Sequence selectivity of the resolvase catalytic domain: implications for Z-resolvase design

Marko Milenkov Prorocic
B.Sc. M.Res

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

University of Glasgow
Faculty of Biomedical and Life Sciences
Department of Molecular and Cellular Biology

May 2009

Summary

The extent of sequence specificity of the Tn3 resolvase catalytic domain was investigated by creating libraries of Tn3 site I variants in which all of the central 16 bp were systematically randomised in overlapping 4 bp blocks and recombination deficient and recombination proficient site I variants were selected using two different independent selection strategies employing an activated Tn3 resolvase mutant NM. A degree of flexibility in the sequences permitted in the central 16 bp of the Tn3 site I was observed especially at the positions 4, 7 and 8, but accumulating changes was found to be in general detrimental to recombination. The data was compared to the naturally occurring site I sequences associated with proteins from the Tn3 resolvase family, and integrated with the available structural information revealing a number of residues in the extended arm region that could account for the sequence selectivity observed. The sequence selectivity of the activated Tn3 resolvase NM catalytic domain was tested in the Z-resolvase context employing a similar but less exhaustive selection strategy using a purified Z-resolvase Z-R(NM). Z-resolvases with sequence selectivity that is different to that of Z-R(NM) were constructed using catalytic domains of activated mutants of Sin and Tn21 resolvases and their *in vivo* and *in vitro* properties were tested, highlighting the universality of the Z-resolvase approach and its potential for the future applications. A number of issues concerning the Z-resolvase design such as the optimum length of Z-sites, what is the effect of the Zif268 DNA-binding domain on catalytic activity i.e. is it activating or inhibiting, is symmetry a prerequisite in the design of Z-sites or can a Z-resolvase catalyse recombination on sites with an odd number of bases between Zif268 binding sites i.e. one half-site longer than the other, what is the relative influence of the Z-resolvase linker length, and can Z-resolvase be complemented by resolvase and catalyse recombination on appropriately designed hybrid sites were explored. The sequence selectivity of catalytic domains of Sin and Tn21 resolvases was compared using a combination of a mutant library selection strategy and the Sin-Tn21 resolvase hybrid experiments. An attempt to change the sequence selectivity of Tn3 resolvase catalytic domain into that of Sin resolvase, both in the resolvase and Z-resolvase context by mutating the specific residues, implicated in catalytic domain sequence selectivity was performed. The sequence selectivity of activated Tn3 resolvase catalytic domain was successfully changed into that of Sin resolvase.

Acknowledgements

While performing the experiments was an exciting and fulfilling experience writing this thesis was probably the hardest and most challenging thing that I have ever done. I would like to thank all the people who helped me along the way, supported me and were there for me in my darkest hours. I would like to thank my lovely Sarah for immense patience and belief in me that she has shown in putting up with my endless writing and my endless complaining about it. My best friend Caroline for her selfless support and encouragement over the years: Thank you for bringing me food, for waiting for me to finish experiments and for never ever letting me down. Gero, Femi and Duncan , thank you for all your valiant efforts in helping me get this done. Marshall for giving me an opportunity to work on such an exciting project, for helpful discussions and many great suggestions and ideas. Sally and Martin for taking care of me in the lab, you guys were like my second parents, I will never forget you. Sean, Al, Amy, Elizabeth, Arlene, Chris, Mary, Aram, Lilly, Marcelo and numerous others that have made the resolvase group such a positive and thriving place to work in. I'd like to thank the Wellcome Trust without whose generous financial support none of this would be possible.

Last but by no means least I'd like to thank my family. Although you were far away your thoughts were always with me. Хвала вам на свему. Без вас не бих успео да ово завршим.

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

Marko M. Prorocic

Abbreviations

Units

k	10^3	g	grams
c	10^{-2}	m	metres
m	10^{-3}	l	litres
µ	10^{-6}	C°	degrees Celsius
n	10^{-9}	M	molar
p	10^{-12}	mol	moles
A	Ampères	bp	basepairs
V	Volts	rpm	revolutions per minute
W	Watts	OD	optical density
h	hours	min	minutes

Chemicals/Reagents

AcOH	acetic acid
APS	ammonium persulphate
ATP	adenosine triphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid (disodium salt)
EtBr	ethidium bromide
KOAc	potassium acetate
PAGE	polyacrylamide gel electrophoresis
SDS	sodium dodecyl sulphate
TAE	tris-acetate-EDTA (electrophoresis buffer)
TBE	tris-borate-EDTA (electrophoresis buffer)
TMED	*N,N,N',N'*-tetramethylethylenediamine
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

Symbol	Meaning	Origin of designation
G	G	Guanine
A	A	Adenine
T	T	Thymine
C	C	Cytosine
R	G or A	puRine
Y	T or C	pYrimidine
M	A or C	aMino
K	G or T	Ketone
S	G or C	Strong interaction (three H bonds)
W	A or T	Weak interaction (two H bonds)
H	A or C or T	not-G, H follows G in the alphabet
B	G or T or C	not-A, B follows A
V	G or C or A	not-T (not-U), V follows U
D	G or A or T	not-C, D follows C
N	G or A or T or C	aNy

Extended IUPAC DNA notation (Recommendations 1984, *Biochem J,* 1985)

1 Chapter 1: Introduction

1.1 Recombination: types and natural roles

DNA, the main storage of genetic information in cells is not a static unchanging molecule but a dynamic entity that is subject to modifications, rearrangements and topological changes brought about by the process of recombination.

There are two main types of recombination, namely homologous and site-specific recombination, both of which are found in prokaryotes and eukaryotes. There are other kinds of DNA rearrangements that are usually grouped together with the two recombination types already mentioned such as transposition (Siefert, 2009), retro-transposition (Callinan & Batzer, 2006) and VDJ recombination (Jones & Gellert, 2004) however these phenomena although fascinating in their own right are not the subject of this review.

1.1.1 Homologous recombination

Homologous recombination is an exchange between two regions of DNA that is characterised by its requirement for extensive sequence similarity (identity) between the recombining sequences along with the need for the complex protein machinery. The two regions of identical or highly similar sequence align with each other, and with the help of protein machinery form a structure called the Holliday intermediate (Holliday, 1964; Paques & Haber, 1999; Ariyoshi *et al.*, 2000; Barzel & Kupiec, 2008; San Filippo *et al.*, 2008). In this structure two strands of the two homologous DNA duplexes form a crossover while becoming partially paired to the other two complementary strands of the parental duplex.

Homologous recombination plays an essential role in generating genetic diversity (meiosis), maintaining and propagating genomes (DNA repair and restarting replication forks), controlling gene expression and even moving DNA of some viruses into and out of host chromosomes, however as it is not a subject of this study it will not be discussed further.

1.1.2 Site-specific recombination

In contrast to homologous recombination, in site-specific-recombination, DNA strand exchange takes place between segments possessing only a limited degree of sequence homology (Kolb, 2002; Coates *et al.*, 2005). Site-specific recombinases perform rearrangements of DNA segments by recognising and binding to short DNA sequences

(sites), at which they cleave the DNA backbone, exchange the two DNA helices involved and rejoin the DNA strands.

In some site-specific recombination systems having just a recombinase enzyme together with the recombination sites is all that is required to be able to perform all these reactions, but in some other systems a number of accessory proteins and accessory sites are also needed.

The recombination sites are typically between 30 and 200 nucleotides in length and can consist of a range of subsites. Recombination takes place at the site (or a subsite) consisting of two motifs with a partial inverted-repeat symmetry, to which the recombinase binds, which flank a central crossover sequence. The pairs of sites between which the recombination occurs are usually identical, but there are exceptions e.g. *att*P and *att*B of λ integrase (Landy, 1989).

The reaction catalysed by the recombinase may lead to the excision of the DNA segment flanked by the two sites, but also to the integration or inversion of the orientation of the flanked DNA segment (Fig 1.1). The outcome of the reaction is dictated by the relative location and the orientation of sites that are to be recombined, and also by the innate specificity of the recombination system in question.

Excisions occur if the recombination takes place between two sites that are found on the same molecule (intramolecular recombination), and if the sites are in the same (direct repeat) orientation. Inversions on the other hand take place if the recombination takes place between two sites that are found on the same molecule and if the sites are in an opposite orientation (inverted repeat).

Alternatively, if the recombination occurs on sites that are situated on two different DNA molecules (intermolecular recombination), an integration (fusion) reaction occurs provided that at least one of these molecules is circular. Most site-specific systems are highly specialised catalysing only one of these different types of reaction and have evolved to avoid recombination between sites that are in a ‘wrong’ relationship.

In nature, site-specific recombination systems can be highly specific, fast and efficient, even when faced with complex eukaryotic genomes (Bode *et al.,* 2000; Sauer, 1998). As such, they are employed in number of processes such as: bacterial genome replication, differentiation and pathogenesis, movement of genetic elements such as transposons,

plasmids, phages and integrons (Nash, 1996) and present an attractive starting material for development of potential 'genetic engineering' tools (Akopian & Stark 2005).

1.2 Classification of site-specific recombinases: tyrosine vs. serine recombinases

Based on amino acid sequence homology and mechanistic relatedness, site-specific recombinases are grouped into one of two families: the tyrosine recombinase family or the serine recombinase family. The names stem from the conserved nucleophilic amino acid residue that the recombinase uses to attack the DNA and which becomes covalently linked to the DNA during strand exchange. The serine recombinase family used to be known as resolvase/invertase family, while tyrosine recombinases were known as the integrase family, which reflected the types of reaction that most known members in each family have evolved to catalyse. However, since serine recombinases that catalyse integration and tyrosine recombinases that catalyse resolution reactions have been discovered these old recombinase family names are becoming increasingly historic.

Typical examples of tyrosine recombinases are the well known enzymes such as Cre (from the P1 phage), FLP (from yeast *S. cerevisiae*) and λ integrase (from λ phage) while well-known serine recombinases include enzymes such as: γδ resolvase (from the Tn*1000* transposon) and ϕC31 integrase (from the ϕC31 phage) (Nash, 1996; Stark & Boocock 1995).

Although individual members of the two recombinase families can perform reactions with the same practical outcomes (e.g. excision or inversion), the two familes are unreleated to each other, having different protein structures and reaction mechanisms.

Unlike tyrosine recombinases, serine recombinases are highly modular, with their catalytic and DNA binding functions largely separated on different domains, as was initially hinted by biochemical studies (Abdel-Meguid *et al.*, 1984) and later shown by crystallographic structures (Yang & Steitz, 1995; Li *et al.*, 2005; Mouw *et al.*, 2008); a fact which could prove useful when attempting to reengineer these proteins as tools for genetic manipulation.

1.3 Mechanism and chemistry of the site-specific reaction

Recombination between two DNA sites begins by recognition and binding of these sites by the recombinase protein (Fig. 1.2). This is followed by synapsis i.e. bringing the sites

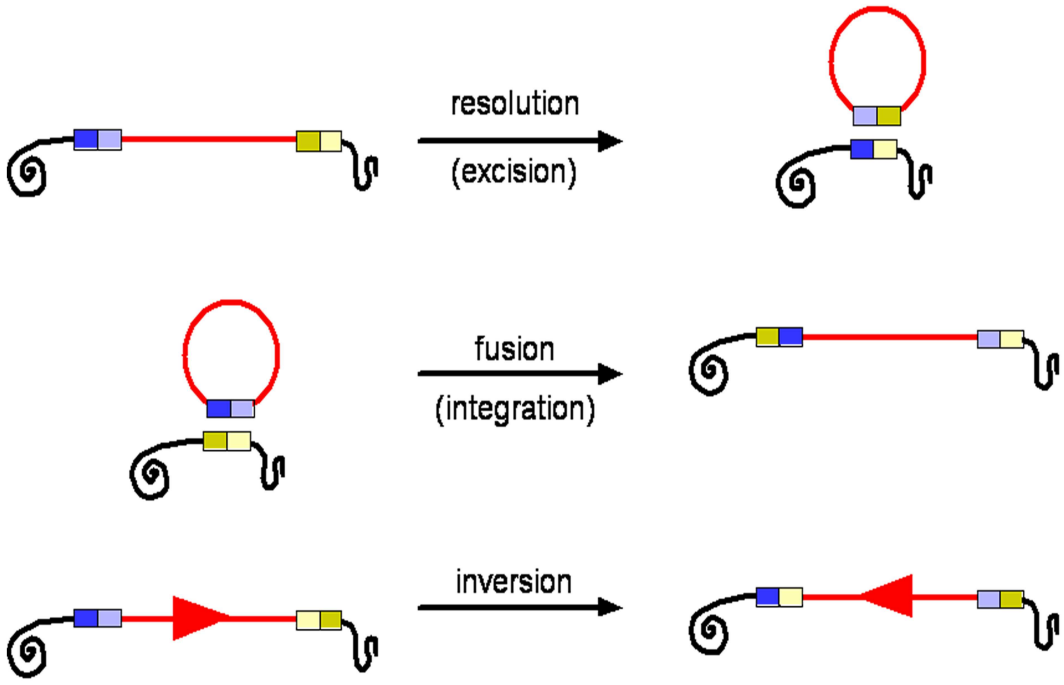


Figure 1.1- A diagrammatic representation of reactions catalysed by site-specific recombinases. Recombination sites are presented by blue and brown squares, intervening DNA is given in red while the flanking DNA is given in black. At the top of the figure resolution or excision reaction is portrayed, in which the 'red' DNA found between two directly repeated recombination sites is cut out from the 'black' DNA molecule, forming a circle. In the middle of the figure the fusion of the integration reaction is shown, which is essentially the reverse of resolution or excision. At the bottom of the figure an inversion reaction is shown. The orientation of the 'red' DNA segment (indicated by the red triangle) changes due to the recombination between recombination sites in inverted repeat.

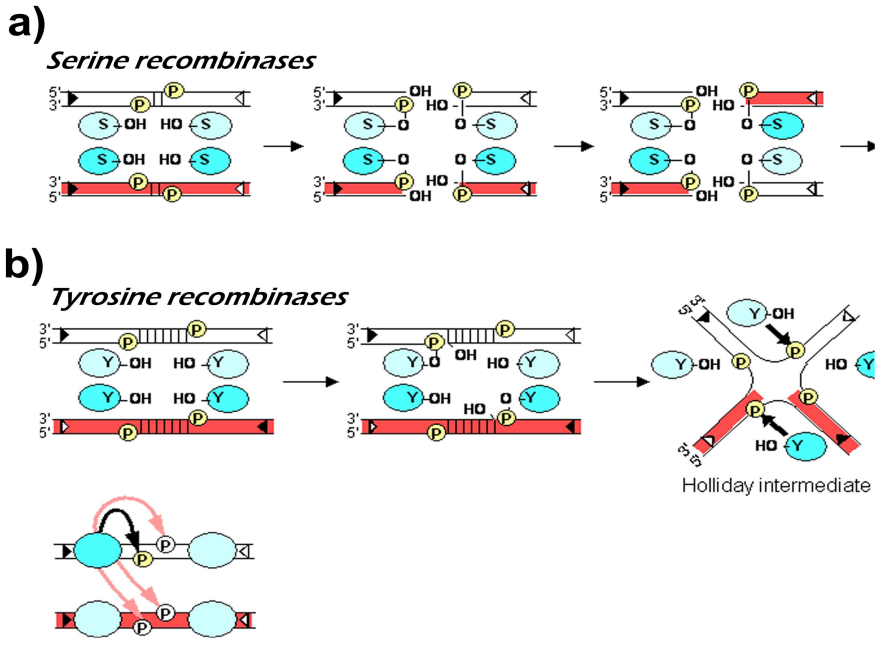


Figure 1.2- Mechanism of recombination by serine and tyrosine r
The scheme illustrates the cleavage, subunit rotation, and religatior
recombination catalysed by the serine recombinases. The cartoon s
complex, formed from a pair of resolvase-bound site Is. Only the c
of the resolvase subunits (blue ovals) are shown. The small yellow
a letter "P" represent the phosphate groups attacked by the recombi
inverted black and white arrowheads represent the ends of site I. Tl
are shown as red and white rectangles delineated by black lines. St
assumed to involve rotation of the resolvase subunits in covalent at
DNA, followed by the religation step. **b)** The scheme illustrates the
strand exchange by the tyrosine recombinases. The synaptic compl
DNA duplexes (red and white) bound by four recombinase subunit
arranged in a head-to-tail fashion. One strand from each duplex is c
exchanged and ligated to form a Holliday junction intermediate. Isc
this junction switches catalytic activity in the synapse between the
recombinase subunits. (Adapted from Stark *et al.,* 1992).

together to form the synaptic complex. It is within this synaptic complex that strand exchange takes place, as the DNA is cleaved and rejoined by controlled transesterification reactions. During strand exchange, the DNA cut at fixed points within the crossover region of the site releases a deoxyribose hydroxyl group, while the recombinase protein forms a transient covalent bond to a DNA backbone phosphate. This phosphodiester bond between hydroxyl group of the nucleophilic serine or tyrosine residue and the DNA conserves the energy that was expended in cleaving the DNA. Energy stored in this bond is subsequently used for the rejoining of the DNA to the corresponding deoxyribose hydroxyl group on the other site. The entire process therefore proceeds without the need for external energy-rich cofactors such as ATP.

As stated previously, natural recombination sites are asymmetric, which allows the enzyme to tell apart the left and right ends of the site. When generating products, left ends are always joined to the right ends of the partner sites and vice versa. This causes the recombination sites to be reconstituted in the recombination products. Joining of left ends to left or right to right is avoided due to the asymmetric “overlap” sequence between the staggered points of top and bottom strand exchange. Left-left or right-right half-site recombinants would contain mismatched base pairs (Stark & Boocock, 1995).

Although the basic chemical reaction is the same for both tyrosine and serine recombinases there are marked differences.

Tyrosine recombinases, such as Cre or FLP, cleave one DNA strand at a time at points that are staggered by 6-8 bp, linking the 3’ end of DNA to the hydroxyl group of the tyrosine nucleophile (Van Duyne, 2002). Strand exchange than proceeds via an intermediate analogous to the Holliday junction (Holliday, 1964; Grainge & Jayaram, 1999) in which only one pair of strands has been exchanged.

Conversely, serine recombinases like Tn3 resolvase, which are the subject of this work, cut all four DNA strands simultaneously at points that are staggered by 2 bp (Stark *et al.*, 1992). During cleavage, a protein-DNA bond is formed via a transesterification reaction in which a phosphodiester bond is replaced by a phosphoserine bond between a 5’ phosphate at the cleavage site and the hydroxyl group of the conserved serine residue (S10) in resolvase (Reed & Grindley, 1981; Reed & Moser, 1984). It is still not entirely clear how the strand exchange occurs after the DNA has been cleaved. It has been shown however that the strands are exchanged while covalently linked to the protein, with a resulting change in the DNA structure equivalent to a net rotation of 180° (Stark *et al.*, 1989; Stark

& Boocock 1994). Two current models can account for this, namely the subunit rotation and the domain swapping model (Fig. 1.3b and 1.3a, respectively)(Sarkis *et al.*, 2001).

In both of these models, DNA duplexes are situated outside of the protein complex, and large movements of protein are needed to achieve the strand exchange, both of which are in stark contrast to the mechanism employed by the tyrosine recombinases. Models are summarised in Fig. 1.3. While the recent crystal structure (Li *et al.*, 2005) seems to favour the subunit rotation model, it is not absolutely incompatible with the domain swapping model. The single molecule experiments in order to establish which of the two strand-exchange models is the correct one are currently underway (G. Schlötel, personal communication).

Tn3 resolvase has been extensively studied both *in vivo* and *in vitro* and is a recombinase for which there is a wealth of structural information (Grindley, 2002; Rowland & Stark, 2005).

1.4 Tn3 resolvase resolution system

Tn3 resolvase or *Tnp*R is one out of three proteins that are encoded on the 4957 bp Tn*3* transposon (Fig 1.4a). Apart from Tn3 resolvase or tnpR, this transposon also carries a transposase (tnpA gene) and a β-lactamase (βla gene) that confers resistance to β-lactam antibiotics. Initially discovered as a repressor of transposase, resolvase also plays a role in facilitating Tn3 replicative transposition (Sherratt, 1989).

Tn3 replicates by intermolecular integrative transposition (Fig 1.4b). This process is catalysed by the Tn3 transposase and results in a fusion of the original host DNA with the target DNA molecule, creating a “cointegrate” along with the replication of the transposon. To separate the host and target molecules, Tn3 resolvase executes site-specific recombination between the old and new copies of the transposon at a specific site called *res*, which is present in each copy of the transposon.

The *res* site that Tn3 resolvase acts on is 114 bp long and contains three sub-sites, namely sites I, II and III (Fig 1.5). Each of these sites is of a different length (28, 34 and 25bp, respectively) and they are unevenly spaced, with 22 bp separating sites I and II but only 5 bp between sites II and III. The sites consist of 6 bp inverted repeat motifs flanking a central sequence of variable length. These motifs act as binding sites for resolvase, so that each site binds a resolvase dimer but with varying affinity and probably a slightly different

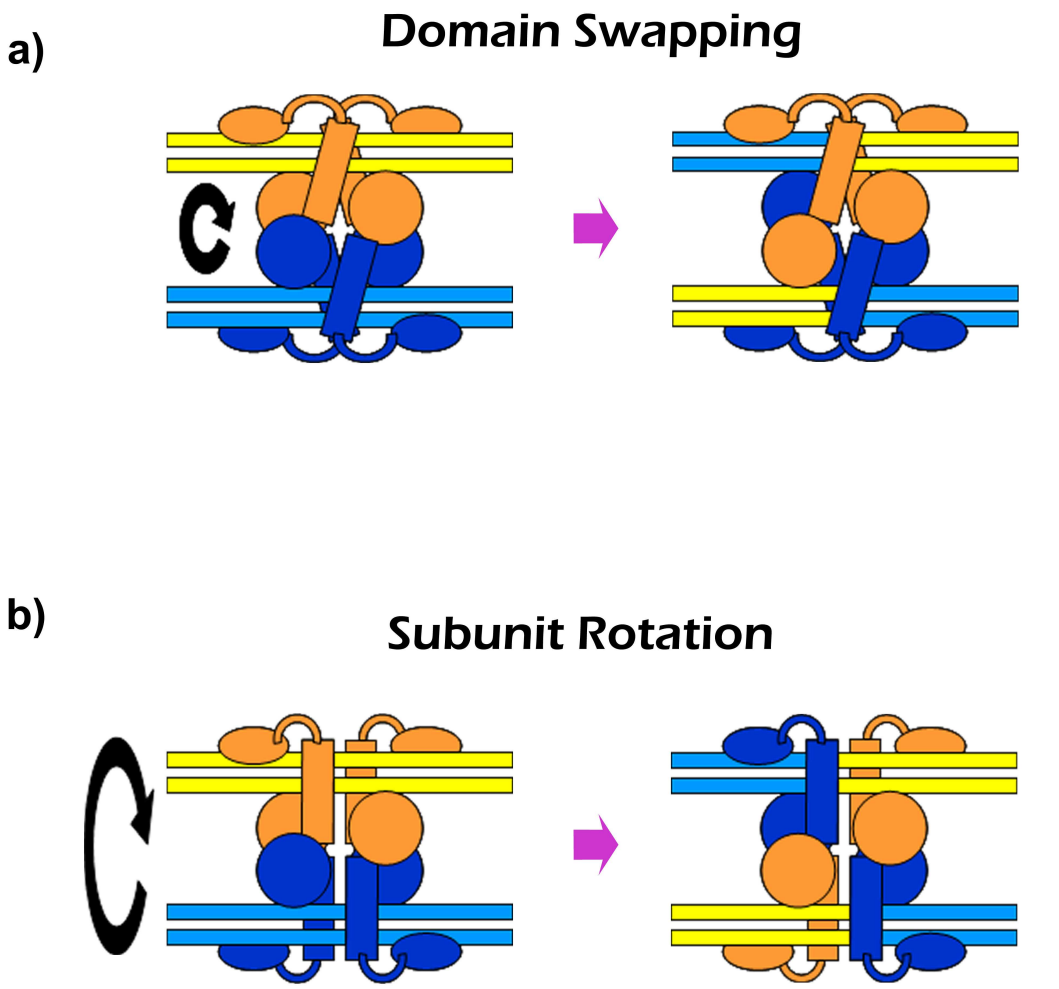


Figure 1.3- Serine recombinase strand exchange models a) The domain swapping model. In this model, following the strand cleavage, the DNA strand covalently linked to the resolvase catalytic domain is swapped by a swivelling motion of the catalytic domain around a hinge (~ residue 100 in Tn3 resolvase) bringing it into a recombinant configuration. For this domain to work C-terminal DNA-binding domain must release the DNA that is about to be exchanged. b) The subunit rotation model. In this model, the resolvase subunits that form the site I synapse catalyse double-strand cleavage of both crossover sites with each resolvase monomer covalently attached to the half-site at which it is bound. Following cleavage, a 180° right-handed rotation of one half of the complex along with the attached half-sites bring the cleaved DNA molecules into a recombinant configuration, which is followed by religation and the dissociation of the synapse releasing the recombinant catenated products.

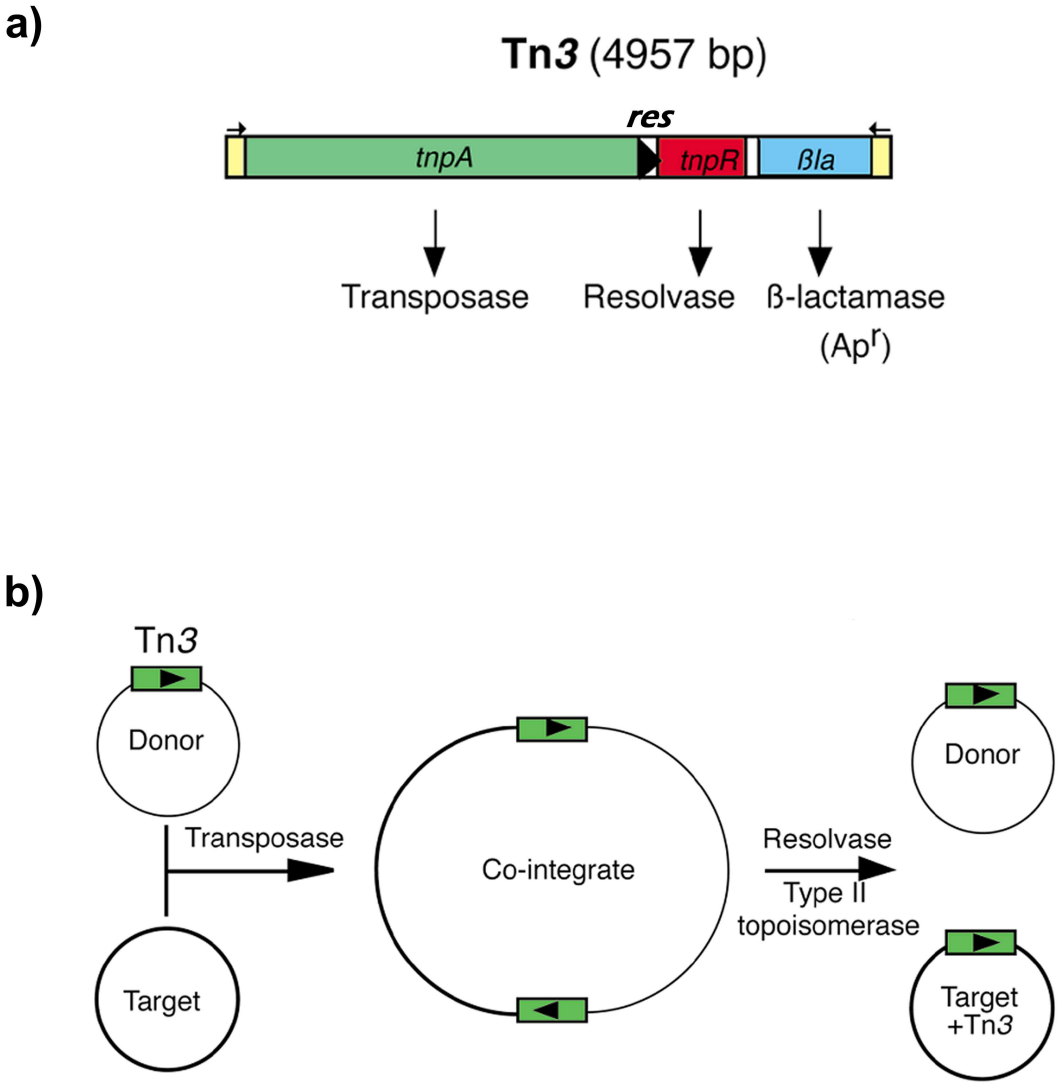


Figure 1.4- Schematic representation of Tn*3* transposition and cointegrate resolution. a) Structural organisation of Tn*3* transposon. The green, red and blue sections represent genes coding for transposase, resolvase and β-lactamase, respectively. The black right pointing triangle represents the *res* site, while yellow segments with small facing black arrows represent the transposon ends. b) Schematic illustration of the Tn*3* transposition pathway. The transposase catalyses the duplication and insertion of Tn*3* into a target plasmid, forming a cointegrate, which is then separated by resolvase into two plasmids, each containing a copy of Tn*3*. (Figure from Blake, 1993).

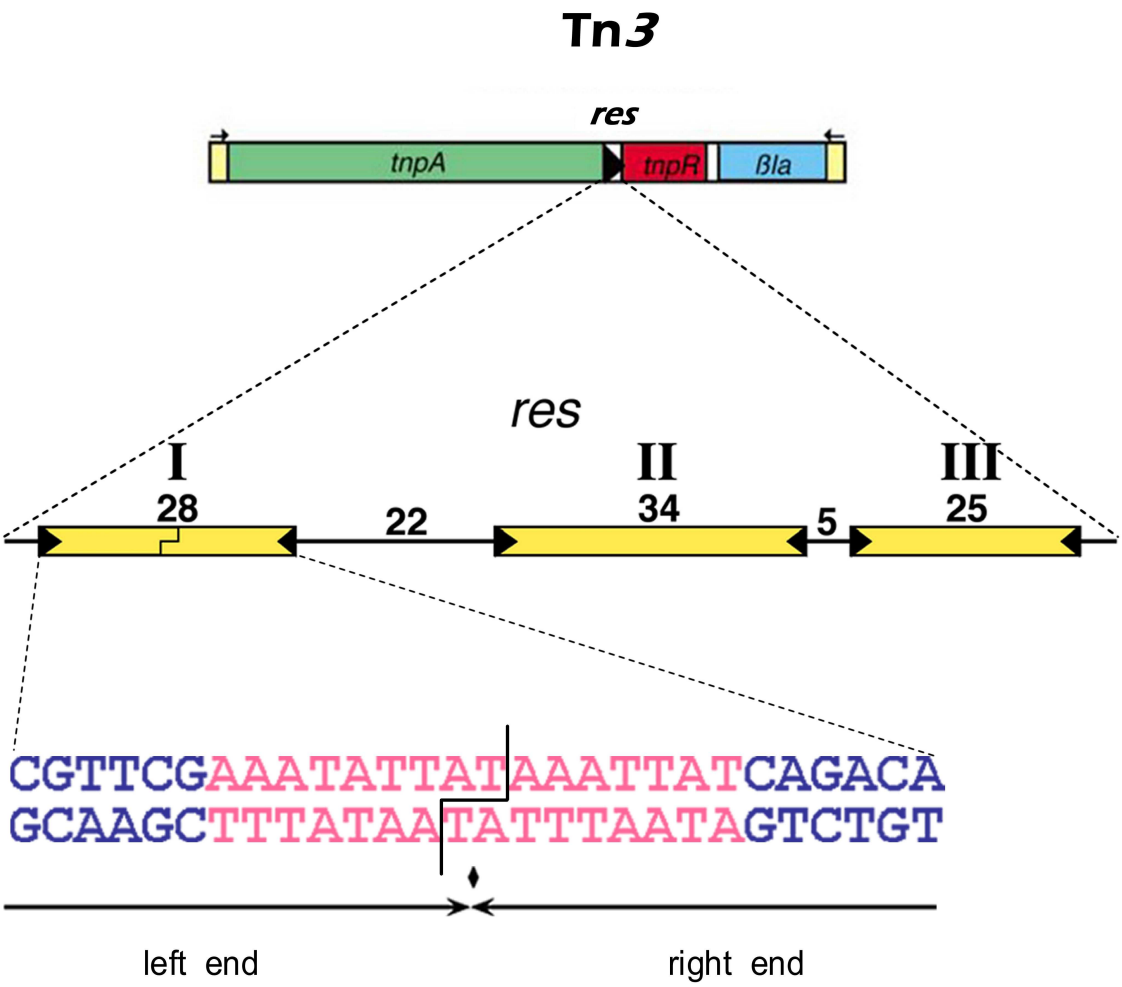


Figure 1.5- Schematic representation of Tn*3* *res* site. The figure indicates the relative location of the res site within the Tn*3* transposon (top). In the middle section res site is shown diagrammatically with sites I, II and III represented as yellow boxes flanked by black, facing triangles (representing inverted repeats) with the numbers above denoting the length of respective fragments in basepairs. The sequence of the recombination site (site I) is given directly below. The cleavage point is marked by the black lines. The "Left" and "Right" end refer to the conventional representation of site I within *res*, where the binding sites are in the order I-II-III left to right. The central 16 base pairs are given in pink while the outermost 6, recognised by the resolvase C-terminal helix-turn-helix domain, are in blue. (Figure adapted from Blake, 1993; and Burke *et al.,* 2004)

protein-DNA complex architecture (Abdel-Meguid 1984; Blake *et al.*, 1995). The presence of all three sites (I, II and III) is essential for recombination.

At recombination, two directly repeated *res* sites with resolvase dimers bound to each of the sites I, II and III , come together to form a large complex structure called the synaptosome (Fig. 1.6, step 1). Resolvase bound to sites II and III initiates the assembly of this complex. In the synaptosome, the exact architecture of which is still unclear, two *res* sites are intertwined in such a way as to juxtapose two copies of site I, allowing resolvase dimers bound to each site I to form a tetramer (Fig 1.6, step 2). Interaction between the resolvase dimers bound at the accessory sites (sites II and III) and resolvase at site I is thought to be required for the two dimers to synapse and form a tetramer. After the tetramer is formed it becomes activated and the top and bottom DNA strands are simultaneously cleaved in the middle of the site I with a 2bp overhang. The strand exchange ensues by as yet unknown mechanism with a resulting net rotation of 180°. The strand exchange is then followed by the religation (Fig 1.6, step3) (Stark *et al.*, 1992).

Recombination between the two directly repeated *res* sites in a cointegrate transposition intermediate separates, or resolves, the “cointegrate” into two molecules, each one now containing a copy of the Tn*3* transposon (Fig 1.6, step 4). After resolution, these two molecules remain linked as a simple two-noded catenane which can be easily separated *in vivo* by a type II topoisomerase (Grindley, 2002). The entire resolvase recombination reaction can be reproduced *in vitro,* requiring only resolvase, a substrate DNA and multivalent cations (Reed, 1981).

A number of “deregulated” or “hyperactive” Tn3 resolvase mutants that have lost the requirement for the accessory sites have been isolated. These mutants are capable of catalysing recombination between two copies of site I only, which basically reduces the recombination site size from 114bp to only 28bp (Arnold *et al.*, 1999; Burke *et al.*, 2004; Olorunniji *et al.*, 2008). Further more these mutants have no supercoiling or conectivity requirements (Arnold *et al.*, 1999) and have been shown to work in mammalian cells (Schwikardi & Droge, 2000). Hyperactive resolvase mutants have so far proven invaluable in creating resolvases with altered sequence specificity (Akopian *et al.*, 2003) but also in structural work (Li *et al.*, 2005), both of which will be discussed further later.

1.5 Structure of resolvase

The resolvase proteins from different Tn3-like transposons form a large family, which can be subdivided into a series of complementation groups. Within each group, resolvase

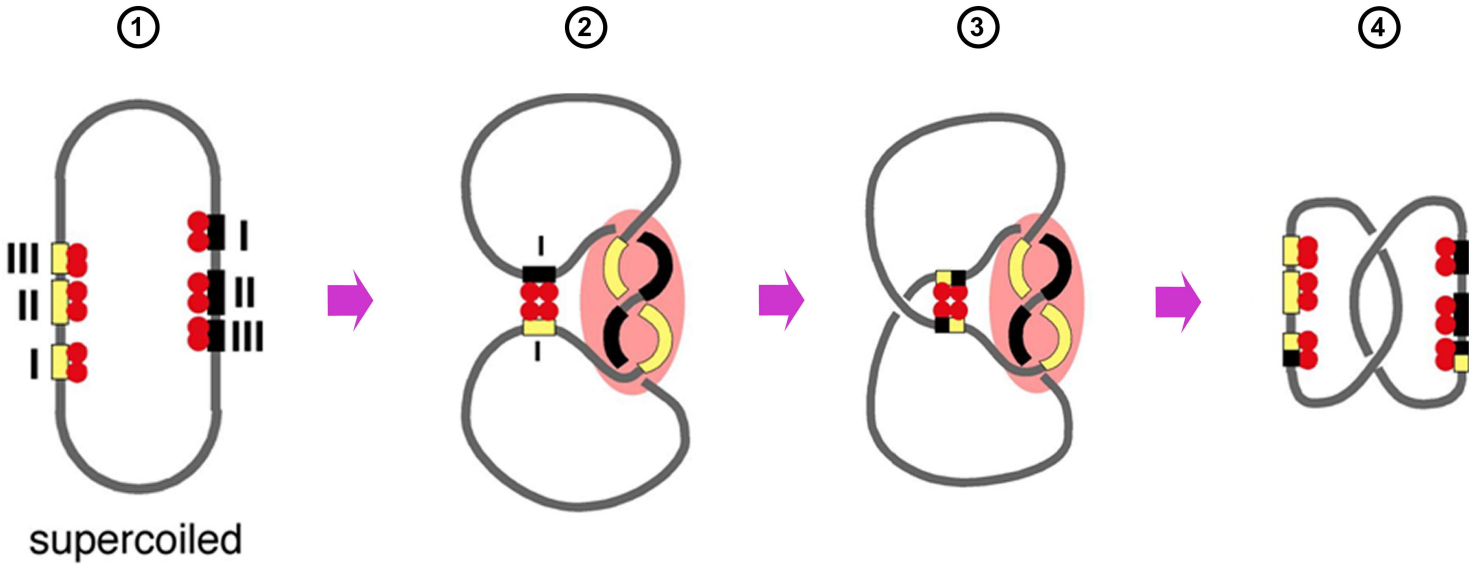


Figure 1.6- Schematic representation of synapsis and recombination by Tn3 resolvase. The cartoon shows the hypothetical pathway for the assembly of the res synapse and the subsequent strand exchange. Red spheres represent resolvase subunits bound to the recombining res sites (represented by yellow and black boxes) on the supercoiled substrate. 1) Resolvase bound to the substrate. Following the binding resolvase-DNA complexes form a synapse, first involving only accessory sites (not shown) and then also site I to create: 2) Pre-reaction res synapse. After the res synapse is formed the resolvase proceeds with recombination catalysis and strand exchange within the res synapse. (the pink shading indicates that the exact nature of the complex is unknown) 3)Post-reaction res synapse. Following the recombination reaction the res synapse dissociates to form catenated recombinant products as depicted under 4).

proteins are highly homologous to each other, with typically over 80% amino acid identity. A resolvase from one group can carry out recombination between *res* sites from any member of the same group (Ackroyd *et al.,* 1990; Avila *et al.,* 1990).

Tn3 resolvase and γδ resolvase are members of the same complementation group. The conclusions from experiments carried out using one of these proteins generally apply to both. γδ resolvase is a recombinase encoded by the transposon *gd* (Tn*1000*) which is a natural component of the *E. coli* F plasmid.Tn3 and γδ resolvases are identical at 147 out of 185 (183 in γδ) amino acid residues, and their structures are believed to be very similar (Grindley, 2002).

Even before the crystal structures were available, there were strong indications as to the modular nature of γδ resolvase. Chymotrypsin digestion of the 20.5 kDa γδ resolvase monomer revealed that this protein can be separated into two distinct domains (15.5 and 5 kDa – residues 1-140 and 141-183, respectively), the smaller of which was shown to bind the DNA specifically, albeit with a lower affinity than the intact protein (Abdel-Meguid *et al.,* 1984).

The small domain was not seen in the early *γδ* resolvase structures obtained from crystals grown without DNA, as only the first 120 amino acid residues of resolvase were ordered (Rice & Steitz, 1994). It was inferred that this domain is very flexible and probably only partially folded, assuming the proper fold only upon binding to DNA. Bearing in mind that the three sub-sites in *res* differ substantially in length and in sequence, this did not come as a surprise (Blake *et al.,* 1995). The first 120 amino acid residues of γδ resolvase, folded into a globular domain containing the catalytic serine (Ser-10) (Rice & Steitz, 1994).

A structure of a γδ resolvase dimer bound to site I (Fig. 1.7) revealed that γδ resolvase contains two distinct functional domains (N-terminal catalytic and C-terminal DNA binding domain), linked by a short but conformationally extended linker region (Yang & Steitz, 1995). The N-terminal catalytic domain spans the residues 1-137. It consists of a five-stranded mixed β sheet that is surrounded by three α helices on one side (αA, αB and αC) and one helix on the other (αD).

The interfaces that are implicated in intersubunit contacts are found on this domain such as the dimer interface and the 2-3’ interface. The dimer interface is the surface of contact between two resolvase subunits bound to the same site (Fig. 1.7), while the synaptic

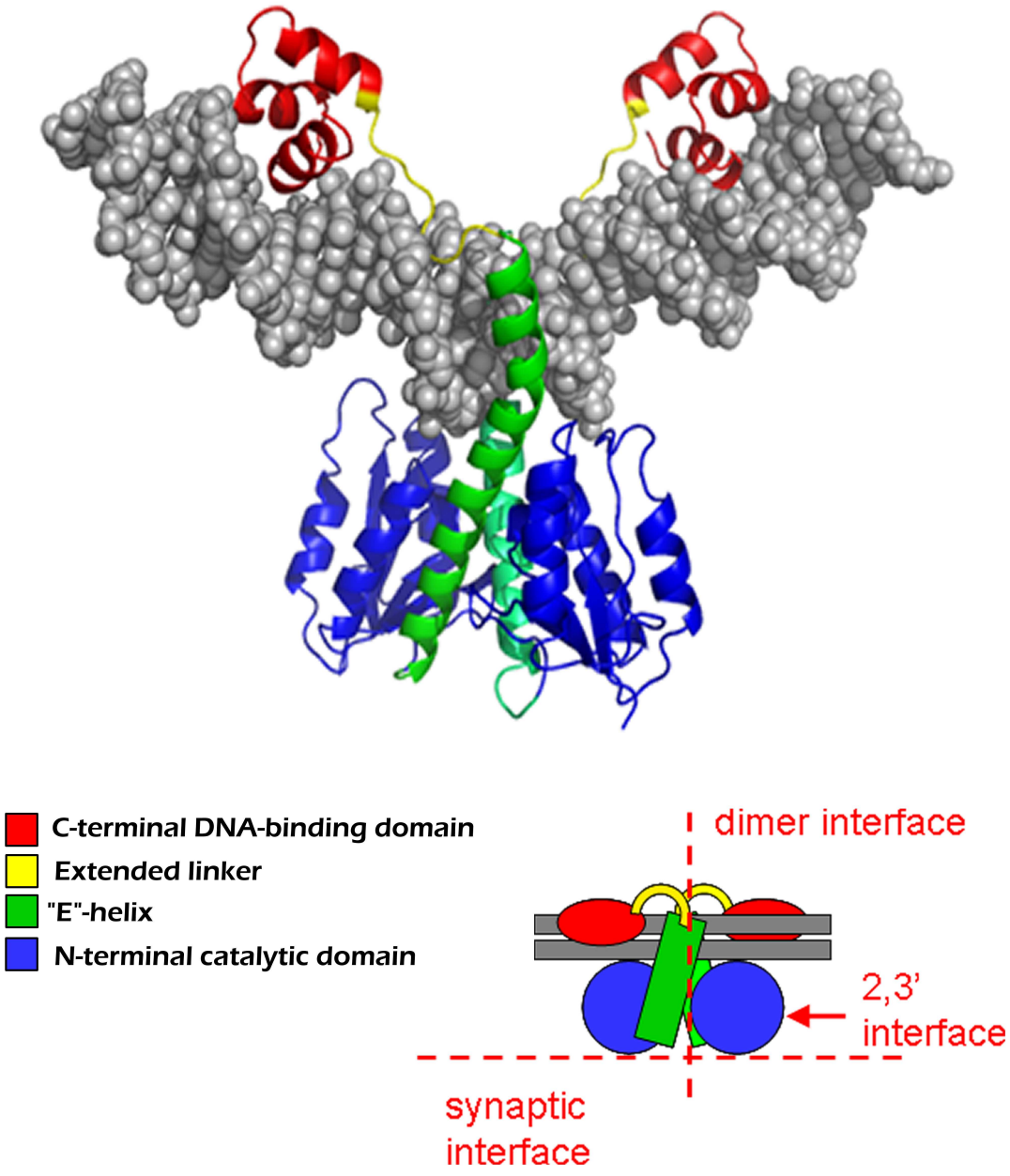


Figure 1.7- Three-dimensional structure of γδ resolvase dimer bound to DNA. The different functional regions of the protein are coloured as follows: N-terminal catalytic domain is blue, E-helix is green, the extended portion of the arm region is yellow and the C-terminal helix-turn-helix DNA-binding domain is red. DNA is coloured grey. Using the cartoon (bottom right) important interfaces discussed in the text are marked. The image was generated using PDB coordinates of 1GDT (Yang & Steitz, 1995).

interface represents the contact area between the two dimers bound to the sites forming the synapse. The 2-3' interface is situated directly opposite to the dimer interface and it is probably where resolvase subunits bound to other sites of *res* (likely site III) contact the resolvase on site I. Residues 1-100 create the globular core of this domain whereas the remaining residues (101-137) are folded into a large "E" helix. The "E" helix is a very important structural motif. The interaction between hydrophobic residues of crossed "E" helices on two adjacent monomers bound to site I is crucial for the dimer interface. This, along with a hydrogen bond between a pair of Glu-118 residues, is what locks the dimer together.

The "E" helices grip the DNA over the minor groove in a simple right-handed 45° crossover, analogous to a pair of chopsticks, helping to position the C-terminal domains and allow them to bind the DNA.

It is interesting to note that only the first half of the "E" helix (up to residue 120) is ordered when protein is not bound to DNA (Rice & Steitz, 1994). Formation of the complete "E" helix is believed to be induced by DNA binding, analogous to the leucine zipper bZIP motif (Weiss *et al.*, 1990).

Due to this disordered character it was decided to include the second portion of the "E" helix in what was called an "arm" region (Yang & Steitz,1995). Basically an arm region is a part of the protein that connects the catalytic core domain with the DNA-binding domain and contacts the site I DNA in the minor groove. The "arm" region comprises the second half of the E helix (residues 121-137), and what was called the linker in the text so far (residues 138-148). A sharp 90° turn induced by a highly conserved residue G137 delineates the end of the "E" helix and the start of the extended linker.

The arm region is responsible for positioning the DNA-binding domain, which it achieves through extensive minor groove interactions. After the sharp bend at residue 137, resolvase continues as a short extended strand (residues 138-140) and then it twists again at residue G141, entering deep into the minor groove. The next residue, R142 interacts extensively with the DNA backbone, but also makes sequence-specific interactions with a specific T·A base pair.

The extended arm region continues until the start of the C-terminal DNA-binding domain at residue 148. The DNA-binding domain is comprised of three helices, namely αF, αG and αH, arranged in a bundle around a hydrophobic core. The residues in this domain,

apart from three in the hydrophobic core, are very poorly conserved across the resolvase/invertase family. This is not surprising as it is this domain that makes sequence-specific interactions with the DNA, and the sequences that different members of the family recognise are highly diverse.

Helices “G” and “H” of the γδ resolvase DNA binding domain form a typical helix-turn-helix motif (HtH) which constitutes the main source of sequence-specific DNA recognition. No less than eight hydrogen bonds and salt bridges are formed between the residues R148, S162, A171, T174, Y176 and K177 and the phosphate groups of the DNA backbone. These interactions firmly anchor the C-terminal domain inside the major groove allowing the side chains of the residues R172, S173, Y176 and K177 to establish further, base-specific interactions.

When the structure of the site I- resolvase dimer is compared to the structure of the tetramer in a DNA-cleaved synaptic intermediate (Li *et al.,* 2005) dramatic changes in the tertiary and quaternary structure of resolvase are observed. These structural changes are achieved by altering the relative orientations of the catalytic domain, the “E” helix and the DNA binding domain, by rotating these domains at linkers that connect them. Most striking of these changes is the bending and 40° rotation of the “E” helix with respect to the “D” helix of the globular core of the N-terminal domain, which brings the catalytic S10 in the appropriate position to form a covalent link with the DNA. This rotation, which is pivoted by a hinge at residues 101 and 102, also opens up the “E” helix chopsticks to an angle of 100°, which in turn creates a uniquely flat interface that is proposed to support subunit rotation and therefore strand exchange (Fig. 1.8).

These structural changes are tightly regulated in the native resolvase protein, requiring the existence of the full synaptosome and probably appropriate interactions at the 2-3’ interface. This is why crystal structures of resolvase tetramer were achieved only when “hyperactive” mutants were used (Fig. 1.9).

Some of the mutations that produce hyperactivity are proposed (Burke *et al.,* 2004) to stabilise an early synaptic interface such as D102Y. The other mutations might make the changes in the orientation of domains easier. This second group is situated either at the hinge connecting “E” helix (G101S, M103I, Q105L) or in the arm region that helps position the C-terminal domain (E124Q) (Li *et al.,* 2005; Arnold *et al.,* 1999; Burke *et al.,* 2004). Mutations at the 2-3’interface (e.g. R2A, E56K) although not activating themselves

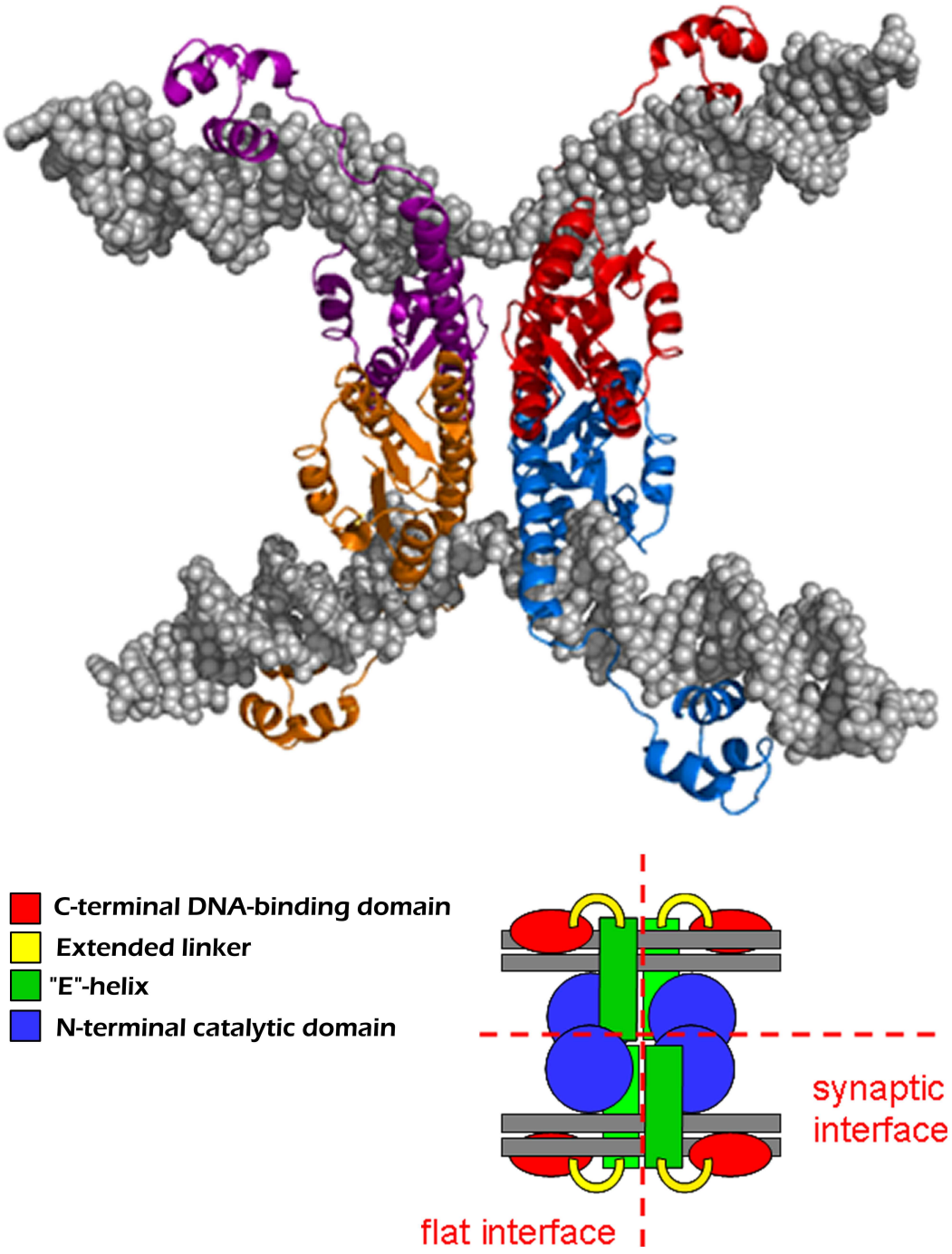


Figure 1.8- Three-dimensional structure of γδ resolvase post-cleavage synaptic complex. The different resolvase monomers are coloured purple, red, orange and light blue. DNA is coloured grey. The functional regions of the protein are coloured in the cartoon (bottom right) as in Fig. 1.7: N-terminal catalytic domain is blue, E-helix is green, the extended portion of the arm region is yellow and the C-terminal helix-turn-helix DNA-binding domain is red. Using the cartoon important interfaces discussed in the text are marked. The image was generated using PDB coordinates of 1ZR4 (Li et al., 2005).

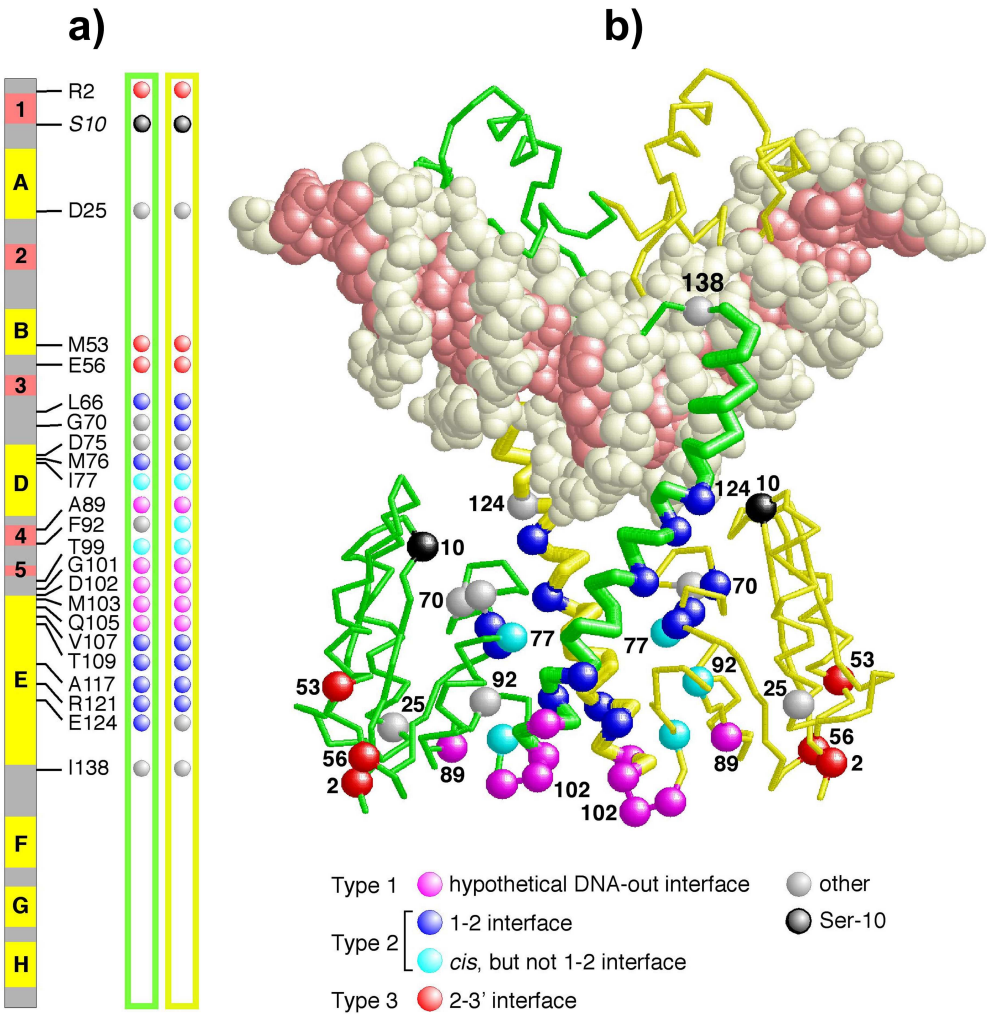


Figure 1.9- Positions of the activating mutations found in Tn3 resolvase. **a)** The positions of the activated residues are marked on Tn3 resolvase amino acid sequence which is presented as a bar (the secondary structure elements are marked as in Yang & Steitz, 1995. The residue type is indicated by the coloured spheres as in the crystal structure image (panel b) right) with a green box indicating subunit A, and the yellow box subunit B. **b)** The γδ resolvase dimer (subunits A and B, green and yellow, respectively) bound to the DNA (pink and cream spacefill representation) based on 1GDT coordinates (Yang & Steitz, 1995). The α-carbons of the residues where activating mutations are found are shown as spheres, colured according to type indicated and as described in Burke *et al.,* 2004. The "hypothetical" DNA-out interface comprises of residues predicted to mediate dimer-dimer interactions in the site I synapse (Sarkis *et al.,* 2001). Please note that this interface is no longer hypothetical as it has been confirmed in the later work (Nollmann, *et al.,* 2004; Li *et al.,* 2005). The cis interface includes contacts made by the globular part of the catalytic domain with the E-helix of its own subunit. The figure is adapted from Burke *et al.,* 2004.

can enhance the hyperactive phenotype in the already activated resolvase (Olorunniji *et al.*, 2008)

One analogy for the changes in orientation between different domains of the resolvase protein and the mutations that facilitate these changes would be to regard resolvase as a door hinge. If the door hinge is stiff, it can be oiled to loosen it up or, more radically, attacked with a hammer. Thus if reaction conditions are carefully adjusted ('oiling'), the resolvase can catalyse reactions that it normally fails to do (Bednarz *et al.*, 1990). If mutations ('hammer blows') are introduced at specific positions that disrupt the dimer interface (hinge closed) and encourage the synaptic interface (hinge opening), resolvase can work well on site I only. Hammer attacks are effective in making the hinge turn more easily, but they also make the hinge more rickety and unstable, and could potentially break it. A rickety hinge is less likely to get stuck in a closed position, but it might also sometimes open when its not supposed to (leading to reactions at pseudo-sites), or it might not be close properly (leading to inversions and intermolecular recombination products). However if it is well oiled (correct reaction conditions used) it can work perfectly well (Olorunniji, unpublished).

1.6 Uses of site-specific recombinases in genetic manipulation

In this genomic age, as our understanding of genome architecture and functions grows, possibilities for altering genomic DNA sequence to our benefit are becoming ever more apparent. The ability to implement precise, controlled genome wide changes in essentially any organism depends on the development of new 'genetic engineering' protein tools, capable of targeted manipulation of DNA, and efficient methods by which these tools can be delivered inside the cells *in vivo*.

Being able to target the activity of these enzymes specifically to chosen sequences within an organism's natural genome could herald a new era in medicine, commercial biotechnology and experimental genetics. Applications such as human gene therapy, new treatments for retroviral diseases, animal transgenesis and numerous others would become more efficient, safer, and easier to accomplish.

In order to create tools capable of controlled genome alteration a number of diverse avenues have been explored, including the use of: homing group II introns (Guo *et al.*, 2000), vectors designed to exploit cells own DNA repair machinery to promote integration

(Ng & Baker, 1999), viral vectors with site-specific integration ability (Monahan & Samulski, 2000), and the site-specific recombinases.

Site-specific DNA recombinases can promote excision of specific DNA segments or insertion of new DNA segments in specific places. With the advent of mass sequencing, many prokaryotic genomes have been sequenced and catalogued. This enormous amount of data represents a treasure trove containing hundreds of site-specific recombinases. This treasure is yet to be used to even close to its capacity. Many of these site-specific recombinases catalyse recombination at unique sequences, some of which might be very attractive with respect to specific applications.

Currently, by far the most commonly used site-specific recombination system in molecular biology today is the Cre/*loxP* system, closely followed by Cre's distant relative FLP (Nagy 2000, Bode *et al.*, 2000). The recombinase Cre works on a relatively short (34bp) site called *loxP*, consisting of two inverted repeats flanking a central sequence. The *loxP* site and the analogous site for FLP (FRT) sequences are sufficiently long as to be quite rare or absent in natural genomes. Therefore, although many of these enzymes work perfectly well both in prokaryotic and eukaryotic cells, most applications of the Cre or FLP systems require that their specific recognition sites be first introduced into the genome that is to be modified. This is achieved by the classic homologous recombination methods, which dramatically reduces the overall efficiency of the procedure. This same problem applies to the potential application of any site-specific recombinase with a recombination site of similar length. Furthermore, the activity of natural site-specific recombination systems is usually tightly regulated, and the type of rearrangement they bring about (excision, inversion, or integration) is usually specified. These constraints can cause problems for those who wish to subvert the recombinases to become tools for genetic manipulation.

The alternative approach lies in trying to make recombinases function on sequences native to the genome. Although detailed structural information is available for tyrosine and serine recombinases (Fig.1.10) (Guo *et al.*, 1999; Chen *et al.*, 2000; Yang & Steitz 1995; Li *et al.*, 2005), the basis of sequence recognition by Cre and FLP is still unclear, as no clear distinction can be made between the residues that are involved in catalysis and DNA binding. In proteins such as resolvase on the other hand this distinction is clearer with the "DNA-binding" and the "catalytic" function being largely confined to the distinct domains. Taking into account the innate differences between serine and tyrosine recombinase systems, two different approaches for altering the specificity of these proteins have been adopted, namely the directed evolution and the modular/re-engineering of recombinases.

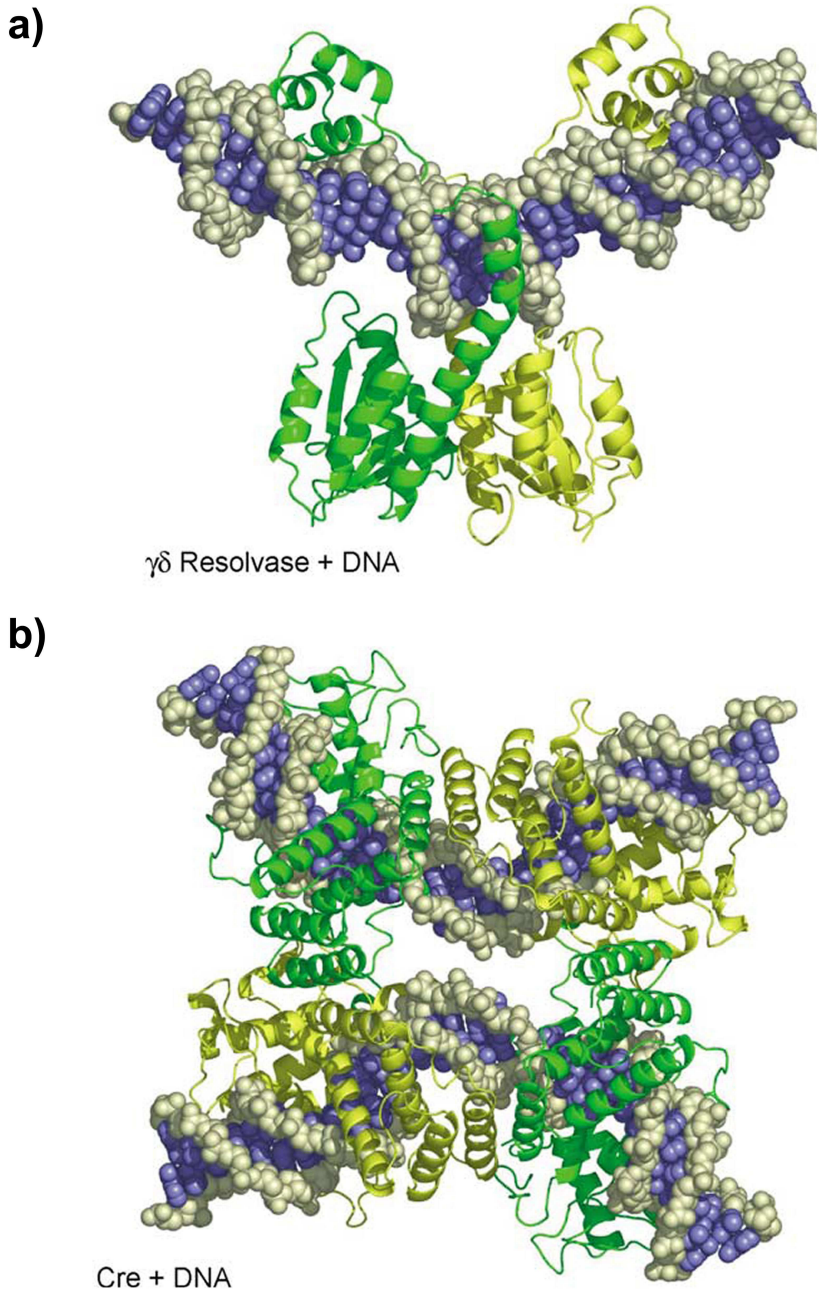


Figure 1.10- Crystal structures of site-specific recombinases bound to DNA. **a)** A dimer of the serine recombinase γδ resolvase, bound to site I where strand exchange occurs (Yang & Steitz, 1995). **b)** A tetramer of the tyrosine recombinase Cre in a synapse with two loxP sites (Guo *et al.,* 1999). The DNA (cream and blue) is in spacefill representation, and the protein (green and yellow) is in a ribbon representation. The two loxP sites in b) are in approximately anti-parallel alignment. The figure is from Akopian & Stark, 2005.

1.6.1 Directed evolution of recombinase specificity

With Cre recombinase the approach was to select a site that is similar to the natural *loxP* site i.e. a pseoudo-site, and then attempt to produce a Cre variant that can catalyse recombination on this site. This typically involved several rounds of a mutagenesis-selection procedure where a recombinase mutant library is tested against an *in vivo* test substrate containing these pseudo-sites and an observable reporter gene (Bucholz & Stewart, 2001). This kind of mutagenesis usually leads to relaxed specificity instead of a shift in specificity (going back to the hinge-hammer analogy) which could be detrimental for potential applications. To avoid this outcome these techniques have been refined as to also select for loss of activity on the natural site (Voziyanov *et al.,* 2002) and to speed up the selection process, which was supplemented by DNA shuffling (Stemmer 1994; Voziyanov *et al.,* 2002; Voziyanov *et al.,* 2003).

A similar approach has been successfully employed also with a very distant serine recombinase relative of resolvase ϕC31 integrase (Groth & Calos, 2003).

1.6.2 Re-engineering modular serine recombinases

Due to their modular structure with their "DNA-binding" and "catalytic" function being largely confined to the distinct domains, Tn3 resolvase and related resolvases/invertases lend themselves to more rational approaches to changing the sequence specificity. Even before structural information became available, experiments in which parts or the entire C-terminal domain have been exchanged between different related wild-type resolvases (Tn3 and Tn21) that recognise very different sites were performed (Ackroyd *et al*., 1990; Avila *et al.,* 1990). These DNA-binding domain swapping experiments were analogous to the first attempts at engineering protein-DNA interactions involving 434 repressor (Wharton and Ptashne, 1985). Similarly, as in experiments with 434 repressor, domain swapping with resolvases did create proteins with different binding specificity. However, although wild-type Tn3 and Tn21 resolvases with swapped DNA-binding domains were able to bind (to an extent) to the non cognate *res* site they were found to be unable to catalyse recombination on the new site.

More recently in our lab, more radical domain-swapping experiments have created a series of chimaeric proteins with the generic name Z-resolvase. Z-resolvase architecture and properties are discussed in the following section.

1.7 Z-resolvase recombination system

Z-resolvase is a chimaeric protein construct assembled from the catalytic domain of a resolvase and a completely unrelated DNA-binding domain from a mouse transcription factor Zif268 joined together by a synthetic linker (Akopian *et al.*, 2003; Akopian, 2003).

The first Z-resolvase was created by A. Akopian using the catalytic domain of NM resolvase, which is an activated mutant of Tn3 resolvase (R2A, E56K, G101S, D102Y, M103I, Q105L) (Akopian, 2003; Olorunniji *et al.*, 2008). NM resolvase is capable of acting on Tn3 *res* site I, without the need for accessory sites, negative supercoiling or specific relative orientation of the recombination sites, making it an ideal candidate for the catalytic domain donor. The catalytic domain cut-off point in the earlier Z-resolvase design was at the residue R148 which was later changed to R144.

The DNA-binding domain used in these and all subsequent Z-resolvases designed in our group was the tridactyl zinc finger (three fingers) DNA-binding domain of a mouse transcription factor Zif268. This domain is relatively small being comprised of ~90 amino acid residues and it functions as a monomer, making it suitable to act as a replacement for the native resolvase HtH domain. Other criteria for choosing this domain were that its structure in the complex with the DNA is known (Pavletich & Pabo, 1991) and that it is amenable to re-engineering to create variants with altered sequence specificity (Griesman & Pabo, 1997).

The Zif268 DNA-binding domain is folded into three zinc finger motifs consisting of two anti-parallel b-sheets and an a-helix that are held together by coordinating Zn^{2+} ion (Elrod-Erickson *et al.*, 1996). Each of the zinc fingers interacts with 3-4bp of DNA in the major groove by a direct read-out via the recognition helix. Only residues at six positions (–1, 1, 2, 3, 4, 6) of the recognition helix interact with DNA which makes this DNA-binding domain so attractive as a template for designing DNA-binding domains with novel sequence specificities. By mutagenising these six residues it should be possible to make variants of this DNA-binding domain that can recognise essentially any sequence (Dreier *et al.*, 2000; Dreier *et al.*, 2001; Dreier *et al.*, 2005). This is just what is currently being done in numerous labs throughout the world and variants of Zif268 that successfully and selectively bind various sequences have been selected using methods such as phage display (Rebar *et al.*, 1996; Beerli & Barbas 2002; Reynolds *et al.*, 2003)

Several different linker sequences were used to connect the NM catalytic domain to the Zif268 DNA-binding domain. In the earliest Z-resolvases made by Akopian the linkers

were loosely based on the sequence of a linker in a domain swapping enzyme T4 endonuclease VII (Akopian 2003). This sequence was chosen for its flexibility and it was followed by a TS sequence which created a SpeI site at the DNA level. The SpeI site was used in cloning to connect the linker with the Zif268 DNA-binding domain. This linker design was then superseded by fully synthetic flexible linkers of varying lengths and sequences (S and G rich) (Akopian, 2003; Akopian *et al.,* 2003).

The Z-resolvases created by A. Akopian showed no activity on substrates containing Tn3 site I but acted on substrates containing specially designed sites called “Z-sites” (Akopian *et al.,* 2003). In a Z-site, the six outermost base pairs of site I each end are replaced by the 9 bp sequence that Zif268 naturally binds to (Fig. 5.1). However, this 9 bp sequence has to be spaced out from the centre of site I by a further 3bp at each end for Z-resolvase to work well, making the total length of the Z-site 40 bp as opposed to 28 bp of native Tn3 site I. These extra 3 bp are required irrespective of the length or the sequence of the linker that was used. Longer linkers were not shown to be able to catalyse recombination on longer Z-sites. However, there is some flexibility in the Z-site length requirement, as Z-sites with an extra two or four base pairs spacer sequence (total length 38 bp or 42 bp) do show some activity. Z-resolvase failed to catalyse recombination with substrates containing Z-sites that are shorter than 38 bp or longer than 42 bp of the total length.

The next generation of Z-resolvases was created by A. McLean. In these proteins the catalytic domain of NM resolvase was again employed (to the residue 148) but the linker sequence was dispensed with and the Zif268 DNA-binding domain was joined onto it just by the sequence TS (containing SpeI site; see above). These proteins showed a comparable or possibly better level of activity to the Z-resolvases with longer linker sequences (A. McLean, unpublished) and were subjected to mutagenesis to try and improve them further.

1.8 Aims of the project

The main aim of this project was to establish the extent and the basis of sequence specificity of the resolvase protein catalytic domain. The attempts to make the resolvase catalytic domain recombine sites with non-cognate centres have so far proven unsuccessful. The crystal structures (Yang & Steitz, 1995; Li *et al.,* 2005; Mouw *et al.,* 2008) show fairly few interactions between the central sequence of site I and the catalytic domain, and it is still not clear what sequence features are important for efficient catalysis/recombination. However even before crystal structures were available (Ackroyd *et al.,* 1990) it was suggested that there are additional sources of sequence specificity in Tn3 resolvase beyond the specificity imposed by the DNA-binding domain.

In Chapter 3 a systematic analysis of how sequence in the central 16 bp affects recombination by activated Tn3 resolvase mutant NM was carried out. All 16 bp were systematically randomised in overlapping 4 bp blocks and recombination deficient and recombination proficient site I variants were selected using two different independent selection strategies. Comparing these mutually exclusive sets of site I variants to each other, and to the naturally occurring site I sequences associated with proteins from the Tn3 resolvase family, and integrating this data with the available structural information should identify the basepairs and/or sequence motifs that influence the sequence selectivity by the resolvase catalytic domain.

In Chapter 4, the sequence selectivity of NM resolvase catalytic domain was tested in the Z-resolvase context employing a similar but less exhaustive selection strategy using a purified Z-resolvase Z-R(NM). Comparing the results of these selections with the findings of Chapter 3 should provide a powerful insight into what sequences could be used as the potential centres of future Z-sites. In the same chapter Z-resolvases with sequence selectivity that is different to that of Z-R(NM) were constructed using catalytic domains of activated mutants of Sin and Tn21 resolvases and their *in vivo* and *in vitro* properties were tested. Creating novel Z-resolvases that are not based on Tn3 resolvase should highlight the universality of the Z-resolvase approach and its potential for the future applications.

A number of unresolved issues concerning the Z-resolvase design remained after work of Akopian and MacLean and arose since publication of related work (Gordley *et al.*, 2007; Gordley *et al.*, 2009) such as the optimum length of Z-sites, what is the effect of the Zif268 DNA-binding domain on catalytic activity i.e. is it activating or inhibiting, is symmetry a prerequisite in the design of Z-sites or can a Z-resolvase catalyse recombination on sites with an odd number of bases between Zif 268 binding sites i.e one half-site longer than the other, what is the relative influence of the Z-resolvase linker length, and can Z-resolvase be complemented by resolvase and catalyse recombination on appropriately designed hybrid sites. All of these issues are explored in Chapter 5.

Finally in Chapter 6, the sequence selectivity of catalytic domains of Sin and Tn21 resolvases is compared using a combination of a selection strategy similar to the one described in Chapters 3 and 4 and the Sin-Tn21 resolvase hybrid experiments. In the second part of this chapter, an experiment to change the sequence selectivity of Tn3 resolvase catalytic domain into that of Sin resolvase, both in the resolvase and Z-resolvase context, by mutating the specific residues, implicated in catalytic domain sequence selectivity in the previous chapters, is performed. Being able to change the sequence

selectivity of one resolvase catalytic domain into another would be the ultimate proof of understanding the basis of sequence selectivity in the centre of site I.

2 Chapter 2: Materials and Methods

2.1 Bacterial Strains

All bacterial strains used in his project were derivatives of *Escherichia coli* K-12. Strain names, genotypes and sources are provided in the Table 2.1 below.

STRAIN NAME	GENOTYPE	SOURCE
DS941	AB1157, but *rec*F143, *sup*E44, *lac*ZΔM15, *lac*I^{q}	Summers and Sherratt, 1988
DH5α	F^{-}/*end*A1, *hsd*R17 (r_k^-, m_k^+), *sup*E44, *thi*$^{-}$1, Φ80d*lac*ZΔM15, Δ(*lac*ZYA -*arg*F)U169, *deo*R, *rec*A1, *pho*A, *gyr*A96, *rel*A1, λ^{-}	Invitrogen
DH5	F^{-}/*end*A1, *hsd*R17 (r_k^-, m_k^+), *sup*E44, *thi*$^{-}$1, Δ(*lac*ZYA -*arg*F)U169, *deo*R, *rec*A1, *pho*A, *gyr*A96, *rel*A1, λ^{-}	Invitrogen
JC8679	*rps*L31, *tsx*-33, *sup*E44, *rec*B21, *rec*C22, *sbc*A23, *his*-328	Gillen *et al.*, 1981
BL21 (DE3) pLysS	*hsd*, *gal*, (λ*cI ts* 857, *ind*1, *Sam*7, *ini*5, lac_{UV5}-T7 gene-1), T7 lysozyme-expressing plasmid pLysS	Studier *et al.*, 1990

Table 2.1-Bacterial Strains

2.2 Chemicals

Sources of most general chemicals used are listed in Table 2.2. Distilled water was used for making all solutions unless otherwise stated.

