump should be connected to a bottle of ddH_2O.
21. Run the pump wash purifier programme on all inlets to flush out the inlet lines. Set up the spectrophotometer to monitor absorbance at 215, 260, and 280 nm and then connect a pre-packed 1 mL HisTrap HP Ni^{2+} column to the system.
22. Run five column volumes (5 mL) of water through the column at a flow rate of 1 mL/min, to remove the preservation buffer.
23. Pass five column volumes (5 mL) of Buffer A through the column at a constant flow rate of 1 mL/min.
24. Pass five column volumes (5 mL) of Buffer B through the column at a constant flow rate of 1 mL/min.
25. Pass at least ten column volumes (10 mL) of Buffer A through the column at a constant flow rate of 1 mL/min.
26. Just prior to loading the cell extract onto the column, filter it through a 0.22 μm cellulose acetate syringe filter attached to a 25 mL syringe. Fill the syringe completely with supernatant from step 19, using a wide-bore needle to reach the bottom of the tube if necessary. Retain ~200 μL of sample for later analysis by SDS PAGE. Connect the syringe filter to the syringe, and filter the sample into a fresh container. Wear eye protection in case the syringe becomes disconnected and extract leaks out at high pressure.
27. Load the filtered protein extract from step 26 onto the nickel column at a flow rate of 1 mL/min. Collect the flow-through and store it at -20 °C for later analysis by SDS-PAGE. Wash the column with a total of 25 mL Buffer A (or more) at a flow rate of 1 mL/min, until a steady baseline absorbance is reached at 260 nm and 280 nm.

28. Elute the His-tagged protein with a gradient from 50 mM to 500 mM Imidazole: set the pump to deliver a linear gradient from 100% Buffer A to 100% Buffer B over 25 minutes at a flow rate of 1 mL/min.
29. Start fraction collection (1 mL fractions) at the start of the gradient elution. Collect all fractions and store them on ice.
30. Run samples of each fraction (1-10 μL) on SDS-PAGE along with protein markers to select fractions to dialyse. Fractions that correspond to peaks with high A280/A260 ratio should be selected for further processing, since these are indicative of the presence of proteins in the fraction. If desired, pool similar fractions together for dialysis.
31. Transfer selected fractions (1-3 mL) into dialysis cassettes and dialyse against 500mL of Recombinase Dialysis/Dilution Buffer (RDB) for at least 6 h at 4 °C. Mix the buffer with gentle rotation (100 rpm), using a magnetic stirrer. Change the buffer by transferring the cassettes into fresh RDB (500 ml), and dialyse for a further 6 h at 4 °C. Sample volume will decrease due to the glycerol in the dialysis buffer.
32. Collect the dialysed protein solutions. Centrifuge at 10000 rpm for 1 min. Transfer the supernatants to fresh tubes and store at -20 C, discard the pellet. Proteins are used without removing the His-tags.
33. Measure A_{280} of protein solutions (or 1 in 10 dilutions in RDB). Zero the spectrophotometer using RDB as blank. Estimate the protein concentration using the calculated molar extinction coefficient for the φC31 Int (78755 M^{-1} cm^{-1}, at 280 nm). Make stock dilutions of proteins at 16 μM, 8 μM, 4 μM and 2 μM and store in screw-capped microfuge tubes at -20 ºC.

2.19 SDS-PAGE

SDS-PAGE was run using commercially available pre-cast SDS-polyacrylamide gels (Thermofisher), according to the manufacturer's instructions, with the following modification:

- 100 μL of 1x SDS-PAGE Loading Buffer was added to the samples before loading. 10 μL of the resulting mix was loaded onto the gel.

A standard SDS-PAGE gel was run for 40 min at 180 V. Gels were stained by incubation in a Coomassie blue staining solution for 20 min in a low-speed shaker.

2.20 Ethanol-mediated DNA precipitation

100% Ethanol was added to the DNA samples in amounts equalling to half the volume of the original sample. 2 M Ammonium acetate was added to the mix, in amounts equalling tenth of the volume of the original sample. The samples were mixed by vortexing and incubated at -20 ºC for 15 min. Samples were then spun down at maximum rpm for 30 min. Supernatant was carefully discarded, and the DNA pellet was resuspended in a desired volume of ddH_2O.

2.21 *In vitro* Integrase reaction assay

1. Prepare reaction mastermix (amounts shown per one sample, final volume 20 µL per sample):
    - 500 ng of plasmid DNA containing desired pair of *att* sequences
    - 5 µL of 4x Assembly Buffer
    - fill to 20 µL with ddH_2O
2. Aliquot the mastermix into Eppendorf tubes (20 µL per sample).
3. Add 2 µL of desired Integrase stock to the corresponding sample. Use 2 µL of RDB as negative control.
4. If testing the reaction in the presence of RDF, add 2 µL of the desired RDF stock to the corresponding samples.
5. Mix by vortexing. Gently spin down with a 3 second pulse.
6. Incubate in a 30 ºC water bath for 1 h.
7. Heat inactivate the reaction at 70 ºC for 2 min.
8. Separate the sample into two 11 µL aliquots. Set aside one aliquot to run on the Agarose gel.
9. Restriction digest of recombination reaction products is necessary to separate any product DNA molecules locked together in a catenane. Proceed with the restriction digest as per Chapter 2.14. NruI was used for experiments using pFM16 and pFM52.
10. Run the digest samples on Agarose gel as per Chapter 2.12 and image as per Chapter 2.13.

2.22 Binding assay

1. Clean and assemble a 40 mL polyacrylamide gel cast (glass plates, rubber seals and plastic spacers). Use 1.5 mm and a 20 well comb.
2. Pour freshly mixed 6% Polyacrylamide Gel (see Chapter 2.5) into the cast. Leave to set for 1 h.
3. Transfer the gel to a cold room (4ºC). Assemble the vertical electrophoresis kit in the cold room, and fill both chambers with 1x TB Synapsis Buffer. Remove the comb and flush out the wells using a syringe.
4. Pre-run empty gel at 200 V for 30 min.
5. Prepare the binding mixture (25 µL needed for each reaction):
    - 1/5th of the final volume of 5x Binding Buffer + $MgCl_2$, for a final concentration of 1x
    - 10 nM final concentration of Cy5' labelled *att* site dsDNA linear fragment
    - 20 nM final concentration of unlabelled, different *att* site dsDNA linear fragment
    - Fill with ddH_2O to 25 µL per each reaction
6. Aliquot 25 µL of the binding mixture into fresh 0.5 mL Eppendorf tubes. Add 2.8 µL of corresponding Integrase stocks. Mix by aspiration and briefly spin down.
7. Incubate the reaction for 30 min at 30ºC. Store samples on ice after the incubation is completed.
8. Flush out the wells of the pre-run gel using a syringe. Load 20 µL of each reaction onto the gel and run for 4 h at 200 V in the cold room (4ºC).
9. Carefully extract the gel from its cast, leaving only the back plate. Transfer the back plate with polyacrylamide gel into an imaging system to take pictures. The Typhoon™ FLA system was used for imaging of all binding assay gels.

3. RESULTS

3.1 Choice of the samples and the hypothesis

As described in more detail in Chapter 1.7, 13 residues of interest were identified by initial comparative analysis of available LSR data. However, because all mutant enzymes surveyed in this project needed to be cloned, expressed and purified on-site (see Chapter 2.18; SDS-PAGE gel images obtained during the purification process can be found in Appendix E1 and E2), it was impossible to produce and examine mutants of all identified residues due to project's time constraints. Instead, residues L474 and Y475 were chosen as the main focus of the study for two reasons. Firstly, the previous LSR alignments predict their position to be near the tip of the αK helix of the φC31 Int CC Domain (see Figure 5), which is a region which was identified as important to CC-CC interactions [Rutherford and Van Duyne, 2014; Gupta *et al.*, 2017]. Secondly, those residues are thought to correspond to LI Int L368 and Y369, both of which have been shown to reside at the core of the CC-CC interface and are known to facilitate CC-CC interactions of that Integrase (see Figure 6) [Gupta et al., 2017; Mandali and Johnson, 2021], allowing for easy comparison of results with existing LI Int data. In addition to alanine mutants of both residues, Y475H was created and examined as well. The main motivation in examining the histidine mutant was a unique set of overlapping properties between histidine and tyrosine - both residues have a large, aromatic sidechain, however, tyrosine is hydrophobic, while histidine is positively charged. This partial overlap in properties allowed to narrow down which properties of the tyrosine sidechain are important for its role in CC-CC interactions (more detailed discussion of differences between the two residues, as well as analysis of the results in light of those differences can be found in Chapter 4.1.3). Additionally, Y475H mutant has been examined in an unrelated study of the Stark-Colloms research group in the past, meaning that the DNA sequence of the mutant was already available in the laboratory DNA banks.

The hypothesised outcome of the study was that L474A, Y475A and Y475H mutants would show decreased recombination efficiency in comparison to the WT φC31 Int. Additionally, since geometry of the CC-CC interface differs depending on the direction of the reactions (see Figure 6), it also appeared plausible that the loss of efficiency induced by those mutations might differ depending on reaction conditions, with especially interesting potential

result being partial or total loss of the unidirectionality or partial or total loss of the ability to conduct RDF-induced reverse reaction. It is known from previous studies that ΔCC mutants of A118 Int are unable to facilitate recombination under any conditions [Mandali *et al.,* 2017], therefore no activity was expected from the surveyed ΔCC φC31 Int mutant as well.

3.2 Recombination reaction efficiency study

The recombination efficiency of ΔCC, L474A, Y475A and Y475H mutants was studied using *in vitro* Integrase reaction assay (see Chapter 2.21), alongside the WT φC31 Int as a positive control. Products of the reaction were digested using NruI in order to linearise the resulting plasmids and separate any created catenanes. Plasmids pFM16 and pFM52 (see Appendix C.2 and C.3) were used as *attP* x *attB* and *attL* x *attR* testing plasmids respectively (Figure 7). The used sequences of the *att* sites can be found in Appendices A.1 through A.4.

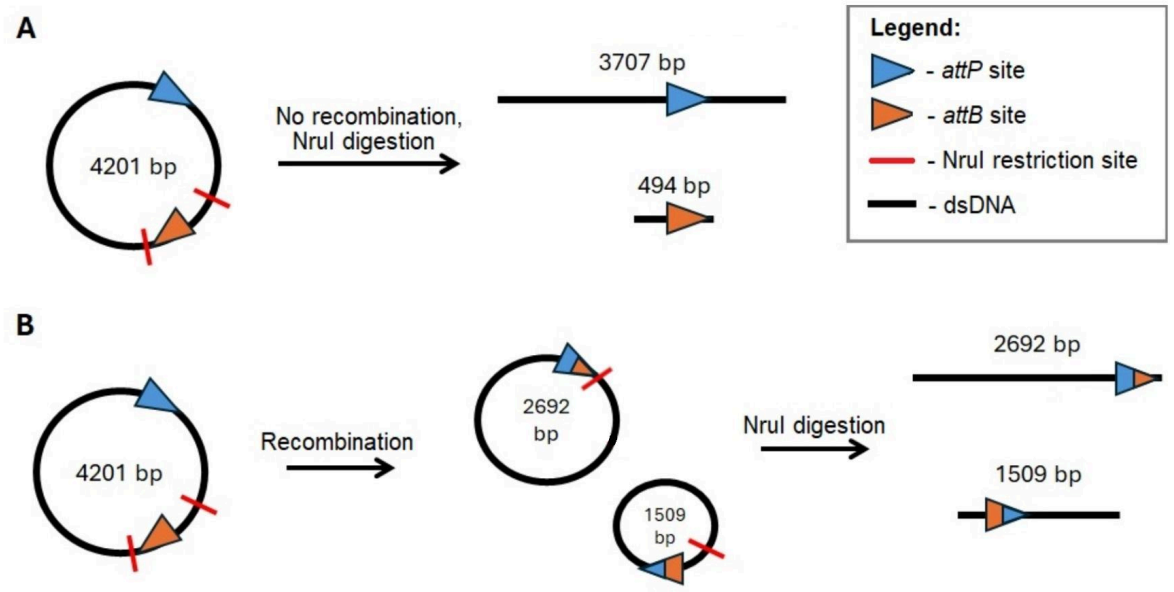


Figure 7. A diagram of substrate plasmids used in the *in vitro* Integrase recombination efficiency assay, and the way they were processed during the experiment. Diagram based on the example of pFM16 (*attP* x *attB* substrate). A - the linear DNA products resulting from NruI digestion if no recombination has occured. B - the product plasmids created from recombination of the substrate plasmid, and the linear products of their digestion using NruI. Note the size difference between products of A and B. The assay using *attL* x *attR* substrate followed the same pattern, as the only difference between sequences of pFM16 and pFM52 is presence of *attR* and *attL* sites in pFM52 at the respective location of *attP* and *attB* in pFM16. Full plasmid maps can be found in Appendix C.2 and C.3.

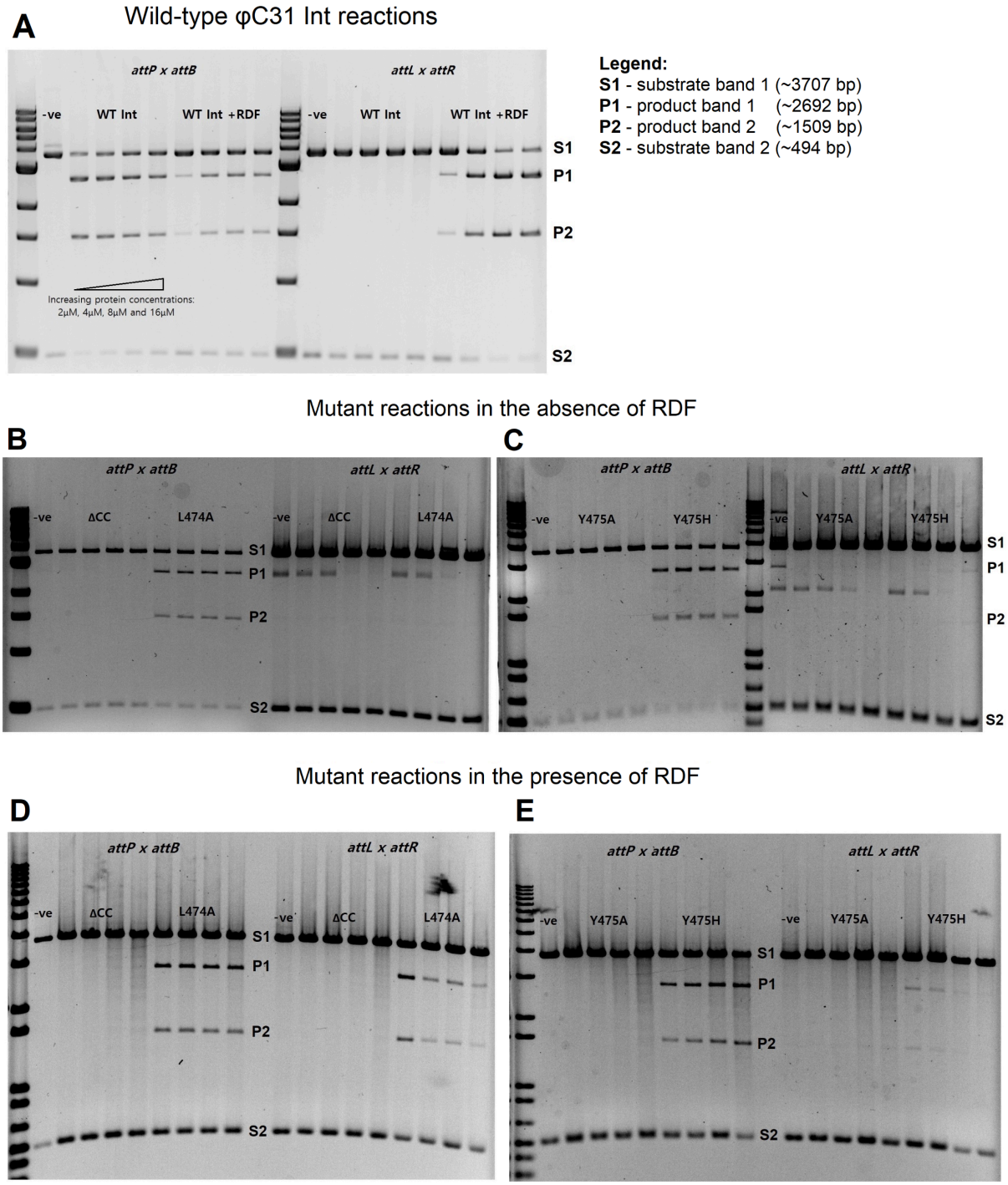


Figure 8. Gel images produced by the *in vitro* Integrase recombination efficiency assay. Bands corresponding to the sizes of the digested fragments of substrate plasmids (see Figure 7 A) are labelled as S1 and S2. Bands corresponding to the sizes of linearised products are labelled as P1 and P2 (see Figure 7 B). Each φC31 Int variant was tested in four concentrations (2μM, 4μM, 8μM and 16μM), loaded onto the gel in ascending order (see A).

A - WT φC31 Int recombination results. WT Int is able to recombine *attP* x *attB* substrate plasmid both in absence and in presence of RDF, albeit its efficiency in presence of RDF is hindered (as evident by fainter product bands). *attL* x *attR* recombination is not occuring at all in absence of RDF, but is very efficient in its presence, with substrate bands becoming visibly fainter at higher enzyme concentrations, implying substrate depletion.

Figure label continued on the next page.

Figure 8. ***continuation.***
B - Recombination of ΔCC and L474A mutants in absence of RDF. The ΔCC mutant shows no signs of activity under any conditions. L474A is able to facilitate recombination of *attP* x *attB* substrate plasmid to an easily observable degree. *attL* x *attR* results are inconclusive as negative control displays anomalous banding (presumably due to very high concentration of substrate plasmid in the reaction mix). However, distinct lack of P2 band suggests that no reaction occured in any sample.
C - Recombination of Y475A and Y475H mutants in absence of RDF. Y475A mutant shows no signs of activity under any conditions. Y475H is able to recombine *attP* x *attB* substrate. *attL* x *attR* reactions show the same anomalous banding as samples in gel B. However, a very faint product banding can be spotted in the 8μM and 16μM samples of Y475H, implying that, unlike WT, this mutant is able to perform some trace recombination of this substrate in absence of RDF.
D - Recombination of ΔCC and L474A mutants in presence of RDF. The ΔCC mutant shows no signs of activity under any conditions. L474A shows some activity on both *attP* x *attB* substrates and *attL* x *attR* substrates, however, the reaction efficiency seems to be much lower for *attL* x *attR* recombination, which is a contrast to the pattern observed in the WT.
E - Recombination of Y475A and Y475H mutants in presence of RDF. Y475A mutant shows no signs of activity under any conditions. Y475H is able to recombine *attP* x *attB* to a meaningful degree, however it shows only trace activity on *attL* x *attR* substrates.

The resulting products were run on agarose gels (Figure 8). Qualitative analysis of those gel results imply that ΔCC and Y475A are unable to facilitate the recombination under any conditions. L474A and Y475H showed a degree of activity on *attP x attB* substrate, both in the presence and absence of RDF. Interestingly, both mutants seem to favour *attP x attB* recombination even in the presence of RDF, with Y475H showing only extremely faint evidence of *attL x attR* recombination even in presence of RDF.

