Engineering the Zinc Finger Recombinase for use in Targeted Genomic Editing

Jeffrey Louis Kentner

Honours Bachelor of Science

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Institute of Molecular, Cell and Systems Biology

College of Medical, Veterinary and Life Sciences

University of Glasgow

November 2015

Abstract

The zinc finger recombinase (ZFR) is a chimeric enzyme system for use in targeted genomic editing. The ZFR is comprised of a recombinase catalytic domain, which is able to catalyse recombination reactions between DNA molecules, and a zinc finger array DNA-binding domain, which is able to target the enzyme to a desired genetic sequence. Currently the ZFR is in an early stage of development and will require several crucial improvements before it can be adopted as a useful genome editing tool by researchers. Two major challenges involve two important parameters of ZFR-catalysed integration reactions: specificity of the orientation of the integration, and stability of the integration. Currently, the ZFR system is unable to select the orientation of integrations, and because the recombination reactions are reversible, the integrations are not stable (and, in fact, are stochastically disfavoured). This project aimed to impart the ZFR system with the ability to perform stable, orientation-specific integrations.

In order to achieve the project aims, the experiments of this project sought to generate new pairs of ZFR mutants that were differentially modified at either their catalytic domain or DNA-binding domain, and characterize their behaviour in an *E. coli*-based recombination assay. The overall strategy was to exploit the interactions within the protein-protein interfaces of the ZFR tetramer to produce selective compatibility, and to generate differences in enzyme activity when two mutants (one active, and one inactive or less active) were either paired as heterodimers or as homodimers. During ZFR recombination reactions heterodimers rearrange to form homodimers, and thus the production of a significant difference in activity between heterodimers and homodimers represents a recombination reaction directionality bias. Both the catalytic domain modification and DNA-binding domain modification approaches proved able to produce the desired bias in the directionality of ZFR recombination reactions, which is predicted to lead to both stability of an integration, and specificity of integration orientation through a stochastic process. Of particular note, was a strategy utilizing a heterodimer that consisted of one ZFR subunit targeted specifically to the DNA, paired with a recombinase subunit (with its native DNA-binding domain) targeted non-specifically to the DNA. The difference in activity between these subunits paired in heterodimer and homodimer configurations appeared to produce a completely irreversible recombination reaction

without any apparent reduction in recombination reaction efficiency. Furthermore, the results of the catalytic domain modification and DNA-binding domain modification experiments suggest that it should be possible to generate a combination strategy in which the recombinase subunit with the native DNA-binding domain is catalytically inactive unless operating within an intended heterodimer, overcoming the potential problem of unwanted off-target activity from homodimers of this subunit.

The success of this work in producing ZFR reactions with the potential to catalyse stable, orientation-specific integration reactions potentially represents a major leap forward in ZFR research; however, these results must be further validated in a mammalian cell system. Although genome editing systems such as the CRISPR-Cas9 RNA-guided endonuclease now allow researchers to modify genomes within embryos and cell culture with ease, off-target effects, reliance on endogenous homology directed repair (HDR) activity, unfavourable ratios of HDR to non-homologous end joining (NHEJ) activity at the target site, and low efficiency make targeted endonuclease technology impractical for use in *in vivo* gene therapy applications and genome editing in some cell types (e.g. non-dividing cells such as neurons and myocytes). Therefore, the ZFR is envisioned as a genome editing tool that can fill this vacant niche for gene editing in non-dividing cell types and human *in vivo* gene therapy.

Table of contents

List of tables

List of figures

Accompanying material

The attached DVD contains sequences relevant to the work carried out in this project, these include: DNA sequences for zinc finger domains and the structural elements used to construct them, Z-sites, expression plasmids, catalytic domains (Tn3[NY], Tn3[NM], and Tn3[G70C]), and an amino acid sequence for a zinc finger recombinase (Tn3[NY]-Z3).

Acknowledgement

This thesis is dedicated to

Rachel, for your companionship, patience, support, and love throughout this adventure,

my parents for your guidance, love, support, and faith in me throughout my life,

and my late granddad, Ormond Kentner, for the fine example of curiosity and careful intelligence you set for me to follow.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature _____

Printed name _____

Abbreviations

2MutHomDim	two Mutant HomoDimer-binding sites
2MutHetDim-DR	two Mutant HeteroDimer-binding sites in Direct Repeat
2MutHetDim-IR	two Mutant HeteroDimer-binding sites in Direct Inverted
AAV	Adeno-Associated Viruses
AHD	Alternative Homodimers Dimer-dimer (configuration)
Amp	Ampicillin
B2H	Bacterial-2-Hybrid
bp	base pairs
Chl	Chloramphenicol
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
ddH_2O	double distilled water
DEL	DELetion (excision) product
DEV	Double Expression Vector
DSB	Double-Strand Break
DNA	DeoxyriboNucleic Acid
ELISA	Enzyme-Linked Immunosorbent Assay
ESC	Embryonic Stem Cell
eZFR	enhanced ZFR
F1	Finger 1
F2	Finger 2
F3	Finger 3
h	hours
HDR	Homology-Directed Repair
HIV	Human Immunodeficiency Virus
HSV	Herpes Simplex Virus
HTH	Helix-Turn-Helix
IDLV	Integrase-Deficient Lentiviral Vector
INV	INVersion product
ILV	Integrating Lentiviral Vector
iPSC	induced Pluripotent Stem Cell

ITR	Inverted Terminal Repeats
Kan	Kanamycin
kb	kilobases
M	G101S + D102Y + M103I + Q105L
M	Molar
mRNA	messenger RNA
ml	millilitre
MTR	Mutant Tn3 Resolvase
N	R2A + E56K
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
MHD	Mirrored Heterodimers Dimer-dimer (configuration)
min	minutes
neo	neomycin
nm	nanometres
nt	nucleotides
OD_{600}	Optical Density at 600 nm
OPEN	Oligomerized Pool ENgineering
PAM	Protospacer-Adjacent Motif
PCR	Polymerase Chain Reaction
PNA	Peptide Nucleic Acid
SDF	Small DNA-Fragment
SFHR	Small Fragment Homologous Replacement
ssODN	single-stranded Oligo-DeoxyriboNucleotide
RABD	Reduced Affinity Binding Domain
RGEN	RNA-Guided Endonucleases
RHD	Reversed Heterodimers Dimer-dimer (configuration)
RNA	RiboNucleic Acid
RVD	Repeat-Variable Diresidue
s	seconds
SA	Signal Attenuation
SCID	Severe Combined ImmunoDeficiency
SEV	Single Expression Vector

SDF	Small DNA-Fragment
sgRNA	single guide RNA
SSB	Single-Strand Break
TALE	Transcription Activator-Like Effectors
TALEN	TALE Nuclease
TALER	TALE Recombinase
TFO	Triplex-Forming Oligonucleotide
TGR	Targeted Gene Repair
tk	thymidine kinase
topo I	topoisomerases I
tracrRNA	*trans*-activating CRISPR RNA
tru-sgRNA	truncated sgRNA
UNC	UNChanged substrate
Y	D102Y
ZFA	Zinc Finger Array
ZFC	Zinc Finger Consortium
ZFN	Zinc Finger Nuclease
ZFR	Zinc Finger Recombinase
µl	microlitre

Chapter 1: *Introduction*

1.1 The dream of targeted genomic sequence editing

Coding the instructions for all the processes, in all of life's organisms, is DNA. Structural materials and enzymes, how they are assembled and modified, and the immense regulatory networks in which they are contained, are all guided by DNA instructions. That all of this reduces down to a single type of code contained within a single type of molecule is profound; more profound, is the idea that we might be able to intelligently alter this code at will. The creation of powerful new enzymes with engineered programmable DNA-binding domains, such as chimeric endonucleases and chimeric recombinases have, for the first time, allowed humans to easily perform a variety of genetic sequence manipulations, including targeted sequence replacement. Thus, the nascent field of targeted genomic sequence editing is beginning to open new doors in the fields of experimental genetics, genetic engineering, and gene therapy. However, this technology is far from perfected, and many challenges must be overcome before it can achieve its full potential. Because the applications of targeted genomic sequence editing technology are vast, this chapter will primarily focus on its development and use as it relates to the modification of human cells and gene therapy. The realization of targeted genomic sequence editing represents an unparalleled opportunity to investigate and decipher the instructions of life, alter the organisms in the world around us, and better the human organism itself.

1.2 Current gene therapy

1.2.1 Gene therapy background

Gene therapy is a medical approach that aims to deliver genes with therapeutic effects in order to treat a wide range of pathologies including cancer, genetic disorders, and heart disease. Currently, over 60% of the ongoing gene therapy clinical trials worldwide are focused on the treatment of cancer (Wirth et al., 2013). Although over 1800 gene therapy clinical trials have been conducted or are ongoing, few gene therapy products have actually made it to market (Wirth et al., 2013). The first gene therapy ever to be approved for use was Gendicine® (SiBiono GeneTech Co., Ltd; Shenzhen, China), which was approved in China, in 2003, for treating head and neck squamous cell

carcinomas (Wilson, 2005, Pearson et al., 2004). Gendicine® is based on a non-replicative adenoviral vector and therapeutically expresses functional p53 tumour suppressor in tumour cells. Following Gendicine®, in 2005, another cancer gene therapy, Oncorine® (Sunway Biotech Co., Ltd; Shanghai, China), was also approved for use in China (Wirth et al., 2013). Oncorine® is based on a conditionally replicative adenovirus that contains a deletion in the viral genome allowing the virus to replicate only in p53-deficient cells. In 2012, Glybera® (UniQure; Amsterdam, Netherlands), a gene therapy designed to treat familial lipoprotein lipase deficiency, became the first to be recommended for approval in the EU (Buning, 2013, Yla-Herttuala, 2012). Glybera® uses an adeno-associated viral vector to express lipoprotein lipase from within muscle tissue. Incidentally, Glybera® will be the most expensive treatment in the world, with an expected price tag of €1.1 million for a one-time treatment that is expected to be permanently corrective (Yla-Herttuala, 2015).

1.2.2 Gene therapy characteristics

Gene therapy may involve either *in vivo* strategies, where therapeutic genes are directly introduced into cells within the body, or *ex vivo* strategies, where cells extracted from the body are reintroduced after treatment with gene therapy vectors. The simplest form of *in vivo* delivery involves direct injection of the gene therapy vector into the target tissue, and this is the method used for the Gendicine®, Oncorine® and Glybera® therapies mentioned above. More complex *in vivo* gene therapy delivery strategies involve systemic administration with the use of vectors that are capable of tissue targeting. Although many gene therapies utilizing systemically administered tissue targeting vectors or *ex vivo* strategies have entered trials, none have yet been approved (Kaufmann et al., 2013). The efficacy of a gene therapy vector will be based on several factors including: targeting to, and internalization within, the target cell type; trafficking within the cell; protection from degradation; nuclear localization; gene expression levels; and stability and persistence of the transfected DNA or transfected cells (Douglas, 2008, Gardlik et al., 2005).

Currently, gene therapy approaches can be broadly categorized into those mediated by viral vectors (such as retroviral, adenoviral, and adeno-associated vectors) and those mediated by non-viral vectors (such as liposomes, charged polymers, lipid polymer combinations, and peptide-based) (Douglas, 2008). Viral vectors often consist of

replication-deficient viruses with recombinant genomes that express a therapeutic gene. Viral vectors have the advantage of introducing DNA to target cells with higher efficiency and produce higher levels of gene expression than non-viral vectors (Gardlik et al., 2005). The drawbacks to using viral vectors can include: possible reactivation of replication competence, pre-existing antibody response from past viral infections, and, in some cases, transient expression (Waehler et al., 2007, Gardlik et al., 2005). Non-viral vectors consist of artificial vesicle-like structures that carry an expression plasmid. Non-viral vectors have the advantage of being less immunogenic than viral vectors, but are significantly limited in nuclear translocation and by transient expression (Douglas, 2008). Another important distinction among gene therapy vectors, which impacts gene expression and stability, is between those that integrate their genetic material into the genome and those that do not. Integrating vectors such as retroviral vectors maintain better gene expression and persistence over time, but suffer the disadvantage that their integration is random, and thus can be oncogenic (Sadelain, 2004).

1.2.3 Targeted genome editing enzymes for gene therapy

Recently, enzymes that mediate targeted genomic sequence editing have presented themselves as a more elegant modality for gene therapy. Targeted genome editing enzymes have the theoretical advantage that they could be used to produce gene insertions with stable, long term gene expression like retroviruses do, but with a reduced risk of random mutagenesis (although this depends largely on the fidelity of the gene editing technology). In addition, these enzymes could be used to perform gene correction, rather than randomly introducing a functional copy of a defective gene into the genome. The gene correction approach would open up the possibility of treating disorders caused by the presence of a dominant negative allele (e.g. Huntington's disease). Gene correction also provides an advantage over current gene therapy modalities in that corrected genes will be expressed in their endogenous genomic context, providing the naturally intended level of expression and regulation for the gene.

Unlike viral vectors, targeted genome editing enzymes do not come with their own cell-delivery system; however, the currently available gene therapy vectors provide options for this delivery (see Section 1.4.5). The DNA molecules encoding targeted genome editing enzymes are essentially therapeutic genes, so, in a sense, this modality

simply represents an enhancement to the cargo options that gene therapy vectors currently carry. The following sections will discuss the development of targeted genomic editing technology while referencing its relationship in gene therapy.

1.3 A brief history of targeted genomic sequence editing

1.3.1 The development of gene targeting

1.3.1.1 Insertion and replacement vectors

The first report of a targeted genomic sequence alteration in mammalian cells was published in the mid eighties by Smithies et al. (1985) (Smithies et al., 1985). Smithies' approach involved transfecting Hu 11 cells (a hybrid murine erythroleukemia cell line containing human chromosome 11) with a plasmid vector containing extensive homology to the β-globin locus, in order to integrate the plasmid at that locus through the endogenous homologous recombination activity of the cell. Because this approach integrated the entire plasmid, it was not gene editing in the sense of a discrete replacement of genetic sequence. However, this proof-of-concept based on homologous recombination, was built upon extensively throughout the late eighties and nineties to allow researchers to perform site-specific deletions, insertions, and sequence replacements of DNA at desired genomic loci, becoming known as 'gene targeting'.

The available suite of gene targeting techniques allow researchers to effectively *disrupt* genes using targeted 'insertion vectors' that rely on a single homologous recombination cross-over event (such as in Smithies' experiment), or by using targeted 'replacement vectors' that rely on two cross-over events in order to effect a cassette replacement (Figure 1-1 A and B, respectively) (Muller, 1999). However, because these recombination events happen at a very low frequency, the use of both insertion vectors and replacement vectors typically involves the permanent introduction of selectable markers (e.g. neomycin resistance) in order to make the techniques practicable.

1.3.1.2 Hit-and-run and tag-and-exchange

Subtle genetic code changes (such as a point mutation) that do not permanently introduce a selectable marker can be accomplished by gene targeting techniques called 'hit-and-run' and 'tag-and-exchange'. The hit-and-run (also known as 'in-and-out') and tag-and-exchange techniques are based on the use of an insertion vector or replacement

vector, respectively, but utilize the transient introduction of a negative selectable marker (e.g. the HSV thymidine kinase gene in conjunction with fialuridine or gancyclovir) and a positive selectable marker (e.g. neomycin).

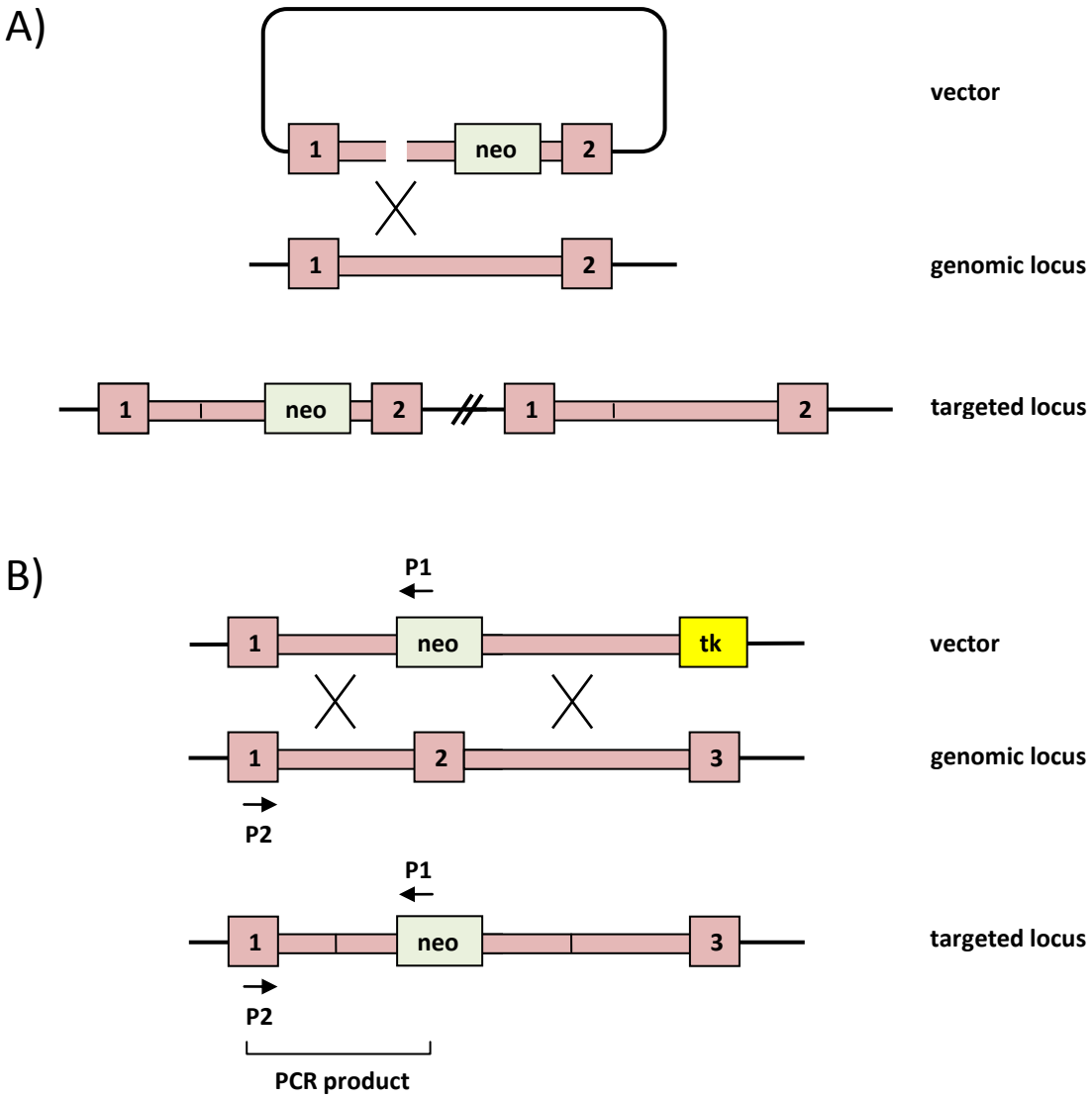


Figure 1-1: Gene targeting strategies using an insertion vector and a replacement vector. A) Insertion vector gene targeting. A vector with homology (indicated by pink) between exon 1 and 2, and carrying a neomycin resistance gene (neo) in order to allow for positive selection, is shown. The break indicates that the vector has been linearized in the region of homology, which increases the efficiency of the homologous recombination. Homologous recombination happens at the crossover site indicated by the crossed lines. The product 'targeted' locus contains some duplication of the target sequence due to total integration of the plasmid vector. **B)** Replacement vector gene targeting. The vector contains a neomycin resistance marker flanked by two regions of homology to the genomic locus. Outside the region of homology the vector also contains an HSV thymidine kinase gene (tk), which acts as a conditional negative selective maker against improper vector integration (e.g. when only one cross-over event takes place) when used in conjunction with fialuridine or gancyclovir. Homologous recombination happens at two crossover sites. In the product targeted locus, exon 2 has been replaced with the neo cassette. PCR primers (P1 and P2) allow for confirmation of correct cassette replacement. Note that the inclusion of exons in the figure is primarily to help with visual orientation. Adapted from Müller (1999).

In the hit-and-run approach (Figure 1-2), the insertion vector contains a genetic sequence that is homologous to the target site but contains a desired mutation, as well as both positive and negative selectable markers within the vector backbone (Muller, 1999). Targeted integration of the insertion vector results in a direct repeat of the target sequence, where one repeat contains the desired mutation and the repeats are separated by the vector backbone containing the selectable markers. After a positive selection step has identified cells that have integrated the insertion vector, a subsequent intrachromosomal homologous recombination event, which deletes the unwanted sequence that includes the two selectable markers, can be detected by a negative selection step. The result is that only the desired sequence replacement remains in the genomes of cells that pass the second selection step.

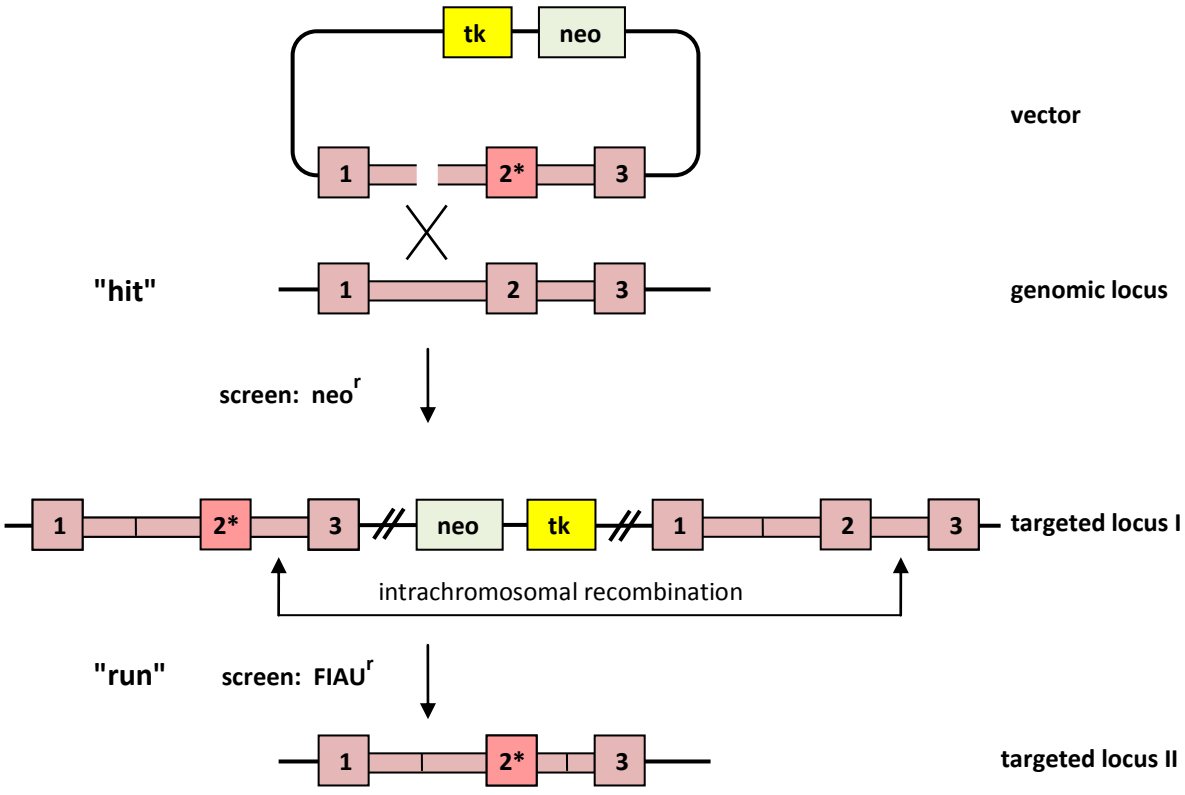


Figure 1-2: Hit-and-run gene targeting. The figure displays the hit-and-run gene targeting approach being used to replace exon 2 at a genomic locus with a variant of exon 2 containing a point mutation. The insertion vector contains an exon 2 variant with a point mutation (2*) flanked by sequences homologous to the genomic locus (pink), as well as genes for HSV thymidine kinase (tk) and neomycin resistance (neo) in the vector backbone. In the 'hit' stage, a homologous recombination crossover event introduces the entire plasmid into the genomic locus. Screening for successful integration is carried out with neomycin. In the 'run' stage, a subsequent intrachromosomal recombination event between the repeat regions removes the repeat, as well as the neomycin and thymidine kinase genes. Selection with fialuridine (FIAU) is used to detect the recombination event. It should be noted, that depending on the location of the intrachromosomal crossover event, the wild-type allele may also be restored (not shown). Adapted from Müller (1999).

In the tag-and-exchange approach (Figure 1-3), a replacement vector is first used to facilitate a cassette exchange that replaces the target endogenous sequence with only the positive and negative selectable markers (Muller, 1999). After positive selection has identified cells that have undergone this cassette exchange, a second replacement vector is used to replace the positive and negative selectable marker cassette with the desired replacement sequence via another cassette replacement event, which can be detected by negative selection. The result is that only the desired sequence replacement remains in the genomes of cells that pass the second selection step. Although the hit-and-run and tag-and-exchange methods are able to effectively 'edit' genomic sequences for cell culture experiments or for genetically engineering murine embryos, the positive and negative selection steps make these techniques inapplicable to gene therapy, and these techniques are labour-intensive.

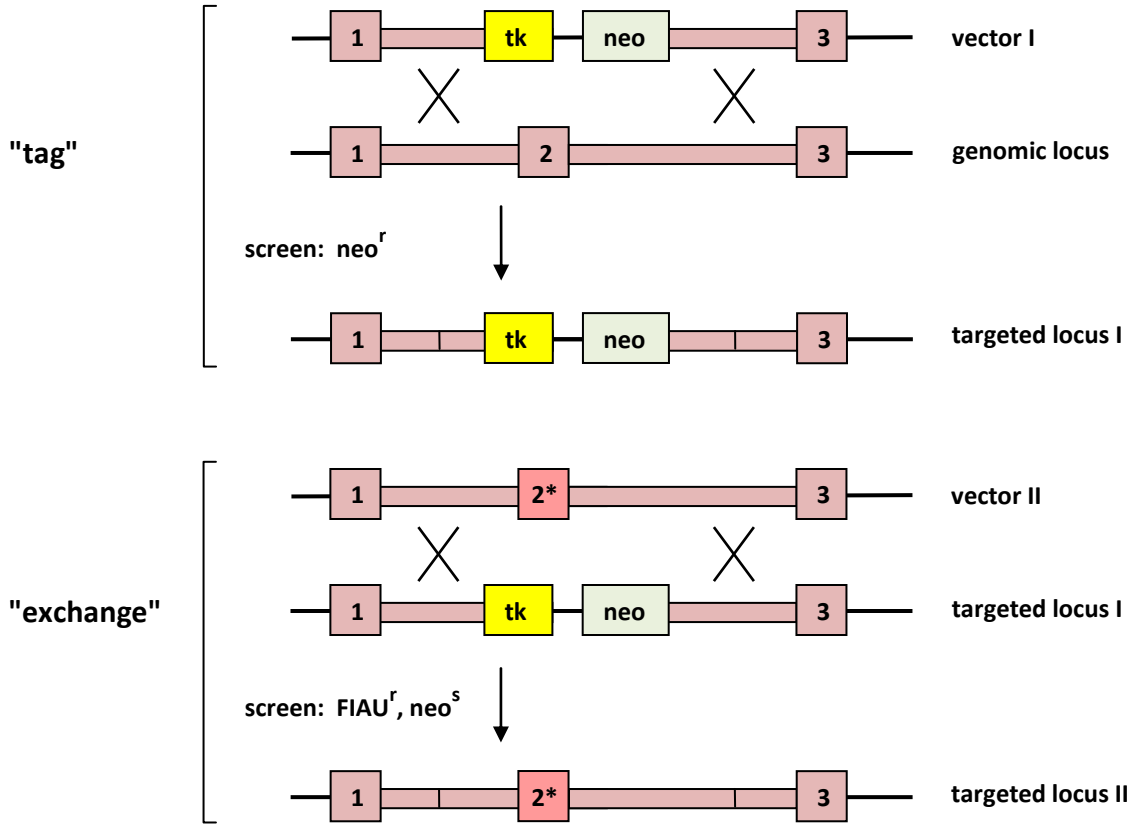


Figure 1-3: Tag-and-exchange gene targeting. The figure depicts tag-and-exchange being used to replace exon 2 at a genomic locus, with a variant of exon 2 containing a point mutation. In the 'tag' stage, the replacement vector contains thymidine kinase (tk) and neomycin (neo) genes flanked by regions of homology to the genomic locus (pink). Two homologous recombination crossover events replace exon 2 with a thymidine kinase and neomycin resistance cassette. Screening for successful cassette replacement is carried out using neomycin. In the 'exchange' stage, a second vector is used to replace the kinase and neomycin resistance gene with a variant of exon 2 containing a point mutation (2*) via the same type of homologous recombination-mediated cassette exchange as in the previous stage. Screening for successful cassette replacement is carried out with fialuridine (FIAU). Adapted from Müller (1999).

1.3.1.3 Site-specific recombinase-enhanced gene targeting

Site-specific recombinases, such as Cre recombinase, FLP recombinase, and ΦC31 integrase, can also be used for gene targeting using a strategy that is similar to hit-and-run and tag-and-exchange, but with greatly increased efficiency (Muller, 1999). These site-specific recombinases act on specific target sequences to produce a DNA recombination event, with higher efficiency than that of endogenous intracellular homologous recombination. For example, the Cre recombinase acts on two 34 bp targets called *loxP* sites (Figure 1-4 A), and will either excise or invert an intervening sequence sandwiched between *loxP* sites depending on whether the sites are arranged in direct or inverted repeat, respectively (Figure 1-4 B).

In a Cre recombinase-based gene targeting strategy (Figure 1-4 C), a replacement-type plasmid vector is used to facilitate a cassette exchange, which replaces the target DNA sequence with a sequence containing the desired mutation as well as positive and negative selectable markers (Muller, 1999). Additionally, the selectable markers are flanked by a direct repeat of *loxP* sites. After detection of the initial cassette replacement by positive selection, transient transfection with a Cre recombinase expressing plasmid causes the positive and negative selectable markers to be excised with high efficiency. Negative selection can then be used to detect this excision event, which will take place with greatly improved efficiency compared to the homologous recombination excision that the previously described tag-and-exchange approach relies on at this stage.

However, Cre recombinase-based gene targeting does not result in a discrete DNA sequence edit because it leaves behind a genetic 'scar' consisting of one *loxP* site. Thus, although gene targeting strategies involving natural site-specific recombinases have a performance improvement due to the increased efficiency of the second step, they are still not applicable for some applications, including gene therapy, due to their use of selection steps and the sequence scar that is left behind.

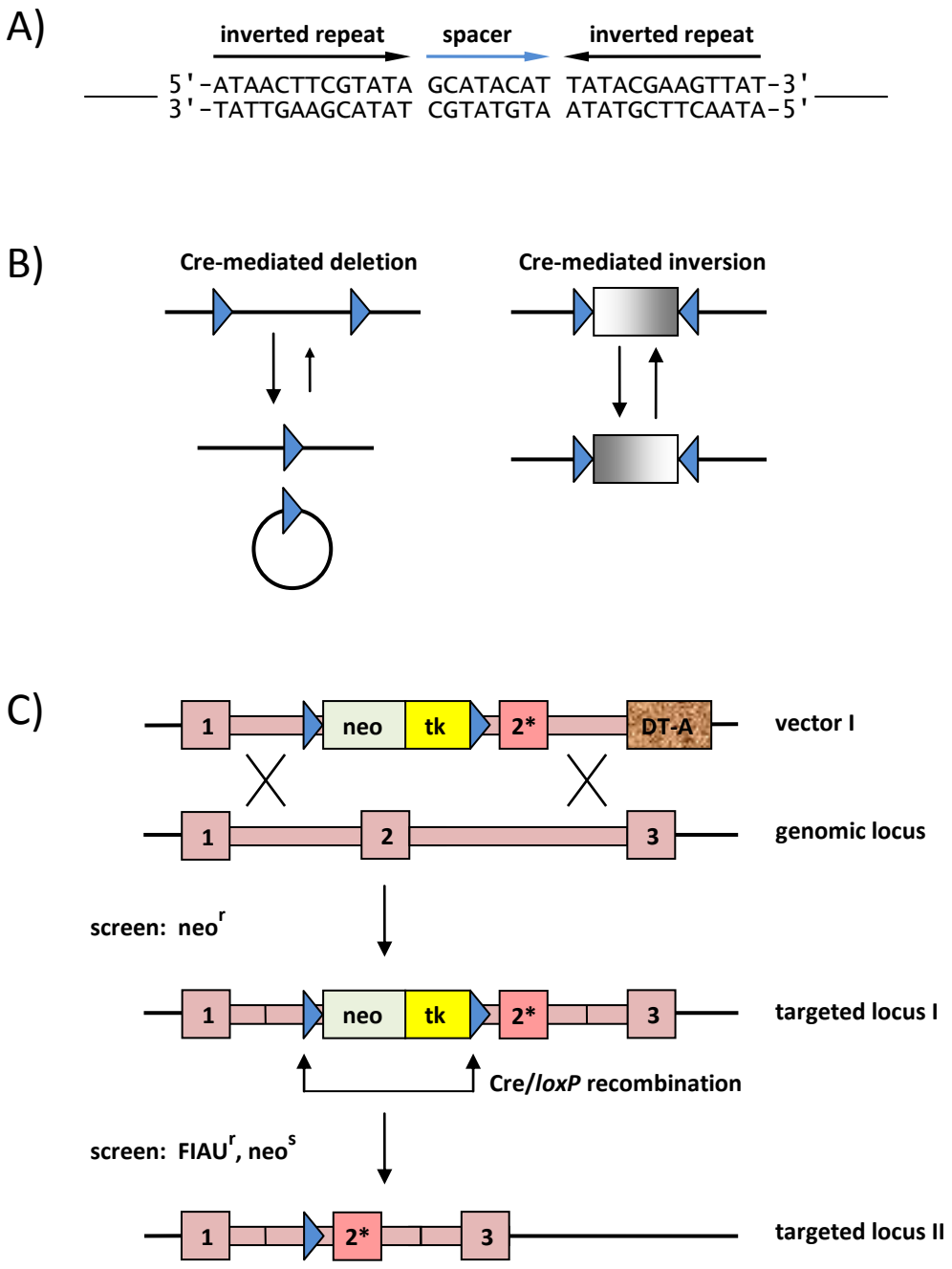


Figure 1-4: Cre/*loxP* system in gene targeting. A) Structure of a *loxP* site. The Cre recombinase binds two 13-bp inverted repeats separated by an 8-bp spacer sequence (blue arrow). The spacer sequence determines the orientation of the *loxP* site. **B)** Cre recombinase-mediated recombination reactions. When *loxP* sites (blue triangles) are arranged in direct repeat, Cre recombinase will mediate an excision reaction where the intervening sequence is circularized. One *loxP* site will remain in both the substrate molecule and the circularized excision product. When *loxP* sites are arranged in inverted repeat, Cre recombinase will mediate an inversion reaction where the intervening sequence becomes reversed in orientation. **C)** Cre/*loxP*-mediated gene targeting. In the first stage a replacement vector is used which contains a cassette with genes for neomycin resistance (neo) and HSV thymidine kinase (tk) flanked by *loxP* sites, as well as an exon 2 variant with a point mutation (2*). This cassette is further flanked by regions of homology (pink) to a genomic locus. Outside of these regions of homology, the replacement vector also contains a gene coding for diphtheria toxin A-fragment (DT-A), which kills cells that undergo improper integration of the vector (e.g. when only a single crossover event takes place). Homologous recombination at two crossover sites replaces exon 2 in the genomic locus with the cassette. Screening with neomycin confirms cassette replacement. In the second stage, Cre recombinase is introduced by transient transfection, causing the neomycin resistance and thymidine kinase genes to be excised, leaving only the exon 2 variant with the point mutation and one *loxP* site. Screening with FIAU confirms Cre recombinase-mediated excision. Introns are not necessarily truncated in this process as shown. Adapted from Müller (1999).

1.3.2 'Classical' targeted gene repair

During the mid 1990s through to the mid 2000s another method for performing simple gene edits was developed in the form of 'targeted gene repair'(TGR). The early TGR strategies involved the introduction of short single- or double-stranded oligonucleotides that form heteroduplexes with target DNA and facilitate small sequence alterations (usually a point mutation) via the activity of the various endogenous DNA repair pathways. The TGR strategies have an advantage over the previously discussed gene targeting strategies in that they do not involve the introduction of selectable markers and are thus more suitable for gene therapy applications. Screening for a successful chromosomal sequence change can be accomplished by PCR assay or commercial gene sequencing. However, the frequency of successful chromosomal sequence change using the early targeted gene repair strategies is low enough to limit their application for *in vivo* gene therapy.

Four noteworthy targeted gene repair methods that will be discussed below involve the use of: single-stranded oligo-deoxyribonucleotides (ssODNs), small DNA-fragments (SDFs), triplex-forming oligonucleotides (TFOs), and peptide nucleic acids (PNAs). The specific interactions with endogenous DNA repair pathways involved in each of these TGR approaches are incompletely understood, but a detailed review of what is currently known about them has been published by Jensen et al. (2011) (Jensen et al., 2011). It should be noted that the currently popular programmable site-specific nuclease gene editing techniques, which rely on homologous repair, are also an important TGR strategy, but since they require a more in-depth discussion, they will be covered separately in a Section 1.4.