CHEMICALS	SOURCE
General chemicals, biochemicals, organic solvents	Sigma/Aldrich, BDH, May and Baker
Media	Difco, Oxoid
Agarose, acrylamide	FMC, Biorad, Flowgen
Restriction enzyme buffers	NEB, Promega, Roche
Ligase buffer	Promega

Table 2.2- General Chemicals

2.3 Bacterial growth media

E. coli liquid cultures were routinely grown in L-Broth (10 g bacto-tryptone, 5 g bacto-yeast extract, 5 g NaCl, made up to 1 litre with deionised water, and adjusted to pH 7.5 with NaOH). *E. coli* on solid medium were cultured on L-Agar (L-Broth with 15g/l agar).

Growth media were sterilised at 120 °C for 15 minutes. Other specialist growth media that were used are: 2×YT broth and MacConkey Galactose Agar. 2×YT broth was used in the recovery step in the DNA transformation protocol (see Section 2.9) while MacConkey Galactose Agar was used for *in vivo* recombination and binding assays (Sections 2.23 and 2.24 respectively). The composition of 2×YT broth is: 16 g bacto-tryptone, 10 g bacto-yeast extract, 5 g NaCl, made up to 1 litre with deionised water and adjusted to pH 7.0 with NaOH. MacConkey Galactose Agar was prepared from a premix powder supplied by Difco (17 g bacto-peptone, 3 g bactoprotease peptone, 1.5 g bacto bile salts No.3, 5 g NaCl, 13.5 g bacto agar, 0.03 g neutral red, 0.001 g bacto crystal violet) by boiling 4 g of premix in 100 ml of deionised water and addition of 5 ml of 20% aqueous solution of D-galactose, giving a final concentration of 1% w/v in the agar.

2.4 Antibiotics

The antibiotics used in liquid and solid selective growth media, and the concentrations in the media and in stock solutions are given in Table 2.3.

ANTIBIOTIC	STOCK CONCENTRATION (1000 X)	FINAL CONCENTRATION IN MEDIUM
Ampicillin (Ap)	100 mg/ml in H_2O	100 μg/ml
Kanamycin (Km)	50 mg/ml in H_2O	50 μg/ml
Chloramphenicol (Cm)	25 mg/ml in ethanol	25 μg/ml

Table 2.3- Stock and final concentrations of antibiotics

2.5 Oligonucleotides

All oligonucleotides that were used in plasmid constructions or as sequencing primers were purchased from MWG Biotech AG. Oligonucleotide names, length, sequence and purpose are listed in Table 2.4.

Oligo Name	Length	5'-3' Sequence	Purpose	Reference
uni -43	23 bp	AGGGTTTTCCCAGTCACGACGTT	Sequencing primer	Sections 3.7, 4.2, 6.2
rev -49	22 bp	GAGCGGATAACAATTTCACACAGG	Sequencing primer	Section 2.7
22F	21 bp	CGCCAGGGTTTTCCCAGTCAC	Sequencing primer	Section 2.7
23R	23 bp	TCACACAGGAAACAGCTATGACC	Sequencing primer	Sections 2.7, 6.2
3resF	30 bp	AGTCGAGCTCACTAGTACTGCAACCGTTCG	Primer used to amplify Tn3 res site from pDB34 for cloning into pMP211	Section 2.7
3resR	24 bp	GCGCGAATTCTGTACCTTAAATCG	As above	Section 2.7
GSGF'	49 bp	GGCCGCAGGCGTACCGTGGACAGGGGCTCTGGCGGTTCCGGCGGCTCTA	GSG linker cloning to create pMP386	Sections 2.7, 4.5
GSGR'	49 bp	CTAGTAGAGCCGCCGGAACCGCCAGAGCCCCTGTCCACGGTACGCCTGC	GSG linker cloning to create pMP386	Sections 2.7, 4.5
hpF'	32 bp	ACGTGTACAAAGGCCGGAAGAAATCCCTGTCG	Primer used to amplify DNA-binding domain of Tn21 resolvase and allow creation of pMP264 and pMP265	Sections 2.7, 6.2
OCP1	18 bp	TTTGAGACACAACGTGGC	Sequencing primer	Section 3.5
OCP2	18 bp	AGTGCTGAACGAAACTCC	Sequencing primer	Section 3.5
R16RF'	40 bp	CACTAGTCTGTCTGAAATATTATAAATTATCAGACATAGG	Recombination site R16R. Cloned into pMP60	Sections 2.7, 5.7
R16RR'	48 bp	AATTCCTATGTCTGATAATTTATAATATTTCAGACAGACTAGTGAGCT	Recombination site R16R. Cloned into pMP60	Sections 2.7, 5.7
R18RF'	42 bp	CACTAGTCTGTCTGCAAATATTATAAATTATGCAGACATAGG	Recombination site R18R. Cloned into pMP61	Sections 2.7, 5.7
R18RR'	50 bp	AATTCCTATGTCTGCATAATTTATAATATTTGCAGACAGACTAGTGAGCT	Recombination site R18R. Cloned into pMP61	Sections 2.7, 5.7

Table 2.4-Oligonucleotides used in this study

Oligo Name	Length	5'-3' Sequence	Purpose	Reference
R20RF'	44 bp	CACTAGTCTGTCTGGCAAATATTATAAATTATGCCAGAC ATAGG	Recombination site R20R. Cloned into pMP62	Sections 2.7, 5.7
R20RR'	52 bp	AATTCCTATGTCTGGCATAATTTATAATATTTGCCAGAC AGACTAGTGAGCT	Recombination site R20R. Cloned into pMP62	Sections 2.7, 5.7
R22RF'	46 bp	CACTAGTCTGTCTGAGCAAATATTATAAATTATGCTCAG ACATAGG	Recombination site R22R. Cloned into pMP63	Sections 2.7, 5.7
R22RR'	54 bp	AATTCCTATGTCTGAGCATAATTTATAATATTTGCTCAG ACAGACTAGTGAGCT	Recombination site R22R. Cloned into pMP63	Sections 2.7, 5.7
RBO1	66 bp	GCTTCTAGAATTCCTATGTCTGATAATTTATAATNNNNC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 1	Sections 2.28, 3.3
RBO2	66 bp	GCTTCTAGAATTCCTATGTCTGATAATTTATANNNNTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 2	Sections 2.28, 3.3
RBO3	66 bp	GCTTCTAGAATTCCTATGTCTGATAATTTANNNNATTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 3	Sections 2.28, 3.3
RBO4	66 bp	GCTTCTAGAATTCCTATGTCTGATAATTNNNNATATTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 4	Sections 2.28, 3.3
RBO5	66 bp	GCTTCTAGAATTCCTATGTCTGATAANNNNTAATATTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 5	Sections 2.28, 3.3
RBO6	66 bp	GCTTCTAGAATTCCTATGTCTGATNNNNTATAATATTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 6	Sections 2.28, 3.3
RBO7	66 bp	GCTTCTAGAATTCCTATGTCTGNNNNTTTATAATATTTC GAACGGACTAGTGAGCTCCCGGGTACC	Tn3 site I random block oligo 7	Sections 2.28, 3.3
RSLP	25 bp	GGTACCCGGGAGCTCACTAGTCGCG	Primer for making random libraries of recombination sites	Sections 2.28, 3.3, 4.2, 6.2
Sin18N	68 bp	GCTTCTAGAATTCTGTGAANNNNNNNNNNNNNNNNNCA TACACGCGACTAGTGAGCTCCCGGGTACC	Sin site I random block oligo	Sections 2.28, 6.2
Sin6NC6N	68 bp	GCTTCTAGAATTCTGTGAANNNNNNGTACACNNNNNNCA TACACGCGACTAGTGAGCTCCCGGGTACC	Sin site I random block oligo	Sections 2.28, 6.2

Table 2.4-Continued.

Oligo Name	Length	5'-3' Sequence	Purpose	Reference
SinZF	32 bp	GTACAAAGGCCGTCGCCGTACCGTGGACCGTA	Sin Z-resolvase linker	Section 4.4
SinZR	32 bp	CTAGTACGGTCCACGGTACGGCGACGGCCTTT	As above	Section 4.4
Tn2118N	68 bp	GCTTCTAGAATTCCGTCAGNNNNNNNNNNNNNNNNNCT GATGCGCGACTAGTGAGCTCCCGGGTACC	Tn21 site I random block oligo	Sections 2.28, 6.2
Tn216NC6N	68 bp	GCTTCTAGAATTCCGTCAGNNNNNNGCATACNNNNNNCT GATGCGCGACTAGTGAGCTCCCGGGTACC	Tn21 site I random block oligo	Sections 2.28, 6.2
Tn21ZF	34 bp	GGCCGCAAAAAAAGCCTGTCCAGCGAACGCATTA	Tn21 Z-resolvase linker	Section 4.5
Tn21ZR	34 bp	CTAGTAATGCGTTCGCTGGACAGGCTTTTTTTGC	As above	Section 4.5
3toSinF	58 bp	GATCAAGGAGCGCCAAAATGAGGGCATTCAGGAAGCAAA GCTGAAAGGAATCTATAAG	Oligonucleotide introducing mutations L123K, T126Q, R130I, K139Y, F140K into Tn3 resolvase catalytic domain	Section 6.3
3toSinR	58 bp	GGCCCTTATAGATTCCTTTCAGCTTTGCTTCCTGAATGC CCTCATTTTGGCGCTCCTT	As above	Section 6.3
Z22N	75 bp	GCTTCTAGAATTCGCGTGGGCGNNNNNNNNNNNNNNNNN NNNNNCGCCCACGCGACTAGTGAGCTCCCGGGTACC	Tn3 Z-site random block oligo	Section 6.2
Z8NC8NZ	75 bp	GCTTCTAGAATTCGCGTGGGCGNNNNNNNNTTATAANNN NNNNNCGCCCACGCGACTAGTGAGCTCCCGGGTACC	Tn3 Z-site random block oligo	Section 6.2
Z22(Sin)F	49 bp	CACTAGTCGCGTGGGCGAAATTTGGGTACACCCTAATCA CGCCCACGCG	Sin Z-site, cloned into pMP212	Section 4.6
Z22(Sin)R	57 bp	AATTCGCGTGGGCGTGATTAGGGTGTACCCAAATTTCGC CCACGCGACTAGTGAGCT	As above	Section 4.6
Z22(Tn21)F	49 bp	CACTAGTCGCGTGGGCGAGGTTGAGGCATACCCTAACCT CGCCCACGCG	Tn21 Z-site, cloned into pMP2	Section 4.6
Z22(Tn21)R	57 bp	AATTCGCGTGGGCGAGGTTAGGGTATGCCTCAACCTCGC CCACGCGACTAGTGAGCT	As above	Section 4.6

Table 2.4-Continued

Oligo Name	Length	5’-3’ Sequence	Purpose	Reference
Z22(Tn21*)F	49 bp	CACTAGTCGCGTGGGCGTCGTTGAGGCATACCCTAACCTCGCCCACGCG	Tn21 Z-site*, cloned into pMP247	Section 4.6
Z22(Tn21*)R	57 bp	AATTCGCGTGGGCGAGGTTAGGGTATGCCTCAACGACGCCCACGCGACTAGTGAGCT	As above	Section 4.6

Table 2.4-Oligonucleotides used in this study

2.6 Custom DNA synthesis

In order to reduce the cost and increase the efficiency of some plasmid constructions, a pair of precursor plasmids was commissioned from GeneART. These plasmids were designed to contain several fragments of interest arranged in a series which are separated by convenient restriction sites (akin to carriages of a train). The restriction sites enable each of the fragments to be cut out and used independently of others when required. The diagrams illustrating the principle and giving the sequences of the fragments are given in Fig. 2.1 and Fig. 2.2.

2.7 Plasmids

A complete list of plasmids used and generated in the presented work is provided in Table 2.5. The way that specific plasmids or groups of plasmids were constructed is shown in diagrams in Figures 2.3-2.9.

2.7.1 Resolvase expression plasmids

Resolvase expression plasmids created during this study were of two general types, differing in the level of protein expression; namely, low level resolvase expression plasmids, used for the assaying resolvase *in vivo*, and over-expression plasmids used to generate high levels of resolvase needed for protein purification.

2.7.1.1 Low level resolvase expression plasmids

The low level resolvase expression plasmids were constructed based on either pMS140 or pACYC184, which have different replication origins but a comparable copy number. The names of plasmids based on the pMS140 backbone are highlighted in light green, whereas the ones based on pACYC184 are highlighted in dark green, respectively, in Table 2.5.

pMS140 has a *Col*E1 origin of replication and exists at ~20 copies per cell. It carries an ampicillin resistance marker and it expresses a cassetted version of wild-type Tn3 resolvase encoded on an NdeI/Asp718 fragment. Resolvase expression is driven by a cryptic promoter situated within ~400 bp upstream of the NdeI site. The level of resolvase expression is very low, which is suitable for *in vivo* recombination assays (Section 2.23). An important site in pMS140 and its derivatives is the unique EagI site that allows an incision to be made in the resolvase reading frame just after the GR motif (residues G141 and R142) at the base of the extended linker. An incision at this position enables the replacement of the Tn3 resolvase C-terminal domain with other domains, such as Zif268

Figure 2.1- Commercially synthesised plasmid GeneART1. Its small NcoI/BamHI fragment (blue DNA sequence) was used to introduce mutations K129L, Q132T, I136R, Y145K, K146F into pSA9994 (encoding Sin resolvase Q115R) in order to make pMP282, encoding Sin/3 resolvase. Sin resolvase residues are noted in orange on top of the DNA sequence while the mutated residues are given in grey with their respective position within the protein reading frame noted above them. The small EcoRI/SacI fragment (black DNA sequence) was used to clone Sin site I into pMTL23 vector to create pMP387.Restriction sites are marked by pink boxes while the Sin site I sequence is marked by a yellow box.

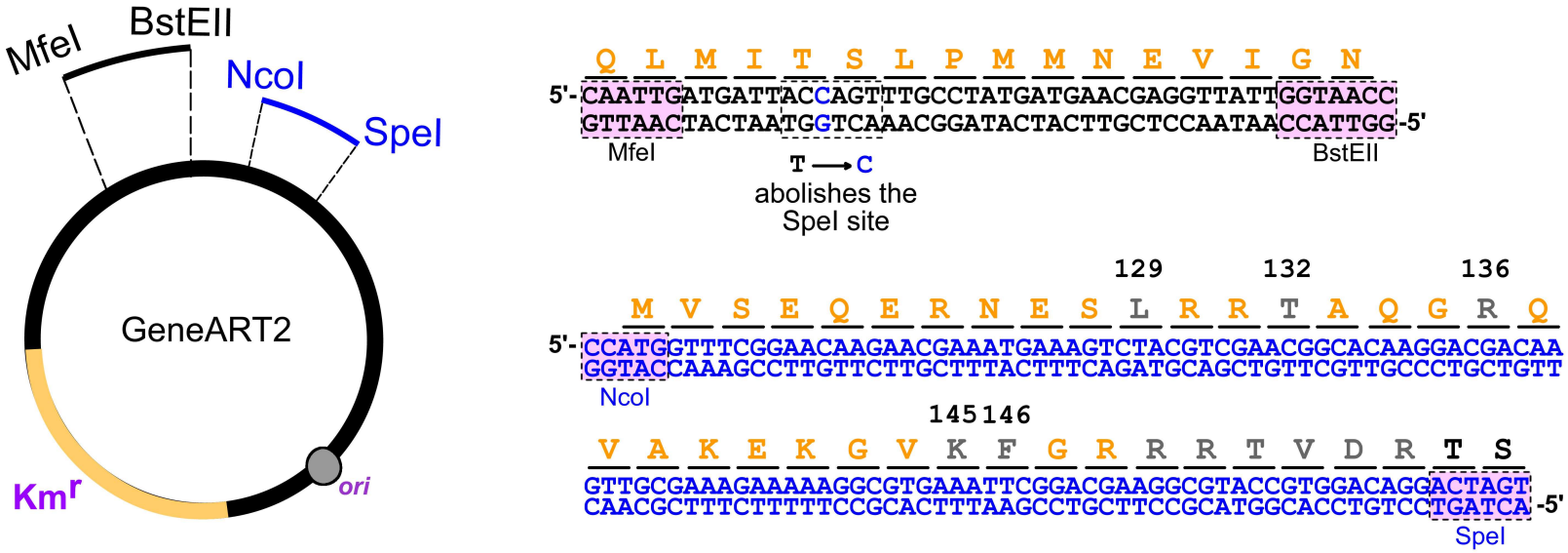


Figure 2.2- Commercially synthesised plasmid GeneART2. Its small MfeI/BstEII fragment (black DNA sequence) was used to abolish one of the SpeI sites within Z-R(Sin) reading frame to create pMP379 (encoding Z-R(Sin) with a unique SpeI site). The mutant base C (top strand) is given in blue. The small NcoI/SpeI fragment (blue DNA sequence) was used to introduce mutations K129L, Q132T, I136R, Y145K, K146F into pMP379 (encoding Z-R(Sin) with a unique SpeI site) in order to make pMP284, encoding Z-R(Sin/3). Z-R(Sin) residues are noted in orange on top of the DNA sequence while the mutated residues are given in grey with their respective position within the protein reading frame noted above them. Restriction sites are marked by pink boxes.

Plasmid	Size (bp)	Antibiotic marker	Description/Derivation	Source/Reference
pACYC184	4244	Cm^r, Tet^r	Cloning vector	Rose, 1988
pAMC11	5636	Ap^r	pMS140 derivative encoding the Z-R(NMF) Z-resolvase (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	A. MacLean / M. Stark
pAN8	5653	Ap^r	pMS140 derivative encoding the AN8 Z-resolvase ORF	A. Akopian
pAN15	5656	Ap^r	pMS140 derivative encoding the AN15 Z-resolvase ORF	A. Akopian
pAN28	5557	Ap^r	pMS140 derivative encoding the AN28 Z-resolvase ORF	A. Akopian
pAT5Δ	5465	Ap^r	pAT5 vector (encoding Tn3 resolvase ORF for low level-expression in *E. coli*) with deletion of its EcoRV-NruI region	M. Boocock
pCO1	2543	Ap^r	pMTL23 with a Tn3 site I cloned between its EcoRI and SacI sites	C. Muir / M. Stark
pCP132	4943	Km^r	*In vivo* recombination substrate with two copies of Z22Z(AA) site in direct repeat flanking a galK gene, pMS183 backbone	C. Proudfoot
pDB34	8371	Km^r	*In vivo* recombination substrate with two copies of Tn3 *res* site in direct repeat flanking a galK gene	D. Blake
pDJS1	5469	Ap^r	pDW21 derivative encoding Tn21 resolvase (E124V) ORF	D. Sneddon / M. Stark
pDJS2	5469	Ap^r	pDW21 derivative encoding Tn21 resolvase (M63T) ORF	D. Sneddon / M. Stark
pDW21	5469	Ap^r	pMS140 derivative encoding cassetted Tn21 resolvase ORF	D. Wenlong/ M. Stark

Table 2.5-Plasmids constructed in this study and their precursors. Names of plasmids are highlighted in different colours according to plasmid type: light green – pMS140 based low level resolvase expression plasmids (Section 2.7.1.1), dark green – pACYC184 based low level resolvase expression plasmids (Section 2.7.1.1), light blue – pSA1101 based resolvase over-expression plasmids (Section 2.7.1.2), grey – pMTL23 based single recombination site plasmids (Section 2.7.2.1), red – two site *in vivo* recombination substrate plasmids (Section 2.7.2.2), purple – two site *in vitro* recombination substrate plasmids (Section 2.7.2.2).

PDW24	5131	Km^r	*In vivo* recombination substrate with two copies of Tn21 *res* site in Direct repeat flanking a *galK* gene, pMS183 backbone	D. Wenlong/ M. Stark
pDW25	4947	Km^r	*In vivo* recombination substrate with two copies of Tn21 site I in Direct repeat flanking a *gal*K gene, pMS183 backbone	D. Wenlong/ M. Stark
pDW27	5039	Km^r	Tn21 *res* site × Tn21 site I, *in vivo* recombination substrate with a *galK* indicator, pMS183 backbone	D. Wenlong/ M. Stark
pFO2	5469	Ap^r	pMS140 derivative encoding Tn3 resolvase mutant NM (R2A, E56K, G101S, D102Y, M103I, Q105L) ORF	F. Olorunniji
pFO5	5469	Ap^r	pMS140 derivative encoding Tn3 resolvase mutant AKSY (R2A, E56K, G101S, D102Y) ORF	F. Olorunniji
pFO32	5469	Ap^r	pMS140 derivative encoding Tn3 resolvase mutant NM, S10A (R2A, S10A, E56K, G101S, D102Y, M103I, Q105L) ORF	F. Olorunniji
pFO102	6692	Km^r	NM resolvase over-expression plasmid derived from pSA1101 and pFO2	F. Olorunniji
pIVS6	3937	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Z22Z Z-site flanking the Km^r marker from pUC71K	D. Wenlong/ M. Stark
pJH2	5469	Ap^r	pMS140 derivative encoding the Tn3 resolvase mutant M (G101S, D102Y, M103I, Q105L) ORF	J. He
pMON1	6220	Cm^r, Tet^r	pACYC184 derivative encoding Z-R(NMF) Z-resolvase (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	M. Sentmanat / M. Stark
pMP1	6859	Km^r	Z-R(NMF) overexpression plasmid derived from pSA1101 and pAMC11	Section 5.5
pMP2	2552	Ap^r	pMTL23 with a Z22Z(Tn21) site cloned between its EcoRI and SacI sites	Section 2.7.2

Table 2.5-Continued.

pMP6	3937	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Z22Z(Tn21) Z-site flanking the Km^r marker from pUC71K	Section 4.7
pMP15	5636	Ap^r	pAMC11 and pMS140 derivative encoding the Z-R(wt) Z-resolvase ORF	Section 5.4
pMP17	5469	Ap^r	pMS140 and pAMC11 derivative encoding the Tn3 resolvase mutant NMF (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	Section 5.6
pMP19	5469	Ap^r	pAMC11 and pFO32 derivative encoding the Z-R(NMF,S10A) Z-resolvase (R2A, S10A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	Section 5.6
pMP20	4897	Km^r	Library precursor plasmid derived from pMS183Δ with one copy of Tn3 site I between BsrGI/NheI sites	Section 3.5
pMP21	4869	Km^r	Library precursor plasmid derived from pMS183Δ with one copy of Tn3 site I between sites Asp718/XbaI sites	Section 3.5
pMP30	2549	Ap^r	pMTL23 with a Z16Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP31	2551	Ap^r	pMTL23 with a Z18Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP32	2553	Ap^r	pMTL23 with a Z20Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP33	2555	Ap^r	pMTL23 with a Z22Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP34	2557	Ap^r	pMTL23 with a Z24Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP35	2559	Ap^r	pMTL23 with a Z26Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP36	2561	Ap^r	pMTL23 with a Z28Z site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP47	3163	Km^r	pCO1 derivative in which Ap^r marker was cut out with BspHI and replaced by BspHI 1628 bp fragment containing Km^r from pUC71K	Section 3.6
pMP50	4931	Km^r	*In vivo* recombination substrate with two copies of Z16Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP51	4935	Km^r	*In vivo* recombination substrate with two copies of Z18Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2

Table 2.5-Continued.

pMP52	4939	Km^r	*In vivo* recombination substrate with two copies of Z20Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP53	4943	Km^r	*In vivo* recombination substrate with two copies of Z22Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP54	4947	Km^r	*In vivo* recombination substrate with two copies of Z24Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP55	4951	Km^r	*In vivo* recombination substrate with two copies of Z26Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP56	4955	Km^r	*In vivo* recombination substrate with two copies of Z28Z site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP57	2549	Ap^r	pMTL23 with a ZR site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP58	2549	Ap^r	pMTL23 with a ZL site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP59	5636	Ap^r	pAMC11 and pFO2 derivative encoding the Z-R(NM) Z-resolvase (R2A, E56K, G101S, D102Y, M103I, Q105L) ORF	Section 5.4.
pMP60	2543	Ap^r	pMTL23 with a R16R site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP61	2545	Ap^r	pMTL23 with a R18R site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP62	2547	Ap^r	pMTL23 with a R20R site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP63	2549	Ap^r	pMTL23 with a R22R site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP78	3919	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Tn3 site I, flanking the Km^r marker from pUC71K	Section 5.5
pMP89	4929	Km^r	*In vivo* recombination substrate with two copies of Z18Z(AA) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP90	4933	Km^r	*In vivo* recombination substrate with two copies of Z20Z(AA) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2

Table 2.5-Continued.

pMP91	4941	Km^r	*In vivo* recombination substrate with two copies of Z24Z(AA) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP92	4945	Km^r	*In vivo* recombination substrate with two copies of Z26Z(AA) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP93	4949	Km^r	*In vivo* recombination substrate with two copies of Z28Z(AA) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.2
pMP94	4943	Km^r	*In vivo* recombination substrate with two copies of Z22Z(Tn21) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 4.6
pMP96	4931	Km^r	*In vivo* recombination substrate with two copies of ZR site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.6
pMP97	4931	Km^r	*In vivo* recombination substrate with two copies of ZL site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.6
pMP120	2554	Ap^r	pMTL23 with a Z21Z(R) site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP121	2554	Ap^r	pMTL23 with a Z21Z(L) site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP122	2556	Ap^r	pMTL23 with a Z23Z(R) site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP123	2556	Ap^r	pMTL23 with a Z23Z(L) site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP129	6220	Cm^r, Tet^r	pMON1 and pMP19 derivative encoding Z-R(NMF,S10A) Z-resolvase (R2A, S10A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	Section 5.6
pMP136	6859	Km^r	Z-R(Tn3/S) overexpression plasmid, derivative of pMP385 and pSA1101	Section 6.4
pMP137	6692	Km^r	Tn21 resolvase (M63T) overexpression plasmid, derivative of pDJS2 and pSA1101	Section 4.7
pMP138	6762	Km^r	Sin resolvase (Q115R) overexpression plasmid, derivative of pSA9994 and pSA1101	Section 4.7
pMP201	4982	Km^r	*In vivo* recombination substrate with two copies of Z21Z(R) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.4

Table 2.5-Continued.

pMP202	4982	Km^r	*In vivo* recombination substrate with two copies of Z21Z(L) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.4
pMP203	4986	Km^r	*In vivo* recombination substrate with two copies of Z23Z(R) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.4
pMP204	4986	Km^r	*In vivo* recombination substrate with two copies of Z23Z(L) site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.4
pMP205	4919	Km^r	*In vivo* recombination substrate with two copies of R16R site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.7
pMP206	4964	Km^r	*In vivo* recombination substrate with two copies of R18R site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.7
pMP207	4968	Km^r	*In vivo* recombination substrate with two copies of R20R site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.7
pMP208	4972	Km^r	*In vivo* recombination substrate with two copies of R22R site in direct repeat flanking a galK gene, pMS183Δ backbone	Section 5.7
pMP211	2633	Ap^r	pMTL23 with a Tn3 *res* site from pDB34 cloned between its EcoRI and SacI sites, pMS183Δ backbone	Section 2.7.2
pMP212	2552	Ap^r	pMTL23 with a Z22Z(Sin) site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP213	5654	Ap^r	Low level-expression plasmid encoding Z-R(Sin) Z-resolvase ORF, derived from pAMC11 and pSA9994	Section 4.4
pMP217	4978	Km^r	*In vivo* recombination substrate with two copies of Z22Z (Sin) Z-site in direct repeat flanking a *gal*K gene, pMS183Δ backbone	Section 4.6
pMP240	3937	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Z22Z (Sin) Z-site flanking the Km^r marker from pUC71K	Section 4.7

Table 2.5-Continued.

pMP243	4960	Km^r	*In vivo* recombination substrate with two copies of Tn3 site I in direct repeat flanking a *gal*K gene, pMS183Δ backbone	Section 3.5
pMP244	5140	Km^r	*In vivo* recombination substrate with two copies of Tn3 *res* site in direct repeat flanking a *galK* gene, pMS183Δ backbone	Section 3.1
pMP245	5050	Km^r	Tn3 *res* site × Tn3 site I, *in vivo* recombination substrate with a *galK* indicator, pMS183Δ backbone	Section 3.1
pMP247	2552	Ap^r	pMTL23 with a Z22Z(Tn21)* site cloned between its EcoRI and SacI sites	Section 2.7.2
pMP248	4978	Km^r	*In vivo* recombination substrate with two copies of Z22Z(Tn21)* site in direct repeat flanking a *gal*K gene, pMS183Δ backbone	Section 4.6
pMP250	5636	Ap^r	pAMC11 and pDJS1 derivative encoding the Z-R(Tn21, E124V, L3) Z-resolvase ORF	Section 4.5.2
pMP251	5645	Ap^r	pAMC11 and pDJS1 derivative encoding the Z-R(Tn21, E124V, L21) Z-resolvase ORF	Section 4.5.2
pMP252	5636	Ap^r	pAMC11 and pDJS2 derivative encoding the Z-R(Tn21, M63T, L3) Z-resolvase ORF	Section 4.5.2
pMP253	5645	Ap^r	pAMC11 and pDJS2 derivative encoding the Z-R(Tn21, M63T, L21) Z-resolvase ORF	Section 4.5.2
pMP254	5636	Ap^r	pAMC11 and pMS68 derivative encoding the Z-R(YQ) Z-resolvase (D102Y, E124Q) ORF	Section 5.3
pMP255	5636	Ap^r	pAMC11 and pMS74 derivative encoding the Z-R(SY) Z-resolvase (G101S, D102Y) ORF	Section 5.3
pMP256	5636	Ap^r	pAMC11 and pFO5 derivative encoding the Z-R(AKSY) Z-resolvase (R2A, E56K, G101S, D102Y) ORF	Section 5.3
pMP257	5636	Ap^r	pAMC11 and pJH2 derivative encoding the Z-R(M) Z-resolvase (G101S, D102Y, M103I, Q105L) ORF	Section 5.3
pMP264	5487	Ap^r	Low level-expression plasmid encoding wild-type Sin/Tn21 resolvase hybrid ORF, derived from pSA9944 and pDW21	Section 6.4
pMP265	5487	Ap^r	Low level-expression plasmid encoding Sin (Q115R)/Tn21 resolvase hybrid ORF, derived from pSA9994 and pDW21	Section 6.2

Table 2.5-Continued.

pMP267	5636	Ap^r	A derivative of pDJS2 and pMP250 encoding the Z-R(Tn21, TV, L3) Z-resolvase (M63T, E124V) ORF	Section 4.5.2
pMP268	5645	Ap^r	A derivative of pDJS2 and pMP251 encoding the Z-R(Tn21, TV, L21) Z-resolvase (M63T, E124V) ORF	Section 4.5.2
pMP269	5469	Ap^r	Low level-expression plasmid encoding the Tn21 resolvase mutant (M63T, E124V) ORF, derived from pDJS1 and pDJS2	Section 4.5.2
pMP282	5539	Ap^r	Low level-expression plasmid encoding Sin resolvase mutant Sin(Q115R)/3 (Q115R, K129L, Q132T, I136R, Y145K, K146F) ORF	Section 6.3
pMP284	5654	Ap^r	Low level-expression plasmid encoding Sin Z-resolvase, Z-R(Sin/3), (Q115R, K129L, Q132T, I136R, Y145K, K146F) ORF	Section 6.3
pMP287	5469	Ap^r	Low level-expression plasmid encoding Tn3 resolvase mutant Tn3/S (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F, L123K, T126Q, R130T, K139Y, F140K) ORF	Section 6.3
pMP368	6877	Km^r	Sin Z-resolvase Z-R(Sin) overexpression plasmid, derivative of pMP213 and pSA1101	Section 4.7
pMP371	6859	Km^r	Tn21 Z-resolvase Z-R(Tn21, M63T, L3) overexpression plasmid, derivative of pMP252 and pSA1101	Section 4.7
pMP378	3897	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Sin site I, flanking the Km^r marker from pUC71K	Section 4.7
pMP379	5654	Ap^r	pMP213 derivative with a SpeI site removed using the MfeI/BstEII GenaART2 cassette	Section 6.3
pMP385	5636	Ap^r	Low level-expression plasmid encoding Tn3 Z-resolvase Z-R(Tn3/S) (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F, L123K, T126Q, R130T, K139Y, F140K) ORF	Section 6.3
pMP386	5660	Ap^r	Low level-expression plasmid encoding Z-resolvase, Z-R(NMF,GSG), (R2A, E56K, G101S, D102Y, M103I, Q105L, V107F) ORF	Section 5.5
pMP387	2542	Ap^r	pMTL23 with a Sin site I cloned between its EcoRI and SacI sites	Section 2.7.2
pMP388	4942	Km^r	*In vivo* recombination substrate with two copies of Sin site I in direct repeat flanking a *gal*K gene, pMS183Δ backbone	Section 4.6

Table 2.5-Continued.

pMP389	6883	Km^r	Z-resolvase Z-R(NMF,GSG) overexpression plasmid, derivative of pMP386 and pSA1101	Section 5.5
pMP390	3907	Ap^r, Km^r	*In vitro* recombination substrate containing two copies of Tn21 site I, flanking the Km^r marker from pUC71K	Section 4.7
pMP393	6859	Km^r	Z-R(NM) Z-resolvase overexpression plasmid, derivative of pMP59 and pSA1101	Section 5.5
pMS68	5465	Ap^r	pMS140 derivative encoding Tn3 resolvase mutant YQ (D102Y, E124Q) ORF	M. Stark
pMS74	6465	Ap^r	encoding Tn3 resolvase mutant YQ (G101S, D102Y) ORF	M. Stark
pMS140	5469	Ap^r	pAT5Δ derivative encoding the wild-type Tn3 resolvase ORF, low level of protein expression, suitable for *in vivo* recombination assays	M. Stark
pMS183	4879	Km^r	In vivo recombination substrate precursor plasmid, allows cloning of recombination sites between BglII/BsrGI and NcoI/XbaI, so that they flank the *galK* indicator	D. Wenlong/ M. Stark
pMS183Δ	4863	Km^r	Derivative of pMS183 with a NdeI/NdeI 16 bp deletion that improves plasmid stability when cloning-in (nearly) palindromic sites	Section 2.7.2
pMTL23	2505	Ap^r	Cloning vector	Chambers *et al.*, 1988
pSA1101	6692	Km^r	Wild-type Tn3 resolvase overexpression plasmid, T7 promoter	Arnold *et al.*, 1999
pSA9944	5535	Ap^r	Low level-expression plasmid encoding wild-type Sin resolvase	S. Rowland
pSA9994	5539	Ap^r	Low level-expression plasmid encoding Sin resolvase (Q115R)	S. Rowland
pSB404	7350	Km^r	pDB34 based in vivo binding assay plasmid with Sin site I overlapping the *gal*K gene promoter	S. Rowland
pSB423	5088	Km^r	*In vivo* recombination substrate with two copies of Sin *res* site in direct repeat flanking a *gal*K gene, pMS183Δ backbone	S. Rowland
pUC71K	3966	Ap^r, Km^r	Cloning vector, a source of Km^r gene	Taylor & Rose, 1982

Table 2.5- Plasmids constructed in this study and their precursors. Names of plasmids are highlighted in different colours according to plasmid type: light green – pMS140 based low level resolvase expression plasmids (Section 2.7.1.1), dark green – pACYC184 based low level resolvase expression plasmids (Section 2.7.1.1), light blue – pSA1101 based resolvase over-expression plasmids (Section 2.7.1.2), grey – pMTL23 based single recombination site plasmids (Section 2.7.2.1), red – two site *in vivo* recombination substrate plasmids (Section 2.7.2.2), purple – two site *in vitro* recombination substrate plasmids (Section 2.7.2.2).

Direct equivalents of pMS140 that encode Tn21 resolvase and Sin resolvase on an NdeI/Asp718 fragment are pDW21 and pSA9944, respectively. pDW21 and its derivatives pDJS1 and pDJS2 have an EagI site at an equivalent position to that in pMS140, allowing for the same kinds of C-terminal domain replacements (Fig. 2.3; Chapters 4 and 6).

pSA9944 and its derivative pSA9994 do not have an EagI site at a position that allows for splitting the N-terminal and C-terminal domains at equivalent position to Tn3 and Tn21 resolvases. The closest site to this position in pSA9944/pSA9994 is BsrGI. In order to replace the C-terminal domain of Sin resolvase with the Zif268 zinc finger DNA binding domain to create a Sin Z-resolvase-expressing plasmid pMP213 (Chapter 4), a synthetic -oligonucleotide linker (made by annealing SinZF and SinZR oligos, Section 4.4) was used (Fig.2.4).

Replacement of the C-terminal domain of Sin resolvase with the equivalent Tn21 resolvase domain to make Sin/Tn21 hybrid resolvase expressing plasmid pMP265 (Chapter 6) was done using PCR (Fig.2.5). The Tn21 resolvase C-terminal domain was amplified from the pDW21 template, using the partially annealing forward primer hpF' and the 23R reverse primer. Partially annealing primer hpF' was designed to introduce a BsrGI site and thus allow joining of the amplified Tn21 resolvase C-terminal domain to the Sin resolvase catalytic domain. Twenty five cycles of PCR were performed in a 30 µl mix (1.25 mM dNTP, 1.25 mM $MgCl_2$, 3 µl of 10% Taq buffer (50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100), 1 unit Taq polymerase). The PCR program was: 94 °C, 2 min; 25 × (94 °C, 30 sec, 55°C, 30 sec, 72°C, 30s). The amplified 233 bp fragment was restricted with BsrGI/Asp718. The resulting 150 bp BsrGI/Asp718 fragment encoding the Tn21 resolvase C-terminal domain was then cloned into BsrGI/Asp718-restricted pSA9944 or pSA9994 vectors, creating Sin/Tn21 resolvase-expressing plasmids pMP264 and pMP265 respectively.

The other type of low level resolvase expression plasmids are plasmids based on the pACYC184 backbone. pACYC184 has a p15A origin of replication, chloramphenicol and tetracycline antibiotic resistance markers and it exists at ~15 copies per cell (Chang & Cohen, 1978). Its derivative, pMON1 encodes Z-resolvase Z-R(NMF) on its NdeI/Asp718 fragment, expression of which is driven by the same cryptic promoter somewhere in the 400 bp preceding the NdeI site as in pMS140. By performing an NdeI/Asp718 fragment swap between pMON1 and the pMS140-based low level resolvase expression plasmidpMP19, pMP129 encoding Z-R(NMF,S10A) Z-resolvase was made. The main use for the plasmids with a pACYC184 backbone was in the *in vivo* complementation

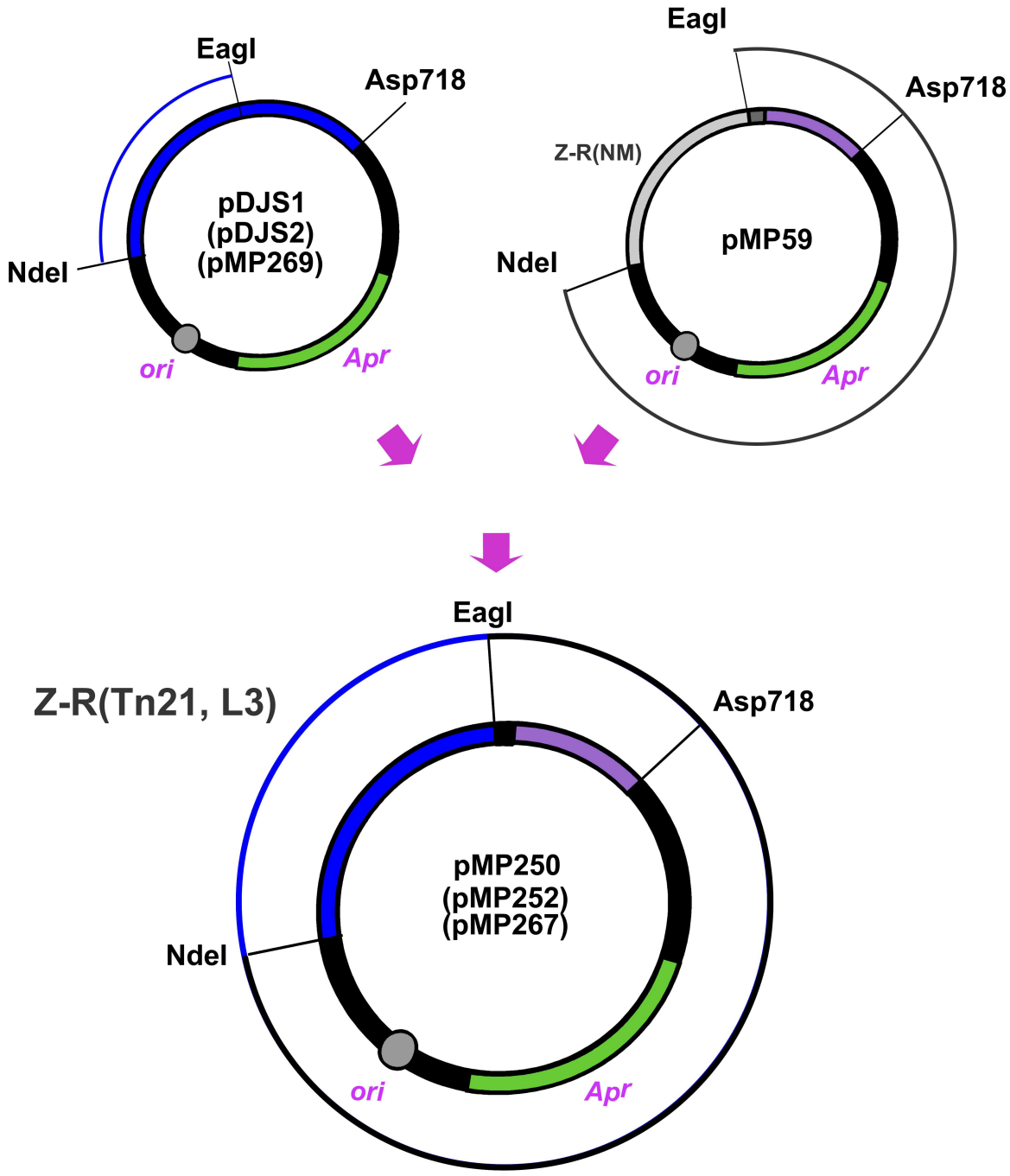


Figure 2.3- Schematic illustration of cloning steps used to generate plasmids pMP250, pMP252 and pMP267 (encoding Z-R(Tn21,124V, L3), Z-R(Tn21,63T,L3) and Z-R(Tn21, 63T/124V, L3), respectively). To make these plasmids a 421 bp Nde/EagI fragment of pDJS1, pDJS2 and pMP269 providing the Tn21 resolvase catalytic domain residues 1-142 (blue fragment), was cloned into a 5125bp NdeI/EagI fragment of pMP59 supplied the backbone and the Zif268 DNA-binding domain residues 2-90 (black fragment).

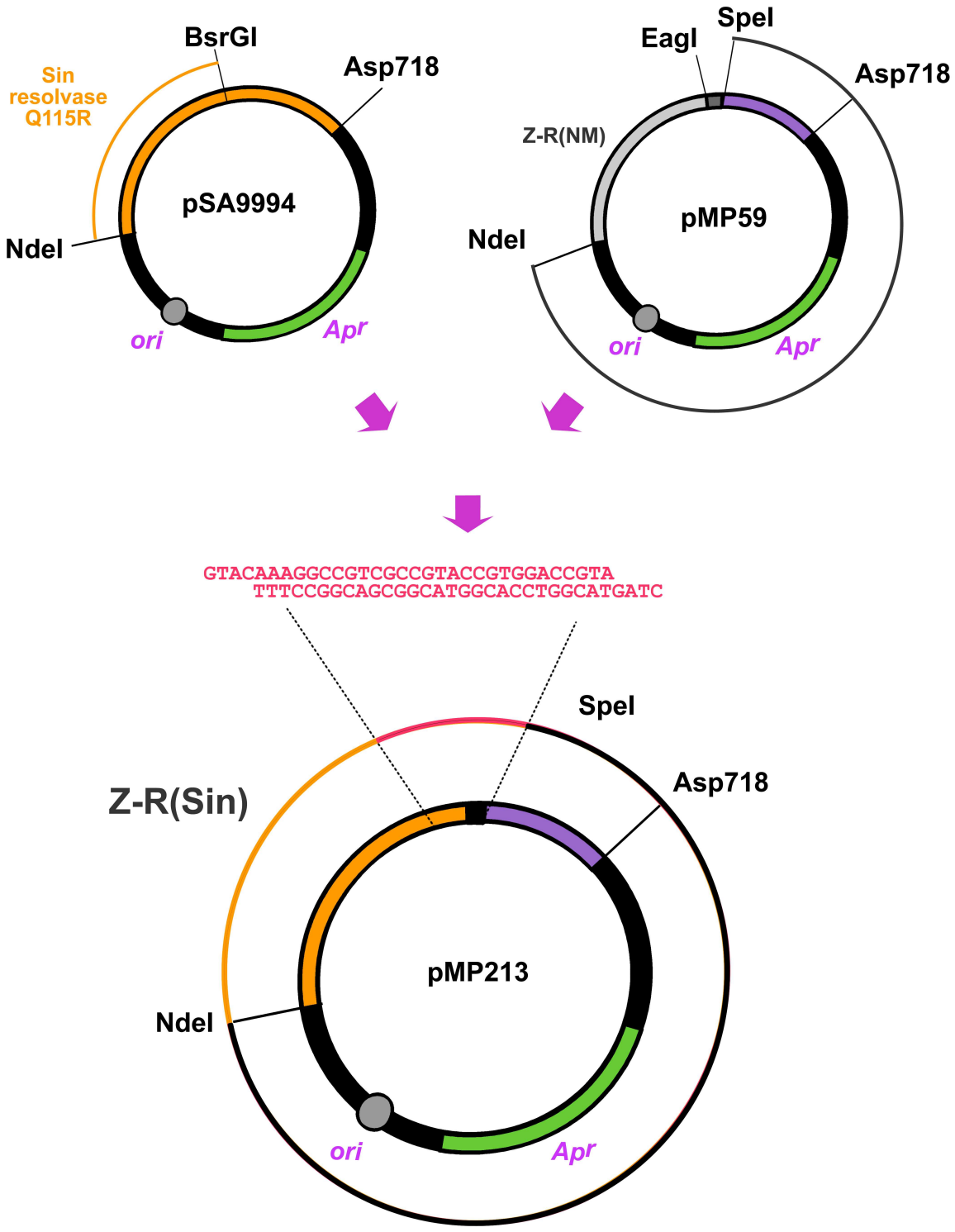


Figure 2.4- Schematic illustration of cloning steps used to generate pMP213 (encoding Z-R(Sin)). To make this plasmid a three piece ligation was performed using a 432 bp NdeI/BsrGI fragment of pSA9994 (orange segment) which provided the catalytic domain for the Z-R(Sin), a 28 bp BsrGI/SpeI linker (pink segment) and a 5190 bp NdeI/SpeI fragment of pMP59 (black segment) acting as a backbone and supplying the part of the Z-resolvase ORF coding for the Zif268 DNA binding domain. The BsrGI/SpeI linker fragment was made by annealing synthetic oligonucleotides SinZF and SinZR (see Table 2.4).

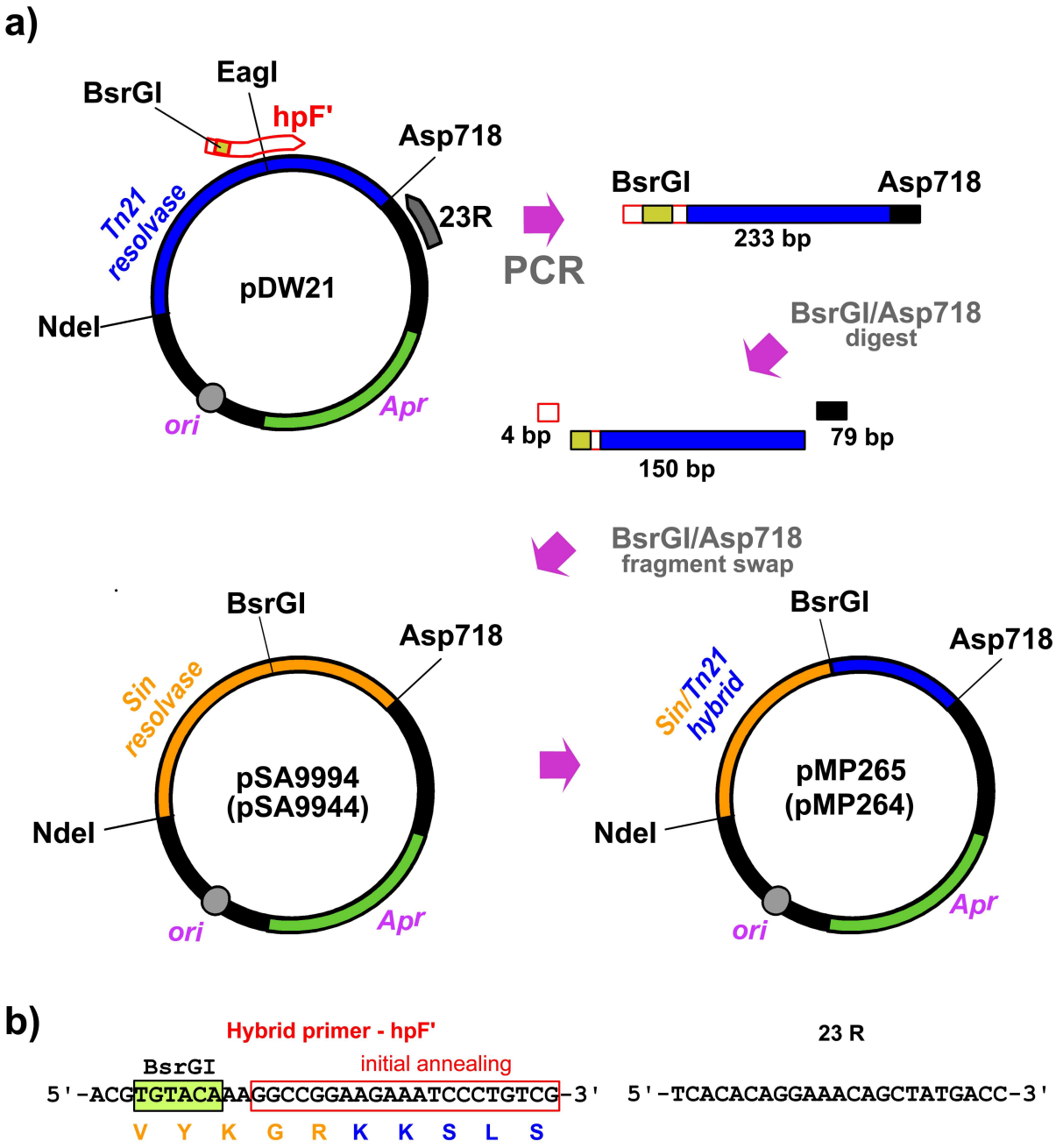


Figure 2.5- a) Schematic illustration of steps used to generate pMP265 (pMP264), encoding Sin-Tn21 resolvase hybrid. Using the primers hpF' and 23R, the Tn21 resolvase C-terminal domain was amplified from the pDW21 template. Partially annealing primer hpF' introduces a BsrGI site allowing joining of the amplified Tn21 resolvase C-terminal domain (blue segment) to the Sin resolvase catalytic domain (orange segment). The amplified 233 bp fragment was restricted with BsrGI/Asp718, resulting in the 150 bp BsrGI/Asp718 fragment encoding the Tn21 resolvase C-terminal domain (blue segment). This fragment was then cloned into BsrGI/Asp718-restricted pSA9994 or pSA9944 vectors, to create Sin-Tn21 resolvase-expressing plasmids pMP265 and pMP264 respectively. **b)** DNA sequence of primers hpF' and 23R. Portion of hpF' is in a red box while the BsrGI site is in the yellow box. Amino acid sequence showing the join between Sin (orange) and Tn21 resolvase (blue) is given below the hpF' DNA sequence.

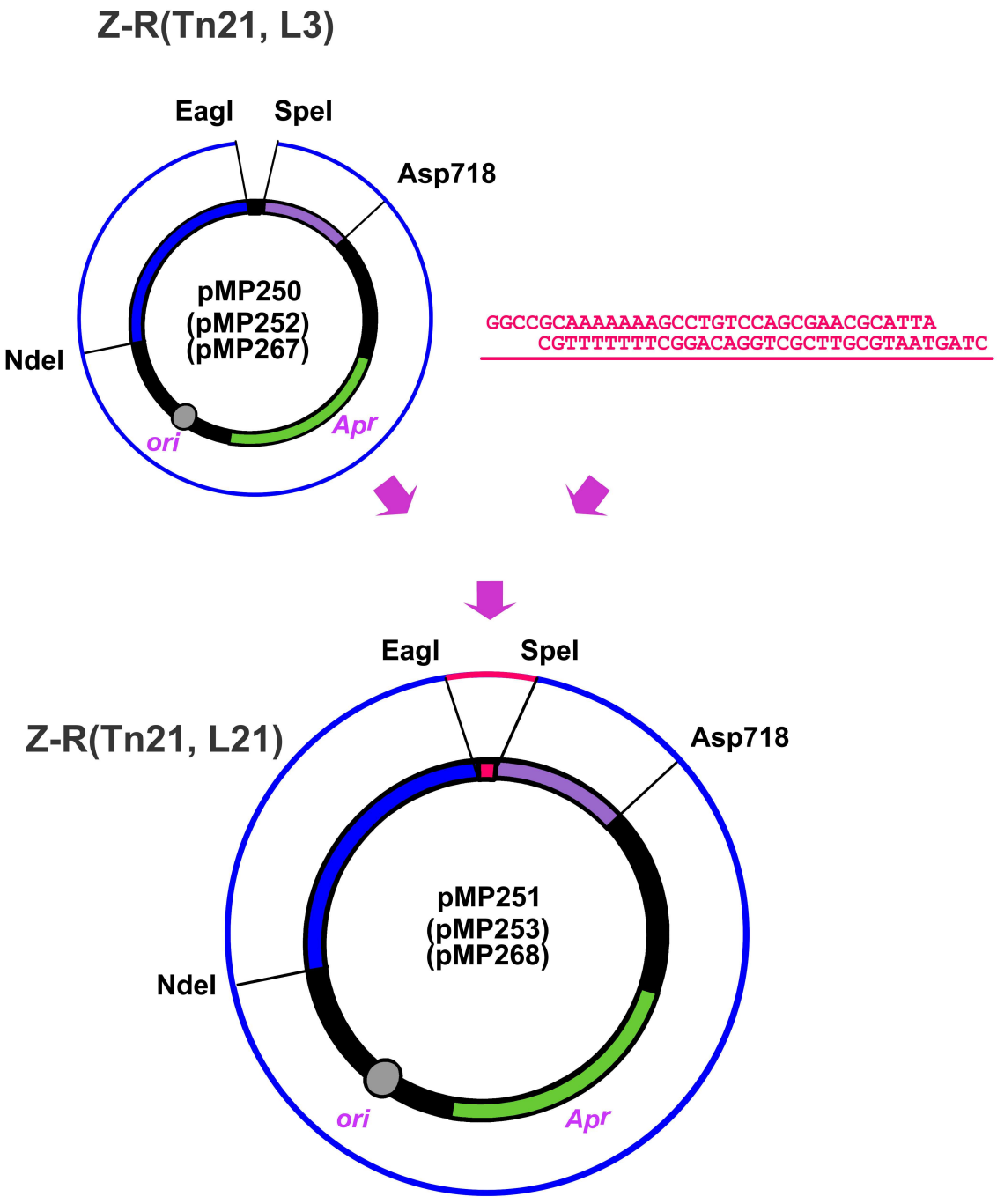


Figure 2.6- Schematic illustrating the cloning steps needed to generate pMP251, pMP253, pMP268 (encoding Z-R(Tn21, 124V, L21), Z-R(Tn21, 63T, L21) and Z-R(Tn21, 63T/124V, L21), respectively). To create these plasmids EagI/SpeI cut pMP250, pMP252 and pMP267 were used as vectors (blue segment), while a short DNA fragment generated by annealing oligonucleotides Tn21ZF and Tn21ZR (pink segment) was used as the insert. This cloning replaces L3 linker with the L21 linker, see text for details (Section 4.5). The drawing is not to scale.

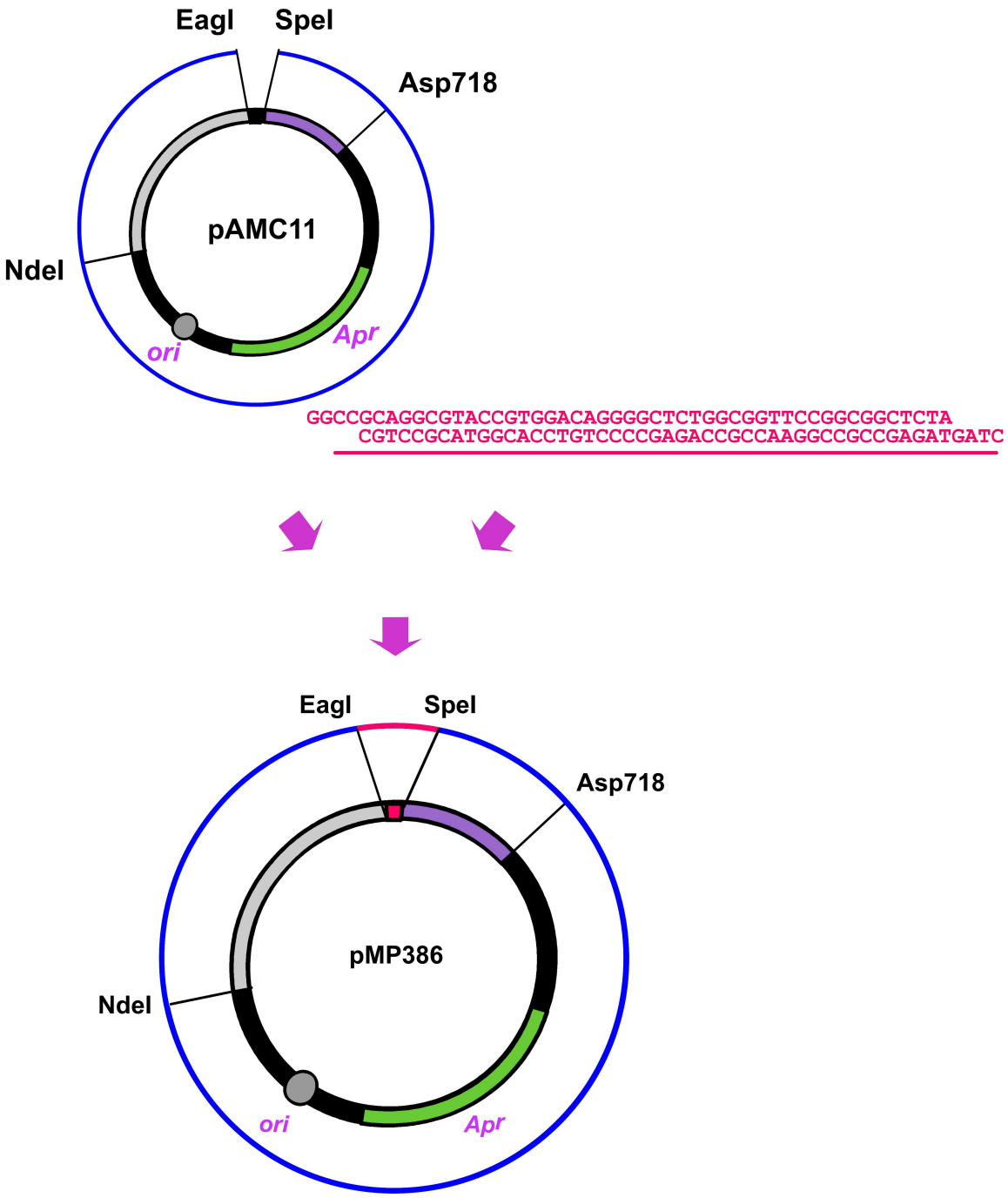


Figure 2.7- Schematic illustrating the cloning steps needed to generate pMP386, the plasmid encoding Z-R(NMF,GSG). To make this plasmid a short double-stranded DNA fragment with EagI/SpeI overhangs (pink segment), created by annealing oligos GSGF' and GSGR'(Table 2.4), was cloned into EagI/SpeI-digested pAMC11 vector (blue segment). This introduced the desired linker, see text for details (Section 5.5). The drawing is not to scale.

experiments (Chapter 5). As their origin of replication is different to the *ColE*1 origin of pMS140-based plasmids, these plasmids can coexist in one cell at the same time, avoiding plasmid incompatibility issues, with each one of them expressing its own resolvase variant, which was vital for these experiments. Although both chloramphenicol and tetracycline antibiotic resistance markers are present on pACYC184-based plasmids, routinely only chloramphenicol resistance was selected for in all the experiments discussed.

2.7.1.2 Over-expression plasmids

All of the over-expression plasmids constructed are based on pSA1101, and their names are highlighted in light blue in the Table 2.5. This plasmid has a ColE1 origin of replication, a kanamycin resistance marker, and has an inducible T7 promoter which drives over-expression of a wild-type Tn3 resolvase gene encoded on an NdeI/Asp718 fragment. All the other over-expression plasmids used in this study were generated by swapping the NdeI/Asp718 Tn3 resolvase-encoding fragment between the pSA1101 backbone and a low level resolvase expression plasmid providing the resolvase variant-encoding insert.

2.7.2 Substrate plasmids

The precursors for both *in vivo* and *in vitro* two-site substrate plasmids were pMTL23-based plasmids with a single recombination site cloned between EcoRI and SacI restriction sites. pMTL23 is an ampicillin resistance marker-carrying cloning vector, with a deregulated ColE1 origin of replication, and thus a very high copy number. This coupled with a large number of useful restriction sites in its polylinker region make this vector ideally suited for cloning of short recombination site containing fragments. The names of these plasmids are highlighted in gray in the Table 2.5.

2.7.2.1 *In vivo* recombination substrate plasmids

In vivo recombination substrate plasmids were based on the pMS183Δ backbone. pMS183Δ is derived from pMS183 by a 16 bp deletion of a NdeI/NdeI region near one of its polylinker regions that improves plasmid stability when cloning (nearly) palindromic sites. pMS183Δ caries a kanamycin resistance antibiotic marker and it has a pSC101 origin of replication. To make the *in vivo* recombination substrate plasmids the 3076 bp BglII/NcoI and 1732 bp BsrGI/XbaI of pMS183Δ were joined in a four-piece ligation to recombination site-containing XbaI/NcoI and BamHI/Asp718 fragments, cut out from a pMTL23-based single site plasmid (Fig. 2.8). This ligation is possible as the BamHI and

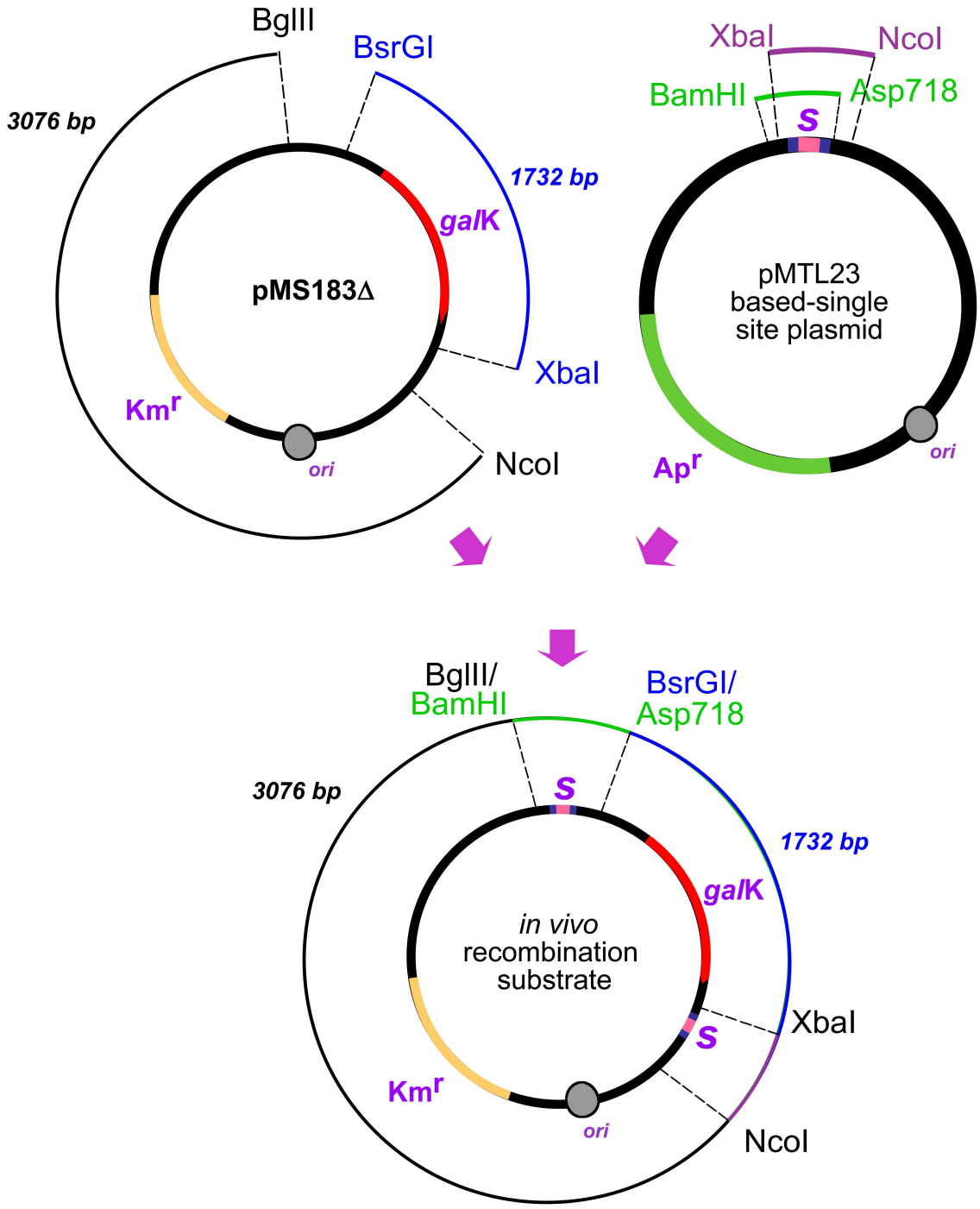


Figure 2.8- General schematic illustrating the cloning steps needed to generate *in vivo* recombination substrate plasmids. To make the in vivo recombination substrate plasmids the 3076 bp BglII/NcoI (black segment) and 1732 bp BsrGI/XbaI of pMS183Δ (blue segment) were joined in a four-piece ligation to recombination site-containing XbaI/NcoI (purple segment) and BamHI/Asp718 (green segment) fragments, cut out from a pMTL23-based single site plasmid. This ligation is possible as the BamHI and BglII, and Asp718 and BsrGI pairs of restriction enzymes make cuts that produce compatible ends. Recombination site is represented by a pink rectangle, flanked by two smaller blue rectangles with a letter "S" above it. The drawing is not to scale.

BglII, and Asp718 and BsrGI pairs of restriction enzymes make cuts that produce compatible ends. The names of the *in vivo* recombination substrate plasmids are highlighted in red in Table 2.5.

2.7.2.2 *In vitro* recombination substrate plasmids

In vitro recombination substrate plasmids were constructed by a three-piece ligation of ~1.7 kb AlwNI/BglII and ~900 bp AlwNI/BamHI recombination site-containing fragments from a pMTL23-based single site plasmid, and a 1264 bp kanamycin resistance-encoding BamHI fragment from pUC71K (Fig. 2.9). Since the BamHI fragment can be cloned in two ways, as BglII and BamHI generated ends are compatible, its orientation was determined by a NruI digest. The BamHI fragment orientation that produces NruI fragments that are ~2.7 kb and 1.1 kb in size instead of ~3.5 kb and 350 bp has been arbitrarily chosen as the correct one. Having all *in vitro* plasmids with a uniform BamHI fragment orientation simplifies the analysis of restricted resolution products after the *in vitro* recombination reaction. The names of the *in vivo* recombination substrate plasmids are highlighted in purple in Table 2.5.

2.8 Preparation of competent E. coli cells

Competent *E. coli* cells for DNA transformations were routinely prepared using the $CaCl_2$ method (see Section 2.8.1). When very high transformation efficiency was required, e.g. in library construction, electro-competent *E. coli* cells were prepared (see Section 2.8.2.) or acquired from Invitrogen.

2.8.1 Chemically competent cells

5 ml of L-broth (with or without antibiotic, as appropriate) was inoculated with 10 µl of an overnight *E. coli* culture and grown at 37 °C with shaking (250 rpm in New Brunswick Scientific Excella E24 Incubator Shaker series) until OD_{600} of 0.4-0.5 was reached. The culture was then chilled on ice and divided into four 1.2 ml aliquots. Cells were harvested by centrifugation in a pre-cooled rotor (9300 g, 1 minute, 4 °C) and all supernatant was removed. Cell pellets were gently resuspended in 1 ml of ice-cold 50 mM $CaCl_2$, and the centrifugation was repeated. After the second centrifugation step, all of the supernatant was removed, and each cell pellet was gently resuspended in 200 µl of ice-cold 50 mM $CaCl_2$. The competent cells thus prepared were then stored on ice until transformation; they maintained their competence for up to 48 hours.

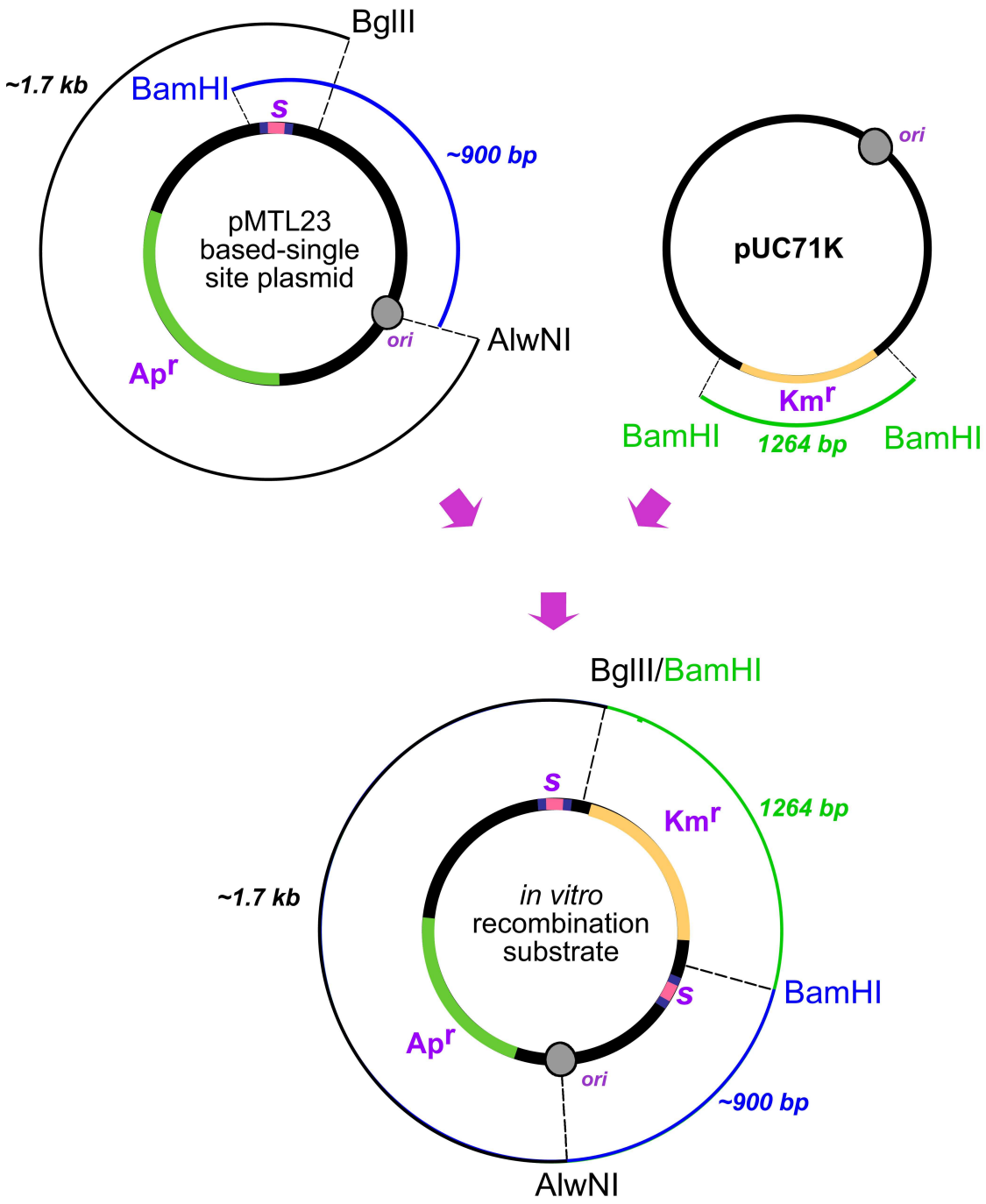


Figure 2.9- General schematic illustrating the cloning steps needed to generate in vitro recombination substrate plasmids. In vitro recombination substrate plasmids were constructed by a three-piece ligation of ~1.7 kb AlwNI/BglII (black segment) and ~900 bp AlwNI/BamHI (blue segment) recombination site-containing fragments from a pMTL23-based single site plasmid, and a 1264 bp kanamycin resistance-encoding BamHI fragment from pUC71K (green segment). This ligation is possible as the BamHI and BglII restriction enzymes make cuts that produce compatible ends. As BamHI fragment can be cloned in two ways, its orientation was determined by a subsequent NruI digest. The BamHI fragment orientation that produces NruI fragments that are ~2.7 kb and 1.1 kb in size was arbitrarily chosen as the 'correct' one. Recombination site is represented by a pink rectangle, flanked by two smaller blue rectangles with a letter "S" above it. The drawing is not to scale.

2.8.2 Electro-competent cells

500 ml of-broth (with or without antibiotic, as appropriate) was inoculated with 5 ml of an overnight *E. coli* culture and grown at 37 °C with shaking (250 rpm in New Brunswick Scientific Excella E24 Incubator Shaker series) until OD_{600} of 0.4-0.5 was reached. Cells were then cooled down to 4 °C and harvested by centrifugation (Beckman Coulter JA-14, 5000 rpm, 4 °C, 15 min). All of the supernatant was removed, the cell pellet was resuspended in 500 ml ice-cold 10% glycerol, and centrifugation was repeated. The resulting cell pellet was then resuspended in 250 ml ice-cold 10% glycerol and centrifuged again. The pellet was resuspended in 25 ml of ice-cold 10% glycerol, transferred to a 30 ml polypropylene tube and centrifuged (Beckman Coulter J-20, 5000 rpm, 4°C, 15 min). The final pellet was resuspended in a final volume of 0.5 ml of ice-cold 10% glycerol. Electro-competent cells thus prepared were divided into 20 µl aliquots and either used for transformation straight away (giving the highest transformation efficiency) or frozen in liquid nitrogen and stored at -70°C.

Alternatively, library quality electro-competent *E. coli* DH5α cells were acquired from Invitrogen.

2.9 Transformation of E. coli cells

Chemically competent *E. coli* cells were transformed as follows: 35-50 µl of cells were mixed with 0.01-0.1 µg of plasmid DNA and incubated on ice for 20 minutes. The cells were 'heat shocked' by incubation at 42 °C for 2 minutes and returned to ice for a further 5 minutes. After the 'heat shock' stage, 1 ml of 2×YT broth was added to the transformation mix. The cells were then incubated at 37 °C for 60-90 minutes ('recovery' stage). After the 'recovery' stage, aliquots of the liquid culture were spread on selective agar plates, which were then incubated at 37 °C overnight.

Electro-competent *E. coli* cells were transformed using the Biorad micropulser. A 20 µl aliquot of cells was added to ~0.01 µg of de-salted DNA on ice and transferred to an ice-cold electroporation cuvette. The transformation mix was tapped to the bottom of the cuvette before it was placed into the slide of the micropulser. The electrical pulse was delivered and 1 ml of 2×YT broth was immediately added to the cells. The resulting liquid culture was transferred into 20 ml polycarbonate tubes and incubated at 37 °C with shaking (250 rpm in New Brunswick Scientific Excella E24 Incubator Shaker series) for

60 minutes ('recovery' stage). After the 'recovery' stage, aliquots of the liquid culture were spread on selective agar plates, which were then incubated at 37 °C overnight.

Commercially supplied electro-competent *E. coli* DH5α cells were transformed by a similar procedure, according to the manufacturer's instructions.

2.10 Preparation of plasmid DNA

Plasmid DNA for most applications was routinely prepared using the small scale method (see Section 2.10.1). If the plasmid DNA was to be used for library selections or *in vitro* reactions, one of the two large scale methods (see Section 2.10.2) was used.

2.10.1 Small-scale plasmid DNA preparation

Plasmid mini-preps were made using the Qiagen mini-prep kit (Cat. No. 27106). For each column, 4.5 ml of overnight culture was used. If the plasmid DNA was prepared from cells grown on plates and then scraped off, a pellet weighing ~0.015g was used, for each column. In both cases, plasmid DNA was prepared using the Qiagen mini-prep kit according to the manufacturer's instructions (Cat. No. 27106). This kit uses a silica-gel membrane system to bind DNA, allowing simple and rapid wash steps to be performed. The purified plasmid DNA was eluted in 50 µl of double-distilled H_2O.