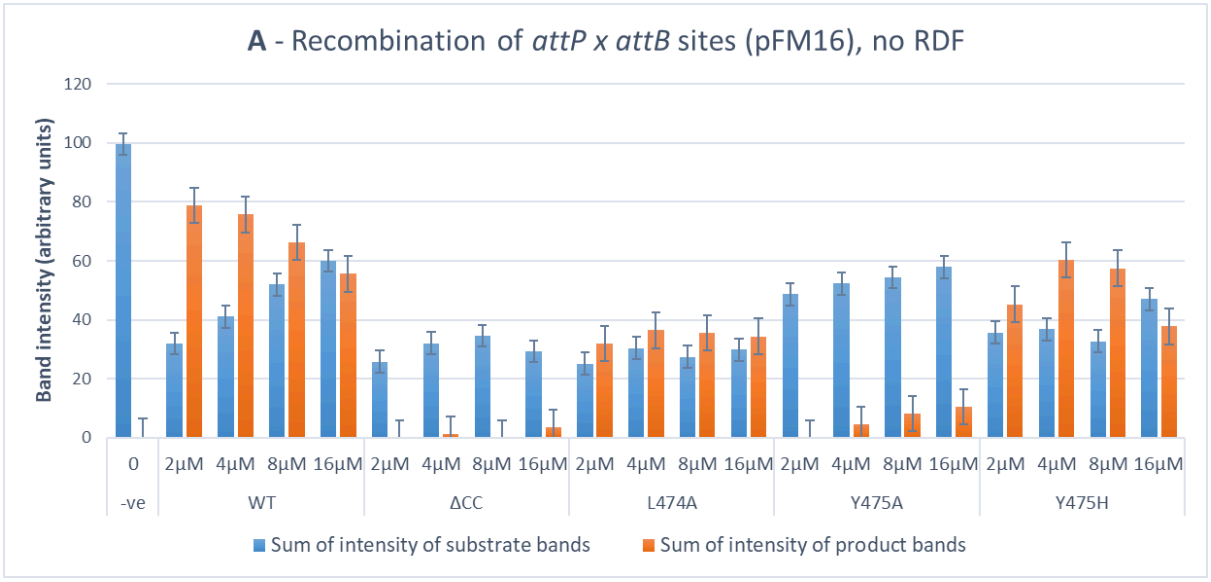
The decrease of recombination efficiency in WT control on *attP* x *attB* substrates with rising integrase concentration in absence of RDF is assumed to be a result of the enzyme oversaturation and consequent aggregation of the dimers into inactive clumps. The opposite case can be observed during the WT reaction on *attP* x *attB* substrates in the presence of RDF. The reaction efficiency increases as the concentration of integrase rises. As RDF concentration in all experiments has been kept at a constant 4 μM, and it is known that RDF facilitates the directionality control via direct binding to the integrase, it is not surprising that the rising integrase concentration was able to overcome its inhibitory properties.

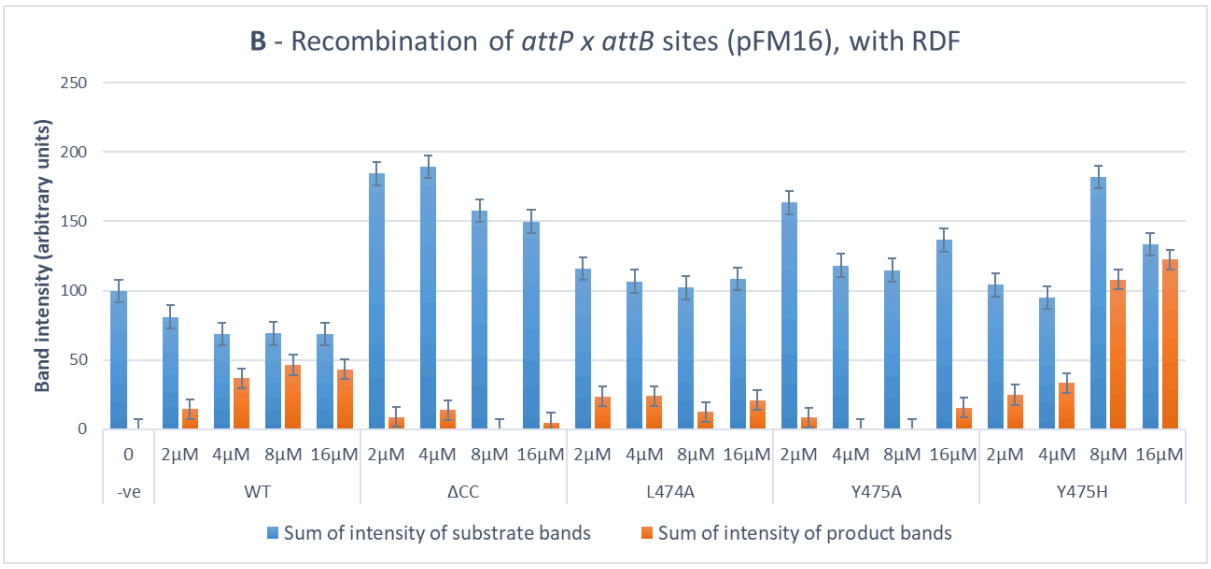
The nature of anomalous banding present in the negative control samples of *attL* x *attR* reactions in Figures 8 B and 8 C is not known. The assumption is that some of the substrate DNA degraded during the course of experiment, potentially due to very high concentration of the plasmid in the sample or contamination of the sample with unrelated buffers or substances. Due to the size of the anomalous band in Figure 8 B being very close

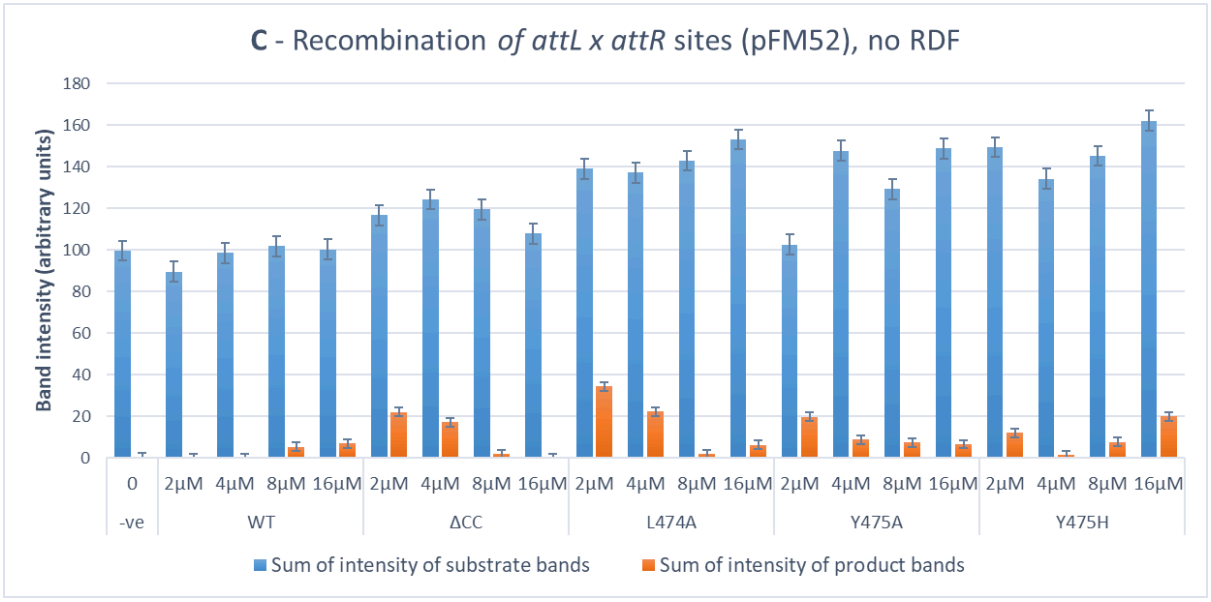
to the size of P1, the reaction efficiency for that experiment was assessed based on the presence (or absence) of the band P2. Anomalous banding in Figure 8 C does not correspond to any relevant band size, and thus is assumed to be an unrelated DNA degradation.

The band intensities of the gels presented in Figure 8 were measured using ImageJ software, in order to estimate the reaction efficiency. The relationship between product and substrate bands is presented in Figure 9. Note that this method of measuring yield is not accurate, and should only be treated as a rough approximation. This is caused by significant variance in background intensity of the gel images, which prevents from establishing an accurate baseline to which the band intensities should be compared. For example, Figures 9 A to C all imply some level of activity in some of the ΔCC mutant reactions. Qualitative data (Figure 8), previous results obtained in the Stark-Colloms lab and the preexisting literature all confirm that ΔCC LSRs are unable to conduct any recombination. Similarly, Figure 9 C implies trace recombination by WT control, which is not visible on Figure 8 A. Both those results are therefore interpreted as an error caused by the background variance. Raw data resulting from those measurements can be found in Appendix D.1.

The ratio of intensities of the substrate and product bands was used to estimate the reaction yield (Figure 10). Similarly to data in Figure 9, those results serve mostly as an approximation and for visualisation purposes, as they were obtained from the same error-prone band intensity measurements. WT Integrase shows efficiency between 50% and 80% in standard conditions for forward and reverse reaction (*attP x attB*, no RDF and *attL x attR*, with RDF respectively). Both L474A and Y475H appear to have significantly reduced efficiency of *attL x attR* reaction in the presence of RDF in comparison to WT, with Y475H efficiency falling as low as ~6.85% at higher enzyme concentration. However, Y475H also appears to resist inhibition of *attP x attB* reaction in the presence of RDF, with no reduction of efficiency at Int concentration of 16 μM.


A - Recombination of attP x attB sites (pFM16), no RDF
Band intensity (arbitrary units)
120
100
80
60
40
20
0
0 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM
-ve WT ΔCC L474A Y475A Y475H
Sum of intensity of substrate bands
Sum of intensity of product bands


B - Recombination of attP x attB sites (pFM16), with RDF
Band intensity (arbitrary units)
250
200
150
100
50
0
0 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM
-ve WT ΔCC L474A Y475A Y475H
Sum of intensity of substrate bands
Sum of intensity of product bands


C - Recombination of attL x attR sites (pFM52), no RDF
Band intensity (arbitrary units)
180
160
140
120
100
80
60
40
20
0
0 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM 2μM 4μM 8μM 16μM
-ve WT ΔCC L474A Y475A Y475H
Sum of intensity of substrate bands
Sum of intensity of product bands

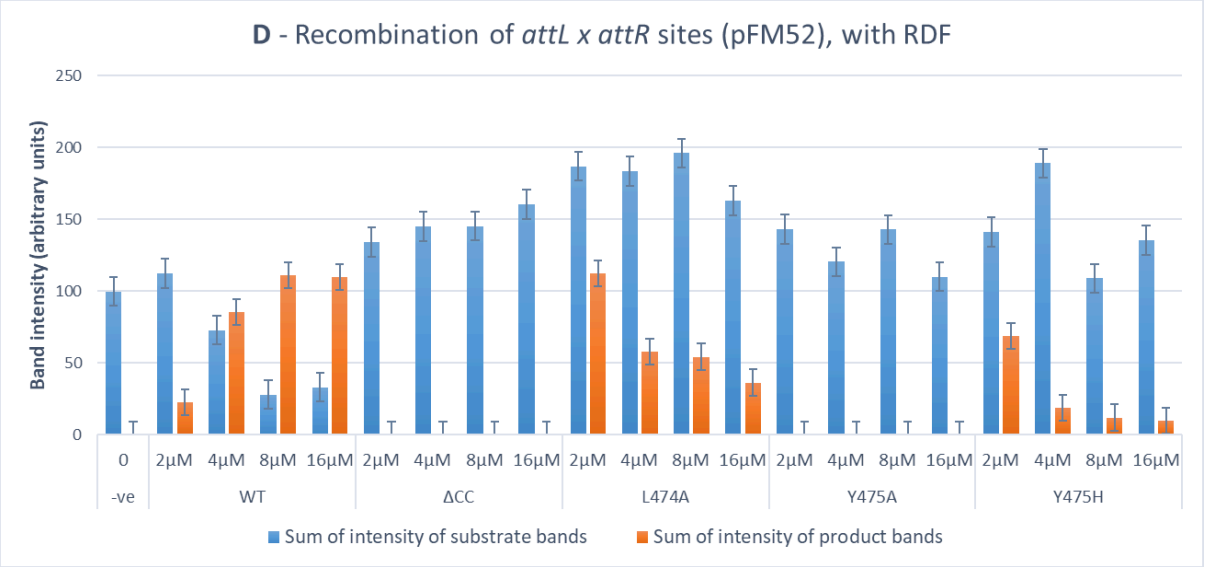


Figure 9. Comparison of band intensity measurements of the Integrase reaction efficiency assay results. Sum of intensity of substrate bands represented in blue, sum of intensity of product bands represented in orange. Note the implied presence of product bands in lanes which qualitatively did not seem to have any banding. This is a symptom of the degree of inaccuracy of presented data - high variance in intensity of the image background makes it difficult to compare bands, even within a single lane. As such, all data presented in this figure should be treated as approximations rather than absolute values.

A - Recombination of the *attP x attB* sites, in absence of RDF. WT, L474A and Y475H all show the ability to recombine over half of the provided substrate, with peak efficiency reached by 2 μM WT sample. ΔCC and Y475A show negligible levels of activity.

B - Recombination of the *attP x attB* sites, in presence of RDF. Visible drop of activity in WT and L474A samples in comparison to dataset A. Y475H seems to suffer a much lesser drop of activity, with levels of activity at higher concentrations comparable to the ones measured in absence of RDF.

C - Recombination of the *attL x attR* sites, in absence of RDF. Negligible levels of activity in all samples except low concentration L474A, which shows some level of recombination. Inaccuracy of the method prevents definitive confirmation of the faint bands noted in Figure 8 E, as the apparent level of activity is comparable to ΔCC and Y475A samples, both of which displayed no product banding visible to the naked eye.

D - Recombination of the *attL x attR* sites, in presence of RDF. High levels of efficiency displayed by high concentration WT samples. Both L474A and Y475H show activity, but significantly reduced in comparison to WT and their own performance against *attP x attB* sites.

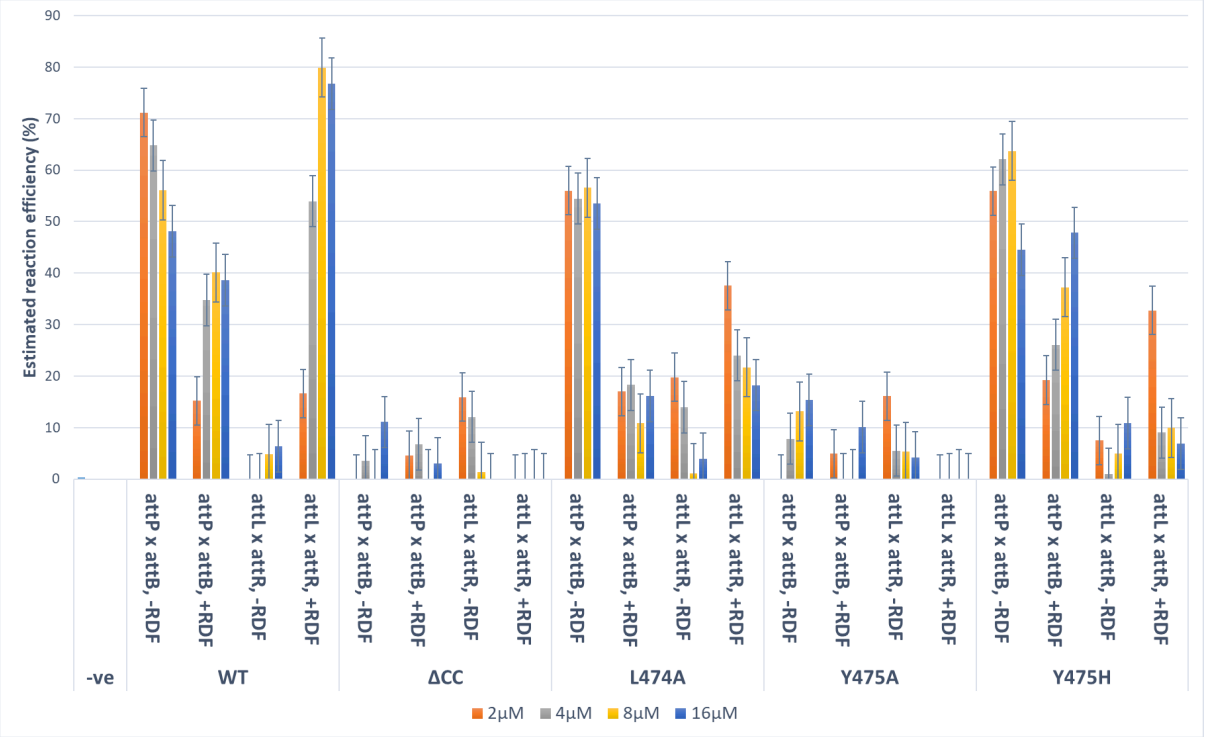


Figure 10. Representation of the reaction efficiency, calculated using yields approximated from the gel band intensity. Integrase concentrations are indicated at the bottom. As band intensity measurements are very error-prone, this data should be treated as a rough estimate. Both active mutants (L474A and Y475H) show significant drop of activity on *attL* x *attR* substrates in the presence of RDF in comparison to WT.

3.3 Binding assay

Integrase DNA binding, dimer formation and tetramer formation/synapsis were studied using a binding assay (see Chapter 2.22). Polyacrylamide gels were run at 4 ºC and in absence of denaturing agents, in order to preserve the complex integrity during electrophoresis. Results of the control experiment for this assay, using the S12A φC31 Int mutant, can be found in Figure 11. S12 is the catalytic serine residue of the CD of φC31 Int. Substitution of this serine residue with alanine results in a mutant which is unable to cleave the DNA after synapse formation, effectively preventing the reaction from proceeding past synapsis. This results in higher accumulation of synaptic complexes in solution than in the case of WT Int (Figure 11). This is because normally the synapse dissociates once recombination is complete (Figure 4).

Figure 12 presents the result of a binding assay conducted with the ΔCC mutant. An increase in the amount of bound monomers in proportion to bound dimers can be observed.

ΔCC appears to be able to form synaptic complexes, albeit with slightly lesser efficiency than observed in S12A, a likewise inactive mutant which was used as the control. Additionally, its appears to form synapses on non-canonical *att* site pairings *attP* x *attL* and *attP* x *attR* much more readily than the S12A, which is in line with findings from preexisting studies on φC31 Int kinetics [Chen *et al*., 2023].

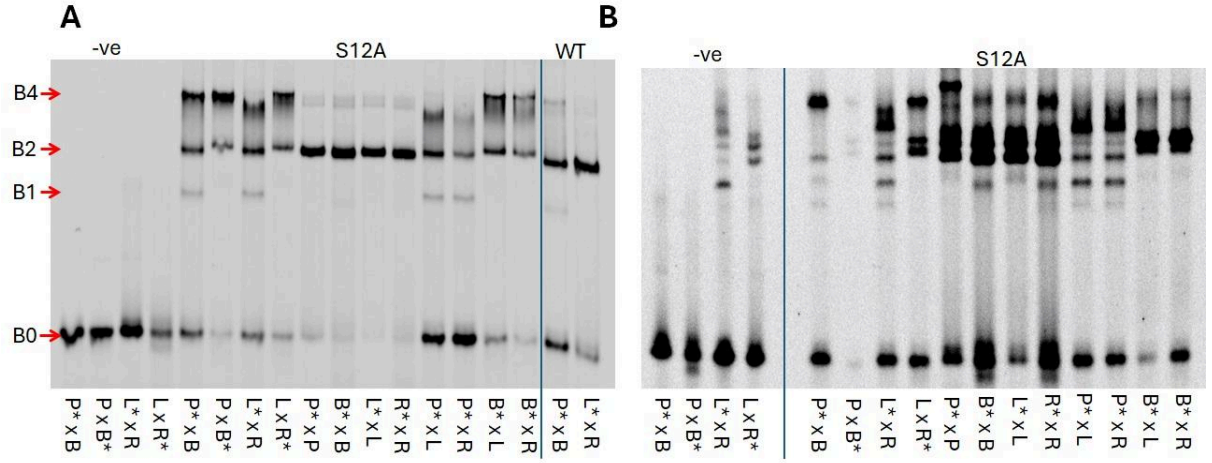


Figure 11. Control experiment for the binding assay. S12 is the catalytic serine residue of φC31 Int. Its substitution (S12A) was used as a control, as inactivation of the CD stabilises synaptic complexes by preventing the recombination from occurring after tetramerisation. This causes higher accumulation of synapses in the reaction mix and thus makes the gel visualisation easier. Bands of interest are labelled as follows: B0 - unbound *att* site. B1 - φC31 Int monomer bound to a single *att* site. B2 - φC31 Int dimer bound to a single *att* site. B4 - φC31 Int tetramer bound in a synaptic complex. Bands were visualised using Cy5*-tagged *att* site sequences. The *att* site pair present in each sample is indicated at the bottom, with the tagged oligo indicated by an asterisk (*).
A - Control binding experiment in the absence of RDF. S12A *attP* x *attB* and *attL* x *attR* samples show high levels of tetramer formation. Single site samples (*attP* x *attP*, *attB* x *attB*, etc.) are almost completely bound in dimers, with negligible tetramer formation. Integrase is not forming synapses on *attP* x *attL/R* pairings, but, notably, is forming synapses on *attB* x *attL/R* pairings.
B - Control binding experiment in the presence of RDF. Two -ve samples, as well as several S12A samples show anomalous ladder-like banding, which makes interpretation of the gel harder. Most samples have double-banding in the region corresponding to bound dimers. Similarly to A, Integrase is not forming synapses on *attP* x *attL/R* pairings, but, notably, is forming synapses on *attB* x *attL/R* pairings.