Single-stranded oligo-deoxyribonucleotides (ssODNs) can be use in targeted gene repair applications where a single base change is desired. Typically, ssODNs consist of several dozen nucleotides with complementarity to a target locus, but containing one mismatch (the desired change) in the centre of the sequence (Aarts and te Riele, 2011, Jensen et al., 2011). Strand invasion leads to the formation of a heteroduplex between the ssODNs and target DNA. The nucleotide excision repair pathway may then act on the mismatch-containing heteroduplex to cause the desired base substitution in the target chromosomal sequence (although, alternative mechanisms involving base excision repair,

replication, transcription, and homology-dependent repair have also been proposed) (Jensen et al., 2011, Aarts and te Riele, 2011). Using the ssODNs approach, gene correction efficiencies of 0.1–5% have been achieved in somatic cells, while efficiencies of approximately 0.1% have been observed in embryonic stem cells (ESCs) (Jensen et al., 2011).

Small DNA-fragments (SDFs) are another early TGR vector, which have the capacity to change the identity of up to four sequential base pairs of a target genomic sequence (Jensen et al., 2011). SDFs typically consist of 400–1000 bp of single- or double-stranded DNA, with homology to the target sequence except for the desired base pair changes that are contained within it. The SDF method is known to exploit the small fragment homologous replacement (SFHR) process within eukaryotic cells in order to effect replacement of the endogenous target DNA with the exogenous SDFs; however, the details of the SFHR process itself are not well understood. SDFs have been used to achieve 0.2–20% gene correction efficiencies in somatic cells, and 0.025% correction efficiencies in ESCs (Jensen et al., 2011).

Triplex-forming oligonucleotides (TFOs) can also be used for TGR. TFOs are short oligonucleotides (RNA, DNA, and synthetic derivatives) of 10–50 bp that bind in the major groove of homopurine tracts of DNA in a triplex fashion via Hoogsteen bonds (Jensen et al., 2011). TFOs were originally envisioned as useful for antigene strategies that interfere with transcription (Giovannangeli and Helene, 1997, Maher, 1996, Praseuth et al., 1999, Vasquez and Wilson, 1998), and targeted mutagenesis by tethering them to mutagens such as psoralen (Wang et al., 1995). However, it was later found that TFO binding, itself, could stimulate DNA repair and homologous recombination as an intrinsic property of triplex formation itself (Wang et al., 1996). The discovery that TFOs could promote DNA repair and homologous recombination led to TGR strategies involving so called bi-functional TFOs. A bi-functional TFO is a TFO linked to a segment of double- or single-stranded DNA with homology to a target sequence, but including one to a few desired base changes (Richardson et al., 2002, Knauert and Glazer, 2001, Culver et al., 1999, Knauert et al., 2006). The DNA repair pathways involved in TFO-mediated TGR are not fully understood but are believed to involve nucleotide excision repair (NER) and homology-directed repair (HDR), but may also involve mismatch repair and non-

homologous end joining (NHEJ) (Jensen et al., 2011, Ricciardi et al., 2014, Richardson et al., 2002, Knauert and Glazer, 2001).

A more recent enhancement to the TFOs strategy is the use of peptide nucleic acids (PNAs). PNAs are synthetic nucleic acid analogues (typically with a length of 12–18 nucleotides) that have had the charged phosphodiester DNA backbone substituted for an uncharged polyamide backbone (Nielsen, 2004). PNAs are nuclease- and protease-resistant, which greatly increases the stability of the molecule. Additionally, some varieties of PNAs allow for additional binding modalities (e.g. a double PNA/DNA duplex, or a PNA/DNA/PNA triplex, where the second strand of the target DNA is displaced as a loop) that increase the strength of binding, and, in the case of double duplex formation, are exempt from the homopurine target restriction of TFOs (Nielsen et al., 1994, Ricciardi et al., 2014, Lonkar et al., 2009, Jensen et al., 2011). Strategies based on TFOs and PNAs have achieved correction efficiencies of 0.1–2.5% in somatic cells (Jensen et al., 2011, Chin et al., 2008, Schleifman et al., 2011).

1.3.3 Adeno-associated viruses (AAVs)

While adeno-associated viruses (AAVs) are frequently associated with episomal gene addition gene therapy (Daya and Berns, 2008), AAVs can also be used as a TGR vector (Jensen et al., 2011, Hendrie and Russell, 2005). The AAV genome consists of a 4.7 kb linear DNA molecule, containing genes for viral replication (rep) and capsid formation (cap), and flanked at both ends by inverted terminal repeats (ITRs). For the purpose of TGR, the entire viral genome, apart from the ITRs that are required for viral packing, can be replaced with a 4.5 kb sequence containing homology to a desired target sequence. This 4.5 kb sequence of homology to a target site may contain subtle sequence changes, such as point mutations, deletions, or small insertions (<20 nucleotides); although, larger insertions are possible when selectable markers are included within the insertion (Hendrie and Russell, 2005).

Recombinant AAV-mediated TGR is thought to rely on the HDR pathways. Strand invasion and heteroduplex formation are believed to precede a recombination event, although repair synthesis may also be involved (Hendrie and Russell, 2005). Additional lines of evidence also suggest the involvement of NHEJ (Jensen et al., 2011). TGR

mediated by recombinant AAVs can reach correction efficiencies of 1% in fibroblasts, ESCs, and iPSCs, without the use of selection (Hendrie and Russell, 2005, Khan et al., 2010). Up to 65% correction efficiencies have been achieved in somatic cells using strategies employing additional methods, such as selection and double-strand breaks (Gellhaus et al., 2010, Jensen et al., 2011). One noteworthy downside of using AAVs for TGR strategies is that random integration of the recombinant AAV DNA into the host genome happens at tenfold the rate of correction, and this occurs in about 10% of the cells that have been successfully corrected (Hendrie and Russell, 2005).

1.4 Where we're at today: current genome editing techniques

1.4.1 Programmable site-specific nucleases and nickases

1.4.1.1 Background

Over the last 10 years, great advances have been made in the development of tools for targeted genome editing. The most popular tools currently in use and under development are the programmable site-specific nucleases and nickases. Programmable site-specific nucleases and nickases are enzymes with nuclease or nickase activity that have DNA binding domains with programmable site-specificity. These enzymes have the ability to catalyse deletion, insertion, and sequence replacement events at desired locations within a target genome through the introduction of single- or double-strand breaks (SSBs and DSBs, respectively), and the subsequent induction of HDR and NHEJ within the cell (Figure 1-5) (Segal and Meckler, 2013). Examples of programmable site-specific nucleases and nickases include: engineered homing endonucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), RNA-guided endonucleases (RGENs) such as the CRISPR-Cas system, and nickase adaptations of all the foregoing nuclease systems. These examples will be described further in the following sections after a brief treatment of the general characteristics of programmable site-specific nucleases and nickases that follows.

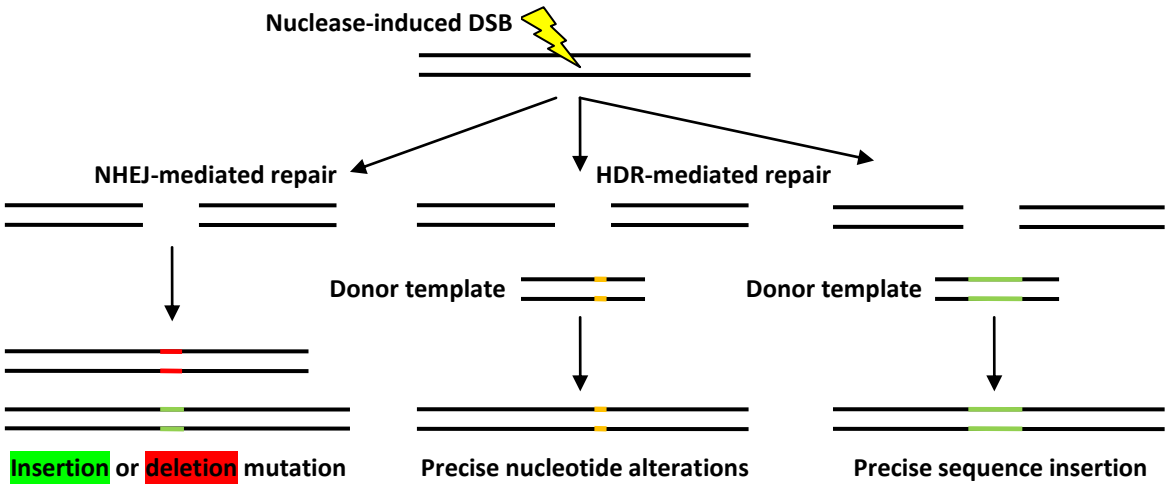


Figure 1-5: Nuclease-induced genome editing. Nuclease-induced DSB can either lead to NHEJ-mediated repair or HDR-mediated repair, depending on whether a donor template is present. When a donor template is absent, NHEJ results in the formation of either small insertion or small deletion mutations. When a donor template is present, HDR can use the donor template to repair the DSB, causing precise nucleotide substitution or precise insertions, depending on the donor template used. Adapted Joung and Sander (2013).

1.4.1.2 Types of sequence edits

Small indels (small insertion or deletions typically of 1–25 bp in length) may be introduced to a target locus through the introduction of a single DSB (Figure 1-5) (Segal and Meckler, 2013). The DNA ends that are produced are then processed predominantly by the NHEJ pathway. This processing typically involves exonuclease and polymerase activities leading to small truncations or additions to the DNA ends before ligation, resulting in the creation of indels. Additionally, programmable site-specific nickases, which place adjacent nicks on opposite strands of the target DNA, may also be used to generate indels. Site-specific nickases are generated by inactivating one domain or subunit (depending on whether the enzyme is monomeric or dimeric, respectively) in a site-specific nuclease system. The nickase strategy has the advantage of reducing off-target DSBs, as individual nickases produce off-target DSBs at a much lower frequency than off-target activity by nucleases (Cho et al., 2014). In one study it was shown that the use of programmable site-specific nickases reduced off-target mutagenesis by over 100-fold compared with their nuclease counterparts (Ran et al., 2013). The use of nickases also has the advantage of increasing the on-target specificity of the enzyme system as twice the number of site-specific proteins (i.e. two or four proteins, depending whether the parent system was monomeric or dimeric, respectively) must bind to the target site in order for a DSB to be achieved, and thus more base pairs are specified at the target site.

When larger deletions are desired, two DSBs may be introduced into a chromosome by site-specific nucleases, which act to delete the intervening DNA segment (Carlson et al., 2012). The resulting DNA ends flanking the deletion are, again, predominantly repaired by NHEJ. However, because of the NHEJ exonuclease and polymerase activity, the boundaries of the deletion may be somewhat imprecise. Inversion of the chromosomal segment flanked by the DSBs is also a potential outcome, although this outcome occurs at a much lower frequency than deletion. In a similar fashion, a pair of site-specific nickases may also be used to generate large deletions of up to 1.1 kb (Cho et al., 2014). In this case, it appears that deletion of the intervening chromosomal segment does not occur as the result of both nicks being converted to DSBs, and it is also unlikely that two SSBs separated by more than 100 bp lead to the production of a single DSB with large overhangs (because the melting temperature is very high under physiological conditions). Instead, deletions have been proposed to take place as a result of a more complex recurrent process involving strand displacement, BER, generation of a single DSB, and NHEJ (Cho et al., 2014).

Precise insertions and DNA sequence edits can also be accomplished through the use of programmable site-specific nucleases and nickases. In this case, DSBs or SSBs are introduced to stimulate HDR, which when accompanied by the introduction of donor DNA molecules (circular, linear, or single-stranded) with homology to a genomic target sequence, can stimulate replacement of a target sequence with the donor sequence (Figure 1-5) (Moehle et al., 2007, Joung and Sander, 2013). The size of the conversion tracts in mammalian cells typically ranges from 100–200 bp (Carroll, 2014, Elliott et al., 1998), but can be as long as 7.6 kb (Moehle et al., 2007, Joung and Sander, 2013). The length of homology typically required for the donor molecule ranges from 1–4 kb (usually distributed equally relative to both sides of the DSB) (Carroll, 2014), although oligonucleotides as short as 40 bp have been used successfully (Carroll, 2014, Chen et al., 2011). The introduction of DSBs can stimulate a 1000-fold increase in HDR activity compared with classical gene targeting strategies (see Section 1.3.1), without the use of selection, and also reduces the length of homologous donor sequence needed for the reaction (Segal and Meckler, 2013).

1.4.1.3 Challenges, and strategies to overcome them

One problem that arises in many cell types when using site-specific nucleases to generate DSBs in order to facilitate gene conversion, is that the activity of the NHEJ pathway is equal to, or dominates over, the HDR pathway, resulting in unwanted indels at the target site instead of the desired gene conversion events (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012). This is further complicated by the fact that in many instances, the product of gene conversion still retains the target site for the site-specific nuclease, allowing the corrected allele to be altered by a subsequent DSB, and NHEJ repair (Segal and Meckler, 2013). Because the indels produced by NHEJ may not leave the site-specific nuclease site intact, this could even lead to stochastic selection for an indel end-product over gene conversion.

For cell culture experiments or *ex vivo* gene therapy applications, screening may be sufficient to identify correctly corrected clones; however, in cell types such as ESCs and iPSCs in which the frequency of nuclease-mediated HDR is low, this may require the screening of 120–240 clones (Segal and Meckler, 2013, Soldner et al., 2011). Alternatively, selection strategies similar to those used in classical gene targeting strategies, involving selectable markers flanked by *loxP* sites that can be introduced and subsequently removed (see Section 1.3.1.3), may also be employed where appropriate (Hockemeyer et al., 2009, Segal and Meckler, 2013). A similar selection strategy has also been devised using the *piggyBac* transposon, which allows the selectable marker to be removed without the generation of a *loxP* sequence scar (see Figure 1-4), provided there is a natural TTAA site within the targeting region (Yusa et al., 2011, Segal and Meckler, 2013).

The generation of site-specific nickases from site-specific nuclease systems is one approach to addressing the problem of unwanted NHEJ activity. Single nickase-induced SSBs can stimulate HDR, while at the same time greatly reducing the level of NHEJ activity (some conversion of SSBs to DSBs does still occur, perhaps as the result of a replication fork proceeding to the SSB) (Segal and Meckler, 2013, McConnell Smith et al., 2009, Sander and Joung, 2014). However, in most cases the level of HDR stimulated by nickases has been shown to be greatly reduced (3–10-fold lower) compared with that of site-specific nucleases (Ran et al., 2013, McConnell Smith et al., 2009, Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012); although, the proportional reduction in NHEJ is far greater (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012). The diversity in the

efficiency of nickase-mediated HDR reactions arises as a result of factors such as: which nicking system was being use, the chosen target site, the cell type the experiments were conducted in, the technique used to measure the efficiency, and whether gene correction or gene addition was being attempted (gene addition appears to be more efficient (Wang et al., 2012)). Additionally, some studies have actually reported high levels of site-specific nickase induced indels at some sites (Sander and Joung, 2014). The use of site-specific nickases may reduce off-target mutagenesis, since off-target SSBs are most likely to be repaired by BER (Dianov and Hubscher, 2013, Hsu et al., 2014); although, off-target DSBs may also be produced at a lower level.

The level of site-specific nickase-induced HDR can be increased to a level comparable to that of site-specific nucleases by using a double nicking strategy. A double nicking strategy generates DSBs, while at the same retaining the benefit of greatly reduced off-target mutagenesis by the individual nickase enzymes (Ran et al., 2013). This approach may be especially worthwhile in light of that fact that, in cell types such as ESCs, the lower HDR activity induced by single nickase-induced SSBs has produced only limited success (Hsu et al., 2013b). Finally, perhaps one of the most interesting applications of site-specific nickases for targeted genome editing, is their use in a quadruple nicking strategy that creates overhangs allowing the precise insertion of a double-stranded oligonucleotide with compatible overhangs (repair is predicted to occur through NHEJ-mediated ligation) (Maresca et al., 2013). In a paper that described this strategy, it was found that 1 out 37 clones screened contained the correct insert (Ran et al., 2013). However, no description was given about the frequency at which aberrant mutagenesis might occur at the target site. Because site-specific nickases do possess the potential for off-target DSB creation, increasing the variety of site-specific nickase enzymes involved in a reaction might increase these events (Sander and Joung, 2014).

1.4.2 Engineered homing endonucleases

1.4.2.1 Engineered homing endonuclease background

Homing endonucleases (also known as meganucleases) are a class of naturally occurring site-specific nucleases that possess high targeting fidelity, and specify long target sites (e.g. I-SceI, 18 bp; I-AniI, 19/20[1] bp; I-CreI, 22 bp). Some of the seminal work

[1] The number of bp recognized by I-AniI is variously described in the literature as being either 19 or 20.

demonstrating that DSBs could enhance gene targeting efficiency by driving HDR was performed with the homing endonuclease I-SceI (Smih et al., 1995), and I-SceI remains the standard for high activity and site-specificity that other site-specific nucleases used for gene targeting are compared to (Silva et al., 2011). Engineered homing endonucleases are homing endonucleases that have had their binding specificity altered in order to achieve desired site-specificity.

1.4.2.2 Engineered homing endonuclease characteristics

There are five families of homing endonucleases classified by their sequence and structural motifs (LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD-(D/E)XK), of which the LAGLIDADG family have the highest level of site-specificity and have been best studied (Silva et al., 2011, Eastberg et al., 2007). Homing endonucleases of the LAGLIDADG family contain a characteristic α/β fold (αββαββα) that is responsible for both DNA recognition and cleavage (and the core subunit-subunit interaction in dimeric proteins) (Silva et al., 2011). All LAGLIDADG homing endonucleases bind to DNA through contacts spread out over two α/β fold motifs; however, these two motifs may be contained within one monomeric enzyme (e.g. I-SceI and I-AniI), or shared between two subunits of a dimeric enzyme (e.g. I-CreI and I-MsoI) (Silva et al., 2011, Eastberg et al., 2007).

Because the DNA-binding activity of the protein is structurally linked to its cleavage activity, retargeting of the enzyme's site-specificity cannot be performed by a modular binding domain swap, as is done in the creation of ZFNs and TALENs (described in Section 1.4.3). Instead, LAGLIDADG engineered homing endonucleases are generated through: recombination of binding elements from several LAGLIDADG homing endonucleases (Chevalier et al., 2002, Epinat et al., 2003, Silva et al., 2006); semi-rational design, where residues are mutated based on structural or functional knowledge (Seligman et al., 2002, Sussman et al., 2004, Arnould et al., 2006, Joshi et al., 2011); directed evolution (Doyon et al., 2006, Chen et al., 2009, Chen and Zhao, 2005); computational approaches involving *in silico* optimization and screening (Ashworth et al., 2006, Ashworth et al., 2010, Thyme et al., 2009); or combinations of the above (Silva et al., 2011, Rosen et al., 2006, Smith et al., 2006, Arnould et al., 2007, Thyme et al., 2014). However, reprogramming the site-specificity of homing endonucleases is nontrivial, and certain characteristics of these enzymes such as differences in binding specificity from the two α/β fold motifs and differential binding and catalytic transition-state contributions

from the two α/β fold motifs (Spiegel et al., 2006, Thyme et al., 2009), balancing of differential thermodynamic contributions to affinity and specificity (Eastberg et al., 2007), and contributions to binding from outside the α/β fold motifs present challenges for construction (Prieto et al., 2007).

Another problem which exists is that when using engineered homing endonucleases based on naturally homodimeric proteins (one α/β fold motif per subunit) to generate site-specific heterodimers capable of recognizing a chosen asymmetric target site, the possibility exists for cleavage activity from three species of dimers (two of which are undesired homodimers). As with other site-specific nuclease systems (see the following Section 1.4.3), obligate heterodimers have been generated by reengineering the dimer interface in order to overcome this problem (Fajardo-Sanchez et al., 2008). Another unique approach to overcoming the unwanted homodimer problem is to generate single protein chains from two site-specific subunits, mimicking the subfamily of monomeric LAGLIDADG homing endonucleases that contain two α/β fold domains in a single protein (Grizot et al., 2009, Li et al., 2009). Finally, the first experiments to generate site-specific nickases (see Section 1.4.1 for a description of the utility of site-specific nickases), were carried out in engineered homing endonuclease systems (Niu et al., 2008, McConnell Smith et al., 2009).

1.4.2.3 Engineered homing endonuclease performance

Engineered homing endonucleases have been successfully used to target about 10 different unique human disease targets (a relatively small number compared the other site-specific nuclease systems) (Segal and Meckler, 2013). In one example, gene targeting frequencies of 6% were achieved at the *RAG1* locus (a gene involved in severe combined immunodeficiency; SCID) in 293H cells (transformed primary embryonic human kidney cells) using a fully redesigned single-chain variant of the I-CreI homing endonuclease (Grizot et al., 2009). This *RAG1*-targeting endonuclease was also shown to have a toxicity and off-target cleavage profile comparable to I-SceI (Grizot et al., 2009, Silva et al., 2011). However, experiments with homing endonucleases and engineered homing endonucleases have demonstrated that the ability to successfully correct a target locus may drop precipitously the farther the polymorphism is located away from the DSB (i.e. shorter gene conversion tracts occur more frequently than longer ones), although overall levels of efficiency can be highly variable between experimental designs (Silva et al.,

2011). Thus, the challenges that exist in designing the site-specificity of engineered homing endonucleases may significantly limit their versatility for gene editing applications.

1.4.3 Chimeric restriction enzymes with programmable binding domains: ZFNs and TALENs

1.4.3.1 ZFN and TALEN background and characteristics

Among the most popular and versatile systems of site-specific nucleases are the ZFNs and TALENs: chimeric enzymes that typically utilize the nuclease catalytic domain from the Type IIS restriction endonuclease *Fok*I combined in a modular fashion with a programmable site-specific DNA binding domain (Figure 1-6). The programmable binding domain of a ZFN or TALEN consists of either a zinc finger array (ZFA) or the binding domain of a transcription activator-like effector (TALE), respectively (these binding domains will be described in detail in Section 1.7). The *Fok*I endonuclease cleaves DNA at a site distal to its recognition sequence. In 1992, Chandrasegaran and colleagues demonstrated that the binding and cleavage activities of *Fok*I were physically separated into two domains of the protein (Li et al., 1992). This discovery lead Chandrasegaran's group to experiments that demonstrated the *Fok*I catalytic domain could be combined with other binding domains, including zinc finger binding domains, in order to retarget the nuclease activity (Kim and Chandrasegaran, 1994, Kim et al., 1996). This became an increasingly important discovery for the field of genome engineering, as in the years that followed, great advances were made in the development of engineered ZFAs with designed site-specificity (Ramirez and Joung, 2013, Ochiai and Yamamoto, 2015). More recently, an even more versatile binding domain from TALEs of bacterial plant pathogens has been discovered, and this has led to the creation of *Fok*I-based TALENs.

The *Fok*I endonuclease exists as monomers in solution (Kaczorowski et al., 1989), but it requires dimerization at its DNA target site for activation and cleavage (Vanamee et al., 2001, Bitinaite et al., 1998). Likewise, site-specific nucleases based on *Fok*I also require dimer formation at the active site for cleavage (Smith et al., 2000). ZFN and TALEN monomers usually recognize 9–18 bp and 15–20 bp (Kim and Kim, 2014), respectively, allowing their active dimers to specify sequences that are statistically unique within a

human-size genome (>17 bp in a $3x10^9$ bp genome)[2]. The inactive nature of the monomer has allowed the development of several useful modifications of the *Fok*I catalytic domain including: a hyperactive mutant (Guo et al., 2010), obligate heterodimer forming monomers (Miller et al., 2007, Szczepek et al., 2007, Doyon et al., 2011, Guo et al., 2010), and an inactive subunit for use in a nickase heterodimer system (Kim et al., 2012, Wang et al., 2012, Ramirez et al., 2012).

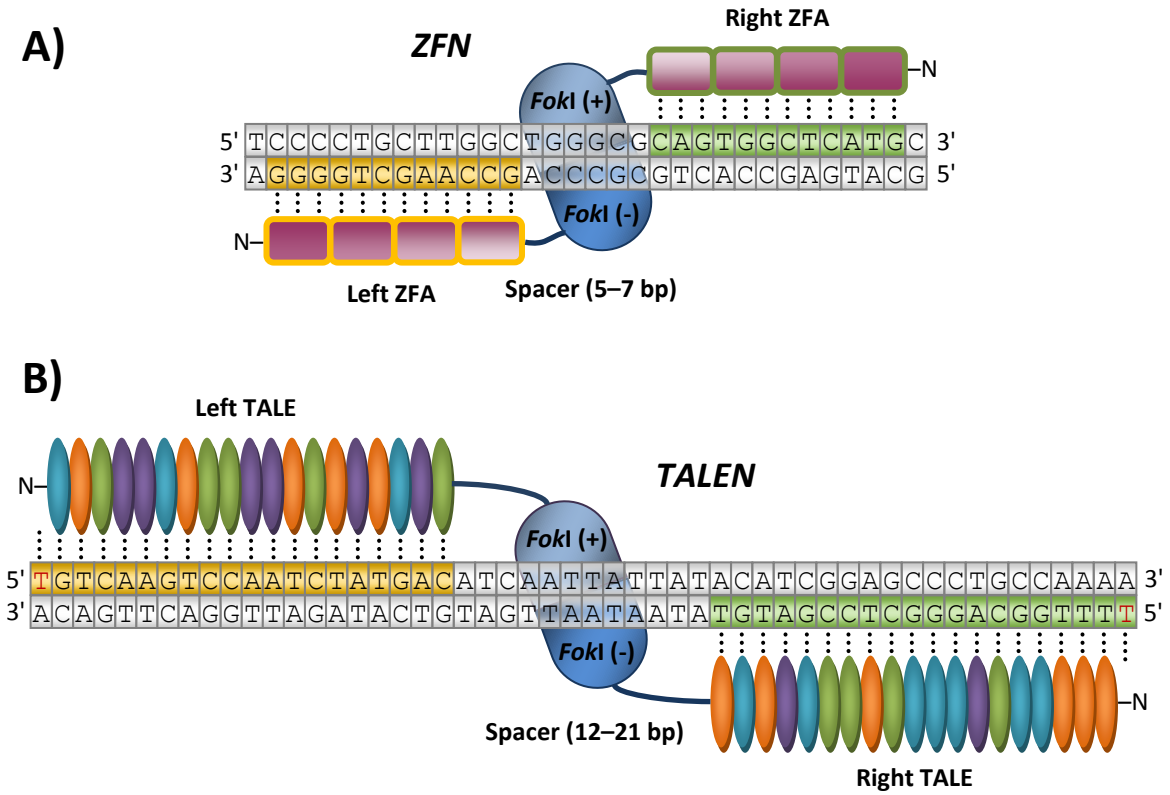


Figure 1-6: Architecture of ZFN and TALEN. A) At the top of the figure is shown a schematic of a ZFN binding as a dimer to its target site. Each ZFN subunit contains a *Fok*I nuclease domain connected to a ZFA via a linker. Each zinc finger in the ZFA specifies 3 bp of the target site. ZFA binding sites (shown in orange and green) are separated by a spacer sequence which is bound non-specifically by the *Fok*I domain. **B)** At the bottom of the figure is shown a schematic of a TALEN binding as a dimer to its target site. Each TALEN subunit contains a *Fok*I nuclease domain connected to a TALE binding domain via a linker. Each repeat module of the TALE domain specifies 1 bp of the target site. TALE binding sites (shown in orange and green) are separated by a spacer when bound non-specifically by the *Fok*I domain. The thymine at the 5′ of each TALE binding site (indicated in red), is specified by an element of the TALE binding domain which is typically not programmable. Adapted from Kim and Kim (2014).

[2] In order to achieve statistical uniqueness in a human-sized genome, a target site must be at least 17 bp. This is determined by considering a target site whose length dictates that the inverse of its probability of occurrence is greater than the sample space of both strands of a 3 billion bp genome ($4^{17} > 3 \times 10^9 \times 2$).

1.4.3.2 ZFN and TALEN performance

Most of the milestones in genome engineering have been accomplished using ZFNs (Segal and Meckler, 2013), and the number of human genes modified by them is too numerous to list here (although they are reviewed in Segal and Meckler, 2013). In one landmark study, ZFNs were shown to be able to mediate gene conversion at the *IL2Rγ* gene (associated with X-linked SCID) in 18% of K562 cells (immortalized myelogenous leukaemia cells) after transfection with both ZFN and donor constructs (Urnov et al., 2005). Although it was necessary for the experimenters to arrest the cell cycle for 30 hours at the G2/M cell cycle boundary to achieve such high efficiency, even in the absence of cell cycle arrest, 10% of cells were modified. In a follow up study, the same group was able to induce gene addition at the *CCR5* gene (whose protein product is associated with HIV cell entry) in 5% of treated ESCs, using a lentiviral vector to deliver the ZFN expression cassettes and donor construct (Lombardo et al., 2007). However, the same study demonstrated only 0.11% efficiency in human $CD34^+$ hematopoietic progenitor cells, indicating that ZFNs may not be universally successful in all cell types.

ZFNs also hold the special honour of being the first site-specific nuclease system to be used in clinical trials. Sangamo Bioscience's (Richmond, CA, USA) SB-728-T is a ZFN-based autologous T-cell therapy designed to disrupt the *CCR5* gene coding for C-C chemokine receptor type 5, the cell surface receptor that HIV uses to gain entry into CD4 T cells (Maier et al., 2013, Tebas et al., 2014). The *CCR5Δ32* loss-of-function mutation is known to impart resistance to the most prevalent strains of HIV, and was the crucial factor in the 2008 'Berlin patient', the first person to be cured of HIV, via a bone marrow transplant from a homozygous *CCR5Δ32* donor (Hutter et al., 2009, Allers et al., 2011). In the latest reports from SB-728-T phase I and II trials, it was reported that *CCR5* was successfully disrupted in 11–28% of patients' CD4 T cells *ex vivo,* using four-fingered ZFNs (Tebas et al., 2014). These T-cells were well tolerated after reintroduction, and viral load was reduced in the majority of the patient cohort that was taken off conventional anti-HIV therapy, with one patient's viral load dropping below detectable levels.

Although ZFNs represent a major breakthrough for targeted genome engineering, there are still significant limitations in the targeting density that ZFNs can achieve, owing to the reliance on ZFAs (ZFAs will be covered in detail in Section 1.7.1). It has been estimated that successful ZFN target sites can be found at a frequency of eight sites per

1 kb of random genomic sequence, with guanine-rich sites being more easily targeted (Kim et al., 2009). However, there are many different methods of ZFA design, and even the most successful of the publically available methods still result in failure rates of 18–77% for the individual ZFN pairs into which they have been incorporated (Gupta et al., 2012, Bhakta et al., 2013, Sander et al., 2011b, Kim et al., 2011, Zhu et al., 2011), and can be >90% using some design methods (Joung et al., 2010).

TALENs have much more versatile targeting capability than ZFNs, as their TALE binding domains can be programmed to recognize virtually any sequence, with the only restriction being a thymine at the 5′ end of the site (TALEs will be described in detail in Section 1.7.2) (Kim and Kim, 2014). One study, in which 96 TALEN pairs were constructed to target an equal number of human genes, found that 84 of these TALEN pairs were successful in producing indels at efficiencies ranging from 2.5–55.8% (Reyon et al., 2012). Furthermore, there is evidence that TALEN targeting success rates can reach 100% if heavily methylated target sites are avoided (Kim et al., 2013). Additionally, TALENs are reported to produce equal or substantially improved off-target cleavage and cytotoxicity profiles compared with ZFNs
(Segal and Meckler, 2013, Mussolino et al., 2011, Kim and Kim, 2014). However, in a comprehensive analysis of TALENs off-target activity, it was found that individual TALENs still may act at hundreds of off-target sites throughout the human genome (Frock et al., 2015). Additionally, a high level of repetition in the TALE binding domain coding sequences makes TALEN construction challenging and requires non-standard molecular biology cloning procedures (see Section 1.7.2.4) (Sander and Joung, 2014).

1.4.3.3 Alternative ZFN and TALEN architectures

Finally, it should be mentioned that ZFNs and TALENs based on the *Pvu*II Type IIP restriction endonuclease have also been reported (Schierling et al., 2012, Yanik et al., 2013). These enzymes were designed to address the off-target activity that arises from *Fok*I-based ZFNs and TALENs as a result of the lack of specificity of the cleavage domain, with possible binding specificity degeneracy of the ZFA or TALE domains allowing unwanted activity at partial recognition sites. The cleavage module of *Pvu*II-based ZFNs and TALENs recognizes and cleaves DNA with high specificity at the 6 bp *Pvu*II recognition site. In spite of their somewhat more restricted targeting capability, these enzymes indeed show notably improved off-target cleavage and toxicity profiles when compared

to *Fok*I-based ZFNs and TALENs targeted to the same sites, and with high fidelity persisting even under conditions of increased enzyme concentration and reaction duration. The *Pvu*II-based TALENs, in particular, had a toxicity profile in HEK293T cells similar to I-SceI (Yanik et al., 2013). Although the use of *Pvu*II-based ZFNs and TALENs has been demonstrated, they have not been widely adopted. An alternative ZFN architecture has been also described in which staphylococcal nuclease, which cleaves as a monomer, has been sandwiched between two ZFAs (Mineta et al., 2008). Additionally, a chimeric Cas9-*Fok*I RGEN has also been described (see Section 1.4.4.4) (Tsai et al., 2014b).

1.4.4 RGENs: the CRISPR-Cas9 system

1.4.4.1 Cas9 RGEN background

Over the last few years a new and important player has emerged as a leading tool for targeted genome editing. The CRISPR-Cas9 system is one of a new class of site-specific nucleases: RGENs (RNA-guided endonucleases). CRISPR, which stands for clustered regularly interspaced short palindromic repeats, is an adaptive immune system that has been discovered in prokaryotes. CRISPR systems function to protection their bacterial and archaeal hosts from viruses and plasmids by retaining exogenous genetic sequences from previous infections (within a genomically encoded 'CRISPR array'), which are then used to transcribe RNA that facilitate targeting of similar sequences upon re-infection. Three types of CRISPR systems (Type I–III) have been described thus far, with the Type II system being the best characterized and least complex. The Type II CRISPR system utilizes a monomeric endonuclease enzyme (Cas9) complexed with two mutually bound strands of RNA (crRNA and tracrRNA; see below), which are used to direct the complex to cleave invading genetic material. The Cas9 RGEN from *Streptococcus pyogenes* has become the most popular and well studied RGEN system for targeted genome editing, and it is this system that will be discussed in the following sections. For a review of CRISPR biology and research history, see Hsu, Lander, and Zhang (2014).

1.4.4.2 Cas9 RGEN characteristics

The advantage of using the Cas9 RGEN for targeted genomic editing is its versatility and ease of use. Unlike ZFNs and TALENs, the targeting capacity of RGENs is afforded by simple Watson-Crick base pairing rules between guide RNA and target DNA, instead of protein-DNA interactions (Sander and Joung, 2014). In the natural CRISPR-Cas9

system, the Cas9 enzyme is bound to a *trans*-activating CRISPR RNA (tracrRNA) that mediates contact between Cas9 and the CRISPR RNA (crRNA), which targets the enzyme to invading DNA (Deltcheva et al., 2011). Typically, in the Cas9 RGEN system used for targeted genomic editing, the tracrRNA and crRNA are fused together to create a single guide RNA (sgRNA) that recognizes a 20 bp target site (Jinek et al., 2012, Sander and Joung, 2014). Because the sgRNA is the primary component of Cas9 RGENs responsible for DNA targeting, programming Cas9 to target desired sequences can be done through the design of a simple sgRNA expression construct. Another feature of versatility in the Cas9 RGEN system is that, owing to the chemical independence of Cas9 and its sgRNA, a single Cas9 protein can be introduced to a cell simultaneously with a variety of sgRNAs to facilitated multiplexed genome editing (Cong et al., 2013).

In addition to the recognition specificity of the crRNA (or sgRNA), Cas9 has another recognition requirement called a protospacer-adjacent motif (PAM) that facilitates target searching and serves to allow Cas9 to discriminate between new invading genetic material and the genetic sequences that have been previously captured into the CRISPR array (Shah et al., 2013). The PAM is a short DNA sequence that is bound by the PAM-interacting C-terminal region of Cas9 (Nishimasu et al., 2014). The PAM sequence for the *S. pyogenes* Cas9 is 5′-NGG (where 'N' can be any base), and must be located directly 3′ of the 20 bp target sequence that is recognized by the sgRNA (Jinek et al., 2012). Strategies to circumvent the PAM restriction include: trading reduced efficiency for alternate PAM recognition (5′-NAG and 5′-NNGG sequences are also accepted); utilizing Cas9 orthologues with altered PAM preference; and orthologous replacement of the PAM-interacting C-terminal region of Cas9 (Hsu et al., 2014, Sander and Joung, 2014). See Figure 1-7 for a schematic of the Cas9 RGEN and its features.