2.10.2 Large-scale plasmid DNA preparation

Plasmid midi-preps were made using the Qiagen midi-prep kit (Cat. No. 12143). Plasmid DNA from cells grown in liquid cultures was purified from 25 ml of overnight culture per column. If the plasmid DNA was prepared from cells grown on plates and then scraped, a pellet weighing ~0.33 g was used, per column. In both cases plasmid DNA was prepared using the Qiagen midi-prep kit according to manufacturer's instructions (Cat. No. 12143). This kit uses a silica-gel beads column system to bind DNA, which allows simple and rapid wash steps to be performed. The purified plasmid DNA was eluted in 400 µl of TE/10 buffer (10 mM Tris-HCl pH 8.2, 0.1 mM EDTA).

Plasmid DNA that was to be used in resolvase *in vitro* reactions was prepared using a method adapted from that used by Birnboim and Doly (1979). 200 ml of overnight *E. coli* culture was grown to stationary phase (~12 h, shaking at 250 rpm, at 37 °C). Cells were harvested by centrifugation (12 000g, 4°C, 5 min.). The cell pellet was resuspended in Doly I buffer (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA), and kept on ice for 5 minutes. 8 ml of freshly made ice-cold Doly II buffer (200 mM NaOH, 1% SDS) was added, and the samples were mixed and kept at 0°C for 4 minutes. 6 ml of ice-cold Doly

III buffer (0.6 vol. 5 M KOAc, 0.115 vol. AcOH, 0.285 vol. H_2O) was added and the sample was mixed and kept on ice. The cell debris was then removed by centrifugation (39 000 g, 4°C, 30 min.) and the supernatant, containing the DNA, was mixed with 12 ml of isopropanol to allow precipitation at room temperature for 15 minutes. The DNA was spun down by centrifugation (27 200 g, 20°C, 30 min.) and the pellet was washed with 70% ethanol, dried and resuspended in 2 ml of TE buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA). Caesium gradient purification was then carried out by adding 4.324 ml aqueous CsCl (equivalent to a solution of 5 g of CsCl with 3 ml of H_2O) to the DNA/TE sample and transferring the mixture into a Beckman ultra-centrifuge tube. 270 µl of 15 mg/ml ethidium bromide was added, the remaining volume was filled with liquid paraffin and heat-sealed. The sealed tube was centrifuged in a Beckman Ti70 fixed-angle rotor (200 000 g, 16 h, at 15°C).

Ethidium bromide-stained DNA bands were visualised using a long-wave UV source (365 nm). Out of two visible bands, the upper band contained chromosomal and nicked plasmid DNA, while the lower, more intense band contained supercoiled plasmid DNA. The lower band was removed using a 1 ml syringe. Ethidium bromide was removed from the DNA solution by a number of consecutive n-butanol extractions. The DNA was recovered by adding 3 volumes of water and 2 volumes of ethanol, followed by mixing and incubating it at 4°C for 30 minutes. Precipitated DNA was pelleted by centrifugation (27 200 g, 4°C, 30 min.). The pellet was then washed with 70% ethanol, dried and resuspended in 500 ml TE buffer. The plasmid DNA was stored at -20°C.

2.11 Ethanol precipitation of DNA

The salt concentration of the DNA solution was adjusted to 0.3 M with ammonium acetate, after which 2 volumes of 100% ethanol were added. The sample was thoroughly mixed and incubated at -20°C or -70°C for ≥ 15 min. The precipitated DNA was pelleted by centrifugation in Eppendorf microcentrifuge (14 000 rpm, 4°C, 30 min.). The resulting pellet was washed with 70% ethanol and centrifuged for a further 5 min (14 000 rpm, 4°C). The ethanol was removed, the DNA pellet was dried in a Gyrovap rotary vacuum drier and resuspended in H_2O or TE buffer.

2.12 Restriction endonuclease digestion of DNA

Restriction digests were performed in the supplier's recommended buffer, using between 2-10 units of restriction endonuclease per microgram of DNA, ensuring complete cleavage of DNA. Digests were routinely incubated at 37°C (unless the manufacturer's instructions

suggested otherwise) for ≥ 1h and were stopped by the addition of SDS loading buffer (see section 1.14). Restriction digests of resolvase *in vitro* reaction products were done in a similar fashion. After the resolvase reaction was stopped by heating it to ≥ 70°C for 5 minutes, the buffer conditions were adjusted appropriately and restriction endonuclease was added and incubated for a further 1 h at 37°C. The reaction was then terminated by the addition of SDS loading buffer (Section 2.14).

2.13 Gel electrophoresis

Depending on the size of DNA fragments that were to be separated, one of two gel systems was used. For large supercoiled or linear fragments, separation was performed on agarose gels, while small (<200 bp) fragments were separated using polyacrylamide gels. Polyacrylamide gels were also used for purifying commercially acquired oligonucleotides (see section 2.18) and for separating proteins (see section 2.26)

2.13.1 Agarose gel electrophoresis

For routine separation of DNA fragments resulting from restriction digests (200-4000 bp), 1.0-1.2% agarose gels were used. These gels were made by dissolving an appropriate amount of agarose (Biorad, 'Ultrapure') in 100 ml of 1×MRT buffer (40 mM Tris-acetate pH 8.2, 1mM EDTA) by mixing and heating in a microwave oven. Hot agarose solution was cooled to ~60°C before it was poured into a gel former fitted with an appropriate comb, and allowed to set at room temperature. Gels were run at room temperature for ~1.5 h at 150 V.

Large 1.2% agarose gels were used for separating supercoiled plasmids isolated from *E. coli* cells after *in vivo* resolution assays. These gels were prepared in similar fashion to the gels described above. The appropriate amount of agarose was dissolved in 300 ml of 1×TAE buffer (40 mM Tris-acetate pH 8.2, 20 mM sodium acetate, 1 mM Na_2EDTA). Gels were run in tanks requiring about 3 litres of 1×TAE buffer at an applied voltage of between 2 and 5 V/cm. Gels were run at room temperature for 4 h at 100V. Identical gels were used for separation of reaction products of *in vitro* cleavage and resolution reactions; however these gels were run at 35 V for 16 h.

Large 0.7% agarose gels were used to separate supercoiled plasmid multimers required for selection of crossover sites from random libraries (Section 3.7).

2.13.2 Polyacrylamide gel electrophoresis

To separate smaller (<200 bp) DNA fragments 12% non-denaturing polyacrylamide gel electrophoresis was used. The gels were prepared by making 30 ml of the appropriate gel solution (12 ml of 30% acrylamide solution (37.5:1 acrylamide:bisacrylamide), 3 ml 10× TBE buffer (890 mM Tris-HCl, 890 mM boric acid, 2 mM EDTA pH 8.3), 15 ml H_2O, 360 µl 10% APS w/v, and 18 µl TEMED (N, N, N', N'-tetramethylethylenediamine). The freshly made acrylamide gel mixture was poured between glass plates that were clamped together with 0.75 mm spacers and sealed with a length of rubber tubing. Immediately after pouring the gel mixture, a well forming comb was fitted and the gel was allowed to polymerise for ~1 h. After polymerisation was complete, the clamps, tubing and comb were removed and the gel was clamped into the vertical electrophoresis kit. The running buffer used was 1× TBE (89 mM Tris-HCl, 89 mM boric acid, 0.2 mM EDTA pH 8.3) and electrophoresis was at a constant 220 V at room temperature for 2.5-3 h. The DNA bands were visualised by ethidium bromide staining (Section 2.16).

2.14 Loading buffers

SDS loading buffer was added to restriction digestions in a 1:5 ratio, prior to loading on an agarose or polyacrylamide gel (50% glycerol, 1% SDS, 0.01% bromophenol blue). SDS/K loading buffer (SDS loading buffer with protease K, 1 mg/ml final concentration) was added to *in vitro* resolution or cleavage reactions prior to agarose gel electrophoresis to ensure that resolvase protein, covalently attached to DNA, is fully digested, so as not to result in a bandshift and obscure the information on the gel. Formamide loading buffer (80% deionised formamide, 10 mM EDTA pH 8.0, 1 mg/ml xylene cyanol, 1 mg/ml bromophenol blue) was used in a 1:1 ratio with oligonucleotide samples prior to denaturing polyacrylamide gel electrophoresis. The samples were heated to 80 °C and then held at this temperature for 5 minutes before being loaded onto the gel.

2.15 DNA molecular weight standards

Invitrogen and NEB 1kb, $1kb^+$, 100 bp ladders (whichever most appropriate) were run on both agarose and polyacrylamide gels to help estimate the size of DNA fragments.

2.16 Ethidium bromide staining of DNA and photography

To visualise the DNA within agarose or polyacrylamide gels, the gels were first stained with 0.6 µg/ml ethidium bromide solution (made by mixing a 15 mg/ml ethidium bromide stock solution with 1 × gel running buffer) for 30-60 minutes. The gels were then thoroughly rinsed and soaked in deionised water for a further 60 minutes in order to remove the excess background ethidium bromide fluorescence. To visualise the ethidium

bromide-stained DNA short wavelength (254 nm) UV illumination was used. In the case of preparative gels from which the DNA was to be extracted, a long wavelength (365 nm) UV source was used.

To photograph the gels, Polaroid type 667 film was used.

2.17 Extraction of DNA from gel fragments

Extracting DNA from agarose gels was done as follows: a band of interest was excised using a scalpel from an ethidium bromide-stained agarose gel, situated on a UV long-wave (365 nm) transilluminator. The DNA was then extracted out of the gel chip using the Qiagen gel extraction kit (Cat No. 28704) as per the manufacturer's instructions. The DNA was eluted in 30-50 µl of H_2O.

Double-stranded DNA molecules shorter than 200 bp were isolated on a non-denaturing 12% polyacrylamide gel. The gel was stained with ethidium bromide (see section 2.16), placed on a UV long-wave (365 nm) transilluminator and the band of interest excised using a scalpel. The gel chip was placed into 1.5 ml eppendorf tube, and crushed using a glass rod, and 750 µl of H_2O was added. The tube containing gel material was incubated at 37°C overnight with vigorous shaking (1000 rpm, Eppendorf Thermomixer compact). Polyacrylamide was removed by centrifugation through a Costar Spin-X 0.22 µm filter (9300 g, 10 min). The filtrate was then dried down in a Gyrovap rotary vacuum drier and resuspended in 10 µl H_2O or TE buffer.

2.18 Purification of oligonucleotides by denaturing PAGE

Comercially synthesised oligonucleotides were purified from prematurely terminated contaminant species using a standard polyacrylamide gel kit (see section 2.13.2). Synthetic single-stranded oligonucleotides were run on 15% polyacrylamide/7M urea denaturing gels (15.75 g urea, 14.75 ml of 40% w/v acrylamide solution (19 : 1 ratio acrylamide : bisacrylamide), 3.75 ml of 10 × TBE, 7.75 ml H_2O, 225 µl of 10% APS w/v, 18 µl TEMED) . The running buffer was 1 × TBE. Samples containing ~2 nmol crude oligonucleotide mix and 1 vol. formamide loading buffer (Section 2.14) were heated to 80 °C and then held at this temperature for 5 minutes prior to loading onto a gel. Electrophoresis was carried out at constant voltage (400-500V) for 2-3 h to ensure that the gel remains hot, preventing the renaturation of single-stranded oligonucleotides.

To visualise the oligonucleotides the gel was stained with cationic carbocyanine 'Stains-all' dye (1-ethyl-2-[3-(1-ethylnaphtho[1,2-d]thiazolin-2-ylidene)-2-

methylpropenyl)naphtho[1,2-d]thiazolium bromide; supplied by Aldrich). This dye can be used to stain DNA, RNA or protein. The oligonucleotide-containing gel was soaked in a staining solution (70 ml of H_2O, 20 ml isopropanol, 10 ml of 0.1% w/v solution of 'Stains-all' in formamide) for ~10 minutes until sufficient staining was observed. The gel was then destained by rinsing several times in water and the slowest migrating, full-size oligonucleotide band was excised using a scalpel. The polyacrylamide gel chip was then crushed and DNA extracted as detailed in section 2.17.

2.19 Annealing oligonucleotides

Pairs of complementary single-stranded oligonucleotides were annealed in 20 µl of annealing buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 50 mM NaCl). The sample was thoroughly mixed, heated to 95°C for 5 minutes and than allowed to gradually cool/anneal at room temperature for ≥ 4 h.

2.20 Ligation of DNA restriction fragments

For cloning, ~1 µg of DNA fragments were ligated in 1× Invitrogen ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM DTT, 5% w/v polyethylene glycol-8 000), using 1 unit of T4 DNA ligase (Invitrogen), in a final volume of 10 µl. The molar ratio of vector to insert was typically 1:3. Ligation reactions were incubated at room temperature overnight and then either used to transform competent *E. coli* cells (Section 2.9) or stored at -20°C until required.

2.21 Estimating DNA concentration by UV spectrophotometry

To estimate the DNA concentration, the absorbance at 260 nm of diluted DNA samples was measured using a Lambda 45 UV/visible spectrophotometer (Perkin Elmer). The concentrations of measured samples were calculated using the approximation that a solution with the absorbance of 1 at 260 nm contains 50 µg/ml double-stranded DNA.

2.22 Sequencing plasmid DNA

DNA samples were sequenced commercially by MWG Biotech AG (Ebersberg,Germany).

2.23 In vivo recombination reactions – "MacConkey assay"

To check for *in vivo* recombination activity, the "MacConkey assay", originally developed by D. Blake was used (Blake, 1993). DS941 (*gal*K$^-$) *E. coli* cells containing a low level resolvase expression plasmid (Section 2.7.1.1) were made competent (Section 2.8),

routinely using the $CaCl_2$ method, and transformed (Section 2.9) with ~0.01 μg of *in vivo* recombination substrate plasmid, containing a kanamycin resistance gene marker and two recombination sites in direct repeat, flanking a *gal*K gene (Section 2.7.2.2). After transformation cells were plated out on selective MacConkey agar (Difco), containing 2% galactose, and incubated overnight at 37°C.

After incubation the plates were photographed using a Fuji Finepix F700 digital camera. A representative sample of cells was scraped from the MacConkey plates and grown in liquid culture, overnight, in the presence of selective antibiotics. Plasmid DNA was isolated from pelleted cells (Section 2.10.1) and analysed by agarose gel electrophoresis (Section 2.13.1).

In the event of successful recombination of the substrate plasmid by the resolvase, the substrate plasmid is resolved into two circular DNA molecules, one containing the kanamycin resistance gene and the origin of replication, and the other carrying the *gal*K gene (Fig. 2.10). As the DNA circle harbouring the *gal*K does not contain an origin of replication it can not be stably maintained and is lost on subsequent cell divisions. MacConkey agar contains a pH indicator 2-methyl-3-amino-6-dimethylaminoaphenazine, which at <pH 6.8 is red and at >pH 8.0 is yellow. MacConkey agar also contains no carbon source other than the added galactose. Therefore, if the resolvase is unsuccessful in resolving the substrate plasmid the *gal*K gene will remain and allow the DS941 cells to metabolise the galactose which results in the colony turning red. In the event of successful resolution of the test plasmid and the resulting loss of the *gal*K gene, the DS941 are forced to metabolise the amino acids in the agar causing an increase in the pH, resulting in the colony turning white (Fig. 2.10).

2.24 In vivo binding reactions

In vivo binding was assayed using the *in vivo* binding test plasmid pSB404, kindly provided by S. Rowland (Fig. 2.11). pSB404 is a pDB34-based plasmid with a Sin site I partially overlapping the *gal*K gene promoter. DS941 (*gal*K^-) *E. coli* cells containing pSB404 were made competent, using $CaCl_2$ method (Section 2.8), and transformed (Section 2.9) with ~0.01 μg of a resolvase expression plasmid. After transformations cells were plated out on selective MacConkey Galactose Agar and incubated overnight at 37°C. After incubation the plates were photographed using a Fuji Finepix F700 digital camera.

If a resolvase binds the Sin site I tightly, it prevents access to the *gal*K gene promoter and the expression of this gene is prevented. In this case, since DS941 cells do not produce *gal*K of their own, they are unable to metabolise galactose in the MacConkey agar and are

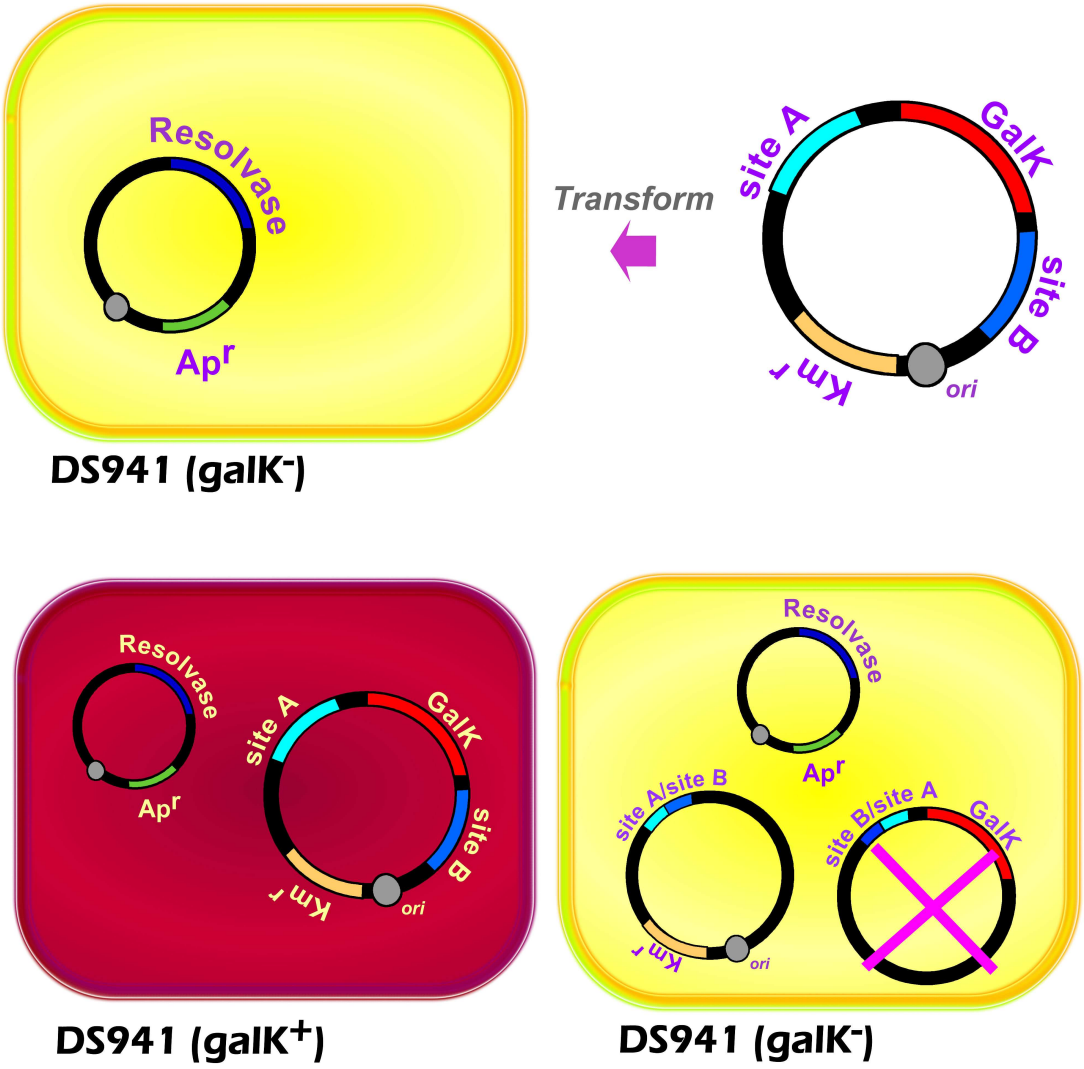


Figure 2.10- Cartoon illustrating the MacConkey assay. In this assay *in vivo* substrate plasmid (top right), carrying the *gal*K gene flanked by two recombination sites (site A & site B), is used to transform competent DS941 cells (*gal*K$^-$) already containing resolvase expression plasmid (top left). Post transformation cells are plated onto selective MacConkey agar plates. In the event of successful resolution (bottom right), *gal*K gene ends up on a circular molecule without an origin of replication (gray sphere marked "ori"), and is therefore lost. The cells are then unable to metabolise galactose and turn white. If the resolution is unsuccessful (bottom left), *in vivo* substrate plasmid is intact, *gal*K gene is expressed and the cells are able to metabolise galactose and turn red, see text for details (Section 2.23). Rounded coloured rectangles represent DS941 cells. The drawing is not to scale.

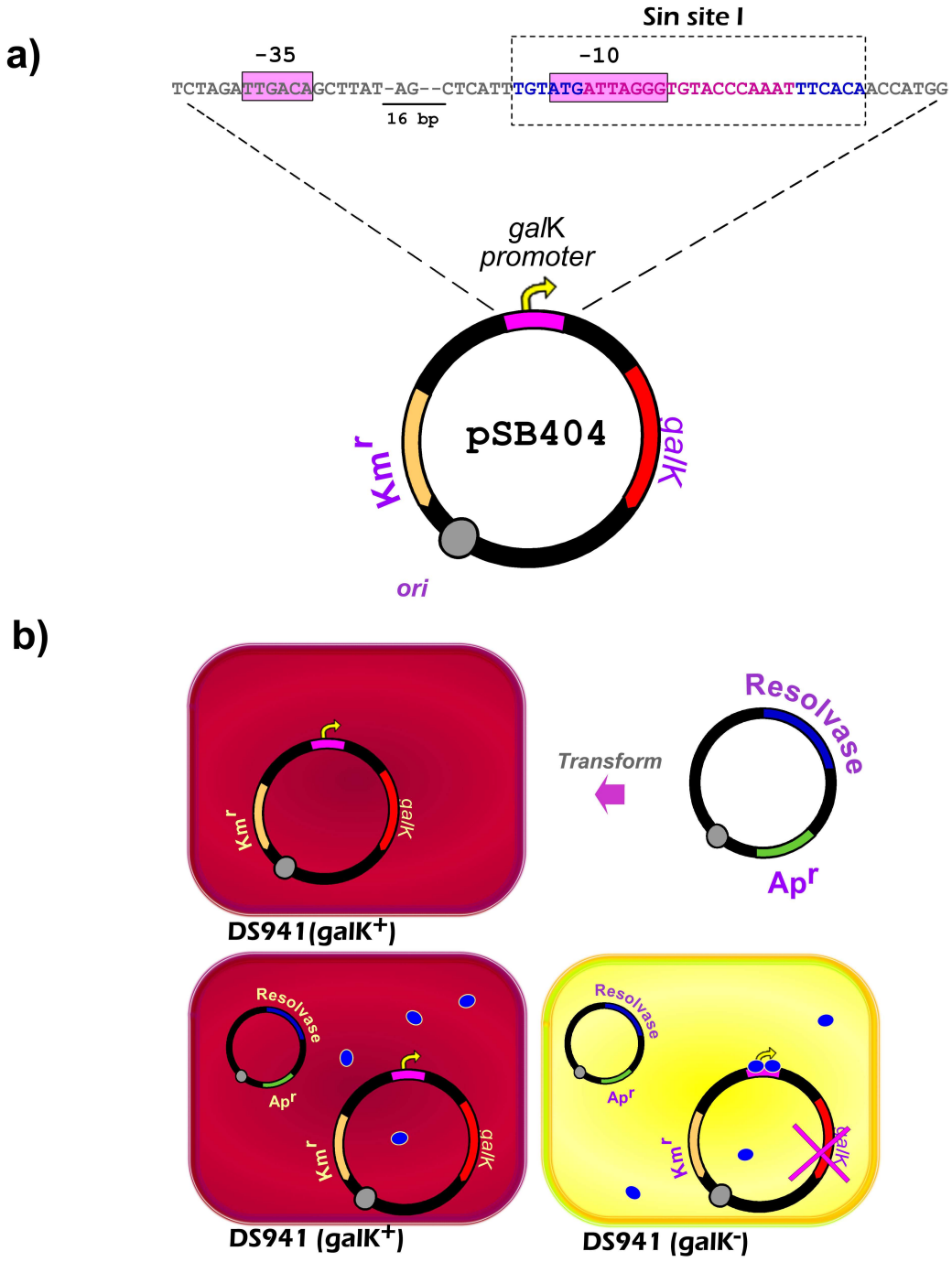


Figure 2.11- Cartoon illustrating the *in vivo* binding assay. a) A schematic of a pSB404 plasmid. The sequence of Sin site I (sequence encircled by a dashed line box) overlapping the galK gene promoter is shown expanded at the top of the plasmid diagram. Pink boxes mark the -35 and -10 promoter elements. The central 18 bp of the Sin site I sequence are coloured pink while the flanking 6 bp on each side that are bound by the Sin resolvase HtH DNA binding domain are in blue. In the plasmid diagram yellow arrow signifies the promoter while the pink segment represents the Sin site I. b) In the *in vivo* binding assay resolvase expressing plasmid (top right), is used to transform competent DS941 cells (*gal*K$^-$) already containing in vivo binding plasmid pSB404 (top left). Post transformation cells are plated onto selective MacConkey agar plates. If the resolvase protein (blue ovals) binds the Sin site I tightly (bottom it occludes the *gal*K gene promoter, this gene is then in turn expressed, the cells remain (*gal*K$^-$) and are therefore white. If the resolvase protein does not bind the Sin site I *gal*K gene promoter is accessible to the RNA polymerase, galK is expressed and cells turn red, see text for details (Section 2.24). Rounded coloured rectangles represent DS941 cells. The drawing is not to scale.

forced to resort to metabolising amino acids in the medium causing an increase in the pH. Increased pH results in the colonies turning white, due to the presence of an indicator (see previous Section). If a resolvase tested fails to bind to the Sin site I, or if the binding is only transient, transcription machinery is able to access the galK gene promoter and this gene is expressed. The DS941 cells expressing the galK gene are able to metabolise galactose in the MacConkey agar causing a decrease in the pH, resulting in the colonies turning red.

2.25 Purification of resolvase proteins

Resolvase and Z-resolvase proteins were purified using essentially the same purification procedure, which was adapted from the previous study (Olorunniji, 2006). In the following sections (2.25.1-2.27), the term resolvase refers to both resolvase and Z-resolvase proteins unless specifically stated otherwise.

2.25.1 Large scale induction of resolvase mutants

E. coli BL21 (DE3) pLysS strain (Studier *et al.*, 1990) was made chemically competent (see section 2.8.1) and transformed with a resolvase over-expression plasmid (Table 2.5). Transformed cells were plated on selective agar containing kanamycin and chloramphenicol and grown overnight at 37°C. Single bacterial colonies were picked and used to inoculate 20 ml starter cultures (L-broth containing kanamycin and chloramphenicol), which were incubated overnight at 37°C. The following day, 8 ml of starter culture were used to inoculate 800 ml of preheated (37°C), selective L-broth and the cells were grown in a flat-bed shaker incubator, at 37°C, 250 rpm to an O.D of 0.4. At this point, (Z-)resolvase expression was induced by adding 0.1 mM IPTG, and the cultures were grown for further 2.5 hours. Cells were harvested by centrifugation (Beckman Coulter JA-14, 10 000 rpm, 4°C, for 10 min.).

2.25.2 Buffers used in the purification of resolvase mutants

Buffers used in the purification of resolvase and Z-resolvase proteins are listed in Table 2.6.

2.25.3 Extraction and purification of resolvase mutants

The protein purification procedure was adapted from Olorunniji (2006). The procedure is based on utilising the difference in solubility between resolvase and most other proteins. Unlike most other proteins, resolvase is insoluble at low salt concentrations but soluble at high salt concentrations (2 M NaCl). The extraction procedure comprised a series of low salt (100 mM NaCl) wash steps designed to remove as much of the soluble protein

PROTEIN PURIFICATION BUFFERS	
Suspension Buffer	20mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 1.0 mM DTT, 0.1 mM EDTA, 1.2 mM PMSF, 1% ethanol
Wash Buffer	20mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 1mM DTT, 0.1 mM EDTA, 1.2 mM PMSF, 1% ethanol, 100 mM NaCl
Solubilisation Buffer	20mM Tris-HCl pH7.5, 1.0 mM DTT, 0.1 mM EDTA, 1.0 mM PMSF, 1% ethanol, 6 M urea, 1.0 M NaCl
Refolding Buffer I	20mM Tris-HCl pH7.5, 1.0 mM DTT, 0.1 mM EDTA, 1.0 mM PMSF, 1% ethanol, 1.0 M NaCl
Precipitation Buffer	20mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 1.0 mM DTT, 0.1 mM EDTA, 1.2 mM PMSF, 1% ethanol
Buffer A	50 mM sodium phosphate pH 7.2, 25 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT, 6M Urea
Buffer B	50 mM sodium phosphate pH 7.2, 1 M NaCl, 0.1 mM EDTA, 0.1 mM DTT, 6M Urea
Refolding Buffer II	20mM Tris-HCl pH7.5, 1.0 mM DTT, 0.1 mM EDTA, 2.0 M NaCl
Refolding Buffer Z	20mM Tris-HCl pH7.5, 1.0 mM DTT, 1.0 M NaCl, 100mM $ZnSO_4$
Resolvase Dilution Buffer (RDB)	20mM Tris-HCl pH7.5, 1.0 mM DTT, 0.1 mM EDTA, 1.0 M NaCl, 50% glycerol

Table 2.6- Buffers used in the resolvase protein purification (Section 2.25).

contaminants as possible while keeping resolvase itself insoluble, followed by solubilisation of resolvase protein in a denatured state (buffer containing 6 M urea, 1 M NaCl). After this resolvase protein was renatured by removing urea by dialysing it out, leaving resolvase in 1 M NaCl.

After the induction of resolvase expression, the cells from 800 ml cultures were harvested by centrifugation (Beckman Coulter JA-14, 10 000 rpm, 4°C, for 10 min.). The resulting cell pellets weighed approximately 2 g. Cell pellets were resuspended in 25 ml Suspension buffer and sonicated in three consecutive 20 second bursts, using a Vibra-cell VC100 sonicator with a micro-probe at 40% amplitude. To prevent overheating, between each sonication burst, the cells/suspension buffer mixture was cooled to 4°C in an ice bath. The resulting crude resolvase extract was centrifuged (Beckman Coulter J-20, 20 000 rpm, 4°C, 15 min) and the supernatant containing soluble protein fraction and general cell debris was removed. The resolvase-containing pellet was resuspended in 20 ml of Wash Buffer, homogenised (~40 strokes) using a Dounce homogeniser and then re-centrifuged (Beckman Coulter J-20, 20 000 rpm, 4°C, 15 min). This wash step was repeated 4 times in order to maximise the removal of contaminating soluble proteins. After the wash steps, the resolvase-containing pellet was resolubilised in 20 ml Solubilisation Buffer in the Dounce homogeniser by vigorous homogenisation (three rounds of ~50 strokes each). Insoluble debris was removed by centrifugation (Beckman Coulter J-20, 20 000 rpm, 4°C, 15 min) and the supernatant containing solubilised denatured resolvase dialysed against 1 litre of Refolding Buffer I for 5 h at 4°C. This dialysis step removes the urea and ensures that most contaminating proteins and other debris that are kept in suspension by 6 M urea now become insoluble. The dialysate was collected and centrifuged (Beckman Coulter J-20, 20 000 rpm, 4°C, 15 min) in order to remove the insoluble pellets. The resolvase-containing supernatant was carefully transferred in fresh dialysis bags and further dialysed overnight against the Precipitation Buffer. The precipitated resolvase resembling snow-like white flakes (Z-resolvase precipitate is much browner in appearance) was collected and pelleted by centrifugation (Beckman Coulter J-20, 20 000 rpm, 4°C, 15 min). The supernatant was removed and the resolvase pellet was resuspended in Buffer A.

In order to remove the remaining non-resolvase protein, nucleic acid and other contaminants, the resolvase extract in Buffer A was subjected to ion-exchange protein chromatography. Chromatography was done using a 1 ml cation exchange SP sepharose chromatography column (HiTrap SP High Performance pre-packed column, GE Healthcare). A cation exchange column is designed to bind positively charged molecules

(such as resolvase), while actively repelling negatively charged molecules (such as nucleic acids and negatively charged protein molecules). The SP sepharose column was connected to ÄKTApurifier (Amersham biosciences) and equilibrated with 10 column volumes of Buffer A, constant flow rate 1 ml/min, prior to loading the resolvase extract solubilised in Buffer A. The resolvase extract was loaded onto the column using a super-loop and absorbance values at 220 nm, 260 nm, and 280 nm were recorded. The eluate was collected as soon as the extract was loaded onto the column, in order to prevent loss of sample due to its possible failure to bind at low salt concentration. Once a stable baseline reading was obtained, a stepwise gradient of increasing salt concentration (achieved by adding increasing proportion of Buffer B to the flow) was used to elute the positively charged resolvase protein. The stepwise gradient consisted of seven, 5% salt concentration increments each lasting for 5 column volumes. After the seventh increment, the Buffer B proportion was increased from 35% to 100% and all the remaining protein was washed off the column.

The flow rate was 1 ml/min during the gradient run with 1 ml fractions collected in 2 ml Nunc tubes and stored on ice. The chromatography procedure was performed at room temperature throughout, including the samples and all the buffers used. At each salt concentration increment some resolvase was eluted from a column; however the strongest resolvase peak was always observed at around 20-25% Buffer B while for Z-resolvase this was usually at 25-30% Buffer B. A stepwise gradient was used instead of a continuous gradient in order to make the resolvase protein elute in a smaller number of highly concentrated fractions. Fractions corresponding to absorbance peaks were run on a discontinuous SDS-polyacrylamide gel to determine their purity.

After the protein elution was completed, the SP sepharose column was regenerated by washing with 10 column volumes of Buffer B, followed by 10 column volumes of buffer A. After this procedure the column was ready for loading the next sample.

Selected pure fractions from the ion exchange column were pooled together and dialysed overnight in 1 litre of Refolding buffer 2 (in the case of Z-resolvase proteins Refolding Buffer Z) containing 2 M NaCl. Keeping the salt concentration this high ensured that resolvase protein stayed in solution, while removing urea allowed the protein to renature into its native conformation. The refolded resolvase was finally dialysed against 1 litre of resolvase dilution buffer for at least 6 hours. Dialysis against this buffer facilitates the addition of glycerol which is required for sample storage at -20°C while also leading to further concentration of the resolvase sample. The concentration and purity of the

resolvase samples were estimated by comparing with standards of known concentration on a discontinuous SDS-polyacrylamide gel. A discontinuous SDS-polyacrylamide gel with purified proteins used in this study is given in Fig. 2.12.

2.26 Discontinuous SDS-polyacrylamide gel electrophoresis

To analyse resolvase protein samples a discontinuous SDS-PAGE system was used (Laemmli, 1970). This system requires running the samples through two gels of different percentage acrylamide and different ionic strength. Initially the sample is run through a stacking gel before entering the resolving gel. The resolving gel used was typically prepared from a solution with a following composition: 15% acrylamide premix (37.5:1 ratio acrylamide to bisacrylamide), 375 mM Tris-HCl (pH 8.8), 0.1% SDS, 0.05% APS, 0.05% (v/v) TEMED. This gel was overlaid with isopropanol to exclude air and allowed to polymerise for 30-45 minutes. After polymerisation, all traces of isopropanol and unpolymerised acrylamide were removed by rinsing the gel surface with water. The polymerised resolving gel was overlaid with stacking gel which was prepared as follows: 5% acrylamide premix (37.5:1 ratio acrylamide to bisacrylamide), 125 mM Tris-HCl (pH 6.8), 0.1% SDS, 0.1% APS, 0.2% (v/v) TEMED. The gel comb was placed immediately and the stacking gel was allowed to polymerise for 30-45 minutes. After polymerisation, the comb was removed and the formed wells rinsed with electrophoresis buffer (25 mM Tris base, 250 mM glycine, 0.1% SDS). Prior to loading, Laemmli loading buffer (50% glycerol, 5% SDS, 200 mM Tris-HCl pH 6.8, 0.1 mM EDTA) was added to protein samples (20% of the total volume) and the samples were boiled for 5 minutes in order to help denature the proteins and reduce disulphide bonds. Samples were loaded and polyacrylamide gels were run at 30-40 mA for 3 hours.

2.27 In vitro recombination reactions

For the purpose of *in vitro* reactions resolvase proteins were diluted with 1× resolvase dilution buffer (20 mM Tris-HCl pH7.5, 1 mM DTT, 0.1 mM EDTA, 1 M NaCl, 50% glycerol) to 200 nM. All dilutions were performed at 0°C and were not stored.

200 nM of resolvase was determined to be the optimal amount required for an efficient recombination or cleavage of 25 μg/ml supercoiled substrate DNA.

All recombination reactions were performed in a standard C8.2 buffer (50 mM Tris-HCl pH 8.2, 10 mM $MgCl_2$ and 0.1 mM EDTA). A typical reaction (40 μl) contained 0.8 μg of plasmid DNA and 4.4 μl of diluted resolvase. Reactions were carried out at 37 °C with

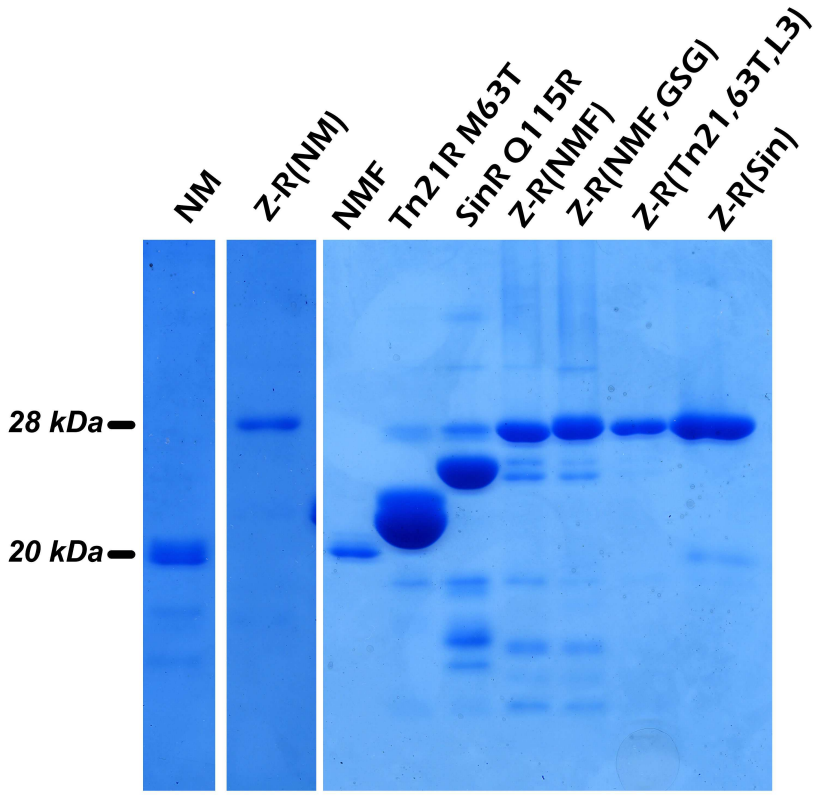


Figure 2.12- Samples of purified proteins used in this study separated on a discontinuous SDS-polyacrylamide gel (Section 2.26), stained with Coomasie blue. The most intense band in each track comes from the resolvase (or Z-resolvase) protein name of which is marked on top of the track. Sizes (Mw) of resolvase and Z-resolvases were validated by comparing to a standard size marker (lane not shown).

duration of the reaction ranging from 0.5 min. to 24 h. The reactions were terminated by heating the sample to 75°C and holding it at this temperature for 5 minutes. After the reaction was terminated it was split into two equal 22 µl aliquots. To analyse the recombination products one of these aliquots was digested with an appropriate restriction enzyme (XhoI or NruI depending on the type of the plasmid) while the other was left untreated (Fig. 2.13 and Fig. 2.14). Restriction proceeded for 45-60 minutes at 37°C. The restricted sample and the untreated sample were analysed by gel electrophoresis. Prior to loading, 5 µl of SDS/K loading buffer was added to all samples.

2.28 Random Library constructions

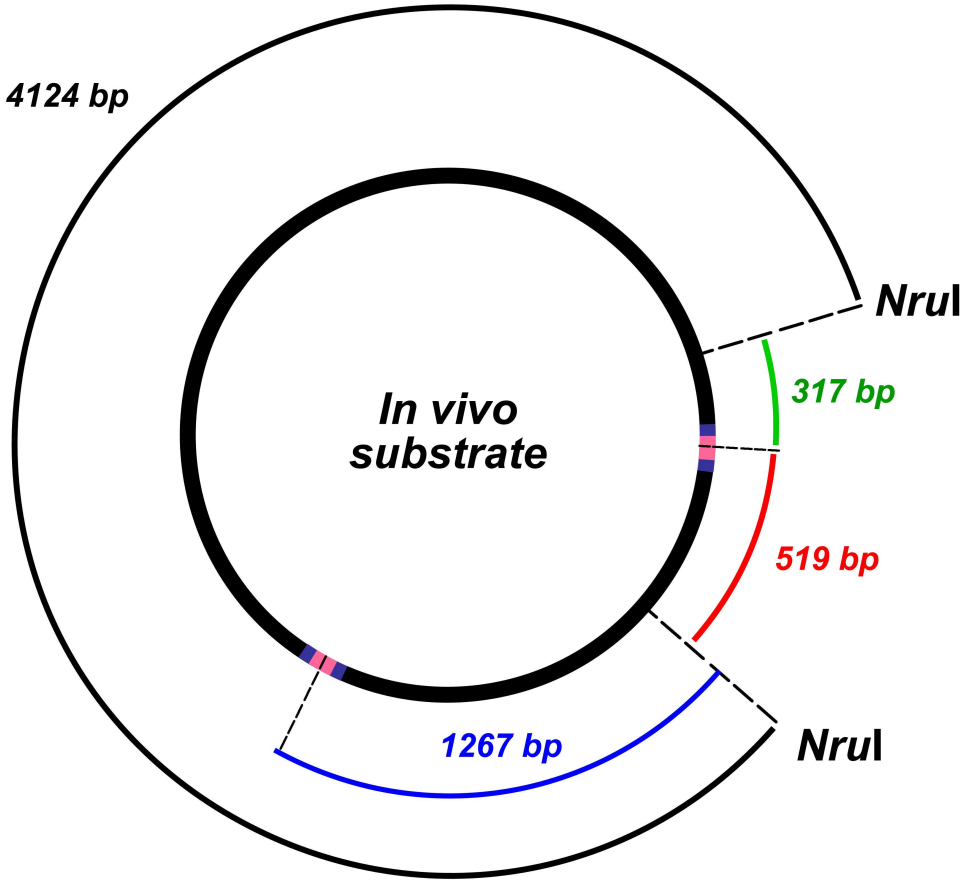
Two types of random libraries were constructed during this work, namely random libraries of recombination sites and random libraries of resolvase mutants.

Random libraries of recombination sites were made with the aim of establishing the sequence requirements in the centre of the recombination site for the resolvase proteins such as Tn3 resolvase (Chapter 3), Sin resolvase and Tn21 resolvase (Chapter 6), and also for the Tn3 Z-resolvase (Chapter 4).

Random libraries of resolvase mutants that are discussed in this work are random libraries of Tn21 resolvase made in order to select Tn21 hyperactive mutants capable of recombination without the requirement for regulatory sites.

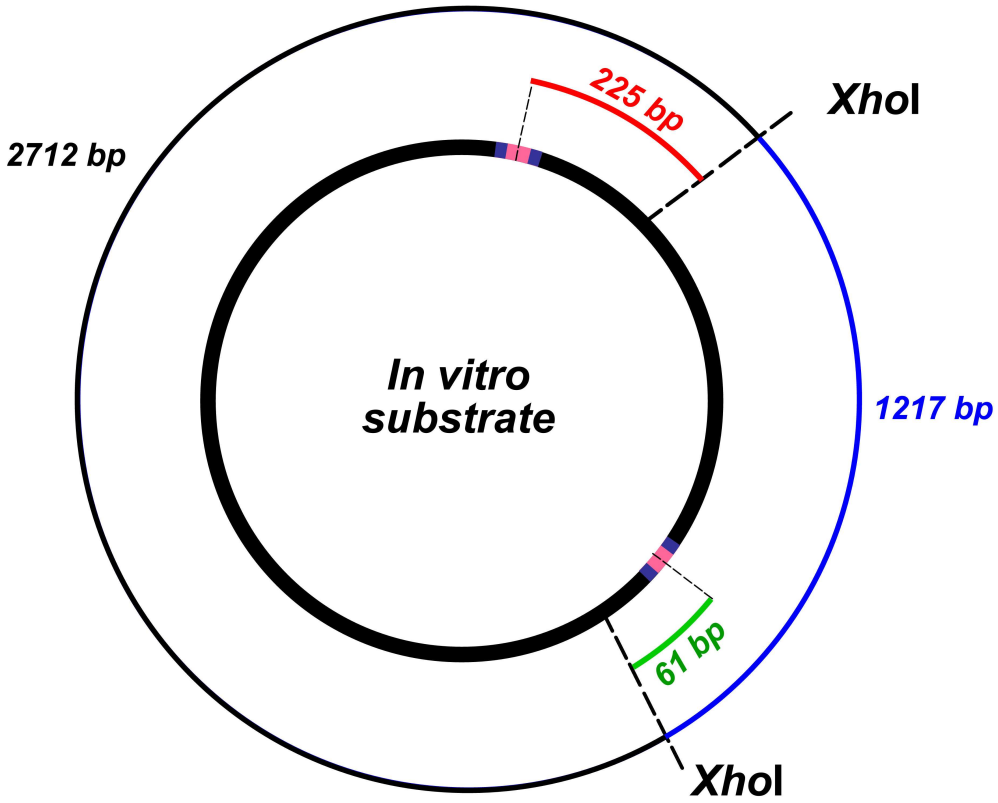
2.28.1 Random libraries of recombination sites

Thirteen separate random libraries designed to introduce differing levels of variation into recombination sites used by the Tn3 resolvase, Sin resolvase, Tn21 resolvase, and Tn3 Z-resolvase proteins were created. To make the libraries (Fig.3.4), the first step was to anneal a synthetic oligonucleotide, containing a recombination site flanked by EcoRI/SacI restriction with the desired changes, to the RSLP primer (for oligonucleotide sequences please refer to Table 2.4). In the 100 µl final volume 10 nM of oligonucleotide and RSLP primer each were mixed with the other reagents (250 µM dNTPs, 10% TPol buffer (50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100), and the reaction mix was heated to 85 °C for 5 minutes. The reaction mix was then allowed to cool down to 68°C so that RSLP primer can anneal to the synthetic oligonucleotide containing the recombination site. Following annealing, the resulting partial duplex was filled in by the *Pfu* DNA polymerase (2.5 units, 90 min at 68°C) generating a double-stranded fragment containing the recombination site with the desired random changes. This double-stranded fragment was



NruI digest	Big circle	Inversion
4124 bp 836 bp	3174 bp	3376 bp 1584 bp
Cleavage	**Small circle**	
2857 bp 1276 bp 519 bp 317 bp	1786 bp	

Figure 2.13-Schematic showing sizes of possible products of resolvase-catalysed *in vitro* recombination reaction using an *in vivo* recombination substrate (Section 3.5), following a NruI digestion. Religated recombinant products are the big and the small circle (3174 bp and 1786 bp, respectively). Inversion products would be (3376 bp and 1584 bp) while the cleavage products are 2857 bp, 1276, 519 and 317 bp. Non recombinant products (NruI digest) are 4124 bp and 836 bp. Recombination sites on the plasmid diagram are marked by a pink rectangle flanked on either side by a smaller blue rectangle.



XhoI digest	Big circle	Inversion
2712 bp 1217 bp	2552 bp	3651 bp 286 bp

Cleavage	Small circle
2487 bp 1156 bp 225 bp 61 bp	1385 bp

Figure 2.14-Schematic showing sizes of possible products of resolvase-catalysed in vitro recombination reaction using an in vitro recombination substrate (Sections 4.7 and 5.6), following a XhoI digestion. Religated recombinant products are the big and the small circle (2552 bp and 1385 bp, respectively). Inversion products would be (3651 bp and 286 bp) while the cleavage products are 2487 bp, 1156, 225 and 61 bp. Non recombinant products (XhoI digest) are 2712 bp and 1217 bp. Recombination sites on the plasmid diagram are marked by a pink rectangle flanked on either side by a smaller blue rectangle.

then restricted with EcoRI/SacI, and finally cloned into a pMTL23-based vector cut with the same restriction enzymes, creating a library of plasmids, each one containing one recombination site with the introduced random mutations (a single site or SS libraries). The list of SS libraries created along with the sequence of the recombination sites is presented in Table 2.7.

SINGLE SITE (SS) LIBRARIES	
Name	**Recombination site DNA sequence (5'-3')**
RB lib 1	TGTCTGATAATTTATAATNNNNCGAACG
RB lib 2	TGTCTGATAATTTATANNNNTTCGAACG
RB lib 3	TGTCTGATAATTTANNNNATTTCGAACG
RB lib 4	TGTCTGATAATTNNNNATATTTCGAACG
RB lib 5	TGTCTGATAANNNNTAATATTTCGAACG
RB lib 6	TGTCTGATNNNNTATAATATTTCGAACG
RB lib 7	TGTCTGNNNNTTTATAATATTTCGAACG
LIB(Z22Z)	GCGTGGGCGNNNNNNNNNNNNNNNNNNNNCGCCCACGC
LIB(Z8TTATAA8Z)	GCGTGGGCGNNNNNNNNNTTATAANNNNNNNNCGCCCACGC
LibS1	TGTGAANNNNNNNNNNNNNNNNNCATACA
LibS2	TGTGAANNNNNNGTACACNNNNNNCATACA
Lib21.1	CGTCAGNNNNNNNNNNNNNNNNNCTGATG
Lib21.2	CGTCAGNNNNNNGCATACNNNNNNCTGATG

Table 2.7-Single site (SS) libraries generated during this study.

These SS libraries were used as a precursor for making recombination substrate libraries with two sites, which were used for the site selections (Section 3.4). Two types of recombination substrate libraries were made, specifically libraries with the wild-type Tn3 site I × random site (Section 3.5), and the libraries with two identical copies of the random recombination site (Sections 3.7, 4.2 and 6.2.1).

2.28.1.1 Tn3 site I x random site library

To create recombination substrate libraries with a wild-type Tn3 site I flanking a galK gene on one side and a random Tn3 site I variant on the other (RB *gal*K Lib 1-7), precursor plasmids pMP20 and pMP21, based on pMS183Δ were made. To construct pMP20 a NheI/BsrGI restricted pMS183Δ backbone was ligated to the wild-type Tn3 site I-containing Asp718/XbaI fragment from pCO1. This ligation is possible as the Asp718 and BsrGI, and XbaI and NheI pairs of restriction enzymes make cuts that produce compatible ends. To make pMP21, pMS183Δ was cut with Asp718/XbaI and ligated to the wild-type Tn3 site I-containing, Asp718/XbaI fragment from pCO1.

Libraries RB galK Lib 1-4 were made by replacing the small Asp718/XbaI fragment of pMP20 with an Asp718/XbaI fragment containing mutant Tn3 site I variants from SS libraries RB Lib 1-4, respectively. Libraries RB galK Lib 4a-7 were constructed by ligation of NheI/BsrGI-restricted pMP21 backbone and the Asp718/XbaI fragment containing mutant Tn3 site I variants from SS libraries RB Lib 4-7, respectively.

2.28.1.2 Random site x random site library

To make the libraries with two identical copies of the random recombination site, SS libraries were used to transform electro-competent JC8679 cells. Following the transformation JC8679 cells were grown at 37°C overnight, on L-agar plates containing ampicillin. Colonies were scraped off the plates using a sterile glass spreader and the plasmid DNA was extracted using a Qiagen Midiprep method (1 Midiprep column/0.33g of cells). JC8679 strain, due to a genetic defect, causes all episomes maintained in it to multimerise. To isolate different size multimers, plasmid DNA was separated by agarose gel electrophoresis (0.7% TAE agarose gel, 17h at 35V). Gels were stained with ethidium bromide and the plasmid dimer band was cut out on the long wave UV-transilluminator. Plasmid dimer was extracted from the gel chip using the Qiagen Gel extraction kit and used to transform electro-competent DH5 cells. Following the transformation DH5 cells were grown at 37°C overnight, on L-agar plates containing ampicillin. To finally obtain the libraries with two identical copies of the random recombination site, DH5 colonies were scraped off the plates using a sterile glass spreader and the plasmid DNA was extracted using a Qiagen Midiprep method (1 Midiprep column/0.33g of cells).

2.28.2 Random libraries of Tn21 resolvase mutants

Libraries of Tn21 resolvase mutants were generated using a two-step mutagenic PCR procedure according to the method adapted from Fromant *et al.*, 1995. In the first step, using pDW21 as a template, the Tn21 resolvase ORF was amplified with primers 22F and 23R (Table 2.4) in 20 µl total reaction volume. A PCR reaction was set up using: 0.25mM dNTPs, 1.5 mM $MgCl_2$, 10% 10x Promega TPol buffer, 0.25 µM primers, 1 unit of Taq(NEB)). To this mixture mutagenic dNTPs (either dPTP or 8-oxo-G base analogues) were added (0.02 mM final) to increase the chances of mutation during the PCR reaction. The PCR program was: initial denaturing for 3 mins at 94°C, followed by 20 cycles of denaturing at 94°C for 30 seconds, annealing for 30 seconds at 53°C and extension for 5 mins at 74°C. Following the initial mutagenic step, a ‘clean up’ PCR reaction was performed using a 1/100 dilution of the mutagenic PCR reaction as the template. The ‘clean up’ PCR reaction was set up using the same concentration of components in a 50 µl

total reaction volume, but without the mutagenic dNTPs. The PCR program was the same as in the initial mutagenic step. The products of the 'clean up' PCR reactions were separated on a 1.2% TAE agarose gel (90 min, 150 V). The gel was stained with ethidium bromide, the amplified product band containing the mutated Tn21 resolvase ORF was cut out on the long wave UV-transilluminator, and finally the DNA was extracted from the gel chip using the Qiagen Gel extraction kit. The fragment was then restricted using NdeI/Asp718 and cloned into a pMS140 backbone cut using the same restriction enzymes. Approximately 0.1 μg of this ligation was used to transform electro-competent DH5α cells (Invitrogen). Following the transformation, DH5α cells were grown at 37°C overnight, on L-agar plates containing ampicillin. DH5α colonies were then scraped off the plates using a sterile glass spreader and the Tn21 resolvase mutant library plasmid DNA was extracted using the Qiagen Midiprep method (1 Midiprep column/0.33g of cells).

2.29 Molecular Graphics

Molecular graphics were generated from PDB coordinates: 1GDT, 1ZR4, 2ROQ and 1AAY (Yang & Steitz, 1995; Li *et al.*, 2005; Mouw *et al.*, 2008; and Elrod-Erickson *et al.*, 1996, respectively) using PyMOL software (www.pymol.org)

3 Chapter 3: Sequence selectivity in the central 16 bp of Tn3 site I

3.1 Introduction

Recombinases of the resolvase family are modular in nature, with an αβ fold N-terminal catalytic domain, a helix-turn-helix (HtH) C-terminal DNA-binding domain, and an extended arm region that connects the two domains and binds DNA around the minor groove (Yang & Steitz, 1995; Smith & Thorpe, 2002). The C-terminal, HtH domains bind DNA by making specific major groove contacts with the six outermost base pairs of each end, flanking the central 16 base pairs of the 28 bp long Tn3 site I. Replacing the HtH domain with another DNA-binding domain can alter the binding specificity of the resolvase protein to that of the new domain (Akopian *et al*., 2003; Gordley *et al*., 2006).

However, binding of a recombinase to the site I does not itself ensure recombination. A wild-type Tn3 resolvase requires the presence of the *res* accessory sites II and III, to which it has to bind in order to be able to bring the two copies of site I together and form a synapse. It is then and only then, that the recombination at site I can proceed (Arnold *et al.*, 1999; Kilbride *et al.*, 1999; Grindley *et al.*, 2006).

A number of activated mutants of resolvase that do not require the presence or binding to the *res* accessory sites II and III in order to catalyse recombination at site I had been isolated (Arnold *et al*., 1999; Olorunniji *et al*., 2008). In the case of such mutants, recombination only ensues on sites with a central 16 bp of sequence that is sufficiently similar to the sequence that the N-terminal domain normally works on (see Chapter 4). This has led us to postulate that there is another aspect of sequence specificity exhibited by these proteins, as well as that conferred by binding of the HtH domains to the ends of site I, possibly accounted for by the arm region minor groove contacts to DNA in the central 16 bp of the site I.

Here we show a series of *in vitro* and *in vivo* assays that were performed to test this hypothesis, and the interactions that have been uncovered. Understanding of these protein-DNA interactions could facilitate the design of recombinases with altered DNA specificity, which could have a variety of uses.

3.2 Rationale and the choice of experimental approach

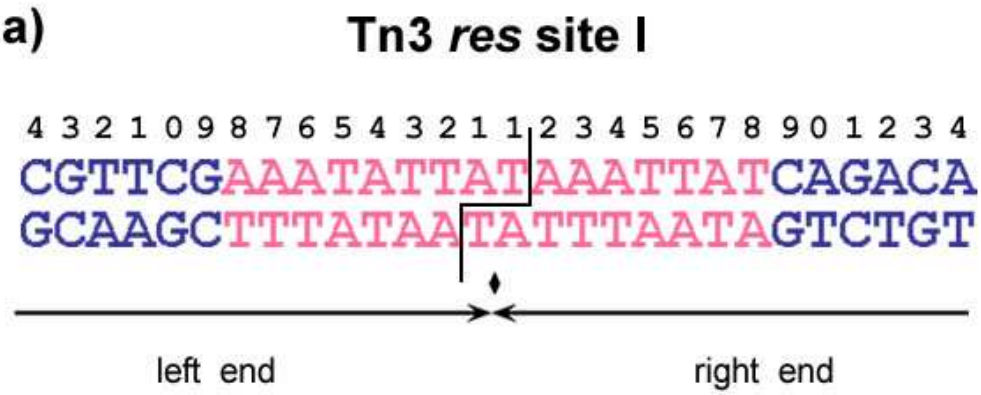
The sequence of the central 16 bp of the Tn3 resolvase site I consists exclusively of AT base pairs (Fig. 3.1a). This AT bias is conserved throughout the Tn3 resolvase family, implying active selection over the evolutionary time scale (Fig 3.1b). It must be borne in mind that, within a complete *res* site in a transposon, this sequence, apart from functioning as a recombination crossover site, also contains a part of the transposase gene promoter (Grindley, 2002). As promoter function and recombination function are quite different, yet both demanded from a rather short sequence, it is conceivable that the current sequence of the Tn3 site I is the product of an evolutionary compromise. This poses a question: to what extent can the sequence composition bias observed be attributed to the requirements for recombination? Furthermore, as the site I in the natural system exists as a part of *res*, it is also possible that the site I sequence is adapted to optimise regulation by the accessory sites II and III. Which of the conserved base pairs are conserved due to the requirements for transposase gene promoter function and which are conserved because they are required for recombinase activity or the need for its regulation by the accessory sites?

No work to date has systematically attempted to answer these questions. Here I have attempted to define the importance of all 16 base pairs in the centre of the Tn3 crossover site for recombination activity.

To achieve this, I decided to randomly mutate all 16 base pairs at the centre of the Tn3 site I, construct libraries of substrate plasmids containing one, or two (identical) mutant site I's, and screen them for recombination proficiency or deficiency using methods described in the following sections. Putting the mutant site I in the context of *res,* and performing selections using wild-type Tn3 resolvase protein was considered; however, in order to avoid the issues regarding regulation, which could add further complexity to the data, it was decided to perform library selections using a hyperactive Tn3 resolvase mutant called NM (R2A, E56K, G101S, D102Y, M103I, Q105L), that recombines at site I without accessory sites, and is not subject to any regulatory pressures (Olorunniji *et al.*, 2008). Also, considering the interest in using resolvase as a precursor for designer recombinases (Sections 1.6 and 1.7), it was concluded that performing selections with the hyperactive protein was more appropriate.

The theoretical diversity of a DNA sequence with 16 random positions is 4^{16} or roughly 4.3 billion combinations. To create a library that would ensure a 95% chance of a sequence being represented, according to the Poisson distribution, at least 3×4^{16} (12.9×10^{9}) clones would be required (Fig 3.2). Creating and screening such a vast library, although

a)



b)

Origin	Accession no.#	4 3 2 1 0 9 8 7 6 5 4 3 2 1 1 2 3 4 5 6 7 8 9 0 1 2 3 4
Tn3	V00613	CGTTCGAAATATTATAAATTATCAGACA
Tn1	L10085	CGTCCGAAATGTTATAAATTATCAGACA
Tn5431	X87814	CGTACGAAATGTTATAAATTATCGGACA
IS101	X01654	TGTCTGATATATCGAAACATAATGTACA
γδ	J01843	CGTCCGAAATATTATAAATTATCGCACA
γδ, *Y. pestis*	2996270	CGTCCGAAATATTATAAGTTGTCGGACA
Y. enterocolitica	37518397	CGTCCGAAATGTTATGAATTATAAGACA
R46	151827	CGTCCGAAATATTATAAATTCTCGGACA
P. syringae	24266782	AGTCTCCTTACTCATAAGTTTGTGGACA
Tn4659	111036276	AGTCTCCTTTCTCATAAGTTTGTGGACA
S. putrefaciens	124545222	CGTCCGAAATGTTATAAATTATCGGACA
C. amoebophilus	189497399	TGTCTACAAAGTTATAATATAATAGACA
Tn5070_Hg	11071875	TGTCTATTAATTTATAACTCTGTGGACA
H. ducreyi	AE017143.	TGTCCAAAAAATTATAAATTTTTTGGAAA
Erwinia SB	AACY01285318	CGTCTCAAATATTATAAATTGTGAGACT
B. bacilliformis	ZP_00947373	TATGTCATAAATTGTAAGATTTGACATA

Figure 3.1- a) The 28 bp sequence of Tn3 site I with the cleavage point marked by the black lines. Numbering of the individual DNA bases is from the centre: 1-14 left and 1-14 right. For bases 10-14 only the second digit is shown, e.g 0 represents base number 10, and so on. The "Left" and "Right" end refer to the conventional representation of site I within *res*, where the binding sites are in the order I-II-III left to right. The central 16 base pairs are given in pink while the outermost 6, recognised by the resolvase C-terminal helix-turn-helix domain, are in blue.
b)An alignment of Tn3 resolvase family site I's along with the sequence's origin and the nucleotide database accession numbers. Site I examples were chosen to point to the similarities and highlight the diversity of recombination site sequences recognised by the resolvases from Tn3 family. All the sequences are found in Tn3-like *res* sites and the proteins associated with them are close Tn3 resolvase relatives according to protein sequence alignment. A grey rectangle behind the sequences marks the central dinucleotide. Alignment performed in Jalview 2.1.2.

m	%=100(1-e^{-m})
0.001	0.1%
0.01	1%
0.1	9.5%
0.3	26%
1	63.2%
3	95%
6	99.8%
10	~100%

Poisson distribution $p_n = \frac{m^n}{n!} e^{-m}$

m = theoretical 'fold' coverage i.e. average (expected) frequency of the event (each clone) in a random sample of known size.

p_n = the probability of n occurrences of the event (clone occuring) in the sample.

$p_{(\geq 1)} = 1 - p_0$ ***the probability that there are NOT zero occurrences of the event (clone) in the sample.***

$p_{(\geq 1)} = 1 - \frac{m^0}{0!} e^{-m} = 1 - e^{-m}$

Figure 3.2- A table showing the relationship, between the 'fold' coverage of the theoretical library size (m), and the calculated probability of a clone occurring in such a library (%) using the Poisson distribution.

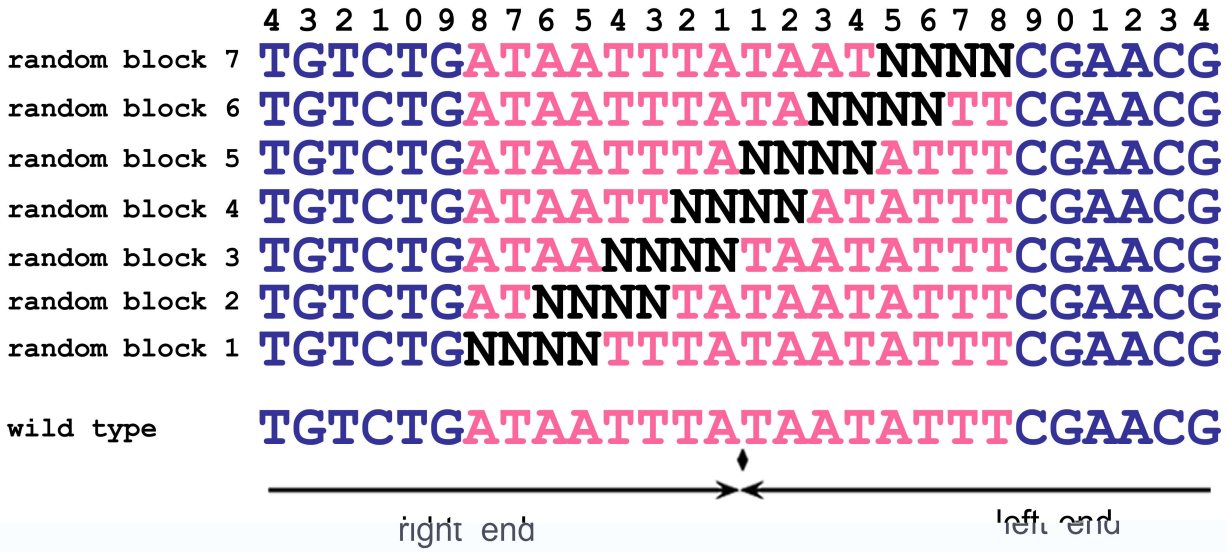


Figure 3.3- The sequences of the top strand of mutated Tn3 site I variants in the seven random block libraries (RB Lib 1-7). The sequence of the central 16 nucleotides is given in pink, outermost six nucleotides, recognised by the resolvase C-terminal helix-turn-helix domain, are given in blue, while randomised positions are in black. The orientation of the site is marked by arrows at the bottom of the figure. Please note that the "Left end" and "Right end" refer to the conventional representation of site I within *res*, where the binding sites are in the order I-II-III left to right.

theoretically possible, would in practice prove challenging. Instead of satisfying ourselves with a much lower percentage clone recovery from a smaller library, an alternative approach was employed.

The central 16 bp of Tn3 resolvase crossover site was subdivided into seven overlapping blocks of four base pairs. Each of these blocks was positioned in such a way that it overlaps with the nearest neighbouring block by 2 base pairs (Fig 3.3). The theoretical diversity of a DNA sequence with 4 random positions is 4^4 or 256 combinations, requiring a much more manageable 768 clones for 95% or ~2500 clones for 99.995% coverage. The overlapping arrangement of blocks was devised in order to allow selections to be self-controlling as one position is mutated in more than one block. The overlap also allows one to look for any neighbour context effects, and it doubles a number of data points per position, making statistical analysis more reliable.

3.3 Library construction

To make the seven random block libraries required to cover the central 16 bp of Tn3 site I, a partial duplex comprising RSLP primer annealed to the RBO1-7 (see Table 2.4) was extended by *Pfu* DNA polymerase, resulting in double-stranded DNA fragments containing random changes in a specific 4 bp block of site I. The resulting double stranded DNA fragments were digested with EcoRI/SacI and cloned into the pMP2 (Table 2.5) backbone by a fragment swap protocol (Fig 3.4). pMP2 was used as the vector because it contained a Tn21 Z-site (Z22Z(Tn21)), which is not bound or recombined by NM resolvase (see Chapter 4). Every effort was taken to achieve complete vector digestion, but even if traces of the original vector plasmid remained in any of the libraries, it would not be selected in any assay asking for resolution by NM resolvase. Each of the seven random block libraries generated (RB lib 1-7) contained around 10^4 clones, well over the ten-fold coverage required to guarantee near 100% representation of the theoretical sequence space. The library DNA was sequenced in bulk using the uni -43 primer (Fig. 3.5, for primer sequence see Table 2.4).

3.4 Site selection strategies

Once the seven random block libraries had been generated, three separate site selection strategies were employed:

- **Strategy A** – MacConkey assay based selection (Section 3.5)
- **Strategy B -** Plasmid fusion-based selection (Section 3.6)

- **Strategy C -** Plasmid multimer resolution-based selection (Section 3.7)

The different strategies approach the problem in different but complementary ways. Strategies A and B assay recombination between a wild-type Tn3 site I and a library of mutant site I variants. Strategy C looks at recombination between two identical mutant sites from a library of mutant Tn3 sites I's.

In theory, all of the three strategies allow for selection either *in vivo* or *in vitro*. *In vivo* and *in vitro* recombination reactions differ markedly in perceived recombination efficiency. The *in vivo* resolution reaction utilises a two-site *gal*K substrate, designed to be resolved into two circular DNA molecules only one of which has an origin of replication. The circle without the replication origin, carrying the *gal*K gene, is selectively lost from the cells on subsequent cell divisions, and is effectively removed from the reaction. Removal of the origin-less DNA circle pushes the reaction equilibrium towards resolution, leading to a perceived recombination efficiency of 100%.

On the other hand, the *in vitro* recombination reaction of a substrate with two site Is by a hyperactive resolvase reaches an equilibrium containing various species: non-recombinant, resolution products, inversion products and products of intermolecular recombination. *In vitro*, no reaction products are removed so the reaction remains in equilibrium, resulting in resolution efficiency reaching 50% at best.

This difference between *in vivo* and *in vitro* reactions can be exploited to select sequences that fail to recombine, when the selection is performed *in vivo*. Conversely, screening the same mutant libraries *in vitro* can be used to select for sequences that are recombination-proficient. Analysing these two mutually exclusive data sets should yield a powerful insight into Tn3 site I DNA sequence requirements for recombination. Although each of the three selection strategies could in theory allow for either *in vivo* or *in vitro* selection, this complementary selection approach was used with Strategy A only, as this strategy was technically most suitable for it.

The three selection strategies employed and the results obtained using them are discussed individually in the following sections.

3.5 Strategy A – MacConkey assay-based selection

Strategy A took advantage of the MacConkey assay (Section 2.23). In this assay, recombination can be visualised by observing the colony colour of *gal*K$^-$ *E. coli* cells in

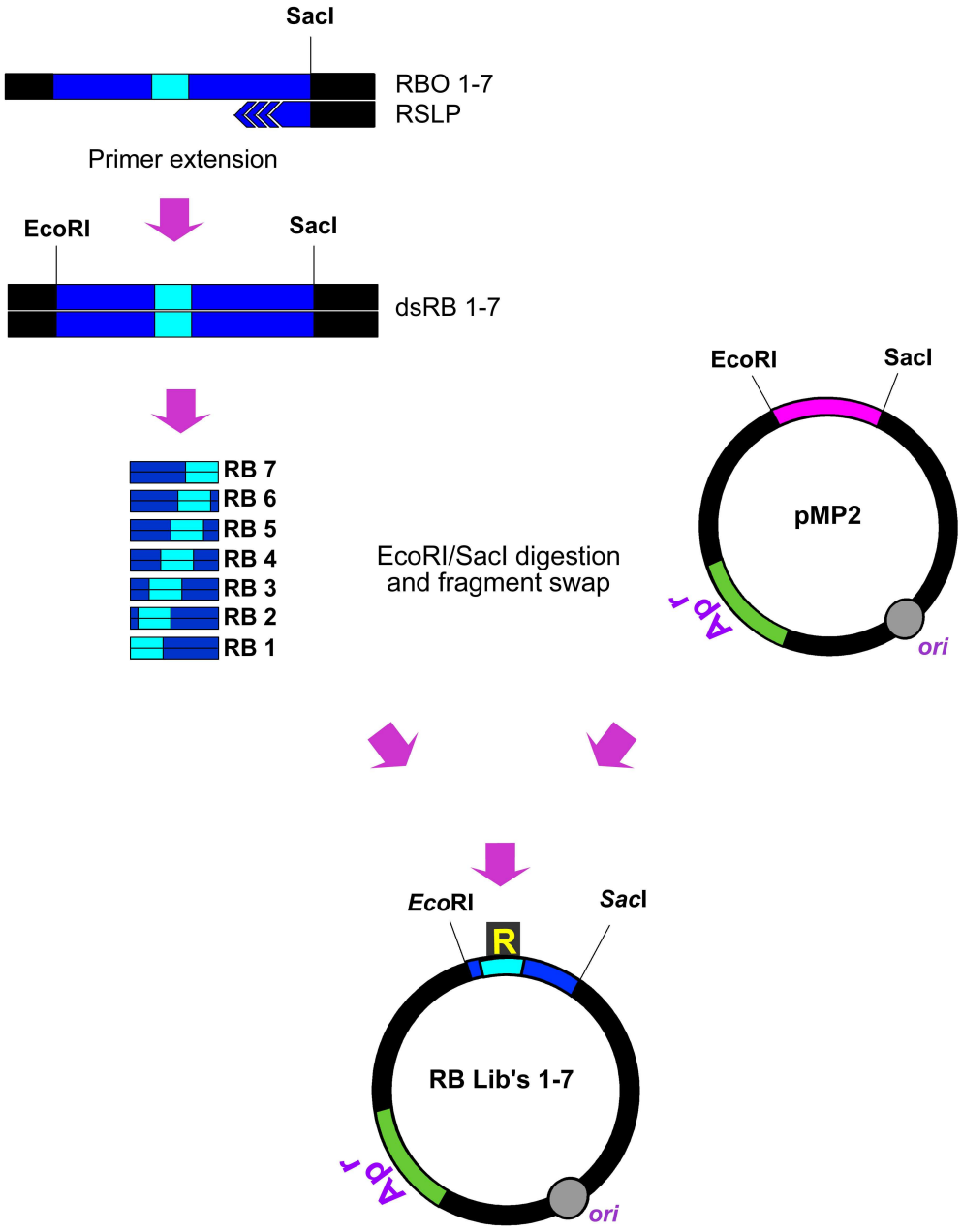


Figure 3.4-A schematic representation of the strategy (not to scale) used to generate seven random block libraries, required to cover the central 16 bp of the Tn3 site I. The mutated 4bp block is in cyan, and is further highlighted with a yellow R on a grey background in the diagram representing the random block library plasmids RB Lib 1 to RB Lib 7.The drawing is not to scale.

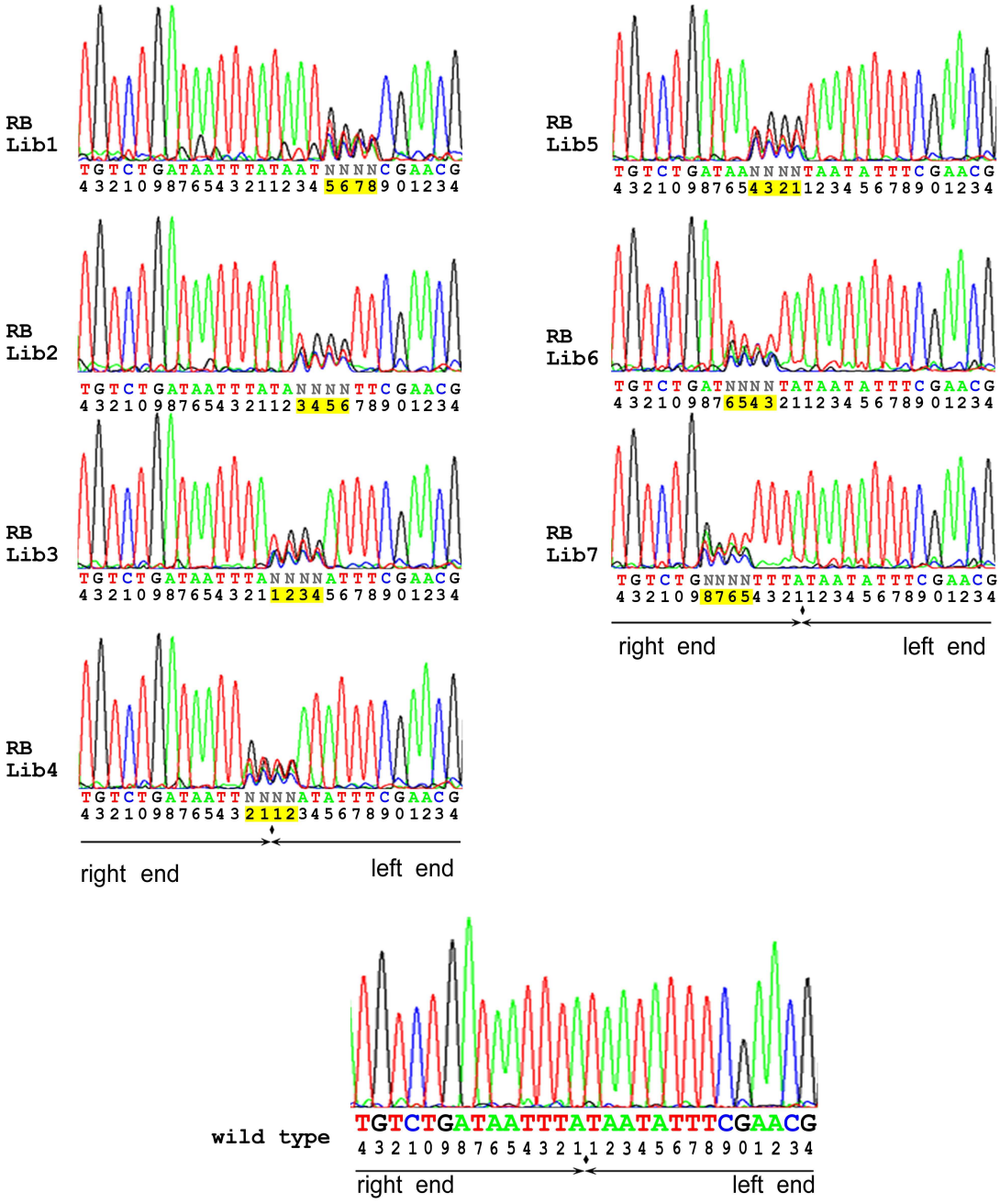


Figure 3.5- Bulk sequencing traces from the random block libraries RB Lib 1 to RB Lib 7 using the primer uni -43. The individual DNA bases are numbered from the centre: 1-14 left and 1-14 right. For bases 10-14 only the second digit is shown, e.g 0 represents base number 10. The position of the random block is highlighted in yellow below the sequencing trace. Each sequencing trace represents combined sequencing result of ~10^4clones. A sequencing trace of wild type Tn3 site I sequenced from a pCO1 template is provided for comparison at the base of the figure.

which resolution of a test substrate is taking place, when grown on selective MacConkey agar. White colonies signify high levels of test substrate resolution, whereas red colonies indicate that the resolution is inefficient. To be able to employ this assay, a set of eight *gal*K test plasmid libraries was constructed, by excising the DNA fragment containing the mutated Tn3 site I from each of the seven random block libraries (Section 3.3) and cloning it into precursor plasmid vectors pMP20 or pMP21 (Fig 3.6). pMP20 or pMP21 vectors contained a *gal*K gene flanked on one side by a wild type Tn3 site I and on the other side by either a Tn21 site I (pMP20) or a short stuffer sequence (pMP21). Precursor plasmids containing these sequences were chosen to ensure that even if traces remained in any of the *gal*K libraries, the plasmid itself could not be selected in an assay requiring resolution by NM resolvase, as neither the Tn21 site I nor the stuffer sequence are amenable to recombination by NM resolvase. Mutant Tn3 site I's were fragment-swapped to replace the Tn21 site I or the stuffer sequence in the precursor vectors, to generate the *gal*K test plasmid libraries (Fig. 3.6). Each library contained ~10-15 000 clones, quite enough to maintain the library diversity. The choice of whether pMP20 or pMP21 was used as vector depended on which random block-containing fragment was to be cloned. pMP20 was used for random blocks 1-4 while pMP21 was used for random blocks 4-7. The resulting site arrangement ensured that, upon resolution, sequence data can be recovered, as the half-site containing the random bases remained on the DNA circle that has an origin of replication (Fig. 3.7). As the random block 4 straddles the very centre of the crossover site, half of it is always lost after recombination. Because of this, random block 4 was cloned into both pMP20 and pMP21 vectors. The *gal*K plasmid library containing random block 4 and pMP20 backbone was called RB GalK Lib 4, while the one with pMP21 backbone was labelled RB GalK Lib 4a. Recombination between a site I with a randomised block and a wild-type Tn3 site I deletes the *gal*K gene from the substrate plasmid. When grown on selective MacConkey agar, the DS941 *E. coli* colonies now containing the resolved substrate plasmid turn pale ("white"), as they are unable to metabolise galactose. Conversely, on the same growth medium the cells containing test plasmids that fail to resolve remain *gal*K^+ and therefore form red colonies.