Figure 13 presents the results of a L474A binding assay. L474A behaves similarly to WT for standard *att* site pairings (*attP x attB* and *attL x attR*), which was expected given the results of the reaction efficiency study (see Chapter 3.2). However, it seems to be making synapses of two identical sites (*attP x attP*, etc.) much more readily than S12A (see Figure 11 A). Formation of synapses on non-standard pairings (*attP x attL*, etc.) follows a similar pattern to that observed in the S12A control, where pairings with *attP* appear to not form the synapse as readily as those with *attB*.

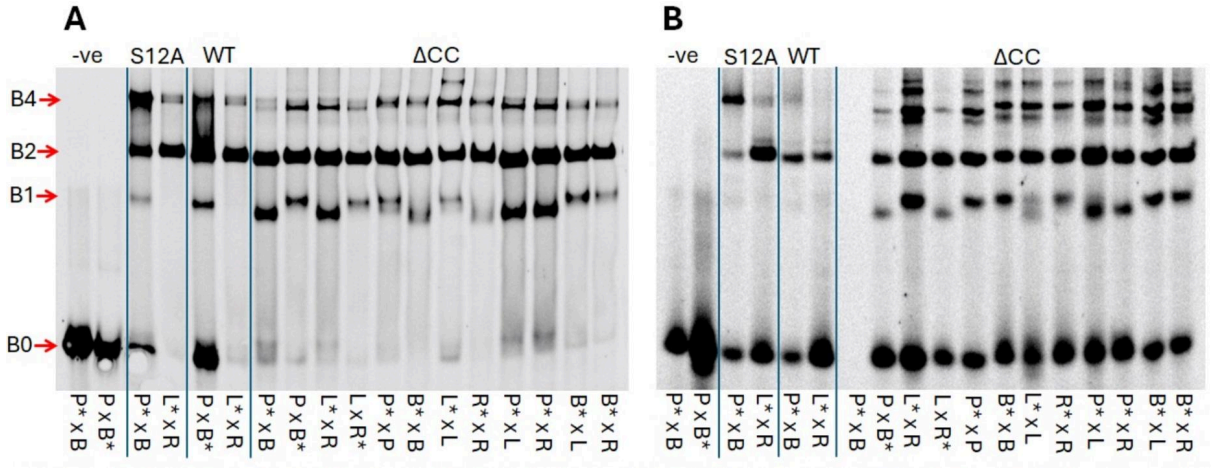


Figure 12. ΔCC binding assay. Bands of interest are labelled as follows: B0 - unbound *att* site oligoDNA. B1 - φC31 Int monomer bound to a single *att* site. B2 - φC31 Int dimer bound to a single *att* site. B4 - φC31 Int tetramer bound in a synaptic complex. Bands were visualised using Cy5*-tagged *att* site. The *att* site pair present in each sample indicated at the bottom, with tagged oligo indicated by an asterisk (*).

A - ΔCC binding experiment in the absence of RDF. There is an apparent increase of bound monomers in proportion to dimers. Synapse formation is occurring, albeit to a lesser degree than in S12A control.

B - ΔCC binding experiment in the presence of RDF. There is an apparent significant increase of bound monomers in proportion to dimers. Anomalous ladder-like banding in the B4 area is obscuring the result (especially in wells corresponding to non-standard *att* pairings), however, the strongest band of the ladder appears to be bigger than expected size of B4.

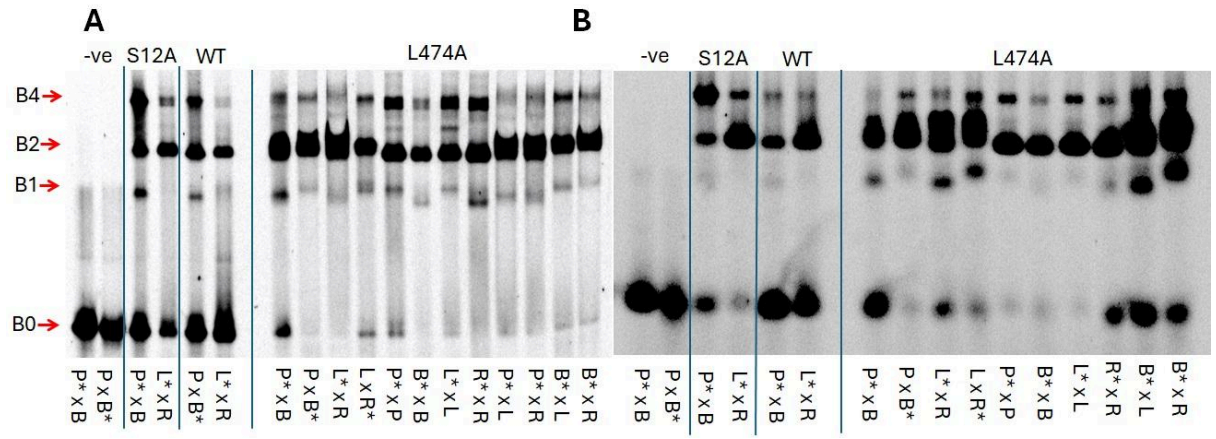


Figure 13. L474A binding assay. Bands of interest are labelled as follows: B0 - unbound *att* site oligoDNA. B1 - φC31 Int monomer bound to a single *att* site. B2 - φC31 Int dimer bound to a single *att* site. B4 - φC31 Int tetramer bound in a synaptic complex. Bands were visualised using Cy5*-tagged *att* site. The *att* site pair present in each sample indicated at the bottom, with the tagged oligo indicated by an asterisk (*).

A - L474A binding experiment in the absence of RDF. L474A behaves similarly to WT for standard *att* site pairings (*attP* x *attB* and *attL* x *attR*), which was expected given results of the reaction efficiency study (see Chapter 3.2). However, it seems to be making synapses of two identical sites (*attP* x *attP*, etc.) much more readily than S12A (see Figure 11 A). Formation of synapses on non-standard pairings (*attP* x *attL*, etc.) follows a similar pattern to that observed in the S12A control, where pairings with *attP* appear to not form the synapse as readily as those with *attB*.

B - L474A binding experiment in the presence of RDF. Interestingly, L474A seems to have a stronger RDF-induced shift towards *attL* x *attR* synapse in comparison to *attP* x *attB* than both WT and S12A. Other trends are consistent with those observed in A.

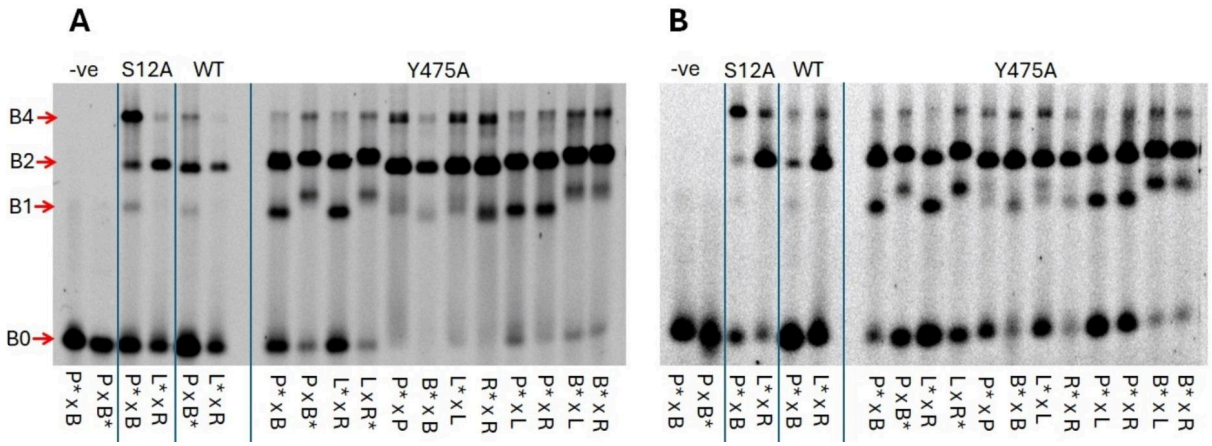


Figure 14. Y475A binding assay. Bands of interest are labelled as follows: B0 - unbound *att* site oligoDNA. B1 - φC31 Int monomer bound to a single *att* site. B2 - φC31 Int dimer bound to a single *att* site. B4 - φC31 Int tetramer bound in a synaptic complex. Bands were visualised using Cy5*-tagged *att* site. The *att* site pair present in each sample indicated at the bottom, with tagged oligo indicated by an asterisk (*).

A - Y475A binding experiment in the absence of RDF. Y475A shows relative lack of synapse formation on standard *att* site pairing (*attP* x *attB* and *attL x attR*) in relation to S12A, another inactive mutant. However, it is able to form a synapse on most self-similar pairings (with exception to *attB x attB*), as well as, to lesser extent, on non-standard pairings *attB* x *attL/R*.

B - Y475A binding experiment in the presence of RDF. Synapsis in all samples is much weaker than in non-RDF samples, but the trend of standard pairings undergoing the least tetramerisation is conserved between the experiments.

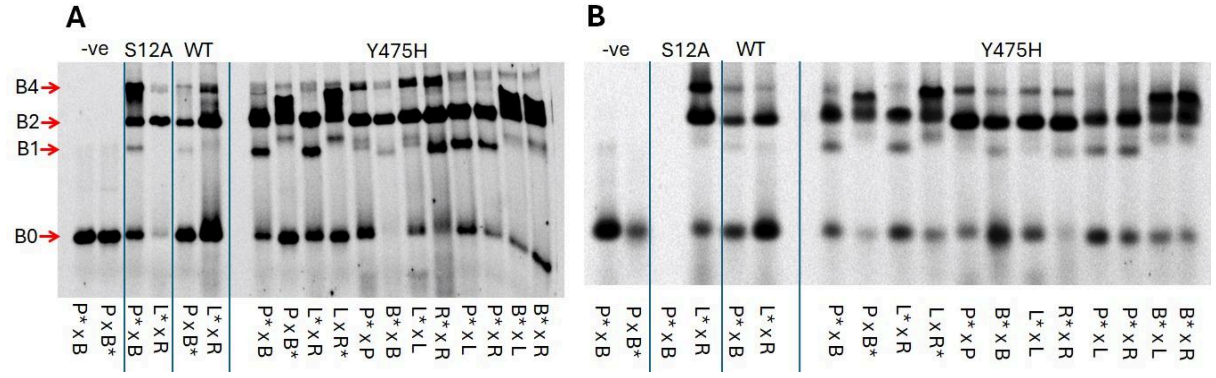


Figure 15. Y475H binding assay. Bands of interest are labelled as follows: B0 - unbound *att* site oligoDNA. B1 - φC31 Int monomer bound to a single *att* site. B2 - φC31 Int dimer bound to a single *att* site. B4 - φC31 Int tetramer bound in a synaptic complex. Bands were visualised using Cy5*-tagged *att* site. The *att* site pair present in each sample indicated at the bottom, with tagged oligo indicated by an asterisk (*).

A - Y475H binding experiment in the absence of RDF. Y475H appears to be more prone to form synapses on self-similar pairs of sites (e.g. *attP x attP*) than S12A control. Concentration of *attP x attB* and *attL x attR* synapses is comparable to WT. Very low synapse formation for all non-standard non-self-similar pairings (*attP/B x attL/R*).

B - Y475H binding experiment in the presence of RDF. General trend appears to be conserved from the non-RDF experiment, with slightly lower synapse concentrations in all samples.

Figure 14 presents the results of a Y475A binding assay. The Y475A mutant was inactive (Figure 8); however, its binding pattern differs from that of S12A. Y475A seems to form more synapses on the self-similar *att* pairings (*attP x attP, attL x attL* and *attR x attR*) than on the standard *attP x attB* and *attL x attR* pairings. The trend of forming some synapse on non-standard *attB x attL/R* pairings seems to be conserved in this mutant. The presence of RDF seems to not incur major changes to those trends, beyond generally reducing the amount of synapses formed across all samples in comparison to both controls.

Figure 15 presents the results of a Y475H binding assay. Synapse concentration on standard *att* pairings (*attP x attB* and *attL x attR*) is lower than in S12A, but in line with WT. This is expected for a mutant which displayed the ability to conduct the recombinations of those sites (see Figures 9 and 10). This mutant also appears to be more prone to form synapses on self-similar pairs of sites (e.g. *attP* x *attP*) than the S12A control, which is the same trend that was observed in the other mutant of the same residue, Y475A. The presence of RDF seems to have no meaningful impact on the binding trends, with the exception of lower synapse and monomer concentrations.

3.4 Site specificity study

A site specificity study was conducted using an *in vitro* Integrase reaction assay. Reactions were set up using *att* sites natively belonging to φBT1 Int, TP901 Int and TG1 Int. Similarly to the recombination efficiency study, substrate *att* site pairs were contained within single plasmids, one for *attP x attB*, and one for *attL x attR*. Maps of these plasmids can be found in Appendices C.5 through C.10. As the method used to check for activity on non-native *att* sites was the same as the one used in Chapter 3.2, Figure 8 A can serve as a model functioning φC31 Int.

Site specificity study results are presented in Figure 16. The results of the site specificity study suggest that none of the selected mutants showed any affinity towards any of the selected non-native sites, with a distinct lack of banding corresponding to the recombination products (see bands P1 and P2 on Figure 8).

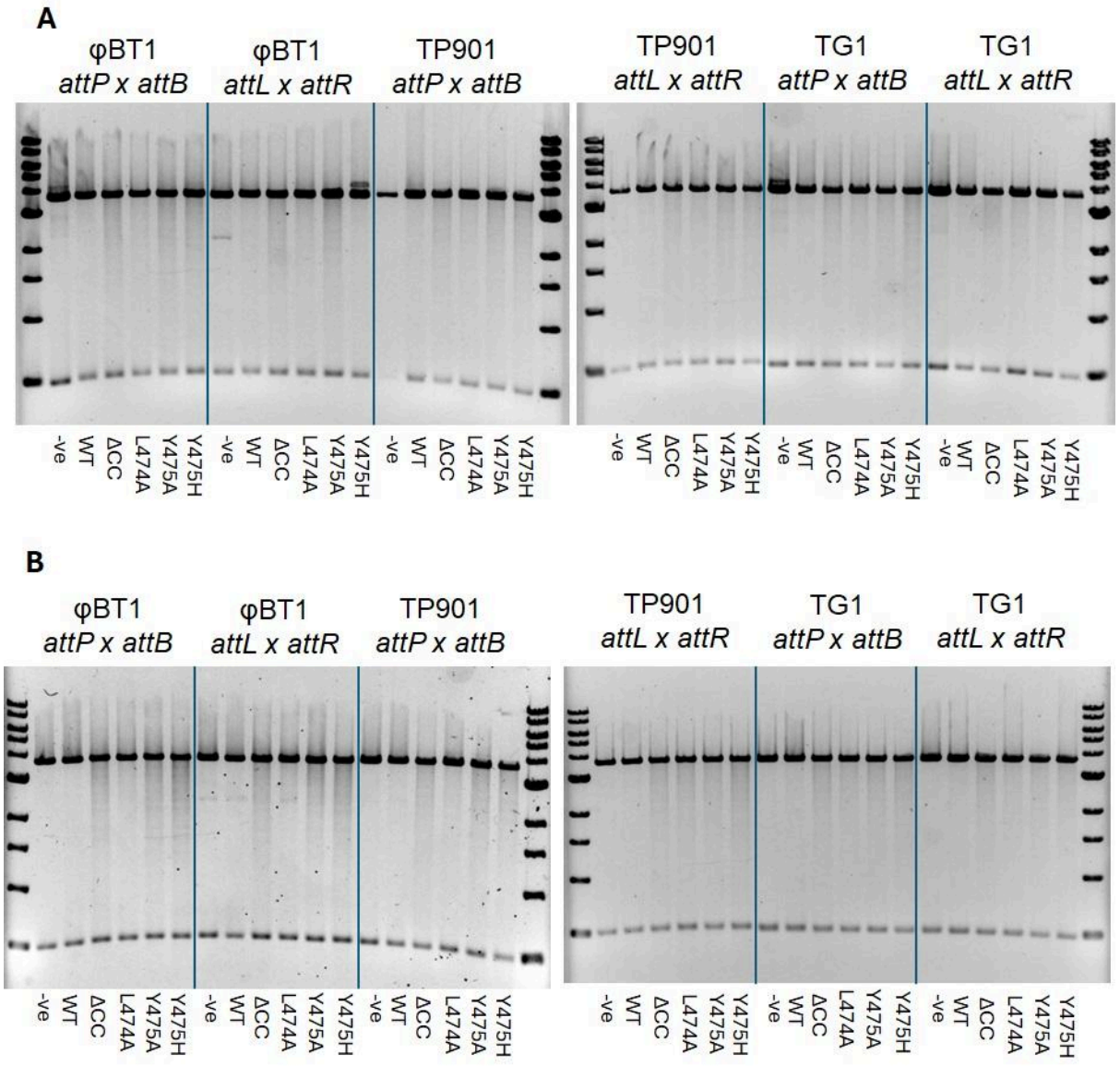


Figure 16. Results of the site specificity study. Gels in section A represent reactions conducted in the absence of φC31 RDF. Gels in section B represent reactions conducted in the presence of φC31 RDF. Substrate sites are indicated above the gel images, φC31 Int stocks used in reactions are indicated below the wells. No sample shows convincing evidence of even trace recombination, with distinct lack of bands corresponding to linearised products. Some reactions conducted on φBT1 Int *attL x attR* sites show a single faint band in the region corresponding to the linearised products. However, both in A and B this band is present in the negative control sample. Coupled with the fact that successful recombination would result in two product bands of different sizes, those bands are thought not to be indicative of any recombination taking place.

4. DISCUSSION

4.1 Interpreting the results

4.1.1 Interpreting the results - ΔCC mutant

The ΔCC mutant showed no activity under any of the experimental conditions. However, the binding study confirms that it is able to form synaptic complexes in the absence of RDF, albeit with significantly lower efficiency than the WT Integrase (Figure 12 A). This suggests that interactions between CDs on their own might be enough to form less stable tetramers (see Figure 4), but CC-CC interactions are either vital for stabilising the complex or significantly increase the dimer-to-dimer binding affinity. Interestingly, the concentration of synapses seems to be consistent in all samples regardless of the *att* site pairings. This contrasts with the S12A control (Figure 11 A), which displays strong preference for canonical pairings (*attP* x *attB* and *attL* x *attR*). This implies that the CC domains play a role in selectivity of synapse formation by inhibiting synapsis of non-canonical site pairings.