The choice of promoter used to express the sgRNA can also impose a restriction on the target site (Sander and Joung, 2014). Promoters such as the RNA polymerase III-dependent promoter or the T7 promoter require that a G or GG, respectively, be located at the 5′ end of the sequence to be transcribed. Therefore, the sequences that can be targeted by the Cas9 RGEN may take the form of GN_{20}-NGG or GGN_{20}-NGG, which would be expected to occur every 32 bp or 128 bp of random genomic sequence, respectively (given that either the Watson or Crick strand could be targeted) (Sander and Joung, 2014).

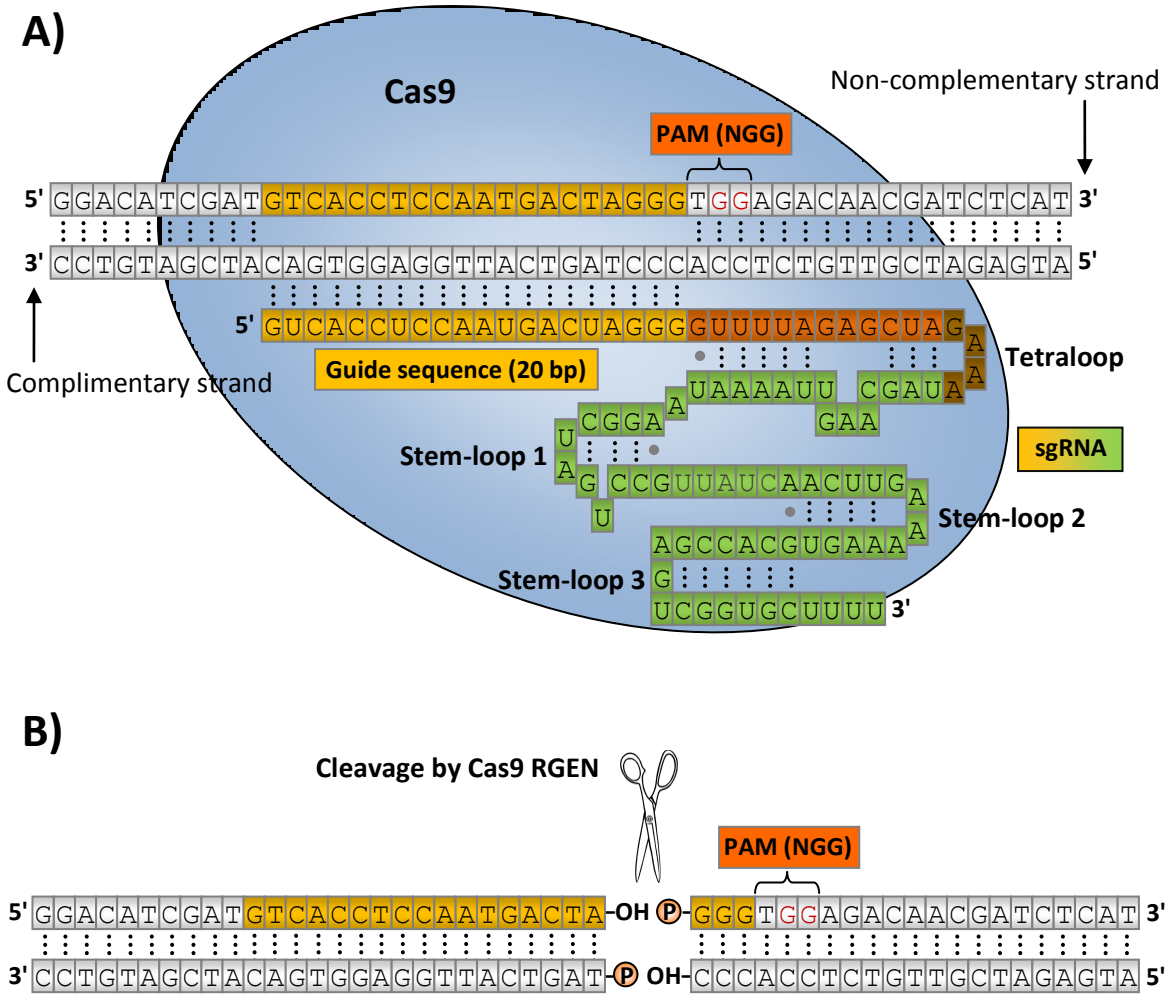


Figure 1-7: The Cas9 RGEN. A) A schematic of the Cas9 RGEN bound to a target site. The complementary strand of the target DNA molecule (grey bases) pairs with the guide sequence (yellow bases) of the sgRNA. The sgRNA is colour-coded to indicate the origin of its segments. Yellow and orange indicate the component which is derived from crRNA, brown indicates the fusion linker, and green indicates tracrRNA. Watson-Crick base pairing is indicated by a series of dots, while single grey dots indicate weak bonding. **B)** Target DNA cleaved by the Cas9 RGEN yielding blunt ends. Adapted from Kim and Kim (2014).

1.4.4.3 Cas9 RGEN performance

Since the first two papers (published simultaneously) demonstrating the use of the Cas9 RGEN for targeted genomic editing were published in 2013 (Cong et al., 2013, Mali et al., 2013), the tool has been rapidly adopted by the scientific community, with papers being published describing its use in a wide variety of organisms and at a wide variety of targets (Sander and Joung, 2014). Cas9 RGENs have been reported to be successful at about 90% of target sites, and can produce targeted gene disruptions with about 20% efficiency (Kim and Kim, 2014). Notably, the Cas9 RGEN was used in the first ever site-specific genomic editing experiment in human embryos (Liang et al., 2015). This experiment attempted to use RGEN-mediated HDR to correct the β-globin gene

(associated with β-thalassaemia, a potentially fatal blood disorder) in non-viable triplonuclear human zygotes; however, the technique was found to be inefficient, and also resulted in substantial off-target activity.

It has been estimated that for any Cas9 RGEN, the number of potential off-target sites may number in the thousands when used in a human-sized genomic context (Sander and Joung, 2014). Furthermore, activity at these off-target sites has been measured at over 2–5%, and can be as high as Cas9 activity at on-target sites (Fu et al., 2014, Fu et al., 2013, Cradick et al., 2013, Sander and Joung, 2014). In the two most robust characterizations of Cas9 off-target activity to date, it was found that Cas9 RGENs may tolerate up to six or seven mismatches in their 20 bp target sequences, in addition to a variety of mismatches in the PAM sequence (Tsai et al., 2015, Frock et al., 2015). These widespread low-level off-target reactions were found to cause genomic rearrangements including translocations, inversions, and large deletions. Furthermore, one of these studies found that in some cases, the Cas9 RGENs were actually more active at off-target sites than at their on-target site (Tsai et al., 2015). However, both studies also found that although some Cas9 RGENs had off-target sites numbering in the hundreds, others showed no detectable off-target activity at all (Tsai et al., 2015, Frock et al., 2015). The reasons for this disparity in off-target potential are not yet clear. Additionally, it was shown that all of the RGENs may produce translocation events between homologous chromosomes (including dicentric chromosome production) as a result of their on-target reactions (Frock et al., 2015). Genomic rearrangements were also found to occur between the on-target sites and naturally occurring breakpoint hotspots within the genome (Frock et al., 2015, Tsai et al., 2015).

1.4.4.4 Improvements to the Cas9 RGEN system

Several modification to the Cas9 RGEN system have been developed to reduce off-target activity, including the generation of Cas9 nickases (see Section 1.4.1 for a description of nickases), and reduction of the sgRNA length (Sander and Joung, 2014). It has been demonstrated that Cas9 is more tolerant of base pairing mismatches at the 5′ end of the 20 bp target sequence (the end most distal to the PAM sequence) (Cong et al., 2013, Fu et al., 2013, Hsu et al., 2013a); therefore, truncated sgRNAs (tru-sgRNAs) have been created that omit 2 or 3 bp of recognition from the 5′ end of the target sequence (Fu et al., 2014). These tru-sgRNAs recognize a 17 or 18 bp target sequence, and facilitate

Cas9 activity with similar efficiency and higher specificity than sgRNAs recognizing a 20 bp target sequence. The two previously mentioned robust characterizations of RGEN off-target activity verified that the Cas9 nickase (Frock et al., 2015) and tru-sgRNAs (Tsai et al., 2015) strategies do indeed substantially reduce off-target activity, although they do not reduce translocation events between homologous chromosomes resulting from on-target activity.

Finally, it should be noted that an additional class of RGENs, consisting of a catalytically inactivated Cas9 fused to a *Fok*I cleavage domain, have been described in the literature (Guilinger et al., 2014, Tsai et al., 2014a). The Cas9-*Fok*I RGENs, which act as dimers, may have the advantage of allowing more sequence selectivity than monomeric Cas9 RGENs, and may allow obligate heterodimer RGENs to be easily produced (as obligate heterodimer variants of the *Fok*I cleavage domain already exist). Although Cas9-*Fok*I RGENs have promising features, they have not yet been widely put into use.

1.4.5 Delivery of site-specific nucleases

1.4.5.1 Cell culture delivery

Delivery of site-specific nuclease and homologous DNA template reagents into cultured cells can be accomplished by plasmid DNA transfection, transfection of *in vitro* transcribed mRNA, viral vector delivery, and direct delivery of protein (Carroll, 2014). Microinjection has also been used to deliver site-specific nuclease constructs to fertilized mammalian embryos for the purpose of generating a variety of genetically modified animals (Carbery et al., 2010, Cui et al., 2011, Geurts et al., 2009, Meyer et al., 2010, Carlson et al., 2012, Niu et al., 2014). A nuclear localization signal (NLS) coding region and poly-A tail are typically incorporated into site-specific nuclease constructs (Gaj et al., 2012, Sanjana et al., 2012, Mali et al., 2013). It has been reported that the ratio of expression plasmid to correction template can affect correction efficiency in site-specific nuclease-mediated HDR reactions (Pruett-Miller et al., 2008b, Porteus, 2006, Connelly et al., 2010). Ideally, a transfection mix should contain 10% expression plasmid and 90% corrective template (Connelly et al., 2010).

Some cell types, such as haematopoietic stem cells, differentiated immune cells, and ESCs, are resistant to common transfection methods, such as electroporation and lipofection (Hsu et al., 2014, Gaj et al., 2013a, Kim and Kim, 2014). AAV has been used successfully as a viral vector for ZFNs (Ellis et al., 2013, Handel et al., 2012), and was the method used to transduce T-cells *ex vivo* with ZFNs in the SB-728-T HIV therapy clinical trials (see Section 1.4.3.2) (Maier et al., 2013, Tebas et al., 2014). However, the sizes of TALEN and Cas9 constructs make them difficult to deliver efficiently using AAV, due to the small packaging limit of AAV vectors (optimally <4.2 kb) (Gaj et al., 2013a, Hsu et al., 2014). This packaging constraint of AAV on the Cas9 RGEN system might be overcome in the future by using smaller Cas9 orthologues that have recently been discovered (Hsu et al., 2014). TALENs and Cas9 RGENs have, however, been successfully delivered to human cells using adenoviral vector delivery (Maggio et al., 2014, Holkers et al., 2013).

Integrase-deficient lentiviral vectors (IDLVs) have been another choice for viral vector delivery of ZFNs (Lombardo et al., 2007). However, IDLVs are incompatible with TALEN constructs, as the template switching activity of reverse transcriptase leads to recombinogenic events among the multiple repeats within the TALE domain during TALEN DNA synthesis in cells (Holkers et al., 2013). Cas9 RGENs have been successfully delivered to human cells using integrating lentiviral vectors (ILVs), producing nearly 100% mutation rates (Shalem et al., 2014, Wang et al., 2014, Kim and Kim, 2014). However, using integration-competent vectors that continually express the RGEN may increase the likelihood of unwanted off-target effects (Kim and Kim, 2014).

Direct delivery of site-specific nuclease proteins into the cell has also been demonstrated. ZFNs, because of the positive charge of the protein, have the innate ability to cross cell membranes, and produce efficient activity when added directly to cell culture (Gaj et al., 2012). But TALENs and RGENs do not have this ability (Carroll, 2014). Instead, TALENs and RGENs can be delivered directly to cells by chemically conjugating them to a cell-penetrating peptide (and in the case of RGENs, additionally complexing the sgRNA to cell-penetrating peptides via charged interaction) (Ramakrishna et al., 2014, Liu et al., 2014). Chemical conjugation of the cell-penetrating peptide was necessary because cell-penetrating peptide fusion proteins proved hard to purify. Additionally, both TALENs and RGENs can be directly delivered to cells using cationic lipid-mediated delivery (Zuris et al., 2015). Direct protein delivery has the advantages of allowing delivery to cell types that

are resistant to viral or non-viral gene delivery, reducing the length of time enzymes are active (thereby reducing the risk of off-target activity), and avoiding unwanted genomic integration of genetic material (Gaj et al., 2013a, Kim and Kim, 2014).

1.4.5.2 *In vivo* delivery

ZFNs and a corrective cDNA cassette have been delivered *in vivo* using an AAV vector, in order to facilitate correction of a human *F9* gene (coding for blood coagulation factor IX) within liver cells in a humanized transgenic mouse model of haemophilia B (Li et al., 2011a). This experiment achieved 3–7% restoration of circulating levels of human factor IX (in humans, 5% restoration would be required to convert severe haemophilia B to a mild form of the disease) (Li et al., 2011a). Additionally, hydrodynamic tail vein injection (rapid injection of a large volume of DNA) of Cas9-expressing plasmids, sgRNA, and ssDNA correction template, has been used to correct the *Fah* (fumarylacetoacetate hydrolase) gene in a mouse model of hereditary tyrosinemia type I (Yin et al., 2014). The initial correction frequency was about 1/250 hepatocytes; however, after 33 days, selective pressure within the liver had expanded the corrected cell population, leading to a 9% correction rate and correction of the disease phenotype.

1.4.6 Limitations of the site-specific nuclease approach

Although engineered site-specific nucleases are powerful tools for genetic research, and, in combination with careful screening procedures, may be successful for *ex vivo* gene therapy strategies, their suitability for *in vivo* gene therapy applications is less clear. The off-target cleavage potential of site-specific nucleases presents an inherent danger to their use within the living human organism. So far, only some RGENs look like they might have the potential to produce on-target-only reactions, and even in this case, chromosomal rearrangements are still possible between homologous chromosomes on which the target site is located (Frock et al., 2015), and between the intended target site and naturally occurring DSBs (Frock et al., 2015, Tsai et al., 2015). Such chromosomal rearrangements might lead to breakage-fusion-bridge cycles (Frock et al., 2015). Breakage-fusion-bridge cycles occur when dicentric chromosomes are generated and stressed between opposite poles during mitosis until further breakage occurs, in a pattern that repeats, leading to the type of gene alterations and amplifications that are key to

oncogenesis and chemotherapy resistance (Hu et al., 2014, Zhu et al., 2002, Difilippantonio et al., 2002).

Another problem that exists with the site-specific nuclease approach to gene repair is that it relies on HDR proteins that are primarily expressed during the G2 phase of the cell cycle (Hsu et al., 2014). This reliance on HDR limits the cell types in which site-specific nucleases can be used to those undergoing mitosis; neurons and cardiac myocytes, for example, are post-mitotic cells. Additionally, reliance on HDR activity often limits gene correction activity to below optimal levels even in cell types in which HDR is active. The site-specific nuclease approach to targeted genome editing is fundamentally limited by the nature of a system, which breaks DNA, but does not have its own DNA repair capability. A comparison of the three most popular site-specific nuclease systems is shown in Table 1-1.

Table 1-1: Comparisons between the three most popular programmable nuclease systems*

	ZFNs	TALENs	RGENs
DNA targeting specificity determinant	ZFAs	TALEs	sgRNA
Nuclease	*Fok*I	*Fok*I	Cas9
Average success rate at any given target site†	Low (~24%)	High (>99%)	High (~90%)
Average mutation rate‡	Low or variable (~10%)	High (~20%)	High (~20%)
Number of bases specified at target site	18–36 bp	30–40 bp	20 bp + 2 bp (PAM)
Target site restrictions	Often G-rich	Start with T and end with A	5′ end NGG or NAG (PAM)⁋
Targeting site density	1/100 bp	All sites¤	1/8 bp (NGG PAM), 1/4 bp (NGG and NAG PAM)⁋
Off-target activity	High	Relatively low	Variable
Cytotoxicity	Variable to high	Low	Low
Size of reading frame	~1 kb × 2	~3 kb × 2	4.2 kb (*S. pyogenes* Cas9) + 0.1 kb (sgRNA)

*This table has been adapted from Kim and Kim (2014), and is based on their data collected in HEK293 cells, published in several of their previous reports; references can be found within (Kim and Kim, 2014). Mutation frequencies are higher in other cell types such as K652 cells and HeLa cells.

†The success rate is defined as the proportion of nucleases that induce mutations at frequencies >0.5% in HEK293 cells.

‡The average mutation rate is based on the frequency of non-homologous end-joining-mediated insertions and deletions obtained at the nuclease target site.

⁋This does not include any consideration of restrictions placed on the sgRNA from its expression promoter (see Section 1.4.4).

¤Given that the length of each TALEN subunit binding site can be varied by 5 bases, all sites should be statistically targetable by using that leeway to circumvent the restriction of the thymine that is required at the 5′ end of the site.

1.5 Chimeric recombinases with programmable binding domains: ZFRs and TALERs

1.5.1 Introduction

Another approach to targeted genome editing, which provides a theoretical advantage over other systems, is the use of chimeric recombinases with programmable binding domains. Recombinases are enzymes that catalyse site-specific reactions with DNA such as insertions, deletions, and inversions (see Section 1.6). These enzymes have the advantage of being able to carry out genome editing-type reactions without the assistance of endogenous cellular processes, thus, making these reactions potentially viable in all cell types, including both mitotic and post-mitotic cells (although, chromatin accessibility may still need to be considered). Recombinases such as ϕC31 integrase, Cre recombinase, and FLP recombinase are already popular tools for molecular biology; however, these recombinases are limited by the integrated site-specificity parameters of the natural enzyme. In an attempt to combine the power of recombinases with the versatility of engineered site-specific binding domains, chimeric recombinases have been developed by fusing a recombinase catalytic domain to engineered DNA-binding domains in a modular fashion.

The best studied chimeric recombinases with programmable binding domains to date consist of a catalytic domain from one of several small serine recombinases fused to either a ZFA (zinc finger recombinase; ZFR) (Akopian et al., 2003, Gordley et al., 2007) or TALE binding domain (TALE recombinase; TALER) (Mercer et al., 2012). Alternatively, chimeric recombinase enzymes based on HIV integrase (Tan et al., 2004), sleeping beauty transposase (Wilson et al., 2005), PiggyBac transposase (Wu et al., 2006, Owens et al., 2013), and ISY100 transposase (Feng et al., 2010) have also been reported. The modular architecture of the recombinases from the small serine recombinase family makes them ideal candidates for the creation of fusion proteins. Additionally, crystal structures have been published for small serine recombinase family members γδ resolvase (Yang and Steitz, 1995), and Gin invertase (Ritacco et al., 2013), aiding chimeric enzyme design. Furthermore, at the time the first ZFR was created, activating mutations for Tn3 resolvase (a close homologue of γδ resolvase) had been characterized (see Sections 1.6.4 and

1.6.5), which facilitated the use of the Tn3 resolvase catalytic domain in ZFR creation (Arnold et al., 1999, Akopian et al., 2003).

The archetypal ZFR architecture consists of a Tn3 resolvase catalytic domain fused via a linker to a Zif268 ZFA (Akopian et al., 2003). The ZFR binds its target (termed a 'Z-site') as a dimer, with site-specific contacts being mediated by both the catalytic module and the Zif268 ZFA DNA-targeting module. The Tn3 catalytic module is comprised of a catalytic domain and an arm region that interacts with the minor groove of the Z-site before connecting, via a short linker, to a ZFA that interacts with the major groove. The Zif268 targeting module may be replaced with other Zif268-like engineered ZFAs, or TALE binding domains, in order to partially reprogram the site-specificity of the chimeric recombinase (see Section 1.7 for details on ZFA and TALE binding domains). Figure 1-8 provides a schematic representation of the ZFR bound to a Z-site.

The Tn3 resolvase catalytic domain of the ZFR carries several activating mutations, which free the domain from its restrictive wild-type regulatory requirements (see Section 1.6 for details). The current ZFR architecture within the Stark lab utilizes a short Thr-Ser linker fused at Tn3 resolvase residue Arg148, which links the catalytic and binding modules. Arg148 has been chosen as the chimeric fusion point for the catalytic module in order to retain the minor groove DNA interactions of the Tn3 arm region, and the crystal structure of γδ resolvase (a close Tn3 resolvase homologue) suggests that this is an ideal location for the linker to cross over the DNA backbone to enable the major groove interactions of the ZFA module. Additionally, experimental evidence has shown this linker arrangement to be superior to several other possible arrangements (Prorocic et al., 2011). Thr and Ser residues are used for the linker because they are small and polar, making the linker both flexible and favourable to its local environment. The Thr-Ser linker also allows the inclusion of a *Spe*I restriction site that facilitates replacement of the binding module.

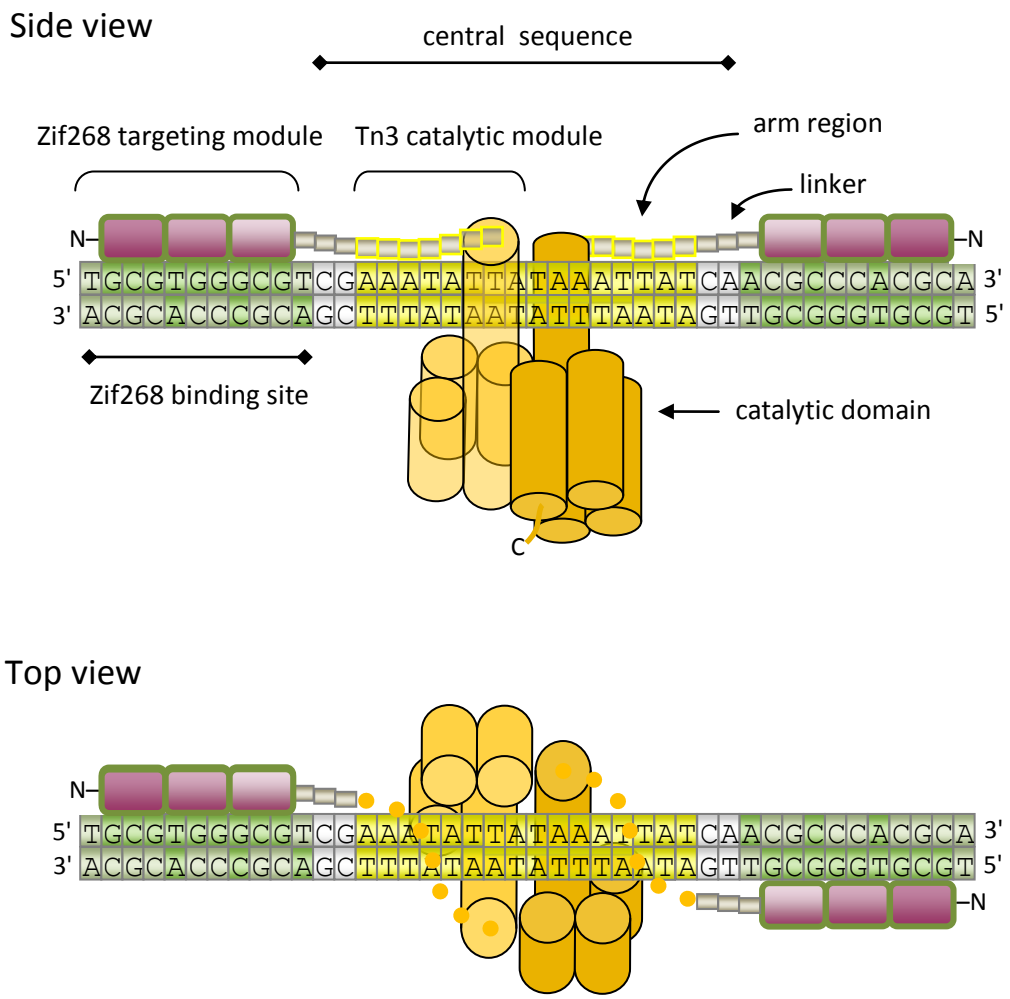


Figure 1-8: Architecture of a ZFR and Z-site. The figure depicts a schematic of a Tn3–Zif268 ZFR bound to its Z-site as a dimer. The two subunits are shown in two shades of orange. Since the catalytic domain of Tn3 resolvase is primarily composed of four α-helices (see Section 1.6), cylinders are used to depict its structure. Yellow bases are used to indicate the portion of the central sequence which is site-specifically contacted by the Tn3 catalytic modules. Green bases indicate the portions of the Z-site which are bound by Zif268 binding modules. Darker green indicates the specific base contacts of Zif268 (see Section 1.7.1.2). Top and bottom views are shown to help with visualization of the structure.

A Z-site is 40 bp and comprised of a central 20 bp region derived from Tn3 resolvase site I, which is bound by the catalytic modules of a ZFR dimer (each specifying 10 bp), which is flanked by 11 bp regions recognized by the ZFA modules that comprise the outermost portion of the Z-site (Figure 1-8). It should be noted, that depending on whether the ZFAs specify a 5′ extended position base, and/or whether finger 1 of the ZFAs have 3′ overlap specificity (see Section 1.7.1.2), the Z-site may also be described as being comprised of a 22 bp central sequence flanked by 9 bp ZFA binding sites. Although the ZFR binds a Z-site with site-specific contacts from both the catalytic module and the ZFA module, only the site-specificity of the ZFA module may be easily reprogrammed, while

the site-specificity from the catalytic module remains a limiting target-site parameter. The ZFA targeting module usually consists of three contiguous zinc finger motifs. Each zinc finger typically specifies up to four bp of DNA but with overlapping recognition sites such that a three-fingered ZFA has up to 10 bp of total recognition specificity, plus 1 bp of possible recognition specificity from the a 5′ extended position base recognition (Section 1.7.1.2). Alternatively, longer ZFAs or a TALE domain may be used to provide additional sequence recognition. Because a three-fingered ZFA allows each targeting module to recognize as many as 11 bp specifically, when ZFR dimers are bound to a Z-site, the targeting modules contribute up to 22 bp of recognition specificity—a statistically unique sequence in a human genomic context (see Footnote 2 on page 42).

The ZFR catalytic module facilitates a recombination reaction between two ZFR-bound Z-sites. Following tetramer formation by two Z-site-bound ZFR dimers, the catalytic modules catalyse cleavage of the DNA phosphodiester backbone at the centre of the Z-site, the ZFR subunits rearrange half-Z-sites, and then re-ligate the DNA to produce recombinant DNA products (Figure 1-9 A; also see Section 1.6.2 for a more detailed description of the reaction). Depending on the initial parameters of the reaction, this ZFR-mediated DNA recombination may lead to inversion, integration, or excision DNA products (Figure 1-9 B; also see Section 1.5.3).

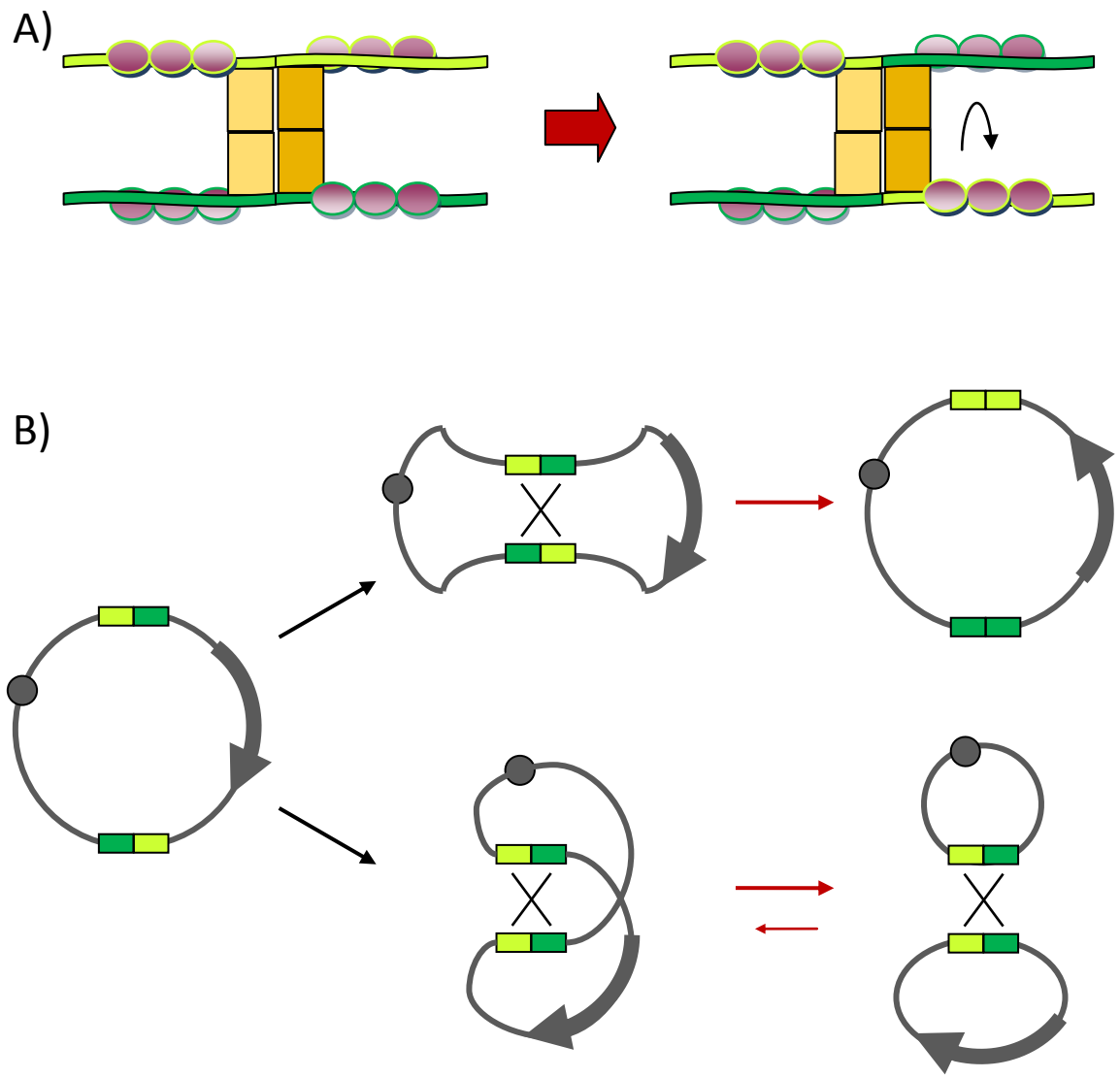


Figure 1-9: Recombination of Z-sites by a ZFR tetramer. A) The figure depicts a simplified representation of a ZFR tetramer bound to two Z-sites. The catalytic modules are depicted in orange, while three-fingered ZFA binding modules are depicted in maroon (the arm region and linker are not shown). The top strand of DNA (yellow) is bound by one ZFR dimer, while the bottom strand of DNA (green) is bound by another, and both Z-site bound dimers have formed a tetramer. The curved arrow beside the Z-site/tetramer complex on the right indicates that catalysis with subunit rotation has taken place, and it can be seen from the colour of the DNA that the Z-site half-sites have been rearranged. **B)** The figure depicts the three types of recombination reactions which can be produced by ZFRs: inversion, excision, and integration. For simplicity, the figure does not display the full complement of plasmid recombination possibilities (these are displayed in **Figure 1-10**). A plasmid is shown with three types of elements indicated: origin of replication (circle), Z-sites (yellow and green rectangles), and a gene (arrow) which indicates the directionality of the plasmid segment which contains it. If the ZFRs bring the Z-sites together directly across the plasmid, a recombination reaction will produce an inversion of the plasmid segment carrying the gene between the Z-sites. However, if the Z-sites are brought together in a reversed orientation in such as way that the plasmid becomes twisted, an excision of the plasmid segment carrying the gene between the Z-sites will take place. Alternatively, ZFRs may also mediate the reverse reaction and integrate a circularized segment of DNA containing a Z-site. Because both integration and excision reactions are possible, both reactions are inherently reversible within the ZFR system. However, since only one of the two circularized segments of DNA contains an origin or replication, the reaction equilibrium will tend toward an excision end product, as the circularized DNA segment without the origin of replication is typically lost as cells replicate.

1.5.2 ZFR studies to date

1.5.2.1 Introduction

Since the first publication describing a ZFR in 2003 (Akopian et al., 2003), efforts have been made to test and improve the system. Experiments have involved assessments of the ability of ZFRs to catalyse reactions in human cells, improvements to the ZFR architecture, and attempts to overcome the sequence-specificity limitations present in the catalytic module. So far only three groups have published ZFR papers (the Stark lab at the University of Glasgow, the Barbas lab at the Scripts Research Institute in California, and the Tamamura lab at Tokyo Medical and Dental University), and of those, only two have published multiple papers on the subject (Stark and Barbas), with most being published by the Barbas lab. With the unfortunate and untimely passing of Carlos Barbas in 2014, it now appears that only Stark's group remains active. However, the lab of Charles Gersbach (a former graduate student of Barbas and author on several ZFR papers), at Duke University, also appears to be actively pursuing ZFR research.

1.5.2.2 Activity in mammalian cells

ZFRs have been used to catalyse both excision and integration reactions in mammalian cells. In two experiments using a reporter cell line in which a genomically integrated cassette containing EGFP was flanked by Z-sites, ZFRs were shown to catalyse excision in 17% and 19% of cells that received ZFR expression constructs via retroviral transduction or plasmid transfection, respectively (Gordley et al., 2007, Nomura et al., 2012). The similarity in efficiency is notable, given that different methods of ZFR delivery and different cell types were used in the experiments (HEK293 cells versus CHO cells). Additionally, several successful experiments demonstrating ZFR- or TALER-mediated excision using episomal reporter assays in human cells, have also been reported (Mercer et al., 2012, Gaj et al., 2013b, Gaj et al., 2014). ZFRs have also been used to perform targeted integration (at both introduced and endogenous sites) in mammalian cells (HEK293 cells, HuH-7 cells, and NIH3T3 cells) with efficiency that ranged from 0.13–1.6% (percentage of successfully transfected cells with insertions), and specificity that range from 8.3–99% (percentage of detectable insertions found at the target site) (Gordley et al., 2009, Gersbach et al., 2011, Gaj et al., 2011, Gaj et al., 2013b, Gaj et al., 2014). Notably, the ZFR-mediated integration efficiency in HEK293 cells is comparable to that reported for ϕC31 integrase (Thyagarajan et al., 2001). This comparable level of

integration efficiency is perhaps somewhat surprising, given that ϕC31 integrase produces a non-reversible integration reaction, while ZFR integration is reversible and excision is the stochastically favoured outcome (see Figure 1-9, and Section 1.5.3).

1.5.2.3 Architecture optimization

Various experiments have been conducted to optimize or alter the architecture of chimeric recombinases with programmable binding domains including: the inclusion of catalytic modules from a variety of small serine recombinases, alteration of the targeting module, linker and Z-site central sequence length optimization, and the creation of obligate heterodimers. Experiments have demonstrated the successful construction of active ZFRs utilizing catalytic modules from hyperactivated variants of Tn3 and Sin resolvases, Hin and Gin invertases, and β recombinase (Akopian et al., 2003, Gordley et al., 2007, Sirk et al., 2014). These catalytic modules vary in efficiency and offer differing sequence-specificity requirements.

Attempts to vary the number of zinc finger motifs in the ZFA binding module have demonstrated that ZFAs comprised of four or five fingers may impart ZFRs with higher efficiency and specificity than ZFAs comprised of two, three, or six fingers (Nomura et al., 2012, Gordley et al., 2009). However, this interpretation may be a simplification of the underlying reality as various individual zinc finger motifs may contribute differing levels of affinity and specificity to a ZFA and the effect may be context-dependent (i.e. some motifs may cooperate or interfere with each other; see Section 1.7.1) (Bhakta et al., 2013). Furthermore, it has been suggested by some that longer ZFAs (>3 fingers) benefit from modifications to the ZFA architecture (Klug, 2010). Thus, the number of fingers in a ZFA will not always reliably correlate with affinity and specificity parameters. The ZFA binding module of the ZFR has also been replaced with a TALE binding module to generate a TALER (Mercer et al., 2012). TALERs show reduced recombination efficiency in human cells compared with ZFRs, bind longer target sites than ZFRs, and have been demonstrated to work successfully in ZFR/TALER heterodimer complexes.

The ZFR linker and Z-site central sequence length (the sequence that the ZFA binding sites flank) has also been a target for optimization (Akopian et al., 2003, Prorocic et al., 2011, Nomura et al., 2012). The two formats for Tn3-based ZFR linker and central sequence arrangement that are currently favoured are a SGS linker starting at Tn3 Thr145

with a 20 bp (or 18 bp if ZFAs have 3′ overlap specificity) Z-site central sequence (Gordley et al., 2007), or a TS linker starting at Tn3 residue Arg148 with a 22 bp (or 20 bp if ZFAs have overlap specificity) Z-site central sequence (Prorocic et al., 2011).