After the *gal*K plasmid libraries were generated, initial selections were done *in vivo.* DS941 *E. coli* cells were made competent and transformed with pFO2, a plasmid expressing NM resolvase (Olorunniji, 2006). The resolvase expression plasmid was introduced to the cells first, allowing the resolvase to be expressed so that resolution can ensue as soon as a substrate plasmid becomes available. DS941/pFO2 cells were made competent, transformed with *gal*K plasmid libraries 1-7 (or pMP243 as positive control), and plated on selective MacConkey agar (Fig. 3.8). pMP243, a plasmid with two wild-type

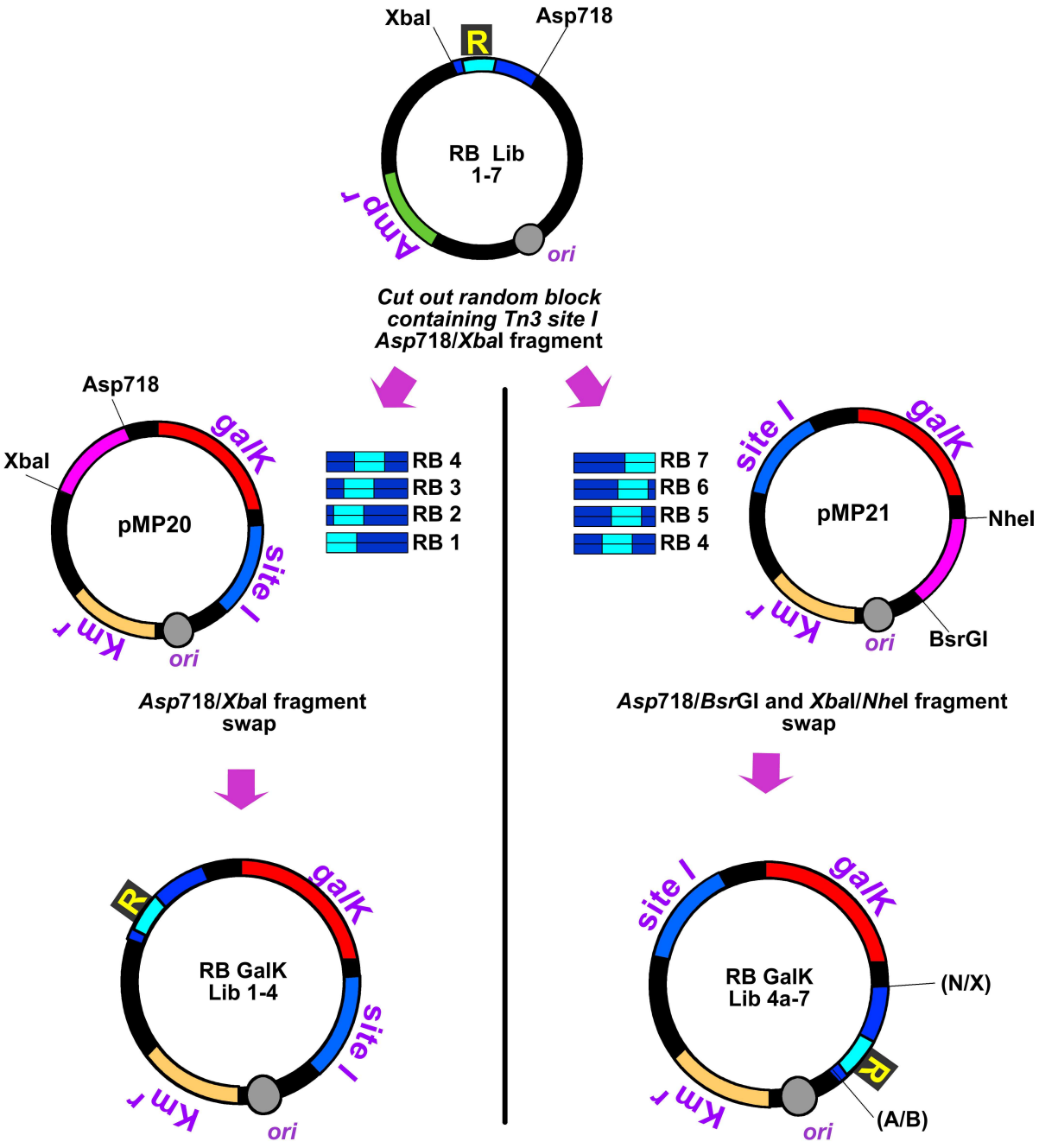


Figure 3.6-A schematic representation (not to scale) of the strategy used to generate eight *gal*K test plasmid libraries. A cloning strategy used to generate RB GalK Lib 1 to RB GalK Lib 4 is given on the left, whilst the strategy employed for constructing RB GalK Lib 4a to RB GalK Lib 7 is presented on the right.The mutated 4bp block is in cyan, and is further highlighted with a yellow R on a grey background in diagrams showing the library plasmids.

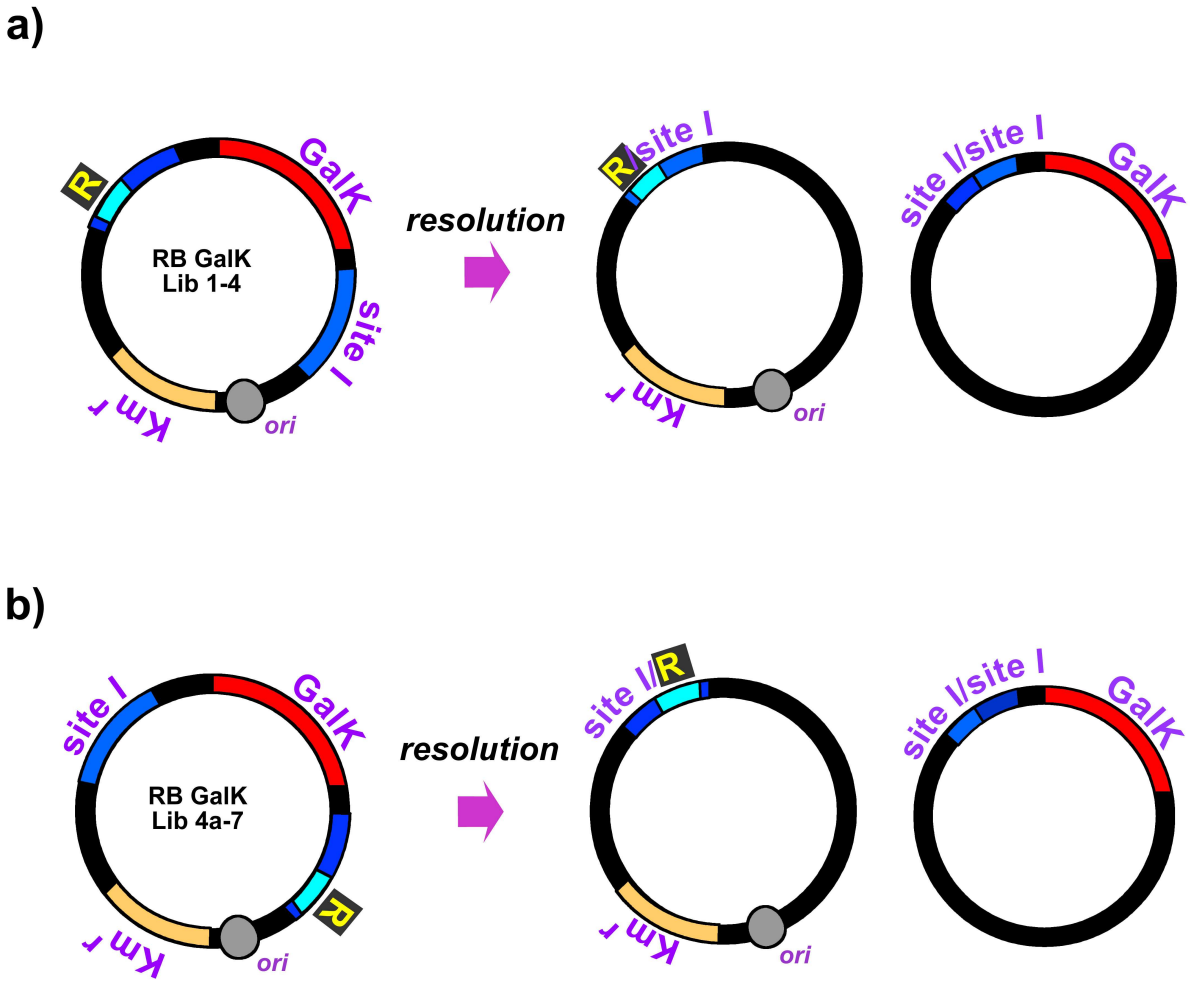


Figure 3.7-A diagramatic representation (not to scale) of the *in vivo* resolution reaction with the eight *gal*K test plasmid libraries used as the substrate. The reactions with RB GalK Lib 1 to RB GalK Lib 4 are given on the top (a), whilst the reactions using RB GalK Lib 4a to RB GalK Lib 7 are presented on the bottom of the figure. In both instances the reaction products are shown on the right. Please note that only the product with an origin of replication (ori) is stably maintained in the cell culture and can therefore be isolated and sequenced. The mutated 4 bp block is shown in cyan, highlighted with a yellow "R" on a grey background.

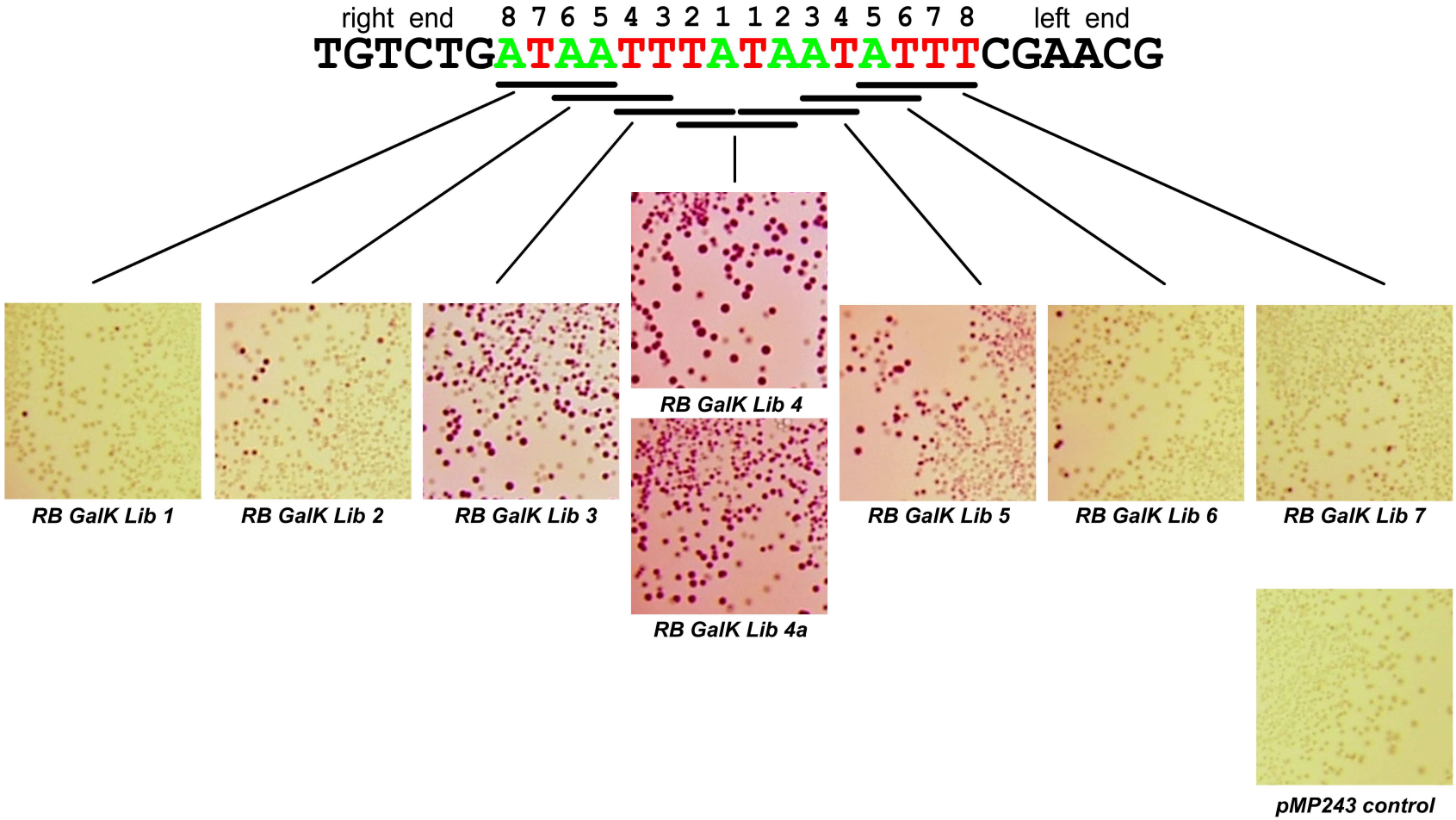


Figure 3.8-*In vivo* selections using the *gal*K test plasmid libraries RB GalK Lib1 to RB GalK Lib7. Photographs of sections of selective MacConkey agar plates, showing the degree of substrate plasmid resolution for the libraries tested, are presented below the sequence of Tn3 site I.The position of the random block for each library is indicated below the sequence by a thick black line, which is connected to the corresponding selection plate photograph by a thin black line. A photograph of the section of a MacConkey plate showing the extent of resolution of the positive control pMP243 containing two copies of wild-type Tn3 site I is given in the bottom right corner of the figure. Pale (white) colonies indicate sucessful resolution, whereas red colonies signify inefficient resolution reaction.

Tn3 site I's flanking *gal*K, was fully resolved. Resolution of *gal*K plasmid libraries 1-7 was partial as can be seen; libraries containing random blocks closer to the centre of site I were less resolved, while the ones with random blocks closer to the edge of the site were more resolved. Red colonies, containing substrate plasmids that failed to recombine, were picked and streaked out on fresh selective MacConkey plates. DNA from these colonies was extracted and sequenced using primer OCP1 for *gal*K plasmid libraries RB GalK Lib 4a-7 and OCP2 for RB GalK Lib 1-4 (Table 2.4). Sequences of the Tn3 site I variants isolated from these clones are given in Fig. 3.9. The same sequence data have been summarised in histogram form and presented in Fig 3.10. It should be mentioned at this point that as the selection in this assay was based on recombination between an unmodified Tn3 site I and a mutant Tn3 site I, a potential synapse formed between such two sites would be expected to contain sequence modifications in only one out of four half-sites. This assay should therefore select for the very worst Tn3 site I variants containing changes in one half-site that are so unfavourable that they could not be compensated for by the wild-type DNA protein interactions in the remaining three half-sites.

Most recombination-deficient Tn3 site I variants that were isolated were found to have 3 or 4 changes with respect to the wild-type Tn3 site I sequence. Although most of the variants were isolated only once, certain variants were selected up to three times. Irrespective of the number of times that a sequence was isolated, each variant was included only once when the histogram was made, to avoid skewing the data. For libraries RB GalK Lib 1 and RB GalK Lib 7, where red colonies were scarce, the number of individual isolates that were sequenced was much higher than is evident from Fig.3.9. Most of these sequences were discounted however, as they contained either Tn3 site I variants with deletions or insertions, which were due to errors in the oligonucleotide synthesis, or in rare instances, because instead of a Tn3 site I variant, they carried either Tn21 site I (RB GalK Lib1) or the stuffer sequence (RB GalK Lib7), coming from the left-over vector. The high prevalence of these defective sequences indicates that only a very small number of base change combinations in the random blocks 1 and 7 is sufficiently deleterious to abrogate recombination. The most frequently observed changes in the recombination-deficient Tn3 site I variants are mutations of A or T to G or a C, with runs of 2 or more GC's being common. It is worth pointing out that there appears to be a degree of symmetry in the changes observed with regards to the left and right end of the site. As can be seen in Fig.3.10, the most common change in position 5R is to a C which is mirrored in position 5L where the most common mutation is to a G. Similarly, most frequent changes in position 3R are to an A or a G, whilst at 3L A is usually changed to a T or a C.

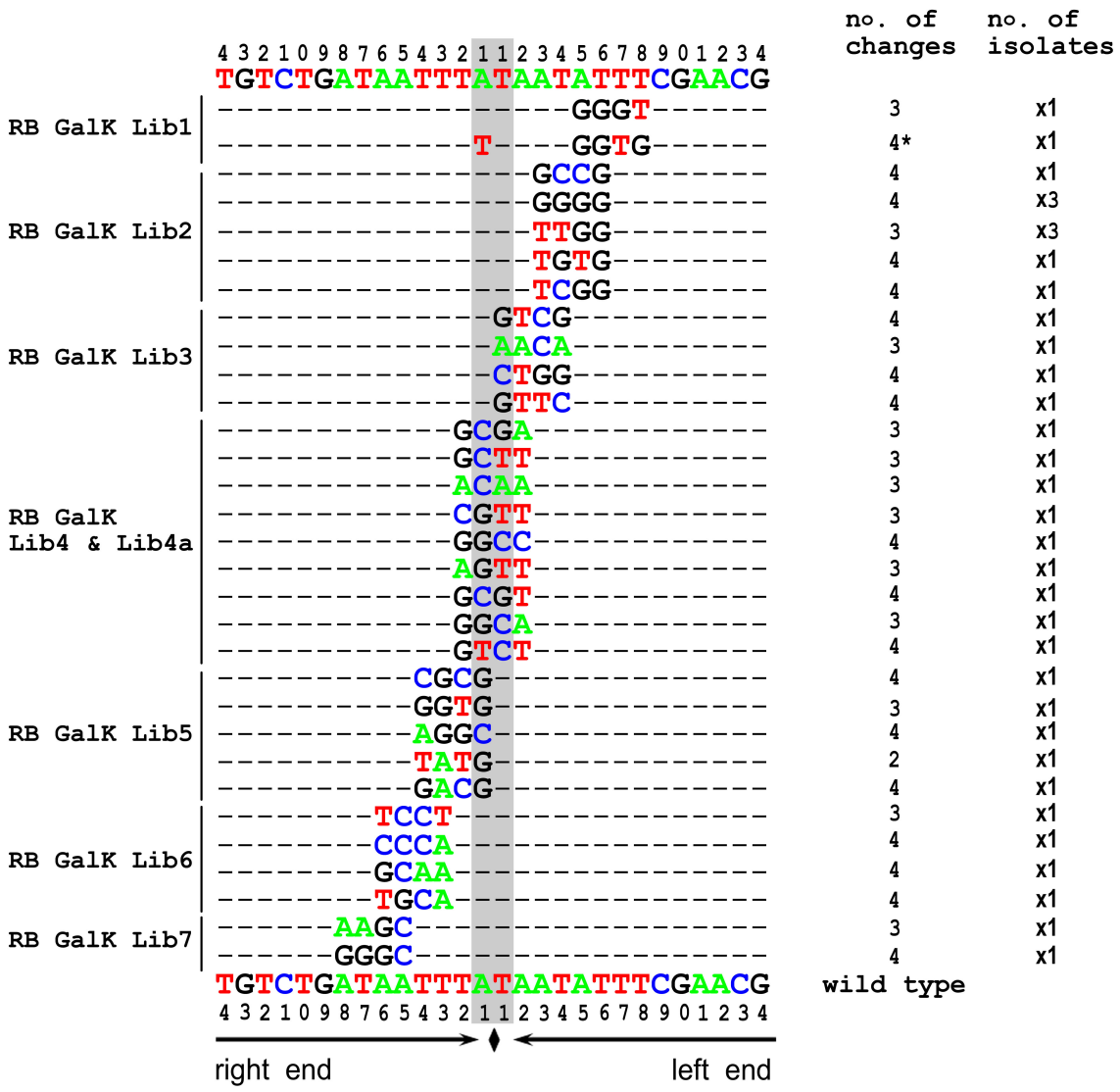


Figure 3.9-An alignment of recombination-deficient Tn3 site I variants selected *in vivo* using RB GalK Lib 1-7 substrate libraries. In the interest of clarity, for each isolated variant only the sequence that differs from the wild-type Tn3 site I is shown. The library that the variant sequence was isolated from is marked on the left of the sequence alignment. On the right of the alignment in the first column the number of sequence changes with regards to the wild-type sequence is provided, while in the second column the number of times a particular variant was isolated is shown. The number of sequence changes in the second variant from the top is marked with an "*" as there is an additional change outside the random block. The central dinucleotide is highlighted by a grey rectangle. Position numbering is from the centre, as in the previous figures.

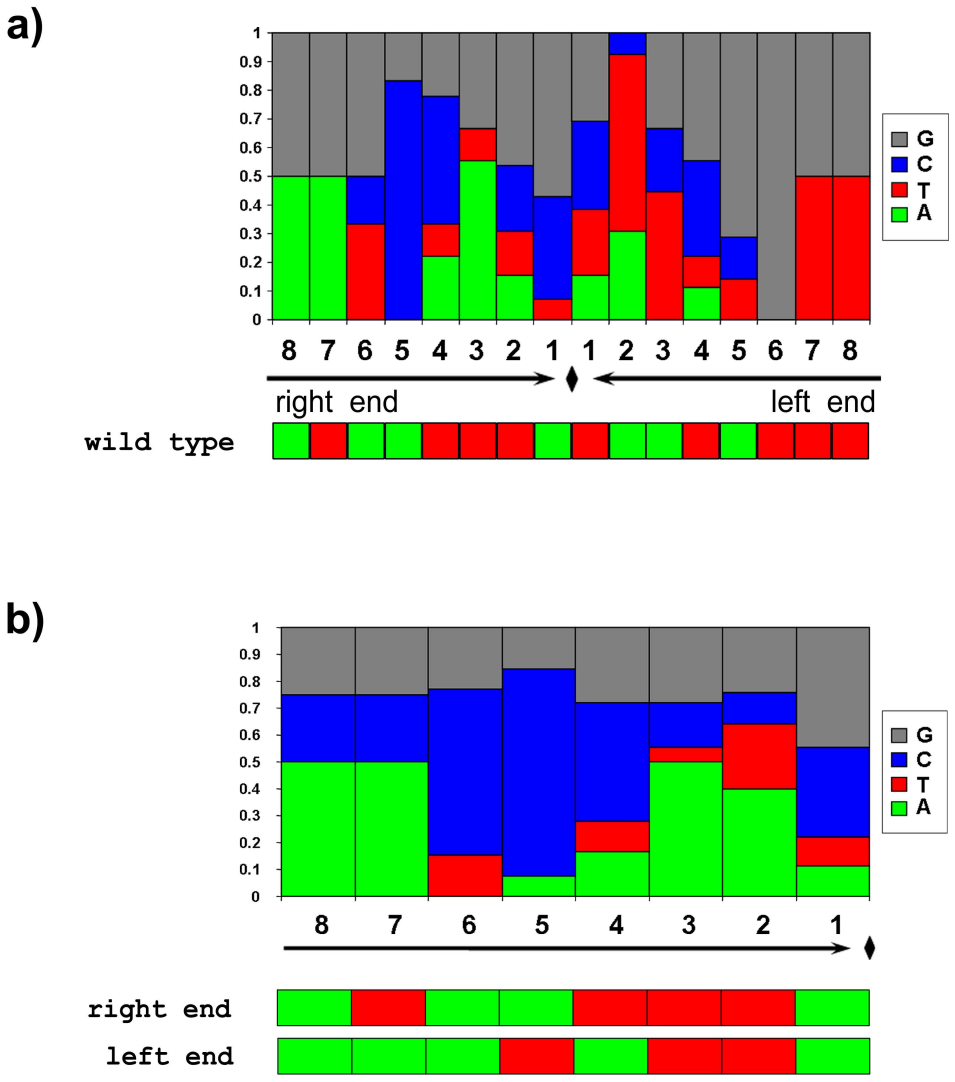


Figure 3.10-a) A histogram showing the relative base frequencies in the recombination-deficient sites presented in the Fig.3.9 Irrespective of the number of times a particular variant was isolated its sequence was used only once in the making of the histogram. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

Since the time required to resolve a substrate plasmid can not be measured in an *in vivo* assay, resolved substrate plasmid DNA from individual white colonies was not isolated nor sequenced. In an *in vivo* assay, a Tn3 site I variant that is resolved slowly would result in a white colony, just like a highly recombination-proficient Tn3 site I variant that is resolved in a fraction of the time, making it difficult to discriminate between the site I variants according to their recombination proficiency. The inability to grade different site I variants based on their recombination proficiency would severely complicate the analysis of requirements for recombination.

In order to be able to select the site I variants that recombine most efficiently, reactions with NM resolvase were performed *in vitro*. It was reasoned that if the reaction time is kept relatively short, only the sequences most suited to recombination will have recombined. To determine the shortest practical reaction time, a time course experiment was set up using NM resolvase and *gal*K plasmid libraries 1-7 as substrates (Fig 3.11). As can be seen on these gels, the NM-catalysed resolution reaction is rather fast, and even after the 0.5 minute time point, some evidence of recombination could be observed. After the *in vitro* recombination reaction, a part of the reaction mix was used to transform electrocompetent DS941 *E. coli* cells. Following transformation the cells were plated out on selective MacConkey agar. Cells transformed with resolved substrate plasmids were expected to give white colonies, and the ones transformed with unresolved substrate plasmids were expected to produce red colonies. The number of red and white colonies was determined, and used to calculate the apparent recombination frequency (see Table 3.1). It should be noted that this determined frequency is likely to be an underestimate as any multiple transformants that contain an unresolved substrate plasmid would also appear red.

In order to isolate resolved products only, the DNA in the resolvase *in vitro* reaction mix was digested with EcoRV restriction enzyme prior to transformation. This enzyme cuts within the *gal*K gene sequence and linearises any unresolved substrate plasmids, preventing them from transforming (Fig 3.12). DNA from resolvase reactions at 0.5, 5, and 60 minute time points was treated in this way, transformed and plated on selective MacConkey agar. The resulting white colonies were scraped off the plates, the DNA from the individual plates was isolated, and then sequenced in bulk, an example of which is shown in Fig 3.13. The bulk sequencing result represents an average of ~1000 different sequences and was intended to provide an overview of results over the entire crossover site by giving an idea of relative DNA base frequencies. As can be seen in the figure, even at an earliest time point there is a rise in the frequency of A and T bases at positions 5L and 6L when compared to the baseline where the G peaks are the highest. This trend continues

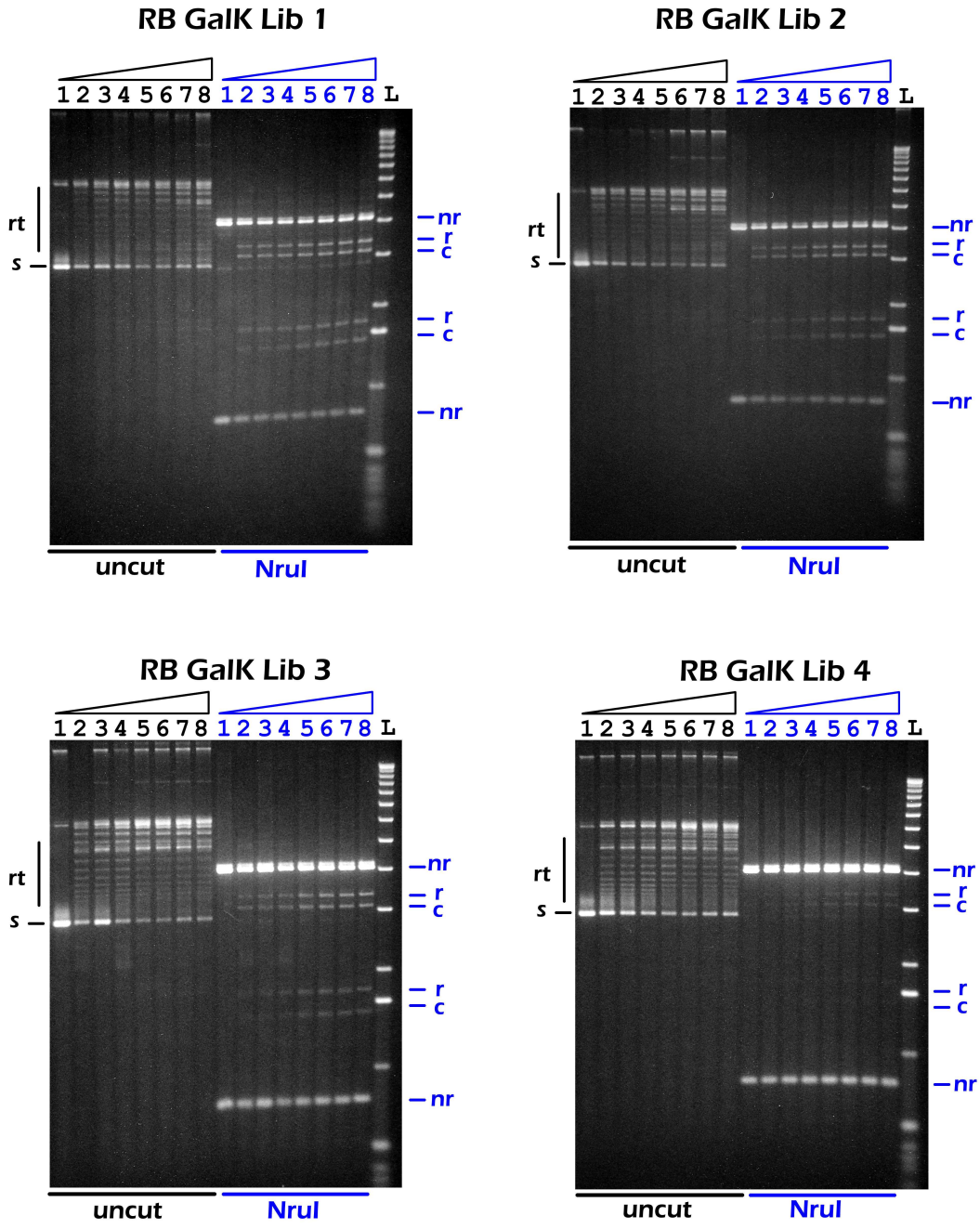


Figure 3.11-Time course of NM resolvase recombining mutant libraries RB GalK Lib1-RB GalK Lib 4. The time points are marked at the top of the gel as follows, 1 (0 min.), 2 (0.5 min.), 3 (1 min.), 4 (2 min.), 5 (5 min.), 6 (15 min.), 7 (30 min.) and 8 (60 min.). For protein and DNA concentration and buffer composition see Section 2.27. The gels on the left show the uncut reactions (black annotation) whereas the ones on the right show the corresponding reaction cut with NruI (blue annotation). The graphic representation of NruI digest of the possible reaction products giving their expected sizes is given in Fig. 2.13. The abbreviations are as follows, "s" (supercoiled substrate), "rt" (recombinant product topoisomers), "nr" (non-recombinant products), "r" (recombinant product), and "c" (cleavage product).

	Time Point	RED	WHITE	TOTAL	% RECOMBINANT
RB GalK Lib 1	0 min	245	0	245	0
	0.5 min	108	25	133	19
	1 min	71	28	99	28
	2 min	29	21	50	42
	5 min	168	69	237	29
	15 min	44	17	61	28
	30 min	80	30	110	27
	60 min	34	20	54	37
	Time Point	**RED**	**WHITE**	**TOTAL**	**% RECOMBINANT**
RB GalK Lib 2	0 min	117	0	117	0
	0.5 min	113	13	126	1
	1 min	97	18	115	16
	2 min	23	6	29	21
	5 min	44	14	58	24
	15 min	32	15	47	32
	30 min	101	39	140	28
	60 min	35	23	58	40
	Time Point	**RED**	**WHITE**	**TOTAL**	**% RECOMBINANT**
RB GalK Lib 3	0 min	394	1	395	0
	0.5 min	112	7	119	6
	1 min	387	20	407	5
	2 min	130	13	143	9
	5 min	200	25	225	11
	15 min	410	70	480	15
	30 min	176	23	199	12
	60 min	103	17	120	14
	Time Point	**RED**	**WHITE**	**TOTAL**	**% RECOMBINANT**
RB GalK Lib 4	0 min	238	4	242	2
	0.5 min	112	4	116	3
	1 min	68	2	70	3
	2 min	44	0	44	0
	5 min	79	4	83	5
	15 min	36	1	37	3
	30 min	120	2	122	2
	60 min	94	5	99	5

Table 3.1- *In vitro* recombination frequencies in a time course experiment with NM resolvase acting on random libraries RB GalK Lib1- RB GalK Lib 7. The table continues on the following page.

	Time Point	RED	WHITE	TOTAL	% RECOMBINANT
RB GalK Lib 4a	0 min	2504	5	2509	0
	0.5 min	831	17	848	2
	1 min	872	38	910	4
	2 min	1222	49	1271	4
	5 min	1521	82	1603	5
	15 min	967	72	1039	7
	30 min	765	53	818	6
	60 min	1393	115	1508	8
RB GalK Lib 5	Time Point	RED	WHITE	TOTAL	% RECOMBINANT
	0 min	2429	3	2432	0
	0.5 min	745	30	775	4
	1 min	895	32	927	3
	2 min	1001	67	1068	6
	5 min	829	100	929	11
	15 min	1300	188	1488	13
	30 min	1381	210	1591	13
	60 min	1063	173	1236	14
RB GalK Lib 6	Time Point	RED	WHITE	TOTAL	% RECOMBINANT
	0 min	1298	2	1300	0
	0.5 min	672	170	842	2
	1 min	482	107	589	18
	2 min	920	480	1400	34
	5 min	486	230	716	32
	15 min	639	365	1004	36
	30 min	102	45	147	31
	60 min	498	384	882	44
RB GalK Lib 7	Time Point	RED	WHITE	TOTAL	% RECOMBINANT
	0 min	179	3	182	2
	0.5 min	88	14	102	14
	1 min	1580	620	2200	28
	2 min	110	28	138	20
	5 min	246	66	312	21
	15 min	328	133	461	29
	30 min	148	69	217	32
	60 min	220	124	344	36

Table 3.1-continued

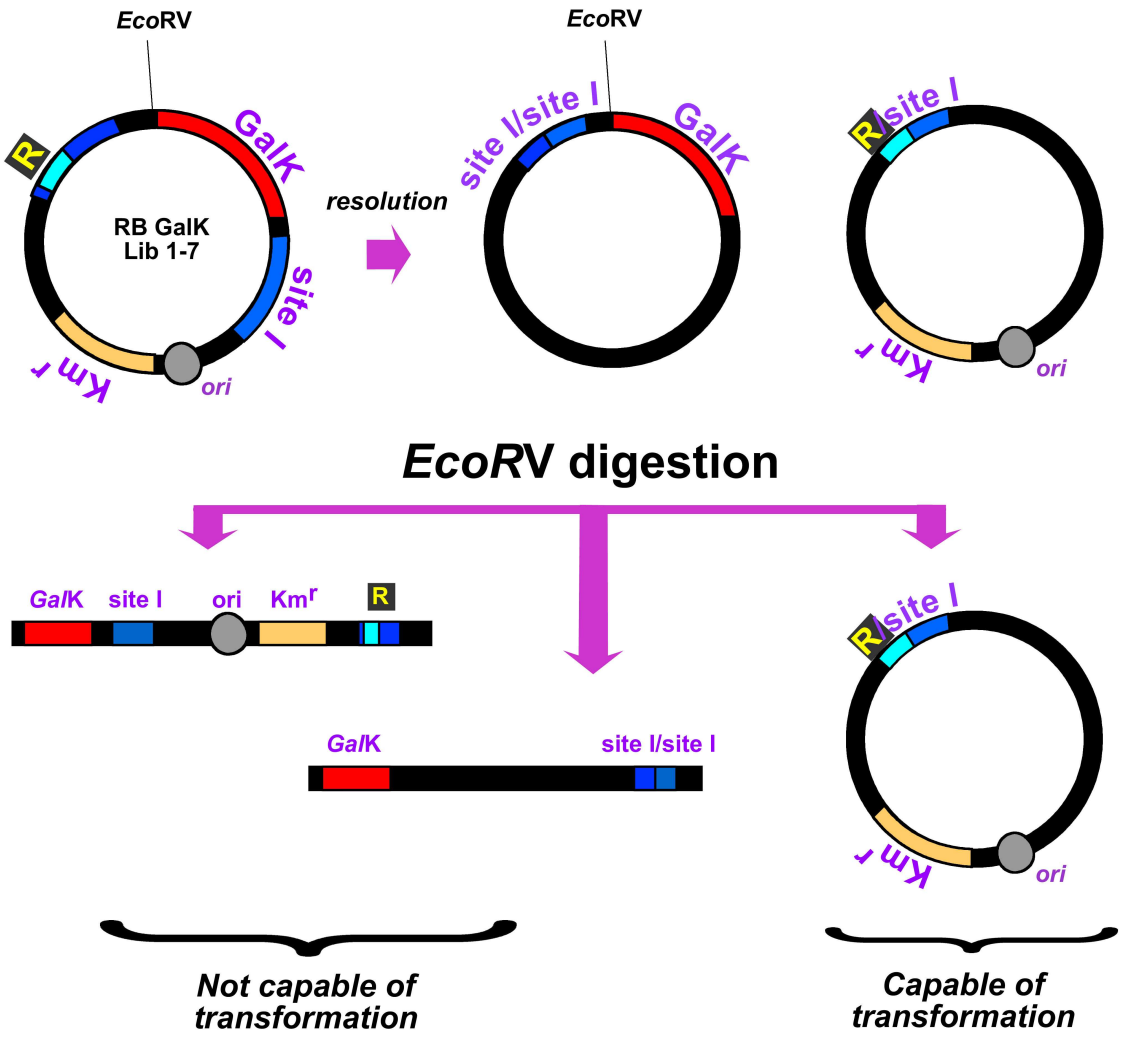


Figure 3.12-A diagramatic representation (not to scale) of the strategy for isolating resolved product containing the Tn3 site I variant out of an NM resolvase-catalysed *in vitro* resolution reaction with the eight *gal*K test plasmid libraries used as the substrate. Both the unreacted library plasmid and the *gal*K gene-containing resolution product are linearised by EcoRV digestion and rendered incapable of transformation. The resolution product containing the recombinant Tn3 site I variant with introduced changes is not cut with EcoRV, remains circular and is therefore capable of transformation . The mutated 4 bp block is shown in cyan, highlighted with a yellow "R" on a grey background.

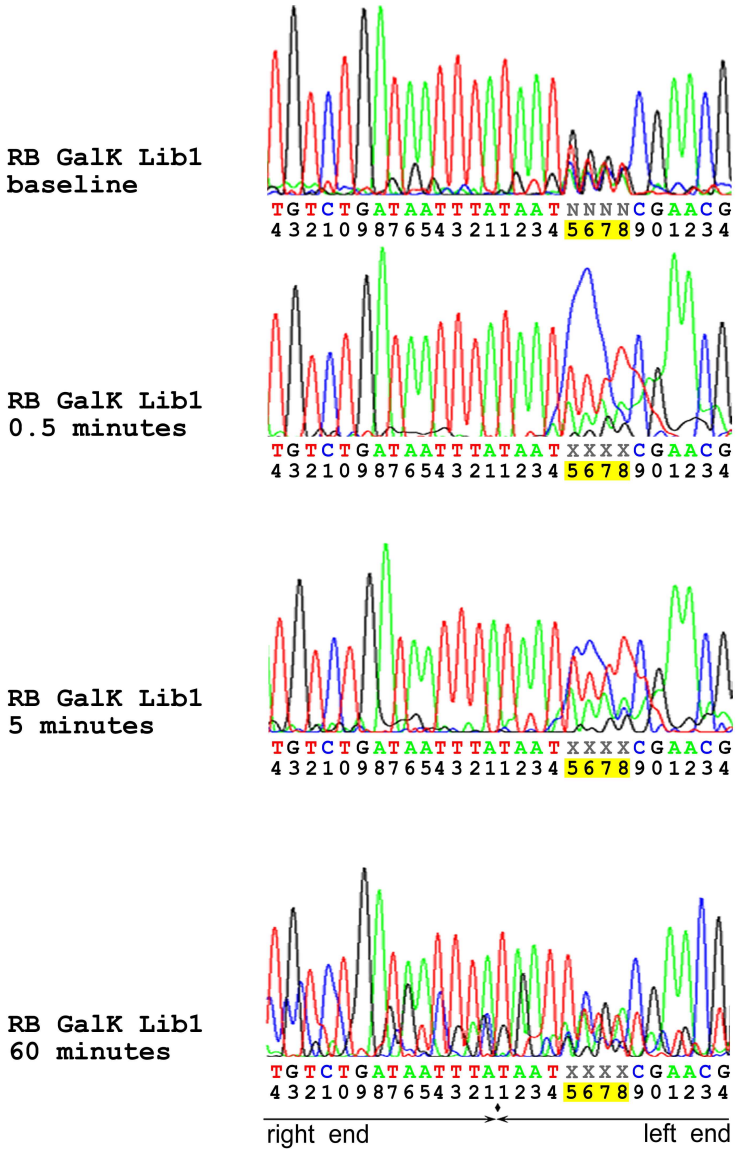


Figure 3.13- Bulk sequencing traces of mutant Tn3 site I variants obtained in selections with NM resolvase using the random block library RB GalK Lib 1. Different sequencing traces correspond to reaction times indicated to the left of the sequencing trace. Sequencing was done using the primer OCP2 (see Table 2.4). The individual DNA bases are numbered from the centre: 1-14 left and 1-14 right. For bases 10-14 only the second digit is shown. The position of the random block is highlighted in yellow below the sequencing trace. The "Left" and "Right" end refer to the conventional representation of site I within res, where the binding sites are in the order I-II-III left to right. Each sequencing trace represents combined sequencing result of ~10^3clones.

throughout the reaction and at the 60 minute time point position 5L is almost exclusively a T. The selection at positions 7L and 8L seems not to be so stringent with the four bases ending up being similarly represented at the final time point, although there seems to be an initial preference for T or A. At the 0.5 minute time point there is a large, wide C peak that is covering all 4 randomised positions, that seems to reduce in prominence at the later time points. There was a suspicion that this peak could be a sequencing artefact caused by the low quality of the plasmid DNA template. Bulk sequencing of plasmid DNA isolated from selections using other *gal*K libraries positions revealed promising glimpses of other base change trends, but similar sequencing artefacts to those seen in Fig.3.13 and poor quality of the sequencing trace in general, hampered the data analysis.

To circumvent these problems and because bulk sequencing can not be easily quantified or subjected to statistical analysis, an alternative method looking at individual Tn3 site I variant isolates was adopted. In order to do this, the resolved plasmid DNA from 20 individual white colonies at the 0.5 minute time point, from each of the *gal*K libraries, was sequenced. The early 0.5 minute time point was chosen, as that resolvase reaction mix should contain the Tn3 site I variants that are the easiest to resolve. These sequences and histograms showing the DNA base distribution for each of the 16 crossover site positions are given in Figs 3.14-3.17.

With the exception of the position 1L which remained a T in every Tn3 site I variant selected, all the other positions appeared amenable to change. Overall, A to a T or a T to an A transversions seemed to be much preferred over substitutions of an A or a T base with a G or a C. This was in stark contrast to the recombination-deficient Tn3 site I variants described previously. Also, the sites containing changes into G or a C usually contained only a single change of that type, as opposed to commonly observed runs of G(C)'s characteristic of the recombination-deficient sites. The most variable positions appear to be positions 8, 7 and 4 on both the left and right end of the site, while the most conserved positions seem to be: 1, 2L, to a lesser extent 2R, and the positions 5 and 6. Conservation of AT at the positions 1R and 1L is unsurprising, bearing in mind that in this assay Tn3 site I variants are expected to recombine with the wild-type Tn3 site I which also has a central AT dinucleotide. Conserving the AT at positions 1R and 1L avoids mismatches in the centre following the recombination reaction, which are known to be unfavourable (Hatfull & Grindley, 1988). Mismatches appear however not to be fatal for recombination as can be seen from the two isolated Tn3 site I variants marked with a "*" in the Fig.3.14. These sequences with a central TT dinucleotide could have arisen from a repaired mismatch in the centre after the recombinant plasmid had been transformed. The adenine base at the

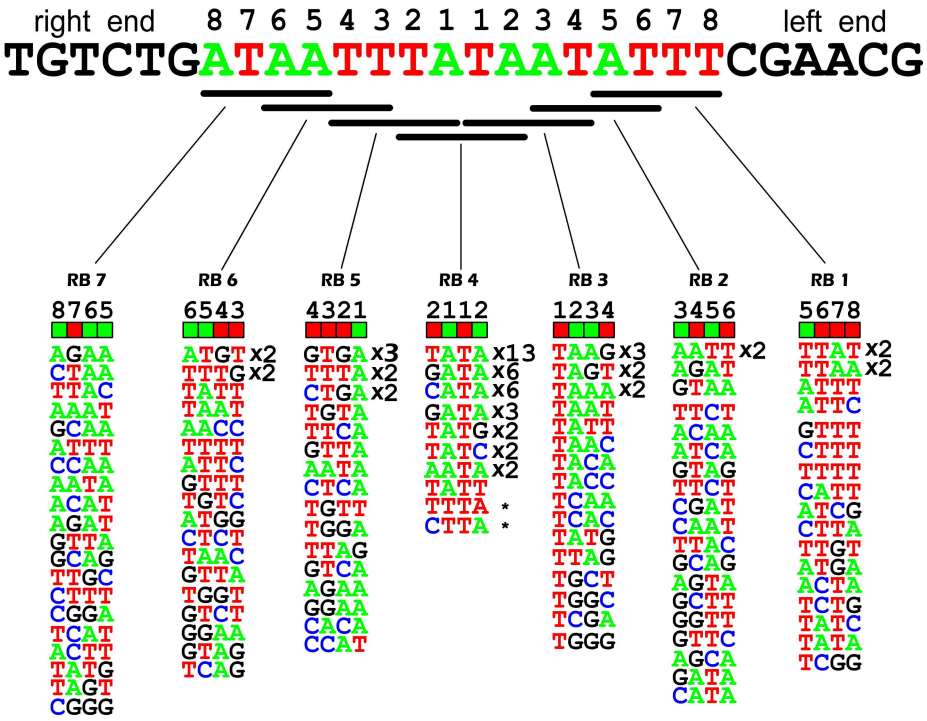


Figure 3.14-Sequences of recombination-proficient Tn3 site I variants selected *in vitro* using RB GalK Lib 1-7 substrate libraries (selection Strategy A, Section 3.5). In the interest of clarity, for each isolated variant only the sequence that differs from the wild-type Tn3 site I is shown. The library that the variant sequence was isolated from is marked on top of the sequence alignment. Coloured squares underneath the library name represent the wild-type Tn3 site I sequence at those positions. The sequences marked with an "*" have a central TT dinucleotide that could have arisen from a repaired mismatch in the centre after the recombinant plasmid had been transformed. Position numbering is from the centre, as in the previous figures.

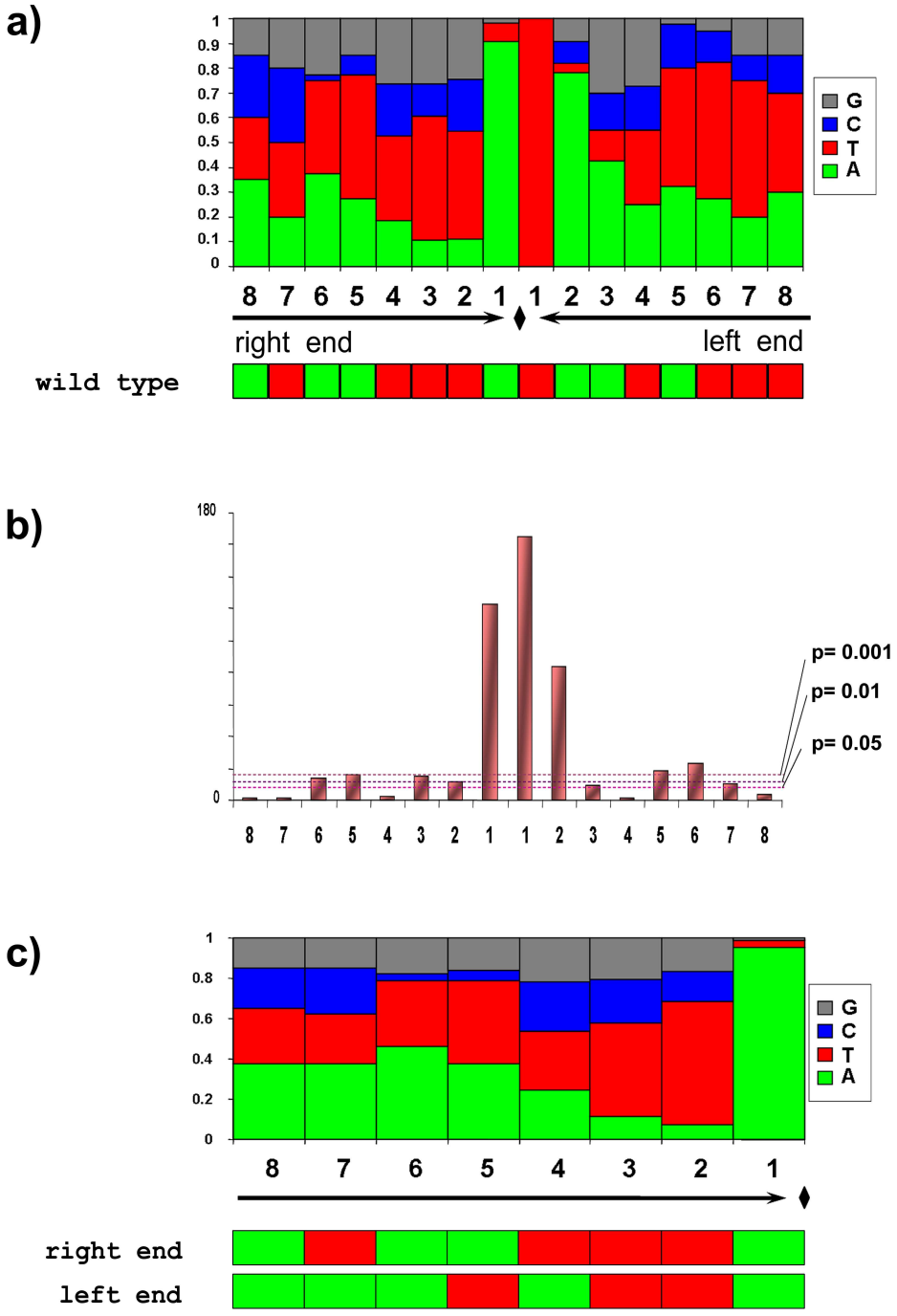


Figure 3.15- a) A histogram showing the relative base frequencies in the recombination-proficient sites presented in the Fig.3.14. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A bar chart showing the $\chi 2$ test values for each of the central 16 site I positions, with the null hypothesis that the DNA base distribution is essentially random, i.e. 25% for each base. The dashed lines show the significance threshold cut-off point as indicated on the right of the bar chart. **c)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

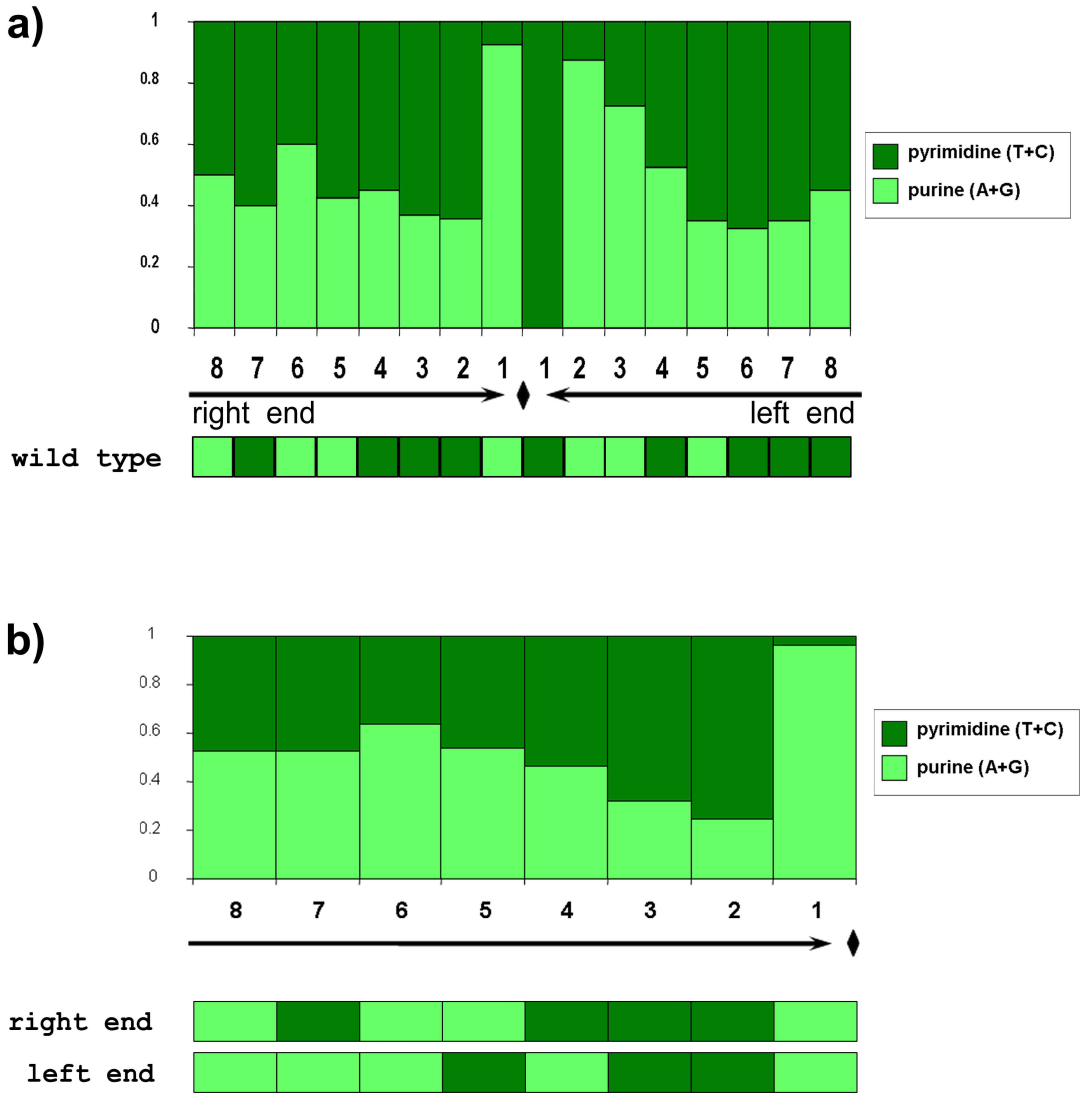


Figure 3.16- a) A histogram showing the purine (light green) and pyrimidine (dark green) frequencies in the recombination-proficient sites presented in the Fig.3.14. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

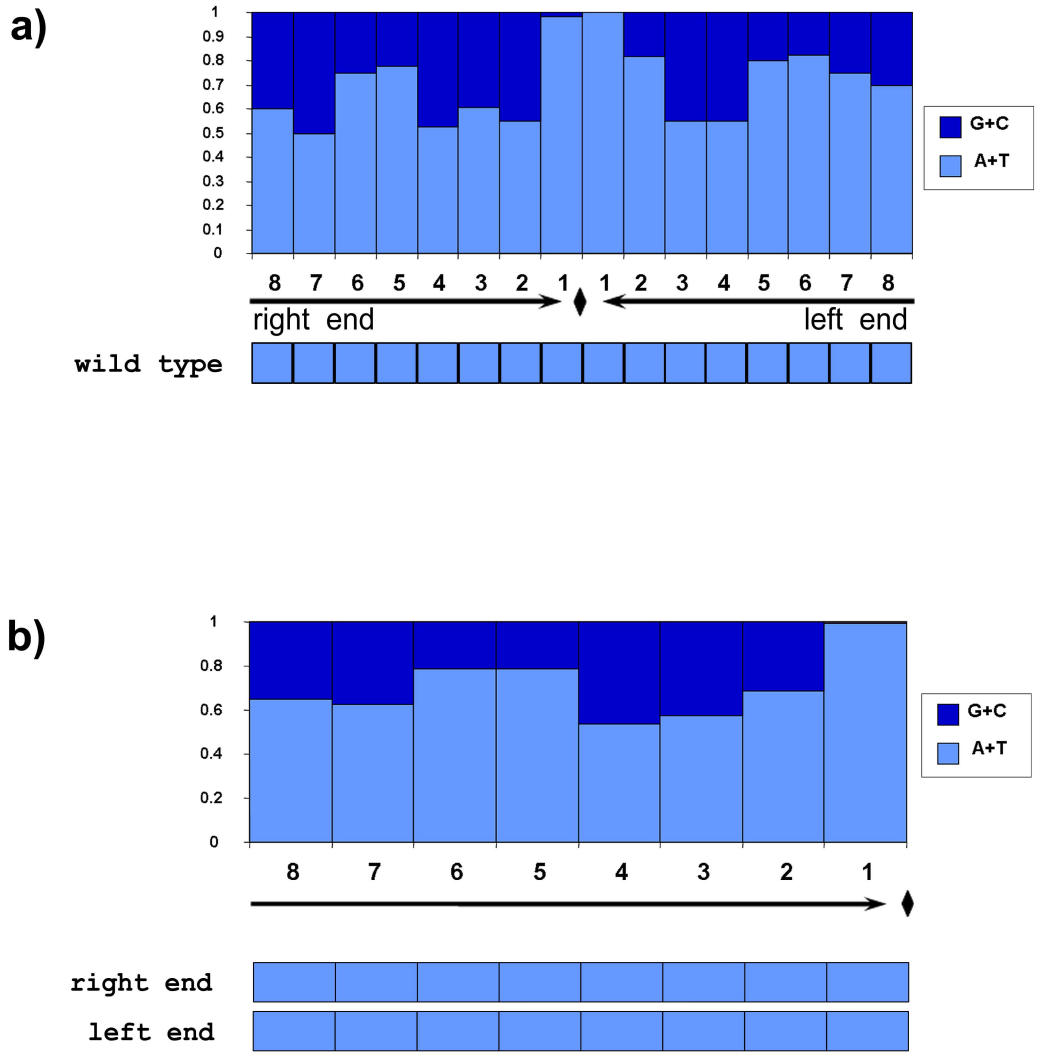


Figure 3.17- a) A histogram showing the A+T (light blue) and G+C (dark blue) frequencies in the recombination-proficient sites presented in the Fig.3.14. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

position 2L seems to be better conserved than its mirror image thymine at the position 2R. In most cases a pyrimidine-purine (Y-R) base step between positions 1 and 2 is preserved on both sides of the site, yet in the instances where it is not, it is always missing only on one side. Positions 5 and 6 are in vast majority of cases an A or a T. If one of them is changed to a G or a C, the other position, apart from very few exceptions, is not. There seems to be a slight preference for position 5 to be a thymine which is extending the run of T's. A χ^2 test was performed for each of the central 16 Tn3 site I positions with the null hypothesis that the DNA base distribution is essentially random, i.e. 25% for each base (Fig 3.15). This test confirms that the strongest selection is seen (in a descending order) at the positions 1L, 1R, 2L, 6L, 5L, 5R, 3R and 6R. The probability that the base frequencies observed at these positions are due to the chance sampling of an equiprobable base set is less than 1%. For positions 1L, 1R, 2L, 6L, 5L this probability is even lower (< 0.1%).

3.6 Strategy B - Plasmid fusion-based selection

In this strategy, the Tn3 site I's capable of recombination were to be selected using a fusion reaction between the seven random library plasmids RB Lib 1-7 (see Section 3.3) and a plasmid pMP47. Plasmid pMP47 has the same pMTL23 backbone as the random library plasmids, but it carries a kanamycin resistance gene instead of a gene for ampicillin resistance, and a wild-type Tn3 site I. The fusion reaction is in essence a reverse of the resolution reaction. The ampicillin-resistance library plasmid should recombine *in vitro*, using NM resolvase, with the kanamycin-resistance plasmid carrying the wild-type Tn3 site I, giving a fusion product with both antibiotic resistances, allowing for selection (see Fig.3.18). The ratio of pMP47 to the library plasmid in the reactions was envisaged to be 1 : 4, in order to minimise recombination between the copies of pMP47 molecule itself. Although these pMP47 fusions with itself could not be selected under ampicillin and kanamycin selection, they are undesirable as they reduce the overall efficiency of the reaction. The original sequence of the mutant crossover site could be reconstructed by sequencing both newly formed recombinant crossover sites in a single product plasmid.

A "proof of concept" reaction between pMP47 and pCO1, both containing the wild-type Tn3 site I, was successfully performed. However, due to time running out, this strategy has not been pursued any further.

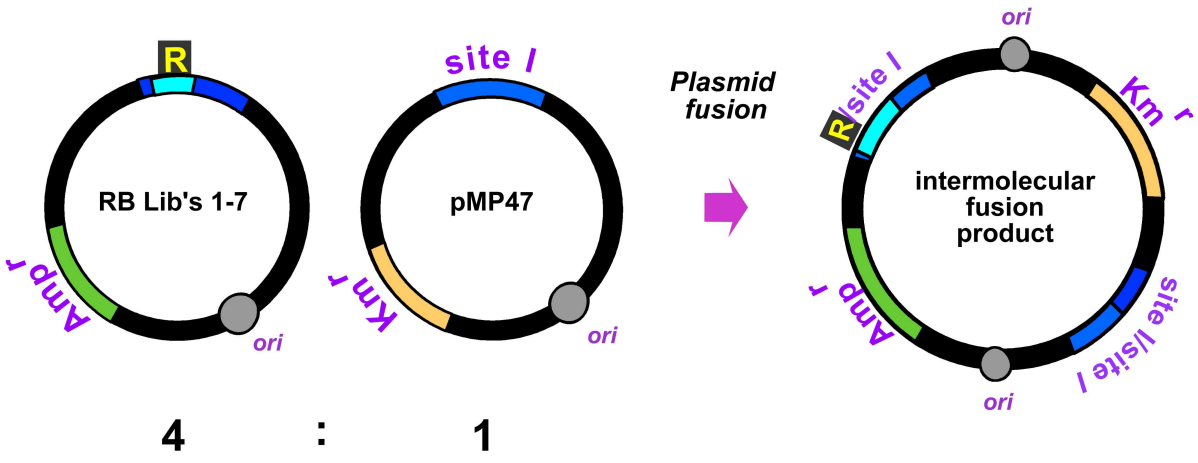


Figure 3.18-A diagrammatic representation (not to scale) of the plasmid fusion reaction which was to form a basis of the Tn3 site I variant selection Strategy B. Intermolecular recombination between the random Tn3 site I variant carried on single site random library plasmid (Amp^r) and a wild-type Tn3 site I found on pMP47 (kanamycin resistance carrying pCO1 analogue) creates a fusion product that carries both ampicillin and kanamycin resistance allowing this product to be selected. The ratio between the single site library plasmids (RB Lib 1-7) and pMP47 should be 4:1 in order to minimise intramolecular recombination between copies of pMP47. The mutated 4 bp block is shown in cyan, highlighted with a yellow "R" on a grey background.

3.7 Strategy C - Plasmid multimer resolution-based selection

The selections discussed in the previous two sections were based on recombination between an unmodified Tn3 site I and a mutant Tn3 site I. This kind of arrangement is likely to be more permissive and selection less stringent as unfavourable changes in one half-site could be compensated for by the wild-type DNA-protein interactions in the remaining three half sites (Hatfull *et al.,* 1988). This approach was informative but it faced some intrinsic limitations, especially at the very centre of the site I. As the wild type Tn3 site I has an AT dinucleotide centre, recombination with a modified site I would, in fifteen out of sixteen times, result in a mismatch. A mismatch in the centre of the site I impairs the religation step of the recombination reaction (McIlwraith *et al*., 1997), and would likely cause the vast majority of site I's with mutated centres to be selected against. As the bases in the central dinucleotide are not directly contacted by the resolvase protein according to the structural evidence (Yang & Steitz, 1995), it seems reasonable to suggest that other central dinucleotides could be as efficient in recombination as AT (Hatfull & Grindley, 1988). Also, it is evident from the alignment of site I's associated with other Tn3 resolvase-like recombinases that alternative central dinucleotides are possible (Fig 3.1). As the wild-type site I's are partially palindromic (Fig 3.1), it is possible that introducing changes at only one side of the site, without the corresponding changes in the other, might not be testing the sequence variability to its full capacity, and that a greater sequence variation is permissible if the introduced changes are symmetrical. Additionally it could be argued that symmetrical negative changes would make the selection more stringent as there would be no possibility of compensatory effect from the wild-type protein-DNA contacts on the remaining half-sites. To test these hypotheses, it was necessary to create resolution substrate plasmids with two mutated crossover sites, containing exactly the same changes in each. To create libraries of such substrates by cloning would be unfeasible. To sidestep this issue, an alternative selection strategy based on plasmid multimer resolution was devised.

The strategy employed is summarised in Figure 3.19. The protocol relied on generating plasmid multimers, using the single site random block libraries RB Lib 1-7 (Section 3.3), in order to create resolution substrate plasmids containing two (or more) mutated Tn3 site I variants, each containing exactly the same changes. Resolution substrates so created would then be assayed for their recombination proficiency *in vitro* using NM resolvase. Resolution of multimer plasmids by NM resolvase should result in plasmid monomers. To

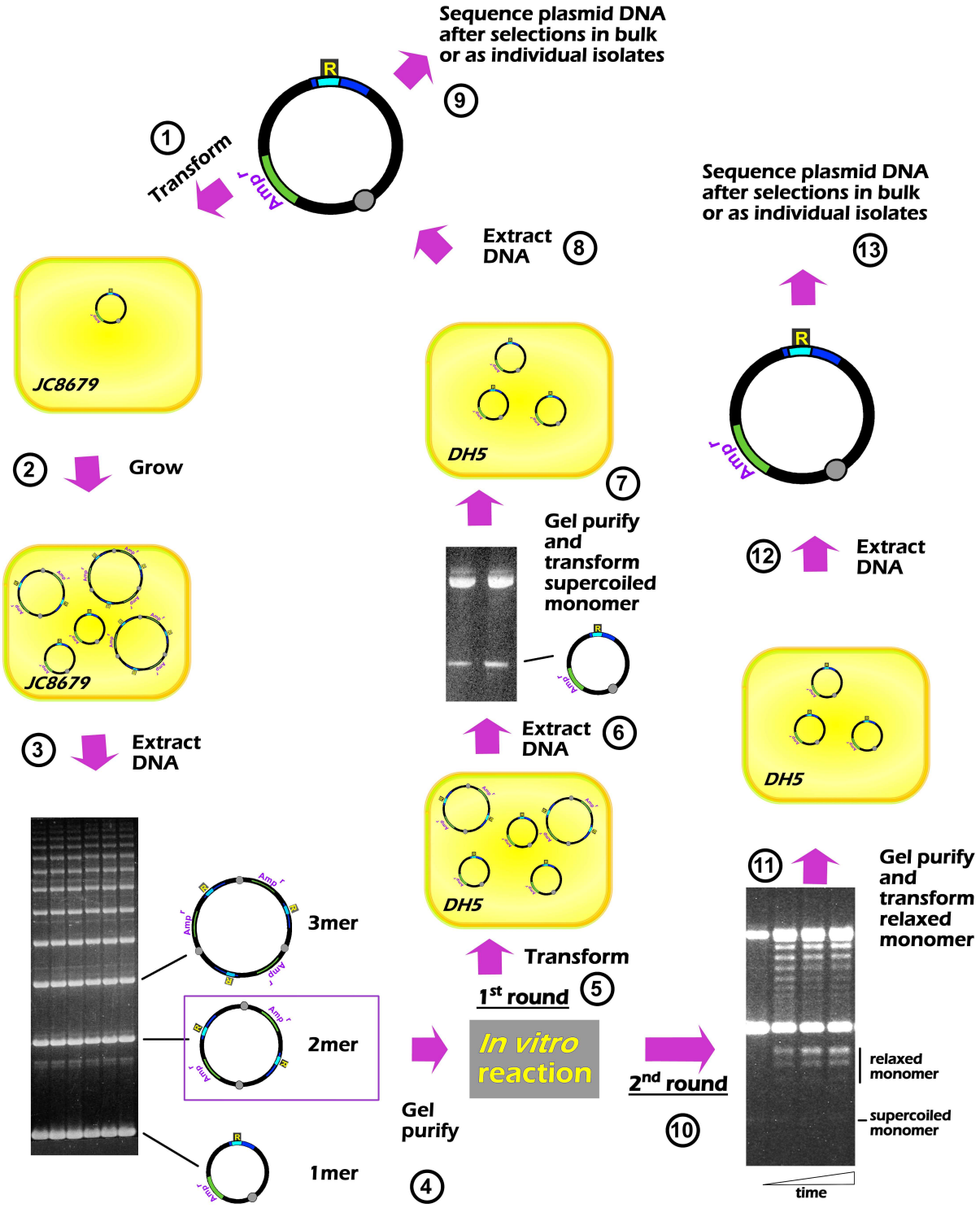


Figure 3.19-A diagrammatic representation (not to scale) of the Tn3 site I variant selection Strategy C. Step 1-Transform the single site random block library monomer plasmid into JC8679. Step 2-Grow JC8679 cells. Step 3-Extract multimerised plasmid DNA and separate using a 1% TAE agarose gel. Step 4-Cut out plasmid dimer and extract out of the agarose gel. Step 5- In vitro recombination reaction using the plasmid dimer followed by transformation into DH5 cells. *The legend continued on the following page...*

Figure 3.19-*The legend continued from the previous page..* Step 6-Following amplification of plasmid DNA in DH5 cells, plasmid DNA is extracted and separated on a 1% TAE agarose gel. Step 7- Cut out supercoiled plasmid monomer extract the DNA out of agarose gel and transform into DH5 cells. Step 8-After amplification in DH5 cells plasmid DNA is extracted. Step 9- Supercoiled monomer plasmid DNA sequenced in bulk or used for a second round of selection. Step 10-Second round of selection, the products of the in vitro recombination reactions are separated on a 1% TAE agarose gel. Step 11- relaxed plasmid monomer topoisomers are cut out of the agarose gel, DNA extracted and used to transform DH5 cells. Step 12-Following the amplification in DH5 cells plasmid monomer DNA is extracted (individually or in bulk). Step 13- Extracted plasmid DNA monomers are sequenced in bulk or as individual isolates.

generate plasmid multimers, *E. coli* strain JC8679 was to be employed. This strain carries mutations in the *sbc*A, *rec*B and *rec*C genes, causing DNA recombination defects (Gillen *et al.,* 1981). These defects lead to the accumulation of multimer forms of any episome that is maintained in these cells.

The seven random block libraries (RB Lib 1-7) were used to transform electro-competent JC8679 *E. coli* cells (Fig 3.19, step 1). Plasmid pCO1 was used as a positive control. pCO1 contains a single wild-type Tn3 site I cloned into the pMTL23 backbone. After transformation, cells were spread on selective L-agar plates and grown overnight (Fig 3.19, step 2). Cells were grown on plates rather than in liquid culture to avoid putting selective pressure on the libraries that could cause a loss in diversity. Bacterial colonies were scraped off the plates and the plasmid DNA was isolated (Fig 3.19, step 3). The number of scraped colonies was about 15 000 for each of the libraries, which is sufficient to maintain the library diversity. Separation of plasmid multimers of different sizes was done by agarose gel electrophoresis. Plasmid dimer, trimer and tetramer were cut out of the gel and purified (Fig 3.19, step 4). To obtain a sufficient quantity of the plasmid DNA for subsequent resolvase reactions, separated multimers were amplified by transformation of recombination-deficient electro-competent *E. coli* strain DH5. Around 15 000 DH5 colonies were scraped off selective L-agar plates for each of the multimer libraries, and the plasmid DNA was prepared using the large scale method (Section 2.10.2).

Analogously to the Strategy A selections (Section 3.5), to determine the shortest time needed for adequate resolution, and the most suitable multimer form to use as the substrate, a time course experiment was performed. After the reactions were terminated, a portion of the reaction mix from each point in the time course was used to transform electro-competent DH5 cells. After transformation and overnight culture on selective L-agar plates, plasmid DNA was isolated from these cells and separated by agarose gel electrophoresis (Fig 3.20). The biggest source of error in the multimer resolution experiments comes from contaminating monomer in the multimer preparation. The bigger the size of the multimer the less likely is the direct contamination with monomer. However, larger multimers might require more than one round of recombination to be resolved to a monomer.

It was decided to use the dimer as the preferred form of multimer substrate. The monomer contamination level was found to be sufficiently low, and as only a single recombination round is required to generate monomers, dimer substrates were deemed superior to trimer. Considering that most substrates with one mutant and one wild-type site I were

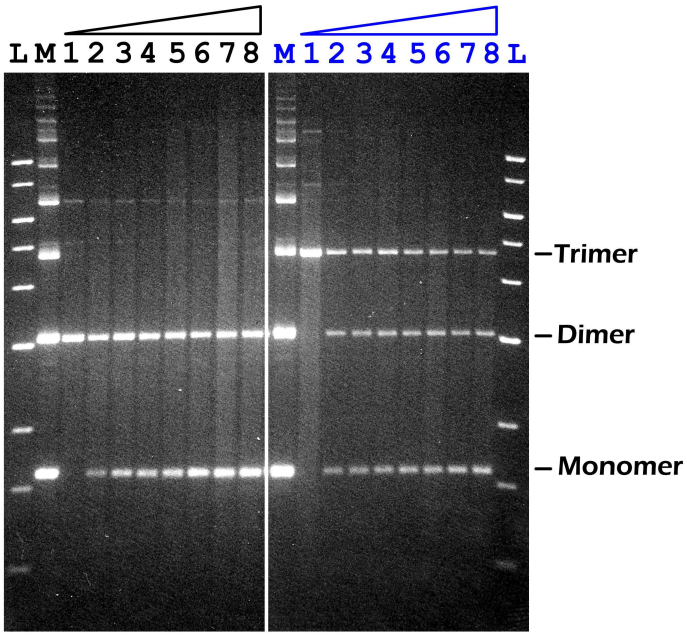


Figure 3.20-Products of the NM resolvase-catalysed pCO1 multimer resolution, post transformation. Samples of in vitro resolution reactions stopped by heat after 0 min. (1), 0.5 min. (2), 1 min. (3), 2 min. (4), 5 min. (5), 15 min. (6), 30 min. (7) and 60 min. (8), using either pCO1 dimer (left panel, black annotation) or pCO1 trimer (right panel, blue annotation) were transformed into recombination deficient DH5 cells and the amplified plasmid DNA isolated and separated using 1% TAE agarose gel. For protein and DNA concentration and buffer composition used for the in vitro reactions see Section 2.27. Lane marked M is a ladder of pCO1 multimers, while the lane marked L is the Invitrogen 1 kb ladder.

recombined slower than the substrates with two wild-type site I's (Section 3.5) it was decided to perform two subsequent selection rounds (Fig. 3.19). The first round (Fig 3.19, steps 1-9) was envisaged as the "enrichment" round, with a sufficiently long reaction time to provide a chance for any crossover sites that are able to recombine to do so. In the second round (Fig 3.19, step 1-4 and 10-13) a more stringent selection, (that is, a much shorter reaction time), was to be employed to select for the most efficient recombination sites.