The binding experiment in the presence of RDF (Figure 12 B) is less conclusive due to anomalous ladder-like banding in the region where the synaptic complex band is expected to appear. One possible explanation of this phenomenon stems from the fact that gp3 (φC31 Int RDF) was found to bind in the hinge region between the CC and the ZD [Fogg *et al.* 2018]. This means that the stability of RDF binding might be significantly decreased in the absence of the CC domain, as it harbours some of the Integrase-RDF interaction sites. Two of the φC31 Int residues known to be important for gp3 binding are I420 and W526 [Fogg *et al.* 2018; see Appendix B.1]. W526 lies within the region excised from φC31 Int in order to create the studied ΔCC mutant (see Appendix B.2). This could mean that anomalous laddering of the band corresponding to the synapse is caused by RDF not being able to bind to the Integrase in a stable manner, resulting in synapses where only some of the subunits have retained the bound RDF. As each bound RDF molecule adds mass to the complex, this would result in five different synapse bands - each corresponding to the amount of RDF molecules bound, from 0 to 4. The gel image in Figure 12 B shows four bands in the region, the sizes of which appear to be consistent between the samples. Additionally, while impossible to ascertain with the quality of current gel images, there might be an extremely

faint fifth band in lanes 11 and 18 (samples *attP** x *attP* and *attB** x *attR*). This faint band is bigger than all four confirmed bands in the region. Therefore if its presence were to be confirmed, it would imply that RDF binding is severely affected, as only a trace amount of synapses with all 4 RDF binding sites was formed. The issue with this hypothesis is that, if it were correct, four out of five bands would be expected to appear smaller than the band corresponding to WT/S12A + RDF synapse, as those synapses are expected to have all 4 RDF binding sites filled properly. However, this isn't the case, as two out of four confirmed bands appear to be larger than the WT/S12A + RDF synapse. Additionally, previous research on A118 Int suggests that the ΔCC mutant of that Integrase is able to form an *attP* x *attB* synaptic complex, but not an *attL* x *attR* one [Rutherford *et al.,* 2013]. While this means that data from other LSR do not support this hypothesis, the possibility that the impact of CC interactions on synapsis differs between the two Integrases cannot be excluded, especially given the known differences between the two Integrases (elaborated on in Chapter 4.2). A future study testing the efficiency of φC31 Int-gp3 binding on a variety of CC truncations could help in confirming the extent to which the binding is stable in the absence of some of the interaction sites, and might also provide information on the minimal region necessary to stabilise the binding.

As mentioned in Chapter 3.2, quantitative measurements of gel band intensity are very error-prone due to high variance in image background intensity. As such, due to lack of any apparent band indicating recombination in any samples containing ΔCC (Figure 8B and 8D), positive values corresponding to those samples that appear in Figures 9 and 10 are thought to be a result of measurement inaccuracies rather than genuine presence of recombination products. This line of interpretation is supported by previous studies, as it is known that the closely related A118 Int is completely unable to facilitate any recombination after introduction of the CC deletion [Mandali *et al.,* 2017]. Complete inability to perform even trace recombination in all tested conditions supports the hypothesis that presence of one of the CC-CC interactions is necessary for enabling the reaction. However, the exact reaction steps enabled by those interactions remain unknown.

As mentioned at the beginning of this Chapter, ΔCC mutants appear to form some synapses, albeit with a much lesser efficiency than the WT Integrase (Figure 12). Those results are in line with previous studies. Examination of the ΔCC φC31 Int and LI Int reaction kinetics have shown that deletion of the Coiled-Coil facilitates formation of higher number of

inactive complexes, as well as slows down their dissociation, severely slowing down the synapsis [Chen *et al.*, 2023]. While it is possible that stability of the synapse is essential for recombination, it appears more likely that, if this was the only issue caused by CC deletion, recombination would still happen, albeit to a much lesser degree. This conflicts with the observed absence of recombination products. Another reaction step which could be disabled by the CC deletion is the DNA cleavage performed by the CD. Structural studies performed with TP901 Int suggest that LSR CDs and resolvase CDs share a high degree of structural similarity [Yuan *et al.*, 2008; Smith, 2015]. This similarity might suggest that the φC31 Int CD is not very likely to need CC interactions to cleave the DNA, as resolvases can perform cleavage despite not having any CCs (see Figure 3). However, if substrates were successfully cleaved without following re-ligation, gels in Figure 8B and 8D would be expected to show 4 bands, as the substrate plasmid would have been cleaved at both NruI sites as well as at *attP/L* and *attB/R* sites. This means that the ΔCC mutant is most probably following one of the two possible reaction sequences: either the DNA is never cleaved by CD, and as such strand exchange cannot occur; or, the DNA is cleaved successfully, but the rotational movement of subunits which facilitates the strand exchange is impossible in the absence of CC interactions, preventing the strand exchange from occurring and leading to re-ligation of the substrate sites.

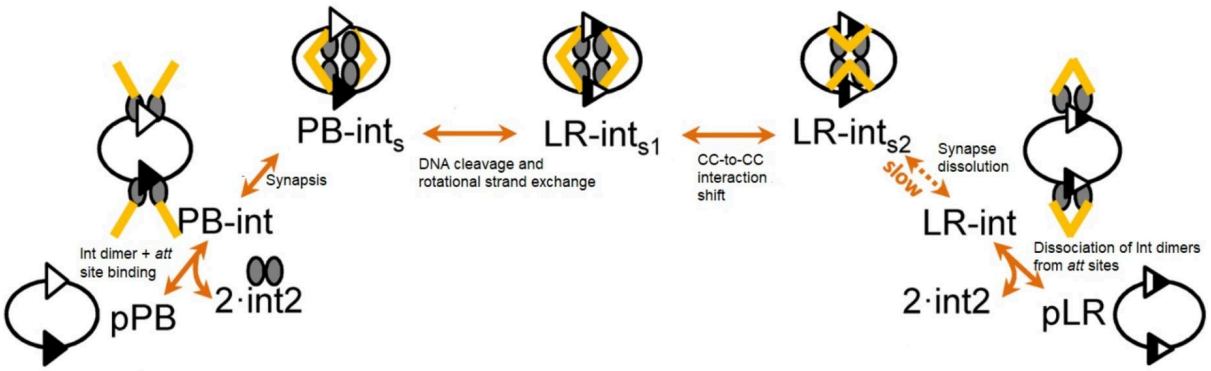


Figure 17. A representation of the φC31 Int reaction predicted by a mathematical model developed by Pokhilko et al. (2017). Coiled-coil domains are represented by yellow rectangles. Note how it is predicted that CC domains cease their interaction with CCs of the opposing dimer in favour of interacting with CCs on the same dimer after the rotational strand exchange is completed, but before the synaptic complex dissolves. Figure adapted from Pokhilko et al. (2017).

LSR CCs are thought to interact with the CCs on the opposite substrate *att* site during synapsis (*trans*-dimeric interaction), whereas after recombination is completed they are thought to interact with the other CC bound to the same product *att* site (*cis*-dimeric

interaction) (Figure 4) [Van Duyne and Rutherford, 2013; Mandali and Johnson, 2021]. The mathematical model of the reaction suggests that the transition between those CC-CC interactions occurs after the rotational strand exchange, but before the synapse dissolves (Figure 17) [Pokhilko *et al*., 2016]. Study of the rotational strand exchange in φC31 Int has shown that, in most cases, the subunit rotation is gated after a single 180^{o} revolution, after which further revolutions appear to be inhibited [Olorunniji *et al.,* 2012]. If the switch of CC interactions from *trans*-dimeric to *cis*-dimeric interactions is occurring rapidly, it would provide an explanation for the mechanism behind this gating.

As mentioned in Chapter 1.6, there are two known geometries of CC-CC interaction in LSRs. In the absence of RDF one of those geometries occurs in *attP* x *attB* synapses and promotes recombination, while the other occurs in *attL* x *attR* synapses and prevents the recombination, as suggested by previous studies in LI Int and A118 Int [Gupta *et al.*, 2017; Mandali and Johnson, 2021]. It could be possible that the shift between *trans*- and *cis*-dimeric interactions occurs after the rotational strand exchange because the exchange results in *attL* x *attR* synapse geometry, making the inhibiting interaction more conformationally favourable. If the reaction step prevented by this inhibiting *cis*-dimeric interaction is the rotational strand exchange, it appears likely that gating of the subunit rotation and the reaction directionality itself are controlled by the same mechanism. Conversely, since *trans*-dimeric interactions promote the reaction, it seems possible that they are responsible not only for stabilising the synaptic complex, but also for facilitating the subunit rotation and the strand exchange. Therefore, I propose that the reaction step most likely to be impossible to facilitate for the φC31 Int ΔCC mutant is the rotational strand exchange.

4.1.2 Interpreting the results - L474A mutant

The L474A mutant is able to facilitate recombination of the expected site pairings, albeit with reduced efficiency in comparison to the WT φC31 Int. This is especially true for reactions of *attL* x *attR* substrates in the presence of RDF. In that reaction, L474A shows a ~50-60% drop in estimated reaction efficiency in comparison to its reaction on *attP* x *attB* substrates in the absence of RDF. This contrasts with WT Int, whose estimated reaction efficiency seems largely similar between those two conditions (peaks of 70% and 80% estimated efficiency for *attP* x *attB* reaction without RDF and *attL* x *attR* in the presence of

RDF respectively). This could suggest that the L474A mutant is less sensitive to RDF than the WT Int.

It is possible that alanine could not facilitate the same interactions as leucine does in WT Int (on account of its much shorter side-chain; Figure 18). However, as both leucine and alanine are non-polar and aliphatic AAs [PubChem, 2025], alanine is also unlikely to disrupt other nearby interactions. Those remaining interactions would compensate for the L474 mutation, explaining why the mutant enzyme retained most of its core functionality.

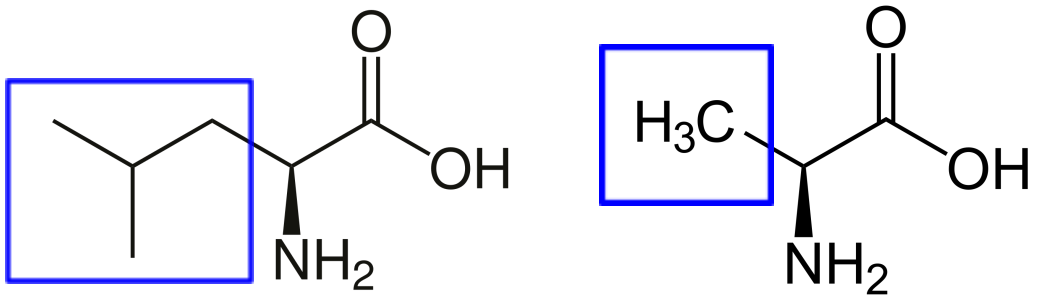


Figure 18. Comparison of the skeletal formulas of Leucine (left) and Alanine (right). Side-chains indicated by the blue box. Note that the side-chains of both amino-acids are simple hydrocarbons, with the Leucine side-chain being 3 Carbon atoms longer and branched. Formulas sourced from PubChem (2025), CID 6106 and CID 5950.

The binding assay results show that L474A forms synapses on pairs of identical sites (ie. *attP* x *attP, attL* x *attL* and *attR* x *attR*; interestingly, *attB* x *attB* pairing appears to be an exception from this trend) much more readily than the S12A control (see Figures 13A and 11A). One possible explanation is that this leucine side-chain is one of the factors which inhibit the synapsis of non-canonical *att* site pairings. As the geometry of the synapsis changes with the *att* sites bound (due to the difference in sizes between different *att* sites) it is possible that one of the functions of the L474 side-chain is blocking other residues from accessing the natural CC-CC interaction sites while in non-canonical synapse geometries. This would prevent those interactions from stabilising the complex on non-canonical *att* sites pairings. As the alanine mutant lacks this side-chain, that inhibitory property would be gone, allowing the L474A mutant to form the observed non-canonical synapses more easily. This line of interpretation is supported by the ΔCC binding assay results, as in the absence of CC domains synapsis appears to be equally likely among all tested *att* site pairings.

4.1.3 Interpreting the results - Y475 mutants

The Y475A mutant was inactive under all tested conditions (Figure 8 C and 8 E), confirming that Y475 is essential for the CC-CC interactions that promote recombination. Interestingly, its binding assay results are similar to that of L474A. Y475A also appears to form synapses between pairs of identical sites much more readily than the S12A control. In fact, both in the presence and absence of RDF, Y475A forms more synapses on non-canonical pairings of *att* sites than on the canonical *attP* x *attB* and *attL* x *attR* pairings (Figure 14). Both of those results point towards Y475 enabling an interaction between CC domains in an *attP* x *attB* synaptic complex in WT φC31 Int, binding to the opposing CC to stabilise the complex and allow the recombination.

The Y475H mutant shows recombination activity on *attP* x *attB* both in the presence and absence of RDF. Additionally, RDF seems to inhibit its *attP* x *attB* activity to a lesser degree than in WT φC31 Int (Figure 10). Y475H is significantly less active than WT on *attL* x *attR* in the presence of RDF. However, it appears to also show some trace activity on those substrates in the absence of RDF, a trait not shown by any other examined sample (Figure 8C and 8E). This could imply that histidine conserves some of the tyrosine's ability to form interactions promoting the recombination, but reduces the specificity of their occurrence.

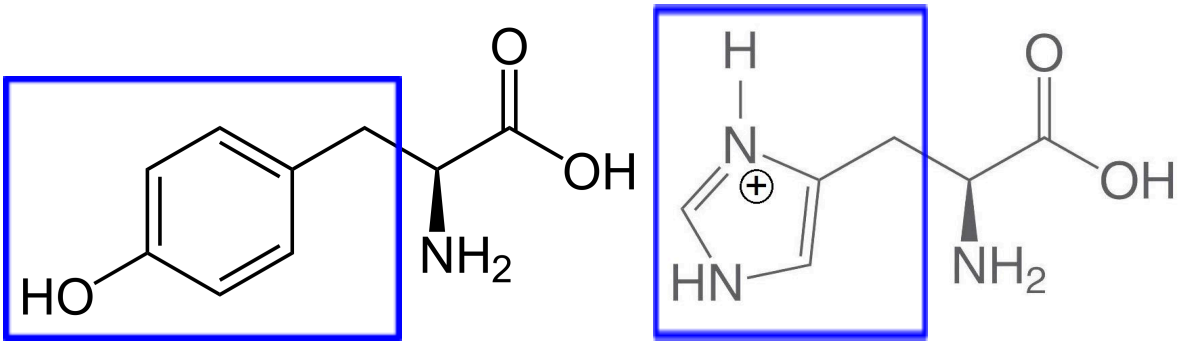


Figure 19. Comparison of the skeletal formulas of Tyrosine (left) and Histidine (right). Side-chains indicated by the blue box. Notice how both side-chains contain an aromatic ring. Formulas sourced from PubChem (2025), CID 6057 and CID 6274.

Tyrosine is a hydrophobic AA while histidine is a positively charged AA. However, they both have a large side-chain containing an aromatic ring (Figure 19). This set of differences and similarities provides a potential explanation for the observed results of

Y475H mutant reactions. If the interaction promoting *attP* x *attB* recombination hinges on either the properties of the π bond of the aromatic ring or the Van der Waals forces, Y475H would retain the ability to contribute to those interactions. Additionally, the hydroxyl group unique to tyrosine might provide additional stability to the interactions, which positively charged histidine would lack, explaining the reduced reaction efficiency of Y475H. If those same ionic interactions are used to prevent the tyrosine side-chain from entering the promoting interactions on *attL* x *attR* pairings, their absence would explain the activity. Y475H retains more activity on the *attP* x *attB* substrate in the presence of RDF, as well as the trace activity it shows on the *attL* x *attR* substrate in the absence of RDF. Unfortunately, comparative analysis using known structures of LI Int CC-CC interactions (see Figure 6) does not provide additional insight into the nature of this interaction in φC31 Int, as neither of the nearby residues in those models are conserved between LI Int and φC31 Int (LI Int R364 and I370 are corresponding to φC31 Int A470 and E476 respectively; see Figure 5).

A deeper investigation into the (unique among all tested mutants) activity of Y475H on *attL* x *attR* substrates in the absence of RDF could lead to a better understanding of how this residue enables recombination. Investigating activity of Y475H on non-canonical *att* sites pairings could result in more insight into the promoting interactions including the residue, as well as the impact of those interactions on overall φC31 Int specificity. Other aromatic AAs, such as phenylalanine or tryptophan, could also be introduced in place of Y475, in order to test the hypothesis that the mode of promoting interaction relies on the presence of the aromatic ring.

The binding assay results confirm the suspicion of decreased specificity of this mutant, and are in line with results observed with other tested mutants. Y475H is able to form synaptic complexes on self-similar *att* site pairings (such as *attP* x *attP*) more readily than the S12A control (Figure 15). Additionally, just as in the case of Y475A, those self-similar pairings appear to form synapses most readily. However, this particular result should not be taken at face value, as Y475H is an active mutant. This means that a significant portion of *attP* x *attB* sites would be recombined, resulting in dissolution of the corresponding synaptic complexes, and therefore weaker bands.

4.1.4 Interpreting the results - Site specificity study

No tested mutant showed any signs of successful recombination of φBT1 Int, TP901 Int or TG1 Int *att* sites under any of the tested conditions (Figure 16).

Lack of any recombination products for Y475H confirm, that despite the fact that this mutation is able to promote recombination of normally inactive φC31 Int *att* substrates, it does not reduce its specificity enough to allow for recombination of *att* sites belonging to different Integrases. This confirms that the tested CC-CC interactions facilitate promotion and inhibition of recombination mainly in relation to directionality control, but not in relation to site specificity.

This result is in line with previous LSR studies. It has been shown that the DNA binding by the ZD of φC31 Int is highly site-specific [McEwan *et al.*, 2011]. The *att* sequences associated with the three tested integrases share little similarity to φC31 Int *att* sites (see Appendix A.1 through A.7). All of the tested *att* sites share less than 40% sequence identity with their φC31 Int equivalents, with a notable exception of TG1 *attP*, which shares 60% sequence identity with φC31 Int *attP*. As such, the result of this specificity study not only falls in line with results previously obtained in ZD-related studies, but also provides evidence supporting the thesis that CC-CC interactions alone are not strong enough to force the φC31 Int to bind to foreign *att* sites. However, experiments using more drastic changes to the CC would be required to definitely confirm this hypothesis, especially in the wake of binding study results, which has shown that synapsis of non-canonical *att* sites pairings is not only possible, but easy to attain with simple point mutations. While ZD specificity normally prevents off-site binding, it could be interesting to see if particularly strong CC-CC interactions would be able to force it to bind into an unfavourable site, and if so, how conserved would that site be in comparison to native *att* sites. The assumption that it could be possible to force the ZD to bind to normally unfavourable sites is not as far-fetched as it might seem. It has been shown that in nature φC31 Int is able to recombine *attP* with various sites not identical to native *attB*, if the *attB* is absent [Combes *et al.*, 2002]. φC31 Int displays 300-fold lesser affinity for those "pseudo-*attB*" sites than for the standard *attB*, but the recombination products are still detectable. The reported pseudo-*attB* sites share only between 50 to 60% identity with canonical *attB*, but the conserved nucleotides appear to be highly conserved between all of them [Combes *et al.*, 2002], suggesting that DNA binding

specificity is guided by several nucleotides of high importance as opposed to general sequence similarity. A study into how many of those conserved nucleotides can be mutated before synapsis of the modified sites is completely impossible could provide insight into both general φC31 Int specificity, and into how drastically it can be affected by changes to the CC domain.

While it was impossible to conduct as a part of this project due to time constraints, it would be appropriate to conduct a binding assay of the reaction mixes containing foreign *att* sites in addition to the presented reaction efficiency assay. TG1 Int *att* sites are the only ones which have the identical central dinucleotide to that of φC31 Int *att* sites (TT; see Appendix A.1 through A.7), meaning that if the φC31 Int CD displays any specificity in relation to the cut site, it would not be able to cleave the φBT1 Int and TP901 Int *att* sites, resulting in a negative result in reaction efficiency assay. A binding assay could provide additional information on whether the dimers are able to bind to, or form the synaptic tetramers on the foreign *att* sites to any extent.

4.2 Notable differences between φC31 Int and LI Int/A118 Int and the resulting bias in the experimental design

As mentioned in Chapter 1.6, the bulk of previous studies on LSRs were conducted using LI Int and A118 Int, near-identical enzymes that are commonly chosen for structural studies due to their smaller sizes and reduced complexity [Rutherford *et al.*, 2013; Gupta *et al.*, 2017]. Because of the higher abundance of structural and biochemical data, the bulk of the experimental design of this project was based on studies made on those two enzymes. The leading assumption was that the homology between LSR domains and functions is conserved enough to ask research questions about φC31 Int based mostly on LI Int data.