Finally, an obligate heterodimer architecture has been designed that allows ZFRs to form intended heterodimers at Z-site, while limiting the occurrence of unwanted off-target homodimer by-products of the reaction (Gaj et al., 2014). The enhanced ZFRs (eZFRs) were generated by making a complementary set of mutations within the ZFR dimer interface. These mutations discourage homodimerization by producing steric clashes between same-type subunits while accommodating alternative-type subunits during heterodimerization. In order to compensate for a reduction in recombination efficiency caused by the dimer interface mutations, the eZFR pairs also include an additional activating mutation that results in eZFR activity comparable to that of the parental non-obligate enzymes when assayed in HEK293 cells. These eZFR pairs show a >200 fold decrease in the formation of unwanted homodimers when compared with the non-obligate parental enzymes. In a test of targeted integration at an endogenous locus on human chromosome 4, the eZFRs were shown to be capable of improved targeted integration specificity and had an improved toxicity profile compared with their parental ZFR enzymes.

1.5.2.4 Catalytic module specificity alteration

One of the primary limitations of the ZFR system is the sequence specificity bias provided by the catalytic module. About half of the ZFR papers published to date have focused on this challenge and have described the creation of catalytic modules with either broadened or altered site-specificity profiles. Random mutagenesis of the ZFR catalytic module followed by selection, has been used to generate catalytic modules with broadened site-specificity that tolerate a range of base pair changes throughout the core of the Z-site central sequence that is contacted by the catalytic module (see Figure 1-8) (Gordley et al., 2007, Proudfoot et al., 2011, Gersbach et al., 2010). One drawback to this approach though, is that the extra site-specificity the catalytic module provides may be desirable in that it increases the overall site-specificity of the ZFR. Because ZFRs tolerate minor variations in Z-site core length (Nomura et al., 2012, Prorocic et al., 2011, Gordley et al., 2007, Akopian et al., 2003), it is also possible that broadening specificity could lead to reactions where the dinucleotide target of catalysis (see Section 1.6) at the centre of

the Z-site core deviates slightly from the intended position, which might lead to undesired codon changes in a targeted reading frame. Therefore, other studies have focused on switching the catalytic module site-specificity rather than abrogating it.

ZFR catalytic module variants with switched site-specificity have been generated through the use of semi-rational design, and the incorporation of catalytic modules from other small serine recombinases into the ZFR system. Site-directed saturation mutagenesis of specificity-determining DNA-binding residues of the ZFR arm region was used to generate Tn3 resolvase- and Gin invertase-based ZFRs that could recombine Gin- and Tn3-based Z-sites, respectively (i.e. they could recombine the non-cognate site), with the same level of efficiency as their parent enzymes on the cognate Z-site (Gaj et al., 2011). The core of the central sequences of the Gin invertase- and Tn3 resolvase-based Z-sites share only 30% sequence identity. These switched specificity ZFRs also showed an inability to recombine the cognate Z-sites of their parent enzymes, with a level of inhibition comparable to that of the parent enzymes on the non-cognate Z-sites. A pure rational design approach, involving swapping key specificity-determining arm region residues, has also been used to successfully switch the specificity bias of a Gin invertase-based ZFR to that of Tn3 resolvase-based ZFR (Sirk et al., 2014).

A subsequent study set out to generate ZFR catalytic modules that could target endogenous sites in the human genome (Gaj et al., 2013b). Gaj *et al.* first characterized the Z-site core sequence specificity tolerance of a Gin-based ZFR. This work showed complete single-base change tolerance in positions 10, 9, 8, and 7 (the four base pairs flanking the ZFA binding sites), single A to T base pair substitution tolerance at positions 6, 5, and 4, limited dinucleotide substitution tolerance at positions 3 and 2 (4 of 16 possibilities tolerated), and broad dinucleotide substitution tolerance in the central dinucleotide of the core sequence (i.e. positions -1 and 1; 12 of 16 possibilities tolerated). Random mutagenesis of the ZFR arm region residues responsible for contact at positions 2 and 3 was then used to generate Gin invertase-based ZFRs that could target five new dinucleotide substitutions at those positions. Using a combination of the Gin-base ZFR sequence specificity profile they had previously generated and the new ZFR catalytic modules, they then successfully demonstrated targeted integration at non-protein coding regions on chromosomes 4 and X, but with relatively low specificity (8–14%).

In order to overcome the A or T sequence requirements at core sequence positions 4, 5 and 6 detected in the preceding study, a subsequent study attempted to generate ZFRs from the catalytic domains of other small serine recombinases with different binding specificity biases (Sirk et al., 2014). The Sin resolvase and β recombinase were chosen as the basis for new ZFRs because they naturally prefer guanines and cytosines at positions 4, 5 and 6 within the core for their binding sites. The Sin resolvase and β recombinase catalytic modules were hyperactivated by a directed evolution strategy, and binding specificity profiles were generated as done in the previous study. A β recombinase-based ZFR was then tested in *E. coli* using split gene reassembly (see below) on 17 Z-sites with cores sequences based on human genomic loci, and was found to be active on six of them, with recombination efficiencies ranging from 3–95%.

Finally, the directed evolution strategy that was used to generate both broadened site-specificity catalytic domains (Gersbach et al., 2010), and switched site-specificity catalytic domains (Gaj et al., 2011, Gaj et al., 2013b, Sirk et al., 2014), itself has technical merit worth describing. This approach called 'split gene reassembly' involves the use of a 'substrate-linked' ZFR expression vector; that is, an expression vector that also acts as a ZFR substrate for recombination. The substrate-linked expression vector contains a β-lactamase gene (conferring resistance to β-lactam antibiotics such as carbenicillin) that is interrupted by a GFPuv cassette flanked by Z-sites. Recombination of the Z-sites deletes the GFPuv cassette, and reconstitutes the β-lactamase gene. Libraries of ZFR mutants are cloned into the substrate-linked expression vector, and active mutants cause deletion of the GFPuv cassette and restoration of β-lactamase function. After transformation of cells with a library of mutants encoded in the substrate-linked expression vector, and incubation, DNA is harvested from cells. The DNA can then be retransformed into cells, which are plated on selective (carbenicillin) and non-selective media, in order to ascertain the ratio of active to non-active mutants, and identify successful active mutant clones. Successful clones can then be amplified, and the ZFR coding sequences extracted for new rounds of selection in order to enrich the library. The ZFR recombination reaction time can also be varied by reducing the incubation time, in order to increase the site-recognition stringency of the enrichment procedure.

1.5.2.5 Remaining challenges

In order for ZFRs to become successful tools for targeted genomic editing, several problems with the system must first be overcome. Zinc finger recombinases are currently not able to specify the orientation of the integrations that they catalyse. Additionally, the recombination activity in the ZFR system is currently bidirectional; that is, the reactions ZFRs catalyse are reversible. This is problematic because one of the chief goals for the ZFR system is to use it for integration reactions. If an integration reaction is reversible, then the product of the reaction will not be stable, and the reverse excision reaction will be stochastically favoured (see Figure 1-9). Given that integration reactions can currently be catalysed with up to 1.4% efficiency in human cells (Gordley et al., 2009), while excision reactions can be catalysed with up to 19% efficiency (Nomura et al., 2012), there appears to be substantial room to improve the system.

It is perhaps surprising that integrations catalysed using the eZFR (obligate heterodimer) architecture (Section 1.5.2.3) do not result in a preponderance of one insertion orientation (Gaj et al., 2014). ZFRs may integrate DNA in one of two possible orientations. When ZFR heterodimers are used to perform an integration reaction, the product Z-sites flanking the insertions will be heterodimer-binding Z-sites if the DNA is integrated in one orientation, or homodimer-binding Z-sites if the DNA in integrated in the other orientation (see Section 1.5.3). Because the homodimer-binding Z-site products of one insertion orientation would be bound by incompatible eZFRs, this might have made those reactions irreversible, leading to a preponderance of that insertion orientation in the results, and this did not occur. This effect might also have been expected to produce a notable increase in integration efficiency, as half of the insertions produced by the eZFRs would be stable, and this also did not occur. In any case, the use of eZFRs does not appear to be a successful strategy for either controlling the orientation of insertions or increasing their stability.

Finally, although ZFRs with altered catalytic module sequence specificity have been generated, these often act with a reduced overall specificity compared to unmodified ZFRs (Gaj et al., 2013b), and thus, the sequence specificity bias of the catalytic module continues to present a challenge to the effective and flexible use of the ZFR system.

1.5.3 ZFR system parameters and outcomes: dimer-dimer orientation specificity and reaction directionality

Although three types of recombination reactions are possible using ZFRs (inversion, excision and integration), each has unique system configuration parameters that affect the outcome. These system parameters are primarily: the two possible orientations of the dimers with respect to one another during tetramer formation (Figure 1-10 A); and whether recombination reactions may be followed by a directionally reversed reaction. ZFRs rely on hyperactivated recombinase catalytic domains, which have lost much of the regulatory control that exists in the wild-type systems from which they were taken (see Section 1.6.4). Currently, neither dimer-dimer orientation during tetramer formation, nor reaction directionality may be controlled. These are important limitations to the current ZFR system as they greatly impinge on the reaction outcomes that may be generated, as will be discussed below.

Dimer-dimer orientation during tetramer formation, and reaction directionality, are both important parameters for inversion reactions (Figure 1-10 B). If the dimer-dimer orientation in a ZFR reaction is unconstrained, the generation of inversion end products will be limited to 50% efficiency, as 50% of the products of the reaction will be excision products. If the ZFR recombination reaction is reversible, production of inversion end products will also be limited to 50% efficiency, as 50% of the reaction products will be transformed back into their original state. Additionally, if a ZFR reaction has neither dimer-dimer orientation specificity constraint, nor reaction directionality control, then inversion end products will be heavily disfavoured as excision end products will dominate the reaction outcome (see below).

Although ZFR excision reactions are affected by both dimer-dimer orientation and reaction directionality, these parameters do not limit excision reactions in the way that they do other types of ZFR reactions (Figure 1-10 B). Excision reactions, alone, produce a stable end product in any ZFR reaction in which dimer-dimer orientation specificity is unconstrained and the reaction is bidirectional. If all of the ZFR reaction outcomes are reversible and there is no dimer-dimer orientation control, then the excision outcome in the reaction will be uniquely stable, as the replication of the cells which host the reaction will cause the excision product (which cannot replicate) to be lost. However, if a ZFR

system were used in which the reactions were irreversible, but there was no dimer-dimer orientation control, then the lack of dimer-dimer orientation control would limit the efficiency of excision reactions, as half the reaction outcomes would get stuck as inversion products.

In the case of ZFR-mediated integration reactions, both dimer-dimer orientation and reaction directionality are key parameters to the outcome (Figure 1-10 B). If dimer-dimer orientation during tetramer formation is not constrained, the orientation of an integrated DNA segment will be stochastically specified. In such cases, 50% of the reaction products will likely contain integrations in one orientation, and 50% will contain integrations in the other orientation. This is obviously an undesirable feature of the current ZFR system, as many foreseeable applications of the technology would require control over the orientation of the integrated genetic material. Reaction directionality is also a key parameter for ZFR-mediated integration reactions. If a ZFR integration reaction is reversible, then as mentioned above, reactions will tend towards excision products, as the excision product is more stable. This also an undesirable feature of the current ZFR technology because it likely greatly reduces the efficiency of integration reactions. Since integration reactions are perhaps the most important type of reaction the ZFR system could be used to catalyse, these limitations in the system represent important challenges to its utility. The need to provide exogenous DNA to be integrated also constitutes a system parameter for ZFR-mediated integration reactions.

It should be noted that in some recombinase systems, the central dinucleotide of the catalytic site will prevent reactions from completing in one of the two possible dimer-dimer orientations. If both bases of the central dinucleotide are the same, then the DNA can only be re-ligated if the reaction happens in one dimer-dimer orientation. However, the catalytic site of Tn3 resolvase contains an AT dinucleotide, and thus, the DNA may be re-ligated in either dimer-dimer orientation. Operating over the central dinucleotide is not considered to be a good option for controlling the orientation of reactions, such as integrations, in the ZFR system, because the DNA can still be cut, it just cannot be re-ligated.

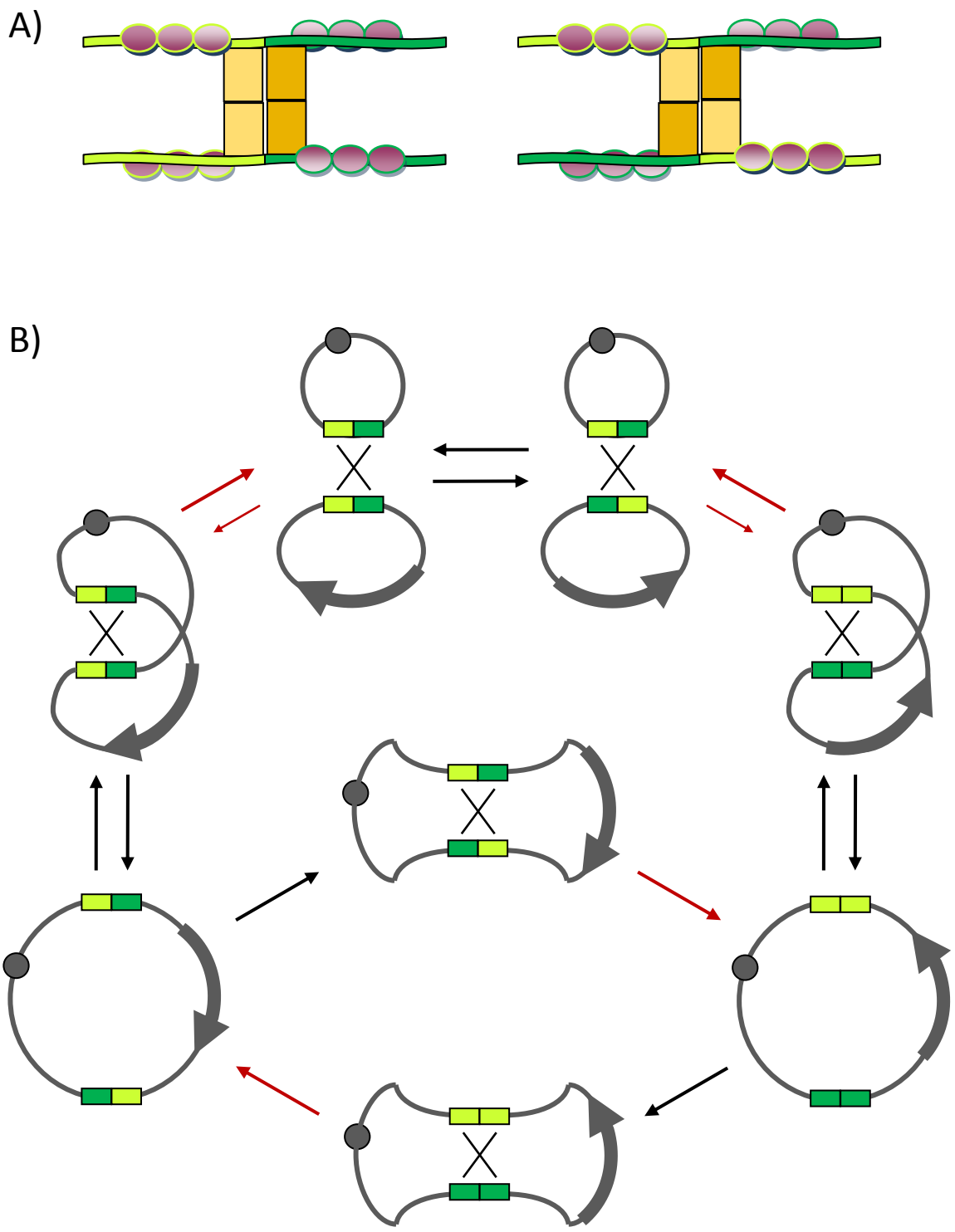


Figure 1-10: ZFR dimer-dimer orientation and reaction outcomes. A) A simplified representation of two ZFR tetramers bound to DNA, depicting both possible dimer-dimer orientations of the tetramer. The DNA is shown by yellow and green bars, the ZFA modules of the ZFR are depicted as maroon ovals, and the catalytic modules are depicted as orange rectangles. The colour of the DNA half Z-sites and the shade of the catalytic domains, provide the visual cues for recognizing the two possible orientations of the ZFR dimer-dimer interaction. **B)** A diagram showing all possible recombination pathways for ZFR reactions. The top half of the figure depicts integration and excision, and the bottom half of the figure depicts inversion. Plasmids are depicted in dark grey, with the origin of replication indicated by a circle, Z-sites depicted as yellow and green rectangles, and a gene, which provides a sense of orientation of the segment which contains it, is depicted with a thick curved arrow. X indicates a ZFR-mediated DNA recombination crossover event. Black straight arrows indicate steps in which the shape of the plasmids is altered in order to bring the Z-sites together, and/or which rearrange the orientation of the Z-sites with respect to one another. In the steps where the Z-sites are brought together, the colours of the half Z-site indicate the orientation of the dimer-dimer interaction of the ZFR tetramer, as the ZFRs themselves are not shown. Red arrows indicate steps after DNA recombination has taken place, with smaller arrows indicating less favourable reaction equilibrium.

1.6 Tn3 resolvase

1.6.1 Origin

Tn3 resolvase is encoded by the Tn3 transposon, which is involved in transmitting β-lactam antibiotic (e.g. ampicillin) resistance between bacteria (Lett, 1988). The Tn3 transposon is a 4957 bp mobile gene element containing three genes, *bla* (β-lactamase), *tnpA* (Tn3 transposase), and *tnpR* (Tn3 resolvase), which are collectively flanked by terminal inverted-repeats. Tn3 resolvase binds a site called '*res',* which overlaps the promoters of the *tnpA* and *tnpR* genes, repressing them. During replicative transposition, Tn3 transposase facilitates the integration of the donor plasmid (containing the transposon) into a host plasmid, to produce a 'cointegrate' plasmid. This process results in the replication of the transposon. The cointegrate plasmid contains two *res* sites, which allow Tn3 resolvase to perform a 'resolution' DNA recombination reaction, resulting in the formation of two plasmids, each containing one copy of the Tn3 transposon.

1.6.2 Structure of Tn3 resolvase

Tn3 resolvase is a 183 amino acid protein from the small serine recombinase family. It contains a C-terminal helix-turn-helix (HTH) DNA-binding domain connected by a flexible 'arm region' to a catalytic N-terminal domain. The catalytic domains is comprised of a long α-helix, known as the E-helix, which terminates around position 102, and is followed by a large N-terminal ββαβαβαβ-fold sub-domain (hereafter referred to as the α/β sub-domain) comprised of three α-helices (A, B, and D) and a β-sheet (see Figure 1-12 in Section 1.6.3). The mechanisms that regulate Tn3 resolvase are incompletely understood, although it is generally accepted that: Tn3 resolvase forms dimers on DNA; has three distinct dimer binding sites, which are separated by spacers, known as site I, II and III (collectively termed *res*); forms a complex DNA-bound structure involving three tetramers called the synaptosome (see below); cleaves the phosphodiester backbone of the DNA via nucleophilic attack from residue Ser10; and rearranges the cleaved DNA via subunit rotation (Figure 1-11). The synaptosome is comprised of dimers bound to two res sites, two of which form the catalytic module bound to site I, and four of which form a regulatory module bound to sites II and III (see (Rowland et al., 2009) and (Rice et al., 2010) for more information on the small serine recombinase synaptosome structure). There is also strong evidence to support an activating interaction between the catalytic

tetramer and the accessory tetramers of the synaptosome, at a location known as the 2-3′ interface of the N-terminal domain (Hughes et al., 1990, Murley and Grindley, 1998).

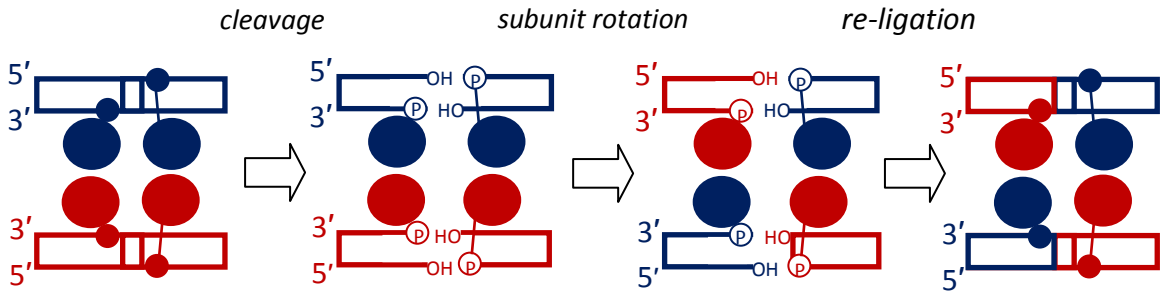


Figure 1-11: Subunit rotation model of Tn3 resolvase recombination. DNA bound dimers (drawn in blue and red) come together to form a tetramer. Synapsis is followed by strand cleavage via nucleophilic attack of the DNA phosphodiester backbone by Ser10, which results in a covalent bond between DNA and protein. A 180° subunit rotation then occurs around the central vertical plane, temporarily separating the dimer interfaces. The DNA is rejoined and strand exchange is complete. Adapted from Olorunniji and Stark (2009).

1.6.3 Crystal structures of γδ resolvase

Tn3 resolvase has a close homologue called γδ resolvase with approximately 80% sequence similarity. However, most of the sequence difference lies in the C-terminal binding domain, and throughout the catalytic domain there are very few amino acid changes. There are several crystallographic structures of γδ resolvase, which greatly aid in the understanding of Tn3 resolvase structure and function. Two structures of particular note are the 1GDT dimer structure (Yang and Steitz, 1995), and the 1ZR4 tetramer structure (Li et al., 2005) (see Figure 1-12). The 1GDT dimer structure depicts resolvase in an 'inactivated' form, having not yet undergone necessary conformational changes that lead to tetramer formation and strand cleavage. The 1GDT structure contains useful information regarding the dimer interface, as well as depicting the DNA-binding contacts to site I. Conversely, the 1ZR4 structure depicts an activated resolvase tetramer that is covalently bound by Ser10, post-cleavage, to the phosphodiester backbone of the DNA. Several large conformational changes are seen in the tetramer structure compared with the dimer structure, including a large change in the position of the N-terminal domain that brings it within reach of the DNA backbone, and a splaying of the E-helices that now make the primary contacts of the tetrameric interface between antiparallel-oriented E-helices of counterpart dimers (see Figure 1-13).

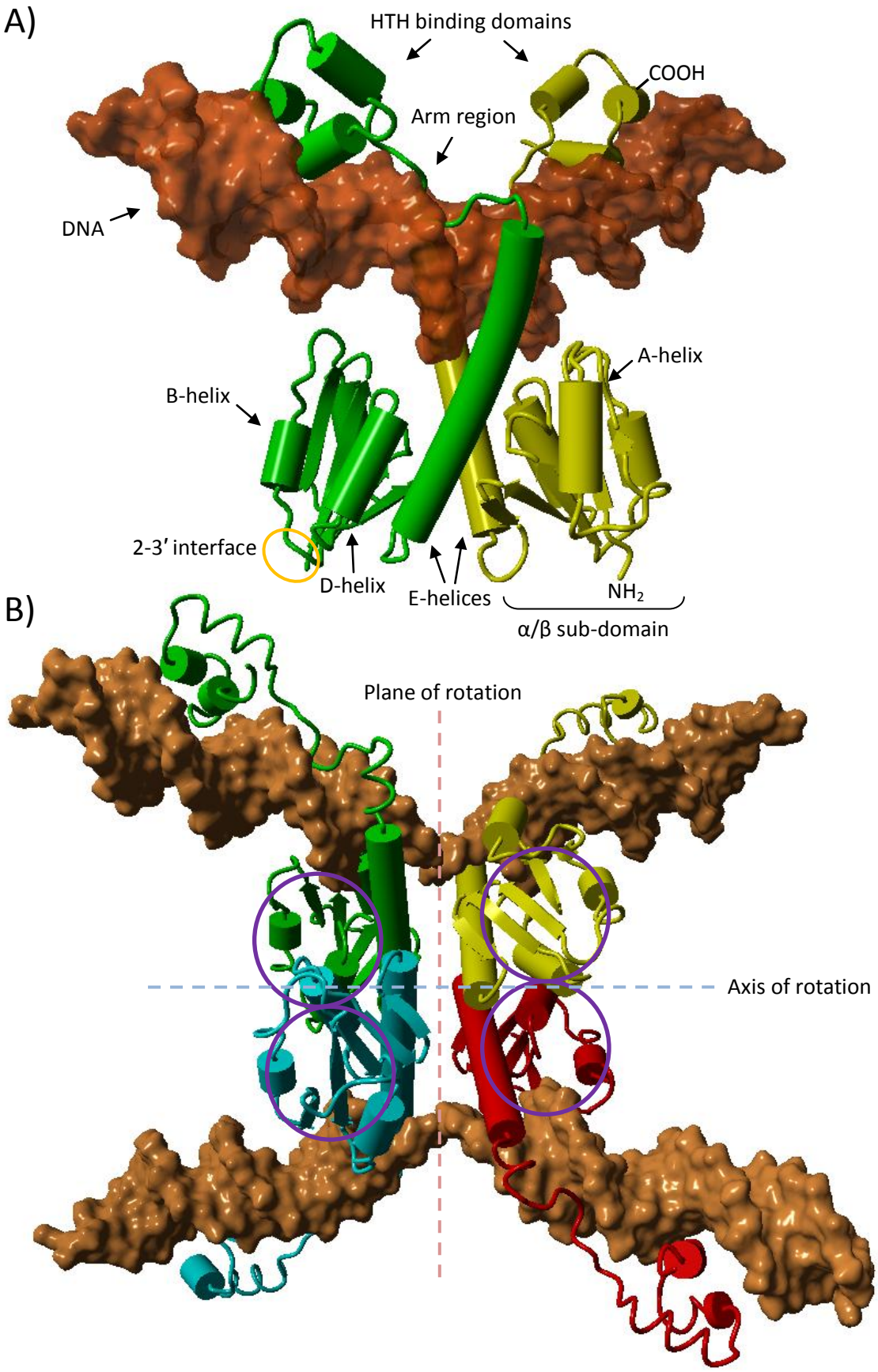


Figure 1-12: Crystal structures of γδ resolvase. The resolvase subunits are displayed in cartoon using green, yellow, cyan, and red. The DNA is shown using molecular surface and is orange-brown. **A)** The 1GDT structure shows γδ resolvase bound to DNA (site I) as an inactive dimer. **B)** The 1ZR4 structure shows γδ resolvase in an activated tetramer conformation (again bound to site I) with Ser10 covalently linked to phosphate at the 5′ ends of the cleaved DNA backbone. The purple circles indicate the approximate location of a residue cluster which facilitates catalysis. Adapted from Olorunniji and Stark (2009).

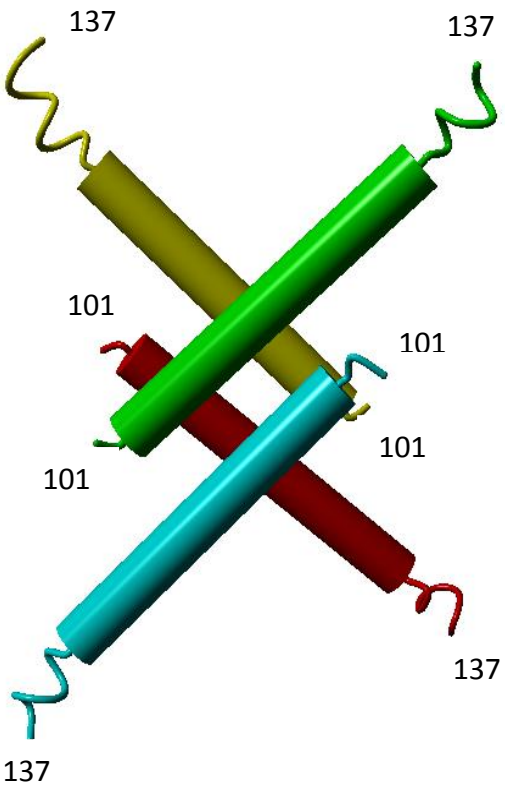


Figure 1-13: Antiparallel interaction of E-helices across the tetrameric interface. The E-helices of the subunits of dimer pairs (displayed in yellow and green, and red and blue) make contacts with their counterparts in antiparallel fashion creating a two-fold symmetry from the centre of the interaction. Residue numbers are given to indicate where the portion of the structure which is displayed terminates. Adapted from Kamtekar et al. (2006).

1.6.4 Hyperactive Tn3 resolvase mutants

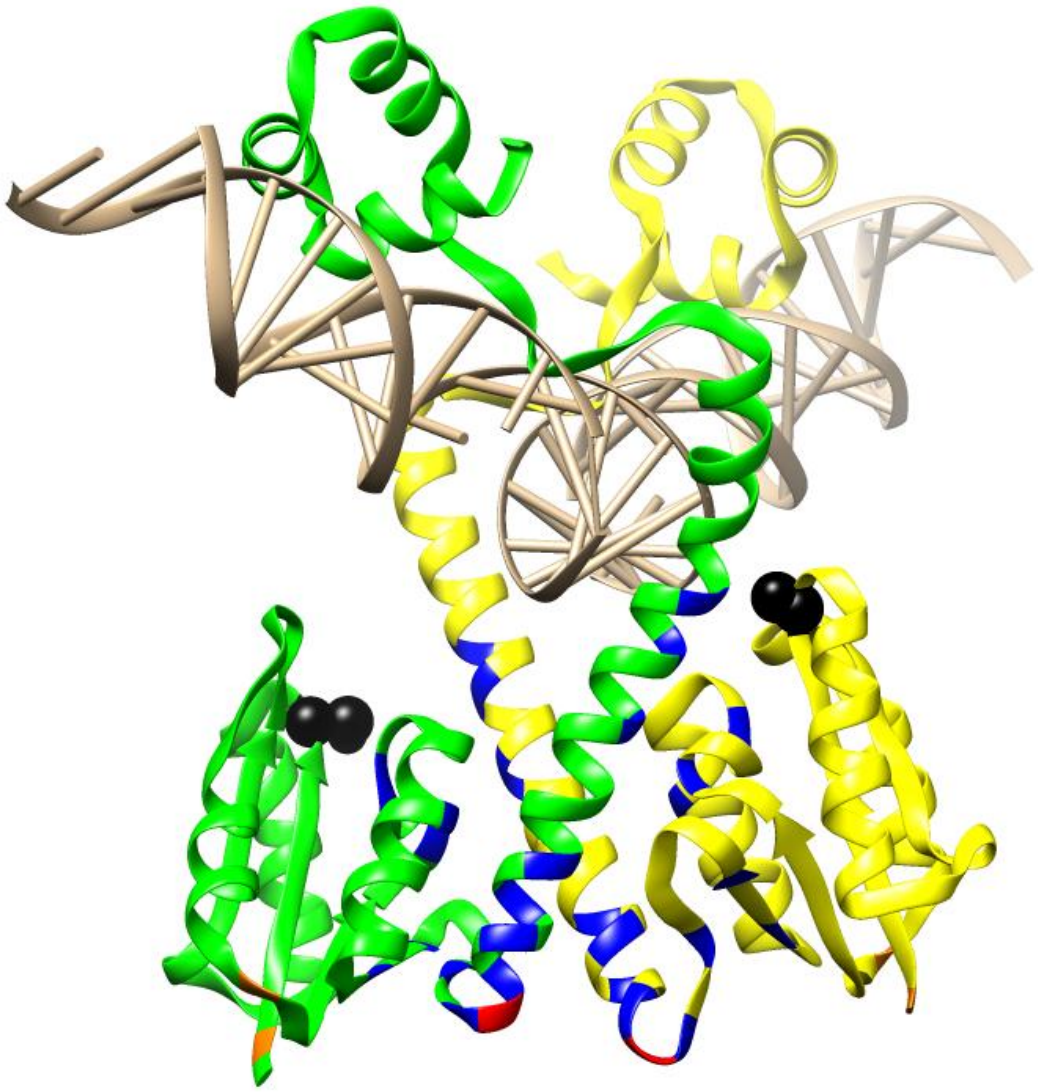
Previous work within the Stark lab has identified several mutations that lead to a so-called 'hyperactive' phenotype (Burke et al., 2004). Originally the term 'hyperactive' was used to describe Tn3 resolvase mutants that had the capability to produce recombination of either a site I x *res* substrates, or a site I x site I substrates. However, this term has since fallen out of favour within the Stark lab, and will be repurposed within this work to specifically describe Tn3 resolvase mutants that have the ability to recombine site I x site I substrates, with lesser levels of activation simply being referred to as 'activated'. This definition of 'hyperactive' specifically implies that the resolvase no longer has any need of the subunits bound to accessory sites, and thus, is the activation threshold where mutations that knock out the 2-3′ interface increase the activity of the protein rather than abolishing it (see Section 1.6.5). Tn3 resolvase always requires at least two mutations for hyperactivity. The most important mutation is made at Asp102, and a tyrosine is usually used for this. There are a number of secondary mutations, which complement D102Y to produce hyperactivation, located in various regions throughout Tn3 resolvase (see the following Section 1.6.5). Commonly within the Start lab, the hyperactivation mutation set of choice is R2A + E56K (collectively termed 'N') and G101S +

D102Y + M103I + Q105L (collectively termed 'M'), which were derived in the Burke *et al.* study (Burke et al., 2004).

Hyperactivation greatly simplifies the requirements for the recombination reaction and makes hyperactivated Tn3 resolvase an attractive potential tool for directing genetic modification. However, much of the regulatory function of Tn3 resolvase is provided as a result of the complex interactions in the wild-type synaptosome utilizing the resolvase subunits bound to the accessory sites. For example, the synaptosome-substrate arrangement ensures the recombination reaction between the sites of DNA catalysis is unidirectional, and also specifies the dimer-dimer orientation within the catalytic tetramer. By freeing the resolvase of its cumbersome accessory subunit aids, these features are lost as well.

1.6.5 'Primary', 'secondary', and 'tertiary' activating mutations

For the purpose of this project, it will be helpful to place the known activating mutations into three categories based on their behaviour in the mutant screen performed by Burke et al. (2004) (Burke et al., 2004): 'primary' activating mutations, 'secondary' activating mutations and 'tertiary' activating mutations (Figure 1-14). Primary activating mutations are any of the activating amino acid substitutions at position 102, which are required in order to hyperactivate Tn3 resolvase (e.g. D102Y). Secondary activating mutations are those that are required in addition to the primary activating mutation, and these can be found at various positions throughout the protein (e.g. G70C, G101S, M103I, and Q105L). Tertiary activating mutations are those that will produce an activating effect, but only if the hyperactivation threshold has already been met through the inclusion of primary and secondary activating mutations (e.g. R2A and E56K). Tertiary mutations will, in fact, kill the activity of Tn3 resolvase if the hyperactivation threshold has not been achieved. The reason that tertiary activating mutations can kill the activity of the resolvase is that these mutations knock out the 2-3′ interface interaction that allows the regulatory module of the synaptosome to activate the catalytic module. Thus, without hyperactivation, Tn3 resolvase still requires the 2-3′ interface in order to function (Burke et al., 2004).



Primary (red)	Secondary (blue)	Tertiary (orange)	Catalytic (black spheres)
D102Y/I/F/V/T/W	L66I/F, G70A/C, I77T, A89T, F92S, T99S, G101S/C, M103I/V, Q105L, V107M, T109I, A117V, R121K/M/S, E124Q	R2A, E56K	S10 (A)

Figure 1-14: Locations of Tn3 resolvase activating mutations and categories. The figure shows a γδ resolvase dimer bound to site I (PDB: 1GDT) in order to indicate the locations of activating mutations of Tn3 resolvase (a close homologue of the γδ resolvase). The ribbon representation of the two resolvase dimer subunits (green and yellow) is coloured to indicate the residue locations of activating mutations. A table below the image indicates the identities of activating mutations, and provides the categories they have been placed into within this work. The table is used to indicate the identities of the residues as there are too many to clearly label in the figure. The catalytic residue Ser10 is also indicated in the figure. An S10A mutation at this location is often used to inactivate the resolvase. This list of activating mutations was derived from the work of Burke et al. (2004).

1.7 Binding domains

1.7.1 Zinc finger arrays

1.7.1.1 Introduction

Developments in the engineering of ZFA DNA-binding domains with designed site-specificity now permit the design of chimeric enzymes that can be targeted to unique locations within a genome. These ZFAs have been used to create designer transcription factors, as well as genetic manipulation tools such as ZFNs and ZFRs (Gersbach et al., 2014). Zinc finger proteins are a class of naturally occurring DNA-binding proteins. The ZFAs that are constructed for targeted genomic editing purposes are primarily based on Zif268-like structures. Zif268 (also known as Egr-1, NGFI-A, and others) is a transcription factor from mice, although it has homologues in many species, including humans (EGR1) (Sukhatme et al., 1987, Sukhatme et al., 1988).

1.7.1.2 Structure and DNA binding

Zif268 is the archetypal ZFA and is comprised of three tandem zinc finger motifs that adopt a ββα fold (Figure 1-15 A and B) (Pavletich and Pabo, 1991). This ββα fold is comprised of about 30 amino acids and is tetrahedrally coordinated by a zinc ion bound to two Cys and two His residues; hence, this family of proteins is referred to as Cys_2-His_2 zinc finger proteins. The crystal structure of Zif268 (Elrod-Erickson et al., 1996) reveals that each zinc finger module in the array binds specifically to 3–4 bp of its target sequence through amino acid contacts to bases via the major groove of the DNA (Figure 1-15 C and D). However, there is overlap in these protein-DNA contacts, with Finger 3 supplying cooperative contacts to the first base pair bound by Finger 2, and Finger 2 supplying cooperative contacts to the first bp bound by Finger 1. Thus, Zif268 makes specific contacts to 10 bp DNA.