For the first selection round, dimers of the seven random block libraries were treated with NM resolvase for 1 hour. The reaction was terminated, and a portion of the reaction mix was used to transform electro-competent DH5 cells (Fig 3.19, step 5). The plasmid DNA was isolated from ~15 000 DH5 colonies for each of the libraries and separated by agarose gel electrophoresis (Fig 3.19, step 6). Plasmid monomer was cut out of the gel, extracted and amplified in DH5 strain (Fig 3.19, step 7). The plasmid DNA was then again isolated from ~15 000 DH5 colonies (Fig 3.19, step 8) and used to transform electro-competent JC8679 (Fig 3.19, step 1) for a second round of multimerisation. This first round monomer DNA was also sequenced in bulk (Fig 3.19, step 9) using the universal -43 primer (Fig. 3.21).

For the second round, multimers were isolated and separated as described previously. Dimers were amplified in DH5 cells as described above. Dimer DNA was then treated with NM resolvase for 5 minutes. The recombination reaction was terminated, and the reaction products were separated using agarose gel electrophoresis (Fig 3.19, step 10). The ladder of relaxed monomer topoisomers was cut out of the gel and extracted (Fig 3.19, step 11). The topoisomers were used rather than the supercoiled monomer, as the topoisomerase activity was evidence that these molecules were substrates for resolvase, whereas the non-relaxed supercoiled monomer might still contain contaminant non-substrate monomers. The gel-extracted monomer topoisomers were used to transform electro-competent DH5 cells. Monomer plasmid DNA was purified (Fig 3.19, step 12) from ~15 000 DH5 colonies for each of the random block libraries using the large scale method (Section 2.10.2). This DNA was sequenced in bulk (Fig 3.19, step 13) using the universal -43 primer (Fig. 3.21). Apart from the bulk sequencing (summarised in Fig. 3.22), plasmid DNA from twenty randomly selected DH5 colonies from each library was sequenced individually (Fig 3.19, step 13). The sequences of these selected monomers are shown in Fig. 3.23. Sequence data from all the library selections were pooled together, and histograms showing the DNA base distribution for each of the 16 central site I positions were made (Figs. 3.24-3.26).

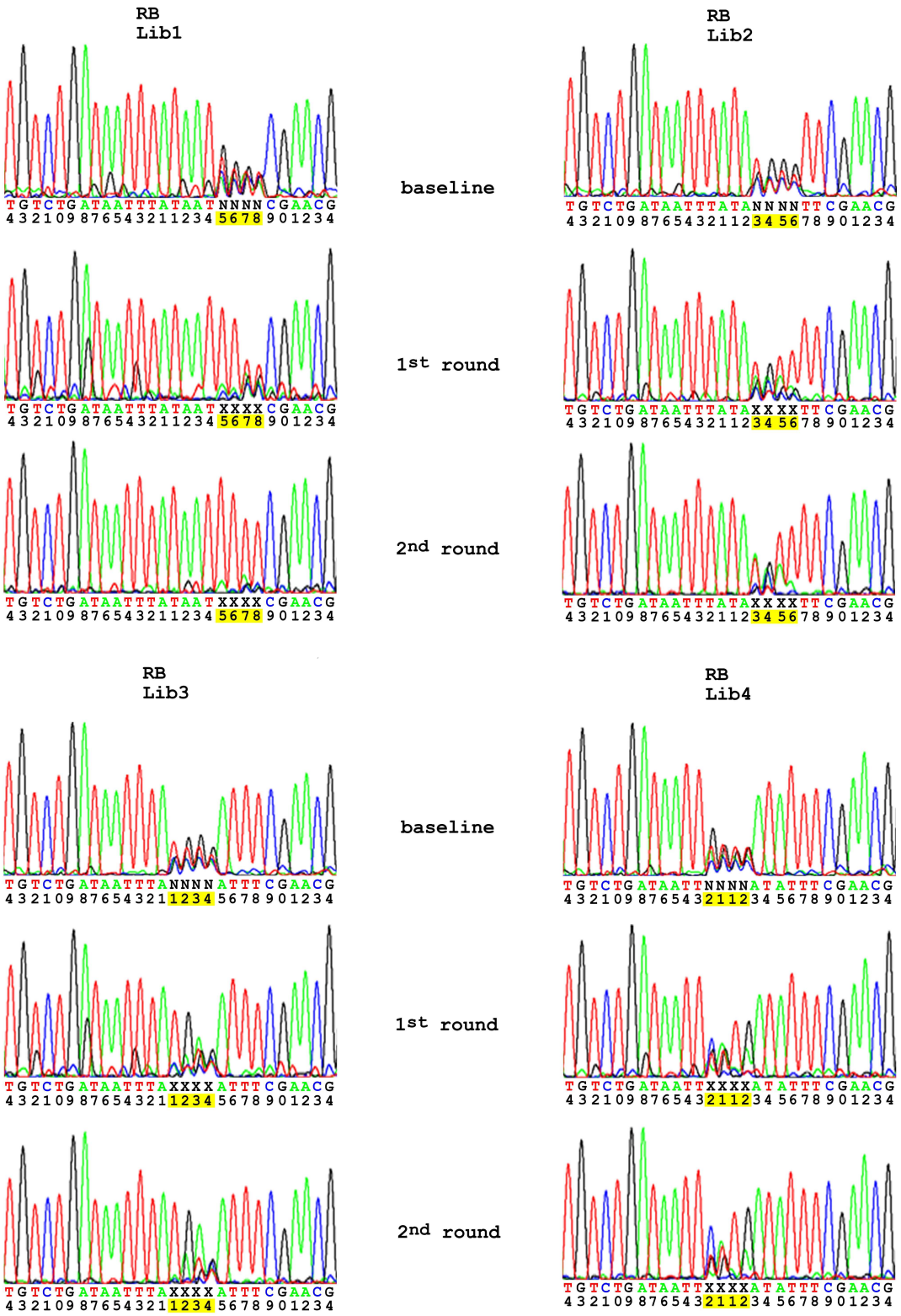


Figure 3.21- Bulk sequencing traces of mutant Tn3 site I variants obtained in selections with NM resolvase using the random block libraries RB 1-7 *Continued on the next page...*

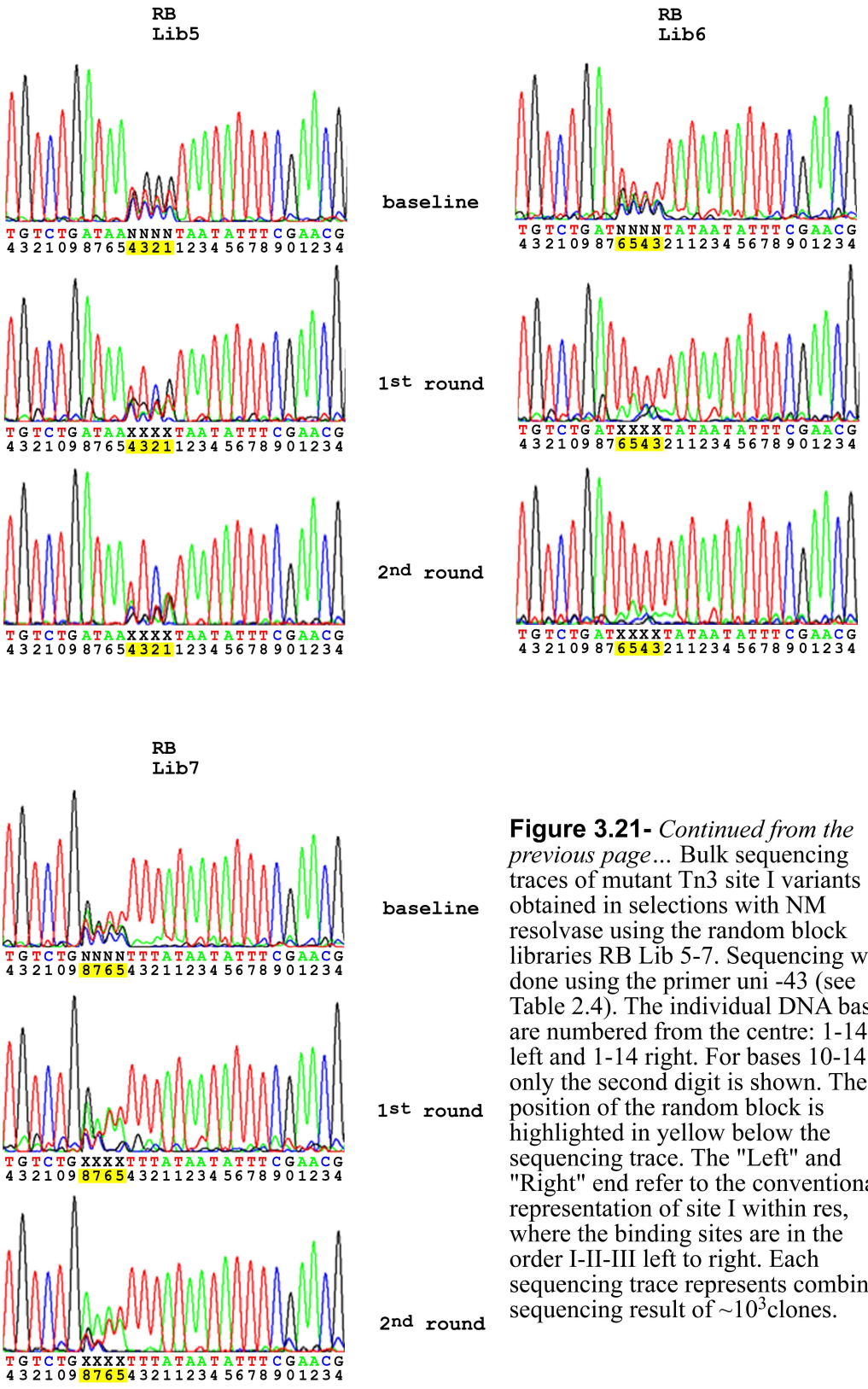


Figure 3.21- *Continued from the previous page…* Bulk sequencing traces of mutant Tn3 site I variants obtained in selections with NM resolvase using the random block libraries RB Lib 5-7. Sequencing was done using the primer uni -43 (see Table 2.4). The individual DNA bases are numbered from the centre: 1-14 left and 1-14 right. For bases 10-14 only the second digit is shown. The position of the random block is highlighted in yellow below the sequencing trace. The "Left" and "Right" end refer to the conventional representation of site I within res, where the binding sites are in the order I-II-III left to right. Each sequencing trace represents combined sequencing result of ~10^3clones.

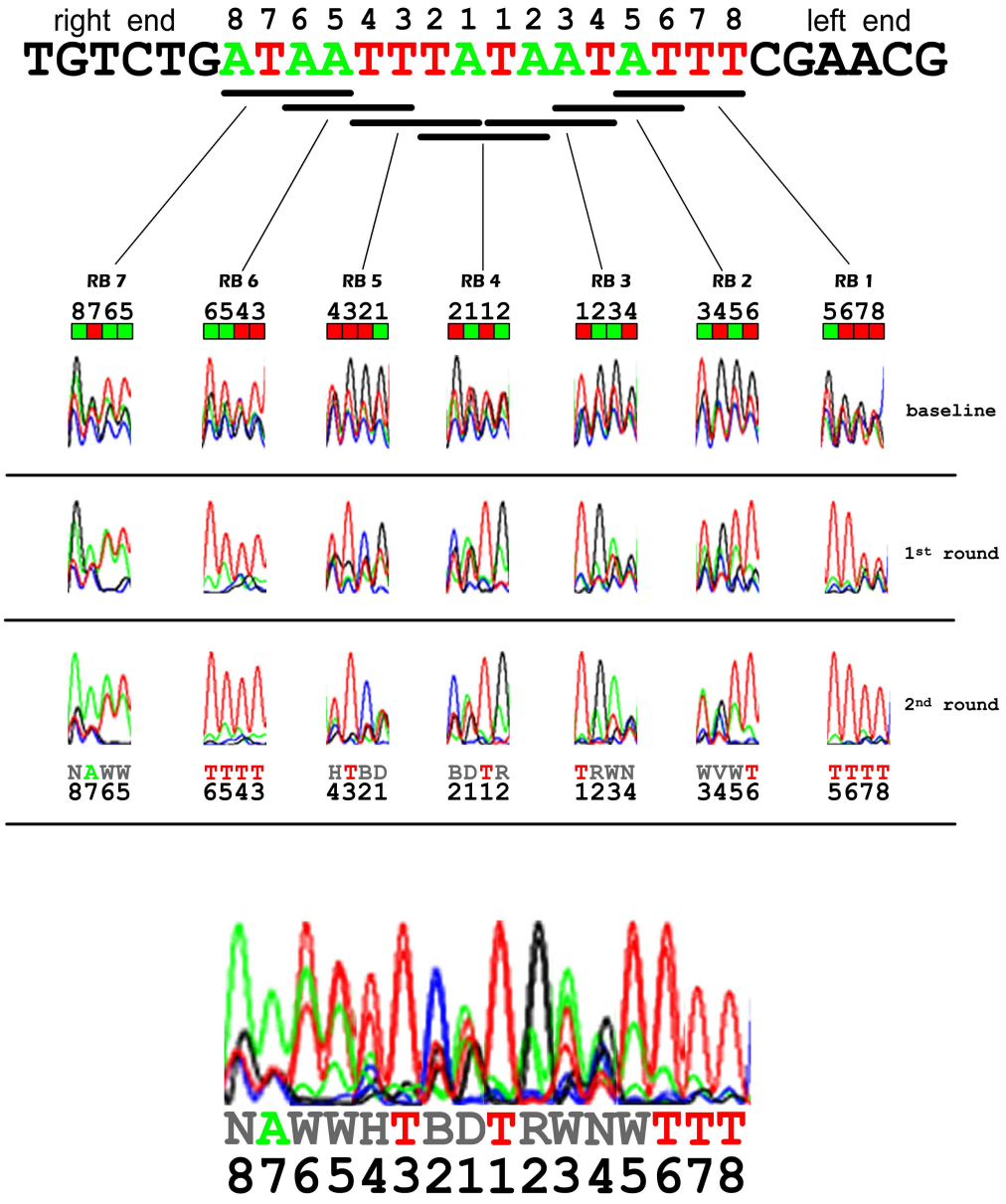


Figure 3.22- A summary of bulk sequencing data shown in Fig. 3.21. In the interest of clarity, for each library only the sequencing traces of the random block regions are shown. The library that the variant sequences were isolated from is marked on top of the sequence alignment. Coloured squares underneath the library name represent the wild-type Tn3 site I sequence at those positions. Underneath the bulk sequencing traces from the 2nd round of selections, the sequence of the selected blocks of four is given using the extended IUPAC DNA notation (see Abbreviations). At the foot of the figure, sequencing traces are merged together and the sequence of the whole Tn3 site I so derived is given using the extended IUPAC DNA notation. Position numbering is from the centre, as in the previous figures

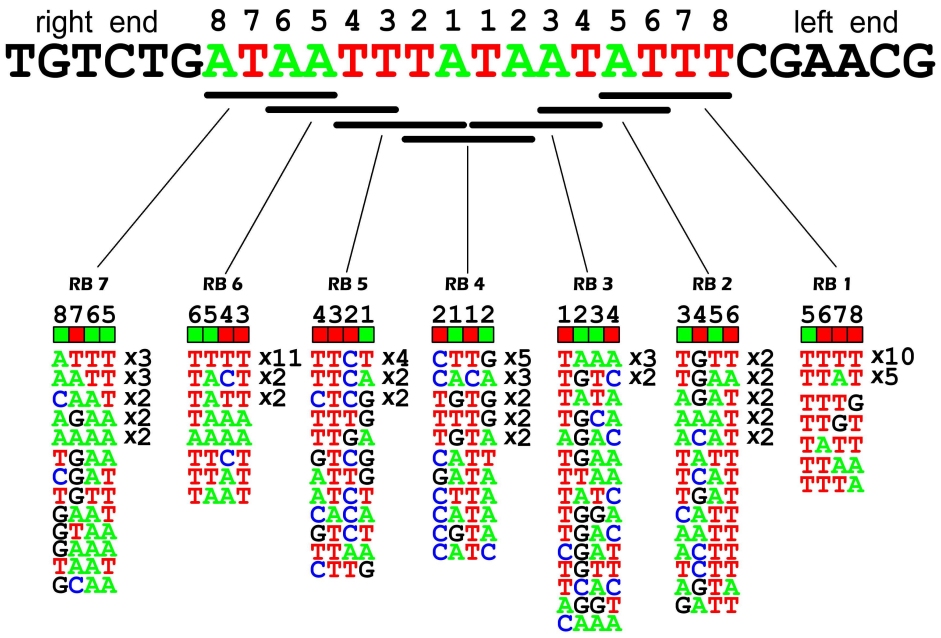


Figure 3.23- Sequences of recombination-proficient Tn3 site I variants selected *in vitro* using the selection Strategy C (Section 3.7). In the interest of clarity, for each isolated variant only the sequence that differs from the wild-type Tn3 site I is shown. The library that the variant sequence was isolated from is marked on top of the sequence alignment. Coloured squares underneath the library name represent the wild-type Tn3 site I sequence at those positions. Position numbering is from the centre, as in the previous figures.

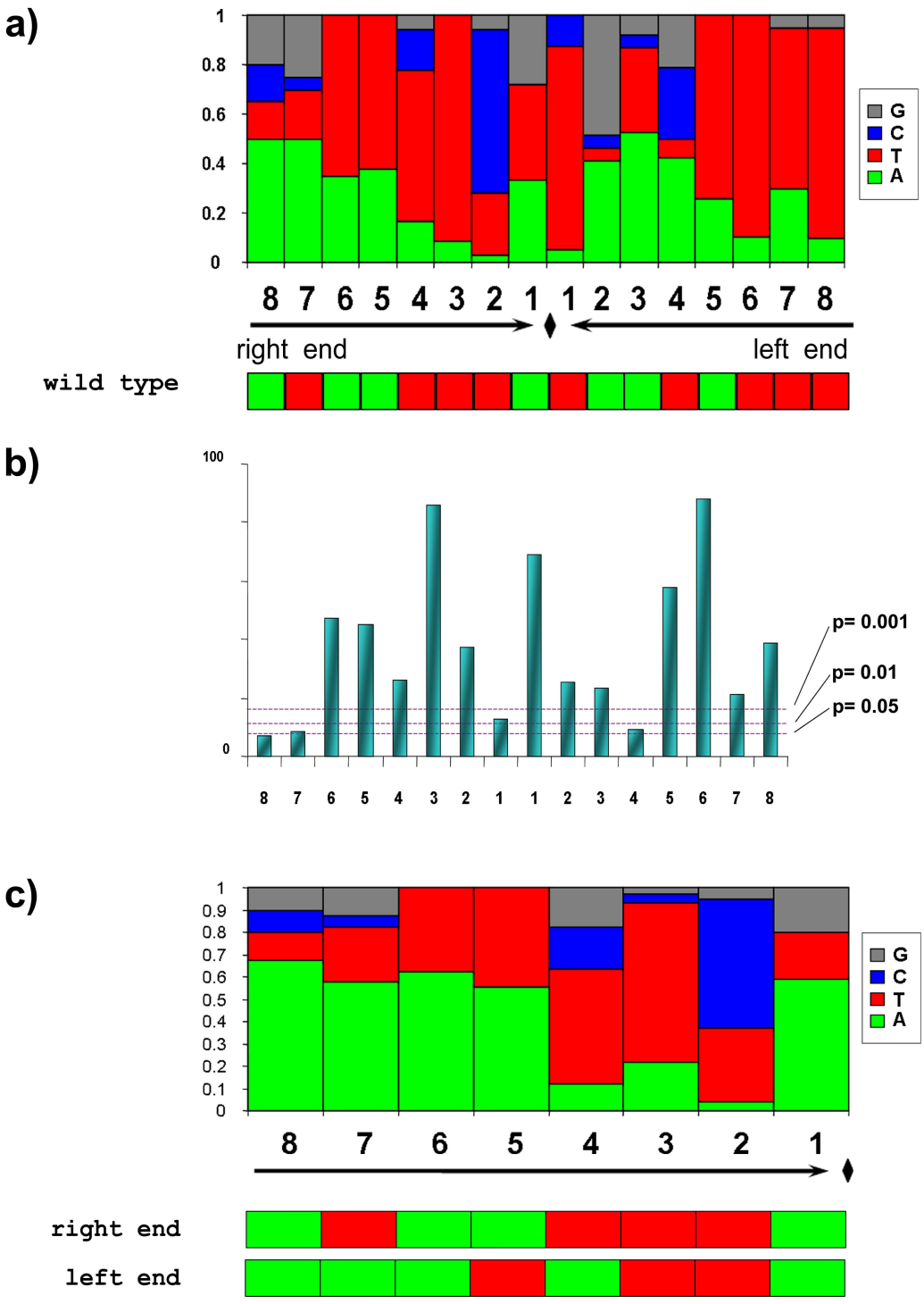


Figure 3.24- a) A histogram showing the relative base frequencies in the recombination-proficient sites presented in the Fig.3.23. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A bar chart showing the $\chi 2$ test values for each of the central 16 site I positions with the null hypothesis that the DNA base distribution is essentially random, i.e. 25% for each base. The dashed lines show the significance threshold cut-off point as indicated on the right of the bar chart. **c)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

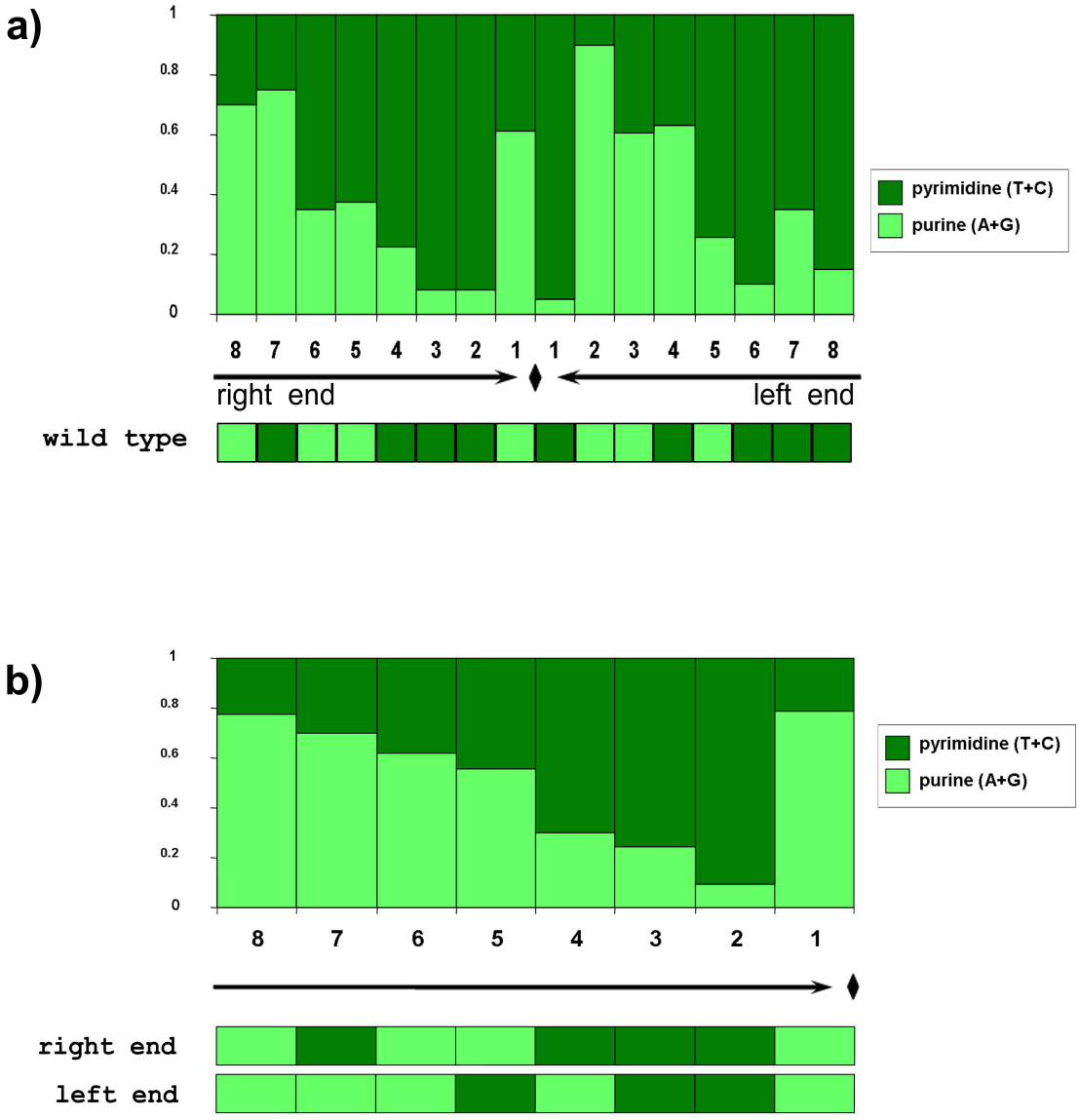


Figure 3.25- a) A histogram showing the purine (light green) and pyrimidine (dark green) frequencies in the recombination-proficient sites presented in the Fig.3.23. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

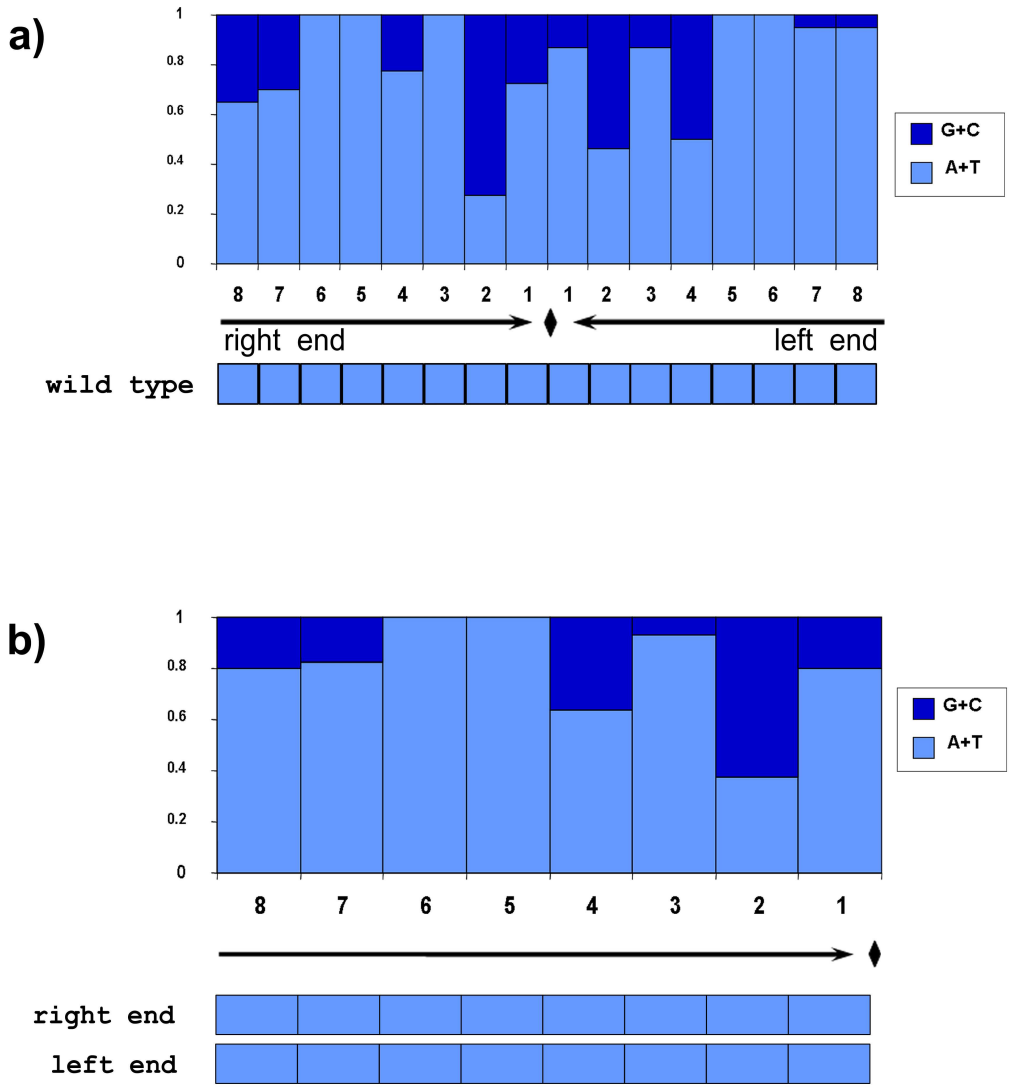


Figure 3.26- **a)** A histogram showing the A+T (light blue) and G+C (dark blue) frequencies in the recombination-proficient sites presented in the Fig.3.23. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the histogram for comparison. **b)** A histogram showing the summary of the same data as a) but symmetrised with the respect to the right end of Tn3 site I. Coloured bars under the histogram represent the wild-type sequence of the right end (positions 1-8R) of the Tn3 site I, and the inverted sequence of the left end (positions 1-8L) for comparison.

In general the results obtained using this selection strategy correspond very well to the findings described in Section 3.5. In contrast to Strategy A however, bulk sequencing of the first and second round selected monomers proved very successful, most likely due to the template plasmid DNA being of higher quality. Sequencing in bulk gives a qualitative overview of base frequencies at each position, which is an average of ~ 15 000 individual Tn3 site I variant isolates. Although peak heights representing different DNA bases at different base positions can not be easily compared, comparing the peak heights at the same position gives a good idea of the relative representation of a particular DNA base at that position in relation to the other three possible bases. In this experiment the bulk sequencing result very graphically illustrates the selection taking place, especially at positions already identified in the previous screen (Section 3.5) such as 1L, 2L, 5, 6 and 3R. A striking feature of the bulk sequencing is just how superimposable the results are for the same base positions obtained in independent selections using random libraries containing overlapping random blocks. This is especially true in the centre of the site where the result for positions 1 and 2 is virtually identical for libraries containing random blocks 5, 4 and 3.

Selections using libraries containing random blocks 6 and 7 or 1 and 2 reveal a possible "context effect"; when the positions 5R and 6R are selected from the library containing random block 7, both A and T appear to be acceptable at these positions, yet the same positions when selected from the random block 6 are almost exclusively a T. If both 5R and 6R are changed into a T a run of 6 T's is made. Indeed, when the individual second round Tn3 site I variants from library containing RB 6 were sequenced, 11 out of 20 sequenced variants had a T at both 5R and 6R. Even more strikingly, all 20 individual Tn3 site I variants isolated from the library containing RB 1 had a T at position 5L, and 10 out of those also had a T at 6L, creating a run of 5 T's. When the library containing changes in RB 2 was used, position 5L is an A in half of the individually sequenced second round isolates. This propensity for creating runs of T's has been noted in the selections described in the previous section, but in this assay it appears to be far more pronounced. Also, while in the selections using the Strategy A, G and C bases were found to be tolerated to a degree at positions 5 and 6, these bases were found to be absolutely not tolerated in the plasmid multimer-based selections.

One of the main reasons for employing Strategy C was to allow the central dinucleotide to vary without having to worry about the issue with a mismatch in the centre of the site. As expected a number, of Tn3 site I variants with a central dinucleotide other than AT were selected, namely TT, GT and AC. Interestingly, central dinucleotide TT was also observed

in the selections using Strategy A, where it was most likely due to mismatch repair in DS941 cells post-transformation. Again a T at a position 1L was found to be better conserved than its corresponding mirror image A at a position 1R, with only 6 out of 40 Tn3 site I variants selected from libraries containing RB 4 or 3 having a different base at this position.

Just like in the Strategy A-selections a purine-pyrimidine step between positions 1 and 2 was found to be preserved at least at one side in all Tn3 site I variants selected, with most isolates having the purine-pyrimidine step on both sides of the centre of the site. Position 2 shows much greater degree of variability than was observed in the previous selections, but displays a similar impression of symmetry with 2R preferably being a C and almost never an A or a G, and position 2L favouring a G but very rarely a C or a T.

A thymine base at position 3R was much more reliably conserved than an A at the position 3L, which agrees well with the similar finding from the Strategy A selections. Finally, the selection at positions 7L and especially 8L seemed to be much more stringent than the selection at the positions 7R and 8R, with the wild-type base being preferred to other possible bases. This difference was much more pronounced in this set of results than in the Strategy A selections.

A χ^2 test was performed for each of the crossover positions with the null hypothesis that the DNA base distribution is essentially random; i.e. 25% for each base (Fig 3.24). The test shows that at 12 positions out of 16 the probability that the base frequencies observed at these positions are due to the chance sampling of equiprobable base set is less than 0.1%. The strongest selection is seen (in a descending order) at the positions 6L, 3R, 1L, 5L, 6R, 5R, 8L, 2R, 4R, 2L, 3L, and 7L.

3.8 Discussion

In the preceding sections, NM resolvase was used to select Tn3 site I variants that are either recombination-proficient or deficient in two independent assays, using libraries of recombination substrates containing sites in which every DNA base in the central 16 bp of the Tn3 site I was allowed to vary. In one assay (Strategy A), a recombination reaction was designed to take place between a wild-type Tn3 site I and a library of mutant site I variants, whereas in the other protocol (Strategy C) a recombination between two mutant site I variants with identical changes was assayed. In both strategies DNA bases in the centre of site I were changed in overlapping blocks of 4 bp, which was intended as a way to cover the possible sequence space in a manageable self-checking manner that could highlight

possible “context effects”. In this section results of these selections will be discussed and compared with the relevant previously published findings, and related to the available structural information in order to attempt to interpret the basis of Tn3 resolvase selectivity in the centre of site I at the molecular level.

Overall, Tn3 site I variant sequences selected as recombination proficient, bore a strong resemblance (Fig. 3.27) to the known site I sequences associated with the Tn3 resolvase family, affirming the validity of the selections. The greatest difference to the wild-type sequences was observed at position 2 in sites selected using Strategy C. Although this position appeared to be allowed to vary, a characteristic pyrimidine-purine (Y-R) base step found between this position and position 1 was found to be very well conserved, and in all cases present at least at one of the two position 2/position 1 junctions. In the crystal structure of a γδ resolvase dimer bound to site I (1gdt; Yang & Steitz, 1995), these Y-R base steps coincide with the positions of the scissile bonds (Fig. 3.28). They display the highest roll angles (40° and 20°) and are making a contact with the residue T126 (Fig. 3.29), causing partial unstacking, unwinding by ~10° and widening of the minor groove from the average 5.1 to 7.6 Å, which results in the overall DNA bend in site I. The observed conservation of the Y-R base step highlights the functional importance of this sequence feature, whereas the existence of recombination-proficient sites with only one Y-R base step conserved, might support the proposal of a non-concerted cleavage reaction mechanism (Yang & Steitz, 1995).

A number of different central dinucleotides apart from AT were selected including AC, GT and TT. AC and GT preserve a purine-pyrimidine (R-Y) base step that is seen in the wild-type AT, whilst this is not the case for TT. This central dinucleotide R-Y base step seems to be less important than the Y-R base step at the scissile bond positions. An alternative central dinucleotide TT has been shown to be tolerated reasonably well in the context of *res* using either Tn3 or γδ resolvase systems (Hatfull & Grindley, 1988).

Further away from the centre of the site, results obtained using the two independent selection methods were remarkably similar and self-consistent. However, it was found that the selection for any particular sequence was generally less stringent using Strategy A. A possible explanation for this is that the wild-type DNA-protein interactions in the remaining three half-sites that do not contain sequence changes might be acting in a compensatory fashion, and that therefore only the most inhibitory changes are likely to have a sufficient effect.

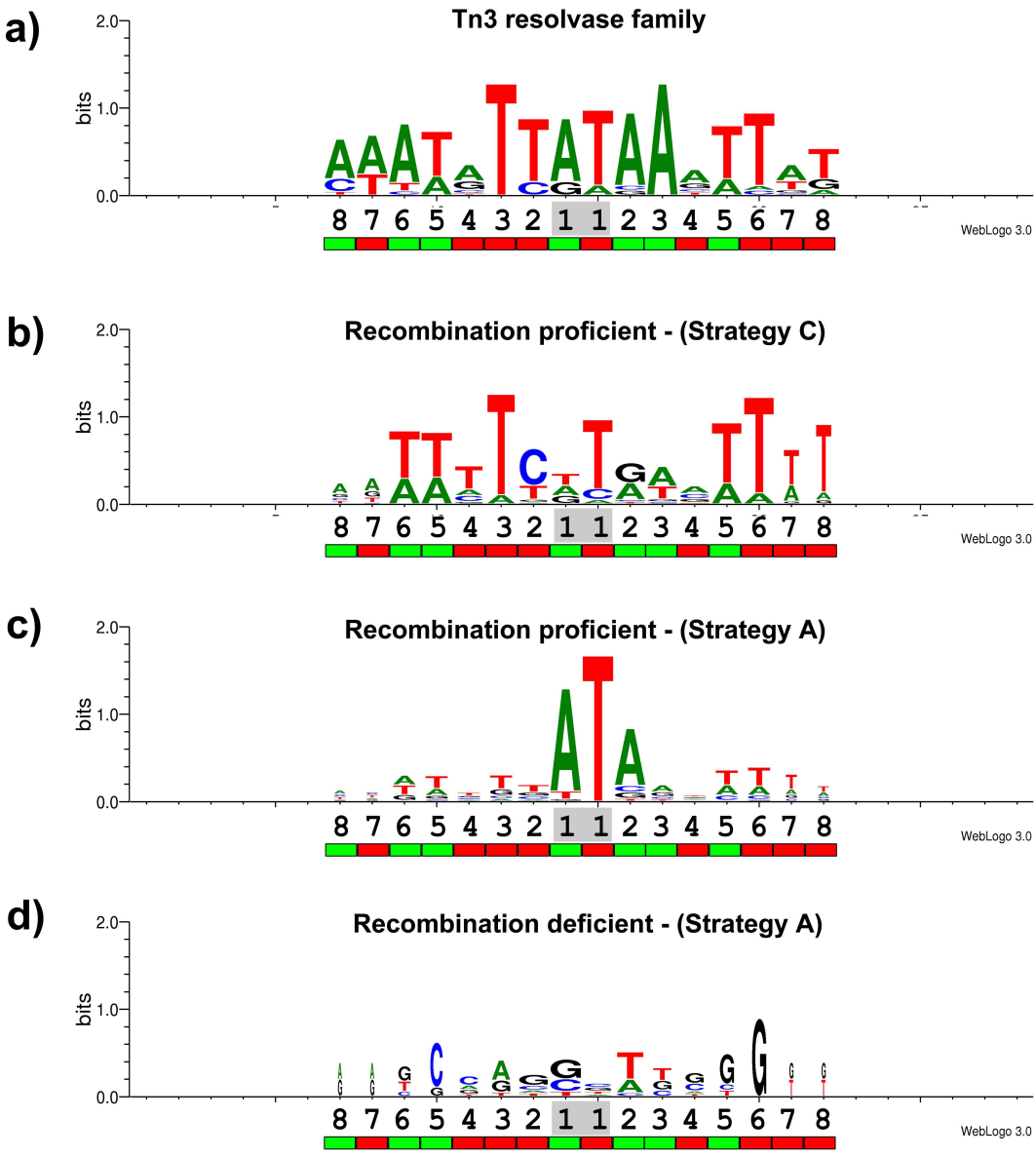


Figure 3.27- Sequence logos summarising the sequences of **a)** Tn3 like site I's (Fig. 3.1), **b)** recombination-proficient Tn3 site I variants selected using Strategy A (Fig. 3.14), **c)** recombination-proficient Tn3 site I variants selected using Strategy C (Fig 3.23), and **d)** recombination-deficient Tn3 site I variants selected using Strategy A (Fig. 3.9). Sequence logos are graphical representations of multiple sequence alignments of DNA bases (Schneider & Stephens, 1990). Each logo consists of stacks of symbols, one stack for each position in the centre of Tn3 site I. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each DNA base at that position. A coloured bar representing the wild-type Tn3 site I sequence (positions 1-8R and1-8L) is provided under the logo for comparison. Sequence logos were generated using the ¬WebLogo 3.0 server.

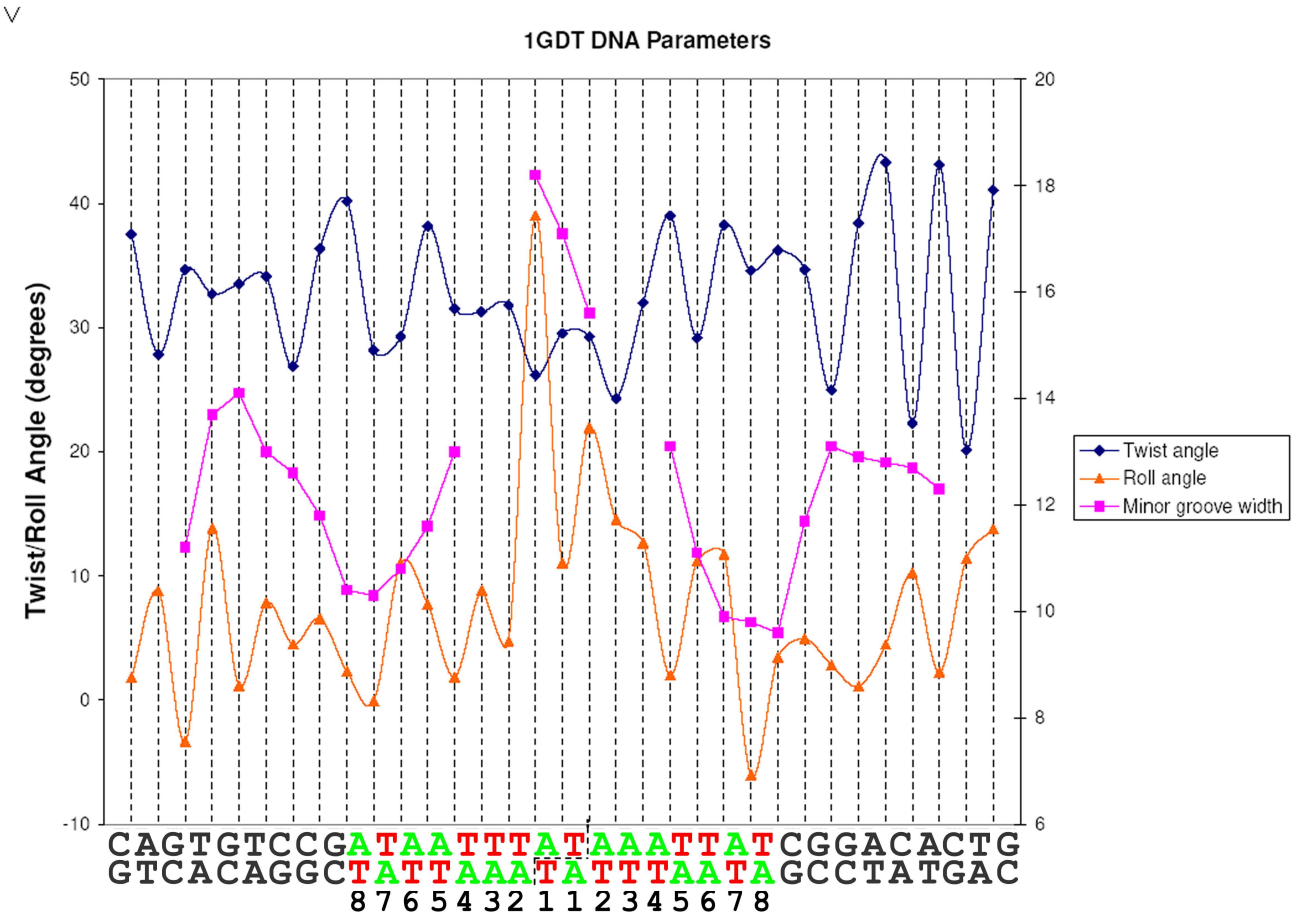


Figure 3.28- A graph showing the calculated DNA parameters for the symmetrised site I sequence in 1gdt structure (Yang & Steitz, 1995). Note that this sequence is slightly different to wild-type Tn3 site I sequence, however most of the differences are found outside the central 16 bp (coloured sequences). Parameters plotted are the twist angle (blue line), roll angle (orange line) and the minor groove width (magenta line).The numbering of bases is from the centre as in the previous figures.

The findings show that most positions can be changed without abolishing recombination if changed individually, but that accumulating several changes is detrimental. In general changes from an A to a T, or a T to an A were much better tolerated than substitutions of an A or a T with a G or a C, which is consistent with the recognition in the minor groove. This is consistent with the available γδ resolvase-DNA co-crystal structures which show extensive minor groove contacts between the extended arm region and bases at positions 1-8 (Figs. 3.29 and 3.30)(Yang & Steitz, 1995, Li *et al.,* 2005).

Certain positions were found to have a very strong preference for certain types of bases, especially when recombination between two mutant site I variants with identical changes was demanded. Generally the selection in the left half-site appeared to be more stringent and bases at key positions more faithfully conserved. Since the binding of the resolvase HtH DNA-binding domain to the outermost 6 bp of the left site I half-site is quite weak (Bednarz, 1990) it is possible that the contribution to binding and sequence recognition by the arm region-DNA contacts is far more crucial at this half-site. This suggestion seems plausible, considering that the individual contribution to binding of the similar arm region of the DNA binding domain of Antp, a *Drosophila* transcription factor, has been calculated to be greater than the contribution of binding Helix-3, the structural element usually regarded as a sole mediator of homeodomain specificity (Crane-Robinson *et al.*, 2006).

Positions that were found to be most strongly selected were 5R, 5L, 6R, 6L, 3R, 1L, 7L and 8L, while position 8R and 4L appeared to be practically random. This result hints that there might be a degree of asymmetry in the way resolvase subunits interact with left and right half-site of site I, a possibility that was raised previously in the literature (Yang & Steitz, 1995).

It is interesting that positions that are contacted in the 1gdt structure such as 4 and 7 seem not to be particularly well conserved in the selection experiments, which is in concordance with an earlier study of binding of 43 amino acid C-terminal α-chymotrypsin fragment of γδ resolvase (residues 141-183) and intact γδ resolvase to a series of mutant γδ site I half-sites (Rimphanitchayakit *et al.,* 1990). Position 4 is contacted by G141, where the glycine carbonyl oxygen forms a hydrogen bond to the ribose ring oxygen of T, whilst R142 forms numerous hydrogen bonds to the TA base pair and the DNA backbone at position 7 (Fig. 3.30). While lack of strong influence of position 4 in the study of Rimphanitchayakit *et al. i*s easily discounted due to the limited contacts it can make with the resolvase fragment whose N-terminus is at G141, their result that any substitution at position 7 is tolerated is more puzzling. Apart from the specific hydrogen bond contacts to the AT base pair at

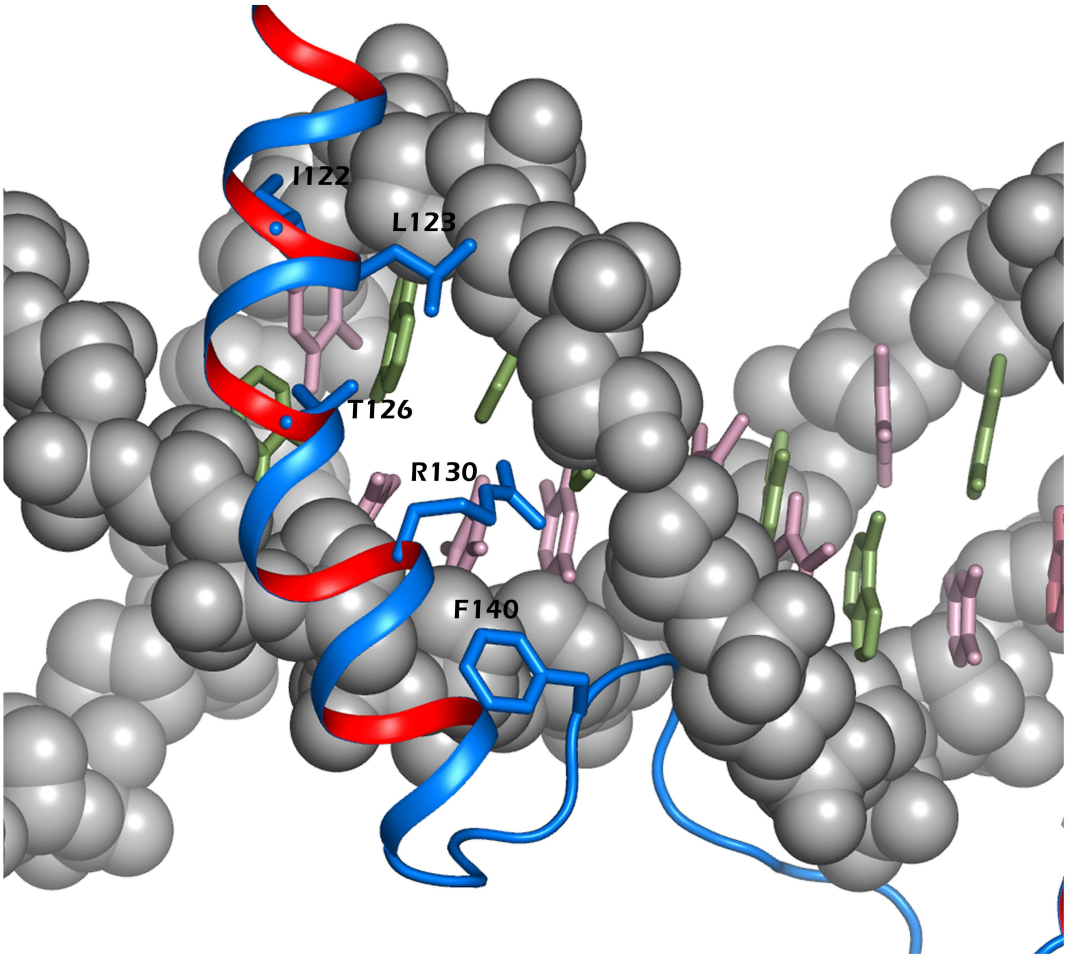


Figure 3.29- Resolvase residues in the extended arm region making contacts with the DNA backbone in the centre of site I. Residues I122, L123, T126, R130 and F140 are given in sticks representation. The rest of resolvase protein is in cartoon (blue and red), while the DNA backbone is gray and in spacefill. DNA bases are in sticks and are coloured as follows; adenine (A, olive green) and thymine (T, lilac). The figure was made using 1gdt coordinates (Yang & Steitz, 1995) and program PyMol 0.99.

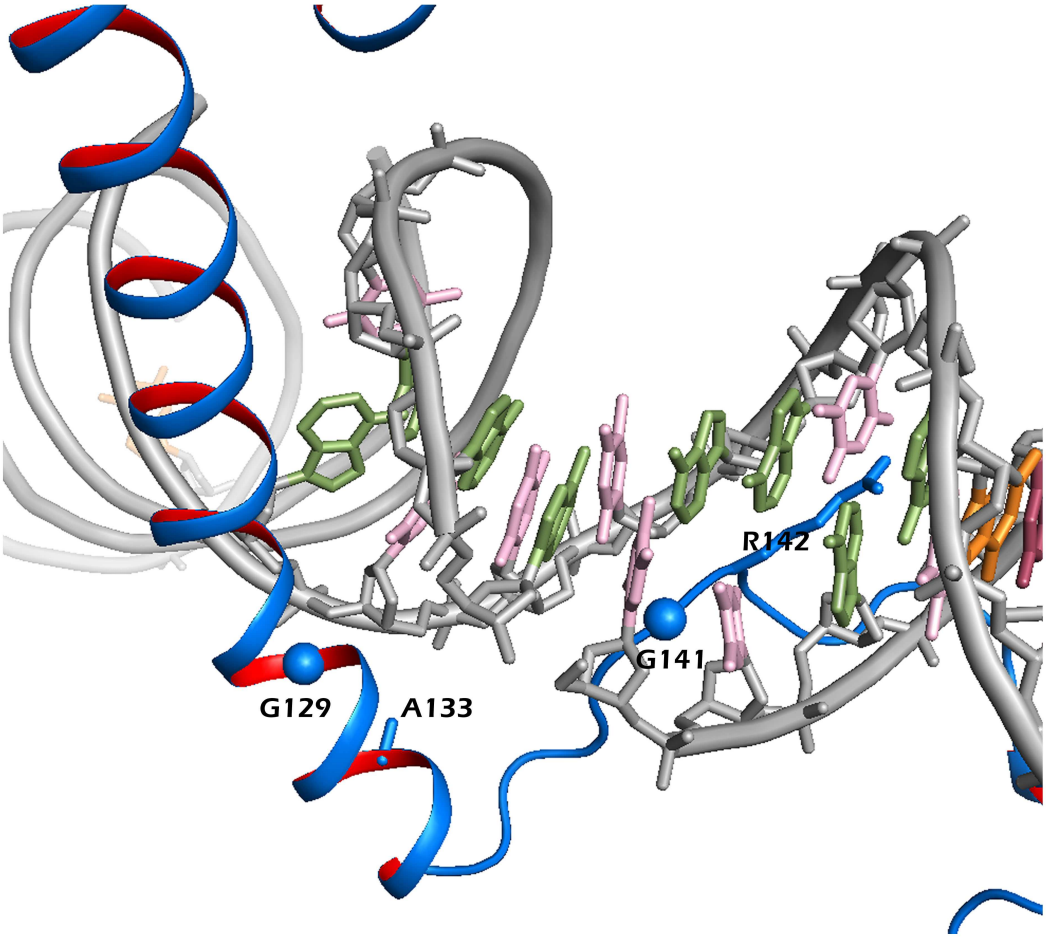


Figure 3.30- Resolvase residues in the extended arm region making contacts with the DNA backbone in the centre of site I. Residues G129, A133, G141 and R142 are blue. The glycines are given in spacefill while alanine and arginine are given in sticks representation. The rest of resolvase protein is in cartoon (blue and red), while the DNA backbone is gray and in sticks. DNA bases are also in sticks and are coloured as follows; adenine (A, olive green) and thymine (T, lilac). The figure was made using 1gdt coordinates (Yang & Steitz, 1995) and program PyMol 0.99.

position 7, the aliphatic portion of R142 makes extensive van der Waals contacts with the carbon moieties of the DNA backbone between positions 4-8 (Fig. 3.30). It seems that these other interactions are sufficient to compensate for the possible substitutions at position 7.

If positions 5 and 6 were allowed to vary on either side of the site, the bases chosen at these positions were exclusively either an A or a T. In addition a propensity for extending the existing runs of T's was also noted. Rimphanitchayakit *et al.,* observed that while an A to T transversion at position 6 has no adverse effect on binding, an A to a G or C substitution strongly inhibits it. For position 5 they reported a similar effect for a G or C substitution, with the A to a C change being especially adverse, but also a milder binding inhibition for an A to T transversion. They showed that A5C causes the site to be less bent causing an incorrect structural configuration of site I, which in turn inhibits *res* × *res* recombination. They argued that a wild-type AT base pair at position 5 provides a narrow and compressible minor groove that plays an important part in the formation of the resolvase-induced bend at site I, and that replacing it with a GC base pair would widen it and reduce its compressibility, which would cause a loss of affinity as well as reducing bending of the complete site.

Indeed, one of the commonest changes at position 5 that is found in the selected recombination-deficient Tn3 site I variants is the change of A at 5R into a C, and corresponding A at 5L into a G. This is illustrated particularly clearly in Fig.3.27d in which the sequences of the selected recombination deficient-Tn3 site I variants were combined to make a sequence logo (Schneider & Stephens, 1990; Crooks *et al.,* 2004). A to T transversion at this position was not selected against, and it seemed to actually be preferred (Section 3.7) which is in contrast with the observations of Rimphanitchayakit *et al.* As the assays reported in this chapter were recombination assays and not binding assays like the ones published previously it is difficult to say whether the A to T transversion at position 5 is inhibitory to binding by NM resolvase. As Tn3 resolvase contacts these two positions in the minor groove it is difficult to see how it could distinguish an AT or a TA base pair either directly or indirectly. A to T changes preserve the minor groove hydrogen bond donor/acceptor pattern and also the increased propeller twist that leads to a narrow minor groove. The direct recognition of the hydrogen bond donor/acceptor pattern would be difficult as the pattern presented by TA or AT base pair in the minor groove is very similar (Seeman *et al.,* 1976).

Changes of A into a T at position 5R or 5L create a run of 4 or 5 consecutive T's, respectively. These homopolymer runs are known to be rigid and to influence sequence induced bending of DNA at their ends (Koo *et al*., 1986; Barbic et al., 2003). They are known to cause progressive narrowing of the minor groove, to the 3' end of the A run (Burkhoff & Tulius, 1987), which in the case of the Tn3 site I means that the minor groove would get progressively narrower towards the centre of the site. As the narrow minor groove is already a feature in this region of Tn3 site I (Fig.3.28) further narrowing should be tolerated.

The pattern of base conservation at positions 8L-5L (8R-5R) suggest that resolvase uses a GR motif (G141, R142, Fig. 3.30) to recognise this AT-rich sequence in a fashion very similar to the AT-hook motifs (Huth *et al.,* 1997). The main force driving the binding of AT-hook motifs, which use an arginine residue in a similar fashion to resolvase, comes from the entropy increase resulting from displacement of ordered water molecules that are uniquely characteristic of the AT-rich minor groove (Dragan *et al.,* 2003) such is found in site I. Only AT-rich regions can achieve a narrow minor groove as they have a greater degree of propeller twist between adjacent base pairs. In a sequence containing CG, an N2 amino group of the guanine clashes with the base in the n+1 position in the opposite strand, preventing the narrowing of the groove. As Tn3 site I variants containing G or C substitutions at the positions 8-5 might not be able to assume a minor groove conformation that is narrow enough to allow for the formation of a spine of ordered water molecules, the entropy increase that would ensue from their replacement and drive the binding of R142 into the minor groove in these mutant site I variants might also be lacking. Therefore I postulate that the inhibition of recombination by G or C substitutions at positions 8-5 is caused by reduction in the binding affinity. This could be tested by using inosine base analogues instead of guanine to replace A or T residues at positions 8-5, as was done for a tyrosine recombinase Flp (Whiteson & Rice, 2008). As the inosine base lacks the N2 amino group, the sites containing inosine replacements should be able to assume a similar level of propeller twist as the wild-type sequence.

Another base position that was shown to be very well conserved is position 3. In the crystal structure this position and position 4 are contacted by R130 (Fig 3.29) which makes specific hydrogen bonds with the N3 atoms of adenine at the opposite DNA strand (Yang & Steitz, 1995). It has been suggested that these contacts are used by resolvase to exclude GC-containing sequences (Yang & Steitz, 1995). This seems likely as other resolvases that have a G or a C at this position have an I or an L at an equivalent protein residue, respectively (M. Boocock, personal communication). Interestingly, no resolvase with an A

base at position 3 has been characterised. So, unlike the affinity-increasing interactions described for positions 5-8, this interaction seems to be to determining specificity. Other residues that contact the edges of the base pairs in the minor groove are F140, L123 and I122 (Fig 3.29). These interactions seem to be specialised van der Waals interaction with base positions 3-5. Residue 123 is variable in different resolvases; it is an R or a K in resolvases associated with a site I which has a string of G's at positions 3-5. Position 3 was especially well conserved in the right half-site of Tn3 site I. As the resolvase HtH binds more tightly to the outermost 6 bp of the right half-site (Bednarz, 1990), it is possible that the contribution of the arm region minor groove contacts at positions 8-5 is less crucial, increasing the relative importance of the specific interaction at position 3.

Sequence discrimination by NM resolvase could take place in the binding and/or catalytic steps. As the assays in this study were based on full recombination, the distinction between these steps can not be made with certainty. Nevertheless it seems that specificity in the central 16 bp of Tn3 site I is imparted by a combination of direct contacts at position 3 and indirect readout based on an AT-hook motif-like binding in the minor groove between positions 8 and 5. NM resolvase seems to interact with the left and right half-sites of Tn3 site I in a slightly different manner. At the left half-site, presumably due to the weak binding of the HtH, site contacts between positions 8-5 and the existence of a Y-R base step between positions 1 and 2 appear to be the most important. At the right half-site, although the aforementioned interactions are important, specific contacts to the position 3 seem to take precedence.

Although sequences that are quite different to the native Tn3 site I sequence, i.e. the ones containing runs of T's or G/C at positions 2 or 4 can be recombined quite efficiently by NM resolvase (Fig. 3.23) they seem to be conspicuously rare in the wild type site I's associated with resolvases from the Tn3 resolvase family (Fig. 3.1 and 3.27). It is possible that the reason for their relative scarcity is connected to the promoter function of site I as mentioned in Section 3.2. Site I sequences containing G's and C's are likely to be poor gene promoters as they deviate from the Pribnow box consensus (Pribnow, 1975; Harley & Reynolds, 1987) which might be the reason for them to be selected against in nature although they seem not to be detrimental to recombination.

4 Chapter 4: Sequence selectivity of Z-resolvases

4.1 Introduction

In the previous chapter I have shown that sequence specificity of hyperactive Tn3 resolvase is influenced by combination of direct and indirect protein-DNA contacts at the centre of site I. The results showed that there is a degree of flexibility in what sequence is tolerated in this region, with most of the 16 mutated positions tolerating more than one type of basepair. However, it was also evident that although individual changes are tolerated relatively well, the accumulation of changes is ultimately detrimental to the recombination reaction.

In this chapter, I present the experiments assessing to what extent these observations apply to a Z-resolvase protein working on a Z-site. Z-resolvases are chimaeric resolvase derivatives in which a catalytic domain is provided by a resolvase, while the DNA-binding function is provided by a zinc finger DNA-binding domain (Akopian *et al.*, 2003; Gordley *et al.*, 2007). A Z-site is a derivative of *res* site I in which a sequence normally bound by the resolvase DNA binding helix-turn-helix (HtH) domain has been replaced by the sequence recognised by a zinc finger DNA binding domain. Using Z-resolvase effectively removes all, even indirect, influence of the native resolvase HtH domain (as this domain is no longer present) on sequence selectivity, leaving the selectivity of the resolvase catalytic domain unobscured. Also, improving the understanding of the sequence selectivity of Z-resolvases could potentially lead to the design of more specific hybrid recombinases with various future applications.

In the first part of the chapter a mutant library selection strategy used to select Z-site variants recombined by Z-R(NM), a Z-resolvase with a catalytic domain from NM resolvase, is described. Later in the chapter, the design, construction and properties of Z-resolvases with different sequence specificity, specifically Z-resolvases with the catalytic domains of Sin and Tn21 resolvase, are discussed.

4.2 Z-R(NM) sequence selectivity

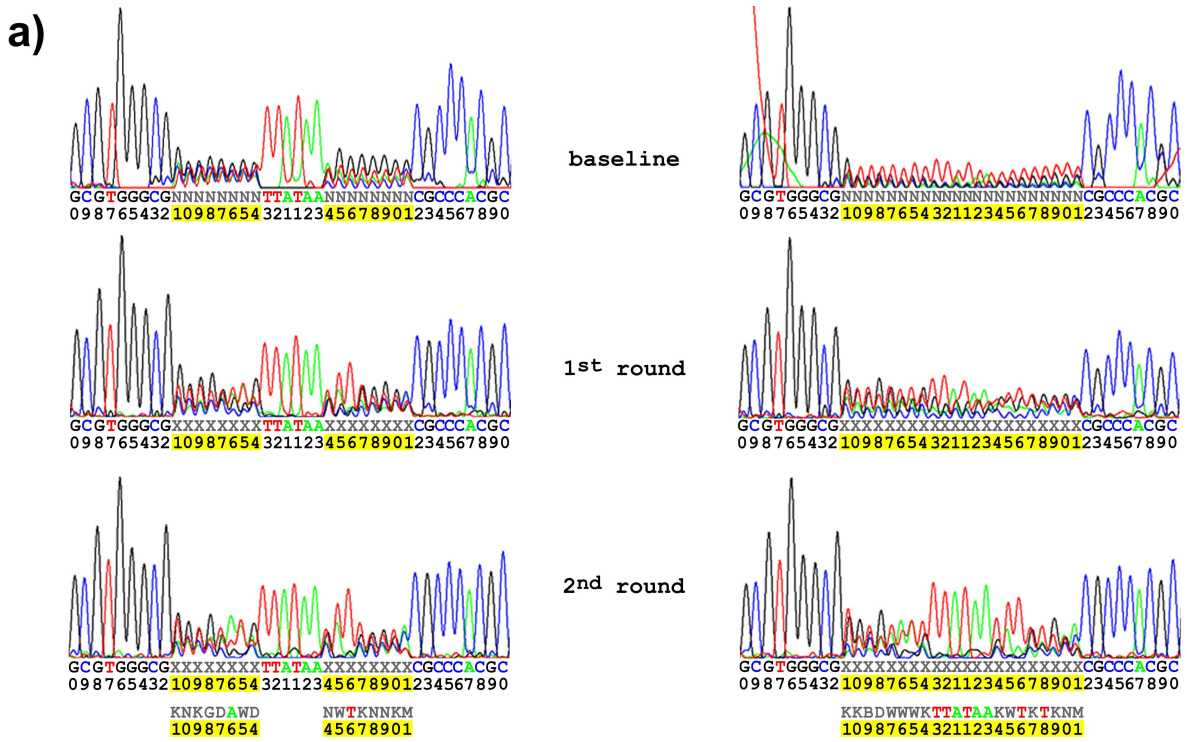
To test the sequence specificity of a Z-resolvase a mutant library selection similar to the one described in Section 3.7, was devised. Instead of creating a series of overlapping libraries (as in Chapter 3), a less exhaustive sequence space scan using only two random Z-site libraries was employed. This approach, although less thorough than the overlapping

libraries method, should generate sufficient data to allow for comparison between the recombination requirements of resolvase and the equivalent Z-resolvase.

The two random Z-site libraries that were created were called LIB(Z22Z) and LIB(Z8TTATAA8Z). Libraries were made using Z22NZ and Z8NC8NZ oligonucleotides respectively (Table 2.4) according to the method described in Section 3.3. LIB(Z22Z) contained a site in which all 22 bases in the centre of the Z-site are random, while library LIB(Z8TTATAA8Z) had 8 random bases on each side flanking the central sequence 5'-TTATAA-3' as is found in the wild-type Tn3 site I (Fig. 4.1a). Each of the libraries generated consisted of ~ 10^5 clones. This number of clones, though orders of magnitude smaller than the theoretical library diversity, should still be large enough to allow for selection of sites that can be recombined.

Selections of recombination proficient sites were performed using purified Z-R(NM) protein (Section 2.25). Z-R(NM) consists of 1-148 amino acid (a. a.) residues of the hyperactive Tn3 resolvase NM, a 2 a. a. residue linker (TS) and the Zif268 DNA-binding domain (residues 2-90). Using a plasmid multimer resolution-based selection strategy analogous to the one described in Section 3.7, recombination proficient Z-sites were selected in two selection rounds. Unlike the selection described in Section 3.7, *in vitro* resolution was allowed to proceed for 1 h in both the first and the second round. The allowed resolution time was kept long in order to maximise the number of recombinants, which are likely to be rarer than in the Chapter 3 experiments due to the larger variation of the library. The resulting monomer DNA was sequenced in bulk (Fig 4.1a) using the uni -43 primer. Also, plasmid DNA from ten randomly selected DH5 colonies from each library was sequenced individually. The sequences were used to produce a sequence logo (Fig 4.1b).

The results of selections using either of the two Z-site libraries were strikingly similar. The bulk sequencing traces (Fig. 4.1) reveal that Z-R(NM) has a strong preference for TTATAA in the centre of the Z-site. The selection for these bases at positions 1-3L and 1-3R was far stronger than was observed in the Tn3 site I variant selections using NM resolvase. Also, there was no difference in the intensity of the selection in the right or left side of the Z-site. Positions 5 and 6 were found to be exclusively an A or a T which was also in concordance with the findings discussed in the previous chapter. No base preference was observed outside positions 7R or 8L. The stringency of the selection in the centre argues that binding by Z-resolvase is equally strong on both half sites and that the Z-



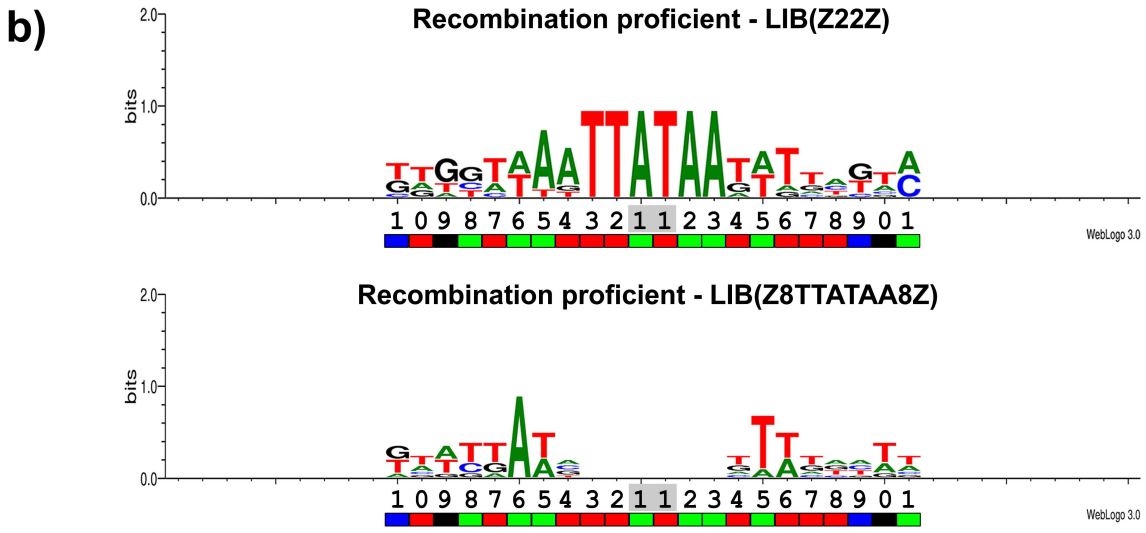


Figure 4.1-a) Bulk sequencing traces from the site selections using Z-R(NM) from the random block libraries LIB(Z22Z) and LIB(Z8TTATAA8Z). Sequencing was done using the primer uni -43. The individual DNA bases are numbered from the centre: 1-20 left and 1-20 right. For bases 10-20 only the second digit is shown. The position of the random block is highlighted in yellow below the sequencing trace. Underneath the bulk sequencing trace from the 2nd round of selections, the sequence of the randomised regions of the sites is given in the extended IUPAC DNA notation. **b)** Sequence logos made using the 10 individual sequences from the second round of selection from the libraries LIB(Z22Z) (top) and LIB(Z8TTATAA8Z) (bottom).A coloured bar representing the wild-type Tn3 site I sequence (positions 1-11R and1-11L) is provided under the logo for comparison. Sequence logos were created using the WebLogo 3.0 server.

resolvase recognises the centre of the site using extended arm contacts in a very similar manner to NM resolvase which was used to construct it.

Thus if the Z-resolvase has the same sequence preference in the central 16 bp of the Z-site as the resolvase that provided its catalytic domain, in order to create a Z-resolvase that can target sequences that are significantly GC-rich, the catalytic domain of a resolvase that intrinsically recognises GC-rich sequences needs to be used.

4.3 Designing new Z-resolvases

The resolvases that were chosen for providing catalytic domains for new Z-resolvases were Sin and Tn21 resolvase. These proteins were deemed to be ideal candidates since their native site I central sequences are GC-rich, and thus quite different from Tn3 site I. In addition, Sin and Tn21 resolvase had been previously studied in our group, so a number of useful plasmid constructs were available. As the sequence preference of the catalytic domains of these proteins is different than that of NM resolvase, Z-resolvases constructed using Sin or Tn21 catalytic domains are likely to display a sequence preference that is very different from that observed for Z-R(NM) resolvase.

Z-resolvase is supposed to catalyse a recombination reaction on a Z-site which is a direct analogue of the recombination reaction at *res* site I. Therefore, in order to construct a Z-resolvase that is capable of recombination, the first requirement is a suitably activated catalytic domain. This domain should come from an activated resolvase mutant that is able to catalyse recombination between its native crossover sites without needing the accessory sites. Fortunately, for Sin and Tn21 resolvase a number of such mutants were already available (Rowland *et al.*, 2009; and D. Wenlong & D. Sneddon, unpublished).

Provided that an activated mutant was available, the next challenge would be to decide at what point to dissect the resolvase protein, i.e. how much of the N-terminal portion to include in the new Z-resolvase protein. Choosing the correct point to create a fusion is critical, as this could influence both the protein's function and its structural stability. In case of Tn3 resolvase, this decision was greatly simplified by the availability of high-resolution structures of the very closely related γδ resolvase (Yang & Steitz, 1995, Li *et al.*, 2005). A high-resolution structures for Sin resolvase bound to DNA (site II) became available during this work (Mouw *et al.*, 2008). Comparison of the structures of γδ and Sin resolvases reveals that although the difference in primary sequence is high (only 31.68% identity), the tertiary structure of the protein bound to DNA is remarkably similar. No high-resolution structure of Tn21 resolvase is currently available; however, based on the

previous observation, it would not be not be unreasonable to assume that Tn21 resolvase has a similar final fold also. Assuming that it does, a decision as to how much of the N-terminus of Tn21 to include when constructing a Z-resolvase could be made with the help of the primary protein sequence alignment (Fig. 4.2).

After deciding at what residue to terminate the resolvase catalytic domain portion of the Z-resolvase, the next issues arising are the sequence and length of the protein linker that is to connect the catalytic domain to the zinc finger domain of the Z-resolvase. The linker needs to be of sufficient length and flexibility to allow the Z-resolvase's zinc finger domain to bind the DNA when the N-terminal domain of the protein is positioned correctly at the centre of the Z-site.

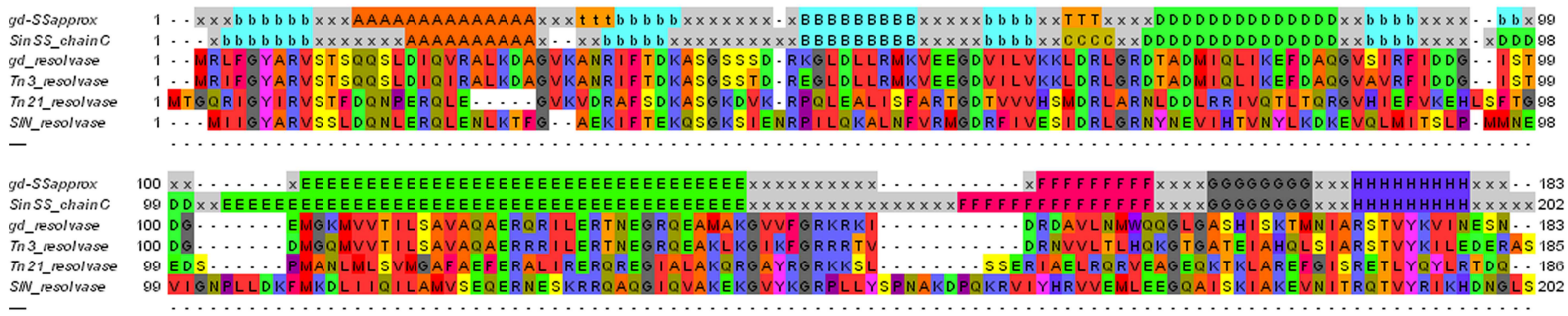
Finally, to test the new Z-resolvases for recombination activity, new *in vivo* test substrate plasmids, containing suitable Z-sites, will need to be constructed. If further characterisation of these proteins is to be carried out *in vitro*, protein over-expression constructs would need to be made, proteins purified, and then tested using *in vitro* test substrate plasmids that would also require construction.

4.4 Sin Z-resolvase, Z-R(Sin)

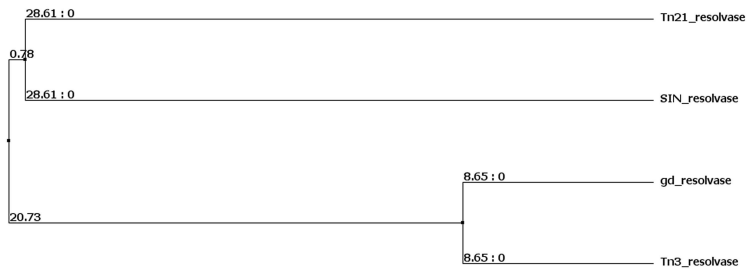
To create a Sin resolvase-based Z-resolvase (Z-R(Sin), from here on) the issues discussed in the previous section had to be tackled. First of all, a hyperactive mutant of Sin resolvase was chosen to provide the catalytic domain. The mutant selected was Q115R (Rowland *et al.*, 2009). This single mutant is sufficient to make Sin resolvase hyperactive so that it no longer requires the presence of the accessory sites, and as such it was deemed to be a suitable candidate for turning into a Z-R(Sin).