However, there exists evidence that CC domain function differs to a significant degree between the two enzymes, despite the fact that their position and orientation in the geometry of the reaction complex is assumed to be identical. The result of the recombination reaction efficiency study of the ΔCC mutant shows that the CC domain is essential for recombinase activity in φC31 Int. This is in line with previous studies on φC31 Int, which suggest that the enzyme cannot facilitate recombination in the absence of the CC [Chen *et al.*, 2023]. However, it contrasts with previous studies in LI Int, which found that deletion of LI Int CC

results in a mutant that is not only able to conduct intramolecular recombination, but also results in a more promiscuous enzyme, able to facilitate *attP x attB, attL* x *attR*, *attP* x *attP* and *attB x attB* recombination in the absence of RDF. [Rutherford *et al.,* 2013]. The results gathered for the Y475A mutant during this project suggest that in the case of φC31 Int, not only is the CC domain's presence needed for the recombination, but also very specific CC-CC interactions are essential as well. This puts the nature of CC control over integrase function between these two enzymes in direct contrast to each other.

Evidence exists that φC31 Int CC-CC interactions can both inhibit and promote recombination [Fogg *et al.*, 2018], which is also the case for LI Int as well [Rutherford *et al.,* 2013]. However, the fact that the presence of CC-CC interactions is essential to φC31 Int function implies that the way in which these interactions affect the behaviour of the protein can be different from the way they do in LI Int. The unmodelled region between the predicted locations of αK and αL (i.e. region between G445 and A531, see Figure 5) might be a promising candidate for examining the source of this discrepancy, as it is unique to φC31 Int. At the time of writing, a new study including high resolution structures of the φC31 Int CC domain has been submitted for peer review [Sun *et al*., 2025]. Its results show that D477, a residue which is not conserved between phage LSRs (Figure 5), is taking part in CC-CC interactions of φC31 Int. As none of the residues traditionally aligned with D477 are observed to take part in CC-CC interactions in the LI Int (Figure 6) [Gupta *et al.*, 2017; Mandali and Johnson, 2021], it provides further evidence that the mode of CC-CC interactions between different Serine Integrases varies more than anticipated.

The CC-CC interactions appear to be different between the two Integrases even when considering a single isolated conserved AA residue as well. Qualitative results of the recombination reaction efficiency study on the φC31 Int Y475A mutant suggest that it is unable to recombine any substrates under any conditions (Figure 8C and 8E; Chapter 4.1.3). However, an alanine scanning study in LI Int has shown that the alanine mutant of the corresponding LI Int residue Y469 is able to recombine the *attL* x *attR* substrates in presence of LI Int RDF [Mandali and Johnson, 2021]. The modelled structure of LI Int CC-CC interactions provided in the same study (Figure 6) suggests that LI Int (Y469) would interact with R364, L468 and I370. However, two of those interactions would be impossible in φC31 Int (assuming that the CC structure is conserved in accordance with the AA sequence alignment), as LI Int R364 and I370 are thought to correspond to φC31 Int A470 and E476

respectively (Figure 5). Isoleucine and Glutamic acid both have long, aliphatic side-chains, but the former is hydrophobic and nonpolar, while the latter bears a strong negative charge. Arginine and Alanine bear even less resemblance, as the former has a long, positively charged side-chain, while the latter is nonpolar and has the smallest side-chain among all canonical AAs [PubChem, 2025]. Interestingly, despite those differences in nature and impact of the interactions, the tyrosine residue corresponding to φC31 Int Y475 is heavily conserved among the LSRs (Figure 5) [Van Duyne and Rutherford, 2013]. This suggests that the presence of this residue might have been important to the function of an ancestral LSR, and the exact nature of how it interacts with surrounding residues slowly evolved over time, creating the observed discrepancies between various contemporary LSRs.

4.3 Relevance of presented findings to φC31 Int applications

As mentioned in Chapter 1.5.4, most biotechnological applications of φC31 Int rely on its high site specificity and controllable directionality [Nern *et al.*, 2011; Farrugio *et al.*, 2012]. However, a designer enzyme with reduced site-specificity could be used to attempt genetic modifications (such as large fragment insertions) without screening the target genome for *att* sites. The results of the binding assay confirm that various CC mutations can induce creation of non-canonical synapses, while the recombination reaction efficiency study on Y475H confirmed that creating mutants which are active on normally silent site pairings is possible. In addition, previous studies showed that even WT φC31 Int is able to recombine some foreign sites with *attP*, albeit with greatly reduced efficiency [Combes *et al.*, 2002]. Deeper investigation into the nature of those phenomena, such as the ones proposed in Chapters 4.1.3 and 4.1.4, could provide the insight necessary for creation of such a designer Integrase.

In addition, the Y475H results show decreased affinity for *attL* x *attR*, even in the presence of the RDF. This confirms that the regulation of φC31 Int can be tightened, which might be a desirable trait for projects working with synthetic biology and biocomputing applications of the Integrase. Locating other residues which selectively reduce affinity for particular sites, as well as their mutants which result in highest recombination rates, could provide φC31 Int with a library of modular mutations, allowing the industrial and research users of the Integrase to tailor the specificity and directionality control of the enzyme to their current needs.

APPENDIX A - DNA sequences:

A.1: φC31 Int *attP* site sequence

Sequence of φC31 Int *attP* site, as visible on the map of plasmid pFM16 (53 bp).

Highlighted in red is the minimal site length necessary for efficient recombination (39 bp) [Groth *et al.*, 2000]. Underlined is the sequence at which crossover occurs during recombination [Combes *et al.*, 2002].

```
5′ - AGTAGTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCGTAGGG - 3′
3′ - TCATCACGGGGTTGACCCCATTGGAAACTCAAGAGAGTCAACCCCCGCATCCC - 5′
```

A.2: φC31 Int *attB* site sequence

Sequence of φC31 Int *attB* site, as visible on the map of plasmid pFM16 (55 bp).

Highlighted in red is the minimal site length necessary for efficient recombination (34 bp) [Groth *et al.*, 2000]. Underlined is the sequence at which crossover occurs during recombination [Combes *et al.*, 2002].

```
5′ - CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCT - 3′
3′ - GGCGCCACGCCCACGGTCCCGCACGGGAACCCGAGGGGCCCGCGCATGAGGTGGA - 5′
```

A.3: φC31 Int *attL* site sequence

Sequence of φC31 Int *attL* hybrid site, as highlighted on the map of plasmid pFM52 (55 bp).

Highlighted in blue is the half-site originating from *attP*, highlighted in orange is the half-site originating from *attB* [Stark, 2017]. Underlined is the sequence at which crossover occurs during recombination [Combes *et al.*, 2002; Stark, 2017].

```
5′ - AGTAGTGCCCCAACTGGGGTAACCTTTGGGCTCCCCGGGCGCGTACTCCACCT - 5′
3′ - TCATCACGGGGTTGACCCCATTGGAAACCCGAGGGGCCCGCGCATGAGGTGGA - 5′
```

A.4: φC31 Int *attR* site sequence

Sequence of φC31 Int *attR* hybrid site, as highlighted on the map of plasmid pFM52 (55 bp).

Highlighted in blue is the half-site originating from *attP*, highlighted in orange is the half-site originating from *attB* [Stark, 2017]. Underlined is the sequence at which crossover occurs during recombination [Combes *et al.*, 2002; Stark, 2017].

```
5′ - CCGCGGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTAGGG - 3′
3′ - GGCGCCACGCCCACGGTCCCGCACGGGAACTCAAGAGAGTCAACCCCCGCATCCC - 5′
```

A.5: φBT1 Int *att* sites sequences

Sequences of φBT1 Int *att* sites. Underlined is the sequence at which crossover occurs during recombination [Gregory *et al.,* 2003].

φBT1 Int *attP:*

5′ - AGACGTTTCGGGTGCTGGGTTGTTGTCTCTGGACAGTGATCCATGGGAAACTACTCAGCACCACCAATGTTCC - 3′

φBT1 Int *attB:*

5′ - GCCCGCTGCCGTCCTTGACCAGGTTTTTGACGAAAGTGATCCAGATGATCCAGCTCCACACCCCGAACGCGAG - 3′

φBT1 Int *attL:*

5′ - GCCCGCTGCCGTCCTTGACCAGGTTTTTGACGAAAGTGATCCATGGGAAACTACTCAGCACCACCAATGTTCC - 3′

φBT1 Int *attR:*

5′ - AGACGTTTCGGGTGCTGGGTTGTTGTCTCTGGACAGTGATCCAGATGATCCAGCTCCACACCCCGAACGCGAG - 3′

A.6: TP901 Int *att* sites sequences

Sequences of TP901 Int *att* sites. Underlined is the sequence at which crossover occurs during recombination [Christiansen *et al.*, 1994].

TP901 Int *attP:*

5′ - GAGTTTTTATTTCGTTTATTTCAATTAAGGTAACTAAAAAACTCCTTTTAAGG - 3′

TP901 Int *attB:*

5′ - ATGCCAACACAATTAACATCTCAATCAAGGTAAATGCTTTTTGCTTTTTTTGC - 3′

TP901 Int *attL:*

5′ - ATGCCAACACAATTAACATCTCAATTAAGGTAACTAAAAAACTCCTTTTAAGG - 3′

TP901 Int *attR:*

5′ - GAGTTTTTATTTCGTTTATTTCAATCAAGGTAAATGCTTTTTGCTTTTTTTGC - 3′

A.7 TG1 Int *att* sites sequences

Sequences of TG1 Int *att* sites. Underlined is the sequence at which crossover occurs during recombination [Hirano *et al.*, 2011].

TG1 Int *attP*:

5′ - GTTCCAGCCCAACAGTGTTAGTCTTTGCTCTTACCCAGTTGGGCGGGATA - 3′

TG1 Int *attB*:

5′ - TCGATCAGCTCCGCGGGCAAGACCTTCTCCTTCACGGGGTGGAAGGTCGG - 3′

TG1 Int *attL*:

5′ - TCGATCAGCTCCGCGGGCAAGACCTTGCTCTTACCCAGTTGGGCGGGATA - 3′

TG1 Int *attR*:

5′ - GTTCCAGCCCAACAGTGTTAGTCTTTCTCCTTCACGGGGTGGAAGGTCGG - 3′

A.8: Sequence of the restriction cassette

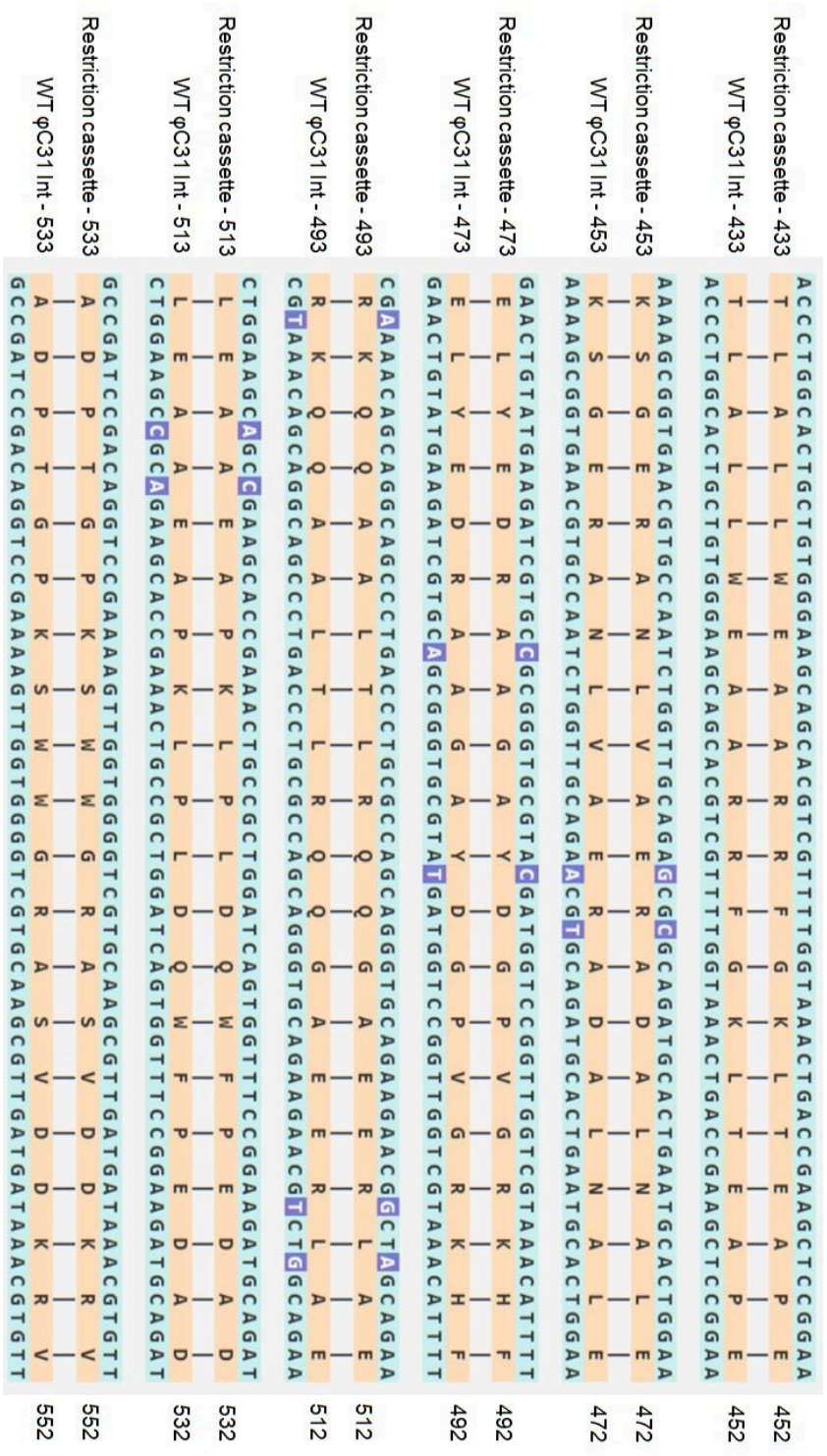
Sequence of the restriction cassette, corresponding to the C-terminal end of the φC31 Int gene sequence. Substituted nucleotides (in comparison to WT) are underlined. Introduced restriction sites are highlighted in the following order:

BssHII site - orange
SacII site - yellow
BsiWI site - green
TaqI site - cyan
NheI site - indigo
EagI site - purple

GGATCCGGCAACAGTTATGCGTATTCTGCGTGATCCGCGTATTGCCGGTTTTGCAGCAGAAG
TGATCTACAAAAAAAAACCTGATGGCACCCCGACCACCAAAATTGAAGGTTATCGTATCCAG
CGCGATCCGATTACCCTGCGTCCGGTTGAACTGGATTGTGGTCCGATTATTGAACCGGCAGA
ATGGTATGAACTGCAGGCATGGCTGGATGGTCGTGGTCGCGGTAAAGGTCTGAGCCGTGGTC
AGGCAATTCTGAGCGCAATGGATAAACTGTATTGTGAATGTGGTGCCGTTATGACCAGCAAA
CGTGGTGAAGAAAGCATCAAAGATAGTTATCGTTGTCGTCGTCGTAAAGTTGTTGATCCGAG
CGCACCGGGTCAGCATGAAGGCACCTGTAATGTTAGCATGGCAGCACTGGATAAATTTGTTG
CCGAACGCATCTTTAACAAAATTCGTCATGCCGAAGGTGATGAAGAAACCCTGGCACTGCTG
TGGGAAGCAGCACGTCGTTTTGGTAAACTGACCGAAGCTCCGGAAAAAAGCGGTGAACGTGC
CAATCTGGTTGCAGAGCGCGCAGATGCACTGAATGCACTGGAAGAACTGTATGAAGATCGTG
CCGCGGGTGCGTACGATGGTCCGGTTGGTCGTAAACATTTTCGAAAACAGCAGGCAGCCCTG
ACCCTGCGCCAGCAGGGTGCAGAAGAACGGCTAGCAGAACTGGAAGCAGCCGAAGCACCGAA
ACTGCCGCTGGATCAGTGGTTTCCGGAAGATGCAGATGCCGATCCGACAGGTCCGAAAAGTT
GGTGGGGTCGTGCAAGCGTTGATGATAAACGTGTTTTTGTGGGTCTGTTCGTGGATAAAATT
GTGGTTACCAAAAGCACCACCGGTCGCGGTCAGGGTACACCGATTGAAAAACGTGCAAGCAT
TACCTGGGCAAAACCGCCTACCGATGATGATGAAGATGATGCACAGGATGGCACCGAAGATG
TTGCAGCAACTAGT

A.9: Alignment of the modified section of the restriction cassette to the corresponding WT φC31 Int gene fragment

DNA and AA sequence of the restriction cassette (top) aligned to WT φC31 Int gene sequence (bottom). The shown fragment lies between T433 and V552 (see Appendix B.1).



APPENDIX B - Amino-acid sequences:

B.1: Amino-acid sequence of WT φC31 Int

AA sequence of WT φC31 Int used in this study, including a 6xHis-tag at the C-terminal end (indicated in blue). The CC domain is highlighted in purple, based on the pre-existing alignments with other LSRs [Gupta *et al.*, 2017]. Residues of interest listed in the Abstract and Chapter 1.7 are indicated in **bold**. L474 and Y475 are indicated by **underscore**. Residues I420 and W526 mentioned in Chapter 4.1.1 are indicated in red. Region excised to create the ΔCC mutant used in this study is found between the brackets (see Appendix B.2).