Often, zinc finger modules are described as binding up to three base pairs specifically, but cooperative contacts to a fourth base pair, such as observed in Zif268, sometimes also need to be considered. The base pair contacts are made by four zinc finger recognition helix residues at positions conventionally described as -1, 2, 3, and 6 (with contacts from position 2 mediating fourth base pair overlapping specificity; Figure 1-15 C). Additionally, residues at positions 4 and 5 may help to enforce target specificity as well (Gersbach et al., 2014). It should also be noted, that although the foregoing is the

generally reported pattern of residue-base contact for Zif268-like ZFAs, other quite different patterns of contact (e.g. specification for different base positions by residues -1 and 2, base contacts from residue 1, and fifth base pair overlap specificity) have also been observed for engineered Zif268-like ZFAs (Wolfe et al., 2001). Thus, alteration of the key residues within the zinc finger recognition helix can facilitate the creation of zinc finger motifs with novel DNA binding specificities. It is also worth noting, that at least two studies have demonstrated that some natural three-fingered ZFAs, including Zif268, have an extended recognition specificity preferences for the base pairs directly 5′ and 3′ of the sequence to which protein-DNA contacts are made (especially the 5′ base pair; hereafter referred to as the 5′ extended base position) (Swirnoff and Milbrandt, 1995, Wolfe et al., 1999). In Zif268, this extended base pair specificity is not mediated by such obvious contacts as the core 10 bp of recognition sequence. Engineered Zif268-like ZFAs can be constructed by linking individual zinc finger motifs together using canonical Zif268 linkers TGEKP and TGQKP, and the construction of arrays with up to six-fingers have been reported (Liu et al., 1997).

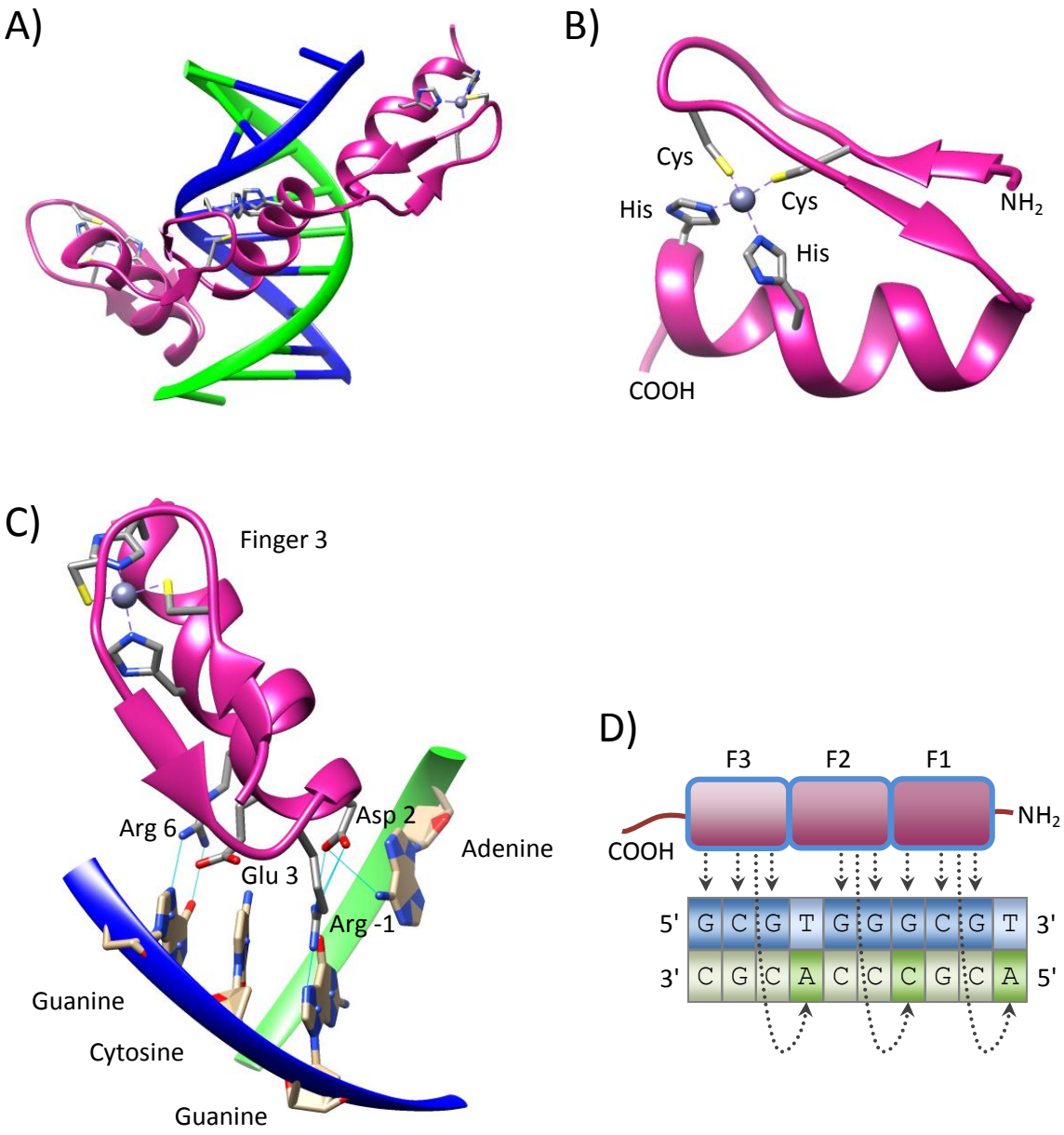


Figure 1-15: The structure and DNA binding of the Zif268 ZFA. A) Crystal structure of Zif268 bound to its target site (PDB: 1AAY). **B)** The ββα fold conformation of Zif268 with tetrahedral coordination by a Zn^{2+} ion bound by two Cys and two His ligands. **C)** The Zif268 recognition helix from finger 3 of the array, bound to its target site. Recognition helix residues Arg 6, Glu 3, and Arg 1 are shown making contacts with bases guanine, cystosine, and guanine, respectively, while Asp 2 is shown making 'overlap' contact to adenine on the opposite DNA strand (the first base in the target of finger two; finger 2 is not shown). Hydrogen bonds are shown using light blue lines (there are two hydrogen bond contacts from Arg -1 to guanine but they cannot be easily seen due to the orientation of the structure). Asp 2 can also be seen providing stabilizing hydrogen bonds to Arg -1. Glu 3 only provides specificity through hydrophobic contact, to cytosine. Non-contacted bases have been removed for clarity. **D)** A schematic representation of Zif268 binding to its 10 bp target site. Bases contacted by finger 1, 2, and 3 (F1, F2, and F3) are shown in darker colours.

1.7.1.3 Construction methods

The labs of Aaron Klug (University of Cambridge, UK), Carlos Barbas (The Scripps Research Institute, USA), and Keith Joung (Harvard University, USA), as well as others, have over the past 20 years generated a wealth of published data on building engineered ZFAs from customized Cis_2-His_2 zinc fingers with designed site-specificity. Customized zinc finger modules are generated through the use of phage display, rational design, and exploitation of naturally occurring zinc finger proteins, and these techniques have been used to generate zinc fingers specifying all GNN, ANN, CNN, and several TNN DNA triplets (Segal et al., 1999, Dreier et al., 2000, Dreier et al., 2001, Dreier et al., 2005, Blancafort et al., 2003, Liu et al., 2002, Bae et al., 2003, Segal et al., 2003, Beerli et al., 1998, Carroll et al., 2006, Desjarlais and Berg, 1992, Sera and Uranga, 2002). Additionally, for triplets for which there are no available zinc finger modules, a 'structured linker', which can skip a triplet, can be generated by mutating all of the base contacting residues of the recognition helix (-1, 2, 3, and 6) to serine; however, this strategy does not appear to have been deployed outside of its original publication (Moore et al., 2001a, Klug, 2010). Many of the known zinc finger modules have been amassed into publicly available online databases that allow users to input desired DNA target sequences and receive an output of amino acid sequences for fingers that bind to the DNA targets (Mandell and Barbas, 2006, Fu et al., 2009). However, for some individual triplets there may be several available zinc finger modules available, and not all of them are necessarily active in combination with one another. Thus, several methods exist for assembling functional ZFAs, including modular assembly, modular assembly with specialized libraries, and context-sensitive assembly (Ramirez and Joung, 2013).

The simplest method for ZFA construction is modular assembly (Bhakta and Segal, 2010). Using this method one simply links together the known zinc finger modules and tests the ZFAs against target sites. Although the modular assembly method of ZFA construction is simple, in some cases high failure rates have been reported (e.g. 76% average failure rate, and 100% failure rate for the ZFA lacking GNN modules) (Ramirez et al., 2008, Joung et al., 2010). It has been suggested that these high failure rates result from a lack of consideration of context-sensitive effects (i.e. target site overlap and side-chain interference between neighbouring zinc finger modules) in the modular assembly approach (Ramirez et al., 2008, Joung et al., 2010, Cathomen and Joung, 2008). However, there is also evidence to suggest that modular assembly failure is the result of insufficient

cumulative affinity of some combinations of zinc finger modules (Sander et al., 2009, Lam et al., 2011), and thus, the problem may be corrected simply by constructing longer ZFAs (Bhakta et al., 2013, Kim et al., 2009), or by affinity balancing through a module affinity-scoring scheme (Sander et al., 2009).

One approach to improving the success rates of ZFAs constructed by modular assembly has been the generation of specialized libraries composed of select zinc finger modules that have a higher frequency of occurrence in successful ZFAs (Kim et al., 2009, Kim et al., 2011, Zhu et al., 2011), or which are known to work well with each other (Gupta et al., 2012, Sander et al., 2011b, Zhu et al., 2013). The benefit of these specialized library strategies is that they are simple, and they achieve high rates of success compared with traditional modular assembly strategies. The Context-Dependent Assembly (CoDA) library, for instance, produces ZFAs with over 76% chance of activity (as measured by >3-fold transcriptional activation in a B2H assay) (Sander et al., 2011b). The drawback to using specialized libraries is that they are much more limited than traditional modular assembly in the number of sequences that can be hypothetically targeted.

Several approaches exist to generate ZFAs through context-dependent selection and assembly; these include: sequential optimization (Greisman and Pabo, 1997), bi-partite selection (Isalan et al., 2001), bacterial two-hybrid (B2H) selection (Hurt et al., 2003), and bacterial one-hybrid (B1H) selection (Meng et al., 2008). These approaches will be briefly described below.

Sequential optimization involves successive rounds of *C-terminal* ZFA finger truncations and *N-terminal* ZFA finger additions to generate a completely novel ZFA with a position-specific finger arrangement (i.e. the fingers used at each position in the array were designed for that position only) (Greisman and Pabo, 1997). First, three libraries of Zif268-based ZFAs are generated, where one module at each of the three finger positions (F1, F2 and F3) has had the key residues of its recognition helix randomized. These libraries are selected for site-specificity using phage display. Randomized F1 modules are then used to replace the Zif268 F3 module of the ZFA (generating an F1-F2-**F1** array), and the new ZFAs are tested for site-specificity using phage display. Next, the Zif268 F1 module of the new ZFAs is removed, and randomized F2 modules are fused to the opposite end of the array, connecting them to the randomized F1 modules (generating an

F2-**F1-F2** array). After further selection, a similar step is carried out to remove the last Zif268 module and fuse randomized F3 modules to the randomized F2 modules (generating a **F1-F2-F3** array). These ZFAs are then selected for the desired final site-specificity, and thus, novel ZFAs with a F1-F2-F3 configuration are generated.

There are two finger-finger junctions within a three-fingered ZFA. In bi-partite construction, a three-finger ZFA is parsed into two 1.5 finger units, each of which *contains* those finger-finger junctions (Isalan et al., 2001). Two libraries of Zif268-based ZFAs, containing amino acid randomization at the base-contacting residues of either F1 and half of F2, or (the remaining) half of F2 and all of F3, are selected for desired site-specificity by phage display. The novel 1.5 finger units are then recombined and selected on the desired target site, again using phage display. Thus, the bi-partite method of construction accounts for compatibility and overlapping base specificity between adjacent fingers, by avoiding construction steps that create unproven finger-finger junctions.

B2H selection is an *E. coli*-based method which makes use of ZFAs that are tethered to a Gal11P protein, and an RNA polymerase α-subunit tethered to a Gal4 protein (Gal11P and Gal4 are found in yeast) (Hurt et al., 2003). Gal11P binds to Gal4, allowing the ZFA-Gal11P fusion to act as a transcription factor (because it recruits the Gal4-RNA polymerase α-subunit fusion) when correctly bound to a site upstream of a reporter gene. The selection begins with three BCR-ABL ZFAs (a previously constructed ZFA based on three modified repeats of the Zif268 finger 2) (Choo et al., 1994), where either finger 1, 2, or 3 of BCR-ABL has been replaced with a novel module library. These BCR-ABL-based ZFAs with new modules at position 1, 2, or 3 of the array, are then selected for binding activity using the B2H assay. The successful fingers from the first round of selection are then recombined in a position-specific fashion to generate entirely novel ZFAs, which are tested for activity, again using the B2H assay (Hurt et al., 2003). A B1H-based construction system also exists, which is like the B2H strategy, but uses a canonical Zif268 ZFA instead of BCR-ABL as the starting ZFA framework, and employs a selection system where the ZFAs are directly tethered to the ω-subunit of RNA polymerase (Meng et al., 2008).

Currently, the most popular publically available system for context-specific construction is Oligomerized Pool ENgineering (OPEN), which is base on the B2H methodology (Maeder et al., 2008). The primary advance over the original B2H system is the provision of large plasmid libraries of context-specific modules archived by the Zinc Finger Consortium (www.zincfingers.org) and made available through Addgene (Cambridge, MA, USA; www.addgene.org/zfc). Additionally, the Zinc Finger Consortium also supplies an online tool (ZiFiT) to help streamline target-site identification, and this tool includes an affinity-based activity scoring algorithm to predict outcomes (Sander et al., 2010, Sander et al., 2007). Technically, all of the ZFAs produced using OPEN are active, because the final selection stage tests for their activity in a B2H assay (successful ZFAs are ones detected to produce >3-fold activation of transcription of a reporter gene in *E. coli*) (Maeder et al., 2008). However, whether these ZFAs that are active in *E. coli*, are active when used in fusion proteins, perhaps in mammalian cells, is another matter. Maeder *et al.* demonstrated that 75% of ZFNs incorporating OPEN ZFAs showed activity in human cells, compared to 18% of the ZFNs incorporating ZFAs constructed by traditional modular assembly. The construction of ZFAs through OPEN is more complicated than construction through modular assembly, but whether it is more labour intensive is debatable given the potential failure rates of traditional modular assembly ZFAs.

Finally, Sangamo Biosciences (Richmond, CA, USA) has developed a proprietary archive of zinc finger modules and proprietary method for constructing ZFAs. Custom ZFAs, in the form of a ZFN pair, can be purchased through Sangamo's licensed partner Sigma-Aldrich (St. Louis, MO, USA) under the CompoZr™ product line for $6000 (Segal and Meckler, 2013). Sigma-Aldrich claims to have readymade ZFN pairs that target all genes in the human, mouse, and rat genomes (Segal and Meckler, 2013). The Sangamo ZFAs are reported to be highly active and have featured in most of the reports of ZFN-mediated endogenous gene modification to date (Urnov et al., 2010, Perez-Pinera et al., 2012), including the first clinical trial of a ZFN-based gene therapy (see Section 1.4.3) (Tebas et al., 2014).

Some of the Sangamo design strategy can be gleaned from papers published by senior scientists who worked for the company (e.g. Sir Aaron Klug and Yen Choo who also founded the zinc finger technology company Gendaq, which was later bought by Sangamo), and from analysis of published reports that describe Sangamo ZFA architecture

(Ramirez and Joung, 2013, Klug, 2010). Sangamo utilizes a library of pre-validated two- and three-finger arrays that have been designed using phage display and the bi-partite construction method (which accounts for cross-contact interactions) (Isalan et al., 2001, Isalan and Choo, 2001, Jamieson et al., 2003, Klug, 2010). The two-finger arrays can be obtained from a library of three-finger arrays constructed using bi-partite assembly (Klug, 2010). These two- and three-finger arrays are then fused using extended linkers (containing one or four residue additions within canonical Zif268 linkers) to generate four- to six-finger ZFAs (Moore et al., 2001b, Jamieson et al., 2003, Perez et al., 2008, Hockemeyer et al., 2009, Klug, 2010). Using ZFAs comprised of independent sub-arrays, separated by extended linkers, produces better target discrimination (Moore et al., 2001b). These ZFAs have better target discrimination because the K_d of two-fingered sub-arrays is more greatly affected by target site mismatches than that of longer contiguous arrays. Inclusion of a single Gly or Ser into a canonical Zif268-derived linker (e.g. TG**G**QKP) also serves to reset the register between zinc fingers and DNA, as the periodicity of the ZFAs does not quite match the periodicity of DNA, and strain accumulates when more than three fingers bind DNA (Moore et al., 2001b, Jamieson et al., 2003, Nekludova and Pabo, 1994, Perez et al., 2008, Hockemeyer et al., 2009, Klug, 2010). Inclusion of GGGS into a canonical linker (e.g. TG**GGGS**QKP) can reset the register and skip a base in the target sequence (Hockemeyer et al., 2009, Jamieson et al., 2003, Moore et al., 2001b). Using extended linkers that skip a base can also help accommodate sub-arrays with recognition helix residue 2 overlap specificity (see Section 1.7.1.2) incompatible with the following sub-array (Klug, 2010). It can also be observed that Sangamo ZFAs sometimes contain linkers within which the canonical Lys has been substituted for an Arg (Hockemeyer et al., 2009), and this perhaps serves some function.

The site-specificity of Sangamo ZFAs is thought to generally be higher than that of ZFAs produced by other methods. In one study that compared a pair of high-performing Sangamo ZFNs (based on four-fingered ZFAs; the same pair used in the HIV clinical trials, see Section 1.4.3.2) to a pair of high-performing OPEN ZFNs (based on three-fingered ZFAs), it was found that the Sangamo ZFNs may have 37 potential off-target sites in the human genome, while the OPEN ZFNs may have 2,652 (Pattanayak et al., 2011). In another study, CompoZr ZFNs (Sangamo via Sigma-Aldrich) were compared to ZFNs with CoDA derived ZFAs (Sood et al., 2013). This study showed that the CoDA ZFNs required a five-fold higher mRNA dose to achieve mutagenesis, and the somatic lesion frequency

produced by CoDA ZFNs was <5%, compared with 9-98% using the CompoZr ZFNs. However, a protocol has been published for optimizing ZFAs constructed through public methods, which may help improve their specificity (Herrmann et al., 2011). This protocol is based on semi-randomization of the key residues in the recognition helices, and selection using an adaptation of a commercial yeast one-hybrid kit. Crucially, this method overcomes a limitation of many publically available methods, allowing users to add additional fingers to their constructs to create ZFAs of four to six fingers in length (OPEN, CoDA and bi-partite construction, for example, allow only three-fingered ZFAs to be constructed).

1.7.2 TALE array binding domains

1.7.2.1 Introduction

Although methods to engineer custom ZFAs have improved greatly over the last 20 years, they are still limited in the range of sequences that they can target. Recently, TALE binding domains, which have much more versatile targeting capability than ZFAs, have been used to build chimeric site-specific enzymes such as TALERs and TALENs (Mercer et al., 2012, Joung and Sander, 2013). TALEs are transcriptional activators derived from bacterial plant pathogens of the genus *Xanthomonas* (TALE-like proteins are also found in phytopathogenic bacteria of the genus *Ralstonia*) (Kim and Kim, 2014), which bind and regulate genes within host plants (Boch and Bonas, 2010, Bogdanove et al., 2010, Kay and Bonas, 2009, Kay et al., 2007, Romer et al., 2007).

1.7.2.2 Structure

The TALE binding domain is comprised of arrays of 33–35 amino acid repeats (with the exception of the final repeat that is a half repeat) each of which recognize a single base pair (Mak et al., 2012, Deng et al., 2012a). These TALE repeat domains specify bases through contacts made by hypervariable residues at positions 12 and 13, known as the repeat-variable diresidues (RVDs) (Boch et al., 2009, Moscou and Bogdanove, 2009) RVDs recognize DNA via a simple recognition code (Boch et al., 2009, Moscou and Bogdanove, 2009, Cong et al., 2012, Streubel et al., 2012, Yang et al., 2014, Miller et al., 2015), which is an attractive feature for engineering site-specific binding domains. The TALE repeats take the form of two helices connected by a short loop containing the RVD (Mak et al., 2012, Deng et al., 2012a). The RVD loop sits in the major groove of the DNA, and the array

of TALE repeats coils around the DNA molecule in a right-handed superhelical structure (Figure 1-16).

1.7.2.3 Using TALE binding domains for chimeric enzyme creation

In order to use TALE binding domains for chimeric enzyme creation, a truncation architecture is usually designed to reduce the size of the protein and reveal the minimal domain required for efficient DNA binding (Mussolino et al., 2011, Miller et al., 2011, Mercer et al., 2012, Bedell et al., 2012, Christian et al., 2010). The truncation architectures also remove unnecessary native TALE elements such as the C-terminal transcriptional activation domain, and the N-terminal region up to residue 152 specifying transport into plant cells. These truncation architectures must account for the fact that different fusion points will lead to different degrees of proper protein folding and DNA binding, and they must be compatible with the catalytic domain that they are to be fused to. TALEN fusion protein designs include the addition of an N-terminal NLS (the native C-terminal NLS is usually removed with the truncations), while the only published TALER design does not include an NLS. It may be noted that full length TALEs (no truncation architecture) may also be used to generate TALENs (Li et al., 2011b, Li et al., 2011c).

Because TALENs have the binding module fused to the catalytic domain by its C-terminus, while TALERs have the binding module fused to the catalytic domains by its N-terminus, the optimal truncation architectures might not be directly compatible between both systems. However, the TALER designed by Mercer *et al.* used the C-terminal TALE truncation point from a TALEN designed by Miller *et al.*, and selected an N-terminal truncation point similar to that of the TALEN construction by Mussolino *et al.* (Mercer et al., 2012, Miller et al., 2011, Mussolino et al., 2011). An appropriately designed amino acid linker to mediate fusion to the catalytic domain must be also generated, in addition to the use of appropriate spacers between TALE binding sites within the target sequence, with consideration given to the chimeric enzyme system being used.

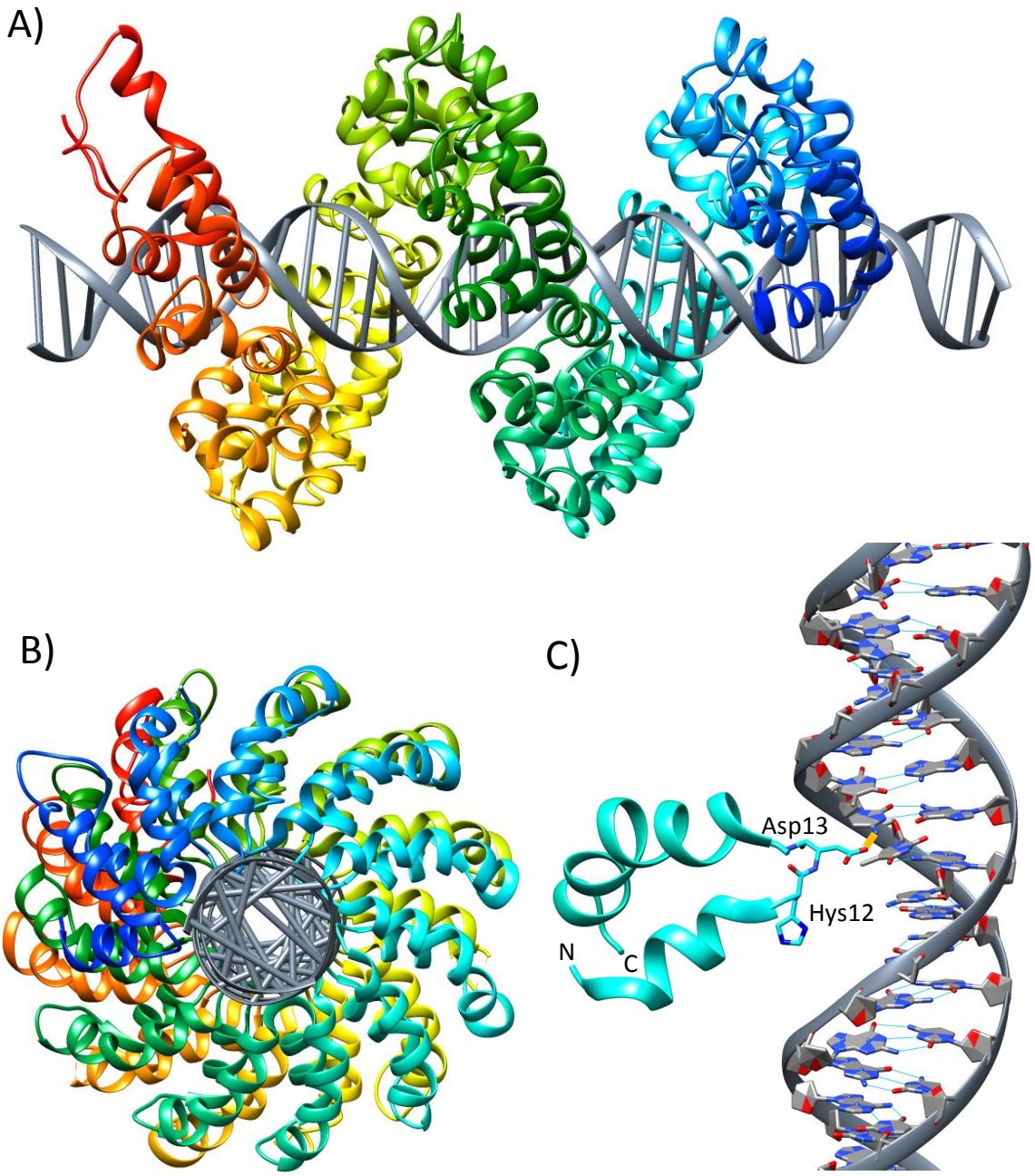


Figure 1-16: Structure and DNA binding of a TALE binding domain. A) The Image shows the crystal structure of the PthXo1 TALE binding domain bound to its DNA target from a side view (PDB: 3UGM). A rainbow colour transition across the protein is used to aid with clarity of the image. **B)** The Image shown is the same structure as in A) but rotated 90° looking at the DNA head on. **C)** A single 34 amino acid TALE repeat binding to DNA via the major groove. The two RVD residues are shown and labelled. Bases of the DNA helix are shown, and hydrogen bonds between base pairs are indicated with light blue lines to aid the clarity of the image. A hydrogen bond contact from RVD residue Asp13 to a cytosine base is indicated by a short orange line.

Using one of the aforementioned architectures, one may then assemble an array of TALE repeats with RVDs specifying a desired base pair sequence. The target sequence for a TALE must be at least 10–12 bp, but is usually designed to be 15–21 bp (Carroll, 2014). The only limitation in the sequence design is presented by two cryptic modules

(resembling TALE repeats in structure but less so in sequence) at the N-terminus of the array, which specify a thymine directly 5′ of the designed target sequence (Mak et al., 2012). However, efforts have been made to overcome this limitation by mutating residues in one of the cryptic modules (Lamb et al., 2013, Doyle et al., 2013). Another challenge that may present itself is that methylation of carbon-5 of a cytosine within a target sequence interferes with binding by TALE arrays. However, this limitation may be overcome through the use of a TALE repeat that recognizes thymine, since the RVD recognition of thymine occurs at its carbon-5 methyl group, which is indistinguishable from that of methylated cytosine (Deng et al., 2012b, Valton et al., 2012). It should be noted that this solution results in some loss of specificity, as either thymine or methylated cytosine may now be recognized at these positions. Additionally, because the RVDs that specify thymine make lower affinity contacts, some loss of TALE binding domain affinity may also occur.

1.7.2.4 Construction of TALE binding domains

Because of the highly repetitive nature of the TALE array, they cannot be generated by ordinary gene synthesis, and assembly from repeat monomers is somewhat nontrivial. Several methods for TALE array construction have been described in the literature, including: iterative hierarchical cloning (Sander et al., 2011a, Huang et al., 2011), Golden Gate-based and Golden Gate-like cloning (Cermak et al., 2011, Kim et al., 2013, Liang et al., 2014, Li et al., 2011c), ligation-independent cloning (Schmid-Burgk et al., 2013), PCR preparation and ligation assembly (Yang et al., 2013, Zhang et al., 2013), and solid-phase assembly (Reyon et al., 2012, Briggs et al., 2012). For these methods, cloning steps are the factor limiting how quickly a chimeric gene-editing enzyme incorporating a TALE domain can be generated. The number of cloning steps required varies from 1–6 steps depending on the method used (see Table 1-2). TALE array assembly methods require libraries of TALE repeats, with the faster versions of each method being enabled by libraries of 2–4 TALE repeat modules. These libraries, which must either be generated or acquired, range in size from 12–424 modules. In addition, several of the TALE assembly methods, including two of the Golden Gate-based approaches (Kim et al., 2013, Liang et al., 2014), ligation-independent cloning (Schmid-Burgk et al., 2013), and the solid-phase assembly approaches (Reyon et al., 2012, Briggs et al., 2012), can be adapted for high-throughput to produce large numbers of TALE arrays at once. Figure 1-17 shows a basic diagram illustrating three common assembly

platforms: iterative hierarchical cloning, Golden Gate cloning, and FLASH (solid-phase assembly).

Table 1-2: Comparison of TALE assembly methods

Source	Method	Module size	Library size	HT	# of cloning steps
Sander et al. (2011)	IHC	1	34† (12)‡		6
Huang et al. (2011)	IHC	2	16		5
Li et al. (2011)	GGC℗	1	48		2
Cermak et al. (2011)	GGC	1	44 (55)º		2
Kim et al. (2013)	GGC	1–3	424	✓	1
Liang et al. (2014)	GGC	1 & 2	308	✓	1
Schmid-Burgk et al. (2012)	LIC	2	64	✓	2
Zhang et al (2013)	PCR	4 (1–4)*	256 (340)*		1 (>1)*
Yang et al. (2013)	PCR	3	64		1
Reyon et al. (2012)	SPA	1–4	376	✓	1
Briggs et al. (2012)	SPA	1	12	✓	1

Abbreviations: IHC = iterative hierarchical cloning, GGC = Golden Gate cloning, LIC = ligation independent cloning, PCR = PCR preparation and ligation assembly, SPA = solid-phase assembly, HT = high-throughput.

Module size indicates the number of repeats in the modules of the library.

† The published library is missing three N- and C-terminal modules specifying G, and A and G, respectively. The library size shown assumes the inclusion of the missing modules.

‡ The published library utilizes module variants with backbone sequences varied through codon redundancy to reduce the repetitiveness of the sequences. However, this was an early report on TALE assembly and subsequently published methods of TALE assembly do not make use of this. Thus, the library might be reduced to 12 modules by eliminating the backbone variants.

℗ This method was published before the Golden Gate cloning method but is Golden Gate-like.

º This library includes an extra 11 modules with alternative RVD specificity for G. However, these modules could be omitted.

*Optionally, this library may include modules of 1–3 repeats (in addition to the 4 repeat modules); however, using these may increase the number of cloning steps required.

Finally, TALE arrays (often in the form of TALEN pairs) may also be purchased for $1500–$5000 from commercial vendors such as Life Technologies (Carlsbad, CA, USA), Cellectis Bioresearch (Romainville, France), ToolGen (Soul, Korea), CoWin Biotech (Beijing, China), Transposagen Biopharmaceuticals (Lexington, KY, USA), and others (Segal and Meckler, 2013). Additionally, several American universities (e.g. the University of San Francisco, the University of Utah, and the University of Wisconsin) now have dedicated TALEN production facilities.

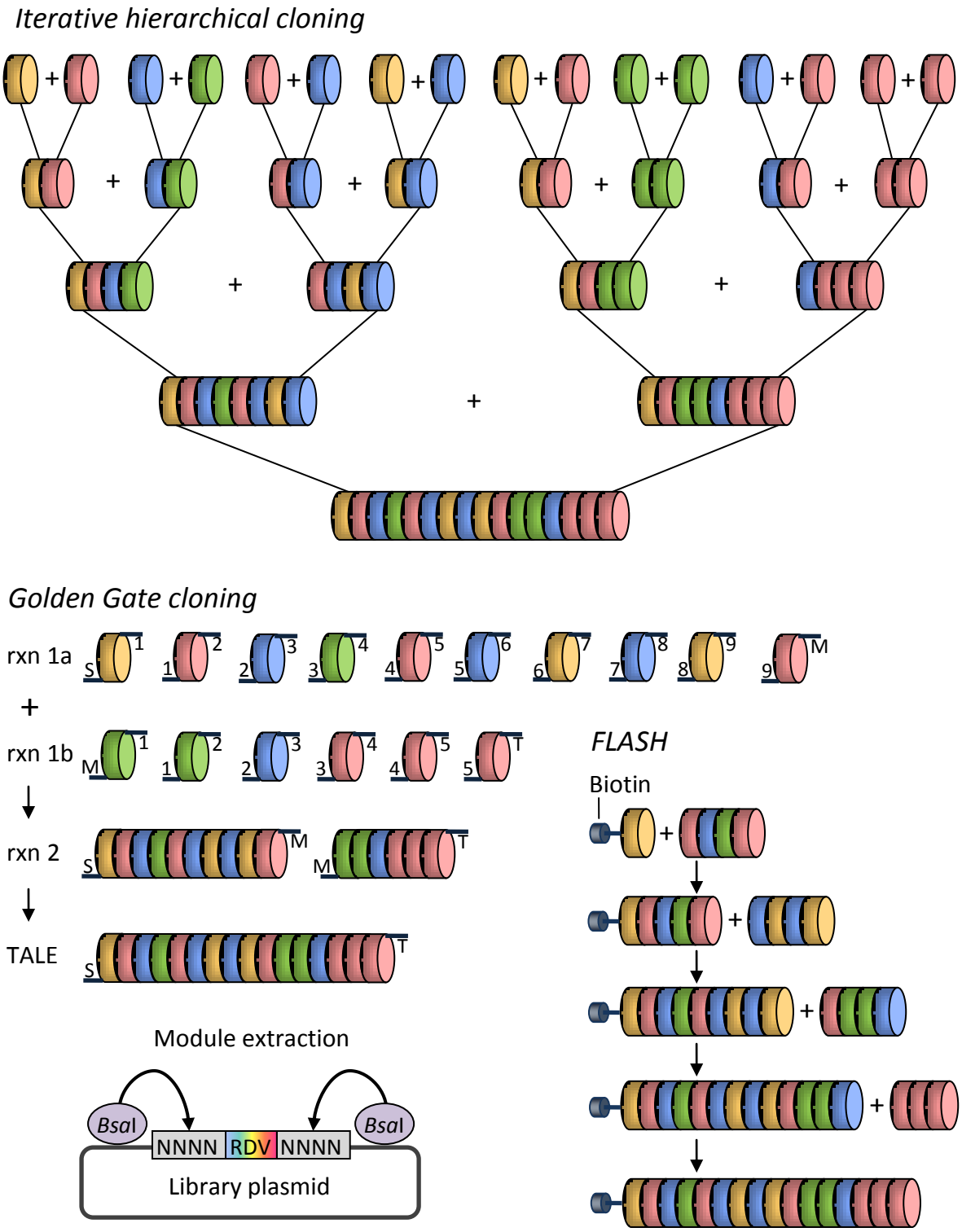


Figure 1-17: Three common TALE array construction platforms: Iterative hierarchical cloning, Golden Gate cloning, and FLASH. For all examples, four different coloured discs (yellow, red, blue, and green) are used to represent four different varieties of TALE repeat modules. **Iterative hierarchical cloning (top).** Coding sequences for four module types with the same pair of compatible 3' and 5' end overhang sequences are used to assemble monomers in parallel. Each iteration of the process doubles the size of the units being ligated until a 16 mer TALE array is achieved. The N- and C-terminal modules contain unique 5' and '3 overhangs (not shown), which allow the TALE array to be cloned into a desired fusion context (e.g. ZFN framework) containing the TALE N- and C-terminal regions of the chosen TALE architecture (also not shown). **Golden Gate cloning (bottom left).** Coding sequences for modules which contain 10 pairs with unique 3' and 5' end overhangs are extracted from library plasmids using TypeIIS restriction enzymes and used to assemble up to 10 modules at once. Two parallel reaction pots (rxn 1a + rxn 1b) are used to produce one 10-mer and one 6-mer array (rxn 2) which are then assembled into a 16-mer array (TALE). **FLASH (botton right)** Iterative ligations are performed on a streptavidin-coated solid-phase surface using modules containing up to four TALE repeats. Adapted from Joung and Sander (2014).

1.8 Project objectives

This project attempts to deduce structural and mechanistic information about the Tn3 resolvase-based ZFR reaction so that three enhancements to the system can be made to enable its use for targeted gene editing applications: capability for orientation-specific integration, irreversibility of ZFR-catalysed reactions, and relaxation of the ZFR target-site specificity requirements.