Following the choice of the hyperactive mutant, the second issue was the N-terminus cut-off point, i.e. a decision on how much of the Sin resolvase N-terminus to include in the Z-R(Sin). This was based on the previous work on creation of Z-resolvases with a Tn3 resolvase catalytic domain (Section 1.7), and the available structures of γδ and Sin resolvase bound to DNA (Yang & Steitz, 1995; Mouw *et al.*, 2008). In γδ resolvase, the arm region of the protein adopts an extended conformation between residues 140 and 148 that follows the minor groove closely until it reaches the F helix of the C-terminal domain. In Z-R(NM), residues 140-148 have been retained, as it was observed that replacing these residues with alternative sequences created less active recombinases (A. MacLean, unpublished). The equivalent extended linker region in Sin resolvase is 11 residues longer (Fig 4.2). Although site I of Sin resolvase is 2 bp longer than the Tn3 site I, the number of

a)



b)



	% identity
Tn3 *vs.* γδ resolvase	80.33
Tn3 *vs.* Tn21 resolvase	33.33
Tn3 *vs.* Sin resolvase	31.86
γδ *vs.* Tn21 resolvase	30.69
γδ *vs.* Sin resolvase	31.68
Tn21 *vs.* Sin resolvase	33.00

Figure 4.2-a) Protein sequence alignment of γδ resolvase, Tn3 resolvase, Tn21 resolvase and Sin resolvase. Secondary structure elements for gd resolvase (gd_SS approx) and Sin resolvase (Sin_SS_Chain C) are marked at the top of the alignment based on the structures 1gdt and 2roq (Yang & Steitz, 1995; Mouw *et al.,* 2008). **b)** An evolutionary tree (left) showing the relationship between the γδ resolvase, Tn3 resolvase, Tn21 resolvase and Sin resolvase (generated by Jalview 2.1.2). Branch length corresponds to the calculated evolutionary distance. A list showing the percentage identity based on the primary protein sequence between the γδ resolvase, Tn3 resolvase, Tn21 resolvase and Sin resolvase is given on the right.

additional residues seems too large to be explained by the need to bind to a longer site. These residues might be involved in synapsis at the accessory sites for Sin, which are quite different from the Tn3 accessory sites (Mouw *et al.,* 2008). Therefore it was decided to include the Sin resolvase N-terminal domain up to the last residue that appears to specifically interact with the DNA (R148); In Tn3 resolvase this residue is R142, the importance of which was discussed in the previous chapter.

In order to link the Sin residue 148 to the zinc finger domain, instead of using an arbitrary sequence as a linker it was decided to utilise the natural Tn3 sequence (residues 143-148) as shown in Fig.4.3a. This sequence has been shown to work well as a linker in Z-R(NM) so it seemed reasonable to use it for Z-R(Sin) also.

To make an expression plasmid (pMP213) expressing Z-R(Sin), a three piece ligation was performed using a 432 bp NdeI/BsrGI fragment of pSA9994, a 28 bp BsrGI/SpeI linker and a 5190 bp NdeI/SpeI fragment of pMP59 (Fig. 2.4). pSA9994 was kindly donated by S. Rowland. This plasmid carries a cassetted Sin resolvase ORF containing a Q115R mutation caused by an A to G transition at position 344. Its 432 bp NdeI/BsrGI fragment provides the catalytic domain for the Z-R(Sin). pMP59 is a plasmid expressing Z-R(NM); its 5190 bp NdeI/SpeI fragment acts as a backbone and it supplies the part of the Z-resolvase ORF coding for the Zif268 DNA binding domain. In order to bridge the gap between the BsrGI and SpeI sites and introduce the linker amino acid residues a 28 bp DNA fragment was used. This linking fragment was made by annealing synthetic oligonucleotides SinZF and SinZR (Table 2.4).

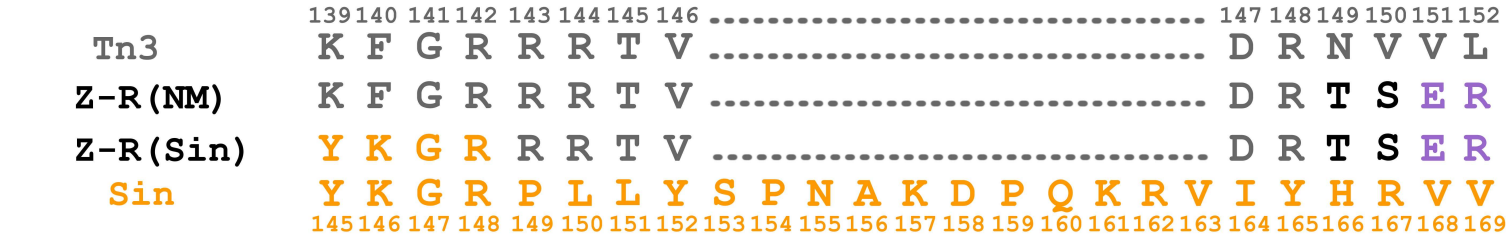
The resulting pMP213 expresses Z-R(Sin), a Z-resolvase assembled from Sin Q115R residues 1-148, linked by the sequence RRTVDRTS to residues 2-90 of the Zif268 DNA binding domain.

4.5 Tn21 Z-resolvase, Z-R(Tn21)

Tn21 resolvase is a distant relative of Sin and Tn3 resolvase (Fig 4.2b) that has been well characterised (Hall & Halford, 1993). Although there is no structure available, the amino acid sequence of Tn21 resolvase aligns well with the other resolvases whose structure is known (Fig 4.2a). No hyperactive mutants of Tn21 resolvase have been reported in the literature to date.

In the earlier work of our group (A. Bednarz, D. Wenlong and D. Sneddon, unpublished) a cassetted ORF of Tn21 resolvase was synthesised and cloned into a pMS140 backbone to

a)



b)

Figure 4.3- Protein sequence of the linker joining the catalytic domain of resolvase to the Zif268 DNA-binding domain. **a)** An alignment of the linker sequences of Z-R(NM) and Z-R(Sin) along with the corresponding sequences from Tn3 and Sin resolvases. Tn3 resolvase residues and numbering is given in grey. Sin resolvase residues and numbering are given in orange. A sequence "TS" that introduces a SpeI site required for joining the catalytic domain of resolvase to the Zif268 DNA-binding domain is in black. The first two residues of Zif268 DNA-binding domain are given in purple. **b)** An alignment of the linker sequences of Z-R(NM) and Z-R(Tn21, L3) and Z-R(Tn21, L21) along with the corresponding sequences from Tn3 and Tn21 resolvases. As above, the Tn3 resolvase residues and numbering is given in grey. Tn21 resolvase residues and numbering are given in blue. A sequence "TS" and the first two residues of Zif268 DNA-binding domain are in black and purple, respectively, as in a). In both a) and b) the "GR motif" at the end of the resolvase N-terminal domain and the first residue of the zinc finger domain are taken as the anchor points for the alignment.

create pDW21. The Tn21 resolvase was mutagenised using either PCR-based mutagenesis or *in vivo* by using the mutagenic XL-1 Red *E. coli* strain (*mut*S$^-$, *mut*D$^-$, *mut*T$^-$) (Stratagene), and tested for resolution of *in vivo* substrates using the MacConkey assay (Section 2.23). The *in vivo* resolution substrates used were pDW24 (*res* × *res*), pDW27 (site I × *res*) and pDW25 (site I × site I). Two hyperactive single mutants of Tn21 resolvase were selected that were capable of recombining pDW25, namely E124V and M63T (Fig.4.4). Plasmids containing these Tn21 resolvase hyperactive mutants were named pDJS1 and pDJS2, respectively.

The Tn21 resolvase mutant libraries that these mutants were selected from contained only a few thousand clones. Since these libraries were so small there was a real possibility that other 'better' activating mutations were missed. Since M63T and E124V hyperactive mutants were already available, it was decided to proceed with them as the source of the catalytic domain for the new Tn21 Z-resolvases, but also concomitantly to construct bigger Tn21 resolvase mutant libraries and attempt to select further hyperactive mutants.

4.5.1 Tn21 resolvase hyperactive mutant selections

Selection of hyperactive mutants of Tn21 resolvase was done using the MacConkey assay (Section 2.23). A wild-type Tn21 resolvase reading frame from pDW21 was mutagenised by mutagenic PCR (Section 2.28.2). Two mutant libraries using di-oxo-G (LIB(Tn21O)) and dPTP (LIB(Tn21P)) nucleotide analogues were created. The concentration of nucleotide analogues was adjusted to result in a mutation frequency of ~one mutation per reading frame. Each of the libraries had around 60 000 clones.

The libraries of plasmids expressing mutated Tn21 resolvase were tested using the *in vivo* resolution substrate plasmids pDW24 (*res* x *res*) and pDW25 (site I x site I). Plasmid pDW24 is resolvable by the wild-type Tn21 resolvase as it contains full *res* sites, whereas pDW25 should be only resolved by mutants that no longer require the presence of accessory sites in order to catalyse recombination. To test for resolution, Tn21 resolvase mutant libraries LIB(Tn21O) and LIB(Tn21P) were used to transform DS941 cells already containing the test plasmids. Following transformation, the cells were plated onto selective MacConkey agar and incubated overnight at 37°C. White colonies on plates with pDW25 test substrate plasmid were picked, and the Tn21 resolvase expression plasmid DNA was isolated. These candidate activated Tn21 resolvase mutants were then re-tested using the same assay, and the ones that reproducibly gave white colonies were isolated and sequenced. The list of the isolated mutants is given in Table 4.1.

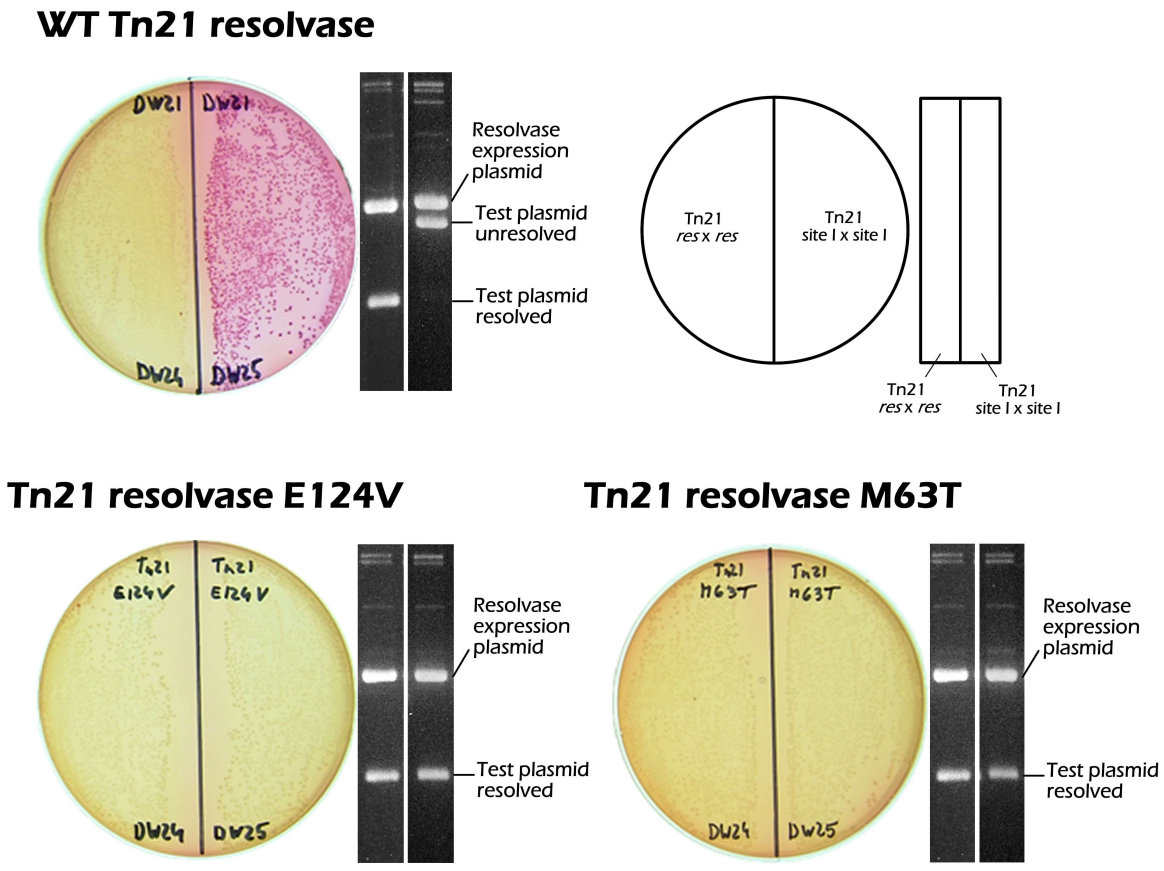


Figure 4.4- A MacConkey assay result using a wild-type Tn21 resolvase (top left), Tn21 resolvase E124V (bottom left) and Tn21 resolvase M63T (bottom right). The DS941 cells containing resolvases acting on pDW24 (*res* × *res*) substrate are plated on the left half of the agar plate while the cells containing resolvases acting on pDW25 (site I × site I) are plated on the right half of the agar plate as indicated in the explanation key (top right). An agarose gel photograph showing uncut plasmid DNA isolated from each half of the agar plate is found next to photograph of the agar plate it corresponds to. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

Mutant isolate	DNA Change	Protein Change
P19	A104**G**, T191**C**, T255C, A334**G**	K33**R**, M63**T**, M111**V**
P20	T109**C**, A190**C**	S36**R**, M63**T**
P32	T191**C**, T442**C**	M63**T**, S147**P**
P33	T93C, A108G, T191**C**	M63**T**
O1	A374**C**	E124**A**
O2	A374**C**	E124**A**
O7	A374**C**	E124**A**

Table 4.1- List of Tn21 resolvase mutants isolated by screening the libraries LIB(Tn21P) and LIB(Tn21O).

As can be seen in Table 4.1 all of the mutants isolated contained a mutation at either position 63 or 124. In the work on Tn3 resolvase, it was observed that combining some of the activating mutations has a synergistic effect (Burke *et al.,* 2004; Olorunniji *et al.*, 2008). To see whether mutations at positions 63 and 124 have a synergistic effect in Tn21 resolvase these two mutations were combined. A PstI/BsrGI fragment containing the M63T mutation from pDJS2 was inserted into pDJS1, a E124V Tn21 resolvase expressing plasmid, which served as a vector. The product of this cloning, pMP269, expressed Tn21 resolvase with both mutations. The double mutant was tested along with both of the single mutants in the MacConkey assay using pDW24 (*res* × *res*), pDW27 (*res* × site I) and pDW25 (site I × site I) substrates. As can be seen on Fig. 4.5 the double mutant, although activated, is less hyperactive than either of the single mutants on their own. Since the two single and the double mutant were all active on the site I x site I substrate pDW25 it was reasoned that each of them could be used to supply the catalytic domain for Tn21 Z-resolvase construction.

4.5.2 Tn21 Z-resolvase construction

Like Tn3 and Sin resolvases, as revealed by the protein sequence alignment (Fig 4.2), Tn21 resolvase has a GR motif that is found at the end of the N-terminal domain (residues G141, R142) just before the start of the extended portion of the arm region. The linker region is three residues longer in Tn21 resolvase than in Tn3 resolvase. This could be due to the necessity to recognise a 30 bp site I, which is 2 bp longer than the site I recognised by the Tn3 resolvase.

Using the strategically placed EagI site both Tn21 and Tn3 resolvase reading frames can be cut at the equivalent place in their amino acid sequences straight after the GR motif. This allowed for simple NdeI/EagI fragment swaps between Tn21 resolvase expression

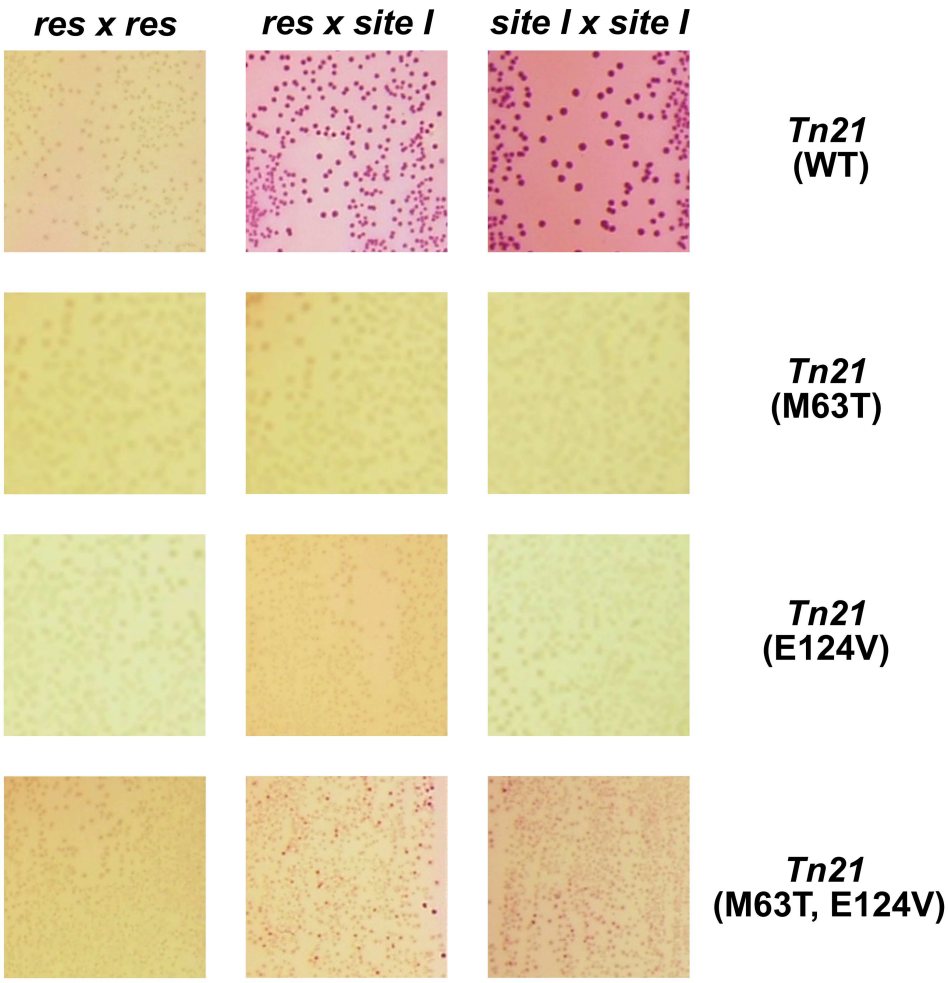


Figure 4.5- A MacConkey assay results using a wild-type Tn21 resolvase, single mutants Tn21 resolvase M63T, Tn21 resolvase E124V and a double mutant Tn21 resolvase M63T, E124V. The photographs of sections of plates with resolvases acting on substrate pDW24 (*res* × *res*), pDW27 (*res* × site I), and pDW25 (site I × site I) are arranged in a grid. The type of substrate is indicated at the top of each column, while the type of resolvase is indicated on the right of each row. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

plasmids pDJS1 (Tn21 resolvase E124V), pDJS2 (Tn21 resolvase M63T), and pMP269 (Tn21 resolvase M63T, E124V) providing Tn21 resolvase residues 1-142, and the Z-R(NM) expressing plasmid pMP59 to make Tn21-Z resolvase expressing plasmids, pMP250, pMP252 and pMP267 (Fig. 2.3). A 421 bp Nde/EagI fragment of pDJS1, pDJS2 and pMP269 provided Tn21 resolvase catalytic domain residues 1-142, and a 5125bp NdeI/EagI fragment of pMP59 supplied the backbone and the Zif268 DNA-binding domain residues 2-90. As a result of using a NdeI/EagI fragment swap with the pMP59 all of these Tn21 Z-resolvases had a L3 linker sequence (Tn3 resolvase residues 145-148 followed by TS), as shown in Fig.4.3.b, just like the previously constructed Z-R(NM) and Z-R(Sin).

Tn21 Z-resolvases expressed from pMP250, pMP252 and pMP267 were called Z-R(Tn21,124V, L3), Z-R(Tn21,63T,L3) and Z-R(Tn21, 63T/124V, L3), respectively.

An alternative second set of Tn21 Z-resolvases with a linker fully based on Tn21 resolvase sequence (L21) was also made (Fig.4.3.b). To do this, a short linking DNA fragment was constructed by annealing oligonucleotides Tn21ZF and Tn21ZR (Table 2.4). This fragment coded for Tn21 resolvase residues 143-151(KKSLSSERI) followed by residues TS. The fragment had staggered ends so that it could be ligated to a vector DNA digested with EagI and SpeI restriction enzymes (Fig. 2.6). This EagI/SpeI fragment was swapped into plasmids pMP250, pMP252, and pMP267, replacing the L3 with the L21 linker. The resulting plasmids were named pMP251, pMP253 and pMP268, and they expressed Tn21 Z-resolvases called Z-R(Tn21, 124V, L21), Z-R(Tn21, 63T, L21) and Z-R(Tn21, 63T/124V, L21), respectively.

4.6 Construction of Sin-Z and Tn21-Z sites

In order to test the Z-R(Sin) and Z-R(Tn21) recombinases, a number of Z-sites for them to work on was designed. When Z-sites were first made to be used with Tn3 resolvase-based Z-resolvases, a set of Z-sites with Zif268-binding sites (ZBS) at different distances from the centre of the site was constructed (Akopian *et al.*, 2003). It was established that the optimal distance between the ZBS is 22 bp (discussed further in Chapter 5). It was unclear whether the same distance between ZBS's would be optimal for Z-resolvases that are based on a different catalytic domain. However, as this distance worked well with Tn3 Z-resolvase, it was decided to proceed with a similar Z-site design (Fig 4.6).

Two different versions of the Tn21 Z-site were designed. A Z-site with 22 bp of Tn21 *res* site I sequence in the centre is a large almost perfect inverted repeat, and as such sequences

are unstable *in vivo*, it proved difficult to clone. Because of this, it was decided to make a site in which the inverted repeat is less perfect. To do this, a version of the Tn21 Z-site (Tn21-Z* site) in which an AG, the outermost two bp on the left side of the site just before the ZBS was replaced with TC, was made (Fig.4.6). These two bases of site I should be contacted by the Tn21 resolvase C-terminal domain, which is no longer present in the Z-resolvase protein. It was shown in Section 4.2 that Z-R(NM) did not seem to make any specific DNA contacts in this region, so it was expected that Z-R(Tn21) proteins might behave similarly. In any case, protein-DNA interactions involving these bases would be by the extended linker in the DNA minor groove. Proteins are generally unable to distinguish between specific bases, but only between types of base pairs in the minor groove (Seeman *et al.*, 1976), so it seemed reasonable to assume that an AG to TC change would not be detrimental. Eventually, through perseverance, I managed to clone both versions of the Tn21 Z-site (Fig 4.6).

To test the activity of the new Z-resolvases I made a series of *in vivo* and *in vitro* substrate plasmids containing these Z-sites. Detailed descriptions of the way these plasmids were assembled are given in Section 2.7.2.

4.7 In vivo and in vitro characteristics of Z-R(Sin) and Z-R(Tn21)

Once all the components, Z-resolvase expression plasmids and Z-site containing resolution test substrate plasmids were constructed, the recombination activity of the new Z-resolvases was tested *in vivo* using the MacConkey assay.

Competent DS941 cells containing the resolution test substrate plasmids with Sin-Z sites (pMP217), Tn21-Z sites (pMP94) and Tn21-Z* sites (pMP248) were transformed with pMP250, pMP251, pMP252, pMP253, pMP267, pMP268 and pMP213, expressing the new Z-resolvases to be tested and plated on selective MacConkey agar plates. As can be seen in Fig 4.7, Z-R(Sin) expressed from pMP213 produced white colonies when acting on a substrate pMP217, with Sin-Z sites. Conversely, most colonies were red when Z-R(Sin) was used on substrates pMP94 and pMP248 containing Tn21-Z sites. None of the Tn21-Z resolvases tested produced purely white colonies when acting on any of the substrates tested. The highest proportion of white colonies was observed with Z-R(Tn21,63T, L3) expressed from pMP252, acting on the pMP217 substrate containing two copies of Sin-Z sites. The same recombinase also gave a small number of white colonies with the substrate pMP94 (Tn21 Z-sites). Z-R(Tn21,63T, L21) expressed from pMP253 produced a similar result with the pMP217 and pMP94 substrates, but with a smaller number of white

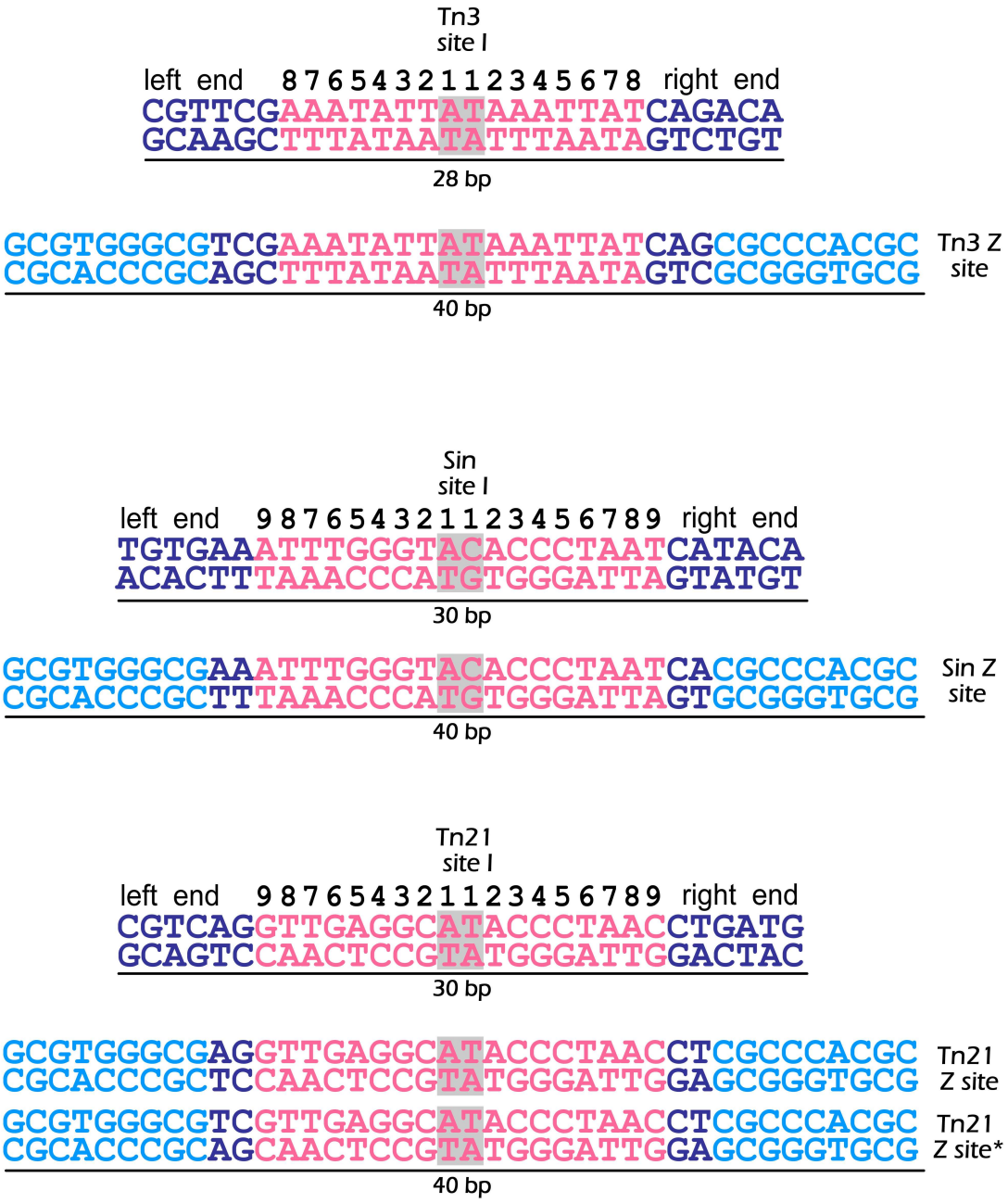


Figure 4.6- The sequences of the Tn3 site I, Tn3 Z-site (Z22Z) (top), Sin site I, Sin Z-site (middle), Tn21 site I, Tn21 Z-site and Tn21 Z-site* (bottom).Sequence given in dark blue is the sequence usually bound by the resolvase HtH domain. Light blue sequence is the sequence bound by the Zif268 DNA-binding domain (ZBS-Zif268 binding site). The central 16 bp in the case of Tn3 site I and Tn3 site I sequence-derived Z-site is pink. For the Sin and Tn21 site I's and their respective Z-sites, the central 18 bp are given in pink. The central dinucleotide is marked by the grey rectangle. Note that AG just 3' of the left ZBS is mutated to TC in the Tn21 Z-site*. The individual DNA bases are numbered from the centre.The "Left" and "Right" end refer to the conventional representation of site I within *res*, where the binding sites are in the order I-II-III left to right. The total length of the site is marked below it.

colonies in each case. No white colonies were observed with Z-R(Tn21,124V, L3), Z-R(Tn21,124V, L21), Z-R(Tn21,63T/124V, L3) or Z-R(Tn21,63T/124V, L21), acting on any of the substrates tested.

As the colony colour changes only if the resolution efficiency is above ~80% (Burke *et al.*, 2004; S. Rowland, personal communication), it was decided to check the plasmid DNA for any detectable signs of substrate plasmid resolution. A representative sample of colonies from each Z-resolvase/substrate combination was scraped off the plates and grown in selective broth overnight. Plasmid DNA was extracted from the cells using the small scale method (Section 2.10.1) and the DNA was separated by agarose gel electrophoresis (Section 2.13.1). This approach revealed that Tn21 Z-resolvases containing the mutation E124V (E124V alone, or E124V, M63T) are inactive on all the substrates tested (Fig.4.7). On the other hand Tn21 Z-resolvases with the M63T mutation (but not E124V) showed recombination activity with all of the substrates tested. The sequence of the linker had only a slight effect on the recombination activity, Z-R(Tn21, 63T, L3) being slightly more active than Z-R(Tn21, 63T, L21). Overall, the most efficient resolution was observed with the substrate pMP217 while the least efficiently resolved substrate plasmid was pMP248 for all Z-resolvases tested.

The following test was to establish whether or not these newly constructed Z-resolvases lose the ability to work on the natural Sin and Tn21 site I's *in vivo*, as was previously observed for Tn3 resolvase-based Z-resolvases (Akopian *et al.*, 2003). As can be seen in Fig. 4.8, activated Tn3 resolvase (G101S, D102Y), Tn21 resolvase (M63T) and Sin resolvase (Q115R) mutants recombine their natural site I's, ignoring the Z-sites derived from the same sequence. The corresponding Z-resolvases: Z-R(SY) (See Chapter 5), Z-R(Tn21, 63T, L3) and Z-R(Sin) do the exact opposite; they recombine substrates containing Tn3, Tn21, and Sin site I derived Z-sites respectively, while ignoring the substrates containing the natural site I sequences. Although Z-R(Tn21, 63T, L3) is not active enough to achieve sufficient resolution of the Tn21-Z sites-containing substrate (pMP94) to produce white colonies, it does indeed partially resolve this substrate as shown previously in Fig. 4.7.

Z-R(Sin), Z-R(Tn21, 63T, L3) and Z-R(Tn21, 63T, L21) were all shown to be at least partially active on all Tn21 or Sin-based Z-sites that were tested, with the resolution of the substrate containing Sin-Z sites (pMP217) being the most complete. This raised a question of whether these Z-resolvases are capable of recombining other Z-sites too. To test whether this was the case, Z-R(Sin), the most active of the new Z-resolvases, was tested using the

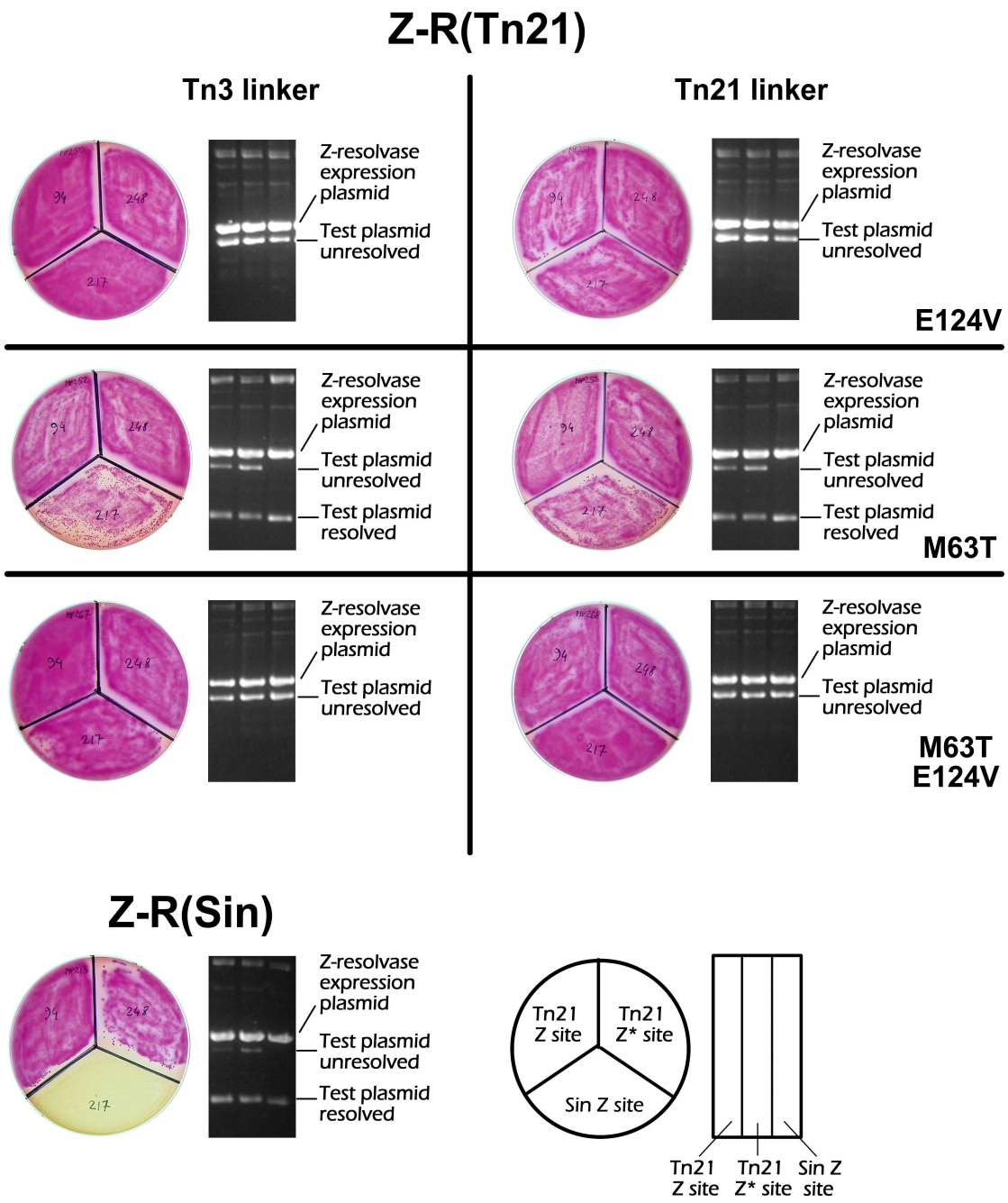


Figure 4.7- A MacConkey assay results with Tn21 Z-resolvases, Z-R(Tn21, 63T, L3), Z-R(Tn21, 63T, L21), Z-R(Tn21,124V, L3), Z-R(Tn21,124V, L21), Z-R(Tn21,63T/124V, L3) or Z-R(Tn21,63T/124V, L21) and Sin Z-resolvase Z-R(Sin) acting on substrates containing two copies of either Sin Z-site, Tn21 Z-site or Tn21 Z-site* (pMP217, pMP94 and pMP248, respectively). How the different substrates that Z-resolvases are acting on are arranged on plates and the corresponding agarose gels is indicated in the explanatory key (bottom right). The plates and the corresponding agarose gels showing the results for the Tn21 Z-resolvases are arranged in a grid. The type of linker used is indicated at the top of each column, whereas the mutations present in the resolvase catalytic domain are indicated on the right of each row. A plate and the corresponding agarose gel showing the result for Z-R(Sin) is found in the bottom left corner of the figure. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

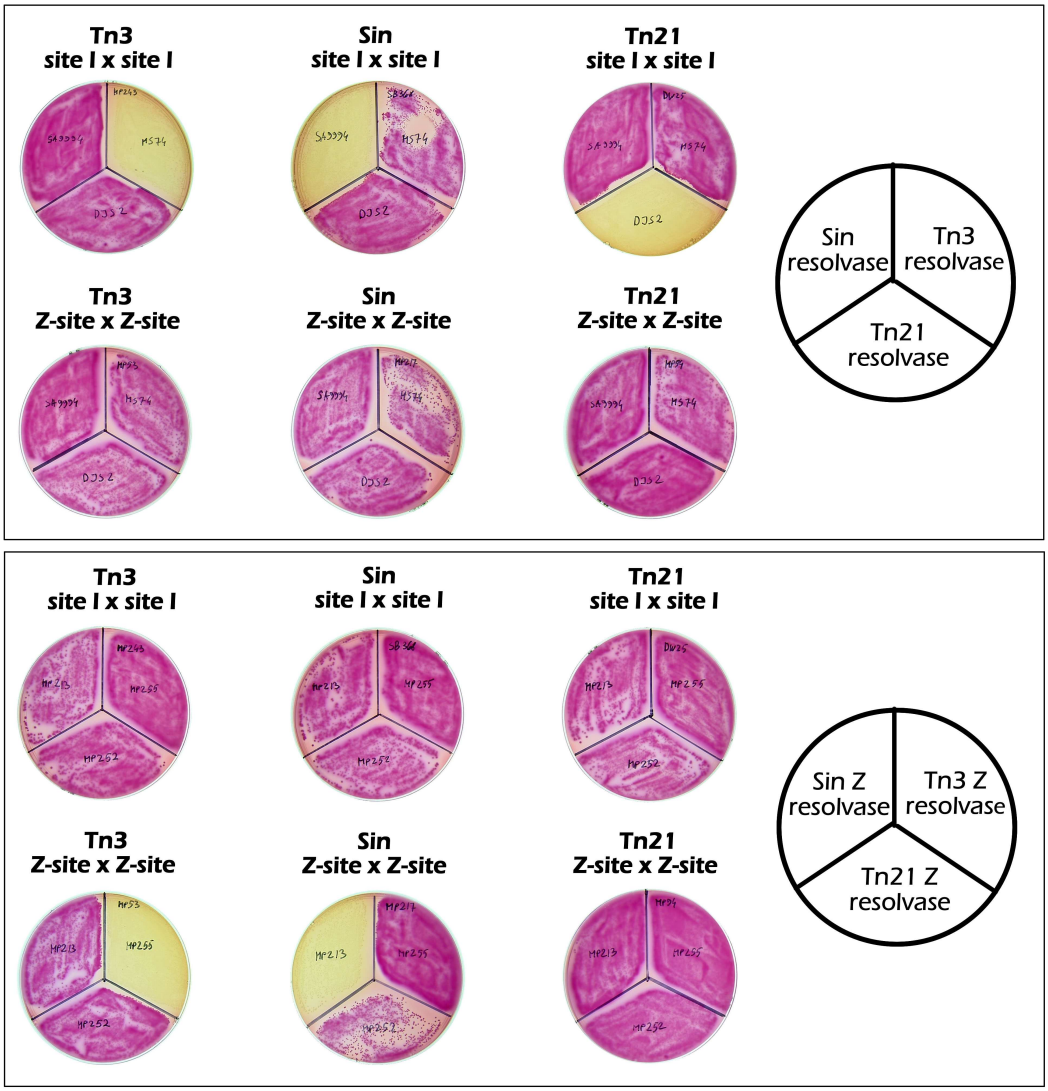


Figure 4.8- A MacConkey assay results with the Tn3 resolvase G101S, D102Y, Sin resolvase Q115R and Tn21 resolvase M63T acting on substrates containing two copies of either Tn3 site I, Sin site I, Tn21 site I, Tn3 Z-site, Sin Z-site or Tn21 Z-site (top panel). In the bottom panel, MacConkey assay results with Tn3 Z-resolvase Z-R(SY), Sin Z-resolvase Z-R(Sin) and the Tn21 Z-resolvase Z-R(Tn21, 63T, L3) acting on the same set of substrates are shown. How the different resolvases or Z-resolvases are arranged on plates is indicated in the explanatory key (to the right of the plate photographs). The type of substrate used is indicated on top of the each of the plate photographs. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

MacConkey assay (Section 2.23) on resolution substrates (pMP52-pMP55) containing the Z-sites Z20Z, Z22Z, Z24Z and Z26Z, with the central sequence derived from Tn3 *res* site I (Fig. 5.1). A range of Tn3 site I sequence-derived Z-sites with differently spaced ZBS were tested in case the 22 bp spacing is actually suboptimal for Z-R(Sin). However, as is evident from Fig. 4.9, no trace of resolution can be observed on any of these substrates even after the plasmid DNA was isolated and analysed using agarose gel electrophoresis.

The fact that no resolution was observed with Z-R(Sin) acting on pMP52-pMP55, substrates containing Z-sites with Tn3 site I-derived central sequence in their centre, suggests that Z-R(Sin) exhibits sequence selectivity at the centre of the Z-site, akin to what was observed with Z-R(NM) (Section 4.2). The ability of Z-R(Tn21, 63T, L3) and Z-R(Sin) to work on each other's sites (Fig. 4.7) shows that their sequence preferences in the centre of the Z-site must be quite similar. To determine the exact nature of these preferences and to better characterise these proteins, it was necessary to purify them and test their activity *in vitro*.

The proteins chosen for purification were Tn21 resolvase M63T, Sin resolvase Q115R and their Z-resolvase analogues Z-R(Tn21, 63T, L3) and Z-R(Sin) as they were shown to be active *in vivo* (Fig. 4.7). Protein over-expression plasmids, pMP137, pMP138, pMP371 and pMP368, expressing these proteins respectively, were constructed as described in Section 2.7.1. The proteins were purified using the adapted denaturing methodology originally designed for purification of Tn3 resolvase mutants (Section 2.25). This protocol proved less than ideal, but due to time constraints was not further optimised. The main problem encountered was the low solubility of the Z-resolvases, with only a fraction of the total going back into solution at the re-solubilising stage.

The proteins were tested for recombination activity in a time course experiment (Section 2.27). Reactions were set up at 37°C, with samples taken at 0.5, 1, 2, 5, 15, 30 and 60 minute time points and stopped by heat. Half of each sample was digested with XhoI enzyme while the other half was left uncut. Uncut and cut samples were analysed using agarose gel electrophoresis.

Disappointingly, the purified Tn21 resolvase mutant M63T and its Z-derivative Z-R(Tn21, 63T, L3) proved inactive when tested with *in vitro* substrate plasmids pMP390 (Tn21 site I x site I) and pMP6 (Tn21 Z site x Z-site), respectively (data not shown). Z-R (Sin) and Sin resolvase Q115R however, were active (Fig 4.10) and recombined *in vitro* substrate plasmids pMP240 and pMP378, respectively. Sin resolvase Q115R was slightly faster at

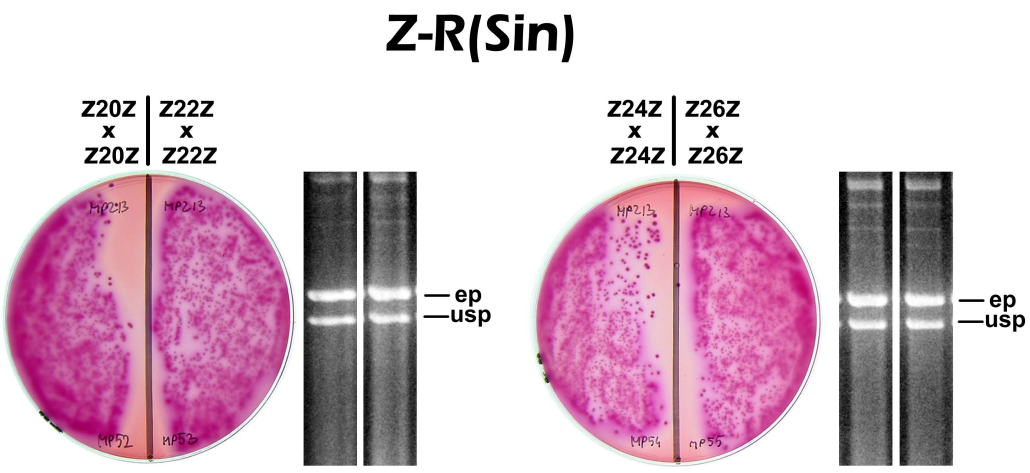


Figure 4.9- A MacConkey assay results with the Sin Z-resolvase Z-R(Sin) acting on substrates (pMP52-pMP55) containing the Z-sites Z20Z, Z22Z, Z24Z and Z26Z, with the central sequence derived from Tn3 *res* site I (for sequences see Fig 5.1). The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, "ep" Z-resolvase expression plasmid, "usp" unresolved substrate plasmid. The type of substrate used is indicated on top of the each of the plate photographs. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

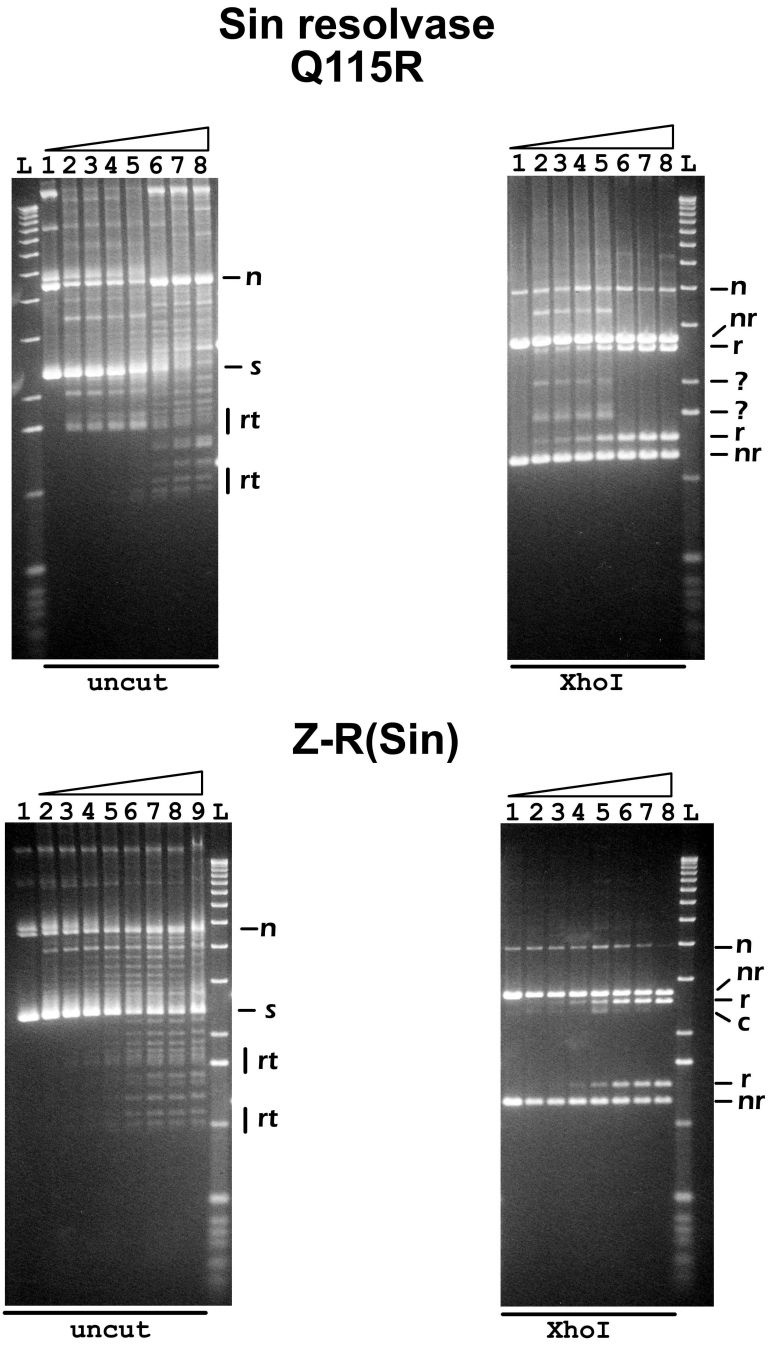


Figure 4.10- In vitro recombination reactions time course catalysed by the activated Sin resolvase Q115R and the Sin Z-resolvase Z-R(Sin), acting on substrates pMP378 (Sin site I × Sin site I) and pMP240 (Sin Z-site × Sin Z-site) respectively. The time points are marked at the top of the gel as follows, 1 (0 min.), 2 (0.5 min.), 3 (1 min.), 4 (2 min.), 5 (5 min.), 6 (15 min.), 7 (30 min.), 8 (60 min.) and 9 (120 min.). For protein and DNA concentration and buffer composition see Section 2.27. The gels on the left show the uncut reactions whereas the ones on the right show the corresponding reaction cut with XhoI. The graphic representation of XhoI digest of the possible reaction products giving their expected sizes is given in Fig. 2.14. The abbreviations are as follows, “s” (supercoiled substrate), “n” (nicked substrate), “rt” (recombinant product topoisomers), “nr” (non-recombinant products), “r” (recombinant product), and “c” (cleavage product). The DNA bands marked with “?”are probably single-stranded circular DNA molecules, see text for details (Section 4.7).

recombining its substrate plasmid (pMP378) than Z-R(Sin) acting on pMP240. However, with both proteins evidence of resolution products were observed even at the earliest time point. As the reaction progressed, more and more of the supercoiled substrate plasmid was converted into recombinant products. In both cases in the unrestricted samples recombinants migrated as a ladder of topoisomers running faster than the supercoiled substrate. In order to simplify the analysis samples were restricted with XhoI which cleaves the reaction products in a diagnostic fashion (Fig 2.14). As can be seen in Fig. 4.10 the main product of the reactions were the big and small recombinant circle. Small amount of cleavage product was generated by Z-R(Sin) at the early time points but after 15 minutes all of the cleavage product appeared to be religated as no cleavage product is visible on the later time points. The bands marked with "?" in the gels showing the Sin resolvase Q115R reactions and the band running slower than the supercoiled substrate band, that are visible at time points at 0.5, 1, 2, and 5 minutes in both the uncut and XhoI digested samples, are most likely due to the single-stranded recombinant and substrate circles, respectively. These single-stranded circles were generated by the denaturation of substrate and recombinant molecules caused by keeping the samples at high temperature for too long after terminating the reaction. As these molecules were single-stranded they could not be cleaved by the XhoI.

4.8 Discussion

In the first part of this chapter, using a multimer resolution-based strategy, recombination-proficient Z-sites were selected using purified Z-R(NM), a Z-resolvase with the catalytic domain of NM resolvase. Analysis of these sequences revealed that Z-R(NM) displays sequence selectivity in the centre of Z-site that is analogous to the sequence selectivity displayed by NM resolvase (see Chapter 3). This suggested that sequence selectivity in the centre of the Z-site may be determined by the same arm region-DNA minor groove contacts as were discussed in Chapter 3, and that if Z-sites with different (e.g. GC-rich) central sequences are to be targeted, new Z-resolvases utilising the catalytic domains of resolvases associated with appropriate (in this example GC-rich) site I's need to be constructed.

In the second part of the chapter a generally applicable step-by-step procedure for generating new Z-resolvases was described (Section 4.3). By following these steps, using the catalytic domains of available hyperactive mutants of Sin resolvase (Q115R) and Tn21 resolvase (M63T; E124V; and the double mutant M63T, E124V), seven new Z-resolvases were constructed.

Surprisingly, only the Z-resolvases with the catalytic domains of Sin resolvase mutant Q115R and Tn21 resolvase mutant M63T were shown to be active *in vivo* (Section 4.7). Z-resolvases constructed using the catalytic domain of Tn21 resolvase containing the mutation E124V, either as a single mutation or in combination with M63T, were shown to be inactive. Although the mutation E124V appears just as activating as M63T in the Tn21 resolvase context, in the Z-resolvase context this mutation is not activating, but inhibiting. The reason for this effect is not yet clear.

The activating mutations E124V and M63T in Tn21 resolvase were shown not to be additive as the double mutant appeared to be less active than either of the single mutants. According to the sequence alignment, (and comparison to the γδ resolvase structures) activating mutations E124 and M63 in Tn21 resolvase are on opposite sides of the interface between the E-helix and the main body of the catalytic domain. Either of these mutations probably increases the flexibility of this interface leading to activation; however, in a double mutant the interface might be too flexible, resulting in a reduction of recombination efficiency. The mutation E124V appears to be dominant over the mutation M63T, especially in the Z-resolvase context, as the Z-resolvases with catalytic domains containing both mutations are inactive, like the ones constructed using a catalytic domain with mutation E124V.

Apart from the mutations in the catalytic domain, another variable in the design of the Z-resolvases with the catalytic domains of Sin and Tn21 resolvase was the length and the sequence of the linker connecting the catalytic domain to the Zif268 DNA-binding domain. In Z-R(Sin), Z-R(Tn21, 63T, L3), Z-R(Tn21, 124V, L3) and Z-R(Tn21, 63T/124V, L3) a linker based on the final four amino acid residues of the extended arm region of Tn3 resolvase followed by the sequence TS was employed, whereas in the Z-R(Tn21, 63T, L21), Z-R(Tn21, 124V, L21) and Z-R(Tn21, 63T/124V, L21) a three amino acid residues longer linker, based on the equivalent Tn21 resolvase extended arm region was used. The Z-resolvases containing the shorter Tn3 resolvase sequence-based linker seemed to be slightly more active than the ones containing the longer linker. This agrees well with the results obtained in the previous work of our group (A. MacLean, unpublished) and the results presented in the following chapter, suggesting that a shorter linker is preferred. A possible explanation for the preference for the shorter linker is presented in Section 5.7.

Although Tn21 resolvase and Sin resolvase are only distantly related, their catalytic domains are able to catalyse recombination using Z-sites derived from either Tn21 site I or Sin site I sequence. This is not replicated when hyperactive mutants of the intact Sin or

Tn21 resolvase are used, with mutants of each of these resolvases catalysing recombination only with substrates containing their native site Is. In contrast, Z-R(Sin) does not recombine substrates containing Z-sites with the central sequence based on Tn3 site I.

These results suggest that the sequence selectivity of the Tn21 resolvase and Sin resolvase catalytic domains might be very similar despite their evolutionary distance, and that this selectivity is quite different to that displayed by the Tn3 resolvase catalytic domain. This possibility is further explored in Chapter 6.

Although in theory any serine recombinase could be used to generate hyperactive mutants that can act as catalytic domain donors, in practice this might prove difficult. Out of the multitude of genes identified as serine recombinase candidates in the DNA sequence databases, only a handful have been studied *in vitro,* which is currently a prerequisite for generating hyperactive mutants that do not require accessory sites. Even if an *in vitro* system is established selecting suitable activating mutations could be quite laborious. The use of known hyperactive mutants of various serine recombinases as a starting point for their Z-resolvase design by Barbas' group (Gordley *et al.,* 2007) highlights this point.

Unlike Sin and Tn21 resolvase which can become activated by a single mutation, there is no known single change that can make Tn3 resolvase hyperactive enough to recombine a substrate containing two copies of Tn3 site I. Serine recombinases that are more repressed, like Tn3 resolvase, might need a combination of mutations in order to be hyperactive enough to be suitable as catalytic domain donors for new Z-resolvases. Additionally, as observed with the Tn21 resolvase mutation E124V, not all activating mutations are suitable in the Z-resolvase context, emphasising the need for further study of the properties of resolvase hyperactive mutants and Z-resolvases in general.

5 Chapter 5: Studies on the interactions of Z-resolvases with Z-sites

5.1 Introduction

Prior to this study, a number of functional Tn3 Z-resolvases have been constructed and tested *in vivo* (Akopian *et al.,* 2003; Gordley *et al.,* 2007, Gordley *et al.,* 2009, A. McLean, unpublished). The Z-resolvases constructed in our group (Akopian *et al.,* 2003, A. McLean unpublished) varied in the N-terminal (catalytic) domain cut-off point, mutations in the catalytic domain used, and the length and the sequence of the linker used to connect the catalytic domain to the Zif268 zinc finger domain.

The design of the Z-site was also varied. Optimal recombination activity was reported with Z-sites that are 40 bp long, that is, with 22 bp of sequence between the two 9 bp zinc finger binding sites (ZBSs) oriented in inverted repeat, irrespective of the Z-resolvase structure (Akopian *et al.,* 2003). In contrast, according to the work by Barbas'group (Gordley *et al.,* 2007; Gordley *et al.,* 2009.) the optimal length of central sequence was found to be 20 bp. The Z-resolvases used by this group were of very similar design to the ones used by Akopian (Akopian *et al.,* 2003) but used a didactyl zinc finger domain (that is, two zinc fingers) instead of tridactyl Zif268. Akopian on the other hand had found that a Z-resolvase of his design with a didactyl zinc finger domain was inactive (Akopian, 2003).

These differences in the reported results highlight a need to clarify these and other outstanding issues regarding Z-resolvase design. In this chapter the following issues will be addressed:

- The length of Z-sites: what is the optimum spacing of the Zif268-binding sites? (Section 5.2)

- The effect of the Zif268 DNA-binding domain on catalytic activity. Does putting a resolvase catalytic domain into a Z-resolvase context have an activating or inhibitory influence? (Section 5.3)

- The effect of asymmetric Z-sites on recombination. Can a Z-resolvase catalyse recombination on sites with an odd number of bases between ZBSs i.e one half-site longer than the other (Section 5.4)

- The relative influences of the Z-resolvase linker length and the V107F mutation. (Section 5.5)

- The possibility of complementation of Z-resolvase by resolvase protein on designed hybrid sites. (Section 5.6)

5.2 The optimal length of Z-sites

Since the Tn3 site I sequence is an imperfect inverted repeat in the centre, and the flanking ZBSs create a further 18 bp of inverted repeat symmetry, the original Z-sites constructed by Akopian (Fig.5.1) (Akopian *et al.*, 2003; marked by AA from now on) contained engineered base changes, introduced in order to avoid problems associated with cloning inverted repeat sequences. In the previous chapter, it was observed that *in vivo* resolution substrates containing Tn21-Z sites* (Section 4.7), a version of a Tn21-Z site with a AG to TC change introduced to reduce the length of the inverted repeat, were recombined less efficiently than substrates containing Tn21-Z sites without such a change. By an accident of Akopian's design, the sequence with 22 bp spacing was also the sequence with the greatest conservation of native Tn3 site I bases in the outer parts of the central sequence (Fig.5.1).

To establish to what extent the recombination efficiency is influenced by the spacing, as opposed to sequence identity to the native crossover site, a series of Z-sites was made with increasing distance between Zif268-binding sites (ZBSs), following the previous design, but without any random bases (Fig 5.1). Inter-ZBS spacing ranged from 16 to 28 bp, each consecutive Z-site being two bases longer. The 2 bp increments were achieved by increasing the extent of Tn3 site I sequence until the final Z-site in which an entire 28 bp Tn3 site I was flanked by ZBSs. A series of *in vivo* resolution substrate plasmids containing these sites was constructed (pMP50-pMP56) and the *in vivo* substrate plasmids containing AA Z-sites were remade (pMP89-93) to allow for direct comparison (Section 2.7.2).

The original plasmids containing AA Z-sites, used by Akopian, were remade as they were found to have a mutation resulting in altered *gal*K expression, thus causing a delay in the colour change in the MacConkey assay, requiring the bacteria to be grown for several days at room temperature before distinct colour phenotype could be observed.

Akopian's tridactyl Z-resolvases AN8 and AN15, and a didactyl AN28 Z-resolvase were tested for recombination *in vivo* using the MacConkey assay on each series of resolution

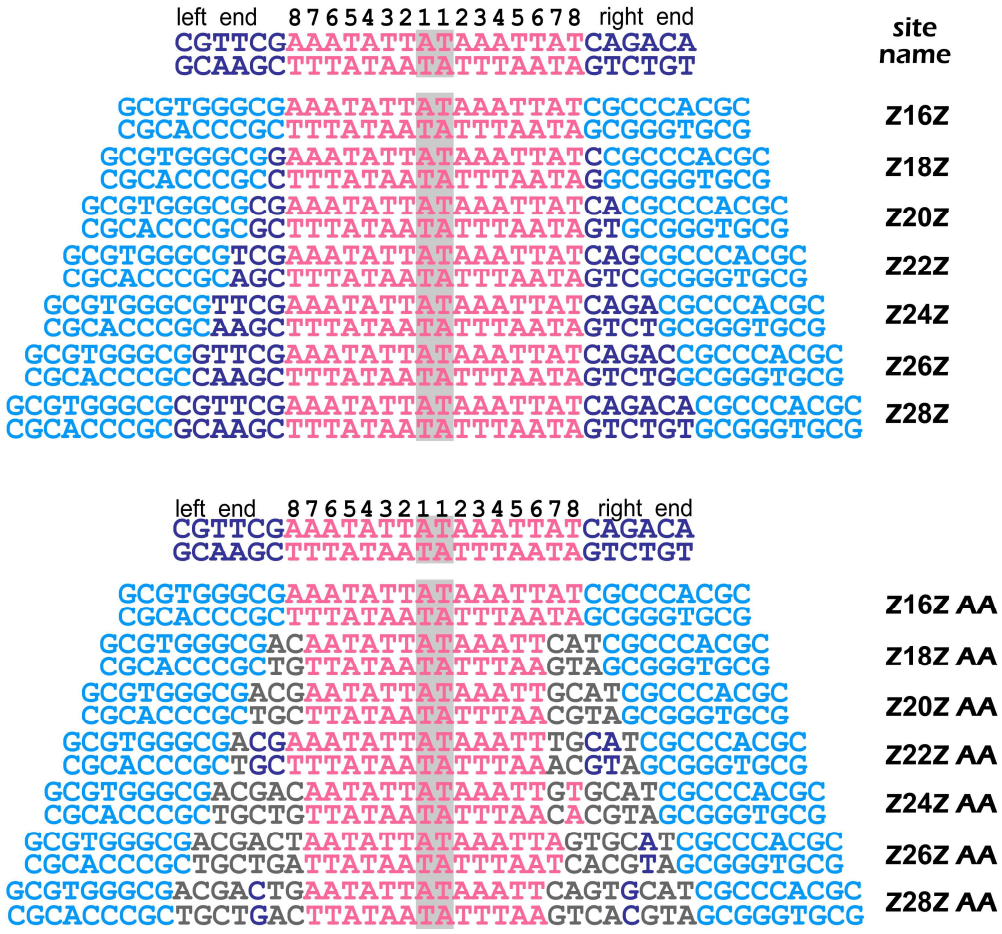


Figure 5.1-Sequences of Tn3 site I-based Z-sites. The sequences of the newly designed Z-sites are given in the top panel (Z16Z-Z28Z) while the original Z-sites designed by Akopian are given in the bottom panel (Z16Z AA-Z28Z AA). Sequence given in dark blue is the sequence usually bound by the resolvase HtH domain. Light blue sequence is the sequence bound by the Zif268 DNA-binding domain (ZBS-Zif268 binding site). The central 16 bp coming from the Tn3 site I sequence is pink. Random base changes introduced by Akopian are given in gray. The central dinucleotide is marked by the grey rectangle. The individual DNA bases are numbered from the centre. The "Left" and "Right" end refer to the conventional representation of site I within *res*, where the binding sites are in the order I-II-III left to right.

substrates (Fig 5.2). AN8 and AN28 Z-resolvase consist of residues 1-148 of NM resolvase and have the linker sequence SSDPTSQTS, differing only in the number of zinc fingers in their C-terminal domain (Zif268 residues 2-90 (3 fingers) and 2-60 (2 fingers) respectively). On the other hand, AN15 Z-resolvase is made from residues 1-144 of NM resolvase with the linker sequence GSGGSGGSGGSGTS connecting it to the Zif268 DNA-binding domain (Zif268 residues 2-90). AN28 Z-resolvase produced red colonies as it failed to recombine any of the substrates tested. This confirmed an earlier result (Akopian, 2003) that Z-resolvases with two zinc fingers are not as active as the ones with three zinc fingers in their DNA binding domain. AN8 and AN15 Z-resolvases acting on substrates pMP52 and pMP53, harbouring Z-sites Z20Z and Z22Z respectively produced white colonies exclusively, indicating efficient *in vivo* resolution. These Z-resolvases also produced a small number of white colonies with a substrate with two copies of Z24Z (pMP54), but resolution was far less efficient than with pMP52 and pMP53. AN8 Z-resolvase recombined substrates pMP90 (Z20ZAA × Z20ZAA) and pCP132 (Z22ZAA × Z22ZAA) as efficiently as it did pMP52 and pMP53, giving white colonies. Resolution of pMP90 by AN15 Z-resolvase was much less efficient than on pCP132, resulting in a mixture of pink and red colonies.

Using the same substrate plasmids the tridactyl Z-resolvase, Z-R(NMF)(A. McLean, unpublished), was also tested (Fig 5.3). At the beginning of this study it was believed that this Z-resolvase was the most active, having been selected in a screen for fast resolution. The N-terminal domain of this Z-resolvase (residues 1-148) is from NM resolvase with the additional mutation V107F. The linker sequence consists of only two amino acids, TS. The results observed with Z-R(NMF) were comparable to those with AN15 Z-resolvase. Z-R(NMF) recombined substrates pMP52, pMP53 and pCP132 producing white colonies. As with AN15, Z-resolvase resolution on pMP90 was less complete, and a mixture composed of mostly red and some white colonies was obtained. In order to check whether any traces of resolution were detectable with any other substrate plasmid, a representative sample of colonies was scraped off the MacConkey agar plates and grown in liquid culture at 37°C overnight. Plasmid DNA was isolated (Section 2.10.1) from the overnight cultures and analysed using agarose gel electrophoresis (Section 2.13.1). The results obtained with agarose gel electrophoresis corresponded very well with the observed colony colours. No trace of resolution was observed with any of the substrates that gave only red colonies. Interestingly, although it seemed that all of the unresolved substrate plasmid has been used up in the case of pMP52 and pMP53, the amount of resolved substrate DNA observed was less than that seen in the pCP132 reaction. It is possible that this effect is due to the

Name	mutations	N-terminus cut-off	Linker sequence	№ of ZF
AN8	NM	R148	SSDPTSQTS	3
AN28	NM	R148	SSDPTSQTS	2
AN15	NM	R144	GSGGSGGSGGSGTS	3

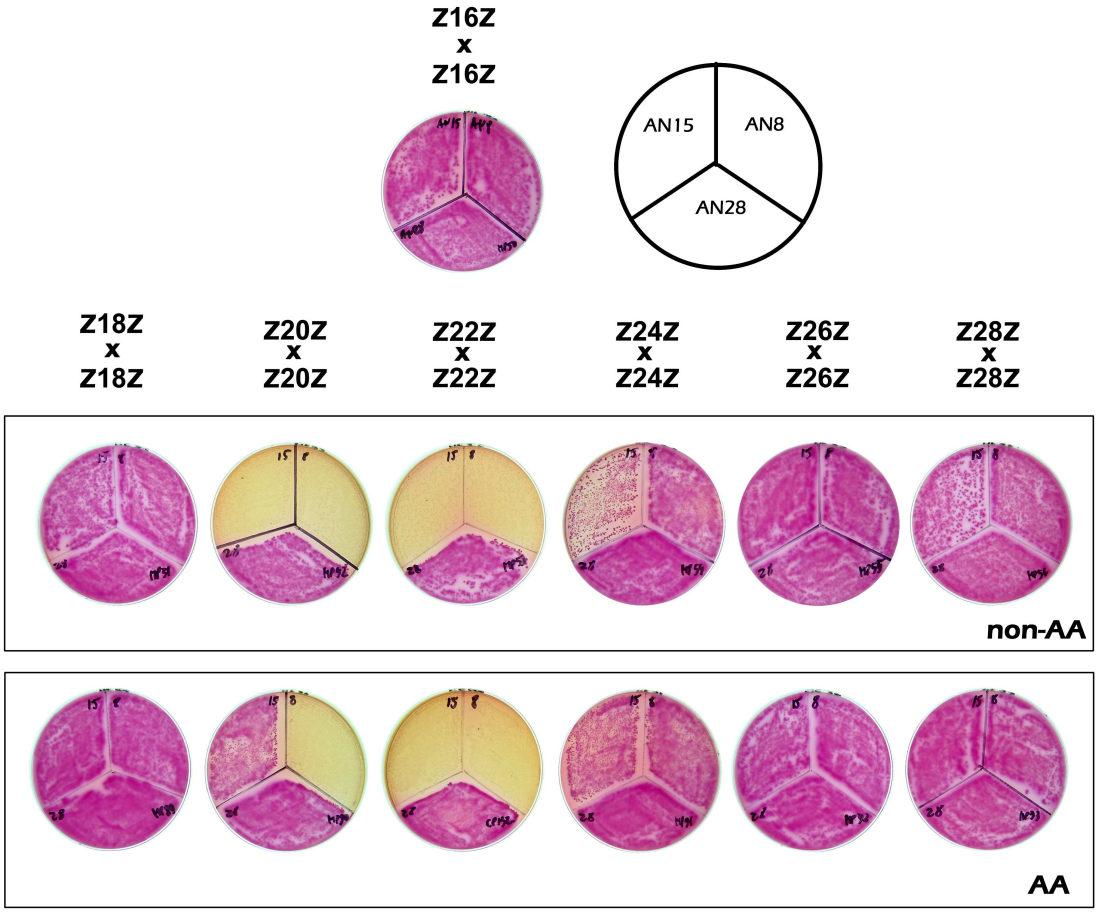


Figure 5.2- A MacConkey assay results with the Z-resolvases AN8, AN15 and AN28 acting on substrates pMP50-pMP56 containing the Z-sites Z16Z-Z28Z and pMP89-93, pCP132 containing Z-sites Z18Z-Z28Z AA (for sequences see Fig 5.1). A table listing the parameters that these Z-resolvase differ in, such as the cut-off point for the resolvase catalytic domain (N-terminus cut-off), the linker sequence and the number of zinc fingers in the DNA-binding domain is given at the top of the figure. How the different resolvases or Z-resolvases are arranged on plates is indicated in the explanatory key (middle right). The type of substrate used is indicated on top of the each of the plate photographs (top panel non-AA, bottom panel AA sites). Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

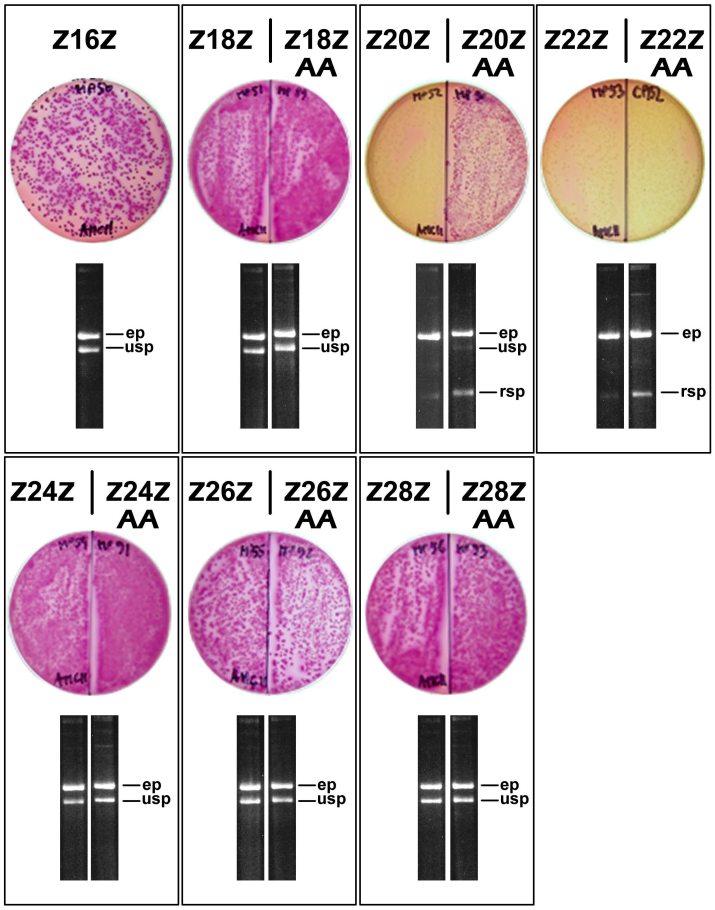


Figure 5.3- A MacConkey assay results with the Z-resolvase Z-R(NMF) acting on substrates pMP50-pMP56 containing the Z-sites Z16Z-Z28Z and pMP89-93, pCP132 containing Z-sites Z18Z-Z28Z AA (for sequences see Fig 5.1). The type of substrate used is indicated on top of the each of the plate photographs. The agarose gels show the uncut plasmid DNA isolated from the plate above. DNA bands are marked as follows, "ep" Z-resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

destructive properties of the NMF catalytic domain. Plasmids that contain Z-sites with purely Tn3 site I-derived sequence in the centre could be more easily resolved than the ones containing base changes such as the AA series of Z-sites. If the Z-resolvase is very hyperactive (‘out of control’) it could be cleaving the substrates without efficiently religating them. Such free cleaved substrates would be likely promptly degraded by the cell which might cause a reduction in the amount of resolved plasmid DNA observed.

Based on the results presented to this point it was difficult to say with certainty which Z-site provides the optimal spacing, the one with 22 or the one with 20 bp of sequence between the two ZBSs. In order to ascertain which is the optimal Z-site architecture, a test using less active Z-resolvases was required. The reasoning for this was that a protein that recombines slowly would provide a far more stringent site selection than the very active Z-resolvases that were tested so far. Less active Z-resolvases were made using the catalytic domains of known less activated Tn3 mutants.

5.3 Effects of DNA binding domain on catalytic activity

Using the Z-R(NM) expressing plasmid pMP59 as a vector, an NdeI/EagI fragment swap was performed with four Tn3 resolvase mutant expression plasmids, namely pMS68 (D102Y, E124Q; “YQ”), pMS74 (G101S, D102Y; “SY”), pFO5 (R2A, E56K, G101S, D102Y; “AKSY”), pJH2 (G101S, D102Y, M103I, Q105L; “M”),(Table 2.5; Olorunniji *et al.,* 2008) to make the Z-resolvase expressing plasmids pMP254, pMP255, pMP256 and pMP257, respectively. Z-resolvases expressed from plasmids pMP254, pMP255, pMP256 and pMP257 were called Z-R(YQ), Z-R(SY), Z-R(AKSY) and Z-R(M).

The Tn3 resolvase mutants used as catalytic domain donors are known to be less active than the NM or NMF mutants used in the Z-resolvases discussed so far (Olorunniji, 2006; Olorunniji *et al.,* 2008). What was unclear was whether the activity of these catalytic domains would be aided or inhibited by the conversion into a Z-resolvase. One could argue that the substitution of a tighter binding, highly specific Zif268 DNA-binding domain might increase recombination efficiency. On the other hand, a fusion protein of two domains that have not co-evolved might be inherently less active than the native protein. To test which of these two alternatives is true, a direct *in vivo* comparison of the activities of the Tn3 resolvase mutants on pMP243 (site I × site I) with the activities of Z-resolvases containing catalytic domains derived from the same mutants on pMP53 (Z22Z × Z22Z) was performed (Fig 5.4).

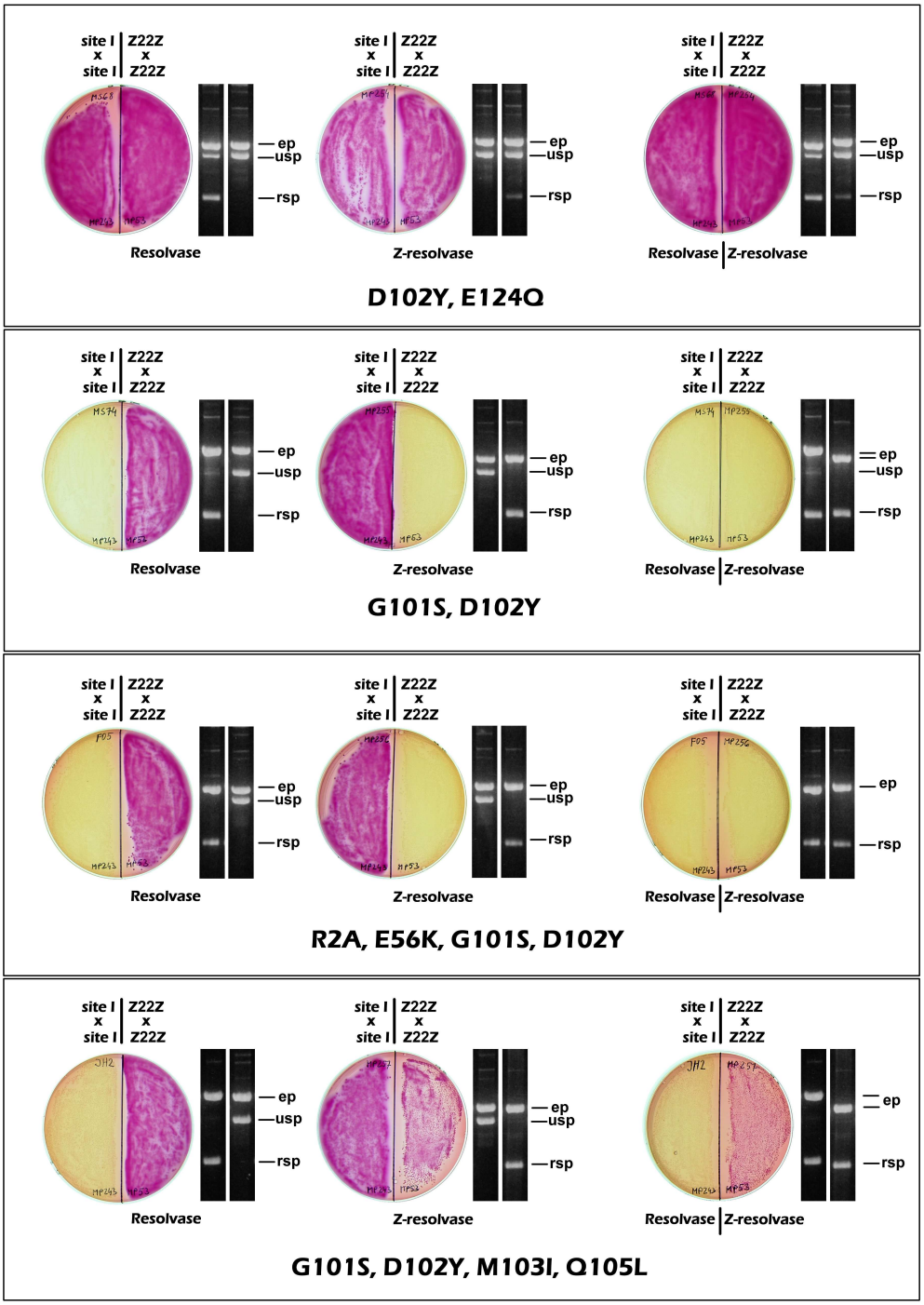


Figure 5.4- An in vivo comparison of Z-resolvases Z-R(YQ), Z-R(SY), Z-R(AKSY) and Z-R(M) acting on a substrate pMP53 (Z22Z × Z22Z) with the activated Tn3 resolvase mutants YQ, SY, AKSY and M acting on substrate pMP243(Tn3 site I × Tn3 siteI). The type of substrate used is indicated on top of the each of the plate photographs. The mutations in the resolvase catalytic domain are marked at the bottom of each panel. Whether the cells on a plate section express a resolvase or a Z-resolvase is marked under the plate photographs. The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, “ep” resolvase (Z-resolvase) expression plasmid, “usp” unresolved substrate plasmid “rsp” resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

As can be seen from the *in vivo* comparison (Fig.5.4), the catalytic domain shows a comparable level of activity when it is in a resolvase or a Z-resolvase context. Putting the catalytic domain into a Z-resolvase context is certainly not activating and in the case of the very weakly activated mutant D102Y, E124Q (YQ) it is actually mildly inhibiting. When the proportion of unresolved to resolved *in vivo* substrate plasmid DNA seen with resolvase YQ is compared with Z-resolvase with the same catalytic domain Z-R(YQ), it is clear that resolution catalysed by Z-R(YQ) is less efficient. Similarly, recombination of pMP53 by Z-resolvase Z-R(M) produced a mixture of red and white colonies, while the resolvase with the same catalytic domain when acting on pMP243 gave only white colonies.