MDTYAGAYDRQSRERENSSAASPATQRSANEDKAADLQREVERDGGRFRFVGHFSEAPGTSA
FGTAERPEFERILNECRAGRLNMIIVYDVSRFSRLKVMDAIPIVSELLALGVTIVSTQEGVF
RQGNVMDLIHLIMRLDASHKESSLKSAKILDTKNLQRELGGYVGGKAPYGFELVSETKEITR
NGRMVNVVINKLAHSTTPLTGPFEFEPDVIRWWWREIKTHKHLPFKPGSQAAIHPGSITGLC
KRMDADAVPTRGETIGKKTASSAWDPATVMRILRDPRIAGFAAEVIYKKKPDGTPTTKIEGY
RIQRDPITLRPVELDCGPIIEPAEWYELQAWLDGRGRGKGLSRGQAILSAMDKLYCECGAVM
TSKRGEESIKDSYRCRRRKVVDPSAPGQHEGTCNVSMAALDKFVAERIFNKIRHAEGDEETL
ALLWEAARRFGKL(TEAPEKSGERANLVAERADA**L**NA**LE**E**LY**EDRAAGAYDGP**V**GRK**HFR**KQ
QAA**LT**LRQQGA**E**ERLAELEAAEAPKLPLDQWFPE)DADADPTGPKSWWGRASVDDKRVFVGL
FVDKIVVTKSTTGRGQGTPIEKRASITWAKPPTDDDEDDAQDGTEDVAATSHHHHHH

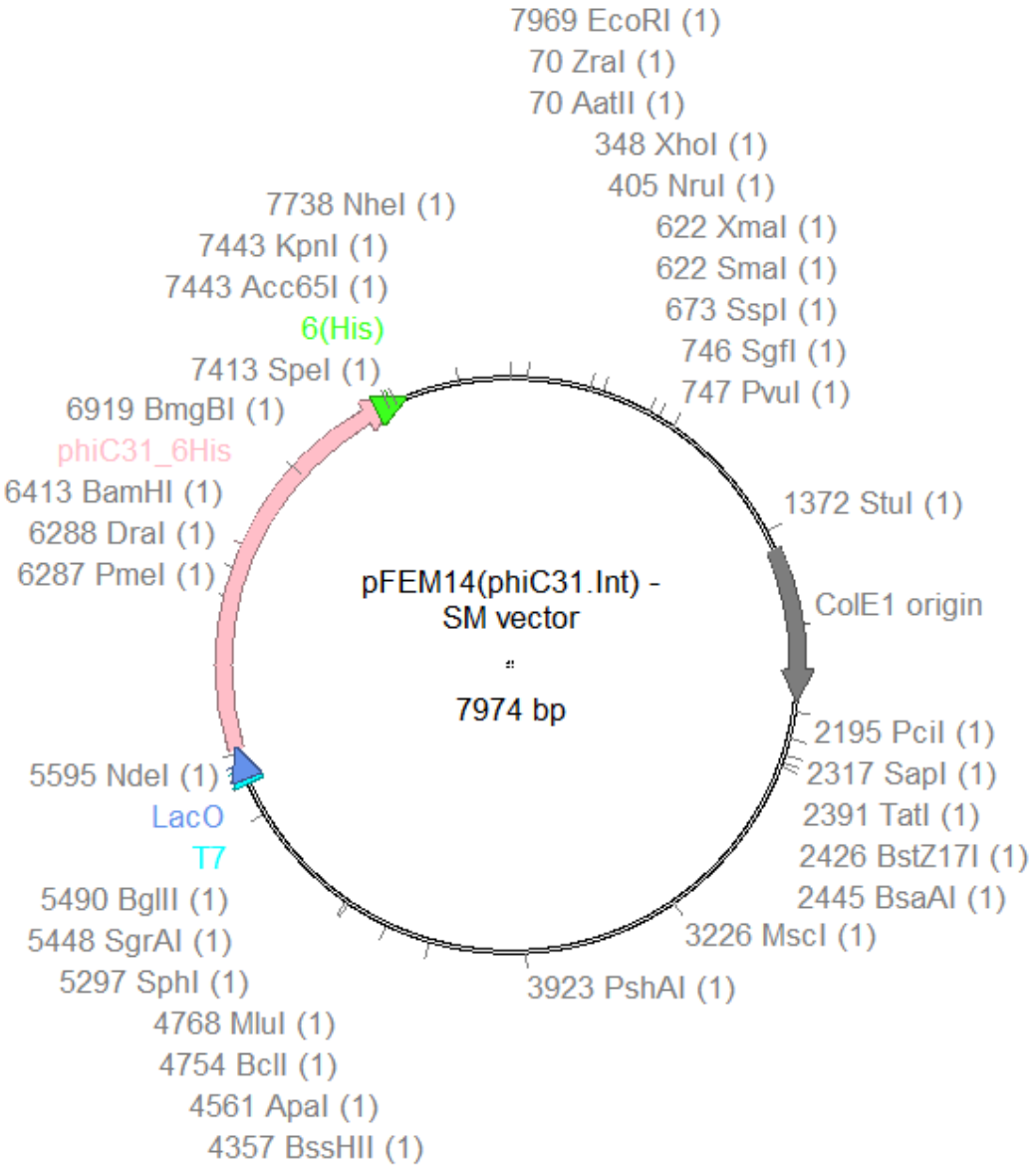
B.2: Amino-acid sequence of ΔCC φC31 Int mutant

AA sequence of the ΔCC φC31 Int mutant used in this study. Region from which the CC domain has been excised is indicated by a vertical line (|). The extra GT residues located after the excised fragment are a result of introduction of the restriction site used for the creation of the deletion.

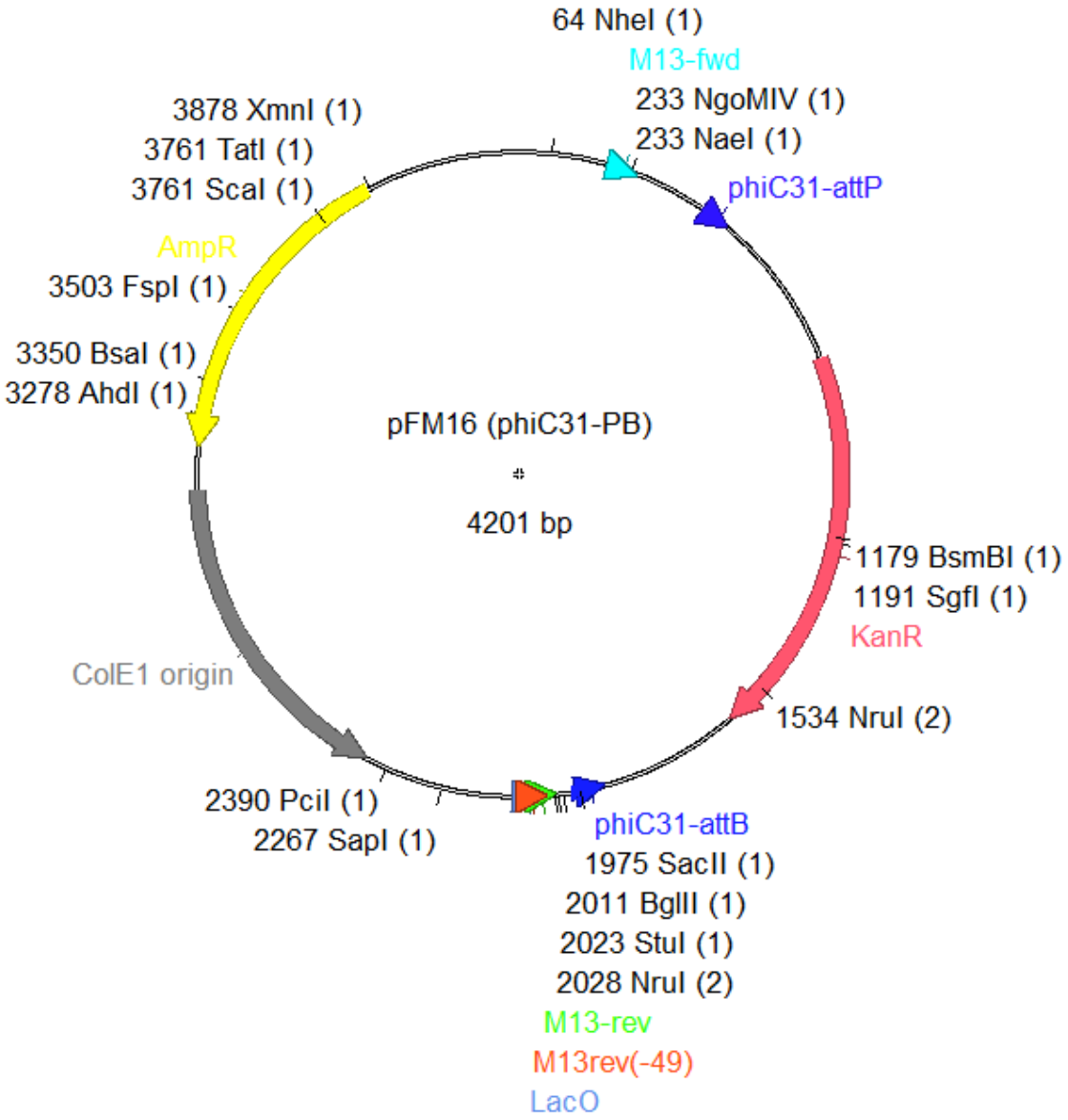
MDTYAGAYDRQSRERENSSAASPATQRSANEDKAADLQREVERDGGRFRFVGHFSEAPGTSA
FGTAERPEFERILNECRAGRLNMIIVYDVSRFSRLKVMDAIPIVSELLALGVTIVSTQEGVF
RQGNVMDLIHLIMRLDASHKESSLKSAKILDTKNLQRELGGYVGGKAPYGFELVSETKEITR
NGRMVNVVINKLAHSTTPLTGPFEFEPDVIRWWWREIKTHKHLPFKPGSQAAIHPGSITGLC
KRMDADAVPTRGETIGKKTASSAWDPATVMRILRDPRIAGFAAEVIYKKKPDGTPTTKIEGY
RIQRDPITLRPVELDCGPIIEPAEWYELQAWLDGRGRGKGLSRGQAILSAMDKLYCECGAVM
TSKRGEESIKDSYRCRRRKVVDPSAPGQHEGTCNVSMAALDKFVAERIFNKIRHAEGDEETL
ALLWEAARRFGKL|GTDADADPTGPKSWWGRASVDDKRVFVGLFVDKIVVTKSTTGRGQGTP
IEKRASITWAKPPTDDDEDDAQDGTEDVAATS

APPENDIX C - Plasmid maps:

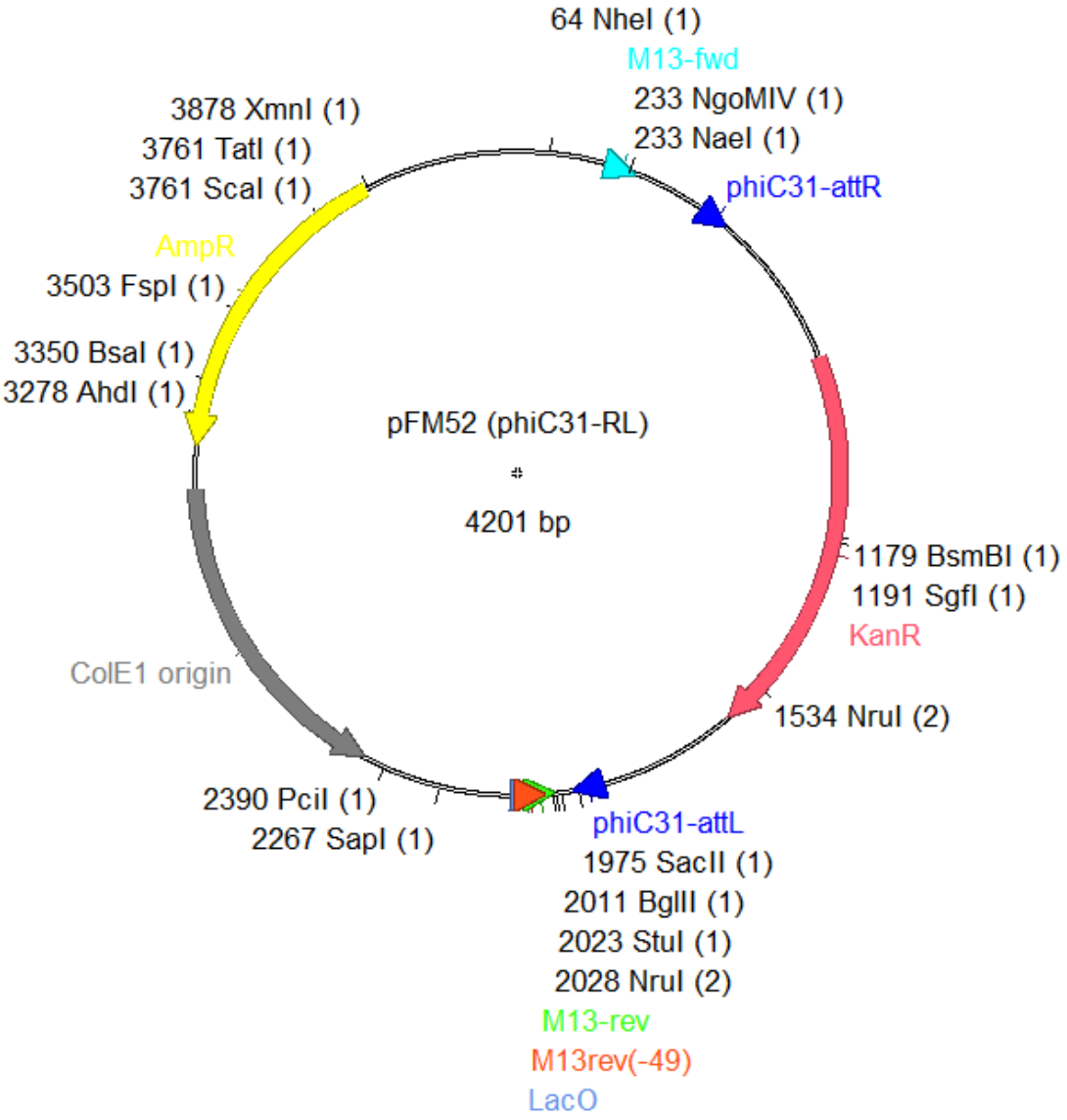
C.1: Map of pFEM14 (vector expressing φC31 Int at a high level when induced IPTG in BL21(DE3)pLysS; used for insertion of pre-existing mutant oligos and their expression)



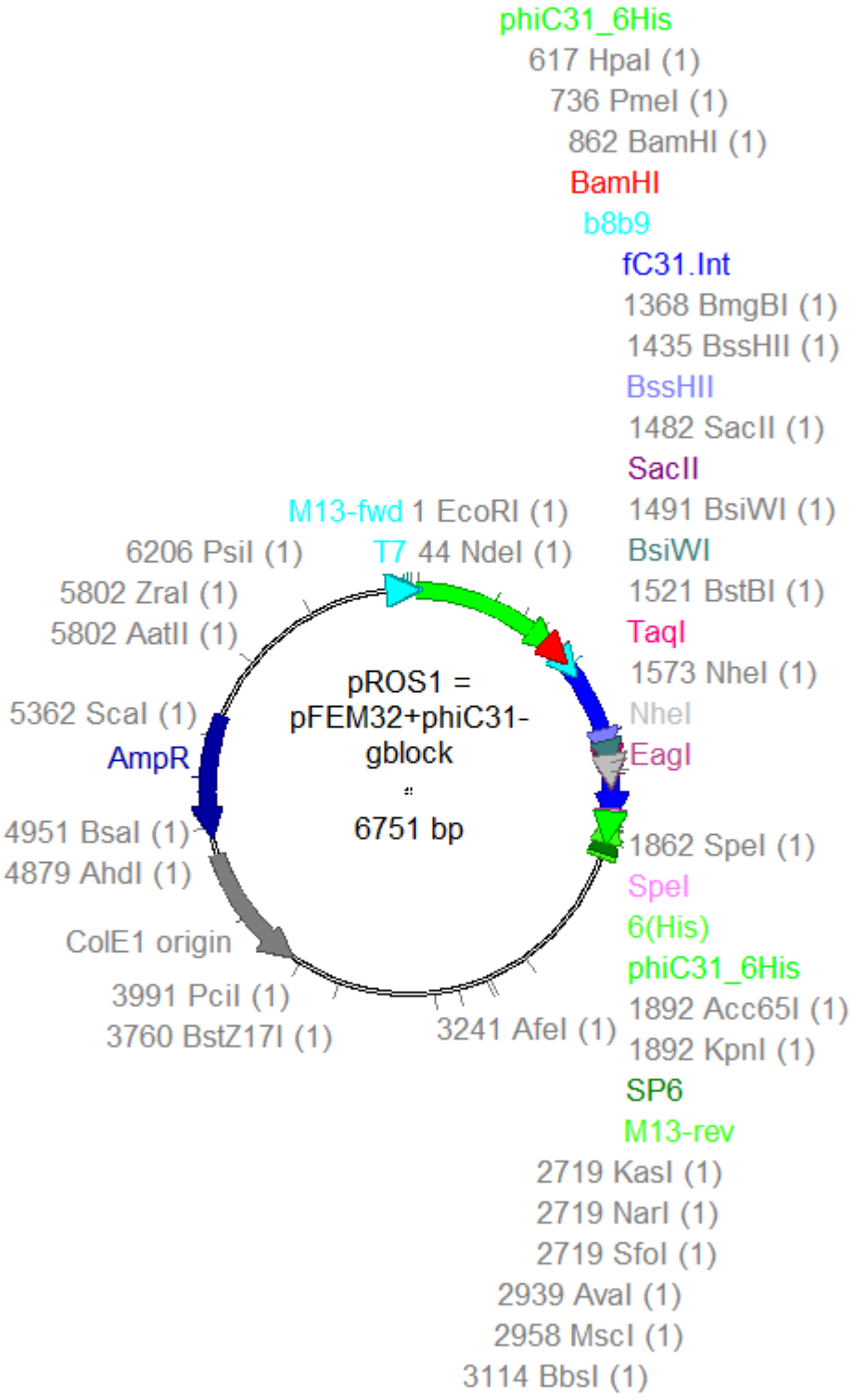
C.2: Map of pFM16 (substrate plasmid for recombination reaction containing φC31 Int *attP* and *attB* site)



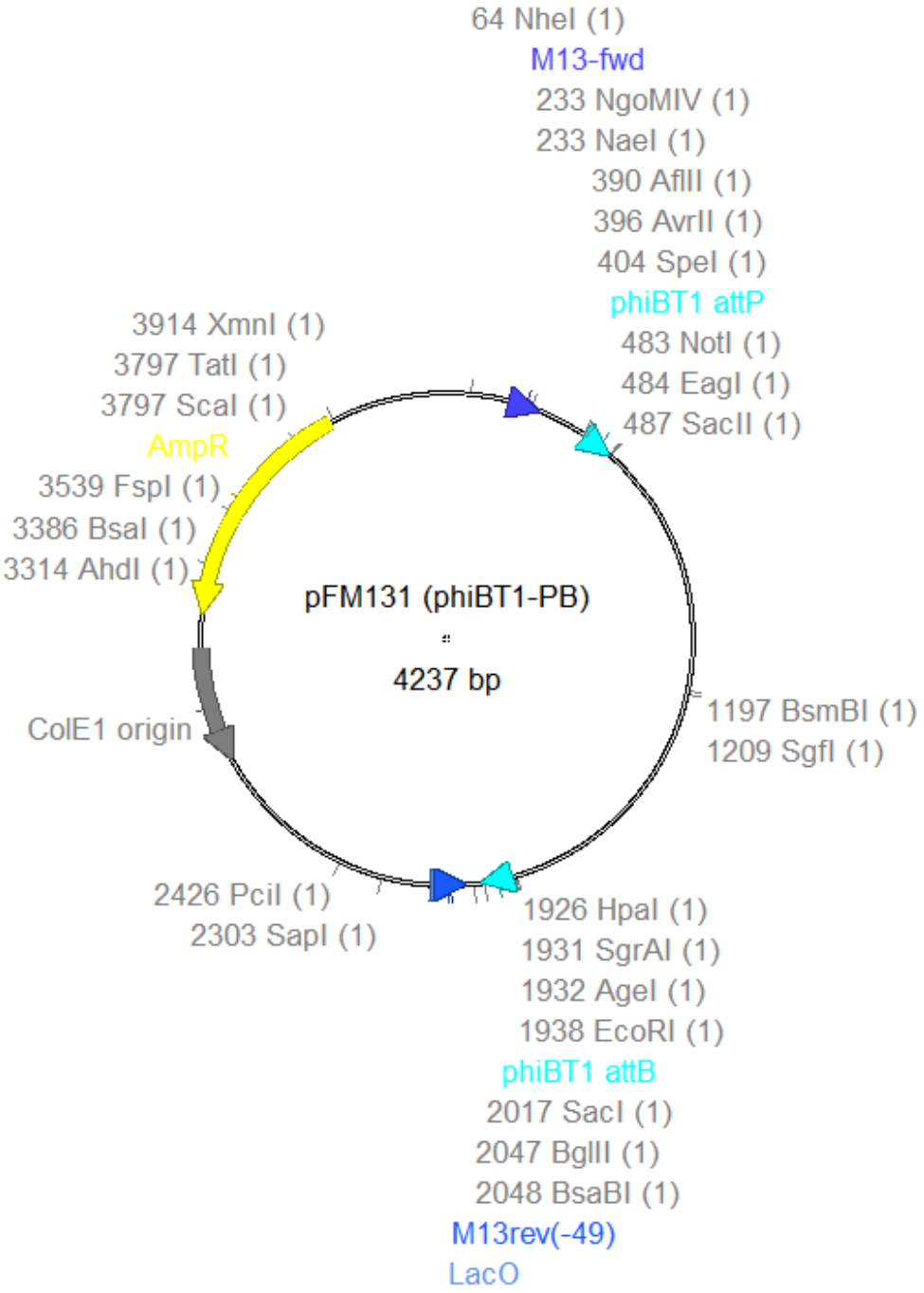
C.3: Map of pFM52 (substrate plasmid for recombination reaction containing φC31 Int *attL* and *attR* site)