Currently, the ZFR system is unable to control the orientation of integration reactions, and insertions are randomly produced in both orientations, resulting in an approximately a 50:50 ratio of both insertion orientations. Two broad approaches were taken in this project to control this parameter of the reaction. The first broad approach was to attempt to generate orientation specificity in the dimer-dimer interface. These efforts included analysis and alteration of residues that may be involved at the initial contact point during dimer-dimer formation, and the use of rationally designed mutations to operate over a hypothesised mechanism of catalytic activation and regulation. The second broad approach was to allow integration reactions to take place in both orientations, but to try to make integration in one orientation irreversible, while leaving integration in the other orientation capable of being reversed. These efforts included attempts to modify the catalytic domain, or binding domain, such that after integration in one orientation, the ZFR no longer functioned on its Z-sites.

Irreversible ZFR reaction catalysis would likely greatly improve integration efficiency, preventing subsequent excision reactions that stochastically favour an excision end product. Two approaches taken in this project to generate unidirectional ZFR reaction are: rationally designed mutations that operate over a hypothesised fundamental aspect of catalytic regulation, and the generation of ZFR subunits with differential binding affinity in order to generate a reaction that produces product Z-sites that are no longer capable of being bound by binding-competent ZFR dimers (this approach is related to the approach for integration orientation specificity outlined above).

Relaxation of ZFR target site recognition requirements would allow the ZFR system to be used to target more sites within a given genome. Currently, the ZFR system is limited in its targeting capability both by the sequence specificity bias of the catalytic

module, and by the lack of available ZFAs. To date, much effort has been focused on attempting to relax or alter the recognition requirements of the catalytic module (Gaj et al., 2011, Gaj et al., 2013b, Sirk et al., 2014, Gersbach et al., 2010, Gordley et al., 2007, Proudfoot et al., 2011). However, it appears that the possibility of reducing the specificity requirements of one binding module in a ZFR dimer has not been considered. This is a hypothetically feasible approach since the combined sequence recognition capability of two catalytic modules and one three-fingered binding module of a dimer, appears to provide enough sequence specificity to target a unique site in a human genome (see Section 5.1.2.2). However, the recognition degeneracy of a ZFR has never been fully characterized (the study by Gaj *et al.* (2013) only characterized a limited set of mutation combinations (Gaj et al., 2013b); see Section 1.5.2.4), and it is not known how many base pair changes a ZFR actually will tolerate within its target site. Tolerance of target site degeneracy will likely depend on the activating mutations incorporated in the catalytic module, and on the identities and positions of the base pair changes themselves.

Since the sequence specificity requirements of the catalytic module must be accommodated anyway, reduction of ZFR site-specificity requirements, by reducing those of a single binding module in a ZFR dimer, could expand the repertoire of target sites that are only limited by the availability of ZFAs. The challenge to this approach is in producing specificity-reduced ZFRs that are able to cooperate as site-specific heterodimers, but are unable to catalyse reactions as less site-specific homodimers (which might participate in off-target reactions in a genomic context). It should be noted, that at the time this project was started TALERs had not yet been developed. However, due to the large size of TALE arrays (TALE repeat motifs are slightly larger than zinc finger motifs, but specify only 1 base instead of 3, and TALE arrays are comprised of many more motifs than ZFAs), the difficulty of their construction, and the potentially reduced activity of TALERs compared to ZFRs (Mercer et al., 2012), the reduced-specificity ZFR approach may still have some value.

<u>Goals summary</u>

- 1 Engineer orientation-specific integration capability.
- 2 Engineer irreversibility of the ZFR recombination reaction.
- 3 Relaxation of the ZFR target site recognition requirements.

Chapter 2: *Methods and Materials*

2.1 Basic procedures

2.1.1 Restriction digests

Restriction digests of plasmid DNA were performed in 15 µL volumes according to the instructions of the restriction enzyme providers (i.e. NEB, Invitrogen, and Promega). NEB enzymes were preferred where an option was available.

2.1.2 Gel electrophoresis

Agarose gels were made at using UltraPure™ Agarose (Life Technologies, Carlsbad, CA, USA). Preparative gels were made at 0.7% agarose, while analytic gels were made at 1.2% agarose. Gels were made approximately 1 cm thick. Gels were run 3.5–4.5 h (e.g. 80 V, 165 mA). Gels were stained for 20 min in 0.6 mg/mL EtBr using the running buffer, and de-stained twice in H_2O for 20 min per wash.

2.1.3 DNA extraction from gels

DNA extraction from gels was performed using either a Qiagen QIAquick Gel Extraction kit or Costar® Spin-X® (Corning®, Corning, NY, USA) centrifuge tube filters (0.45 µm pore size) by spinning for 1 min at RCF = 16,000 × *g.*

2.1.4 Ligations

Ligation of DNA constructs were performed in 20 µL volumes using approximately a 1:1 ratio of DNA molecules for each reaction. Invitrogen T4 ligase was used according to the manufacturer's protocol.

2.1.5 Ethanol precipitation

To 20 µL ligation mixtures was added 1 uL of tRNA (Sigma-Aldrich, R5636 - Ribonucleic acid, transfer from baker's yeast (*S. cerevisiae*), conc. 9-11 mg/mL). Mixtures were then briefly vortexed. Next, 10 µL of 5 M ammonium acetate solution was added to the mixtures, and the mixtures were again briefly vortexed. 70 µL of 100% ethanol was then added to the tube. The ligations were then set on ice for 15 min before being

centrifuged at maximum for 50 min. Afterward the supernatant was removed from the tubes using thin gel loading pipette tips. 100 µL of cold 70% ethanol was added to the tubes to wash the pellet. The tubes were briefly spun again for approximately 5–10 s before the supernatant was again removed from the tubes using the thin gel loading pipette tips. The pellets were then allowed to dry for approximately 15 min before 15 µL of ddH_2O was added to re-suspend the DNA for use for transformations.

2.1.6 High-density overnight cultures

Colonies were picked from plates using toothpicks that were then used to inoculate 5 mL of 2YT in 15 mL conical tubes. Five small holes where punched in the lid for better air circulation. The tubes were set up at an approximately 45° angle in a shaker set at 225 rpm and 37°C. The cultures were collected approximately 16 h later.

2.1.7 Gel image transformation

All gel images were transformed for display using the Quality One GelDoc software (Bio-Rad, Hercules, CA, USA). Within the *transform* menu the *Invert display* option was selected, and then the *Gamma* option was set to either 0.25 or 4.00, depending on whether the *Image Mode* had been set to *UV* or *White,* respectively, at the time the image was taken. Taking images in either *UV* or *White* mode appears to reverse the way the image information is coded, allowing opposite ends of the gamma transform curve to be applied to the image. The preferred method for transforming an image for display was to acquire the image in *White* mode, invert it, and then apply a gamma transformation of 4.00 to it. From the point at which the preferred mode was discovered, all images were prepared this way, though many images were prepared from *UV* mode acquisitions with inversion and a 0.25 gamma transformation, before this point. This type of transformation was used because it was found to display some of the faint bands in the figures most clearly.

2.2 Protocols

2.2.1 Preparation of electrocompetent cells

Step 1) Three x 500 mL of L-broth were placed in three 2 L Erlenmeyer flasks and each were inoculated 500 µL of with fresh overnight *E. coli* culture.

Step 2) Cells were grown at 37°C with shaking (300 rpm). Samples were taken from the cultures periodically and analysed using a spectrophotometer. When cultures reached an OD_{600} between 0.5–0.7 (indicating early- to mid-log phase), they were placed on ice in the cold room for approximately 20 min. Cells were then transferred to centrifuge bottles and spun for at 4000 x g for 15 min at 4°C.

Step 3) The supernatant was carefully poured off the pellets.

Step 4) Each pellet was then gently resuspended in 500 mL of ice-cold 10% glycerol, centrifuged again at 4000 x g for 15 min at 4°C, and again the supernatant was gently poured off.

Step 5) Each pellet was then gently resuspended in 250 mL of ice-cold 10% glycerol, centrifuged again at 4000 x g for 15 min at 4°C, and again the supernatant was gently poured off.

Step 6) Each pellet was then gently resuspended 2 mL ice-cold 10% glycerol and a 20 µL sample was taken from each tube and pulsed with the micropulser. If the pulse time was between 5.7 and 5.9 ms (preferably 5.8 or 5.9 ms) the stocks from which they were taken were marked with the pulse time and kept on ice. If the pulse time was lower than 5.7 ms, a further 20 mL of ice-cold 10% glycerol was added to each stock, and they were again centrifuged again at 4000 x g for 15 min at 4°C, and afterward the supernatant was gently poured off.

Step 7) Each pellet was then gently resuspended in 2 mL of ice-cold 10% glycerol, and the cells were then separated into 100 µL aliquots. 20 µL samples were taken from each lot and pulsed using the micropulser. The pulse times were recording and lots were labelled accordingly.

Step 8) Cell aliquots were then flash frozen in liquid nitrogen and subsequently stored at -70°C until use.

Final note: The efficiency of the cells depended on how thoroughly the supernatant was poured off, how gently the cells were resuspended (waiting a few minutes and then gently swirling worked best; although, gently pipetting up and down with an electronic pipette was also effective; *resuspending the pellet by manual shaking produced poor quality electrocompetent cells*), and whether it was possible to reduce the number of centrifugation steps (see Step 6).

2.2.2 Electroporation

Step 1) Electrocompetent *E.coli* cells (see Section 2.2.1) were thawed on ice for approximately 15 min, and 0.2 cm cuvettes were placed on ice alongside the cells.

Step 2) For single plasmid transformations the cells were diluted tenfold in 10% glycerol. For double plasmid transformations the cells were diluted fourfold in 10% glycerol. For triple plasmid transformations the cells were diluted twofold in 10% glycerol.

Step 3) When experimental work was done from plasmid minipreps 0.5 µL of each plasmid stock was added to 20 µL of cells, the tubes were flicked or gently pipette up and down to mix, and the cells put on ice for 1 min. For ligation transformations, 5 µL of purified DNA in ddH_2O (see Section 2.1.5) was used instead.

Step 4) The cell and DNA mixture was placed in a cuvette, and pulsed using the MicroPulser (setting Ec2). Immediately after the pulse, 1 mL of 2-YT broth was added to the cuvette, and the cell suspension transferred to a to a 5 mL conical tube. The pulse time was recorded.

Step 5) The conical tube containing the cell suspension was placed in the shaker at 37°C, at 225 rpm for 1 h.

Step 6) The cell suspension was placed on selective media and incubated at 37°C overnight.

2.2.3 17 Hour Recombination Assay

For all of the following steps the cells were kept on ice throughout.

Step 1) *E. coli* DS941 cells were simultaneously transformed with one or two expression plasmids and substrate plasmid by electroporation (0.5 µL of each plasmid stock in 20 µL of cells). The cells were then plated on selective media (ampicillin for pMS140-based plasmids; chloramphenicol for pSA3017-based and pSA3022-based plasmids; and kanamycin for pMS184Δ-based plasmids) and incubated for approximately 17 h at 37°C.

Step 2) Cells were then harvested by pipetting 6 mL of cold Luria Broth onto the plates and then resuspending by agitation with the side of a clear gel loading pipette tip (bend the end to avoid tearing into the agar). Approximately 5 mL of cell suspension was then aspirated from each plate and pipetted into 15 mL conical tubes that were set on ice.

Step 3) The optical density of the cell suspensions was measured by spectrophotometry at a 1:100 dilution and reading at OD_{600}. Cell suspensions were vortexed before each reading to thoroughly re-suspend the cells.

Step 4) A calculation (see below) was then performed on the OD_{600} readings in order to determine what volume (V) from the 5 mL cell suspensions was equal to 5 mL of high density (grown in 2YT) overnight culture. (High density overnight cultures had previously been measured to have an OD_{600} of 0.046 at a 1:100 dilution.)

$$V = \frac{5\text{mL}}{\text{OD600}/0.046}$$

Optional step 4b) If a cell passage was required, the volume (V) determined in the Step 4, was removed from each cell suspension added to a new 15 mL conical tube (tube 1) and the volume of suspension was topped up to 5 mL with 2YT broth. The suspension was then vortexed to mix, and 0.5 µL of suspension was then added to a new 15 mL conical tube (tube 2) containing 5 mL 2YT broth, and placed in the shaker at 37°C on 250 rpm.

Step 5) If no cell passage was required, the volume (V) determined in Step 4a, was removed from each cell suspension and a QIAprep Spin Miniprep Kit (Qiagen, Venlo, Netherlands) was used (elusion in 50 µL) in order to harvest DNA from the cells. If a cell passage was required then suspension from tube 1 from Step 4b was used for the miniprep instead. In both cases, the initial centrifugation step for the miniprep was conducted in the 15 mL connical tubes because the volume of suspension surpassed the capacity of the 2 mL Eppendorph tubes (additional 2TY broth was added to the connical tubes if their volumes were sufficiently different to unbalance the centrifuge rotor).

Step 6) 14 µL of each Miniprep was then digested with *Fsp*I + *Kpn*I-HF® + *Pvu*II-HF® (see below for reaction conditions) and analysed by gel electrophoresis on a 1.2% agarose gel. In one instance, when the desired enzymes were not available, an alternative set of enzymes (*Age*I-HF® + *Kpn*I-HF®) were used to achieve similar discrimination of product bands. The same reaction conditions were used as below, but using 1 µL each enzyme.

Cut Smart buffer =	2.00 µL
*Fsp*I =	0.50 µL
*Kpn*I-HF =	0.75 µL
*Pvu*II-HF =	0.75 µL
DNA =	16.0 µL
Total volume =	20.0 µL
Time =	1.5 h

2.3 Constructions

2.3.1 Construction of substrates

2.3.1.1 Four-part oligonucleotide annealing

The double-stranded oligonucleotides containing the Z-sites that were used for substrate construction were constructed through a four-part annealing strategy with two annealing steps (Figure 2-1). Single-stranded oligonucleotides containing Z-site half-sites (Table 2-1) were ordered from Eurofins Genomics (Ebersberg, Germany) and resuspended in Tris-EDTA buffer (10 nM, pH 8.2) to a final concentration of 100 pmol/µL. The first annealing step was designed to created double-stranded oligonucleotides containing Z-site half-sites. 1 µL of oligonucleotide suspension for the oligonucleotides corresponding

to both top and bottom strands of the left or right halves of a Z-site (see Table 2-1) were mixed in a 1:1 ratio in an Eppendorf tube and combined with 20 µL of 0.5 M NaCl and 78 µL of Tris-EDTA buffer (10 nM, pH 8.2) to make 100 µL total volume. The annealing mix was heated on a heating block at 85°C for 10 min, then covered with tin foil (to slow cooling) and left to cool overnight. The second annealing step was designed to generate double-stranded oligonucleotide Z-site full-sites. The double-stranded oligonucleotides generated in the first annealing step, which comprise the left and right halves of Z-sites were mixed in 1:1 ratios corresponding to full Z-sites in new Eppendorf tubes, and then heated to 49°C and cooled using the same procedure as before. A lower temperature (49°C instead of 85°C) was used in the second annealing step in order to prevent the melting of the desired intermolecular bonds between oligonucleotides that had formed during the first annealing step (see Figure 2-1). Oligonucleotide melting temperatures were calculated using Oligo Calc: Oligonucleotide Properties Calculator (Northwestern University Feinberg School of Medicine, Chicago, IL, USA; http://www.basic.northwestern.edu/biotools/oligocalc.html). A list of all double-stranded oligonucleotides that were generated is given in Table 2-2.

Table 2-1: Single-stranded Z-site oligonucleotides

ID and sequence	Position*	site
Z1o1: CTAGACTTGCGTGGGCGTCG	Top-left	Z1
Z1o2: AAATATTATAAATTATCAGCGCCCACGCAACTAGTG	Top-right	Z1
Z1o3: TGATAATTTATAATATTTCGACGCCCACGCAAGT	Bottom-left	Z1
Z1o4: GTACCACTAGTTGCGTGGGCGC	Bottom-right	Z1
Z2o1: CTAGACTTGACGCTGCTCCG	Top-left	Z2
Z2o2: AAATATTATAAATTATCAGTCCTCACTCAACTAGTG	Top-right	Z2
Z2o3: TGATAATTTATAATATTTCGGAGCAGCGTCAAGT	Bottom-left	Z2
Z2o4: GTACCACTAGTTGAGTGAGGAC	Bottom-right	Z2
Z3o1: CTAGACTTGAGTGAGGACCG	Top-left	Z3
Z3o2: AAATATTATAAATTATCAGTCCTCACTCAACTAGTG	Top-right	Z3
Z3o3: TGATAATTTATAATATTTCGGTCCTCACTCAAGT	Bottom-left	Z3
Z3o4: GTACCACTAGTTGAGTGAGGAC	Bottom-right	Z3
Z4o1: CTAGACTTGCGGCGGACCCG	Top-left	Z4
Z4o2: AAATATTATAAATTATCAGGTCCGCCGCAACTAGTG	Top-right	Z4
Z4o3: TGATAATTTATAATATTTCGGGTCCGCCGCAAGT	Bottom-left	Z4
Z4o4: GTACCACTAGTTGCGGCGGACC	Bottom-right	Z4
Z5o1: CTAGACTTTGGGGTGCCCCG	Top-left	Z5
Z5o2: AAATATTATAAATTATCAGGGCACCCCAAACTAGTG	Top-right	Z5
Z5o3: TGATAATTTATAATATTTCGGGGCACCCCAAAGT	Bottom-left	Z5
Z5o4: GTACCACTAGTTTGGGGTGCCC	Bottom-right	Z5

*Position is the same as that depicted in Figure 2-1.

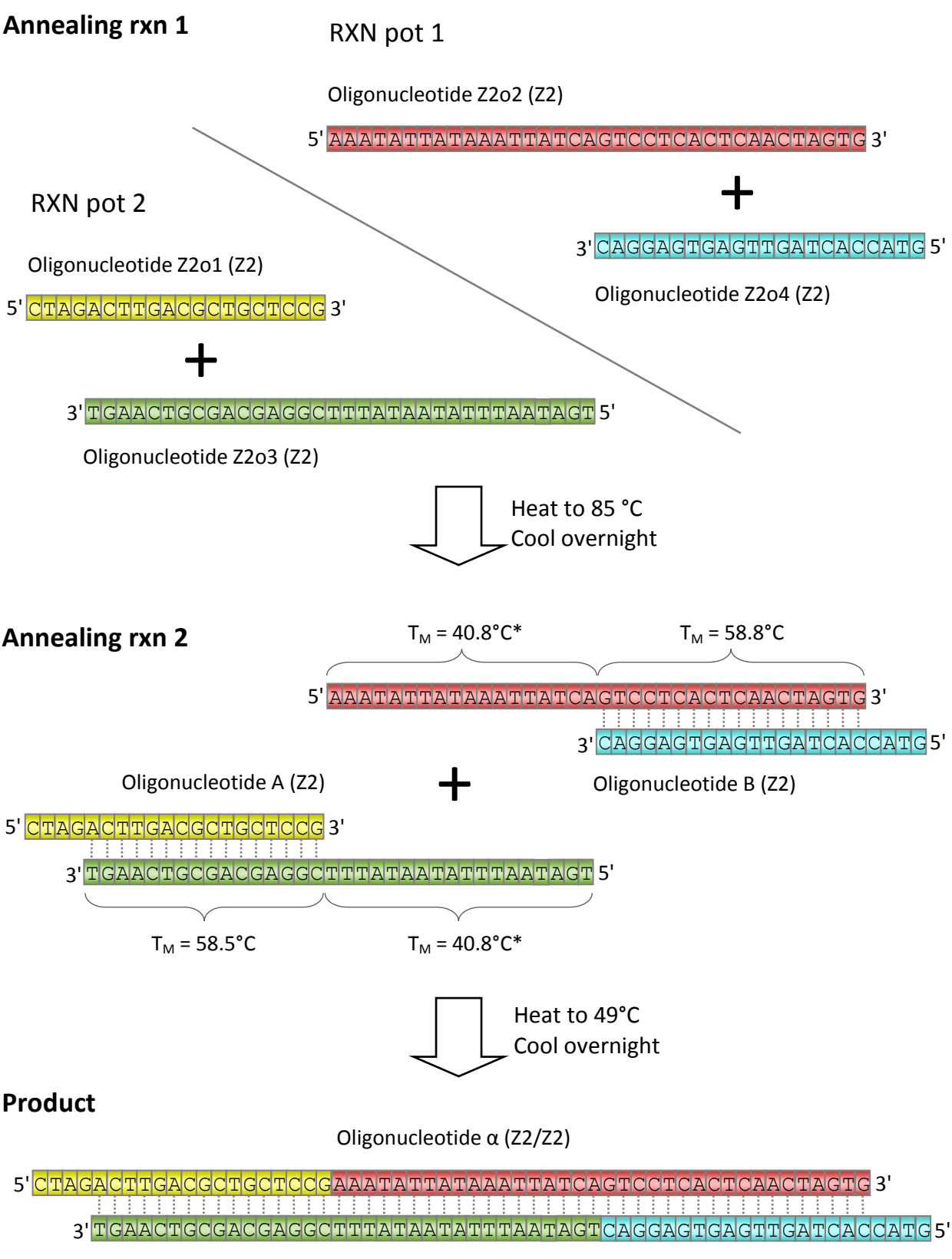


Figure 2-1: Four-part oligonucleotide annealing strategy. This figure depicts the strategy that was used to generate double-stranded Z-site oligonucleotides from four single-stranded oligonucleotides in two steps, using the construction of oligonucleotide α (containing the Z-site Z2/Z2) as an example.

Table 2-2: All constructed double-stranded oligonucleotides

Half-sites construction			Homodimer Z-sites construction			Heterodimer Z-sites construction (set A)			Heterodimer Z-sites construction (set B)		
A	Z2o1 + Z2o3	Z2/	α	A + B	Z2/Z2	ε	A + Δ	Z2/Z3	λ	Γ + B	Z3/Z2
B	Z2o2 + Z2o4	/Z2	β	Γ + Δ	Z3/Z3	ζ	A + Z	Z2/Z4	μ	E + B	Z4/Z2
Γ	Z3o1 + Z3o3	Z3/	γ	E + Z	Z4/Z4	η	A + Θ	Z2/Z5	ν	H + B	Z5/Z2
Δ	Z3o2 + Z3o4	/Z3	δ	H + Θ	Z5/Z5	θ	Γ + Z	Z3/Z4	ξ	E + Δ	Z4/Z3
E	Z4o1 + Z4o3	Z4/				ι	Γ + Θ	Z3/Z5	ο	H + Δ	Z5/Z3
Z	Z4o2 + Z4o4	/Z4	ρ	I + K	Z1/Z1	κ	E + Θ	Z4/Z5	π	H + Z	Z5/Z4
H	Z5o1 + Z5o3	Z5/									
Θ	Z5o2 + Z5o4	/Z5				σ	I + B	Z1/Z2	χ	A + K	Z2/Z1
						τ	I + Δ	Z1/Z3	ψ	Γ + K	Z3/Z1
I	Z1o1 + Z1o3	Z1/				υ	I + Z	Z1/Z4	ω	E + K	Z4/Z1
K	Z1o3 + Z1o4	/Z1				φ	I + Θ	Z1/Z5	ϗ	H + K	Z5/Z1

2.3.1.2 Construction of substrate plasmids

Figure 2-2 provides a diagram of the following construction procedure using the construction of pJK193 (Z2/Z2 x Z2/Z2) as an example. Substrates were constructed by first digesting pMS183Δ with either XbaI + Asp718I or NheI + BsrGI (which cut at two different polycloning sites within pMS183Δ) and then extracting the linear DNA from a 0.7% UltraPure Agarose gel. The double-stranded oligonucleotides containing the various Z-sites were then ligated into the polycloning sites, generating two sets of plasmids with Z-sites located at either of the polycloning sites. After amplifying the products plasmids in *E. coli*, the product plasmids were digested using AlwNI + BlpI, separated on a 0.7% UltraPure Agarose gel, and the fragments containing the Z-sites were extracted. The AlwNI - BlpI and BlpI - AlwNI fragments containing the Z-sites were then ligated to one another to generate substrates containing two Z-sites. A table of all the substrate plasmids constructed by this method can be found later in Chapter 3 (Table 3-2). A diagram of pMS183Δ, and the two polycloning sites into which the Z-sites were ligated, is also given in Chapter 3 (Figure 3-2).

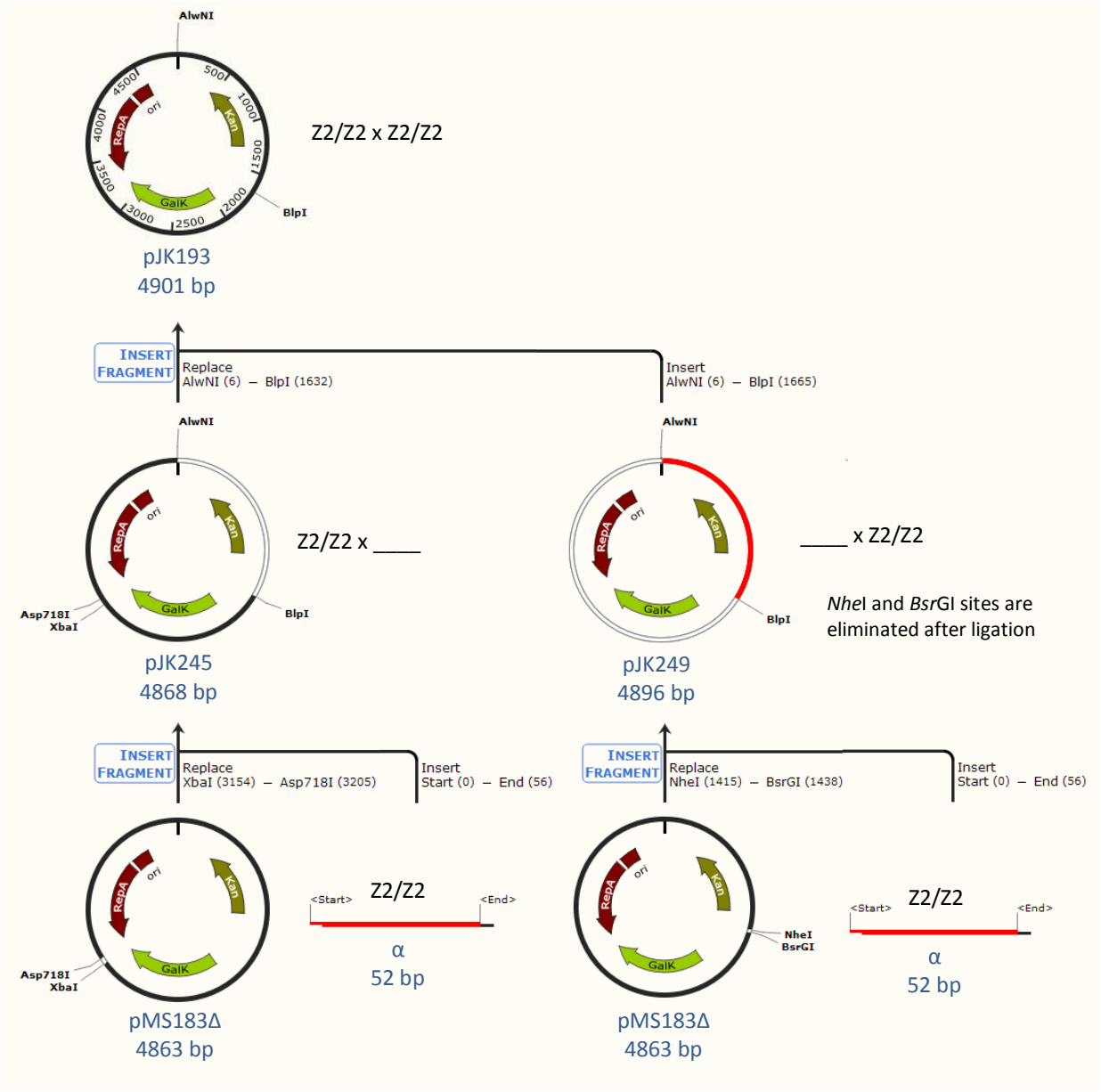


Figure 2-2: Substrate plasmid construction. The figure depicts a diagram of the construction strategy for new substrate plasmids using the construction of pJK193 as an example. Note that the oligonucleotides are indicated as being 52 bp, which does not include the four base overhangs on either side. Also note that the 'insert' in the diagram is labelled 'Start (0) - End (56),' which is the length of the top strand of the oligonucleotide (a labelling convention employed by the program Snap Gene, which was used to generate the figure).

2.3.2 Construction of ZFRs

Figure 2-3 provides a diagram of the following construction procedure using the construction of Tn3[NM]-Z2 as an example. The plasmids p12AAMO2P, p12AAMOZP, p12AAMOYP, and p12AAMOXP (containing Z2, Z3, Z4, and Z5, respectively), along with pAMC11 (containing Tn3[NM V107F-Z1) were digested using *Spe*I and *Sac*I. The DNA fragments from the plasmids containing the Z2, Z3, Z4, and Z5 were then separated by gel electrophoresis on a 3% MetaPhor Agarose gel. The DNA fragments from pAMC11 were separated by gel electrophoresis on a 0.7% UltraPure Agarose gel. The 264 bp bands containing the ZFAs, and a 5.35 kb containing the Tn3[NM V107F] catalytic domain and pAMC11 backbone, were extracted from the gels and ligated. It was necessarily to clone these Z2, Z3, Z4, and Z5 into a pAMC11, or pAMC11-like framework, because not all ZFR expression plasmids in the Stark lab have the required *Sac*I site at the end of the ZFR reading frame. The resulting plasmids containing Tn3[NM V107F]-Z2, Tn3[NM V107F]-Z3, Tn3[NM V107F]-Z4, and Tn3[NM V107F]-Z5 were then digested using their *Bsa*I and *Eag*I restriction sites to replace the Tn3 catalytic module with other varieties of previously generated Tn3 catalytic module, and the multiple restriction sites throughout the ZFR reading frame were used to clone in oligonucleotides in order to generate novel mutants.

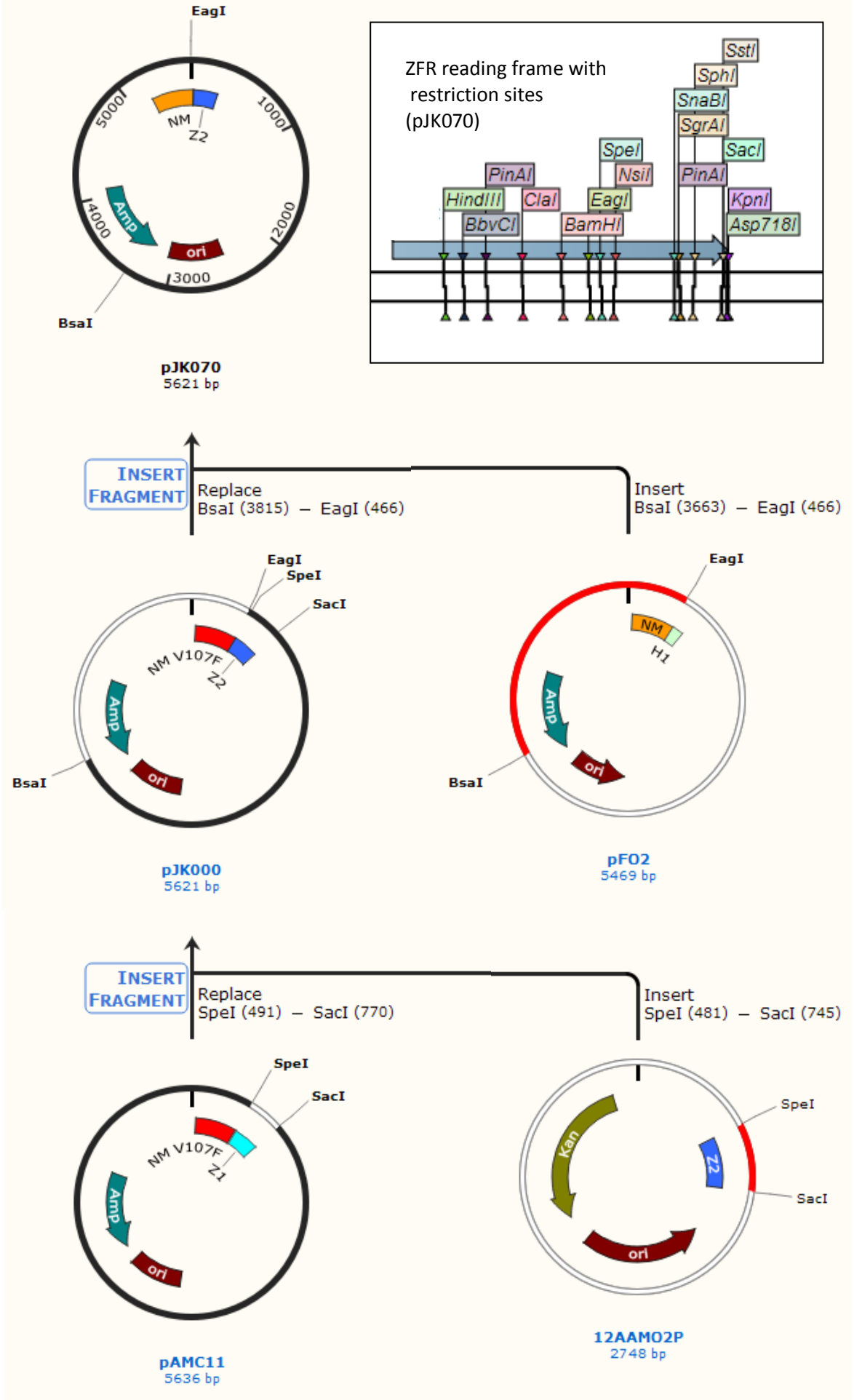


Figure 2-3: ZFR construction method. The figure depicts a diagram of the construction strategy for new ZFRs using the construction of Tn3[NM]-Z2 as an example.