This corresponds well with the Tn21-Z resolvase results described in the previous chapter. Tn21 hyperactive mutants E124V and M63T both gave white colonies in a MacConkey assay when acting on pDW25 (Tn21 site I × Tn21 site I) substrate, whilst Z-resolvases made using the same catalytic domains, Z-R(Tn21, 63T, L3), Z-R(Tn21, 63T, L21), Z-R(Tn21, 124V, L3) or Z-R(Tn21, 124V, L21) never produced purely white colonies on pMP94, the equivalent Tn21 Z-sites-containing substrate.

Since Z-resolvases made using less hyperactive catalytic domains were shown to be active on the potentially most optimal Z22Z site, it was decided to use them to attempt and identify which is the most optimal Z-site architecture. The recombination activity of the four less efficient Z-resolvases, Z-R(YQ), Z-R(SY), Z-R(AKSY) and Z-R(M), was assessed *in vivo* on substrates pMP52, pMP90, pMP53, pCP132, pMP54, pMP91 containing two copies of Z20Z, Z20Z AA, Z22Z, Z22Z AA, Z24Z and Z24Z AA Z-sites, respectively (Fig. 5.5). Based on the colony colours and the analysis of isolated plasmid DNA by the agarose gel electrophoresis in this set of experiments, the Z-site that is the easiest to resolve is Z22Z, as resolution was observed with all Z-resolvases tested. The substrate pMP53 containing two copies of Z22Z is the only site that Z-R(YQ) showed any activity on. With Z22Z as the easiest to resolve, other Z-sites tested could be arranged in the descending order as follows: Z22Z AA, Z20Z, Z20Z AA and finally Z24Z and Z24Z AA as a joint last. Since none of the Z-resolvases tested showed any trace of resolution of *in vivo* substrates containing either Z24Z or Z24Z AA (pMP54 and pMP91, respectively), it was impossible to distinguish between these two sites. From the result of this experiment it is evident that there is a degree of flexibility in the Z-site length requirement with distances of 20 to 22 bp between ZBSs being tolerated, but with 22 bp distance being preferred.

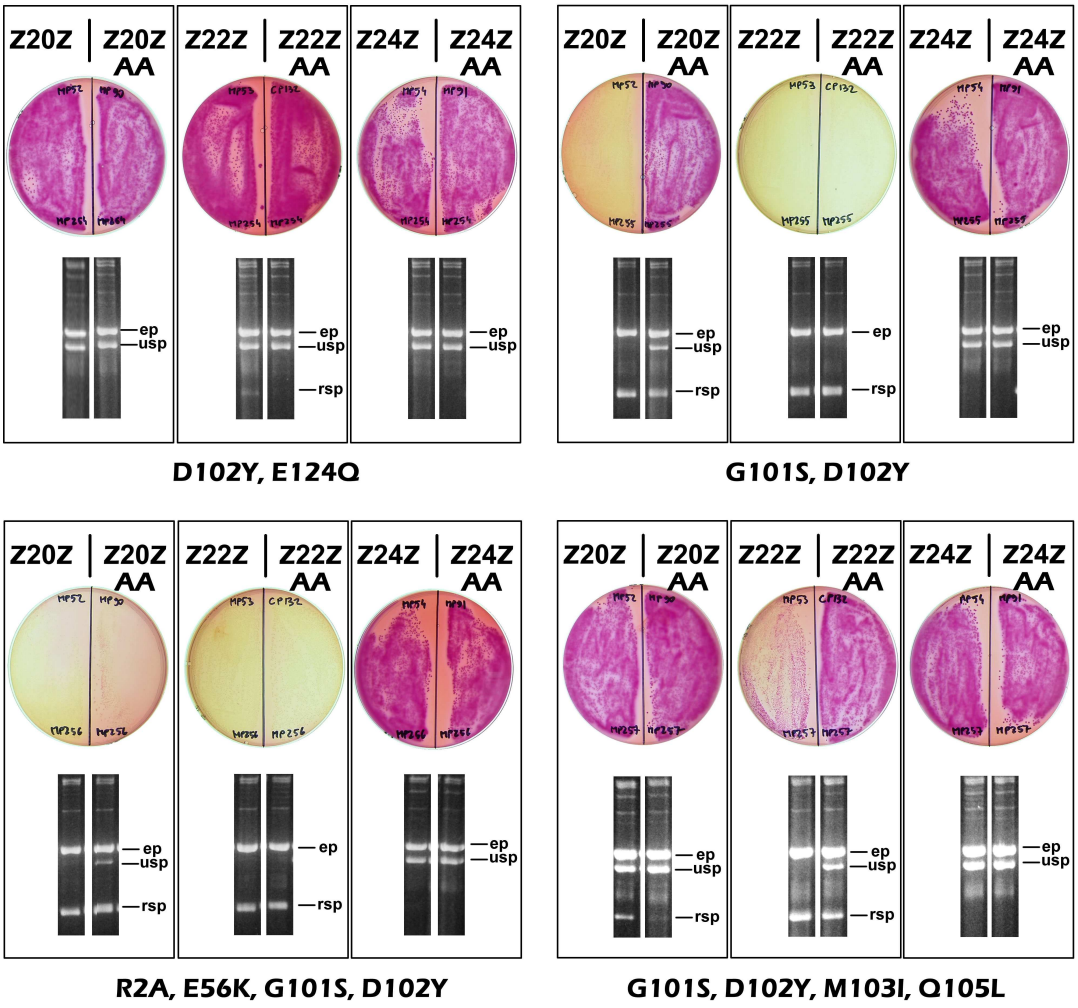


Figure 5.5- A MacConkey assay results with the Z-resolvases Z-R(YQ), Z-R(SY), Z-R(AKSY) and Z-R(M) acting on substrates pMP52 (Z20Z × Z20Z), pMP90 (Z20Z AA × Z20Z AA), pMP53 (Z22Z × Z22Z), pCP132 (Z22Z AA × Z22Z AA), pMP54 (Z24Z × Z24Z), pMP91 (Z24Z AA × Z24Z AA) (for sequences see Fig 5.1). The type of substrate used is indicated on top of the each of the plate photographs. The mutations in the Z-resolvase catalytic domain are marked at the bottom of each panel. The agarose gels show the uncut plasmid DNA isolated from the plate above. DNA bands are marked as follows, "ep" Z-resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

5.4 Recombination on asymmetric sites

As shown in the previous section using the Z-resolvases with mildly hyperactive catalytic domain, the optimum distance between "symmetric" Zif268 binding sites is 22 bp, rather than 20 bp or 24 bp. An intriguing possibility that had not been explored to date was that the optimal site could be asymmetric, with a preferred distance between ZBSs of 21 or 23 bp.

To investigate this scenario new Z-sites were designed, with either 21 and 23 bp of distance between ZBSs (Fig 5.6). Since these sites are not symmetric, two versions of each Z-site length were made, namely Z21ZL, Z21ZR, Z23ZL and Z23ZR. The sites with 21 bp spacer have 10 bp on one side of the centre and 11 bp on the other; those with 23 bp spacer have 11 bp on one side and 12 bp on the other. The letters L and R in the names of these sites signify which, the left (L) or the right (R) half-site is the longer one. Four *in vivo* resolution substrate plasmids containing pairs of these sites were created (Section 2.7.); pMP201 (Z21ZR × Z21ZR), pMP202 (Z21ZL × Z21ZL), pMP203 (Z23ZR × Z23ZR) and pMP204 (Z23ZL × Z23ZL).

Six Z-resolvases that differed in the level of hyperactivity displayed by their catalytic domain were tested for resolution activity using the asymmetric *in vivo* substrates pMP201-204; Z-R(YQ), Z-R(SY), Z-R(AKSY), Z-R(M), Z-R(NM) and Z-R(NMF) (Fig 5.7). The least hyperactive, Z-R(YQ), failed to resolve any of the asymmetric Z-site substrates, producing red colonies with all substrates tested. All the other Z-resolvases tested promoted at least partial resolution of all the substrates used. The most complete resolution was observed with substrates pMP201 and pMP202 containing sites Z21ZR and Z21ZL. Out of the other two asymmetric Z-sites tested, Z23ZL was the harder one to resolve, with all Z-resolvases tested achieving only partial resolution, the only exception being Z-R(NM). One possible explanation for this is that the 12 bp distance between the centre of the site and the ZBS is too far for the proper arm region contacts to be established in the minor groove. While this does not seem to affect the highly hyperactive Z-resolvases such as Z-R(NM), Z-resolvases containing only weakly hyperactive catalytic domains are finding it harder to compensate for the lack of these contacts if the left half-site is the longer one. It could be that the arm region contacts in the left half-site are more important than the ones on the right half-site. This was the case observed with NM resolvase in Chapter 3, where it was reasoned that the stronger binding of the resolvase HtH domain to the outermost 6 bp of the right half-site renders these affinity-enhancing arm region

Figure 5.6- Sequences of asymmetric Z-sites, Z21ZR, Z21ZL, Z23ZR, and Z23ZL. The symmetric Z-site Z22Z is provided for comparison (top). Sequence given in dark blue is the sequence usually bound by the resolvase HtH domain. Light blue sequence is the sequence bound by the Zif268 DNA-binding domain (ZBS-Zif268 binding site). The central 16 bp coming from the Tn3 site I sequence is pink. The central dinucleotide is marked by the grey rectangle.

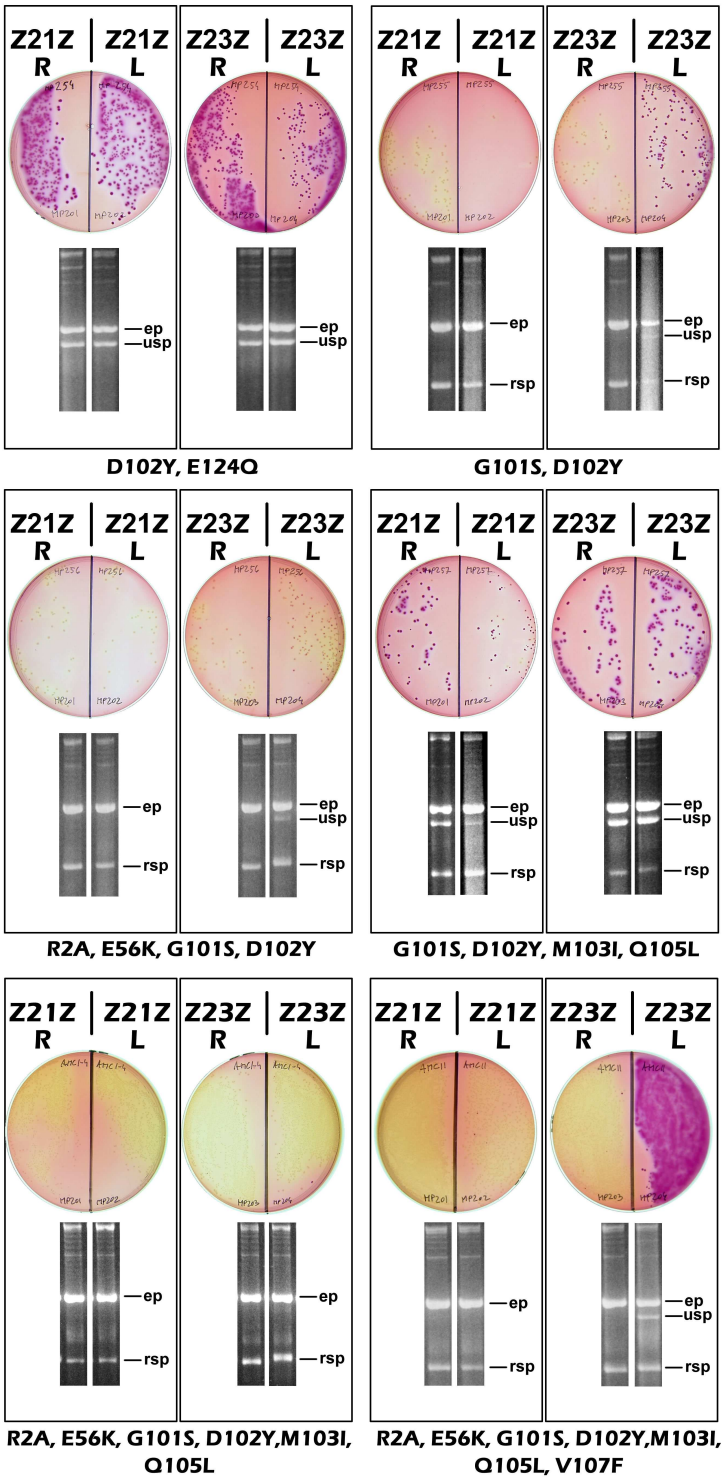


Figure 5.7- A MacConkey assay results with the Z-resolvases Z-R(YQ), Z-R(SY), Z-R(AKSY), Z-R(M), Z-R(NM) and Z-R(NMF) acting on substrates pMP201(Z21ZR × Z21ZR), pMP202(Z21ZL × Z21ZL), pMP203 (Z23ZR × Z23ZR) and pMP204 (Z23ZL × Z23ZL). The type of substrate used is indicated on top of the each of the plate photographs. The mutations in the Z-resolvase catalytic domain are marked at the bottom of each panel. The agarose gels show the uncut plasmid DNA isolated from the plate above. DNA bands are marked as follows, "ep" Z-resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

contacts less critical. The asymmetry between right and left Z-half-sites, observed with Z-resolvases with less hyperactive catalytic domains, could be a result of this inequality between the left and right half-site DNA-arm region contacts. Any affinity-reducing changes in the architecture of Z-sites are likely to have a greater effect on catalytic domains that are only mildly hyperactive than on more hyperactive catalytic domains such as NM.

Based on the results of the experiment in this and previous sections it was concluded that the optimum distance between ZBSs in a Z-site is indeed 22 bp. The Z22Z site is recombined at least partially by all Z-resolvases tested, even Z-R(YQ) which was made using the least hyperactive catalytic domain. The findings also show that there is a degree of flexibility in the Z-site spacing requirement, and that the ability of Z-resolvase to compensate for a suboptimal spacing is dependent on the level of activation of its catalytic domain.

5.5 Issues regarding mutation V107F, the linker and the N-terminus cut-off point

In the previous section it was observed (Fig.5.7) that the resolution of pMP204 (Z23ZL × Z23ZL) with Z-R(NMF) was less efficient than when catalysed by Z-R(NM), with the former producing red colonies while the latter gave white colonies. Even Z-R(AKSY), a Z-resolvase made from a catalytic domain much less hyperactive than NMF, resolved pMP204 more efficiently, as can be deduced from both the resulting mixture of white and red colonies and the resolved *vs.* unresolved substrate plasmid ratio as revealed by the isolated plasmid DNA analysis by agarose gel electrophoresis. This result brought the validity of the V107F mutation as a hyperactivity-enhancing mutation into question. In this section a contribution of this particular mutation, and other issues such as the catalytic domain cut-off point and the length and sequence of the linker between catalytic and Zif268 domains will be assessed.

Directly comparing Z-resolvases that differ in the features noted above, such as Akopian's AN15-Z resolvase, Z-R(NMF) and Z-R(NM), seemed an obvious course of action. AN15 Z-resolvase has a catalytic domain cut-off point at R144, while Z-R(NMF) and Z-R(NM) have a cut-off point at residue R148. In AN15 the catalytic domain is connected to the Zif268 DNA-binding domain by a 14 amino acid linker, while in the other two Z-resolvase the linker is only 2 amino acid residues long. AN15 and Z-R(NM) have the same mutations in their catalytic domain, while Z-R(NMF) has the additional mutation V107F. As can be

seen from the above, a straight three-way comparison between these proteins would be problematic as they vary from each other in more than one parameter.

To overcome this, an additional Z-resolvase Z-R(NMF,GSG) that has the same cut-off point at R148 as Z-R(NMF) and Z-R(NM), the same total protein length as AN15 (250 amino acids), and the V107F mutation in addition to the NM mutations in its catalytic domain, was designed. To make a plasmid expressing this Z-resolvase (pMP386), a short double-stranded DNA fragment with EagI/SpeI overhangs, created by annealing oligos GSGF' and GSGR'(Table 2.4), was cloned into EagI/SpeI-digested pAMC11 vector (Fig 2.7). This introduced the desired linker that is shown in comparison with linkers from Z-R(NM) and AN15 in Fig.5.8.Comparing Z-R(NMF,GSG) to Z-R(NMF) would expose the influence of the linker length, while comparing Z-R(NMF) to Z-R(NM) reveals the contribution of the V107F mutation. A third comparison that would assess the influence of the cut-off point would require construction of a Z-resolvase Z-R(NM,GSG), which would be identical to Z-R(NMF,GSG) but without the V107F mutation, which would then be compared directly with the AN15 Z-resolvase. Due to time constraints this third comparison was not attempted.

As both Z-R(NM) and Z-R(NMF) are very active *in vivo,* it was decided to compare them *in vitro* along with Z-R(NMF,GSG). Although *in vivo* assays can be very informative, the subtle differences between these Z-resolvases are more likely to be detected in an *in vitro* assay. To this end, the protein over-expression plasmids pMP393, pMP1 and pMP389, coding for Z-R(NM), Z-R(NMF) and Z-R(NMF,GSG) (Section 2.7.1) were constructed. Z-R(NM), Z-R(NMF) and Z-R(NMF,GSG) were over-expressed and purified according to the method described in Section 2.25.

To compare their recombination activity a time course resolution reaction using pIVS6, a substrate with two copies of the Z22Z site was employed (D. Wenlong, unpublished). The reaction was set up at 37°C and samples were taken at 0.5, 1, 2, 5, 15, 30 and 60 minute time points, after which they were stopped by heat. Half of the reaction mix volume was digested with XhoI enzyme, and the other half was left uncut. Uncut and cut samples were analysed using agarose gel electrophoresis (Fig.5.9).

Z-R(NMF) was found to be initially faster than Z-R(NM), as it started relaxing the supercoiled plasmid even after 0.5 minutes but it reached a peak in resolution at 15 minute time point, after which the relative amount of recombinant products appeared to decrease. At the 60 minute time point most of the recombinant product is gone, and seems to be

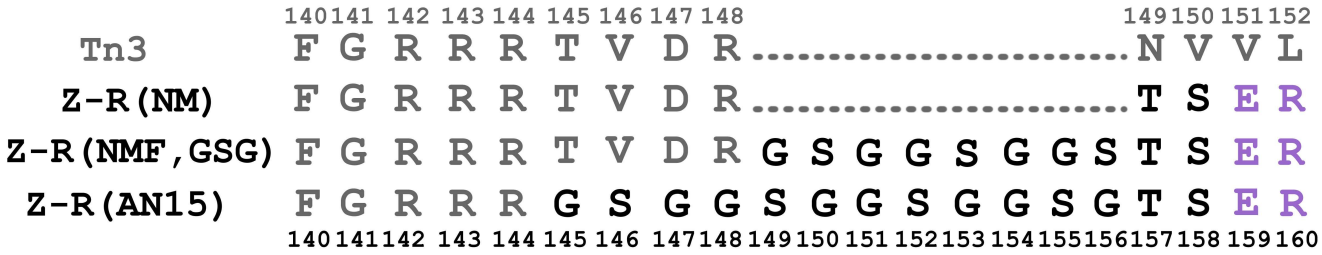


Figure 5.8- Protein sequence of the linker joining the catalytic domain of resolvase to the Zif268 DNA-binding domain. An alignment of the linker sequences of Z-R(NM), Z-R(NMF,GSG) and Z-resolvase AN15 along with the corresponding sequences from Tn3 resolvases. Tn3 resolvase residues and numbering is given in grey. A synthetic linker sequence including a sequence "TS" that introduces a SpeI site required for joining the catalytic domain of resolvase to the Zif268 DNA-binding domain is in black. The first two residues of Zif268 DNA-binding domain are given in purple. The numbering corresponding to the Z-resolvase AN15 is in black.

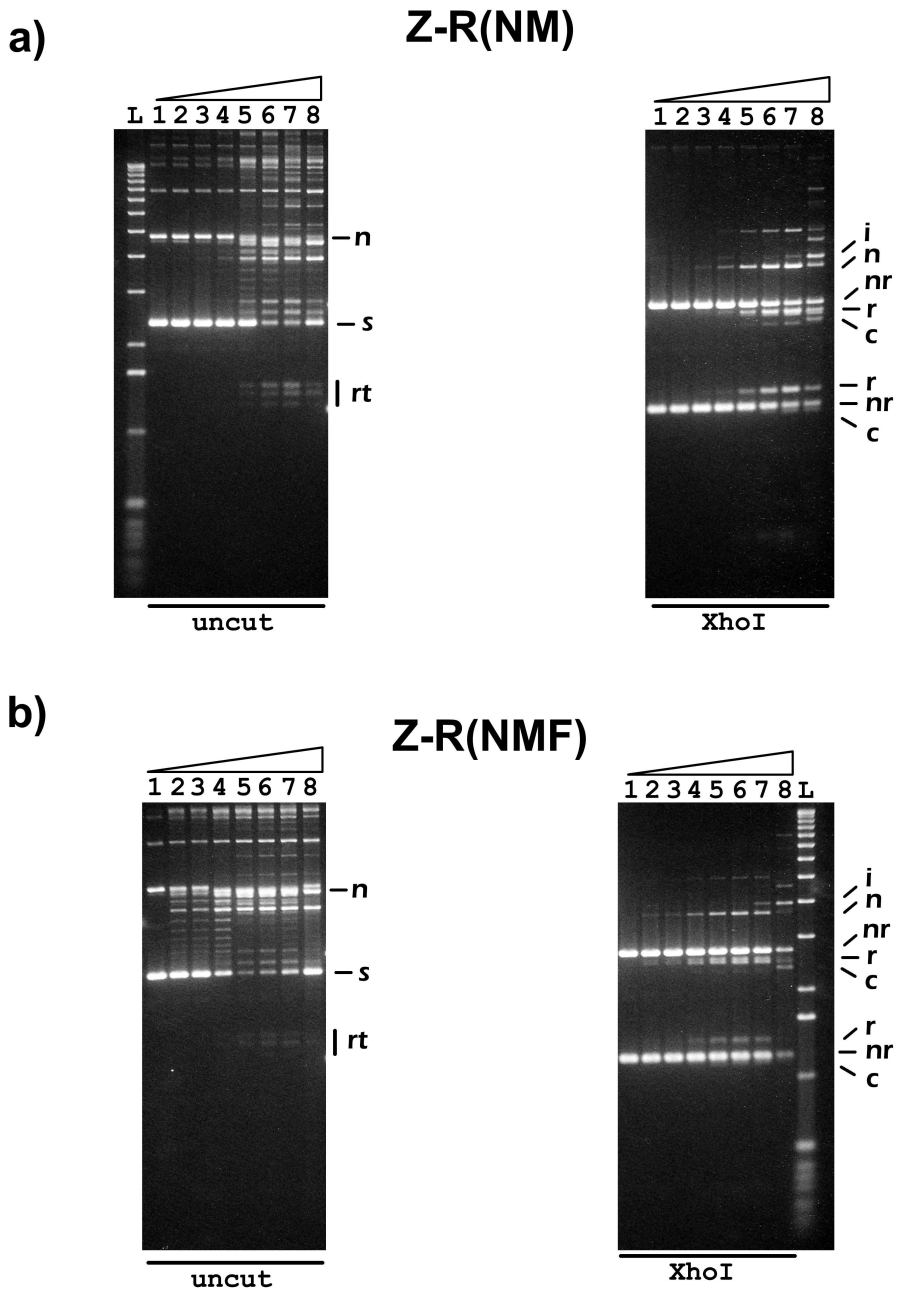


Figure 5.9- *In vitro* recombination reactions time course catalysed by the Z-resolvases **a)** Z-R(NM), **b)** Z-R(NMF) and **c)** Z-R(NMF,GSG) acting on substrate pIVS6 (Z22Z × Z22Z) The time points are marked at the top of the gel as follows, 1 (0 min.), 2 (0.5 min.), 3 (1 min.), 4 (2 min.), 5 (5 min.), 6 (15 min.), 7 (30 min.) and 8 (60 min.). For protein and DNA concentration and buffer composition see Section 2.27. The gels on the left show the uncut reactions, whereas the ones on the right show the corresponding reaction cut with XhoI. The graphic representation of XhoI digest of the possible reaction products giving their expected sizes is given in Fig. 2.14. The abbreviations are as follows, “s” (supercoiled substrate), “n” (nicked substrate), “rt” (recombinant product topoisomers), “nr” (non-recombinant products), “r” (recombinant product), “c” (cleavage product) and "i" (intermolecular product), see text for details (Section 5.5). *Continued on the next page...*

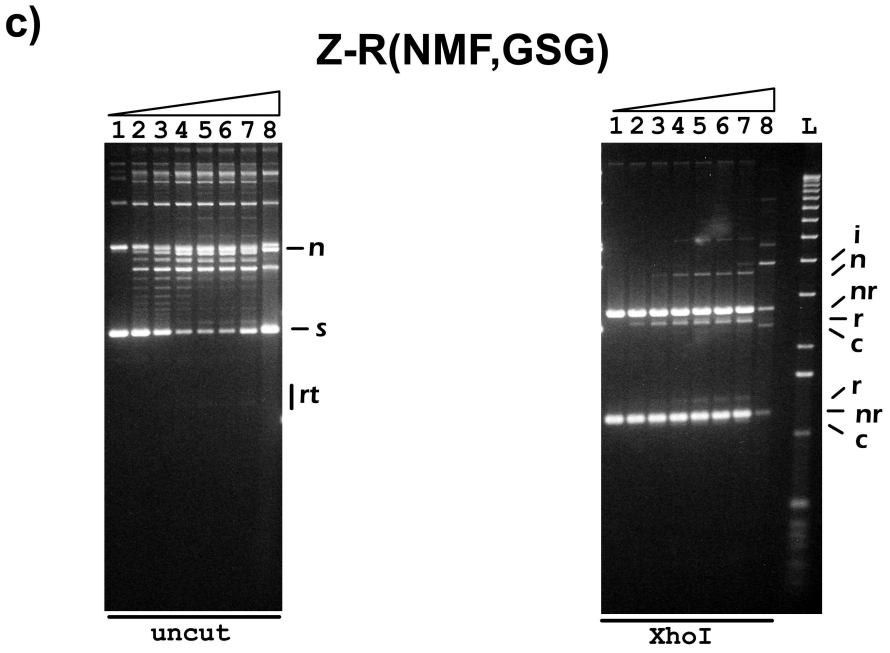


Figure 5.9-*Continued.. In vitro* recombination reactions time course catalysed by the Z-resolvase **c)** Z-R(NMF,GSG) acting on substrate pIVS6 (Z22Z × Z22Z) The time points are marked at the top of the gel as follows, 1 (0 min.), 2 (0.5 min.), 3 (1 min.), 4 (2 min.), 5 (5 min.), 6 (15 min.), 7 (30 min.) and 8 (60 min.). For protein and DNA concentration and buffer composition see Section 2.27. The gels on the left show the uncut reactions whereas the ones on the right show the corresponding reaction cut with XhoI. The graphic representation of XhoI digest of the possible reaction products giving their expected sizes is given in Fig. 2.14. The abbreviations are as follows, “s” (supercoiled substrate), “n” (nicked substrate), “rt” (recombinant product topoisomers), “nr” (non-recombinant products), “r” (recombinant product), “c” (cleavage product) and and "i" (intermolecular product), see text for details (Section 5.5).

replaced by a slow migrating band that could be a product of an intermolecular reaction and the cleavage product bands. Z-R(NM)-catalysed reaction on the other hand, is slower with the recombination products first appearing at 2 minute time point, also at the 60 minute time point the destruction of the resolution products is not as pronounced as with Z-R(NMF). The highest relative proportion of resolution product is observed at the 30 minute time point. The ratio of resolution products to the non-recombinant with Z-R(NM) was in general higher than that observed in reactions catalysed by Z-R(NMF) despite the initial difference in the speed of the reaction. This indicates that Z-R(NMF) continues performing secondary recombination reactions using the resolution products as a substrate but that religation step in these reactions is affected, as suggested by the accumulation of cleavage products. Z-R(NMF,GSG) recombined pIVS6 just as fast as Z-R(NMF) but the ratio of the resolution product to the non-recombinant was lower than that observed in reactions catalysed by either Z-R(NMF) or Z-R(NM). Z-R(NMF,GSG) did not accumulate cleavage products for the first 30 minutes of the reaction but at the 60 minute time point almost all of the resolution product was converted into either intermolecular product or cleavage products.

The results above indicate that although V107F increases the processivity of the Z-resolvase it does so at the expense of the reaction precision. The bulky phenylalanine residue could help destabilise Z-resolvase dimer aiding the processivity but also it could interfere with the recombination reaction at a later stage, accounting for the accumulation of cleavage products. The longer linker as in Z-R(NMF,GSG) brings the overall efficiency of the reaction even lower. In conclusion, both V107F and the longer linker are undesirable and should be avoided in construction of future Z-resolvases.

5.6 Complementation assays

Z-resolvases had so far been tested on a number of Z-sites, with different lengths of central sequence flanked by the Zif268-binding sites. Provided that the distance between the two ZBSs is between 20 and 22 bp, two of these sites should be able to form a synapse prior to recombination, containing four Z-resolvases arranged in a tetramer, analogous to the arrangement seen with hyperactive resolvase mutants (Li *et al.,* 2005; Kamtekar *et al.,* 2006). Would different Z-resolvases with the same catalytic domain but different DNA-binding domains be able to combine on appropriate sites and form heterotetramers capable of catalysing recombination? Or would it be possible to combine Z-resolvase subunits with resolvase subunits, and form heterodimers, or various possible heterotetramers? Both of these possibilities have important implications for the future use of Z-resolvases. Heterotetramers consisting of Z-resolvases with different DNA-binding domains are a

prerequisite for vast majority of possible future applications in biotechnology or molecular medicine, whereas the possibility of making heterodimers or various heterotetramers containing both Z-resolvase and resolvase would enable further detailed studies of recombination mechanism.

The possibility of heterotetramer formation containing Z-resolvase and resolvase subunits was tested *in vitro* in the previous studies of our group (Akopian, 2003). Complementation between Z-resolvases AN8 or AN15 and NM resolvase in an *in vitro* recombination assay was attempted. The results were inconclusive, as NM resolvase was found to be able to catalyse resolution on all substrates that were tested.

In this section, the possibility of formation of a heterotetramer *in vivo*, consisting of two heterodimers each containing both resolvase and Z-resolvase, is explored. In order to induce heterotetramer formation, two new hybrid recombination sites, namely ZL and ZR, were designed (Fig. 5.10). The ZL site is composed of the right half-site of the Z22Z Z-site and the left half-site of Tn3 site I. The arrangement in ZR site is exactly the opposite, the site consisting of the left half-site of Z22Z and the right half-site of Tn3 site I. Either Z-resolvase or NM resolvase should be capable of binding to these sites as a monomer, but should be unable to form dimers, or bring the sites together to form a synapse, as only one half-site in each case should be suitable for binding. On the other hand a combination of NM resolvase and a Z-resolvase, provided that the catalytic domains are positioned in a compatible fashion, should be able to form a heterodimer. Once these heterodimers are formed they should be able to come together to form a synapse and proceed with recombination.

5.6.1 Positive complementation

Complementation experiments were performed *in vivo* using the MacConkey assay (Fig.5.11). To be able to do this, hybrid sites ZL and ZR were cloned into pMS183Δ vector creating pMP96 (ZR × ZR) and pMP97 (ZL × ZL), *in vivo* recombination substrate plasmids (Section 2.7.2). To test for resolution, pMP96 and pMP97 were used to transform DS941 cells already containing either pMON1, expressing Z-R(NMF), or pMP17, expressing NMF resolvase, or DS941 containing both pMON1 and pMP17 together. pMON1 (M. Sentmanat, unpublished) was used to express Z-R(NMF) instead of pAMC11 in order to avoid plasmid incompatibility with pMP17 used to express NMF resolvase. The origin of replication of pMON1 comes from p15A via pACYC184, while plasmids like pMP17 and pAMC11 have a ColE1 origin of replication.

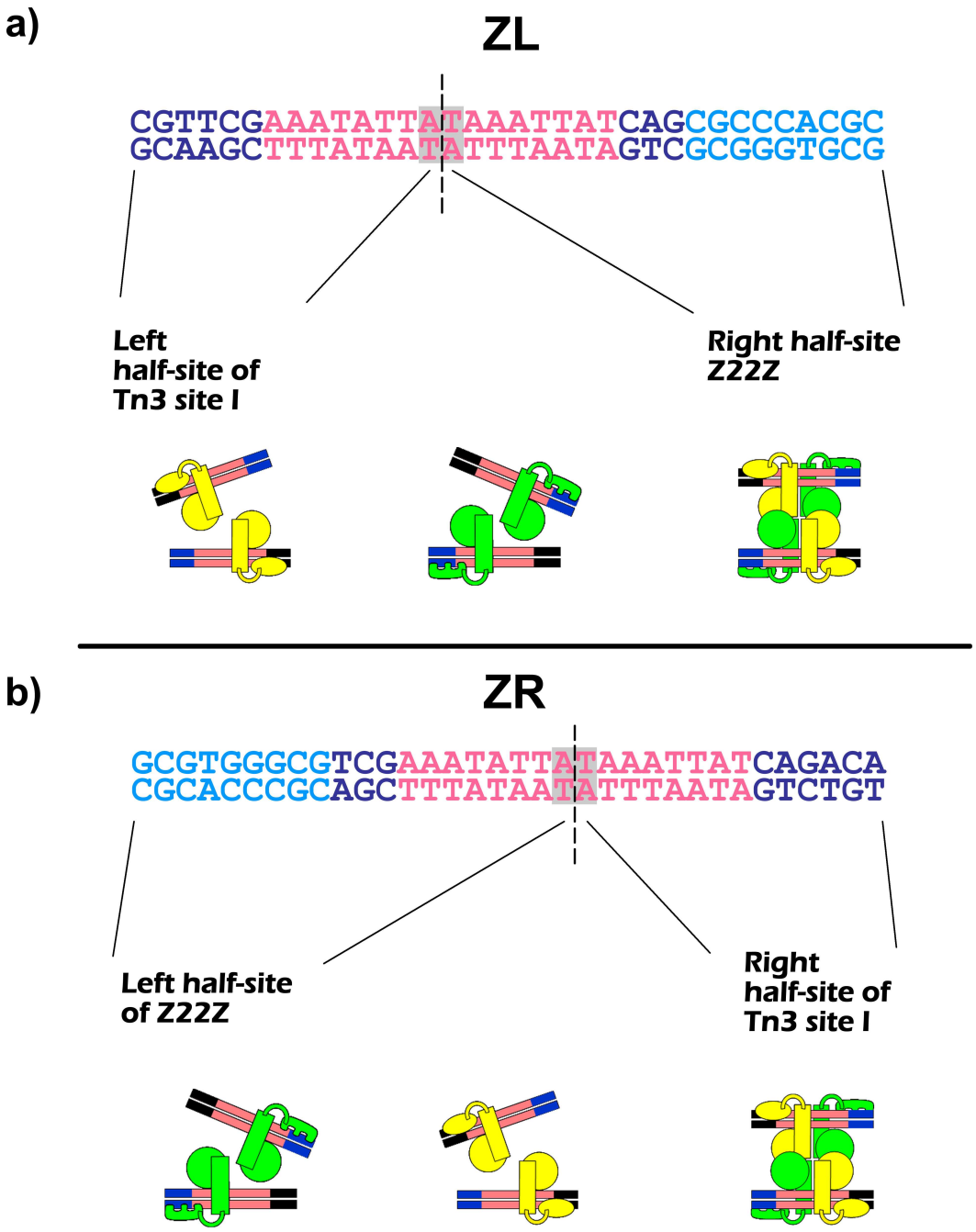


Figure 5.10- Sequences of hybrid sites a) ZL and b) ZR. Sequence given in dark blue is the sequence usually bound by the resolvase HtH domain. Light blue sequence is the sequence bound by the Zif268 DNA-binding domain (ZBS-Zif268 binding site). The central 16 bp coming from the Tn3 site I sequence is pink. The central dinucleotide is marked by the grey rectangle. Underneath the sequences are cartoon representations of expected result in a complementation experiment. Only when both Z-resolvase and resolvase are present is the site synapsis expected (bottom right in each panel). When proteins are expressed on their own only binding by monomers is expected. Z-resolvase monomers are drawn in green and resolvase monomers are drawn in yellow.

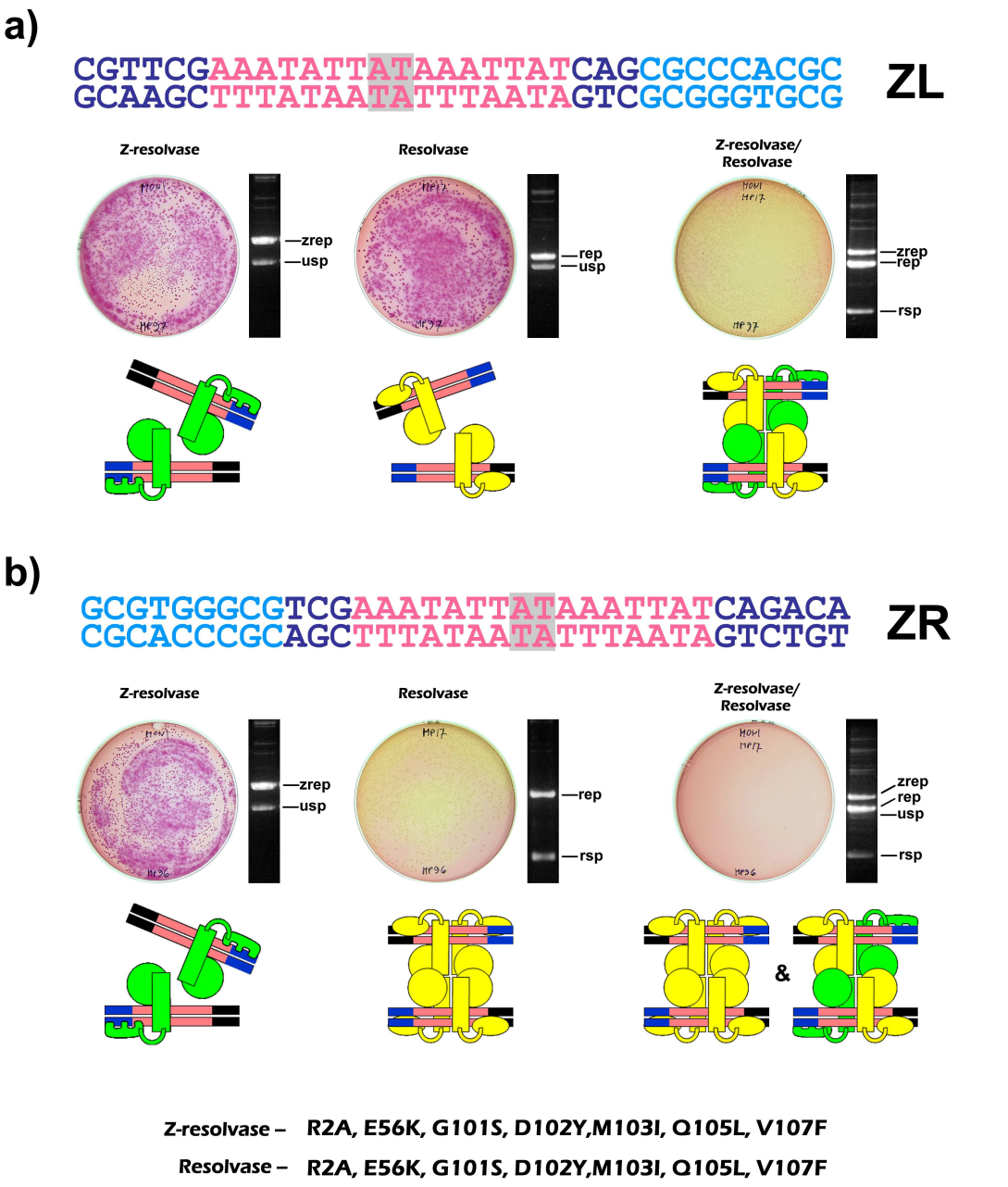


Figure 5.11-Results of the in vivo complementation experiment using activated NMF resolvase and Z-R(NMF) Z-resolvase acting on substrates ZL **a)** and ZR **b)** The sequence of the site is given at the top of each panel. A note indicating whether resolvase or Z-resolvase or both are expressed in the cells on a plate is given above the each plate photograph. The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, “zrep” Z-resolvase expression plasmid, “rep” resolvase expression plasmid, “usp” unresolved substrate plasmid “rsp” resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details). Below the plate photographs are the cartoons representing the interpretation of the observed results, see text for details (Section 5.6). Z-resolvase monomers are drawn in green and the resolvase monomers are drawn in yellow.

As can be seen in Fig. 5.11, Z-R(NMF) expressed from pMON1 produced red colonies when acting on either pMP96 (ZR × ZR) or pMP97 (ZL × ZL). Analysing isolated plasmid DNA by agarose gel electrophoresis confirmed that Z-R(NMF) is unable to resolve either of these substrate plasmids when present on its own. On the other hand, hyperactive resolvase NMF expressed from pMP17 produced white colonies with pMP96, indicating that NMF resolvase is capable resolving pMP96 even when expressed on its own. Conversely, NMF resolvase acting on pMP97 resulted in red colonies. No trace of resolution with this substrate was observed even after the plasmid DNA was isolated and separated on an agarose gel. When Z-R(NMF) and NMF were present in the cells together, Z-resolvase and resolvase were able to complement each other on pMP97 producing white colonies. On pMP96 substrate, expression of both Z-R(NMF) and NMF resolvase resulted in drastically reduced colony numbers. However, all the colonies that were observed with this substrate-protein combination were white.

The resolution of substrate plasmid pMP96 containing two copies of ZR by NMF resolvase is in concordance with the previously reported *in vitro* results (Akopian, 2003). ZR contains the right half-site of Tn3 site I, which is known to provide a better binding site for the resolvase DNA-binding domain than the left half-site (Bednarz, 1990). When a monomer of NMF resolvase binds to the right half-site of Tn3 site I within the ZR site, it is presumably able to recruit a second NMF subunit through cooperativity. Two dimers of NMF resolvase that were formed on two copies of ZR go on to form a synapse and proceed with recombination. This does not happen with pMP97 substrate as it carries two copies of ZL, a site containing the left half-site of Tn3 site I. Binding of the NMF resolvase DNA-binding domain to the left half-site of Tn3 site I must be insufficiently stable to recruit a second NMF resolvase subunit through cooperativity. Hence, the NMF dimer fails to form on the ZL site, no synapsis ensues and no recombination is observed.

Since NMF resolvase is capable of catalysing recombination of pMP96 it is difficult to say whether true complementation is taking place. However, the reduction in colony numbers observed when NMF resolvase is co-expressed with Z-R(NMF) in cells harbouring pMP96 suggests that both recombinases seem to interact with the substrate.

5.6.2 Negative complementation

As it was difficult to say whether NMF and Z-R(NMF) complement each other when acting on pMP96, it was decided to try and test for negative complementation (Fig.5.12).

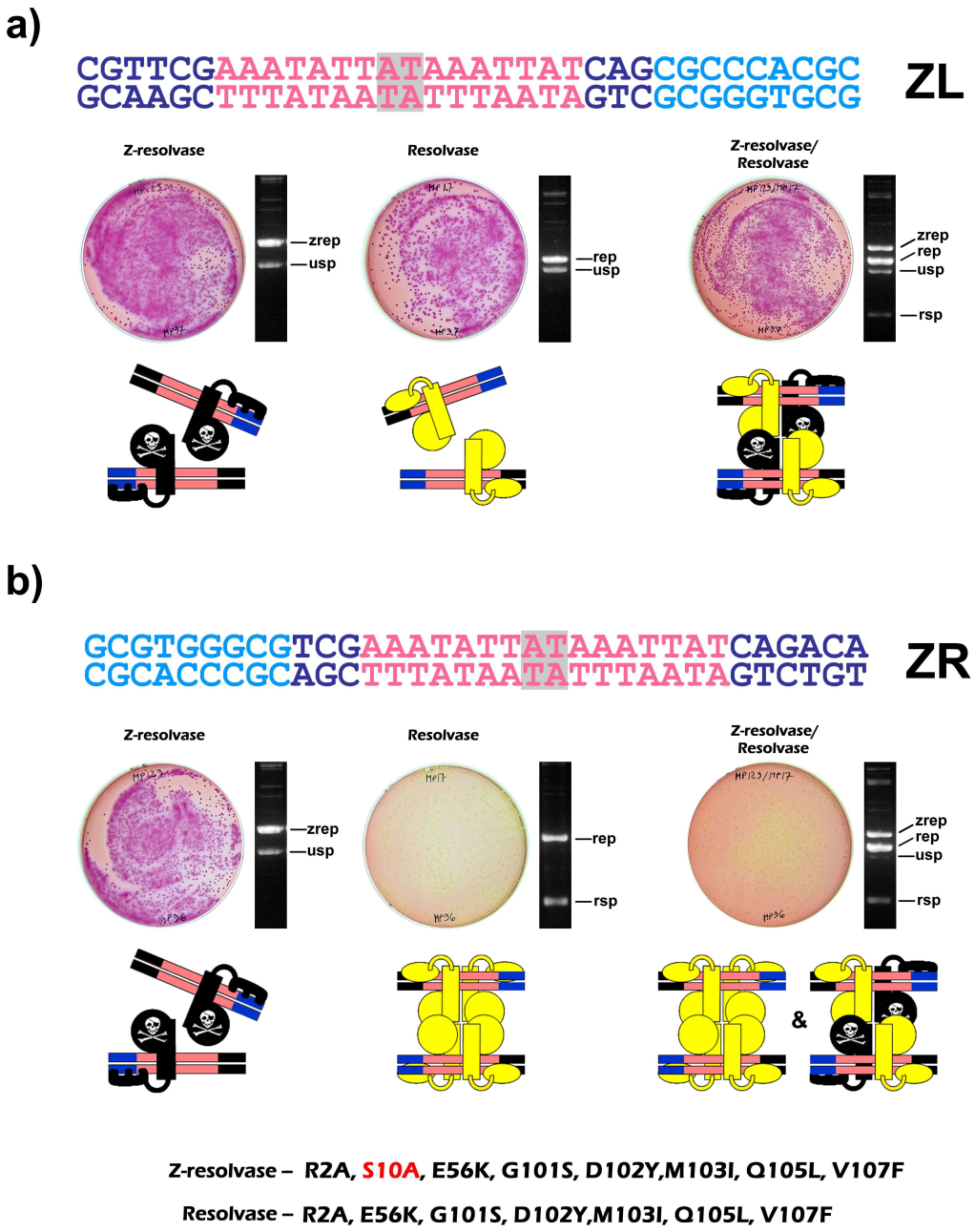


Figure 5.12-Results of the in vivo complementation experiment using activated NMF resolvase and Z-R(NMF,S10A) Z-resolvase lacking the catalytic serine residue acting on substrates ZL **a)** and ZR **b)** The sequence of the site is given at the top of each panel. A note indicating whether resolvase or Z-resolvase or both are expressed in the cells on a plate is given above the each plate photograph. The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, "zrep" Z-resolvase expression plasmid, "rep" resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details). Below the plate photographs are the cartoons representing the interpretation of the observed results, see text for details (Section 5.6). Z-resolvase monomers are drawn in black and the resolvase monomers are drawn in yellow.

Instead of using a catalytically active Z-R(NMF) as in the previous experiment, an inactive version of Z-R(NMF), namely Z-R(NMF, S10A) expressed from the plasmid pMP129 was employed. This inactive Z-resolvase was expected to bind the ZBS-containing left half of the ZR and prevent NMF resolvase from making homodimers on this site.

As expected, Z-R(NMF,S10A) produced red colonies when acting on either pMP96 or pMP97, and no trace of resolution product could be detected by analysing the isolated plasmid DNA by agarose gel electrophoresis. Co-expressing the catalytically inactive Z-R(NMF, S10A) along with NMF resolvase did not abrogate its activity. However, the decrease in the number of colonies was not nearly as dramatic as the one observed when both Z-R(NMF) and NMF resolvase were acting on pMP96 as described in the previous section. This meant that NMF resolvase still managed to out-compete Z-R(NMF, S10A) at binding to ZR, but that some of NMF resolvase was sequestered by Z-R(NMF,S10A) to form a heterodimer, which lessened the impact of the NMF resolvase as indicated by the increased number of colonies.

Unsurprisingly, when pMP97 was transformed into DS941 cells in which NMF resolvase and inactive Z-R(NMF,S10A) were co-expressed, red colonies were produced. However, when the plasmid DNA was isolated and separated by agarose gel electrophoresis, a small amount of resolved substrate plasmid was observed, indicating that even catalytically inactive Z-resolvase can complement hyperactive resolvase, at least to some extent.

A possible explanation would be that as Z-R(NMF, S10A) is able to bind the ZBS on the right half of ZL it can recruit a subunit of NMF to the left hal-site to form a heterodimer. Two of these heterodimers are then able to synapse together and form a heterotetramer containing two catalytically active NMF resolvase subunits and two catalytically inactive Z-R(NMF, S10A) subunits. In the heterotetramer, NMF resolvase is able to cleave two out of four DNA strands, whereas Z-R(NMF, S10A) is unable to cleave DNA as it lacks the catalytic serine residue. After this cleavage of the two strands, the whole complex might be expected to dissociate as the strand exchange is unable to proceed by the normal mechanism. The substrate plasmid containing such two nicks could then be recombined by the host cell using some resident DNA repair or recombination enzyme.

5.7 Discussion

In the preceding sections, different aspects of Z-resolvase recombination system have been investigated. Various Z-resolvase and Z-site designs were tested *in vivo* and *in vitro* and the findings were compared with the previous work of our group (Akopian *et al.,* 2003;

Akopian, 2003; A. Maclean, unpublished). In addition, complementation of a hyperactive resolvase mutant with both catalytically active and catalytically inactive Z-resolvase on specially designed hybrid sites was observed. In this section, these results will be summarised, compared with other relevant published findings (Gordley *et al.,* 2007; Gordley *et al.,* 2009), and their significance for the future work on the Z-resolvase system will be discussed.

An intriguing property of the Z-resolvase system is the apparent lack of correlation between the length of the linker connecting the resolvase-derived catalytic domain to the zinc finger-based DNA binding domain, and the length of Z-site that the protein prefers to work on. Z-resolvases with longer inter-domain linkers are not able to catalyse recombination on Z-sites with a longer sequence between zinc finger binding sites. This is in stark contrast to what is observed with the partially analogous zinc finger nuclease (ZFN) system, where a longer interdomain linker allows for cleavage of sites with longer spacer sequence between the zinc finger binding sites (Shimizu *et al.,* 2009; Händel *et al.,* 2009).

While Barbas and co-workers dismissed the difference in the optimal distance between the zinc finger binding sites observed in their work (Gordley *et al.,* 2007) and that of our group (Akopian *et al.,* 2003) as due to the difference in the inter-domain linker length used in their Z-resolvase ($Tn3Ch15_{\times 2}$) and AN15 Z-resolvase, the work in this chapter clearly shows that that is not the case.

The optimum distance between zinc finger binding sites according to Barbas and co-workers was found to be 20 bp, whereas results from our studies suggest that this distance is 22 bp. Z-resolvase $Tn3Ch15_{\times 2}$ consists of the catalytic domain of a hyperactive Tn3 resolvase mutant (G70S, D102Y, E124Q), residues M1-T145, linked by a five amino acid residue linker SGSTS to a didactyl engineered zinc finger domain. The catalytic domain of AN15 Z-resolvase comes from NM resolvase (residues M1-R144) and is connected by a 14 amino acid residue linker GSGGSGGSGGSGTS to the tridactyl Zif268 zinc finger DNA binding domain. Comparing the length of the linker region literally as done by Barbas and co-workers ignores the issue of the catalytic domain cut-off point. A much more accurate measure in relation to the Z-site inter-ZBS length would be the overall length of the protein prior to the zinc finger domain, according to which AN15 Z-resolvase is 10 amino acid residues longer. Although having an extra 10 amino acid residues is more than enough to span one extra DNA base in a Z-half-site, we believe that this is not the reason for the preference for the longer Z-site.

Z-resolvases such as Z-R(YQ), Z-R(SY), Z-R(AKSY) and all Z-R(M) tested in Sections 4.3-4.4 are of the same length prior to the zinc finger domain as Tn3Ch15$_{\times 2}$, and yet all of them show preference for 22 bp of distance between the zinc finger binding sites within a Z-site. However, it must be noted that Z-resolvase Z-R(NM), containing the highly hyperactive NM resolvase catalytic domain, recombined Z20Z and Z22Z with equal efficiency, pointing to a degree of flexibility in the acceptable distance between the zinc finger binding sites.

This flexibility does not come from the NM resolvase catalytic domain itself as it can be seen from an additional experiment results of which are shown in Fig.5.13. In this experiment, a series of *in vivo* substrate plasmids (pMP205-pMP208) containing RR Tn3 site I variants (R16R, R18R, R20R, R22R) was constructed, and tested for resolution with NM resolvase expressed from pFO2 (Olorunniji, 2006). At the outermost 6 bp of both half-sites, the RR Tn3 site I variants had the sequence that is naturally found in the right half-site of the Tn3 site I, which is known to be very tightly bound by the resolvase HtH domain (Bednarz, 1990). As can be seen in Fig. 5.13 RR sites differed in length by 2 bp, which was achieved by introduction of random bases between the HtH 6 bp binding site and the central 16 bp. Adduction of HtH 6 bp binding sites was analogous to the increasing the inter-ZBS distance in successive Z-sites. NM resolvase only produced white colonies on the substrate containing two copies of R16R (pMP205). When acting on all other substrates tested containing longer RR sites (pMP206-208), NM resolvase produced red colonies. Even upon agarose gel electrophoresis analysis of the isolated plasmid DNA, no trace of resolved substrate plasmid was revealed with pMP206-208. The result of this additional experiment argues that the flexibility as to the length of the recombination site is not a property inherent in the NM resolvase catalytic domain but is a feature enabled when the NM catalytic domain is put into the Z-resolvase context.

The requirement for a distance between zinc finger binding sites in the range between 20 and 22 bp observed with Z-resolvases could be explained by steric hindrance and the requirement of the zinc finger domain to follow the DNA phasing. If the zinc finger-binding sites are too close to the centre of the Z-site, the binding of two subunits of Z-resolvase to such a site could presumably be interfered with by the clash of resolvase catalytic domains attempting to make a dimer. If a dimer could be formed, as the tridactyl zinc finger domain is much larger than the resolvase native HtH DNA-binding domain as can be seen in Fig.5.14, possible clashes between zinc-finger domains when attempting to synapse two Z-resolvase dimers are also likely.

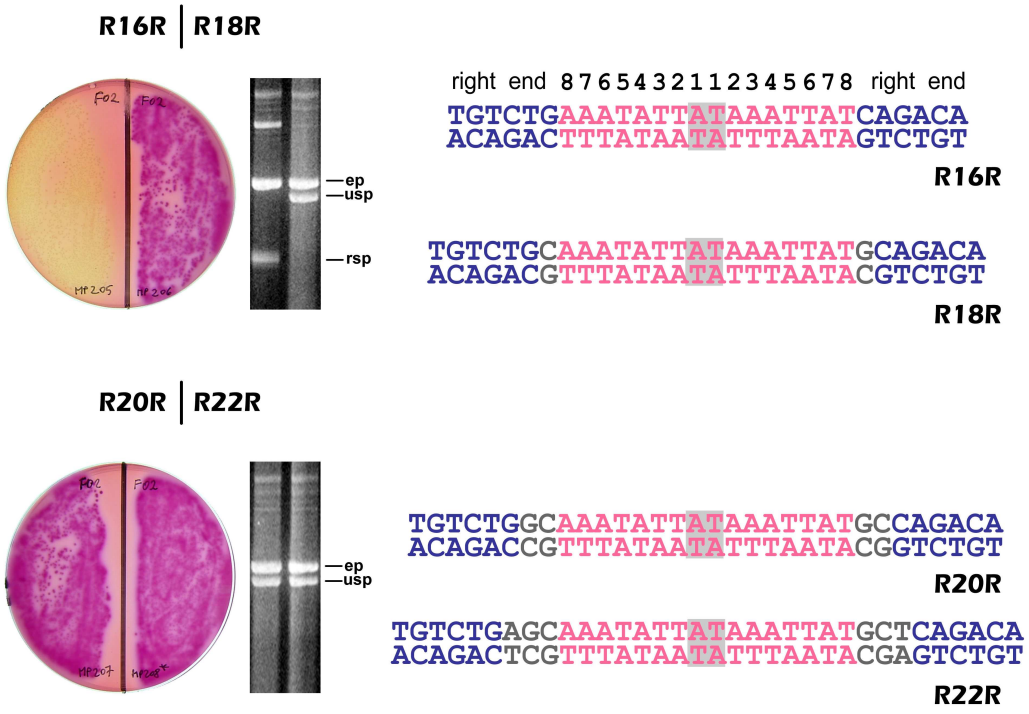


Figure 5.13- A MacConkey assay results with the activated Tn3 resolvase mutant NM acting on substrates pMP205(R16R × R16R), pMP206(R18R × R18R), pMP207 (R20R × R20R) and pMP208 (R22R × R22R). Sequence of sites R16R, R18R, R20R and R22R is given on the right. The dark blue sequence in all these sites is the sequence that Tn3 resolvase HtH domain binds to in the right half-site of the Tn3 site I. The central 16 bp of each of the sites is in pink. The bases that have been inserted in order to extend the site i.e increase the distance of the HtH binding sequence from the centre are given in grey. The central dinucleotide is marked by a grey rectangle. The individual DNA bases are numbered from the centre. On the left are the photographs of plates with the type of substrate used indicated at the top of the each of the plate photographs. The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, “ep” resolvase expression plasmid, “usp” unresolved substrate plasmid “rsp” resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

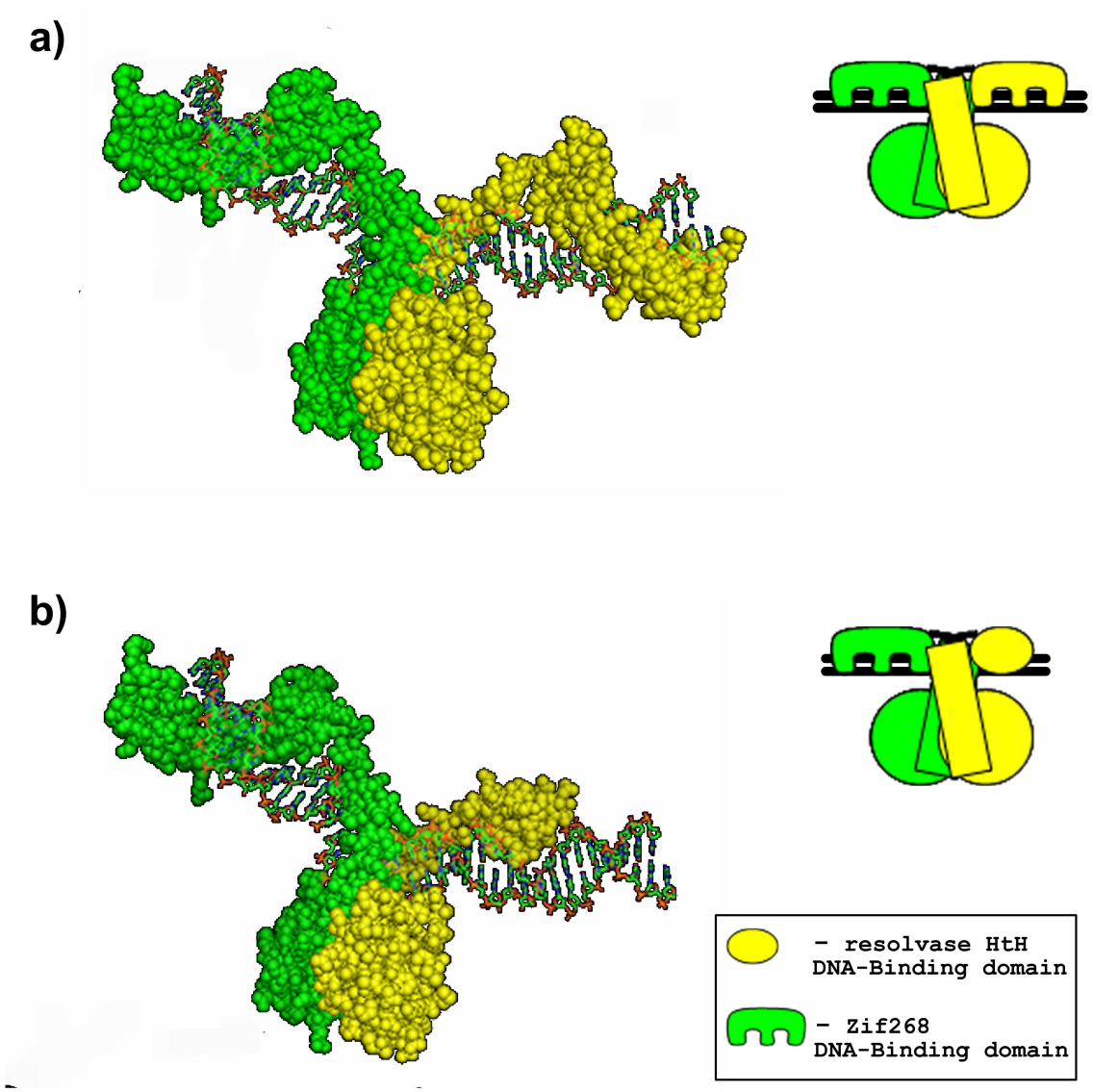


Figure 5.14- A model of **a)** Z-resolvase dimer bound to DNA **b)** a heterodimer consisting of a Z-resolvase (green) and resolvase (yellow) monomers bound to DNA. Z-resolvase and resolvase subunits are shown in spacefill whereas DNA is shown as sticks. Note the difference in size between the resolvase HtH DNA-binding domain and Zif268 DNA binding domain. Models were made by overlaying crystal structures of γδ resolvase dimer (1gdt, Yang & Steitz, 1995) and Zif268 DNA-binding domain (1aay, Elorod-Erickson et al., 1996).

If on the other hand the zinc finger-binding sites are too far apart, correct DNA phasing is likely to play more of a role. As each extra base in a classic B-DNA helix causes around ~36°turn in the DNA, the N-terminal end of a zinc finger DNA binding domain that is meant to connect to the resolvase catalytic domain can end up in a variety of unfavourable orientations. A longer flexible linker between zinc finger domain and resolvase catalytic domain does not seem to successfully compensate for this.

A possible reason for this is to do with the mechanism of zinc finger binding, which relies on coupled folding and binding. Coupled folding and binding is the process in which an intrinsically disordered region of a protein folds into an ordered structure concomitant with binding to its target. Zinc fingers in a zinc finger domain, behave as beads on a flexible string where the function of the linker regions between the fingers is to enable a relatively unhindered spatial search by the attached domains (Zhou, 2003). When a zinc finger domain binds to the major groove it does so in a sequence, with its C-terminal finger binding first. This causes the linkers between the fingers to fold, cap and thereby stabilise the preceding helix in the protein, orientating the next finger correctly for binding in the major groove (Laity *et al.,* 2000a). Extending the length and increasing the flexibility of the linkers between the fingers can inhibit binding as exemplified by the alternative splicing variant of WT1 a zinc finger domain containing Wilms' tumour protein (Laity *et al.,* 2000b).

In a Z-resolvase, the linker between the zinc finger domain and the resolvase catalytic domain could be behaving in the same fashion as the linker regions between the individual zinc fingers. As entropic cost to fold a disordered region of a protein is paid by the binding enthalpy, increasing the length of the linker region between zinc finger domain and resolvase catalytic domain would increase its flexibility and hence the entropic cost of folding this segment, which could have an inhibitory effect on binding, explaining the observed preference for shorter linkers.

Additionally in the resolvase, folding of the whole HtH DNA-binding domain (residues 148-183 in γδ resolvase) and the final portion of the arm region (residues 140-147 in γδ resolvase) is also induced by the binding to DNA (Yang & Steitz, 1995; Rice and Steitz, 1994a, 1994b). The amino acid residues 140-147 increase the affinity of resolvase for its site I through numerous contacts with the DNA in the minor groove, as discussed in Chapter 3. It is possible that binding of a zinc finger domain to DNA is not optimal for inducing the fold of the final portion of the arm region reducing the affinity of Z-resolvase for its site. This could be an explanation for the unexpectedly weak binding of Z-resolvases

in vitro (G. Schlötel, unpublished) and the reason behind the mild inhibitory effect on the resolvase catalytic domain observed in the Z-context (Section 5.3).

Resolvase/Z-resolvase complementation experiments (Section 5.6) indicate the necessity of synapse formation as a prerequisite for recombination and cleavage reactions. The ability of the Z-resolvase to be complemented by the resolvase with the same catalytic domain, on specially designed hybrid sites, is encouraging with regards to the possibility of formation of heterotetramers in which each Z-resolvase subunit has a different DNA-binding zinc finger domain, which could have a variety of uses. As long as the binding affinity of the zinc finger domains used is comparable we are confident that the recombination catalysed by such heterotetramer is possible. Unambiguously confirming the complementation *in vitro* remains a task for the future.

6 Chapter 6: Sequence specificity of Tn21 and Sin resolvase catalytic domains

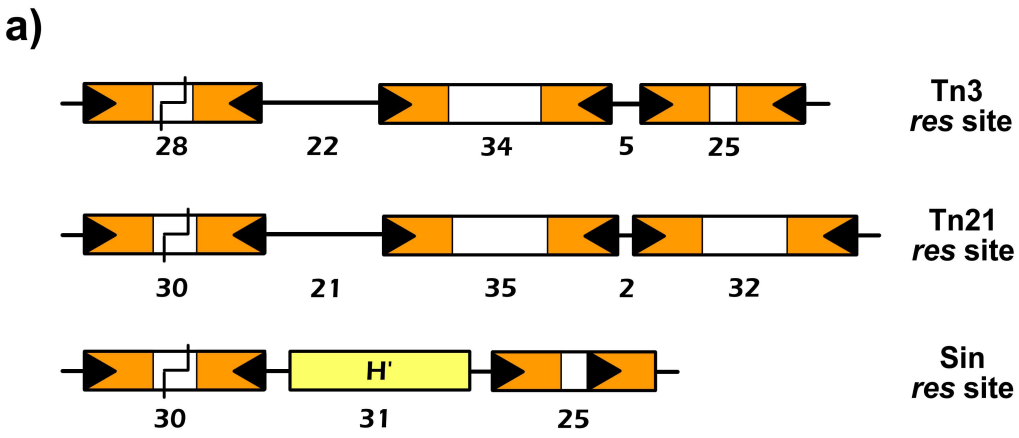
6.1 Introduction

In Chapter 4, catalytic domains of hyperactive mutants of Tn21 resolvase and Sin resolvase were used to make Z-R(Sin) and several Z-R(Tn21) variants. The resulting Z-resolvases showed a degree of cross-activity, with Z-R(Sin) recombining the Tn21 Z-site and *vice versa* (Fig. 4.7). However, neither of these recombinases showed any activity on Z-sites with the central sequences derived from Tn3 site I (Figs. 4.8 and 4.9).

Tn21 and Sin resolvases come from different resolvase families; in fact they are almost as distantly related from each other as they are from Tn3 resolvase (Fig 4.2). Nevertheless, the results of the Z-R(Sin) and Z-R(Tn21) tests suggest that the catalytic domains of Tn21 and Sin resolvase proteins, although distantly related, have similar sequence specificity, which is different from that of Tn3 resolvase.

Wild type Tn21 and Sin resolvase can not utilise each other's *res* sites, which is easily explained by the large differences in the *res* site architecture of these two recombinases (Fig 6.1). The failure of hyperactive mutants of Sin and Tn21 resolvase that do not require accessory sites to catalyse recombination on each other's sites (Fig 4.8) is probably due to the different sequence requirements for their DNA-binding domains. Most differences between the Sin and Tn21 site I sequences are found in the 6 outermost bases at the end of each half-site, the sequence known to be recognised by the resolvase HtH domain (Fig 6.1).

That the HtH domain of Tn21 resolvase does not bind strongly to the Sin site I sequence was confirmed using the *in vivo* binding assay (Section 2.24) in Fig 6.2. In this assay, binding to Sin site I in the plasmid pSB404 (S. Rowland, unpublished) occludes the *gal*K gene promoter that overlaps this site, and this in turn abolishes *gal*K transcription. As the assay is performed in DS941 cells, which are naturally *gal*K$^-$, the lack of *gal*K expression (signifying binding to Sin site I) produces white colonies. Conversely, binding which is not strong enough to prevent access to the *gal*K gene promoter results in the production of red colonies. A negative control for binding using Z-R(Sin), which lacks a HtH DNA-binding domain, is also included in Fig.6.2



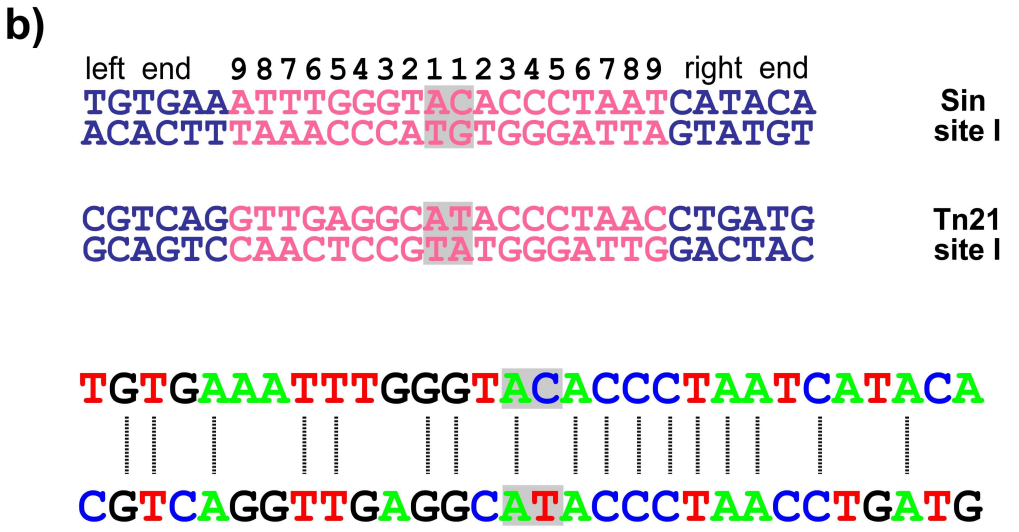


Figure 6.1-a) Architecture of Tn3, Tn21 and Sin *res* sites. The lengths of the individual binding sites and intervening DNA segments are given below the cartoons. Black triangles mark the direction of the repeated sequence that forms the resolvase binding sites. All sites are inverted repeats, expcept Sin *res* site II which is a direct repeat. Yellow box with an "H'" on it found between Sin *res* sites I and II represents the Hbsu binding site. The cleavage point within *res* site I is marked by black lines. **b)** The sequence of Sin *res* site I and Tn21 *res* site I (top of the panel). The dark blue sequence in these sites is the sequence that the resolvase HtH domain binds. The central 18 bp of the sites is in pink. The central dinucleotide is marked by a grey rectangle. The individual DNA bases are numbered from the centre. The "Left" and "Right" end refer to the conventional representation of site I within *res*. Bottom of the panel-An alignment of Sin *res* site I and Tn21 *res* site I top strand with the dashed black lines connecting DNA bases that are identical.

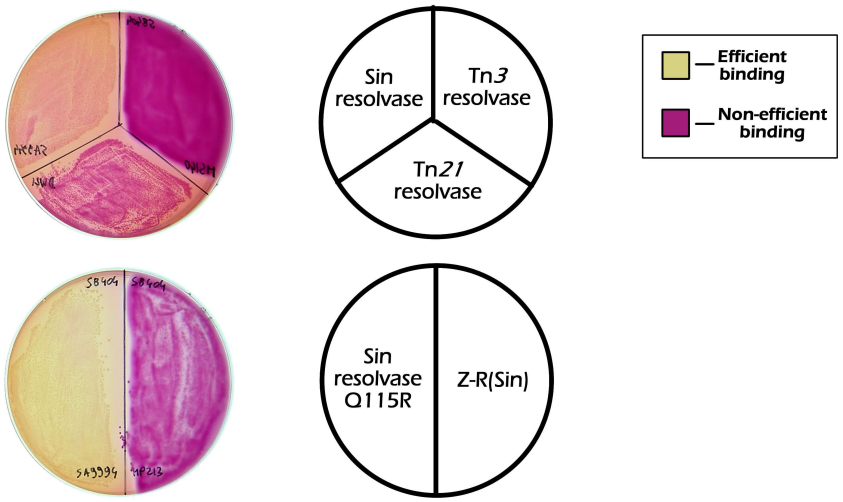


Figure 6.2- Results of the *in vivo* binding assay using the plasmid pSB404 with the wild-type Tn3, Tn21, and Sin resolvases (top plate photograph), and an activated Sin resolvase mutant Q115R and the Sin resolvase based-Z-resolvase Z-R(Sin) (bottom plate photograph). Z-R(Sin) is a negative control as it doesn't have a resolvase HtH DNA-binding domain. How the different resolvases or Z-resolvases are arranged on plates is indicated in the explanatory keys (to the right of the plate photographs). The white colonies signify strong binding to the Sin *res* site I while red colonies signify weak binding (see Section 2.24 for details).

The first part of this chapter presents experiments performed to establish whether Sin and Tn21 resolvase catalytic domains do indeed have similar sequence requirements in the central 18 bp of site I, and what exactly these sequence requirements are. The second part of the chapter deals with attempts to alter the specificity of the Sin resolvase catalytic domain into that of the Tn3 resolvase catalytic domain and *vice versa*, in order to better understand the factors determining the specificity of the resolvase catalytic domain.

6.2 Site I sequence requirements for Sin and Tn21 resolvase

To establish the sequence specificity of the Sin and Tn21 resolvase catalytic domains in the central 18 bp of their crossover sites, two separate, but complementary approaches were adopted. The first approach involved the use of random site library selections employing a multimer resolution-based method as described in Chapters 3 and 4, using the purified Sin resolvase hyperactive mutant Q115R and the Tn21 resolvase hyperactive mutant M63T. The second approach is based on making a hybrid protein consisting of the Sin resolvase Q115R catalytic domain and the Tn21 resolvase DNA-binding domain. This hybrid resolvase was to be tested for recombination activity *in vivo* using substrates such as pDW25 (Tn21 site I × Tn21 site I), pMP388 (Sin site I × Sin site I), pDW24 (Tn21 *res* × Tn21 *res*) and pSB423 (Sin *res* × Sin *res*). If the sequence preference of the Sin and Tn21 resolvase catalytic domains is indeed very similar, the hybrid protein containing the Tn21 resolvase DNA-binding domain should preferentially bind and recombine sites associated with Tn21 resolvase even though its catalytic domain comes from Sin resolvase.