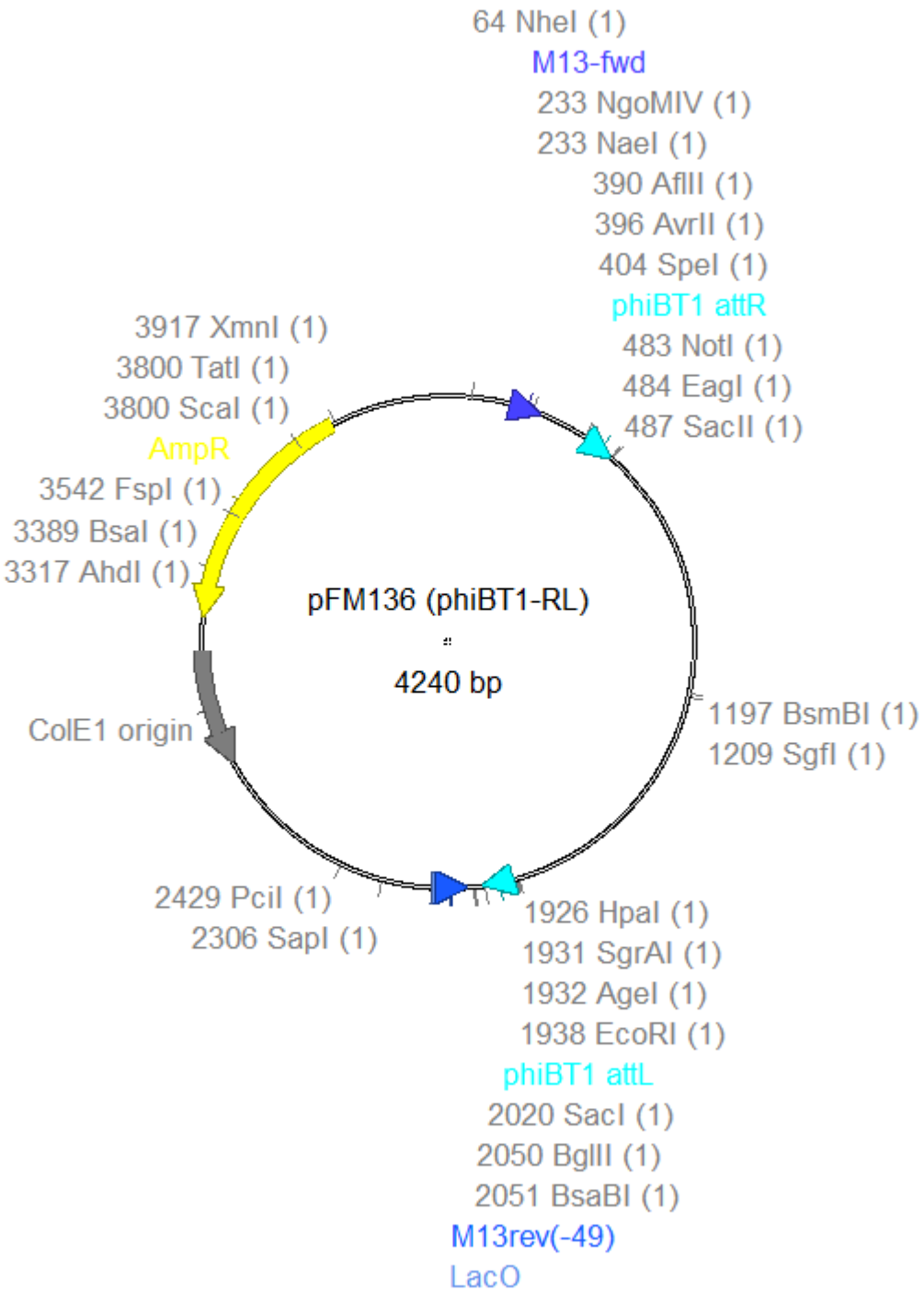
C.4 Map of pROS1 (plasmid vector containing the restriction cassette for generation of new mutants)



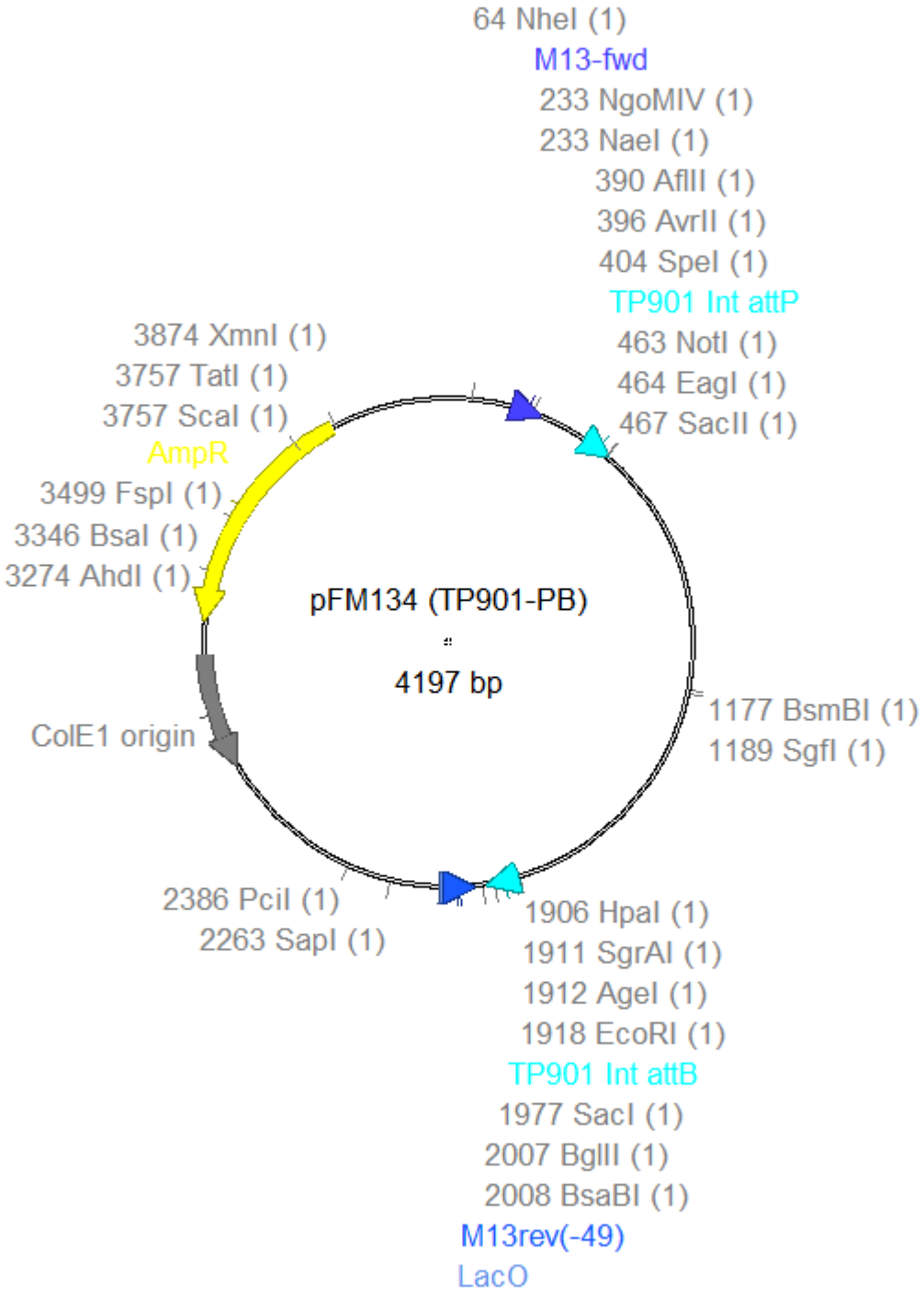
C.5 Map of pFM131 (substrate plasmid for recombination reaction containing φBT1 Int *attP* and *attB* site)



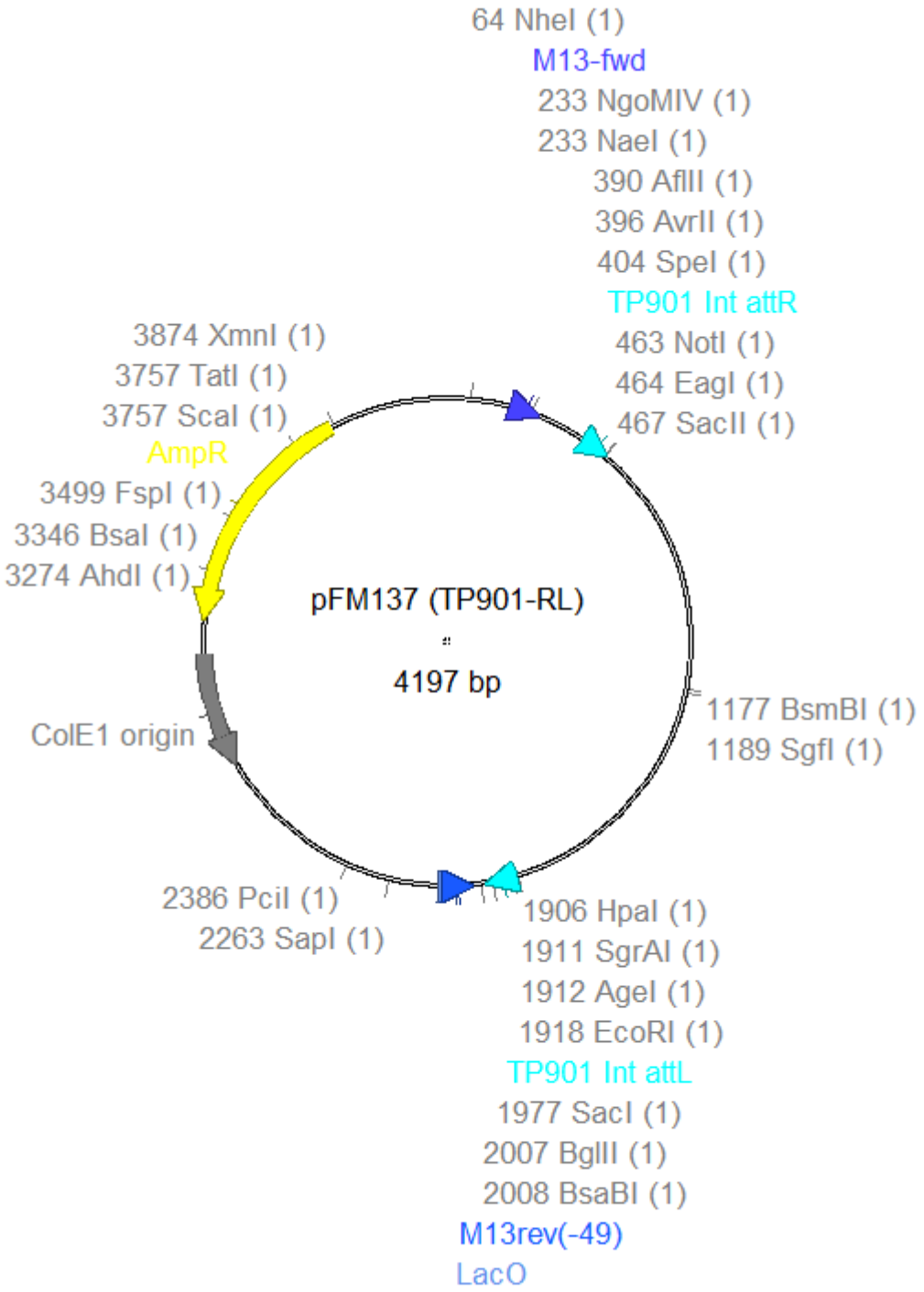
C.6 Map of pFM136 (substrate plasmid for recombination reaction containing φBT1 Int *attL* and *attR* site)



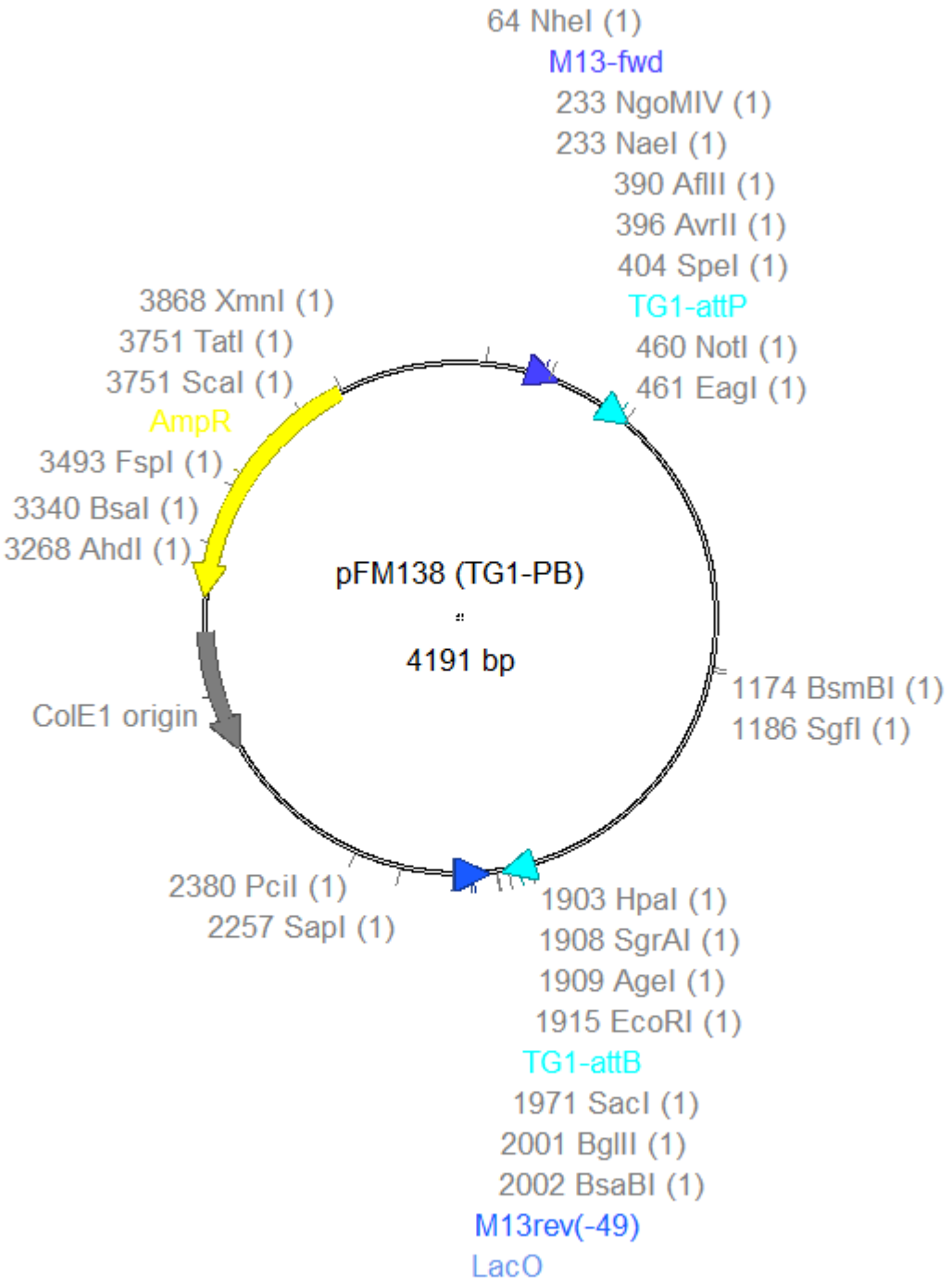
C.7 Map of pFM134 (substrate plasmid for recombination reaction containing TP901 Int *attP* and *attB* site)



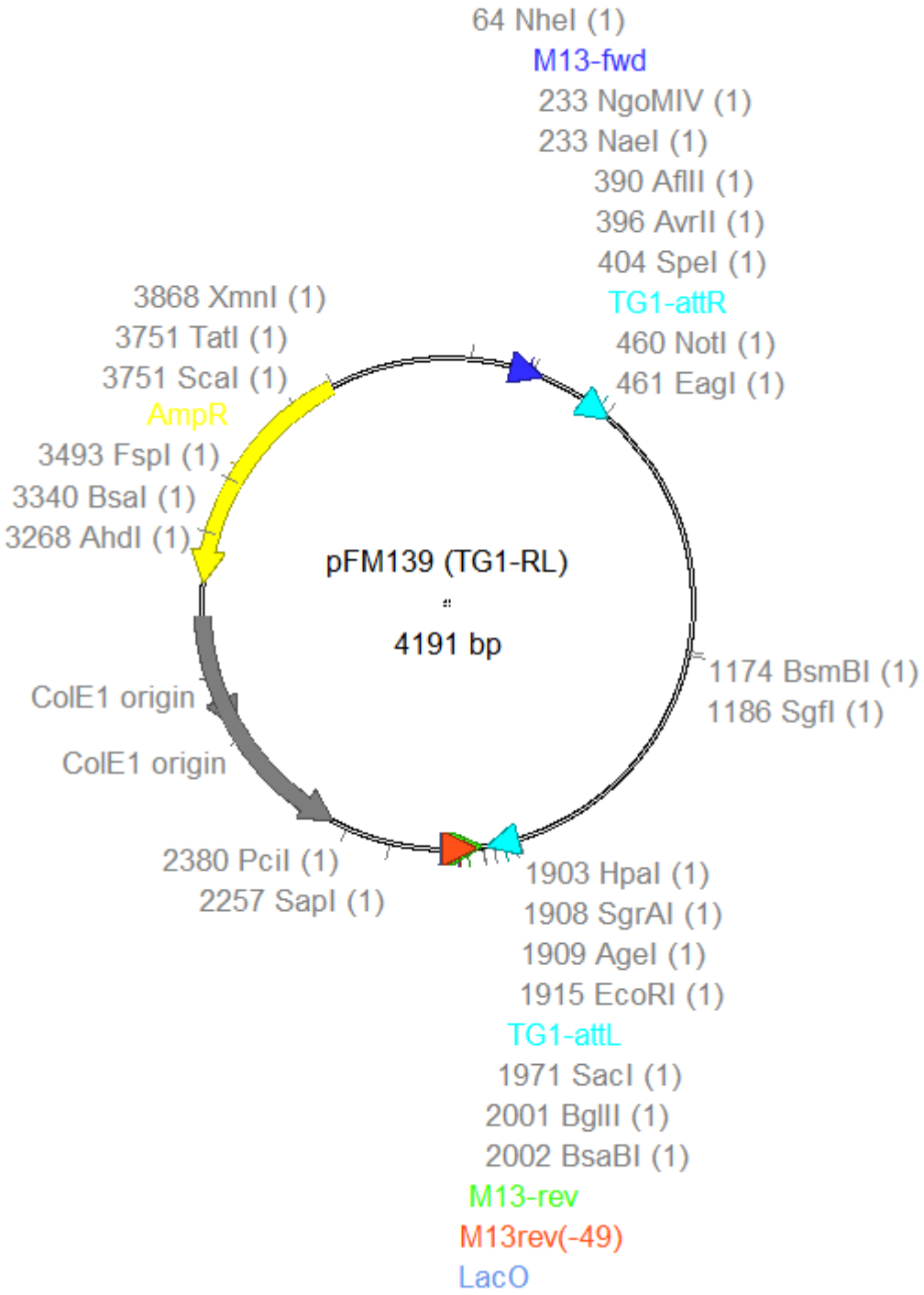
C.8 Map of pFM137 (substrate plasmid for recombination reaction containing TP901 Int *attL* and *attR* site)



C.9 Map of pFM138 (substrate plasmid for recombination reaction containing TG1 Int *attP* and *attB* site)



C.10 Map of pFM139 (substrate plasmid for recombination reaction containing TG1 Int *attL* and *attR* site)



APPENDIX D - Raw quantitative data:

D.1 Band intensity measurements from recombination efficiency study

Reaction	Reaction between *attP* and *attB* sites, no RDF																					
Sample	-ve	WT				ΔCC				L474A				Y475A				Y475H				
Concentration of the enzyme	0	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM
Sum of intensity of substrate bands	99.65	31.95	41.14	51.94	60.03	25.83	32.07	34.6	29.3	25.07	30.43	27.38	29.8	48.63	52.31	54.41	57.87	35.64	36.8	32.73	47.02
Sum of intensity of product bands	0.36	78.74	75.73	66.39	55.59	0	1.17	0	3.65	31.9	36.43	35.65	34.39	0	4.44	8.26	10.48	45.29	60.27	57.48	37.75
Estimated reaction efficiency (%)	0.35996	71.1356	64.7985	56.1058	48.0799	0	3.51986	0	11.0774	55.9944	54.487	56.5604	53.5753	0	7.82379	13.1801	15.3328	55.9619	62.0892	63.718	44.5323