2.4 Plasmids

NAME	DESCRIPTION	AB SELECTION
Chapter 3, Section 3.5		
pJK321	Tn3[NY G70C]-Z2 (pSA3017 vector)	Chl^r
pJK322	Tn3[NY G70C]-Z3 (pSA3017 vector)	Chl^r
pJK323	Tn3[NY G70C]-Z4 (pSA3017 vector)	Chl^r
pJK324	Tn3[NY G70C]-Z5 (pSA3017 vector)	Chl^r
pJK290	Tn3[NY G70C]-Z2 (pMS140 vector)	Amp^r
pJK291	Tn3[NY G70C]-Z3 (pMS140 vector)	Amp^r
pJK292	Tn3[NY G70C]-Z4 (pMS140 vector)	Amp^r
pJK293	Tn3[NY G70C]-Z5 (pMS140 vector)	Amp^r
pJK294	Tn3[NY G70C]-Z2 (pSA3022 vector)	Chl^r
pJK295	Tn3[NY G70C]-Z3 (pSA3022 vector)	Chl^r
pJK296	Tn3[NY G70C]-Z4 (pSA3022 vector)	Chl^r
pJK297	Tn3[NY G70C]-Z5 (pSA3022 vector)	Chl^r
pJK193	Z2/Z2 x Z2/Z2	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK195	Z4/Z4 x Z4/Z4	Kan^r
pJK196	Z5/Z5 x Z5/Z5	Kan^r
Chapter, Section 3.6		
pJK070	Tn3[NM]-Z2	Amp^r
pJK072	Tn3[NM]-Z2	Chl^r
pJK193	Z2/Z2 x Z2/Z2	Kan^r
pJK197	Z2/Z2 x Z3/Z3	Kan^r
pJK198	Z2/Z2 x Z4/Z4	Kan^r
pJK199	Z2/Z2 x Z5/Z5	Kan^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK210	Z2/Z4 x Z2/Z4	Kan^r
pJK211	Z2/Z5 x Z2/Z5	Kan^r
pJK221	Z2/Z3 x Z3/Z2	Kan^r
pJK222	Z2/Z4 x Z4/Z2	Kan^r
pJK223	Z2/Z5 x Z5/Z2	Kan^r
pJK071	Tn3[NM]-Z3	Amp^r
pJK073	Tn3[NM]-Z3	Chl^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK200	Z3/Z3 x Z2/Z2	Kan^r
pJK201	Z3/Z3 x Z4/Z4	Kan^r
pJK202	Z3/Z3 x Z5/Z5	Kan^r
pJK212	Z3/Z2 x Z3/Z2	Kan^r
pJK213	Z3/Z4 x Z3/Z4	Kan^r
pJK214	Z3/Z5 x Z3/Z5	Kan^r
pJK224	Z3/Z2 x Z2/Z3	Kan^r
pJK225	Z3/Z4 x Z4/Z3	Kan^r
pJK226	Z3/Z5 x Z5/Z3	Kan^r
pJK108	Tn3[NM]-Z4	Amp^r
pJK068	Tn3[NM]-Z4	Chl^r
pJK195	Z4/Z4 x Z4/Z4	Kan^r

pJK203	Z4/Z4 x Z2/Z2	Kan^r
pJK204	Z4/Z4 x Z3/Z3	Kan^r
pJK205	Z4/Z4 x Z5/Z5	Kan^r
pJK215	Z4/Z2 x Z4/Z2	Kan^r
pJK216	Z4/Z3 x Z4/Z3	Kan^r
pJK217	Z4/Z5 x Z4/Z5	Kan^r
pJK227	Z4/Z2 x Z2/Z4	Kan^r
pJK228	Z4/Z3 x Z3/Z4	Kan^r
pJK229	Z4/Z5 x Z5/Z4	Kan^r
pJK109	Tn3[NM]-Z5	Amp^r
pJK069	Tn3[NM]-Z5	Chl^r
pJK196	Z5/Z5 x Z5/Z5	Kan^r
pJK206	Z5/Z5 x Z2/Z2	Kan^r
pJK207	Z5/Z5 x Z3/Z3	Kan^r
pJK208	Z5/Z5 x Z4/Z4	Kan^r
pJK218	Z5/Z2 x Z5/Z2	Kan^r
pJK219	Z5/Z3 x Z5/Z3	Kan^r
pJK220	Z5/Z4 x Z5/Z4	Kan^r
pJK230	Z5/Z2 x Z2/Z5	Kan^r
pJK231	Z5/Z3 x Z3/Z5	Kan^r
pJK232	Z5/Z4 x Z4/Z5	Kan^r
Chapter 3, Section 3.7		
pJK325	Tn3[NY G70C]-L6-Z2	Amp^r
pJK326	Tn3[NY G70C]-L6-Z3	Amp^r
pJK327	Tn3[NY G70C]-L6-Z4	Amp^r
pJK328	Tn3[NY G70C]-L6-Z5	Amp^r
pJK193	Z2/Z2 x z2/Z2	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK195	Z4/Z4 x Z4/Z4	Kan^r
pJK196	Z5/Z5 x Z5/Z5	Kan^r
Chapter 3, Section 3.8		
pJK498	Tn3[NY G70C]-Z1 (pSA3017 vector)	Chl^r
pJK322	Tn3[NY G70C]-Z3 (pSA3017 vector)	Chl^r
pJK499	Tn3[NY G70C]-Z1 (pMS140 vector)	Amp^r
pJK291	Tn3[NY G70C]-Z3 (pMS140 vector)	Amp^r
pJK358	Tn3[NY G70C]-Z1 (pSA3022 vector)	Chl^r
pJK295	Tn3[NY G70C]-Z3 (pSA3022 vector)	Chl^r
pMP53	Z1/Z1 x Z1/Z1	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK500	Tn3[NM]-Z1	Amp^r
pJK098	Tn3[NM]-Z1	Chl^r
pJK263	Z2/Z2 x Z1/Z1	Kan^r
pJK264	Z3/Z3 x Z1/Z1	Kan^r
pJK102	Z4/Z4 x Z1/Z1	Kan^r
pJK265	Z5/Z5 x Z1/Z1	Kan^r
pJK269	Z1/Z2 x Z1/Z2	Kan^r
pJK270	Z1/Z3 x Z1/Z3	Kan^r
pJK257	Z1/Z4 x Z1/Z4	Kan^r

pJK271	Z1/Z5 x Z1/Z5	Kan^r
pJK275	Z1/Z2 x Z2/Z1	Kan^r
pJK276	Z1/Z3 x Z3/Z1	Kan^r
pJK258	Z1/Z4 x Z4/Z1	Kan^r
pJK277	Z1/Z5 x Z5/Z1	Kan^r
pJK070	Tn3[NM]-Z2	Amp^r
pJK072	Tn3[NM]-Z2	Chl^r
pJK071	Tn3[NM]-Z3	Amp^r
pJK073	Tn3[NM]-Z3	Chl^r
pJK108	Tn3[NM]-Z4	Amp^r
pJK068	Tn3[NM]-Z4	Chl^r
pJK109	Tn3[NM]-Z5	Amp^r
pJK069	Tn3[NM]-Z5	Chl^r
Chapter 3, Section 3.9		
pJK290	Tn3[NY G70C]-Z2	Amp^r
pJK291	Tn3[NY G70C]-Z3	Amp^r
pJK380	Tn3[NY G70C]-Z2 (slot #1) | Tn3[NY G70C]-Z3 (slot #2)	Amp^r
pJK381	Tn3[NY G70C]-Z3 (slot #1) | Tn3[NY G70C]-Z2 (slot #2)	Amp^r
pJK294	Tn3[NY G70C]-Z2	Chl^r
pJK295	Tn3[NY G70C]-Z3	Chl^r
pJK385	Tn3[NY G70C]-Z2 (slot #1) | Tn3[NY G70C]-Z3 (slot #2)	Chl^r
pJK386	Tn3[NY G70C]-Z3 (slot #1) | Tn3[NY G70C]-Z2 (slot #2)	Chl^r
pJK193	Z2/Z2 x Z2/Z2	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
Chapter 4, Section 4.2		
pJK155	Tn3[N D102D G70C]-Z2	Chl^r
pJK046	Tn3[N D102E G70C]-Z2	Chl^r
pJK510	Tn3[N D102H G70C]-Z2	Chl^r
pJK511	Tn3[N D102K G70C]-Z2	Chl^r
pJK512	Tn3[N D102R G70C]-Z2	Chl^r
pJK151	Tn3[N D102D G70C]-Z3	Amp^r
pJK104	Tn3[N D102E G70C]-Z3	Amp^r
pJK105	Tn3[N D102H G70C]-Z3	Amp^r
pJK106	Tn3[N D102K G70C]-Z3	Amp^r
pJK107	Tn3[N D102R G70C]-Z3	Amp^r
pJK197	Z2/Z2 x Z3/Z3	Kan^r
Chapter 4, Section 4.3		
pJK286	Tn3[NY]-Z2	Chl^r
pJK291	Tn3[NY G70C]-Z3	Amp^r
pJK193	Z2/Z2 x Z2/Z2	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK197	Z2/Z2 x Z3/Z3	Kan^r
pJK200	Z3/Z3 x Z2/Z2	Kan^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK212	Z3/Z2 x Z3/Z2	Kan^r
pJK221	Z2/Z3 x Z3/Z2	Kan^r
pJK224	Z3/Z2 x Z2/Z3	Kan^r
pJK294	Tn3[NY G70C]-Z2	Chl^r

pJK283	Tn3[NY]-Z3	Amp^r
pJK390	Tn3[NY]-Z2 | Tn3[NY]-Z2	Chl^r
pJK379	Tn3[NY G70C]-Z3 | Tn3[NY G70C]-Z3	Amp^r
pJK152	Tn3[N G70C]-Z3	Amp^r
pJK397	Tn3[NY L66I]-Z3	Amp^r
pJK398	Tn3[NY L66F]-Z3	Amp^r
pJK333	Tn3[NY G70A]-Z3	Amp^r
pJK363	Tn3[NY G70C]-Z3	Amp^r
pJK334	Tn3[NY I77T]-Z3	Amp^r
pJK335	Tn3[NY A89T]-Z3	Amp^r
pJK336	Tn3[NY F92S]-Z3	Amp^r
pJK337	Tn3[NY I97M]-Z3	Amp^r
pJK338	Tn3[NY T99S]-Z3	Amp^r
pJK339	Tn3[NY G101C]-Z3	Amp^r
pJK340	Tn3[NY G101S]-Z3	Amp^r
pJK341	Tn3[NY M103I]-Z3	Amp^r
pJK342	Tn3[NY M103V]-Z3	Amp^r
pJK343	Tn3[NY Q105L]-Z3	Amp^r
pJK375	Tn3[NY V107F]-Z3	Amp^r
pJK344	Tn3[NY V107M]-Z3	Amp^r
pJK345	Tn3[NY V108A]-Z3	Amp^r
pJK346	Tn3[NY T109I]-Z3	Amp^r
pJK347	Tn3[NY A117V]-Z3	Amp^r
pJK348	Tn3[NY R121G]-Z3	Amp^r
pJK349	Tn3[NY R121K]-Z3	Amp^r
pJK350	Tn3[NY R121M]-Z3	Amp^r
pJK351	Tn3[NY R121S]-Z3	Amp^r
pJK352	Tn3[NY E124Q]-Z3	Amp^r
pJK071	Tn3[NM]-Z3	Amp^r
pJK451	Tn3[NY L66I]-Z3 | Tn3[NY L66I]-Z3	Amp^r
pJK452	Tn3[NY L66F]-Z3 | Tn3[NY L66F]-Z3	Amp^r
pJK453	Tn3[NY G70A]-Z3 | Tn3[NY G70A]-Z3	Amp^r
pJK454	Tn3[NY I77T]-Z3 | Tn3[NY I77T]-Z3	Amp^r
pJK455	Tn3[NY A89T]-Z3 | Tn3[NY A89T]-Z3	Amp^r
pJK456	Tn3[NY F92S]-Z3 | Tn3[NY F92S]-Z3	Amp^r
pJK457	Tn3[NY I97M]-Z3 | Tn3[NY I97M]-Z3	Amp^r
pJK458	Tn3[NY T99S]-Z3 | Tn3[NY T99S]-Z3	Amp^r
pJK459	Tn3[NY G101C]-Z3 | Tn3[NY G101C]-Z3	Amp^r
pJK460	Tn3[NY G101S]-Z3 | Tn3[NY G101S]-Z3	Amp^r
pJK461	Tn3[NY M103I]-Z3 | Tn3[NY M103I]-Z3	Amp^r
pJK462	Tn3[NY M103V]-Z3 | Tn3[NY M103V]-Z3	Amp^r
pJK463	Tn3[NY Q105L]-Z3 | Tn3[NY Q105L]-Z3	Amp^r
pJK464	Tn3[NY V107F]-Z3 | Tn3[NY V107F]-Z3	Amp^r
pJK465	Tn3[NY V107M]-Z3 | Tn3[NY V107M]-Z3	Amp^r
pJK466	Tn3[NY V108A]-Z3 | Tn3[NY V108A]-Z3	Amp^r
pJK467	Tn3[NY T109I]-Z3 | Tn3[NY T109I]-Z3	Amp^r
pJK468	Tn3[NY A117V]-Z3 | Tn3[NY A117V]-Z3	Amp^r

pJK469	Tn3[NY R121G]-Z3 | Tn3[NY R121G]-Z3	Amp^r
pJK470	Tn3[NY R121K]-Z3 | Tn3[NY R121K]-Z3	Amp^r
pJK471	Tn3[NY R121M]-Z3 | Tn3[NY R121M]-Z3	Amp^r
pJK472	Tn3[NY R121S]-Z3 | Tn3[NY R121S]-Z3	Amp^r
pJK473	Tn3[NY E124Q]-Z3 | Tn3[NY E124Q]-Z3	Amp^r
pJK549	Tn3[NY T109I]-Z2	Chl^r
pJK550	Tn3[NY A117V]-Z2	Chl^r
pJK551	Tn3[NY R121G]-Z2	Chl^r
pJK552	Tn3[NY R121K]-Z2	Chl^r
pJK553	Tn3[NY R121M]-Z2	Chl^r
pJK554	Tn3[NY R121S]-Z2	Chl^r
pJK555	Tn3[NY E124Q]-Z2	Chl^r
Chapter 4, Section 4.4		
pJK155	Tn3[N G70C]-Z2	Chl^r
pJK363	Tn3[NY G70C]-Z3	Amp^r
pJK061	Tn3[N]-Z2	Chl^r
pJK071	Tn3[NM]-Z3	Amp^r
pJK393	Tn3[NY]-Z2	Chl^r
pJK197	Z2/Z2 x Z3/Z3	Kan^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK221	Z2/Z3 x Z3/Z2	Kan^r
Chapter 5, Section 5.2		
pJK513	Tn3[NY G70C]-AAA	Amp^r
pJK514	Tn3[NY G70C]-AAB	Amp^r
pJK515	Tn3[NY G70C]-AAC	Amp^r
pJK516	Tn3[NY G70C]-ABA	Amp^r
pJK517	Tn3[NY G70C]-ABB	Amp^r
pJK518	Tn3[NY G70C]-ABC	Amp^r
pJK519	Tn3[NY G70C]-ACA	Amp^r
pJK520	Tn3[NY G70C]-ACB	Amp^r
pJK521	Tn3[NY G70C]-ACC	Amp^r
pJK522	Tn3[NY G70C]-BAA	Amp^r
pJK523	Tn3[NY G70C]-BAB	Amp^r
pJK524	Tn3[NY G70C]-BAC	Amp^r
pJK525	Tn3[NY G70C]-BBA	Amp^r
pJK526	Tn3[NY G70C]-BBB	Amp^r
pJK527	Tn3[NY G70C]-BBC	Amp^r
pJK528	Tn3[NY G70C]-BCA	Amp^r
pJK529	Tn3[NY G70C]-BCB	Amp^r
pJK530	Tn3[NY G70C]-BCC	Amp^r
pJK531	Tn3[NY G70C]-CAA	Amp^r
pJK532	Tn3[NY G70C]-CAB	Amp^r
pJK534	Tn3[NY G70C]-CBA	Amp^r
pJK535	Tn3[NY G70C]-CBB	Amp^r
pJK536	Tn3[NY G70C]-CBC	Amp^r
pJK537	Tn3[NY G70C]-CCA	Amp^r
pJK538	Tn3[NY G70C]-CCB	Amp^r

pJK539	Tn3[NY G70C]-CCC	Amp^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK501	Tn3[NY G70C]-AA	Amp^r
pJK502	Tn3[NY G70C]-AB	Amp^r
pJK503	Tn3[NY G70C]-AC	Amp^r
pJK504	Tn3[NY G70C]-BA	Amp^r
pJK505	Tn3[NY G70C]-BB	Amp^r
pJK506	Tn3[NY G70C]-BC	Amp^r
pJK507	Tn3[NY G70C]-CA	Amp^r
pJK508	Tn3[NY G70C]-CB	Amp^r
pJK509	Tn3[NY G70C]-CC	Amp^r
Chapter 5, Section 5.3		
pJK294	Tn3[NY G70C]-Z2	Chl^r
pJK295	Tn3[NY G70C]-Z3	Chl^r
pJK329	Tn3[NY G70C]-Z2F2	Amp^r
pJK330	Tn3[NY G70C]-Z3F2	Amp^r
pJK331	Tn3[NY G70C]-Z4F2	Amp^r
pJK332	Tn3[NY G70C]-Z5F2	Amp^r
pJK193	Z2/Z2 x Z2/Z2	Kan^r
pJK194	Z3/Z3 x Z3/Z3	Kan^r
pJK195	Z4/Z4 x Z4/Z4	Kan^r
pJK196	Z5/Z5 x Z5/Z5	Kan^r
pJK197	Z2/Z2 x Z3/Z3	Kan^r
pJK198	Z2/Z2 x Z4/Z4	Kan^r
pJK199	Z2/Z2 x Z5/Z5	Kan^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK210	Z2/Z4 x Z2/Z4	Kan^r
pJK211	Z2/Z5 x Z2/Z5	Kan^r
pJK478	Tn3[NY G70C]-Z2F1	Amp^r
pJK479	Tn3[NY G70C]-Z3F1	Amp^r
pJK480	Tn3[NY G70C]-Z4F1	Amp^r
pJK481	Tn3[NY G70C]-Z5F1	Amp^r
Chapter 5, Section 5.4		
pJK400	Tn3[NY G70C]-H1	Amp^r
pJK396	Tn3[NY G70C]-Z3	Chl^r
pMP53	Z1/Z1 x Z1/Z1	Kan^r
pMP96	H1/Z1 x H1/Z1	Kan^r
pJK270	Z1/Z3 x Z1/Z3	Kan^r
pJK209	Z2/Z3 x Z2/Z3	Kan^r
pJK216	Z4/Z3 x Z4/Z3	Kan^r
pJK219	Z5/Z3 x Z5/Z3	Kan^r
pJK264	Z3/Z3 x Z1/Z1	Kan^r
pJK200	Z3/Z3 x Z2/Z2	Kan^r
pJK201	Z3/Z3 x Z4/Z4	Kan^r
pJK202	Z3/Z3 x Z5/Z5	Kan^r

2.5 Software

Sequence preparation and analysis was carried using CLC Genomics Workbench (CLC bio, Aarhus, Denmark). Molecular modelling was performed using YASARA (http://www.yasara.org/) and UCSF Chimera (Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco; http://www.cgl.ucsf.edu/chimera). The plasmid maps for figures were created using SnapGene (GSL Biotech, Chicago, IL, USA). Oligonucleotide melting temperatures were calculated using Oligo Calc: Oligonucleotide Properties Calculator (Northwestern University Feinberg School of Medicine, Chicago, IL, USA; http://www.basic.northwestern.edu/biotools/oligocalc.html).

Chapter 3: *New Components for the ZFR System*

3.1 Necessity of new zinc finger arrays

In order to achieve the project aims of generating orientation specificity at the dimer-dimer interface and producing a unidirectional reaction, at least *two types of mutant subunits, each with unique binding specificity,* are required. Through their mutual interface, these two subunit types would be able to discriminate between each other within the tetramer, and this differential interaction would give rise to either productive or unproductive reactions. Even more complex reactions could be envisioned that require *four types of subunits (each having unique binding specificity),* where higher degrees of discrimination between the subunit interfaces would lead to either a more robust orientation specificity, or unidirectional reaction effect, or a simultaneous combination of these two effects (see Chapter 6). Therefore, as a starting point, an effort was made to incorporate four new zinc finger array (ZFA) binding domains into the ZFR system. Correspondingly, a new set of substrates was also designed and tested. These substrates, which incorporate new Z-sites cognate to the new ZFAs, allow the detection of active and inactive subunit configurations, with respect to recombination reactions. Finally, in order to enable the expression of four ZFR subunits simultaneously, double expression vectors (DEVs) were also designed and tested.

3.2 New zinc finger arrays

3.2.1 ZFA criteria and background

Ideally, new ZFAs to be incorporated into the ZFR system should possess two qualities. Firstly, they should provide their ZFRs with relatively equal binding affinity, which should translate into relatively equal enzyme activity given the same catalytic domain. Secondly, they should bind to their target sites with high specificity in general, and more specifically, should not produce ZFRs that are cross-reactive on one another's Z-sites.

A previous attempt was made within our group by Chris Proudfoot, to select two engineered ZFAs from the literature (Isalan and Choo, 2001), and incorporate them into the ZFR system (Proudfoot et al., 2011). These domains, called ZifA and ZifB, performed with modest success in the ZFR context, but were not without problems. Notably, the enzyme activity of the ZFR utilizing ZifB was substantially reduced in comparison with that utilizing ZifA. This was in spite of these ZFAs being reported as having similar K_d, as measured by phage ELISA (Isalan et al., 2001). In addition, the ZFR utilizing ZifA appeared to be cross-reactive for the Zif268-based Z-site, although this possibility was not prohibited by the data given in the original paper describing these ZFAs. In fact, this cross-reactivity might not be entirely contrary to expectation, given that one half of ZifA was identical to Zif268, as the result of its bipartite construction (see Section 1.7.1.3 for a description of bipartite construction).

Because of the problems encountered during Proudfoot's previous attempt to incorporate new ZFAs, in addition to searching for domains with equal binding affinity and high binding specificity a further consideration was taken: the new domains should have their binding properties characterized in a system that is as similar as possible to the *in vivo E. coli* system being used for the ZFR experiments throughout this project. This would reduce the possibly for variation between expectation and result that might arise due to unpredictable idiosyncratic effects inherent to different experimental systems.

3.2.2 ZFA selection and rationale

After a thorough review of the current literature on engineered ZFAs, four were selected that had been generated through the OPEN (Oligomerized Pool ENgineering) platform designed by Maeder and colleagues (see Section 1.7.1.3) (Maeder et al., 2008). OPEN involves the generation and screening of three-fingered ZFA libraries by recombining pools of position-specific fingers (i.e. zinc fingers that always occupy the same location in the array). The arrays themselves are architecturally based on a triple repeat of Zif268 F2 (the central finger in the array). The OPEN system generally yields higher quality ZFAs than other open-access platforms (Pruett-Miller et al., 2008a, Maeder et al., 2008). Additionally, OPEN ZFA amino acid sequences are freely available through the Zinc Finger Consortium's web-accessible database, *ZiFDB* (http://www.zincfingers.org/) (Fu et al., 2009), making them accessible. Furthermore,

OPEN constructions employ a bacterial two-hybrid (B2H) binding activity screen carried out using *E. coli*, and so these data may be reliable within our experimental system.

The four ZFAs that were chosen (OZ217, OZ227, OZ246 and OZ345) were selected because Maeder and colleagues themselves had included them among domains they selected for more advanced work from a large number of ZFAs that appeared active in their B2H screen (Maeder et al., 2008). It is possible that unreported results could have contributed to their confidence in these domains (or lack of confidence in the ones that were not carried forward), and certainly their subsequent experiments provide further validation for the chosen ZFAs. Domains OZ217, OZ227, OZ246 and OZ345 also produced relatively similar activity scores in the B2H assay (see Table 3-1). Additionally, two of these ZFAs (OZ217 and OZ227) were further characterized for their binding specificity in a zinc finger nuclease (ZFN) context and compared with two top-performing Sangamo ZFAs in a more recent publication by the same group (Pattanayak et al., 2011). This provided further information about the potential for cross-reactivity with the sites of the other two chosen ZFAs (OZ246 and OZ345). The inclusion of OZ217 and OZ227 in the subsequent study was also taken as a potential indication of an additional degree of confidence in these two ZFAs by this group.

Table 3-1: B2H activity scores (adapted from Maeder et al. 2008)

ID	Mean B2H fold-activation	S.D. of B2H fold-activation
OZ217	7.19	2.15
OZ227	11.59	1.15
OZ246	10.16	0.55
OZ345	8.15	0.73

3.2.3 ZFA integration into the ZFR framework, and sequence modifications

Amino acid sequences for OZ217, OZ227, OZ246 and OZ345 were acquired from ZiFDB (Fu et al., 2013). A DNA sequence was then generated for each amino acid sequence using the reverse translate function of CLC Genomics Workbench with the “Use only the most frequently used codon” for *E. coli* option selected. These sequences were

then used as a starting point for necessary codon changes, coding sequence additions, and the insertion of various useful restriction sites, through silent changes and the addition of flanking sequence.

The amino acid sequences of the new ZFAs were first aligned with that of ZFR: Tn3-Zif268 (where 'Tn3' indicates the type of catalytic domain, and 'Zif268' indicates the identity of the ZFA) to observe how well they fit with the ZFR architecture (see Figure 3-1 A). In order to create the Tn3 arm region / zinc finger junction at the 5′ end of the OZ217, OZ227, OZ246 and OZ345 coding sequences, a small segment of DNA coding for an N-terminal Pro-His (which is not included in the Stark lab ZFR architecture (Prorocic et al., 2011)) was replaced with a sequence coding for Thr-Ser. The Thr-Ser junction sequence also includes a *Spe*I restriction site, which (along with a *Sac*I site at the 3′ end of a ZFA coding sequence), facilitates the removal and insertion of ZFAs within the ZFR framework.

Although the OPEN ZFAs are Zif268-like (based on a triple repeat of the Zif268 F2 finger; see Section 1.7.1.3), several Zif268 residues that are typically included in the C-terminus of the Tn3-Zif268 ZFR architecture, are missing from the amino acid sequences provided by ZiFDB. The coding sequence of these missing C-terminal residues is required for inclusion of the aforementioned *Sac*I site that facilitates swapping of ZFAs in the ZFR architecture. Because the backbone geometry of the *F2 triplet* array used for the OPEN ZFAs might be slightly different from that of the natural Zif268 array (there is variation in the amino acid sequences of fingers F1, F2, and F3, even apart from that within their recognition helices), it is possible that the missing C-terminal residues from Zif268 could be incompatible with the OPEN ZFAs. Therefore, published work on ZFNs that include the OZ217 and OZ227 ZFAs was used as a guide for completing the sequence (see Figure 3-1 A).

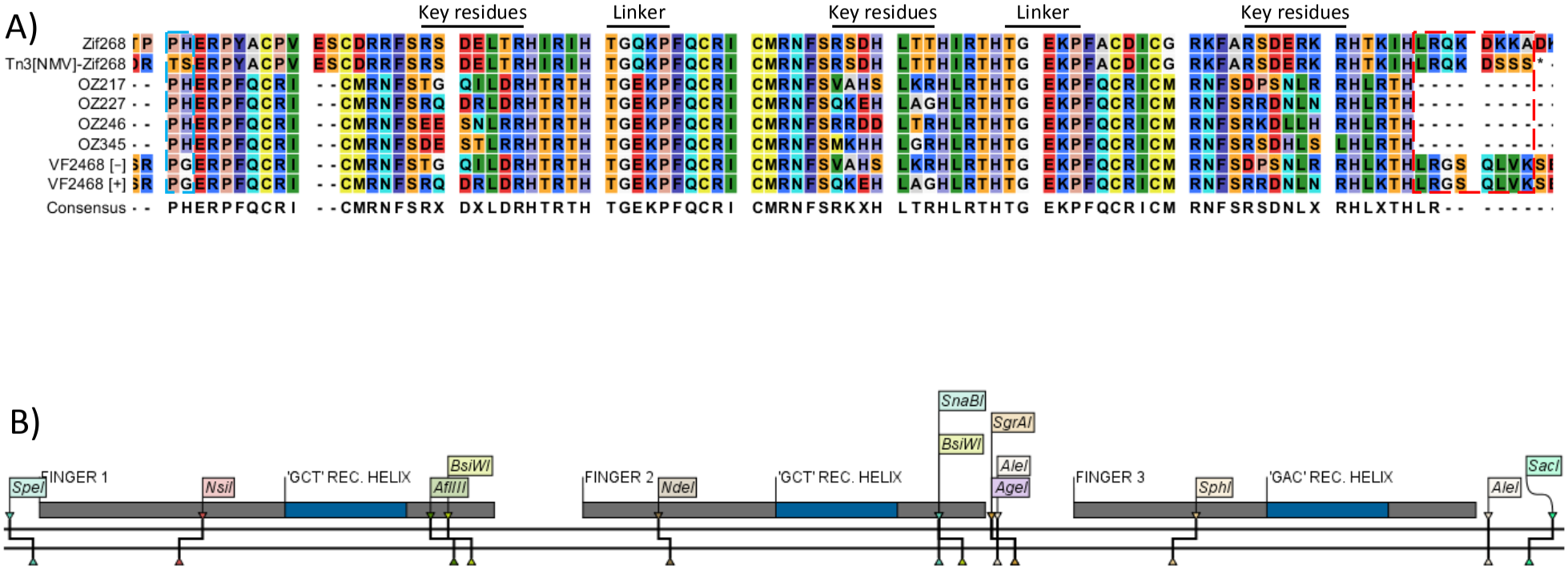


Figure 3-1: ZFA alignment and architecture. A) An alignment of Zif268, Tn3-Zif268, OZ217, OZ227, OZ246, OZ345, VF2468 (–) and VF2468 (+). Blue and red dashed boxes indicate the N-terminal residues replaced with the Thr-Ser linker, and the C-terminal regions, respectively. The locations of the key recognition helix residues and the inter-finger linkers are shown above the alignment. **B)** The architecture of the new ZFAs using Z2 as an example. The nucleotide sequence is represented by two black lines. All relevant features of the aa sequence are shown above the nucleotide representation using labelled and coloured horizontal bars. Locations of restriction sites are indicated within the lines representing the nucleotide sequence.

Two ZFN sequences, which incorporate OZ217 and OZ227 (VF2468 [-] and VF2468 [+], respectively), described in Pattanayak *et al.* (2011), include the first two C-terminal Zif268 residues (Leu-Arg) missing from the ZFA sequences given by ZiFDB (Figure 3-1 A) (Pattanayak et al., 2011). This was taken as evidence they could be included without detriment to function. The C-terminus of the Tn3-Zif268 ZFR sequence has an additional six residues after the Leu-Arg dipeptide, three of which are Zif268 derived, and the final three of which are a non-native serine triplet that allows the inclusion of the *Sac*I site (Figure 3-1 A). Conversely, following the Leu-Arg residues in the sequences from Pattanayak *et al.* (2011), there is a Gly-Ser dipeptide linker before the start of the *Fok*I domain sequence (note that a difference between the ZFN and ZFR architecture is that the ZFA is attached to the C-terminal of the catalytic domain instead of the N-terminal, respectively). Although this Ser-Gly dipeptide is not found in the Zif268 sequence, it does not appear to interfere with function in the ZFN context. By adding an additional serine residue to the end of the Ser-Gly dipeptide, a *Sac*I site can be introduced into the codons for these amino acids such that the site would be frame-compatible with the existing *Sac*I site in the current Tn3-Zif268 ZFR framework. Because the OPEN ZFAs have not been demonstrated as compatible with the six C-terminal ZFR residues of Zif268, the most conservative approach was to deviate as little as possible from the ZFN sequence where functionality has been demonstrated. Additionally, although beginning three residues earlier, this Gly-Ser-Ser C-terminus would be similar to the Ser-Ser-Ser C-terminus of the current Tn3-Zif268 architecture. Therefore, the sequences from the Pattanayak et al. (2011) ZFNs were used as the basis for the C-terminal region of the new ZFAs in the ZFR context, and a sequence coding for Leu-Arg-Gly-Ser-Ser was appended to the 3′ ends of all the coding sequences derived from the ZiFDB ZFAs.

It should be noted that an additional difference between the ZFA sequences OZ217 and OZ227, and the ZFN sequences VF2468 [–] and VF2468 [+] was spotted in the alignment (Figure 3-1 A). Within the ZFN sequences, the final arginine found in the ZFA portion of the sequence, which is part of the otherwise canonical OPEN Zif268 F2-triplet framework, has been replaced by a lysine residue. No explanation for this was provided by Pattanayak et al. (2011), and the canonical arginine was retained for the new sequences.

Additional modifications to the coding sequences were also made to allow alterations of the ZFAs required by later experiments. Restriction sites *Ale*I and *Bsi*WI were introduced through silent DNA sequence changes, to allow the sequential removal of fingers F3 and F2, respectively, while preserving the C-terminal ZFA architecture and corresponding 3' coding sequence containing the *Sac*I site (used to transport the reading frame). Restriction sites *Srg*AI and *Age*I were introduced to allow non-specific fingers (described in Chapter 5) to be introduced to the array. Restriction sites *Nsi*I, *Afl*III, *Nde*I, *Sna*BI, and *Sph*I were also introduced into the regions flanking each of the recognition helices, in order to allow the recognition helices to be replaced, or the ZFA to be otherwise reassembled and modified (see Figure 3-1 B).

After modification, the sequences for OZ217, OZ227, OZ246 and OZ345 were renamed Z2, Z3, Z4 and Z5, respectively, and were synthesized through GeneArt® (Life Technologies, Carlsbad, CA, USA), and delivered within plasmids p12AAMO2P, p12AAMOZP, 12AAMOYP and p12AAMOXP, respectively. The sequences were extracted from their respective plasmids by digestion with *Spe*I and *Sac*I restriction enzymes, separated on a 3% MetaPhor® agarose gel, and ligated between the *Spe*I and *Sac*I sites of the Tn3(NM[3] V107F)-Zif268 expression plasmid pAMC11 (Prorocic et al., 2011), replacing the Zif268 ZFA coding region. The ZFA coding regions were then ligated to coding sequences for mutant catalytic domains to create various novel ZFRs (see Chapter 2).

3.3 New ZFR substrates

3.3.1 Substrate configurations and function

New substrates were also designed to allow experiments with ZFRs utilizing the new ZFAs. These substrates were based on a GalK assay substrate plasmid similar that used by Akopian et al. (2003), Prorocic *et al.* (2011), and Proudfoot *et al.* (2011) (and similar to pDB35 (Arnold et al., 1999)), and were of five configurations (Table 3-2): (1) The single-mutant homodimer Z-site (1MutHomDim) substrates contain two identical ZFR homodimer binding sites (targeted by one of the four ZFAs). (2) The two-mutant homodimer Z-site (2MutHomDim) substrates contain two alternate homodimer binding sites (targeted by two of the four ZFAs). (3) The two-mutant heterodimer Z-sites in direct-

[3] A description of 'NM' can be found in Section 1.6.4.

repeat (2MutHetDim-DR) substrates contain two identical heterodimer binding sites in direct repeat orientation with respect to one another (targeted by two of the four of ZFAs). (4) The two-mutant heterodimer Z-sites in inverted-repeat (2MutHetDim-IR) substrates contain two heterodimer binding sites, arranged in inverted repeat with respect to one another (targeted by two of the four of ZFAs). (5) Finally, the four-mutant heterodimer Z-site (4MutHetDim) substrates contain two different heterodimer binding sites (targeted by four ZFAs). These substrates enable the detection of activity of various tetramer configurations consisting of one to four types of mutant subunits, and are important for the work described in the following chapters. The half Z-sites, containing binding targets for Z2, Z3, Z4 and Z5 domains, were likewise labelled Z2, Z3, Z4 and Z5, respectively, and the identity of full Z-sites is indicated using the convention Z*N*/Z*N* (where *'N'* indicates the numeric designation of the ZFA). Table 3-2 lists all substrate configurations that were constructed at this stage.

Table 3-2: List of generated substrates*

1MutHomDim	2MutHomDim	2MutHetDim-DR	2MutHetDim-IR	4MutHetDim
Z2/Z2 x Z2/Z2	Z2/Z2 x Z3/Z3	Z2/Z3 x Z2/Z3	Z2/Z3 x Z3/Z2	Z2/Z3 x Z4/Z5
Z3/Z3 x Z3/Z3	Z2/Z2 x Z4/Z4	Z2/Z4 x Z2/Z4	Z2/Z4 x Z4/Z2	Z2/Z3 x Z5/Z4
Z4/Z4 x Z4/Z4	Z2/Z2 x Z5/Z5	Z2/Z5 x Z2/Z5	Z2/Z5 x Z5/Z2	Z2/Z4 x Z3/Z5
Z5/Z5 x Z5/Z5	Z3/Z3 x Z2/Z2	Z3/Z2 x Z3/Z2	Z3/Z2 x Z2/Z3	Z2/Z4 x Z5/Z3
	Z3/Z3 x Z4/Z4	Z3/Z4 x Z3/Z4	Z3/Z4 x Z4/Z3	Z2/Z5 x Z3/Z4
	Z3/Z3 x Z5/Z5	Z3/Z5 x Z3/Z5	Z3/Z5 x Z5/Z3	Z2/Z5 x Z4/Z3
	Z4/Z4 x Z2/Z2	Z4/Z2 x Z4/Z2	Z4/Z2 x Z2/Z4	Z3/Z4 x Z2/Z5
	Z4/Z4 x Z3/Z3	Z4/Z3 x Z4/Z3	Z4/Z3 x Z4/Z3	Z3/Z4 x Z5/Z2
	Z4/Z4 x Z5/Z5	Z4/Z5 x Z4/Z5	Z4/Z5 x Z5/Z4	Z3/Z5 x Z2/Z4
	Z5/Z5 x Z2/Z2	Z5/Z2 x Z5/Z2	Z5/Z2 x Z2/Z5	Z3/Z5 x Z4/Z2
	Z5/Z5 x Z3/Z3	Z5/Z3 x Z5/Z3	Z5/Z3 x Z3/Z5	Z4/Z5 x Z2/Z3
	Z5/Z5 x Z4/Z4	Z5/Z4 x Z5/Z4	Z5/Z4 x Z4/Z5	Z4/Z5 x Z3/Z2

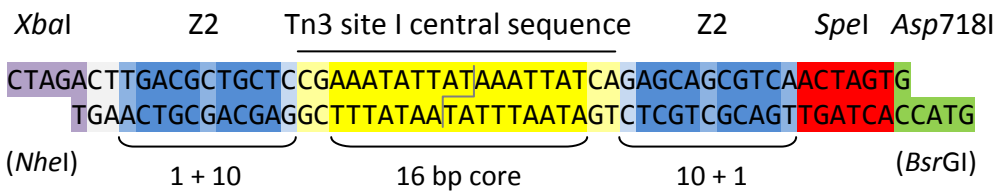
*Substrates are given in the format where the first Z-site is that between *Asp*718 and *Xba*I, and the second Z-site is that between *Bsr*GI and *Nhe*I (see Figure 3-2 B for a plasmid map of the substrate). A useful feature of this format is that the sites are arranged so that those half-sites that are adjacent to the 'x', are those that are on either side of the origin of replication in the plasmid. So, for instance, when an excision reaction is performed on any of the above substrates, the surviving product of the excision reaction will contain one Z-site composed of the two half Z-sites closest to the 'x'.

3.3.2 Z-site design and substrate construction

The Z-sites for the Z2, Z3, Z4 and Z5 ZFAs were designed to contain 20 bp of the Tn3 resolvase site I central sequence (the core 16 bp of which are involved in Tn3 arm region contacts, see Section 1.5.1), flanked by 10 bp sequences that code for the ZFA binding sites, plus 1 additional base pair 5′ of the ZFA binding sites that may influence binding (Figure 3-2 A). Although a three-fingered ZFA is often said to specify up to 9 bp of recognition sequence, 3′ overlap specificity from F1 of the array can increase this length of the binding site to 10 bp (see Section 1.7.1.2). Additionally, some studies suggest that some three-fingered ZFAs may have a preference a base pair at an extended position 5′ of the 10 bp binding site (Swirnoff and Milbrandt, 1995, Wolfe et al., 1999, Pattanayak et al., 2011).