6.2.1 Random site I library selections

To determine the sequence preferences in the centre of the Tn21 and Sin crossover sites, an analogous dimer resolution approach to the one described for Z-R(NM) (Section 4.2) was employed. Two mutant site libraries (LibS1 and LibS2) were designed. In LibS1, all 18 bases in the centre of the Sin site I were randomized. In the LibS2 library the central six base positions were kept as the wild-type Sin site I sequence, while the six bases flanking the central six on either side were randomised. The selection was performed for two rounds using purified Sin resolvase mutant Q115R (Section 2.25) just as described in Section 4.2 for Z-R(NM) selection. Selected monomer plasmid DNA was sequenced in bulk and the sequencing traces for LibS1 and LibS2 library selections are shown in Fig.6.3.

As can be seen from the bulk sequencing traces (Fig. 6.3) the results of selection were quite similar for both libraries. The most striking feature revealed by these selections is the conservation of the triple G motif at the positions 3, 4, 5L and corresponding C bases at

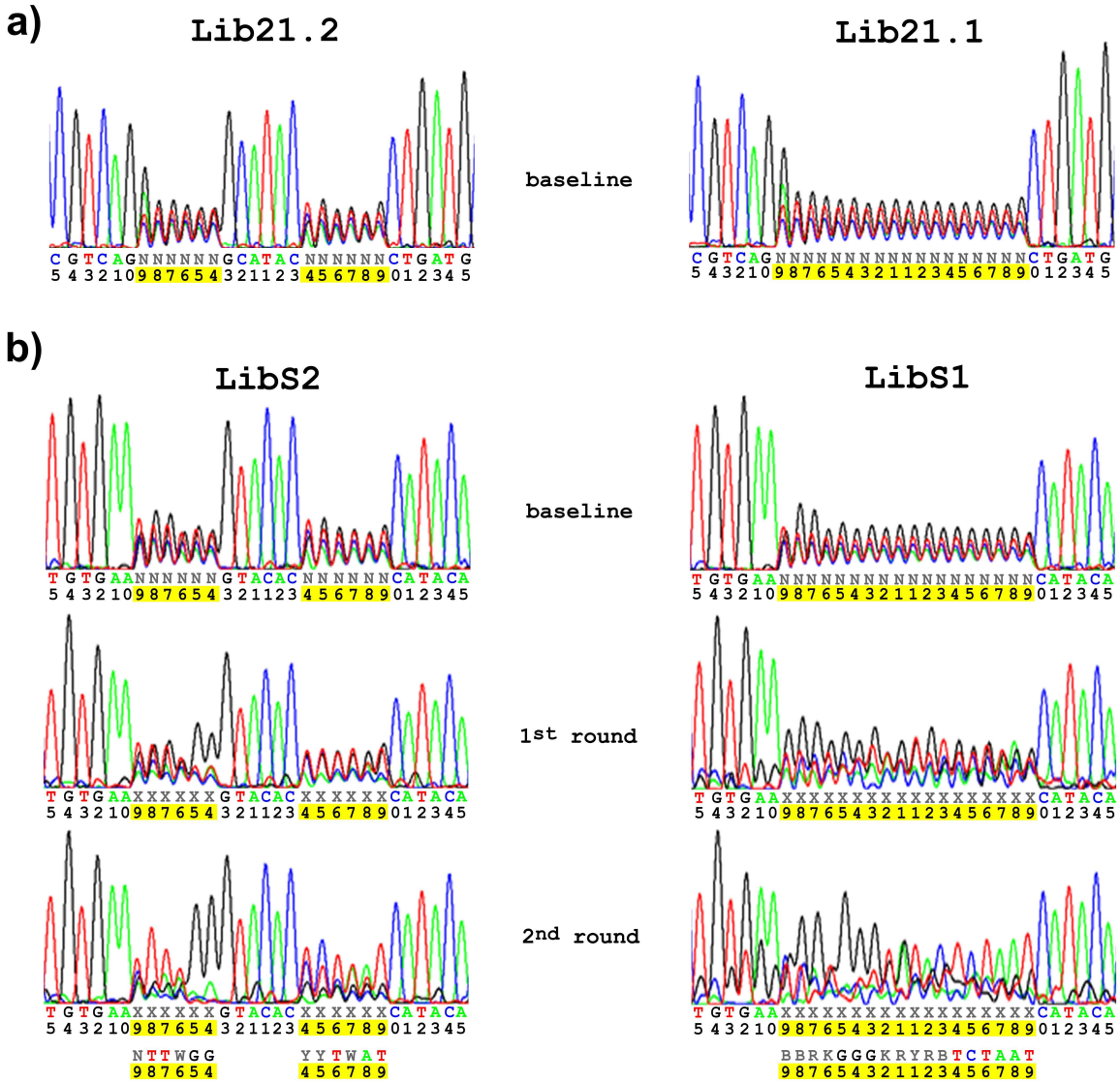


Figure 6.3- a) Bulk sequencing traces of the random block libraries Lib21.1 and Lib21.2 Sequencing was done using the primer uni -43. The individual DNA bases are numbered from the centre: 1-15 left and 1-15 right. For bases 10-15 only the second digit is shown. The position of the random block is highlighted in yellow below the sequencing trace. **b)** Bulk sequencing traces from the site selections using Sin resolvase Q115R from the random block libraries LibS1 and LibS2. Underneath the bulk sequencing trace from the 2nd round of selections, the sequence of the randomised regions of the sites is given using the extended IUPAC DNA notation (see Abbreviations). The individual DNA bases are numbered from the centre: 1-15 left and 1-15 right. For bases 10-15 only the second digit is shown.

positions 3R, 5R and to the lesser extent 4R. As was discussed in Chapter 3, the corresponding positions in Tn3 site I are contacted by residues I122, I123, T126, R130 and F140 in the minor groove, which are also the residues that are variable between resolvases recognising site I sequences that are different from that of Tn3 resolvase (Section 3.8 ;Yang & Steitz, 1995). In the selections using library LibS1 there was a propensity for G bases at positions 9-7L, which is not mirrored in the selections using the library LibS2 where these positions are more likely to be a T. There was more of an agreement between the results of selections using the two libraries at positions 6-9R, with a T base being preferred at positions 6R and 9R, and an A base at position 8R in both libraries. While 7R is found to be predominantly an A in selections using library LibS1, in the selections using library LibS2, A and T bases are equally represented at this position. Selections across the central dinucleotide using the LibS1 library suggest that the natural central dinucleotide of the Sin site I (AC) is underrepresented and that more common sequences at this position are AT or GT. This observation agrees with the experimental data showing that a Sin site I with an AT central dinucleotide is more efficiently recombined than the one with an AC centre (Rowland *et al.,* 2005).

The equivalent Tn21 crossover site libraries were constructed (Lib21.1 and Lib21.2) (Fig. 6.3) with the intention to use them for site selections with an activated Tn21 mutant. However, since purified Tn21 resolvase M63T was shown not to be active (Section 4.7), the selections were not carried out. Therefore, a direct comparison between sequence requirements in the central 18 bp of Sin site I and Tn21 site I was impossible.

6.2.2 Sin-Tn21 resolvase hybrid protein

Concomitantly to the random site I libraries selections, a hybrid resolvase consisting of Sin resolvase Q115R catalytic domain (residues 1-148) and the Tn21 resolvase DNA-binding C-terminal domain (residues 126-186) was designed (Fig. 6.4). The Sin resolvase residue R148 that the catalytic domain in this hybrid resolvase was terminated at, is the same position at which the Sin catalytic domain is terminated in the Z-R(Sin) (see Section 4.4).

To create pMP265, the plasmid expressing the Sin-Tn21 resolvase hybrid, the EagI/Asp718 fragment of pDW21 encoding the Tn21 resolvase DNA-binding domain was amplified by PCR, using the partially annealing forward primer hpF’ and the reverse primer 23R (Table 2.4, Section 2.7.1.1). Amplification using the hpF’ introduced a BsrGI site immediately next to the EagI site, allowing this fragment to be cloned into

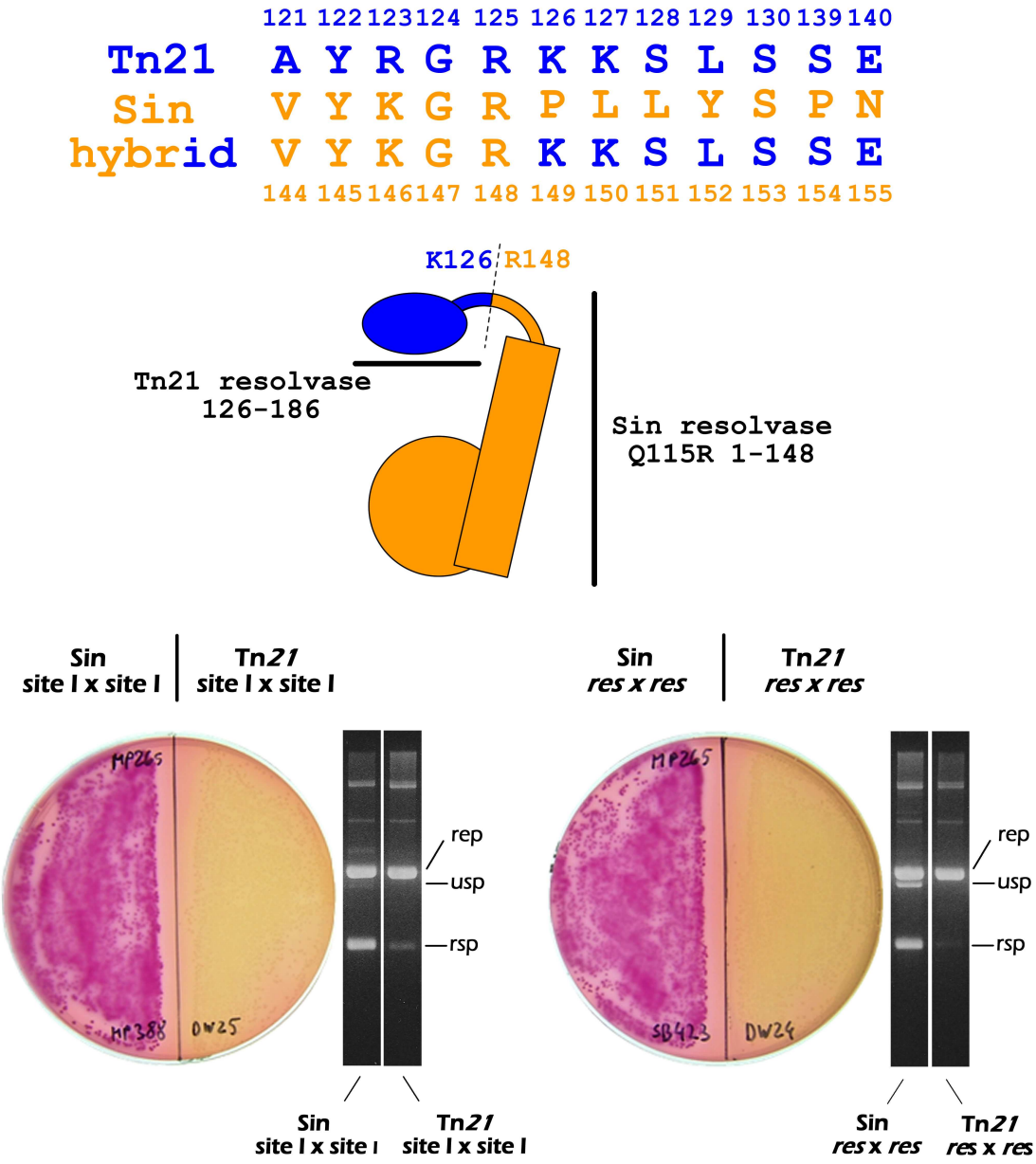


Figure 6.4-The Sin-Tn21 hybrid resolvase experiment. At the top of the figure is an alignment of Tn21 resolvase, Sin-Tn21 hybrid resolvase and Sin resolvase protein sequence to show the join between Tn21 resolvase DNA-binding domain and Sin resolvase catalytic domain. Sin resolvase protein sequence and numbering are given i orange while Tn21 resolvase sequence is given in blue. In the middle, a cartoon showing the Sin-Tn21 hybrid resolvase with its Tn21 resolvase DNA-binding domain coloured in blue and the Sin resolvase catalytic domain coloured in orange. Below this panel are the MacConkey assay result with the Sin-Tn21 hybrid acting on substrates pMP388 (Sin site I × Sin site I), pDW25 (Tn21 site I × Tn21 site I), pSB423 (Sin *res* × Sin *res*) and pDW24 (Tn21 *res* × Tn21 *res*). The type of substrate used is indicated on top of the each of the plate photographs. The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows, "rep" hybrid resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

BsrGI/Asp718-digested pSA9994 vector, effectively replacing the Q115R Sin resolvase DNA-binding domain with that of Tn21 resolvase.

The Sin-Tn21 hybrid resolvase was tested for recombination on substrates pSB423 (Sin *res* × Sin *res*), pMP388 (Sin site I × Sin site I), pDW24 (Tn21 *res* × Tn21 *res*), and pDW25 (Tn21 site I × Tn21 site I) *in vivo* using the MacConkey assay (Fig. 6.4).

When acting on pDW24 and pDW25 Sin-Tn21 hybrid resolvase produced white colonies. Conversely, the equivalent tests with the Sin substrates to recombine substrates pSB423 and pMP388 resulted in red colonies. However, after the plasmid DNA was isolated and analysed by agarose gel electrophoresis it was revealed that the Sin-Tn21 hybrid resolvase resolved pSB423 (Sin *res* × Sin *res*) and pMP388 (Sin site I × Sin site I) quite efficiently, but with some traces of unresolved substrate plasmid still remaining. As can be seen in the lanes showing Sin-Tn21 hybrid resolvase acting on pDW24 (Tn21 *res* × Tn21 *res*) and pDW25 (Tn21 site I × Tn21 site I) all of the substrate plasmid DNA was resolved (Fig 6.4). The reduction in the total amount of substrate plasmid DNA however that can be seen in these lanes is reminiscent of the recombination reactions using Z-R(NMF) (Section 5.2). As in that case, the Sin-Tn21 hybrid resolvase might be cleaving the substrate plasmids containing Tn21 sites without properly religating them. Plasmids that are not religated would presumably be promptly destroyed by the cell, which in turn would lead to the reduction of total resolved plasmid DNA observed.

These results suggest that Sin-Tn21 hybrid resolvase preferentially recombines substrates containing Tn21 resolvase recombination sites, despite having the Sin resolvase catalytic domain. This indicates that the Sin catalytic domain makes very little distinction between the central sequences of Tn21 *res* site I and Sin site I, provided that it can bind to it stably, either by using the Zif268 DNA-binding domain as in Z-R(Sin) (Section 4.7) or the native Tn21 resolvase HtH DNA binding domain as in the Sin-Tn21 resolvase hybrid. This means that the sequence specificity of the Sin catalytic domain must be very similar to that of the Tn21 resolvase catalytic domain.

6.3 Attempts to alter the sequence specificity of the resolvase catalytic domain

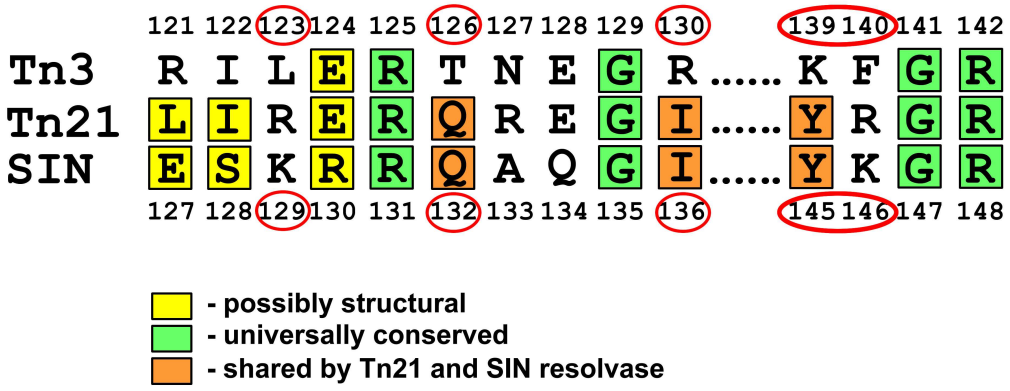
To better understand the factors determining the specificity of the resolvase catalytic domain, an experiment aimed at altering the specificity of the Sin resolvase catalytic domain into that of the Tn3 resolvase catalytic domain and *vice versa* was performed.

It was shown in the previous sections that the sequence selectivity of the Sin resolvase catalytic domain is very similar to that of the Tn21 resolvase catalytic domain but different to that of Tn3 resolvase. Comparing the random site library selection results using NM resolvase in Chapter 3 to the ones obtained using Sin resolvase mutant Q115R (Section 6.2) shows that the most conserved bases in their respective site Is are the ones that are contacted by the residues: 122, 123, 126, 130, 140 (numbering corresponds to Tn3 resolvase positions) in the minor groove (Yang & Steitz, 1995; Mouw *et al.*, 2008), which also happen to be the bases that are highly variable between different resolvase families (M. Boocock, personal communication). As these residues seem to be contacting the DNA bases themselves, it was suggested in the literature (Yang & Steitz, 1995) that they might be likely candidates for imposing sequence specificity on the resolvase catalytic domain.

To test whether the DNA contacts made by the "extended arm" residues 122, 123, 126, 130 and 140 are important for the sequence selectivity displayed by the resolvase catalytic domain it was decided to mutate the candidate residues in Sin resolvase into their Tn3 resolvase equivalents and vice versa, making Sin/3 and Tn3/S resolvase, respectively (Fig.6.5). The decision on which residues to change was partially based on the observation that the Sin and Tn21 resolvase catalytic domains are able to recombine each other's sites (Chapter 4 and Section 6.2) meaning that any differences between the protein residues involved in the DNA contacts in Tn21 and Sin resolvase are likely to be of lesser significance than those that exist between Tn21 (or Sin) resolvase and the Tn3 resolvase. As the residue 122 in both Tn21 and Tn3 resolvases is an isoleucine, and its equivalent 128 in Sin resolvase is a serine, these residues although implicated in DNA contacts, were not mutated. Furthermore, based on the recent Sin resolvase structure, residue S128 appears to play a role in Sin *res* site synapsis (Mouw *et al.*, 2008). The residues K139-F140 in Tn3 resolvase and residues Y145-R146 in Sin resolvase were swapped together as a block, as K139 and R146 seem to act to stabilise the stacking contact that the adjacent residue phenylalanine and tyrosine rings respectively, make with the deoxyribose ring of the DNA backbone (Yang & Steitz, 1995; Mouw *et al.*, 2008).

Plasmids pMP282 and pMP287 expressing Sin/3 resolvase (that is, Sin resolvase with the mutations Q115R, K129L, Q132T, I136R, Y145K, K146F) and Tn3/S resolvase (Tn3 resolvase with the mutations R2A, E56K, G101S, D102Y, M103I, Q105L, V107F, L123K, T126Q, R130I, K139Y, F140K) respectively, were made as follows. pMP282 was constructed by cloning the short NcoI/BamHI fragment from the GeneART1 plasmid (Section 2.6) introducing the mutations K129L, Q132T, I136R, Y145K, K146F into pSA9994, a Sin resolvase Q115R expression plasmid. To make pMP287, oligonucleotides

a)



b)

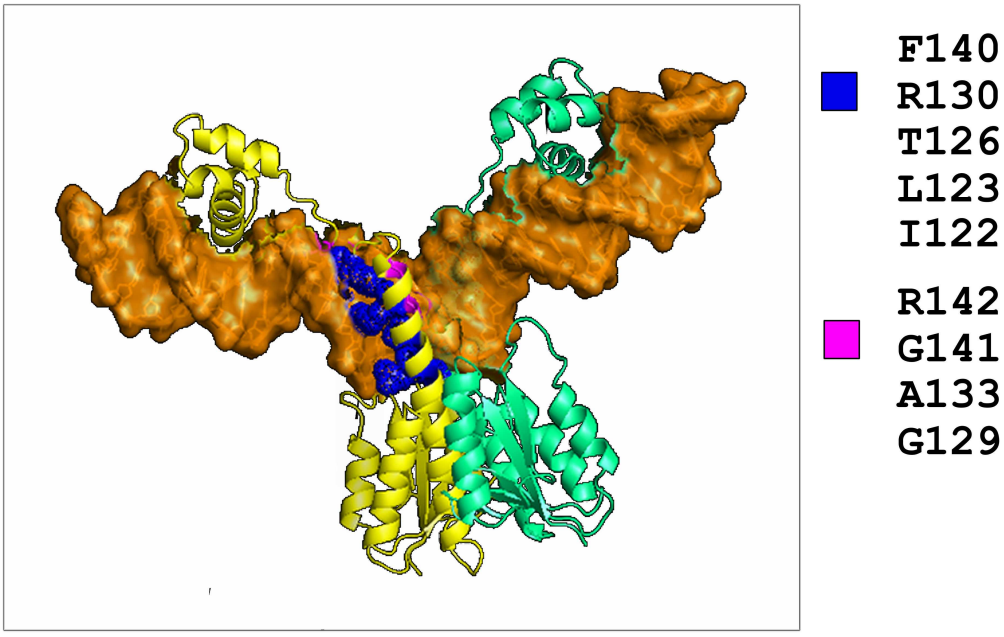


Figure 6.5-a) An alignment of Tn3, Tn21 and Sin resolvase extended arm residues that form contacts with the DNA in the centre of their respective site Is. The numbering at the top of the alignment pertains to Tn3 resolvase whereas the numbering at the bottom of the alignment corresponds to Sin resolvase. The residues that were changed in the experiment aimed at altering the specificity of the Tn3 resolvase catalytic domain into that of the Sin resolvase catalytic domain are circled in red (see Section 6.3 for details). **b)** The positions of the residues that contact the DNA in the centre of the site I are shown in spacefill on the γδ resolvase dimer (cartoon representation) structure bound to DNA (1gdt, Yang & Steitz, 1995). Residues contacting the DNA backbone are coloured magenta, whereas the residues contacting the DNA bases directly are in blue. For a more detailed view please refer to Figs. 3.29 and 3.30.

3toSinF and 3toSinR (Table 2.4) were first annealed, creating BamHI/EagI compatible ends, and then cloned into BamHI/EagI-restricted pMP17 vector, which encodes hyperactive Tn3 resolvase NMF.

After the expression plasmids pMP282 and pMP287 were created, Sin/3 and Tn3/S resolvases were tested for recombination *in vivo*, using substrate plasmids pMP243 (Tn3 site I × Tn3 site I) and pMP388 (Sin site I × Sin site I), in a MacConkey assay (Fig.6.6). *In vivo* assays with Sin resolvase Q115R and NMF resolvase, expressed from pSA9994 and pMP17 respectively, acting on the same substrates were included as a control. As can be seen in Fig.6.6, Sin/3 and Tn3/S resolvase failed to recombine any of the substrates tested, producing red colonies in all cases. Even after the plasmid DNA was isolated from the cells and analysed by agarose gel electrophoresis, no trace of resolution could be observed. Failure of Sin/3 resolvase to recombine pMP243 (Tn3 site I × Tn3 site I), and Tn3/S to recombine pMP388 (Sin site I × Sin site I) is unsurprising and could be attributed to the lack of sufficient binding to the sites. Tn3/S and Sin/3 have the native HtH DNA-binding domains of Tn3 resolvase and Sin resolvase respectively, and are therefore unlikely to bind tightly to the sites that do not have their recognition sequences. More interestingly, the failure of Sin/3 to recombine pMP388 (Sin site I × Sin site I), and of Tn3/S to recombine pMP243 (Tn3 siteI × Tn3 site I) showed that the changes introduced in the extended arm have a negative effect on recombination, reinforcing the importance of these residues.

Tn3/S and Sin/3 could have failed to recombine each other's sites due to lack of binding. However, even if binding were not an issue, Tn3 site I and Sin site I are of an unequal length (28 and 30 bp, respectively), so recombination by Sin/3 and Tn3/S on each other's sites would likely fail. A way to circumvent both of these issues (lack of binding and the unequal length of the sites) was to create Tn3/S and Sin/3 Z-resolvases (Z-R(Tn3/S) and Z-R(Sin/3). In the Z-resolvase context the binding is no longer an issue as both the DNA-binding domains and the ZBS are identical. The length of the Z-sites is also identical (40bp; 22 bp central sequence flanked by two 9 bp ZBSs)

Plasmids expressing Z-R(Tn3/S) and Z-R(Sin/3) were constructed (pMP385 and pMP284 respectively). To make pMP385, oligonucleotides 3toSinF and 3toSinR (Table 2.4) were first annealed creating BamHI/EagI-compatible ends, and then cloned into BamHI/EagI-restricted pAMC11 vector, which encodes Z-R(NMF). pMP284 was made in two steps. First, pMP379 was constructed by cloning the short MfeI/BstEII fragment from the GeneART2 plasmid (Section 2.6) into pMP213, the Z-R(Sin) expression plasmid, in order to remove the second SpeI site. In the second step, a short NcoI/SpeI fragment from the

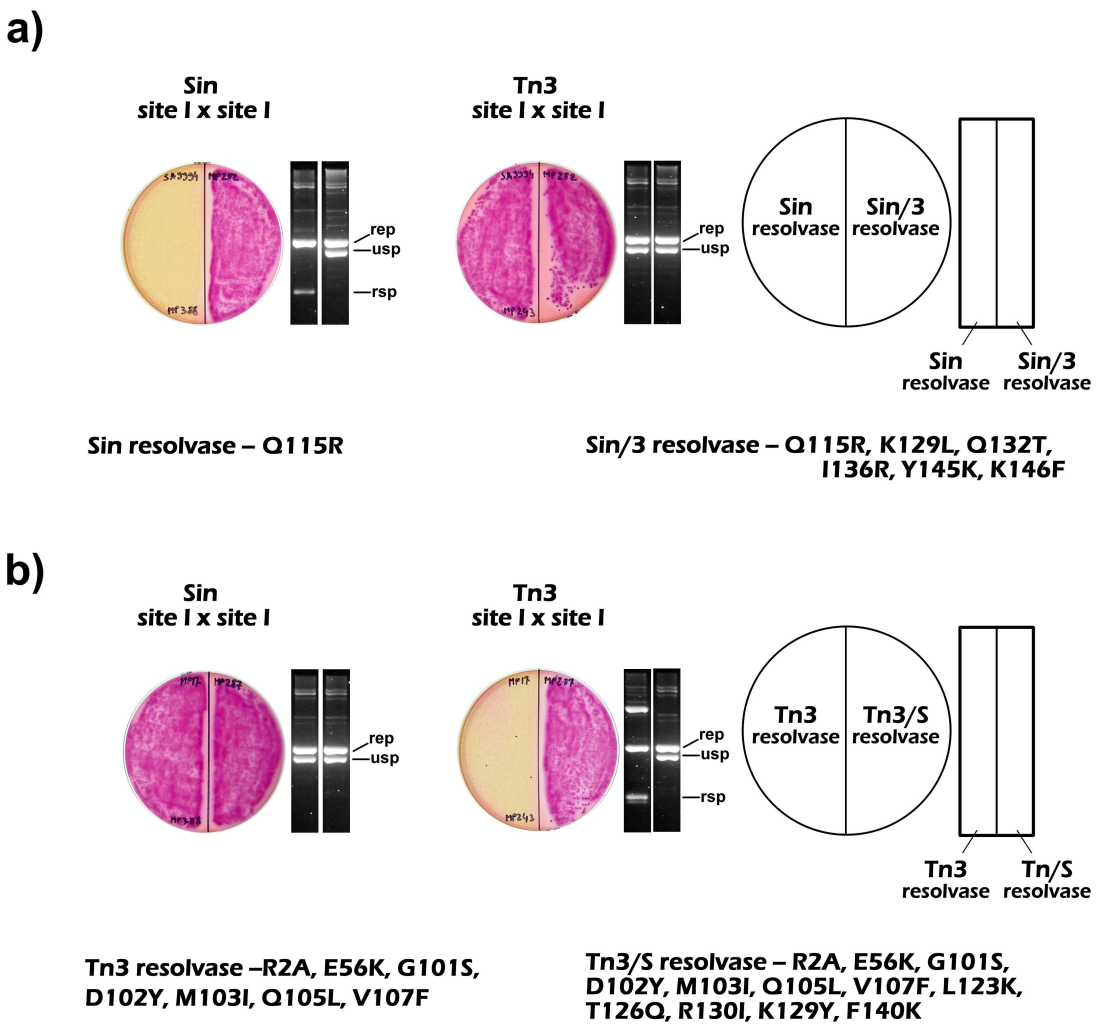


Figure 6.6- A MacConkey assay results with the Sin resolvase Q115R, Sin/3 resolvase **a)** and Tn3 resolvase NMF, Tn3/S resolvase **b)** acting on substrates pMP388 (Sin site I × Sin site I) and pMP243(Tn3 site I × Tn3 site I). The type of substrate used is indicated on top of the each of the plate photographs. The mutations in the resolvase catalytic domain are marked at the bottom of each panel. How the different resolvases are arranged on plates is indicated in the explanatory keys (to the right of the plate photographs). The agarose gels show the uncut plasmid DNA isolated from the plate adjacent. DNA bands are marked as follows; "rep" resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

GeneART2 plasmid introducing the mutations K129L, Q132T, I136R, Y145K, K146F was cloned into NcoI/SpeI-restricted pMP379 vector, creating pMP284.

Z-resolvases Z-R(Tn3/S) and Z-R(Sin/3), expressed from pMP385 and pMP284 respectively, were tested *in vivo* (Fig. 6.7) in a MacConkey assay using substrates with two copies of Z22Z (pMP53) or two copies of the Sin Z-site (pMP217). The equivalent Z-resolvases, Z-R(NMF) and Z-R(Sin) without the 'specificity altering' mutations were used as controls.

As can be seen in Fig. 6.7, Z-R(Sin/3) produced red colonies when acting on either pMP217 (Sin Z-site × Sin Z-site) or pMP53 (Z22Z × Z22Z). No trace of resolution was observed with this Z-resolvase even when the isolated plasmid DNA was analysed using agarose gel electrophoresis.

Z-R(Tn3/S) conferred a lethal phenotype on cells grown on selective MacConkey agar if either of the resolution substrate plasmids was present. If the cells were grown on selective L-agar the lethal phenotype was not observed and colonies of normal size and morphology were observed. A possible explanation for this would be that in the presence of a very active resolvase the substrate plasmid might have a problem establishing itself in the cells post-transformation, as it is attacked and damaged by resolvase as soon as it is enters the cell. Since the actual number of substrate plasmid molecules entering the cell at transformation is not very large a Z-resolvase protein that cleaves the substrate without efficiently religating it might cause a reduction in the amount of colonies observed. This occurrence might be attributable to the influence of the V107F mutation, the effects of which were discussed in the previous chapters. Extraction of the plasmid DNA from these cells followed by separation on an agarose gel revealed that Z-R(Tn3/S) partially resolves pMP217, while failing to resolve pMP53 (Fig. 6.7).

6.4 Discussion

In this chapter it was established that the catalytic domains of Sin and Tn21 resolvase have very similar sequence selectivity despite being very distantly related. The Sin-Tn21 resolvase hybrid constructed from the hyperactive Sin resolvase mutant Q115R catalytic domain and Tn21 resolvase DNA-binding domain recombined substrates containing two copies of Tn21 site I more efficiently than those containing two copies of Sin site I. This was in contrast to the earlier experiments involving resolvase hybrids (Ackroyd *et al.*, 1990; Avila *et al.*, 1990), where hybrids made by joining Tn21 resolvase catalytic domain to the Tn3 resolvase DNA-binding domain and vice versa, were found to be able to bind to

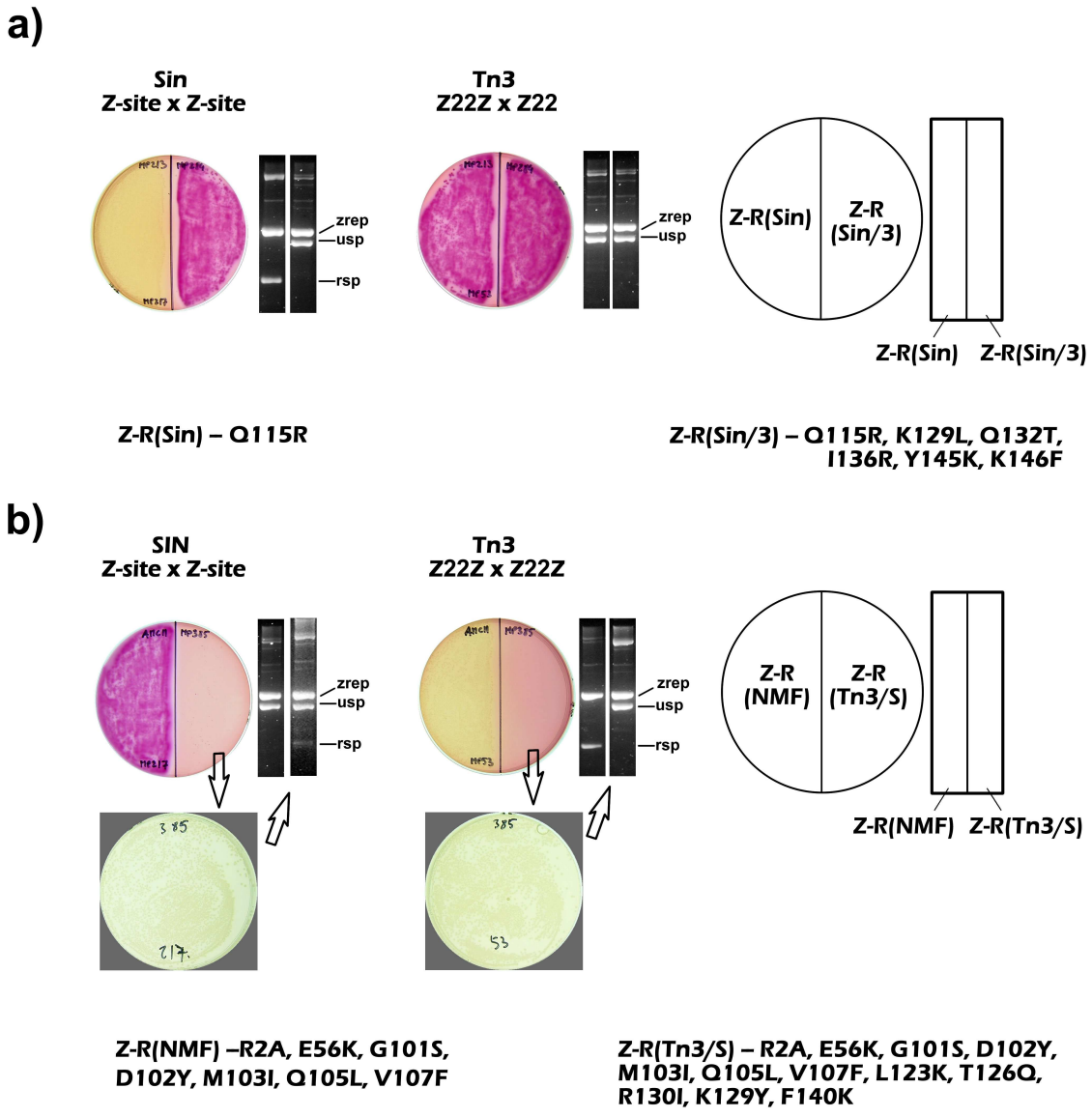


Figure 6.7- A MacConkey assay results with the Z-resolvases Z-R(Sin), Z-R(Sin/3) resolvase **a)** and Z-resolvases Z-R(NMF), Z-R(Tn3/S) **b)** acting on substrates pMP217 (Sin Z-site × Sin Z-site) and pMP53 (Z22Z × Z22Z). The type of substrate used is indicated on top of the each of the plate photographs. The mutations in the Z-resolvase catalytic domain are marked at the bottom of each panel. How the different Z-resolvases are arranged on plates is indicated in the explanatory keys (to the right of the plate photographs). The agarose gels show the uncut plasmid DNA isolated from the plate adjacent(No colonies could be obtained with Z-R(Tn3/S) when grown on MacConkey agar due to toxicity so the DS941 cells was grown on selective LB agar and the plasmid DNA extracted from them). DNA bands are marked as follows; "rep" Z-resolvase expression plasmid, "usp" unresolved substrate plasmid "rsp" resolved substrate plasmid. Red colonies represent inefficient resolution while white colonies represent efficient resolution (see Section 2.23 for details).

non-cognate sites, but were unable to catalyse recombination using these sites. While Tn3-Tn21 resolvase hybrid (catalytic domain of Tn3 resolvase joined to the DNA-binding domain of Tn21 resolvase) was able to catalyse recombination of substrates containing Tn3 *res* sites, the Tn21-Tn3 resolvase hybrid (catalytic domain of Tn21 resolvase joined to the DNA-binding domain of Tn3 resolvase) was not able to recombine substrates containing either Tn3 or Tn21 *res* sites. This result suggests that the Tn21 resolvase DNA-binding domain is more promiscuous than the Tn3 resolvase DNA-binding domain and also that the differences in the sequence selectivity in the centre of site I between Tn21 resolvase and Tn3 resolvase catalytic domains might be too great to allow for recombination of each other's sites. Alternatively, the failure of the Tn21-Tn3 and Tn3-Tn21 resolvase hybrids to recombine each other's sites might be due to the difference in length of Tn3 and Tn21 site I (28 bp and 30 bp, respectively).

It should be pointed out that these earlier experiments (Ackroyd *et al.*, 1990; Avila *et al.*, 1990) were done in the context of *res*, using wild-type catalytic domains which is different to the Sin-Tn21 hybrid resolvase experiment reported in this chapter (Sin resolvase catalytic domain contained an activating mutation Q115R). Also the difference in length of site I's was not an issue in the Sin-Tn21 hybrid resolvase experiment since both Sin and Tn21 site I's are 30 bp long.

In an experiment in which Sin-Tn21 hybrid was made using a wild-type Sin resolvase catalytic domain (pMP264) (data not shown) and tested on the same substrate plasmids as described in Section 6.2.2. this hybrid resolvase was found to be inactive with all substrates tested, suggesting that the hybrid resolvase fails to act properly on the accessory sites of either Sin *res* or Tn21 *res* sites. Since the Sin resolvase system requires C-terminal domain synapsis between resolvase subunits bound at site I and site II (Mouw *et al.*, 2008), the hybrid protein with the foreign Tn21 resolvase C-terminal domain (DNA-binding domain) is likely unable to establish appropriate contacts. To test this hypothesis a resolvase hybrid using Tn21 resolvase wild-type catalytic domain and Sin resolvase DNA-binding domain would need to be constructed and tested to establish whether it is capable of recombination using a Sin *res* × *res* substrate, however due to time constraints this was not attempted.

Since the hyperactive catalytic domains do not require contact with the accessory sites (Rowland *et al.*, 2005) the original Sin-Tn21 resolvase hybrid (Section 6.2.2.) was able to recombine both site I × site I and *res* × *res* substrates with comparable efficiency (Fig. 6.4).

In the second part of this chapter, Sin and Tn3 resolvase and Z-resolvase mutants in which residues implicated in making contacts with the centre of site I were swapped, were created and tested for recombination using substrates containing two copies of Sin site I, Tn3 site I, Sin Z-site and Z22Z (Tn3 site I sequence-derived Z-site). Swapping these residues in hyperactive catalytic domains of Tn3 and Sin resolvase (L123K, T126Q, R130I, K139Y, F140K and K129L, Q132T, I136R, Y145K, K146F, respectively) rendered them inactive with all substrates tested, highlighting their importance. As discussed in Section 3.8 these residues are probably important at the level of binding, which is also in concordance with an earlier study (Hughes *et al.,* 1990), which showed that mutating residue R130 of γδ resolvase into a histidine severely affects binding.

In contrast, swapping the same residues in Tn3 and Sin Z-resolvases yielded contrasting results. While Z-R(Sin/3) failed to recombine any substrates, Z-R(Tn3/S) was able to partially recombine pMP217, a substrate containing two copies of Sin Z-site. Since Z-R(Tn3/S) did not recombine substrate pMP53 (Z22Z × Z22Z) it is evident that sequence specificity of hyperactive Tn3 resolvase catalytic domain in Z-R(Tn3/S) was not relaxed but actually changed into that of Sin resolvase catalytic domain. The recombination efficiency of Z-R(Tn3/S) was much lower than that of either Z-R(NMF) or Z-R(Sin) suggesting that only swapping the residues making direct contacts with the DNA might not be the optimal way of swapping the sequence specificity of the resolvase catalytic domain. It could be that other residues adjacent to the ones swapped play a role in maintaining the correct context for the residues 122, 123, 126, 130 and 140 or their equivalents to be able to make appropriate contacts with the DNA in the centre of site I. Further research will be needed to establish whether this is the case.

The difference between results obtained using Z-R(Sin/3) and Z-R(Tn3/S) could be due to the different levels of hyperactivity displayed by their respective catalytic domains. A single activating mutation Q115 in Sin resolvase catalytic domain could be less activating than a multiple mutations in NMF Tn3 resolvase and therefore not sufficient to overcome the inhibitory effect that sequence selectivity swap appears to have. In recent *in vivo* assays (C. Proudfoot, unpublished) it was shown that the rate of recombination by Z-R(Sin) is slightly slower than that catalysed by Z-R(NMF). The toxicity effect observed with Z-R(Tn3/S) is most likely due to the adverse influence of the V107F mutation that is exacerbated by the presence of specificity altering mutations. The influence of the V107F mutation was discussed in the previous chapter. It would be interesting to see whether removing this mutation from Z-R(Tn3/S) would alleviate the *in vivo* toxicity issues. Alternatively Z-R(Tn3/S) could be tested *in vitro*. To that effect an over-expression

plasmid was constructed (pMP136, see Table 2.5) and the over-expression of Z-R(Tn3/S) was successfully induced. However, due to lack of time Z-R(Tn3/S) was not purified.

In conclusion, although results of this chapter are encouraging with regards to the possibility to engineer the sequence selectivity of a resolvase catalytic domain, they highlight the need for further experiments in order to fulfil this goal.

7 Chapter 7: Conclusions and general discussion

Over the last four chapters, the factors influencing sequence specificity of resolvase catalytic domain have been explored along with the practical aspects of Z-resolvase design. In the following sections the conclusions of the study and the possible future directions for this work are discussed.

7.1 Sequence specificity of resolvase catalytic domain

In Chapter 3, the extent of sequence specificity of the Tn3 resolvase catalytic domain was investigated employing a series of libraries of Tn3 site I variants in which all of the central 16 bp were systematically randomised in overlapping 4 bp blocks. Using an activated Tn3 resolvase mutant NM, recombination deficient and recombination proficient site I variants were selected by two different independent selection strategies.

Comparison between these two mutually exclusive sets of Tn3 site I variants revealed that most positions can be changed without abolishing recombination if changed individually, but that accumulating several changes is detrimental. In general, changes from an A to a T, or a T to an A were much better tolerated than substitutions of an A or a T with a G or a C suggesting the recognition in the minor groove, which is consistent with the available γδ resolvase-DNA co-crystal structures which show extensive minor groove contacts between the extended arm region and bases at positions 1-8 (Figs. 3.29 and 3.30) (Yang & Steitz, 1995, Li *et al.,* 2005). Recombination proficient Tn3 site I variant sequences were found to strongly resemble known site I sequences associated with the Tn3 resolvase family. The particularly well conserved sequence feature was the characteristic pyrimidine-purine step found at the junction between positions 1 and 2, which coincides with the position of the scissile bond. The selection was strongest at the positions 5R, 5L, 6R, 6L, 3R, 1L, 7L and 8L, while position 8R and 4L appeared to be practically random, hinting at a degree of asymmetry in the way resolvase subunits interact with left and right half-site of site I, a possibility that was raised previously in the literature (Yang & Steitz, 1995).

As the assays in this study were based on full recombination it is impossible to be certain whether sequence discrimination by NM resolvase takes place in the binding and/or catalytic steps. Nevertheless it seems that specificity in the central 16 bp of Tn3 site I is imparted by a combination of direct contacts at position 3 and indirect readout based on an AT-hook motif-like binding in the minor groove between positions 8 and 5. NM resolvase seems to interact with the left and right half-sites of Tn3 site I in a slightly different

manner. At the left half-site, presumably due to the weak binding of the HtH, site contacts between positions 8-5 and the existence of a Y-R base step between positions 1 and 2 appear to be the most important. At the right half-site, although the aforementioned interactions are important, specific contacts to the position 3 seem to take precedence.

The results presented in Chapters 4 and 6 show that the types of interactions of resolvase catalytic domain with the recombination site sequence are similar to the ones described in Chapter 3 regardless of whether the catalytic domain is in the resolvase or the Z-resolvase context, or whether the catalytic domain comes from Tn3 resolvase or very distantly related Sin resolvase.

The logical next step in this investigation would be to establish whether the sequence discrimination takes place at the moment of binding, synapsis, cleavage or at an even later catalytic step. To do this further binding assays using recombination deficient site I variants need to be performed, both under binding and synapsis conditions. It is likely that different site I sequence variants are going to be inhibited at different steps in the reaction. Site I variants with mutations at the positions 8-5 that widen the minor groove are likely to display reduced binding affinity, if the AT-hook like binding mechanism is correct. Site I variants with mutations at position 3 could be inhibited later in the reaction, possibly at the point of synapsis, while the mutants that do not have Y-R base step between positions 1 and 2 are most likely inhibited at the cleavage step.

Swapping the extended arm region residues, contacting the site I bases in the minor groove, between Tn3 Z-resolvase and Sin Z-resolvase resulted in the sequence selectivity swap, with Tn3 resolvase-based Z-resolvase containing Sin resolvase arm region residues, Z-R(Tn3/S) being able to recombine a substrate containing Sin-Z sites (pMP217) and not the one containing Tn3 site I-based Z-sites (pMP53). However, recombination of pMP217 catalysed by Z-R(Tn3/S) was far less efficient than recombination of pMP53 catalysed by Z-R(NMF), also it was found that Z-R(Tn3/S) is quite toxic to the *E. coli* cells grown on the MacConkey medium. Furthermore, Z-R(Sin/3), a Sin Z-resolvase containing Tn3 resolvase arm region residues was found to be inactive on any of the recombination substrates tested.

This result highlights that the catalytic domain of resolvase has a degree of sequence selectivity which is beyond and independent on the C-terminal DNA-binding domain and that engineering the specificity of resolvase is possible, but that further work is necessary to fully understand the impact of the arm region changes and the mechanism of sequence

selectivity by the resolvase catalytic domain. The low efficiency of recombination and the toxicity effects could be due to improper context in which swapped residues find themselves. Potentially this could be ameliorated by swapping the whole arm region section instead of just the specific residues.

7.2 Practical aspects of Z-resolvase design and the future prospects

In Chapters 4 and 5 various aspects of Z-resolvase and Z-site design have been investigated. It was shown that given a hyperactive resolvase catalytic domain functional Z-resolvases with corresponding sequence selectivity can be made. While the optimal distance between the zinc finger-binding sites (ZBS) was shown to be 22 bp it was established that there is a degree of flexibility in this distance requirement with distances of 20, 21, 23 and 24 bp being tolerated. Furthermore it was demonstrated that complementation of Z-resolvase with a resolvase was possible.

The ability of Z-resolvase to catalyse recombination on asymmetric sites and to be complemented by the resolvase with the same catalytic domain, on specially designed hybrid sites, is encouraging with regards to the future applications of this system. These abilities greatly extend the range of sites that Z-resolvases can potentially work on. Despite best efforts some DNA sequences are still proving recalcitrant to recognition by zinc fingers (Wu *et al.*, 2007). Having a degree of flexibility in the allowed inter-ZBS distance could prove very useful when trying to avoid sequences that are particularly difficult as zinc fingers targets. Positive complementation result indicates the successful formation of heterotetramers consisting of Z-resolvases with different DNA-binding domains, which in turn means that recombination between sites with a same central sequence but different flanking sequences, is also possible. A very useful feature if Z-resolvases are to be used as tools for genomic surgery, as suggested (Akopian & Stark, 2005).

In the future the Z-resolvase system could develop along two different paths. It could be envisaged that further mutagenesis of the resolvase catalytic domain could ultimately result in a Z-resolvase in which all the sequence selectivity imposed by the catalytic domain has been removed. Such non-selective recombinases would be fully reliant on the sequence specificity that is inherent in their zinc finger DNA-binding domain for its targeting, and would face the same pitfalls as the partially analogous zinc finger nucleases (ZFN) that are currently being developed (Durai *et al.*, 2005; Carroll, 2008) such as aberrant cleavage due to imperfect recognition of target locus by the zinc fingers and the resulting cellular toxicity (Wu *et al.*, 2007; Cornu *et al.*, 2008).

Alternatively, Z-resolvases with custom designed sequence specificity that is not only relaxed but truly altered could be created. This could be achieved either via directed evolution (Fig 7.1) or through protein engineering approach which would require in depth understanding of the molecular basis of the extended arm region interactions with the centre of the recombination site. Such Z-resolvases are still far on the horizon but if they were ever created they would present an alternative to the zinc finger nuclease (ZFN) technology which would not be dependant on cellular pathways of DNA repair.

Custom designed Z-resolvase capable of recognising a specific sequence would find many applications in high end biotechnology (animal transgenesis) and molecular medicine. Initially, the most likely applications of Z-resolvase technology would be the modifications of stem cells or gene correction *ex vivo*, followed by reintroduction of modified cells back into the organism following the treatment. Similar modifications have already been achieved using ZFN technology (Perez *et al.*, 2008).

Provided that Z-resolvases can be delivered to cells, the range of possible applications of Z-resolvase technology would be extended, such as treatment of monogenic disorders or even retroviral diseases such as HIV (Fig. 7.2). Since numerous delivery methods that could be used to deliver genome manipulation tools as: expressable DNA constructs (Sorrell & Kolb, 2005), but also as proteins (Baubonis & Sauer, 1993), or protein-DNA complexes (Goryshin *et al.*, 2000), already exist and are continually being developed, this proposition does not seem too unrealistic. Once delivered to the cells Z-resolvases would face other challenges like reaching the nucleus and recognising their target sequence that could be packed in heterochromatin which could pose a problem. However, recent results (Gordley *et al.*, 2007) suggest that Z-resolvases will be active in human cells.

Apart from being useful for designing better Z-resolvases understanding the sequence selectivity by the resolvase catalytic domain could simplify future mechanistic studies by enabling the creation of fully targetable resolvase subunits. Being able to target resolvase to specific half-sites within *res* could be potentially very useful in experiments aiming to elucidate the structure of *res* × *res* synapse.

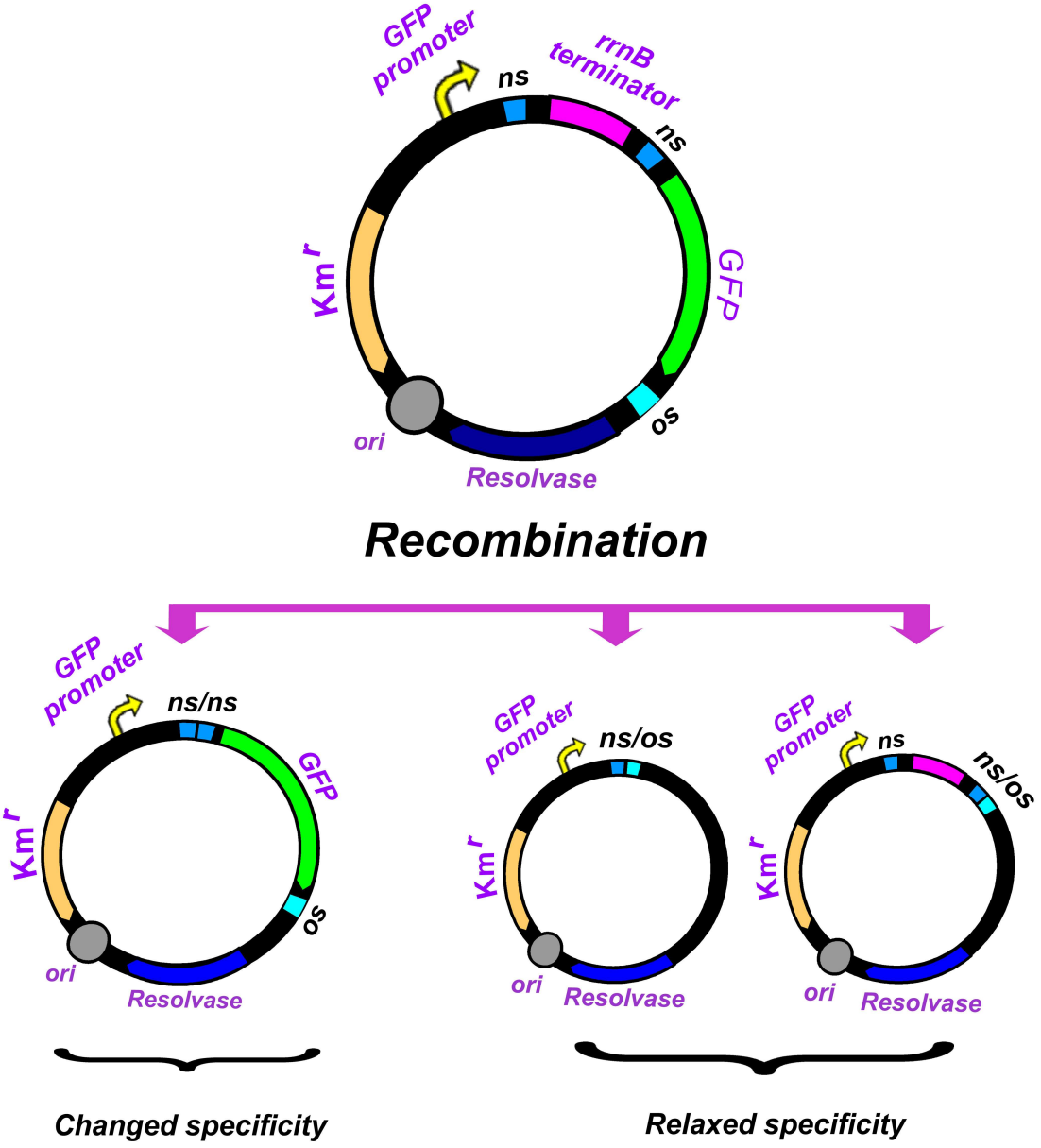


Figure 7.1- A schematic diagram showing a hypothetical plasmid construct that could be used to differentiate between resolvase mutants that have a relaxed sequence specificity, i.e can recombine a new site but retain activity on the original site, and the mutants with altered sequence specificity that are only active on the new site. Recombination between new sites (short blue sections labelled "ns") catalysed by resolvase protein expressed from the same plasmid (dark blue section) eliminates the transcription terminator *rrnB* (magenta section, Sohn & Kang, 2005), which in turn allows transcription and subsequent expression of GFP. Recombination between either of the new sites and the original site (short cyan section labelled "os") eliminates either the GFP ORF or both the GFP ORF and the *rrnB* terminator. In either case GFP is not expressed. Cells expressing GFP and the ones which do not could be relatively easily separated by FACS analysis. Plasmid DNA can then be isolated from GFP expressing cells and the sequence of the resolvase mutant determined

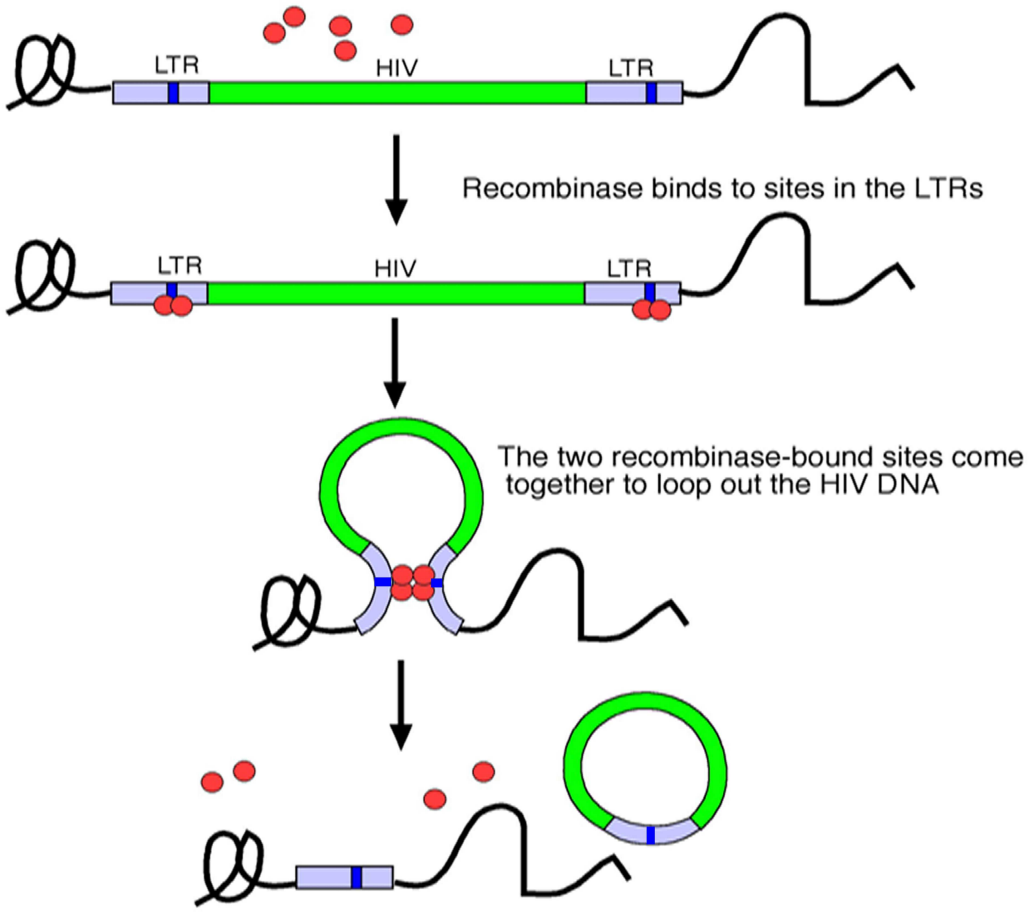


Figure 7.2- A diagram showing a hypothetical application of Z-resolvase to inactivate HIV. HIV DNA (green segment) integrated in the cells genome (black bendy line) is flanked by long terminal repeats (LTRs, light blue segments). Z-resolvase (red circles) is made that can recognise a conserved site (dark blue) situated in the HIV LTR that is less likely to mutate than the rest of the HIV genome, e.g. the HIV promoter. Z-resolvase is delivered to the cell's nucleus and it binds the sites found in the HIV LTRs. It then cuts and rejoins the DNA and effectively removes the HIV leaving the LTR behind. The circular HIV DNA is no longer infective as it is unable to reintegrate and is eventually destroyed by the cell enzymes.

References

1. Abdel-Meguid, S. S., Grindley, N. D., Templeton, N. S. and Steitz, T. A. (1984). Cleavage of the site-specific recombination protein gamma delta resolvase: the smaller of two fragments binds DNA specifically. *Proc. Natl. Acad. Sci. USA*. **81,** 2001-2005
2. Ackroyd, A. J., Avila, P., Parker, C. N., and Halford, S. E. (1990). Site-specific recombination by mutants of Tn21 resolvase with DNA recognition functions from Tn3 resolvase. *J. Mol. Biol.* **216,** 633–643.
3. Akopian, A. (2003). Site-specific DNA recombinases with novel sequence recognition. PhD Thesis University of Glasgow.
4. Akopian, A. and Stark, W. M. (2005). Site-specific DNA Recombinases as instruments for genomic surgery. *Adv. in Genet.* **55,** 1-23
5. Akopian, A., He, J., Boocock, M. R. and Stark, W. M. (2003). Chimeric recombinases with designed DNA sequence recognition. *Proc. Natl. Acad. Sci. USA* **100**, 8688-8691.
6. Ariyoshi, M., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2000). Crystal structure of the holliday junction DNA in complex with a single RuvA tetramer. *Proc. Natl. Acad. Sci. USA*. **97,** 8257-8262
7. Arnold, P. H., Blake, D. G., Grindley, N. D. F., Boocock, M. R. and Stark, W. M. (1999). Mutants of Tn*3* resolvase which do not require accessory binding sites for recombination activity. *EMBO J.* **18**, 1407-1414.
8. Avila, P., Ackroyd, A. J., and Halford, S. E. (1990). DNA binding by mutants of Tn21 resolvase with DNA recognition functions from Tn3 resolvase. *J. Mol. Biol.* **216,** 645–655.
9. Barbic, A., Zimmer, D. P. and Crothers, D. M. (2003). Structural origins of adenine-tract bending. *Proc. Natl. Acad. Sci. USA* **100,** 2369-2373
10. Barzel, A. and Kupiec, M. (2008). Finding a match: how do homologous sequences get together for recombination? *Nat. Rev. Gen.* **9,** 27-37.
11. Baubonis, W., and Sauer, B. (1993). Genomic targeting with purified Cre recombinase. *Nucl. Acids. Res.* **21,** 2025-2029.
12. Bednarz, A. L. (1990). Dissection of the Tn3 resolution site. PhD Thesis University of Glasgow.
13. Blake, D. G. (1993).Binding of Tn3 resolvase to its recombination site. PhD Thesis University of Glasgow.
14. Blake, D. G., Boocock, M. R., Sherratt, D. J. and Stark, W. M. (1995). Cooperative binding of Tn3 resolvase monomers to a functionally assymetric binding site. *Curr. Biol.* **5,** 1036-1046
15. Bode, J., Schlake, T., Iber, M., Schubeler, D., Seibler, J., Snezhkov, E. and Nikolaev, L. (2000). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol. Chem*. **381,** 801-813
16. Burke, M.E., Arnold, P.H., He, J., Wenwieser, S.V.C.T., Rowland, S. J., Boocock, M.R., and Stark, W.M. (2004). Activating mutations of Tn*3* resolvase marking interfaces important in recombination catalysis and its regulation. *Mol. Microbiol.* **51**, 937-948.
17. Burkhoff, A. M. and Tulius, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell*. **48,** 935-943
18. Callinan, P. A. and Batzer, M. A. (2006). Retrotransposable elements and human disease**.** *Genome Dyn*.**1,** 104-115.
19. Carroll, D. (2008). Progress and prospects: Zinc-finger nucleases as gene therapy agents. *Gene Therapy* **15,** 1463-1468
20. Chamber, S. P., Prior, S. E., Earstow, D. A. and Minton, N.P. (1988). pMTL nick-cloning vectors. I. improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene*. **68,** 139-149

21. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., and Rice, P. A. (2000). Crystal structure of a FLP recombinase – Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*. **6,** 885-897
22. Coates, C. J., Kaminski, J. M., Summers, J. B., Segal, D. J., Miller, A. D., Kolb, A. F. (2005). Site-directed genome modification: derivatives of DNA-modifying enzymes as targeting tools. *Trends in Biotech.* **23,** 407-419
23. Cornu T. I, Thibodeau-Beganny S., Guhl E., Alwin S., Eichtinger M., Joung J. K. and Cathomen, T. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther*.**16,** 352-358.
24. Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (ed.) (2002). *Mobile DNA II*. ASM Press, Washington, DC, USA.
25. Crane-Robinson, C., Dragan, A. I. and Privalov, P. L. (2006). The extended arms of DNA-binding domains: a tale of tails. *Trends Biochem. Sci.* **31,** 547-52.
26. Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188-1190.
27. Dragan, A. I., Klass, J., Read, C., Churchill, M. E., Crane-Robinson, C. and Privalov, P. L. (2003). DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. *J. Mol. Biol.* **331,** 795-813
28. Dreier, B., Beerli, R. R., Segal, D. J. Flippin, J. D. and Barbas, C. F., 3rd (2001). Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **276,** 29466-29478
29. Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., Koksch, B. and Barbas, C. F., 3rd (2005). Development of zinc finger domains for recognition of the 5'-CNN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* **280,** 35588-35597
30. Dreier, B., Segal, D. J. and Barbas, C. F., 3rd (2000) insights into molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J. Mol. Biol.* **303,** 489-502
31. Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M. H. and Chandrasegaran, S. (2005). Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.* **33,** 5978-5990
32. Elrod-Erickson, M., Rould, M. A., Nekludova, L. and Pabo, C.O. (1996). Zif268 protein-DNA complex refined a 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure.* **4,** 1171-1180
33. Gillen, J. R., Willis, D. K. and Clark, A. J. (1981). Genetic analysis of *Rec*E pathway of genetic recombination in *Echerichia coli* K-12. *J. Bacteriol*. **145,** 521-532
34. Gordley, R. M., Gersbach, C. A. and Barbas, C. F. 3rd (2009). Synthesis of programmable integrases. *Proc. Natl. Acad. Sci. USA.* **106,** 5053-5058
35. Gordley, R. M., Smith, J. D., Gräslund, T. and Barbas, C. F. 3rd (2007). Evolution of programmable zinc finger-recombinases with activity in human cells. *J. Mol Biol.* **367,** 802-813
36. Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., and Reznikoff, W.S. (2000). Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat. Biotechnol.* **18,** 97-100.
37. Grainge, I. and Jayaram, M. (1999). The integrase family of recombinase: organisation and function of the active site. *Mol. Microbiol.* **33,** 449-456
38. Griesman, H. A. and Pabo, C. O. (1997). A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275,** 657-661
39. Grindley, N.D.F. (2002). The movement of Tn*3*-like elements: transposition and cointegrate resolution. In *Mobile DNA II*, Craig, N., Craigie, R., Gellert, M. and Lambowitz, A. (ed.), pp272-302. ASM Press, Washington, DC, USA.
40. Grindley, N.D.F., Whiteson, K.L. and Rice, P.A. (2006). Mechanisms of site-specific recombination *Annu. Rev. Biochem.*;**75,** 567-605
41. Guo, F., Gopaul, D. N. and Van Duyne, G. D. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl. Acad. Sci. USA.* **96,** 7143-7148

42. Guo, H., Karberg, M., Long, M., Jones, J.P., 3rd, Sullenger, B. and Lambowitz, A.M. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science.* **289,** 452-457.
43. Hall, S. C. and Halford, S. E. (1993). Specificity of DNA recognition in the nucleoprotein complex for site-specific recombination by Tn21 resolvase. *Nucleic Acids Res.* **21,** 5712-5719
44. Hallet, B., and Sherratt, D. J. (1997). Transposition and site-specific recombination: Adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEBS Microbiol. Rev.* **21,** 157-178
45. Händel, E. M., Alwin, S. and Cathomen, T. (2009). Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.* **17,** 104-111
46. Harley, C. B. and Reynolds, R. P. (1987). Analysis of *E. coli* promoters. *Nucleic Acids Res.* **15,** 2343-2361
47. Hatfull, G. F. and Grindley, N. D. F (1988). In Kucherlapati, R. and Smith, G. R. (eds), *Genetic Recombination.* American society for Microbiology, Washington DC, pp. 357-396
48. Hatfull, G. F., Salvo, J. J., Falvey, E. E., Rimphanitchayakit, V., and Grindley, N. D. F. (1988). Site-specific recombination by the 78 resolvase. *Symp. Soc. Gen. Microbiol.* **43,** 149-181.
49. Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genet. Res.* **5,** 282-304
50. Hughes, R. E., Hatfull, G. F., Rice, P. A., Steitz, T. A., and Grindley, N. D. F. (1990). Cooperativity mutants of the γδ resolvase identify an essential interdimer interaction. *Cell* **63,** 1331-1338
51. Huth, J. R., Bewley, C. A., Nissen, M. S., Evans J. N., Reeves, R., Gronenborn, A. M. and Clore, G. M. (1997). The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat. Struct. Biol*. **4,** 657-665.
52. Jones, J. M. and Gellert, M. (2004) The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev.* **200,** 233-248
53. Kamtekar, S., Ho, R. S., Cocco, M. J., Li, W., Wenweiser, S. V., Boocock, M. R., Grindley, N. D. and Steitz, T. A. (2006). Implications of structures of synaptic tetramers of gamma delta resolvase for the mechanism of recombination. *Proc. Natl. Acad. Sci. USA.* **103,** 10642-10647
54. Kilbride, E., Boocock, M. R. and Stark, W. M. (1999) Topological selectivity of a hybrid site-specific recombination system with elements from Tn3 res/resolvase and bacteriophage P1 loxP/Cre. *J. Mol. Biol*. **289,**1219-1230
55. Kolb, A. F. (2002). Genome engineering using site-specific recombinases. *Cloning Stem Cells.* **4,** 65-80
56. Koo, H. S., Wu, H. M. and Crothers, D. M. (1986). DNA bending at adenine . thymine tracts. *Nature.* **320,** 501-506
57. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680-685
58. Laity, J. H., Dyson, H. J. and Wright, P. E. (2000 a). DNA-induced alpha-helix capping in conserved linker sequences is a determinant of binding affinity in Cys(2)-His(2) zinc fingers. *J. Mol. Biol.* **295,** 719-727
59. Laity, J. H., Dyson, H. J. and Wright, P. E. (2000 b). Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. *Proc. Natl. Acad. Sci. USA.* **97,** 11932-11935
60. Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58,** 913-949
61. Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D.F. and Steitz, T.A. (2005). Structure of a synaptic γδ resolvase tetramer covalently linked to two cleaved DNAs. *Science* **309**, 1210-1215.

62. McIlwrath, M.J., Boocock, M.R. and Stark, W.M. (1997). Tn3 resolvase catalyses multiple recombination events without intermediate rejoining of DNA ends. *J. Mol. Biol.* **266,** 108-121
63. Monahan, P. E., and Samulski, R. J. (2000). AAV vectors: Is clinical success on the horizon? *Gene Ther.* **7,** 24-30.
64. Mouw, K. W., Rowland., S. J., Gajjar, M. M., Boocock, M. R., Stark, W. M. and Rice, P. A. (2008). Architecture of a serine recombinase-DNA regulatory complex. *Mol. Cell*. **30,** 145-155
65. Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. *Genesis*. **26,** 99-109
66. Nash, H. A. (1996). Site-specific recombination: Integration, excision, resolution, and inversion of defined DNA segments. In "Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology" (F. C. Neidhart, R. Curtis, J. L. Ingraham, E. C. C. Lin, K. B. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds.), 2nd edn. pp. 2363–2376.American Society for Microbiology, Washington, DC.
67. Ng, P., and Baker M. D. (1999). Mechanism of double strand break repair during gene targeting in mammalian cells. *Genetics.* **151,** 1127-1141
68. Nollmann, M., He, J., Byron, O. and Stark, W. M. (2004). Solution structure of the Tn3 resolvase-crossover site synaptic complex. *Mol. Cell* **16,** 127-137.
69. Olorunniji, F. J. (2006) Tn3 resolvase-catalysed recombination: Assembly and activation of the site I synapse. PhD Thesis University of Glasgow
70. Olorunniji, F. J., He, J., Wenweiser, S. V., Boocock, M. R. and Stark W.M. (2008). Synapsis and catalysis by activated Tn3 resolvase mutants. *Nucleic Acids Res.* **36,** 7181-7191
71. Paques, F. and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cervisiae. Microbiol. Mol. Biol. Rev.* **63,** 349-404
72. Pavletich, N. P., and Pabo, C. O. (1991). Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252,** 809–817.
73. Perez, E. E., Wang, J., Miller, J. C., Jouvenot, Y., Kim, K. A., Liu, O., Wang, N., Lee, G., Bartsevich, V. V., Lee, Y. L., Guschin, D. Y., Rupniewski, I., Waite, A. J., Carpenito, C., Carroll, R. G., Orange, J. S., Urnov, F. D., Rebar, E. J., Ando, D., Gregory, P. D., Riley, J. L., Holmes, M. C. and June, C. H. (2008). Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26,** 808-816.
74. Pfeifer, A., and Verma, I. M. (2001). Gene therapy: Promises and problems. *Ann. Rev. Genomics Hum. Genet.* **2,** 177-211.
75. Pribnow, D. (1975). Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. USA.* **72,** 784-788
76. Rebar, E. J., Griesman, H. A., and Pabo, C. O. (1996). Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. *Methos Enzymol.* **267,** 129-149
77. Recommendations 1984 (1985) Nomenclature for incompletely specified bases in nucleic acid sequences. *Biochem J.***229,** 281-286
78. Reed, R. R. (1981). Transposon-mediated site-specific recombination: a defined in vitro system. *Cell* **25,** 713-719
79. Reed, R. R. and Grindley, N. D. (1981). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell.* **25,** 721-728
80. Reed, R. R. and Moser, C. D. (1984). Resolvase-mediated recombination intermediates contain a serine residue covalently linked to DNA. *Cold. Spring. Harb. Symp. Quant. Biol.* **49,** 245-249
81. Rice, P. A. and Steitz, T. A. (1994). Refinement of gamma delta resolvase reveals a strikingly flexible molecule. *Structure* **2,** 371-384

82. Rimphanitchayakit, V. and Grindley, N. D. F. (1990). Saturation mutagenesis of the DNA site bound by the small carboxy-terminal domain of gamma delta resolvase. *EMBO J.* **9,** 719-725

83. Rowland, S. J. and Stark, W. M. (2005). Site-specific recombination by the serine recombinases. In: *The dynamic bacterial genome,* Mullany, P. (ed.) Cambridge University Press, pp. 121-150.

84. Rowland, S. J., Boocock, M. R. and Stark, W. M. (2005). Regulation of Sin recombinase by accessory proteins. *Mol. Microbiol.* **56,** 371–382

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

86. San Filippo, J., Sung, P. and Klein, H. (2008). Mechanism of Eukaryotic Homologous Recombination. *Annu. Rev. Biochem.* **77,** 229-257.

87. Sarkis, G. J., Murley, L. L., Leschziner, A. E., Boocock, M. R., Stark, W.M. and Grindley,N. D. (2001). A model for the γδ resolvase synaptic complex. *Mol. Cell.* **8,** 623-631

88. Sauer, B. (1998). Inducible gene targeting in mice using the Cre/lox system. *Methods: a Companion Methods Enzymol.* **14,** 381-392

89. Schneider, T. D. and Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097-6100

90. Schwikardi, M. and Droge, P. (2000). Site-specific recombination in mammalian cells catalysed by γδ resolvase mutants: implications for the topology of episomal DNA. *FEBS Letters* **471,** 147-150

91. Seeman, N. C., Rosenberg, J. M. and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA.* **73,** 804-808

92. Sherratt, D. J. (1989). Tn3 and related transposable elements: site-specific recombination and transposition. In Berg, D. E., Howe, M. (eds) *Mobile DNA.* American Society for Microbiology, Washinghton, DC pp. 163-184

93. Shimizu, Y., Bhakta, M. S. and Segal, D. J. (2009). Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg. Med. Chem. Lett.* Mar. 3 E-pub ahead of print

94. Siefert, J. L (2009). Defining the mobilome. *Methods Mol Biol.* **532,** 13-27

95. Smith, M.C. and Thorpe, H. M. (2002). Diversity in the serine recombinases. *Mol. Microbiol.* **44,** 299-307

96. Sohn, Y and Kang, C. (2005). Sequential multiple functions of the conserved sequence in sequence-specific termination by T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA.* **102,** 75-80.

97. Sorrell, D. A. and Kolb, A. F. (2005). Targeted modification of mammalian genomes. *Biotech. Adv.* **23,** 431-469.

98. Stark, W.M. and Boocock, M.R. (1994). The linkage change of a knotting reaction catalysed by Tn3 resolvase. *J. Mol. Biol.* **239,** 25-36.

99. Stark, W.M. and Boocock, M.R. (1995). Topological selectivity in site-specific recombination. In *Mobile Genetic Elements,* Sherratt, D.J. (ed.), pp 101-129, Oxford University Press.

100. Stark, W.M., Boocock M.R. and Sherratt, D.J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8,** 432-439.

101. Stark, W.M., Sherratt, D.J. and Boocock, M.R.(1989). Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reaction. *Cell.* **58,** 779-790

102. Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods. Enzymol.* **185,** 60-89

103. Summers, D. K. and Sherratt, D. J. (1988). Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of Xer site. *EMBO J.* **7,** 851-858

104. Taylor, L. A. and Rose, R. E. (1982). A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. *Nucleic Acids Res.* **16,** 358

105. Van Duyne, G. D. (2002). A structural view of tyrosine recombinase site-specific recombination. In *Mobile DNA II* (N. L. Craig, R. Craigie, M. Gellert, and A.M. Lambowitz, eds.), pp. 93-117. ASM Press, Washington, DC.

106. Voziyanov, Y., Konieczka, J. H., Stewart, A. F. and Jayaram, M. (2003). Stepwise manipulation of DNA specificity in FLP recombinase: Progressively adapting FLP to individual and combinatorial in its target site. *J. Mol. Biol.* **326,** 65-76

107. Voziyanov, Y., Stewart, A. F. and Jayaram, M. (2002). A dual reporter screening system identifies the amino acid at position 82 in FLP site-specific recombinase as a determinant for target specificity. *Nucleic Acids Res.* **30,** 1656-1663

108. Weiss, M. A., Ellenberger, T. E., Wobbe, C. R., Lee, J. P., Harrison, S. C. and Struhl, K (1990). Folding transition in the Dna-binding domain of GCN4 on specific binding to DNA. *Nature* **347,** 575-578

109. Wharton, R. P and Ptashne, M. (1985). Changing the binding specificity of the repressor by redesigning α-helix. *Nature.* **316,** 601-605

110. Whiteson, K. L. and Rice, P. A. (2008). Binding and catalytic contributions to site recognition by flp recombinase. *J. Biol. Chem.* **283,** 11414-11423.

111. Wu, J., Kandavelou, K. and Chandrasegaran, S. (2007). Custom-designed zinc finger nucleases: what is next? *Cell. Mol. Life Sci.* **64,** 2933-2944.

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

113. Zhou, H. X. (2003). Quantitative account of the enhanced affinity of two linked scFvs specific for different epitopes on the same antigen. *J. Mol. Biol.* **329,** 1-8