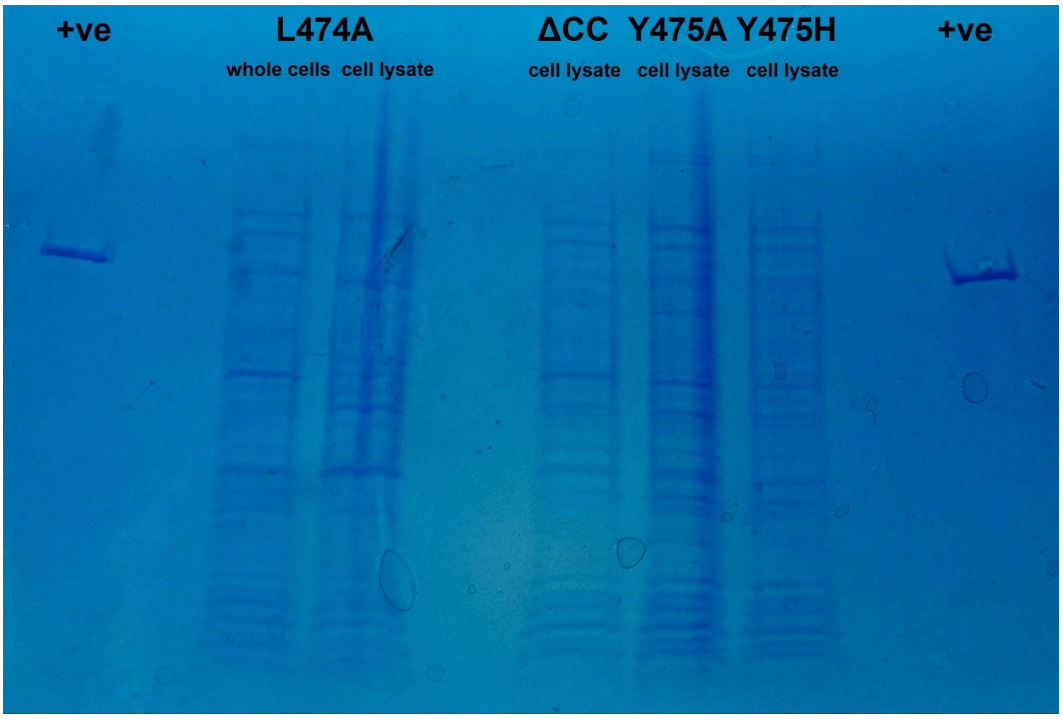
Reaction	Reaction between *attP* and *attB* sites, with RDF																					
Sample	-ve	WT				ΔCC				L474A				Y475A				Y475H				
Concentration of the enzyme	0	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM
Sum of intensity of substrate bands	99.65	81.09	68.75	69.19	68.74	184.45	189.53	157.85	149.83	115.73	106.6	102.19	108.58	163.45	118.24	114.86	136.56	104.2	95.1	181.95	133.46
Sum of intensity of product bands	0.36	14.58	36.64	46.41	43.3	8.89	13.72	0	4.67	23.75	23.87	12.46	20.84	8.44	0	0	15.45	24.81	33.54	108.1	122.3
Estimated reaction efficiency (%)	0.35996	15.2399	34.7661	40.1471	38.6469	4.59812	6.75031	0	3.02265	17.0275	18.2954	10.8679	16.1026	4.91012	0	0	10.1638	19.2311	26.0728	37.2694	47.8183

Reaction	Reaction between *attL* and *attR* sites, no RDF																					
Sample	-ve	WT				ΔCC				L474A				Y475A				Y475H				
Concentration of the enzyme	0	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM
Sum of intensity of substrate bands	99.65	89.43	98.43	101.72	100.25	116.58	124.2	119.42	107.84	138.97	137.14	142.8	152.99	102.47	147.64	129.22	148.66	149.34	134.13	145.06	162.12
Sum of intensity of product bands	0.36	0	0	5.18	6.9	22.04	17.06	1.71	0	34.26	22.23	1.65	6.19	19.64	8.67	7.22	6.53	12.07	1.37	7.53	19.81
Estimated reaction efficiency (%)	0.35996	0	0	4.84565	6.43957	15.8996	12.077	1.41171	0	19.7772	13.9487	1.14226	3.88868	16.0839	5.54667	5.2917	4.20775	7.47785	1.01107	4.93479	10.8888

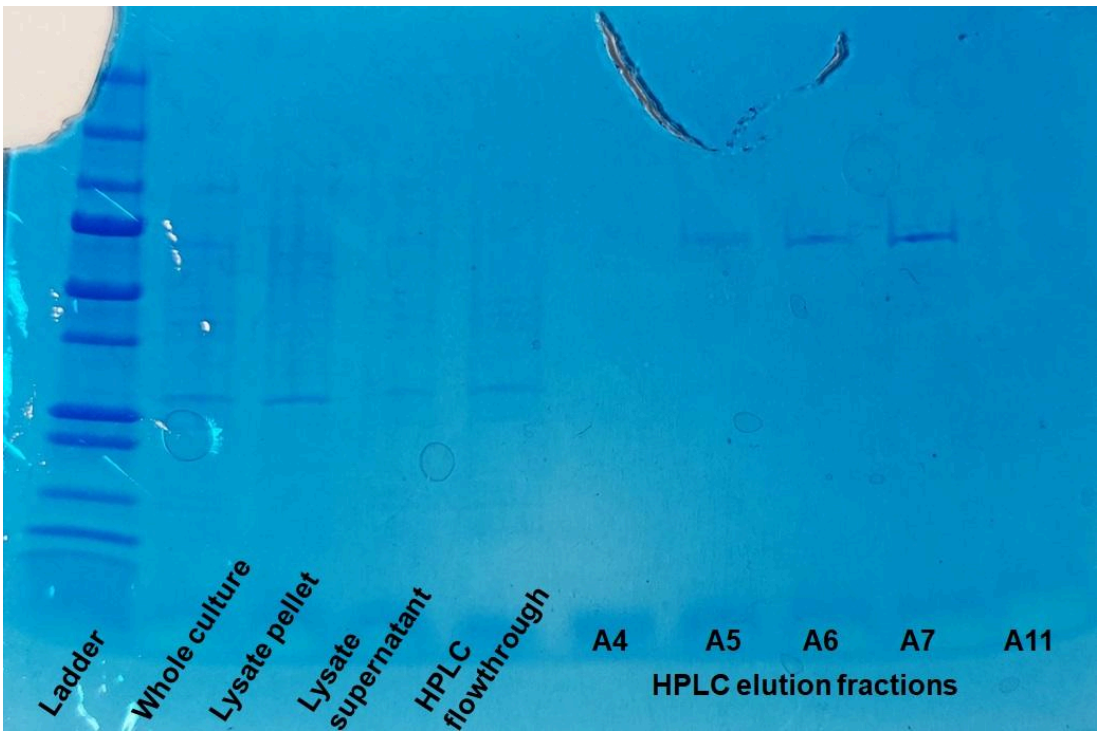
Reaction	Reaction between *attL* and *attR* sites, with RDF																					
Sample	-ve	WT				ΔCC				L474A				Y475A				Y475H				
Concentration of the enzyme	0	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM	2μM	4μM	8μM	16μM
Sum of intensity of substrate bands	99.65	112.37	72.85	27.96	33.08	134.04	145.03	145.11	160.34	186.95	183.4	196.01	162.94	143.1	120.57	142.77	110.05	141.2	189.04	109.03	135.33
Sum of intensity of product bands	0.36	22.41	85.37	111.09	109.72	0	0	0	0	112.4	58.04	54.34	36.4	0	0	0	0	68.84	18.74	12.03	9.96
Estimated reaction efficiency (%)	0.35996	16.6271	53.9565	79.8921	76.8347	0	0	0	0	37.548	24.0391	21.7056	18.2603	0	0	0	0	32.7747	9.01915	9.93722	6.85526

APPENDIX E - SDS-PAGE gel images:

E.1 Unfiltered cell lysates of expression strains used in production of mutants of interest. 8 μM WT φC31 Int stock was used as the positive control, whole cells pre-ligation included for L474A as additional control



E.2 Validation gel for HPLC curves used for purification of samples by fractioning (see Chapter 2.18). The ΔCC HPLC samples were used for this validation gel.



BIBLIOGRAPHY AND LIST OF REFERENCES:

1. Alsaleh, A., Holland, A., Shin, H., Reyes, T. P., Baksh, A., Taiwo-Aiyerin, O. T., Pigli, Y., Rice P. A., & Olorunniji, F. J. (2025). Large serine integrases utilise scavenged phage proteins as directionality cofactors. *Nucleic Acids Research*, 53(3), gkaf050.
2. Bonilla, B., Hengel, S. R., Grundy, M. K., & Bernstein, K. A. (2020). RAD51 gene family structure and function. *Annual review of genetics*, 54(1), 25-46.
3. Bonnet, J., Subsoontorn, P., & 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.
4. Chavez, C. L., & Calos, M. P. (2011). Therapeutic applications of the PhiC31 integrase system. *Current Gene Therapy,* 11(5), 375-381.
5. Chen, Y. W., Su, B. Y., Van Duyne, G. D., Fogg, P., & Fan, H. F. (2023). The influence of coiled-coil motif of serine recombinase toward the directionality regulation. *Biophysical Journal*, 122(24), 4656-4669.
6. Cho, E. H., Gumport, R. I., & Gardner, J. F. (2002). Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *Journal of bacteriology*, 184(18), 5200-5203.
7. Christiansen, B., Johnsen, M. G., Stenby, E., Vogensen, F. K., & Hammer, K. (1994). Characterization of the lactococcal temperate phage TP901-1 and its site-specific integration. *Journal of bacteriology*, 176(4), 1069-1076.
8. Combes, P., Till, R., Bee, S., & Smith, M. C. (2002). The *Streptomyces* genome contains multiple pseudo-*attB* sites for the φC31-encoded site-specific recombination system. *Journal of bacteriology*, 184(20), 5746-5752.
9. Dale, E. C., & Ow, D. W. (1991). Gene transfer with subsequent removal of the selection gene from the host genome. *Proceedings of the National Academy of Sciences*, 88(23), 10558-10562.
10. Farruggio, A. P., Chavez, C. L., Mikell, C. L., & Calos, M. P. (2012). Efficient reversal of phiC31 integrase recombination in mammalian cells. *Biotechnology Journal*, 7(11), 1332-1336.
11. Fogg, P. C., Colloms, S., Rosser, S., Stark, M., & Smith, M. C. (2014). New applications for phage integrases. *Journal of molecular biology*, 426(15), 2703-2716.
12. Fogg, P. C., Younger, E., Fernando, B. D., Khaleel, T., Stark, W. M., & Smith, M. C. (2018). Recombination directionality factor gp3 binds ϕC31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Research*, 46(3), 1308-1320.
13. Gregory, M. A., Till, R., & Smith, M. C. (2003). Integration site for *Streptomyces* phage φBT1 and development of site-specific integrating vectors. *Journal of bacteriology*, 185(17), 5320-5323.

14. Groth, A. C., Olivares, E. C., Thyagarajan, B., & Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Aacademy of Sciences*, 97(11), 5995-6000.
15. Gupta K., Sharp R., Yuan J. B., Li H., Van Duyne G. D. (2017). Coiled-coil interactions mediate serine integrase directionality. Nucleic Acids Research, 45(12), 7339–7353.
16. Hirano, N., Muroi, T., Kihara, Y., Kobayashi, R., Takahashi, H., & Haruki, M. (2011). Site-specific recombination system based on actinophage TG1 integrase for gene integration into bacterial genomes. *Applied microbiology and biotechnology*, 89(6), 1877-1884.
17. Kempe, K., Rubtsova, M., Berger, C., Kumlehn, J., Schollmeier, C., & Gils, M. (2010). Transgene excision from wheat chromosomes by phage phiC31 integrase. *Plant Molecular Biology*, 72, 673-687.
18. Krasnow, M. A., & Cozzarelli, N. R. (1983). Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented res sites. *Cell*, 32(4), 1313-1324.
19. Kuhstoss, S., & Rao, R. N. (1991). Analysis of the integration function of the streptomycete bacteriophage φC31. *Journal of molecular biology*, 222(4), 897-908.
20. Kuhstoss, S. A. (1992). Analysis of the site-specific integration function of the streptomycete bacteriophage phiC31. Doctoral dissertation, Purdue University.
21. Lister, J. A. (2011). Use of phage PhiC31 integrase as a tool for zebrafish genome manipulation. *Methods in Cell Biology*, 104, 195-208.
22. Mandali, S., Gupta, K., Dawson, A. R., Van Duyne, G. D., & Johnson, R. C. (2017). Control of recombination directionality by the Listeria phage A118 protein Gp44 and the coiled-coil motif of its serine integrase. *Journal of bacteriology,* 199(11), 10-1128.
23. Mandali, S., & Johnson, R. C. (2021). Control of the serine integrase reaction: roles of the coiled-coil and helix E regions in DNA site synapsis and recombination. *Journal of Bacteriology*, 203(16), 10-1128.
24. Mason, J. M., & Arndt, K. M. (2004). Coiled coil domains: stability, specificity, and biological implications. *ChemBioChem*, 5(2), 170-176.
25. Masson, J. Y., & West, S. C. (2001). The Rad51 and Dmc1 recombinases: a non-identical twin relationship. *Trends in biochemical sciences*, 26(2), 131-136.
26. McEwan, A. R., Rowley, P. A., & Smith, M. C. (2009). DNA binding and synapsis by the large C-terminal domain of ϕC31 integrase. *Nucleic acids research*, 37(14), 4764-4773.
27. McEwan, A. R., Raab, A., Kelly, S. M., Feldmann, J., & Smith, M. C. (2011). Zinc is essential for high-affinity DNA binding and recombinase activity of φ C31 integrase. *Nucleic acids research,* 39(14), 6137-6147.
28. Mouw, K. W., Rowland, S. J., Gajjar, M. M., Boocock, M. R., Stark, W. M., & Rice, P. A. (2008). Architecture of a serine recombinase-DNA regulatory complex. *Molecular cell*, 30(2), 145-155.
29. Nern, A., Pfeiffer, B. D., Svoboda, K., & Rubin, G. M. (2011). Multiple new site-specific recombinases for use in manipulating animal genomes. *Proceedings of the National Academy of Sciences*, 108(34), 14198-14203.

30. Nöllmann, M., Byron, O., & Stark, W. M. (2005). Behavior of Tn3 resolvase in solution and its interaction with res. *Biophysical journa*l, 89(3), 1920-1931.
31. Olorunniji, F. J., Buck, D. E., Colloms, S. D., McEwan, A. R., Smith, M. C., Stark, W. M., & Rosser, S. J. (2012). Gated rotation mechanism of site-specific recombination by ϕC31 integrase. *Proceedings of the National Academy of Sciences*, 109(48), 19661-19666.
32. Olorunniji, F. J., Merrick, C., Rosser, S. J., Smith, M. C., Stark, W. M., & Colloms, S. D. (2017). Multipart DNA assembly using site-specific recombinases from the large serine integrase family. In *Site-Specific Recombinases: Methods and Protocols* (pp. 303-323). New York, NY: Springer New York.
33. Pokhilko, A., Zhao, J., Ebenhöh, O., Smith, M. C., Stark, W. M., & Colloms, S. D. (2016). The mechanism of ϕC31 integrase directionality: experimental analysis and computational modelling. *Nucleic Acids Research*, 44(15), 7360-7372.
34. PubChem. (2025) *CID 5950. L-Alanine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/5950.
35. PubChem. (2025) *CID 6057. L-Tyrosine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/6057.
36. PubChem. (2025) *CID 6106. L-Leucine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/6106.
37. PubChem. (2025) *CID 6274. L-Histidine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/6274.
38. PubChem. (2025) *CID 6274. L-Isoleucine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/6306.
39. PubChem. (2025) *CID 6322. L-Arginine.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/6322.
40. PubChem. (2025) *CID 33032. L-Glutamic Acid.* National Library of Medicine: National Center for Biotechnology Information. Retrieved 22.09.2025 from: https://pubchem.ncbi.nlm.nih.gov/compound/33032.
41. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S., & Lu, T. K. (2016). Synthetic recombinase-based state machines in living cells. *Science*, 353(6297), aad8559.
42. 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.
43. Rutherford, K., & Van Duyne, G. D. (2014). The ins and outs of serine integrase site-specific recombination. *Current opinion in structural biology*, 24, 125-131.

44. Sarkis, G. J., Murley, L. L., Leschziner, A. E., Boocock, M. R., Stark, W. M., & Grindley, N. D. (2001). A model for the γδ resolvase synaptic complex. *Molecular Cell*, 8(3), 623-631.
45. Siuti, P., Yazbek, J., & Lu, T. K. (2014). Engineering genetic circuits that compute and remember. *Nature Protocols*, 9(6), 1292-1300.
46. Smith, M. C., Burns, R. N., Wilson, S. E., & Gregory, M. A. (1999). The complete genome sequence of the *Streptomyces* temperate phage φC31: evolutionary relationships to other viruses. *Nucleic Acids Research*, 27(10), 2145-2155.
47. Smith, M. C., Brown, W. R., McEwan, A. R., & Rowley, P. A. (2010). Site-specific recombination by φC31 integrase and other large serine recombinases. *Biochemical Society Transactions*, 38(2), 388-394.
48. Smith, M. C. (2015). Phage‐encoded serine integrases and other large serine recombinases. *Mobile DNA iii*, 253-272.
49. Stark, W. M., Boocock, M. R., & Sherratt, D. J. (1989). Site-specific recombination by Tn3 resolvase. *Trends in Genetics*, 5, 304-309.
50. Stark, W. M. (2017). Making serine integrases work for us. *Current opinion in microbiology*, 38, 130-136.
51. Sun, Y. E., Aspinall, L., Joseph, A. P., Colloms, S. D., Stark, W. M., & Spagnolo, L. (2025). Structural basis of DNA recombination catalysis and regulation by ϕC31 integrase. *bioRxiv*, 2025-05.
52. Thorpe H. M., Smith M. C. M. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA*. 95:5505–5510.
53. Thorpe, H. M., Wilson, S. E., & Smith, M. C. (2000). Control of directionality in the site‐specific recombination system of the *Streptomyces* phage φC31. *Molecular microbiology*, 38(2), 232-241.
54. Truebestein, L., & Leonard, T. A. (2016). Coiled‐coils: The long and short of it. *Bioessays*, 38(9), 903-916.
55. UniProt. (2025). *Q9T221 · Q9T221_BPPHC. Streptomyces phage phiC31 (Bacteriophage phi-C31) Integrase*. UniProt: the Universal Protein Knowledgebase. Retrieved 05.05.2025 from: https://www.uniprot.org/uniprotkb/Q9T221/entry.
56. UniProt. (2025). *Q9T216 · RDF_BPPHC. Streptomyces phage phiC31 (Bacteriophage phi-C31) Recombination directionality factor*. UniProt: the Universal Protein Knowledgebase. Retrieved 05.05.2025 from: https://www.uniprot.org/uniprotkb/Q9T216/entry.
57. Van Duyne, G. D., & Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. *Critical reviews in biochemistry and molecular biology*, 48(5), 476-491.
58. Wang, Y., Yau, Y. Y., Perkins-Balding, D., & Thomson, J. G. (2011). Recombinase technology: applications and possibilities. *Plant cell reports*, 30, 267-285.
59. Yuan, P., Gupta, K., & Van Duyne, G. D. (2008). Tetrameric structure of a serine integrase catalytic domain. *Structure,* 16(8), 1275-1286.

60. Zhang, L., Ou, X., Zhao, G., & Ding, X. (2008). Highly efficient in vitro site-specific recombination system based on Streptomyces phage φBT1 integrase. *Journal of bacteriology*, 190(19), 6392-6397.
61. Zhao, J., Pokhilko, A., Ebenhöh, O., Rosser, S. J., & Colloms, S. D. (2019). A single-input binary counting module based on serine integrase site-specific recombination. *Nucleic Acids Research*, 47(9), 4896-4909.