The paper that describes the creation of OZ217, OZ227, OZ246 and OZ345 (Z2, Z3, Z4 and Z5), describes their binding specificity in terms of a 9 bp binding site (3 bp for each finger), and does not mention overlap specificity from F1 to a 10th base pair 3′ end of the 9 base pair binding site, nor the potential importance of an 11th base pair flanking the 5′ end of the binding site (Maeder et al., 2008). The Pattanayak *et al.* characterization of binding specificity for VF2468 [–] and VF2468 [+] (ZFNs based on OZ217 and OZ227, respectively) does, however, indicate that there is a modest selective preference for specific bases 3′ and 5′ of the 9 bp recognition site (Pattanayak et al., 2011). It is interesting to note, that Pattanayak *et al.*'s characterization suggests that the 5′ extended position contributes more specificity than the 3′ F1 overlap position. In both the case of OZ217 and OZ227, the base pair preference is cytosine at the potential 3′ F1 overlap position, and thymine at the extended 5′ position. Therefore, since there was no binding specificity characterization for OZ246 and OZ345, in all of the Z-sites that were created (for the Z2, Z3, Z4 and Z5 ZFAs), a cytosine was used at the potential 3′ F1 overlap position and thymine at the extended 5′ position (Figure 3-2 A).

A)



B)

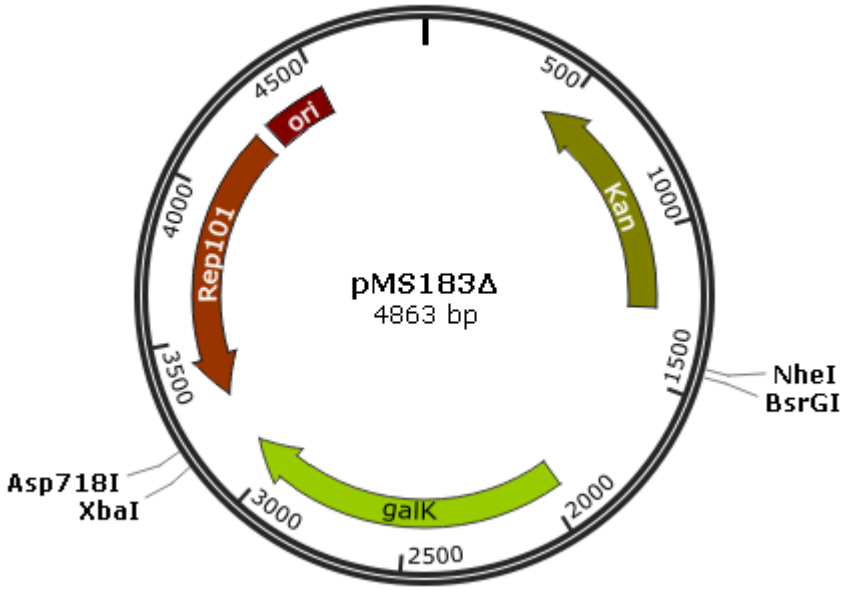


Figure 3-2: Z-site architecture and substrate framework plasmid. A) The figure shows the architecture of the Z-site oligonucleotides using the Z2/Z2 Z-site as an example. The central sequence derived from Tn3 site I is depicted in yellow, with lighter shades used to indicate those bases which are not important for Tn3 resolvase arm region contacts. The ZFA binding sites for Z2 are shown in blue, with lighter shades used to indicate their importance to binding relative to the specificity profile generated by Pattanayak *et al.* (2011). The *Spe*I restriction site, used for confirmation of oligonucleotide insertion during cloning, is indicated in red. The flanking *Xba*I (*Nhe*I) and *Asp*718I (*Bsr*GI) overhangs which are used for cloning into the substrate framework plasmid are depicted in mauve and green, respectively. Finally, faint grey lines are used to depict the central dinucleotide of the crossover site which separates the left and right halves of the binding site. **B)** Shows a plasmid map of pMS183Δ, the framework plasmid used to construct the substrate (Misiura et al. 2013). Note that pMS183Δ does not already have Z-sites within it. A pSC101 origin of replication is shown in burgundy, while the related rep101 gene which controls copy number is indicated in vermilion. The kanamycin resistance gene (aminoglycoside 3′-phosphotransferase) is indicated in olive, and the galactokinase gene, which the Z-sites flank, is indicated in chartreuse. The *Asp*718I and *Xba*I, and *Bsr*GI and *Nhe*I restriction sites, between which the Z-site oligonucleotides are inserted, are also indicated on the map. The restriction sites for *Nhe*I and *Bsr*GI are eliminated when the Z-site oligonucleotides are cloned in.

However, after all the experimental work for this project was completed, it was realized through reading a description in a subsequent publication of the OPEN protocol, that the designers of OPEN do consider the 5′ extended position and the 3′ F1 overlap position to be potentially important, and thus, OPEN ZFAs are selected on 11 bp sites (Maeder et al., 2009). The 11 bp sites for OZ217, OZ227, OZ246 and OZ345 can be found in the supplementary information that accompanies the paper describing their construction (Maeder et al., 2008). It is interesting to note, that the specificity indicated for the 5′ extended position determined by the specificity profile generated by Pattanayak *et al.*, does not match the selection sequences used for the creation of OZ217 and OZ227; however, the base identity at the 3′ F1 overlap position does (Pattanayak et al., 2011, Maeder et al., 2008). Therefore, an obviously better strategy for the creation of the Z4 binding site would have been to include at least the 3′ F1 overlap position base from the sequences on which OZ246 and OZ345 were generated. In the case of the Z5 binding site, the 3′ F1 overlap position in the selection site is also a cytosine (i.e. the base that was used for all of the Z-sites at this position).

The right-side half-Z-sites, within the Z-site construction oligonucleotides, were flanked by a *Spe*I site that was used to confirm insertion of full length Z-sites into a substrate framework plasmid, pMS183Δ (Misiura et al., 2013) (Figure 3-2). The *Spe*I site was then further flanked by a guanine, and then a 5′ overhang that is compatible with both *Asp*718 and BsrGI restriction sites, in order to facilitate cloning into the pMS183Δ polycloning sites (the cytosine retains the *Asp*718 sites but eliminates the *Bsr*GI site after ligation). The left-side half-Z-sites were flanked by an ACT trinucleotide, and then a 5′ overhang that is compatible with *Nhe*I and *Xba*I, in order to facilitate ligation of the full length Z-sites into pMS183Δ polycloning sites (the adenine of the trinucleotide preserves the *Xba*I site but not the *Nhe*I site after cloning).

Oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, Germany), and sites were assembled using a four-part ligation strategy that allowed all combinations of half-sites to be assembled from the minimal number of oligonucleotides. The strategy involved first annealing oligonucleotides to generate half-sites with long 18 bp 3′ overhangs within the central sequence, which is identical for all sites. These segments, containing half Z-sites Z2–Z5, were then further annealed to one another in all permutations to generate all possible full Z-sites. All desired substrate plasmids were

created by sequentially cloning the full sites into the two target regions flanking *galK* in pMS183Δ (see Figure 3-2 B) in a two-step cloning procedure.

3.4 17 Hour Recombination Assay

ZFR-mediated substrate recombination was assayed using the '17 Hour Recombination Assay'. This is a simple bacterial *in vivo* assay that produces a quick turnover of results (within 24 hours) and a reasonably high differential resolution of substrate plasmid recombination products. Briefly, the procedure is as follows: *E. coli* DS941 is simultaneously co-transformed with relevant expression vectors and substrate plasmids, followed by an approximately 17–17.5 hour incubation period on selective plate media. All plasmids are then harvested from the plates, subjected to restriction digest, and subsequently analysed by gel electrophoresis (see Section 2.2.3).

ZFR-mediated recombination of substrate plasmids typically produces two types of products: An **inversion product** plasmid may be produced, where the segment of DNA containing *galK,* which is flanked by the Z-sites, has been reversed in orientation. An **excision product** plasmid may be produced where the segment of DNA containing *galK* has been excised. Additionally, if an interaction between the ZFRs and substrate is not productive, the **unchanged substrate** will be present at the end of the reaction. However, the unchanged substrate form can also be reproduced as inversion products undergo *subsequent* inversion reactions, returning the plasmid to its original form. Productive interaction between the ZFR tetramer and substrate is predicted to produce approximately initially equal quantities of inversion and excision product. However, because the excision product is an end product, and the inversion product can be further converted back into the unchanged substrate form or excision product through subsequent reactions, reactions will tend to produce a preponderance of excision products on extended incubation (see Section 1.5.3). For the restriction digest, enzymes *Fsp*I, *Kpn*I, and *Pvu*II were selected in order to produce a cleavage pattern where the unchanged substrate, the inversion product, and the excision product could be clearly differentiated from one another and from the expression plasmids (see Figure 3-3).

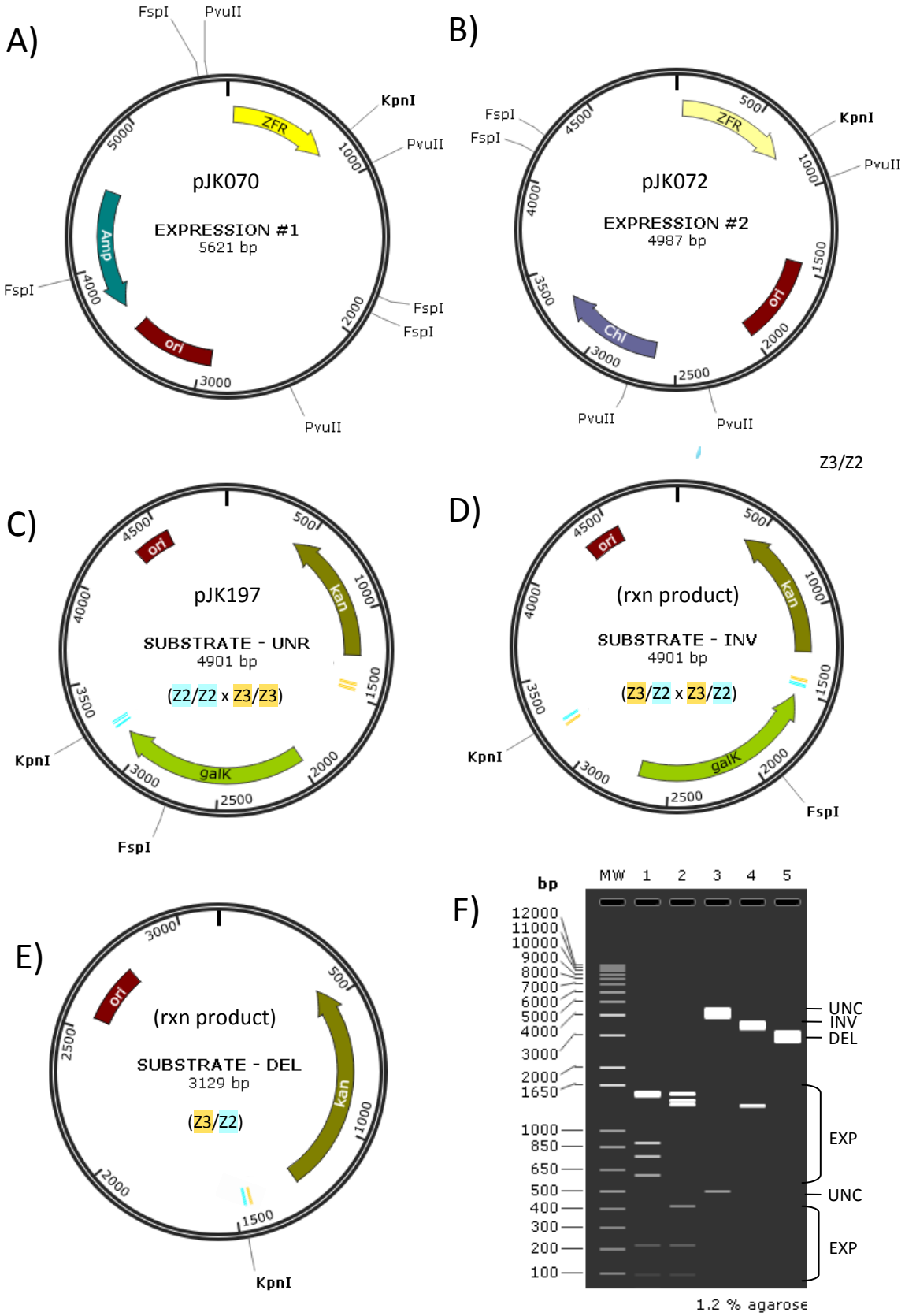

A)
FspI
PvuII
KpnI
PvuII
FspI
FspI
PvuII
pJK070
EXPRESSION #1
5621 bp
ZFR
Amp
ori
B)
FspI
FspI
KpnI
PvuII
PvuII
pJK072
EXPRESSION #2
4987 bp
ZFR
ori
Chl
Z3/Z2
C)
ori
kan
galK
KpnI
FspI
pJK197
SUBSTRATE - UNR
4901 bp
(Z2/Z2 x Z3/Z3)
D)
ori
kan
galK
KpnI
FspI
(rxn product)
SUBSTRATE - INV
4901 bp
(Z3/Z2 x Z3/Z2)
E)
ori
kan
KpnI
(rxn product)
SUBSTRATE - DEL
3129 bp
(Z3/Z2)
F)
bp
MW 1 2 3 4 5
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1650
1000
850
650
500
400
300
200
100
UNC
INV
DEL
EXP
UNC
EXP
1.2 % agarose

Figure 3-3: Expression plasmids and substrate products in the 17 Hour Recombination Assay. Maps for expression vectors and substrate plasmids are shown with relevant features and the 17 Hour Resolution Assay restriction digest sites (*Fsp*I, *Kpn*I, and *Pvu*II) indicated. ZFA binding sites are indicated using small blue (Z2) and orange (Z3) rectangles. The Z-site identities are written on the substrates in the format that is given in **Table 3-2**, in order to allow the reader to understand the convention that was used for describing the plasmids. The Z*N*/Z*N* x Z*N*/Z*N* description reads as if you looking at a plasmid that has been linearized with a cut at the bottom. **A)** Expression vector #1, based on pBR322, confers ampicillin resistance, and is cut into small fragments by the digest of the assay. **B)** Expression vector #2, based on pACYC184, confers chloramphenicol resistance, and is cut into small fragments by the digest of the assay. **C)** The unchanged substrate (UNC) is based on pSC101, confers kanamycin resistance, and is cut at two sites. Pairs of ZFA binding sites are alternatively coloured blue and orange and the identity of the Z-site is provided beside them. The *galK* gene is shown to indicate the orientation of the fragment between the binding sites. **D)** The inversion product product (INV) substrate is the same as the UNC substrate, but the segment of DNA containing *galK* has been reversed in orientation, which changes the restriction pattern, and the pair of blue and orange ZFA sites has been rearranged. **E)** The deletion (excision) product (DEL) substrate is the same as the UNC and INV substrates, but the segment of DNA containing *galK* has been excised and only one restriction site and one pair of rearranged ZFR sites remain. **F)** A virtual agarose gel indicating restriction patterns for each plasmid. Lane 1 = EXPRESSION #1, lane 2 = EXPRESSION #2, lane 3 = SUBSTRATE - UNC, lane 4 = SUBSTRATE - INV, lane 5 = SUBSTRATE - DEL. Note: the slow bands in the image are given extra thickness to indicate signal strength and are not to be interpreted as multiple overlapping bands.

3.5 Binding activity and vector expression level assay

3.5.1 Rationale and design of the binding activity assay

Ideally, all of the new ZFRs would bind their cognate Z- sites with equal affinity. Dissimilar levels of binding activity within complementation experiments could lead to inconsistent results. Additionally, were there some cross-reactivity of the ZFRs on non-cognate sites, dissimilar levels of binding activity would need to be taken into account when designing complementation experiments so as to minimize these effects. For example, it would be a poor experimental design to pair a weakly binding ZFR with a strongly binding ZFR that is cross-reactive for the weakly binding ZFR's binding site. It is therefore desirable to assess the binding activity of new ZFR binding domain variants before their use.

The simplest way to accomplish this test is by indirectly measuring the binding activity through a comparative ZFR recombination activity assay, where the catalytic domain for all ZFRs is of the same type, while binding domain type is used as an independent variable. This assay can also screen for any effects that result not from difference in binding activity, but from unexpected incompatibility between the Tn3 catalytic domain and ZFA. A Tn3 resolvase catalytic domain containing the four mutations

R2A E56K D102Y G70C was selected for this work. As described in Section 1.6.4 of Chapter 1, hyperactivation of Tn3 resolvase is typically accomplished through the introduction of both a primary activating mutation at position 102 (usually D102Y; hereafter referred to simply as 'Y'), and any one of a collection of secondary activating mutations that occur at locations throughout the protein (in this case G70C). In addition, a tertiary set of activating mutations can be included (in this case R2A E56K; hereafter collectively referred to as 'N'), which only produce an up-regulating effect when used in combination with primary and secondary mutations (see Section 1.6.5 for more details). The NY mutations were selected because they are regularly utilized within our group, and are thus predictable in their behaviour. The G70C secondary mutation was selected because it produces a high level of activation and had been chosen for use in much of the experimental work of the following chapters. Therefore, this catalytic domain (NY G70C) was selected for these activity level tests and assembled with Z2, Z3. Z4 and Z5 in expression vector pMS140, to create Tn3[NY G70C]-Z2, Tn3[NY G70C]-Z3, Tn3[NY G70C]-Z4 and Tn3[NY G70C]-Z5 ZFRs.

3.5.2 Rationale and design of the vector expression level test

Within *in vivo* complementation experiments, the expression level of ZFRs is an important factor. Virtually all of the experimental designs in this project involve complementation assays utilizing simultaneous ZFR expression from two vectors. Ideally the level of expression from the two vectors should be equal; dissimilar levels of expression could lead to inconsistent results. Additionally, were any problems in binding fidelity present, unintentionally placing the less specific ZFR in a higher expression vector than the higher specificity ZFR it was complemented with, would exacerbate the problem.

A series of variants of the pACYC184 expression vector with variably attenuated Shine-Dalgarno sequences was previously generated by Sally Rowland within the Stark lab (S. J. Rowland; unpublished work), in order to produce a compatible expression vector with a similar expression level to the pMS140 vector. pACYC184 is based on a p15A replicon, which has a copy number of 18–22 (Chang and Cohen, 1978), while the pMS140 is based on the compatible pMB1/ColE1 replicon, which has an copy number of 15–20 (Bolivar et al., 1977). Two variants of pACYC184 (pSA3017 and pSA3022) appeared, in MacConkey assays Rowland conducted, to have potentially similar activity levels to the

pMS140 expression vector. Of the two, the pSA3017 vector appeared potentially less active, and the pSA3022 vector appeared potentially more active, than the pMS140 vector. However, the comparative levels of expression from expression vectors can be hard to gauge effectively from the MacConkey assay when the activity levels are not drastically different from one another, such as in the case of pSA3017 and pSA3022. Therefore, pSA3017 and pSA3022 were selected for further characterization in an activity level test using the 17 Hour Recombination Assay, which provides a much higher level of resolution in gauging relative activity level. Coding sequences for Tn3[NY G70C]-Z2, Tn3[NY G70C]-Z3, Tn3[NY G70C]-Z4 and Tn3[NY G70C]-Z5 were subsequently inserted into pSA3017 and pSA3022. All ZFRs were assayed against their cognate 1MutHomDim substrates (substrates with two copies of the same homodimer binding site), expressed from all three expression vectors (pSA3017, pMS140, and pSA3022).

3.5.3 Results of the vector expression level test

The results, which are displayed in Figure 3-4, show the outcome of the 17 Hour Recombination Assay as determined by agarose gel electrophoresis. This experiment expressed four different ZFRs, which varied by their ZFA (Tn3[NY G70C]-Z2, Tn3[NY G70C]-Z3, Tn3[NY G70C]-Z4 and Tn3[NY G70C]-Z5), from three expression vectors (pSA3017, pMS140, and pSA3022). The aim was to infer the level of binding activity provided by the ZFAs, and level of expression provided by the expression vectors. With one exception, the comparison indicated the same pattern of activity between the ZFA variants when expressed from each vector (see below).

The result displayed in lane 2, representing Tn3[NY G70C]-Z2 expressed from pSA3017, is not consistent with the rest. In this lane there is a large band between the identifiable 4.4 kb band of the unchanged substrate, and the 5.1 kb marker of the 1 kb ladder. It is unclear how this unexpected large band has arisen; although, it might be single cut (i.e. incompletely digested) unchanged plasmid (e.g. if *Fsp*I didn't not cut; refer back to Figure 3-3), or partially digested expression plasmid, both of which could run at about this location. There is also a faint band just below the 3 kb mark that would suggest some partial digestion of the expression plasmid, as it has fragments that if incompletely digested would run there (e.g. a *Pvu*II to *Pvu*II fragment, if *Fsp*I didn't not cut well; refer back to Figure 3-3). This single lane does not present a problem for interpreting the

results of this experiment though because of the redundancy of information represented by expression of the same comparative set of ZFRs from the two other expression vectors, and because the other lanes within the pSA3017 set show the expected bands.

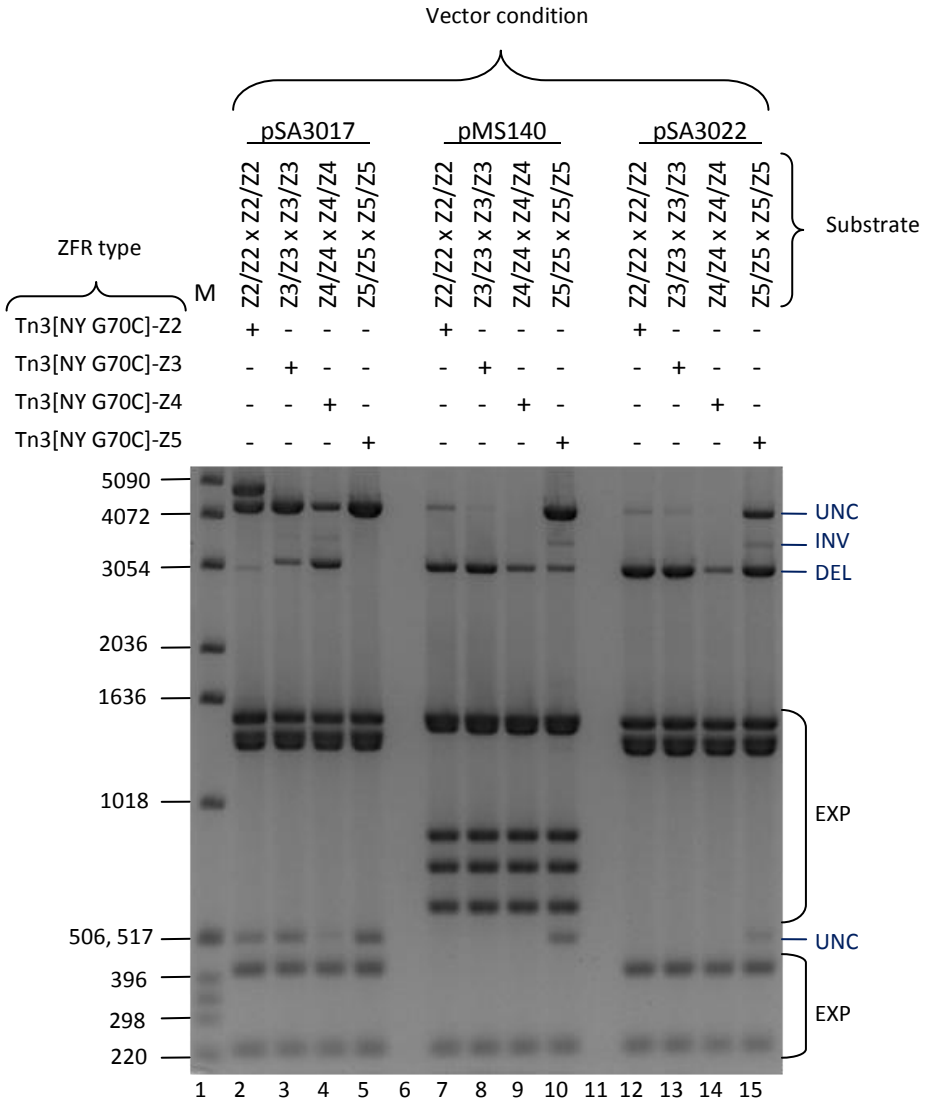


Figure 3-4: New ZFR activity level comparative assay results. The image shows the results of the *Fsp*I, *Kpn*I, and *Pvu*II restriction digest on a 1.2% agarose gel. To the right of the image, indications are given for bands corresponding to the unchanged substrate (UNC), inverted substrate product (INV), deletion substrate product (DEL), and the expression plasmids (EXP). To the left of the image, sizes are given for the 1 kb marker ladder which is contained in lane 1. Lanes 2–5, 7–10, and 12–15 contain the results of 1MutHomo substrate reactions with their cognate ZFR (Tn3[NG70CY]-Z2–Z4) expressed from pSA3017, pMS140, and pSA3022 vectors, respectively, as indicated above the image.

3.5.4 Vector expression level determined

The first piece of information, from the gel shown in Figure 3-4, that is notable is the overall ZFR expression level from each vector. The recombination activity level seen for each ZFR variant when produced from each vector indicates that the expression level from the vectors is ordered as follows: **pSA3017 < pMS140 < pSA3022**. This can be best visualized from the Tn3[NY G70C]-Z5 results (lanes 5, 10 and 15) where the reaction is apparently slowest and the substrate has not been completely converted to product. Because the recombination activity measured in a 17 Hour Recombination Assay is only assessed at the end of the experiment, when some reactions in a comparative set go all the way to completion, discriminative information becomes lost as the slower reactions continue. This is most obvious in a case where the slow reactions subsequently run until completion as well as well as the faster reaction, hiding the difference in reaction rate of the faster reactions. The results of the vector expression level component of this experiment, are consistent with the prediction made based on the previously referred to results of S. Rowland's MacConkey plate assay, but have allowed for a more accurate scaling of the activity of pSA3017 and SA3022 relative to that of pMS140.

3.5.5 Binding activity level determined

The comparative activity level imparted to the ZFRs by their ZFAs is the next piece of information that can be observed from the results in Figure 3-4. The Tn3[NY G70C]-Z2 and Tn3[NY G70C]-Z3 ZFRs appear to have approximately the same level of activity, with Tn3[NY G70C]-Z3 appearing slightly more active when expressed from the pMS140 vector (Figure 3-4, lanes 7 and 8). Although the Tn3[NY G70C]-Z5 ZFR appears to be active, it is notably less so than the ZFR with other ZFAs. Finally, ZFR Tn3[NY G70C]-Z4 appears to be most active, but produces an anomalous effect when expressed from the two vectors with higher expression level. The product bands for Tn3[NY G70C]-Z4 expressed from pMS140 and pSA3022, appear to be significantly attenuated in signal strength (discussed in Section 3.10.5).

3.5.6 Best expression vector / ZFA match for further experiments

The most similar expression vector and binding domain variant combinations are probably Tn3[NY G70C]-Z3 expressed from pMS140, and Tn3[NY G70C]-Z2 expressed from pSA3022. This takes into consideration the higher expression level from pSA3022, and the higher activity level of Tn3[NY G70C]-Z3. The level of activity of Tn3[NY G70C]-Z4 expressed from pSA3017, and Tn3[NY G70C]-Z2 expressed from pMS140, is also broadly comparable. The activity levels produced by Tn3[NY G70C]-Z4 expressed from pSA3017, and Tn3[NY G70C]-Z5 expressed from pSA3022 are similar, but these two vectors cannot be used in conjunction for complementation assays because they are both derivatives of pACYC184 (i.e. they have the same origin of replication and antibiotic resistance).

3.6 Binding specificity test

3.6.1 Rationale and design for the experiment

Another important aspect of ZFR binding domain variants is their binding specificity. Cross-reactivity between ZFR binding domain variants and non-cognate binding sites will inherently confound the results of ZFR complementation experiments, which rely on site-specificity. In order to test the binding fidelity of the new ZFR binding domain variants, an experiment was constructed utilizing three types of complementation substrates: 2MutHomDim, 2MutHetDim-DR, 2MutHetDim-IR (see Table 3-2). In this experiment, ZFR binding domain variant types were expressed in isolation, and reacted with complementation substrates for which they had 50% site-specificity.

The 2MutHomDim substrate contains two different homodimer binding sites. It can detect aberrant activity if a properly bound homodimer forms a productive tetramer with another improperly bound homodimer at the other Z-site. The 2MutHetDim-DR and 2MutHetDim-IR substrates, which are bound by heterodimers, can detect aberrant activity when subunits properly bound at their cognate half-sites, co-opt other same-type subunits to improperly bind the other half of the heterodimer site. The contacts of the dimer interface of the Tn3 resolvase catalytic domain are extensive. In conjunction with the DNA binding contacts of the Tn3 resolvase arm region, which are invariant in these experiments, homodimers might improperly bind heterodimer sites if their DNA binding

domains are sufficiently permissive to allow this. Preliminary experiments (data not shown) demonstrated this possibility in a Tn3[NY G70C]-Zif268 plus Tn3[NY G70C]-HTH complementation (HTH is the Tn3 resolvase binding domain, see Section 1.6.2), where the HTH domain was found to be sufficiently permissive to allow homodimers to improperly bind heterodimer binding sites and produce recombination. The 2MutHetDim-DR and 2MutHetDim-IR substrates were expected to give comparable results, but both were included in order to observe any unexpected effects, and because redundancy in the experiment increases the reliability of the results.

In this experiment, the Tn3[NM] catalytic domain (where 'N' represents the mutations R2A and E56K, and 'M' represents G102S, D102Y, M103I, and Q105L; see Burke et al. (2004)(Burke et al., 2004)) was chosen because it was believed to confer a higher level of recombination activity to the ZFR than the Tn3[NY G70C] domain used for the previous activity level tests. The higher activity of Tn3[NM] increases the sensitivity of the experiment to any cross-reactivity. The Tn3[NM] domain is also considered to be a 'well behaved' high activity catalytic domain, in that it is thought to produce minimal levels of DNA breakage from aberrant catalytic reactions, unlike some other known high activity domains ((Olorunniji et al., 2008), and unpublished work within the Stark lab).

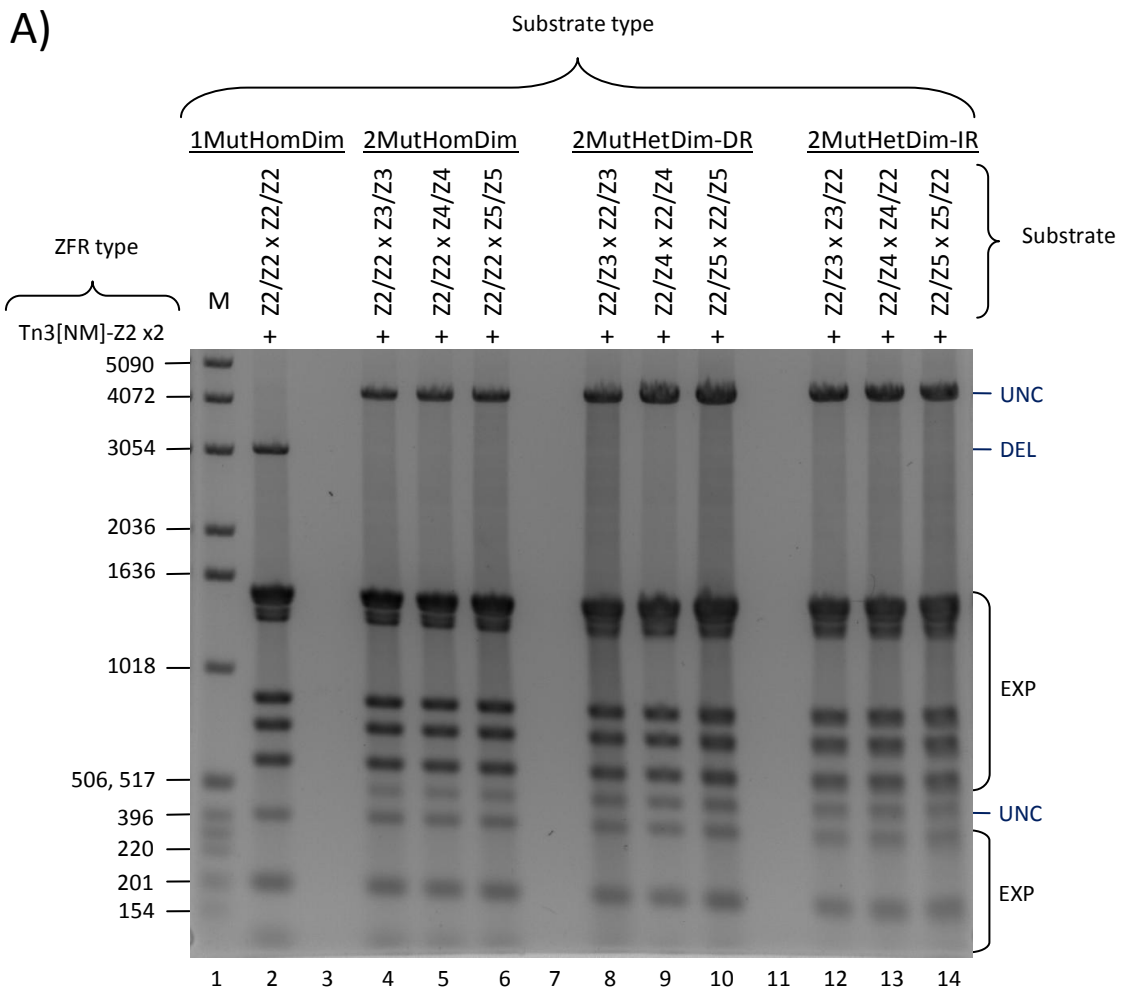
For the binding specificity assay, both pMS140 and pSA3022 vectors were used to simultaneously express the same type of each ZFR binding domain variant. It can be seen from the activity level experiments in Section 3.5 that activity level varies with expression plasmid, and by implication, with the level of enzyme production. This indicates that the binding sites are not saturated with ZFRs at these levels of enzyme production. Therefore, increasing the production of ZFRs within the cell has the potential to increase stringency of the test. A 17 Hour Recombination Assay was thus performed using two expression vectors.

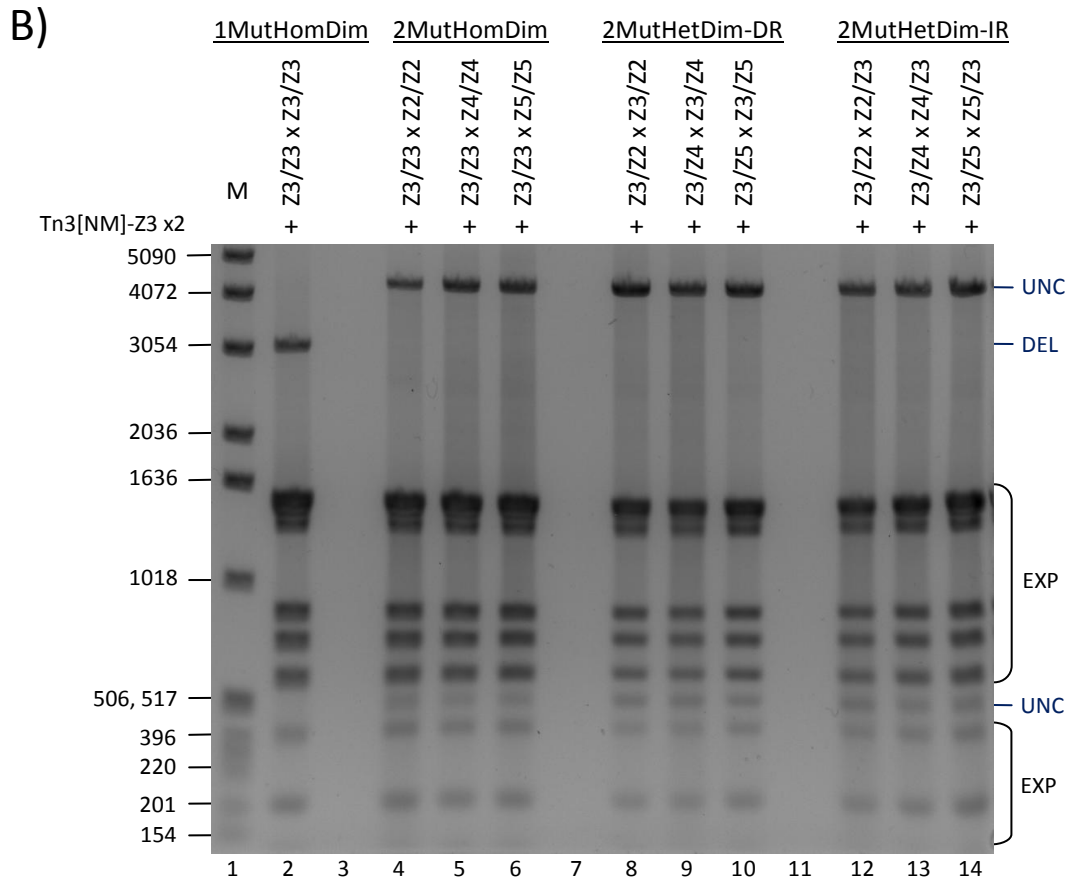
3.6.2 New ZFRs perform with high-level binding specificity

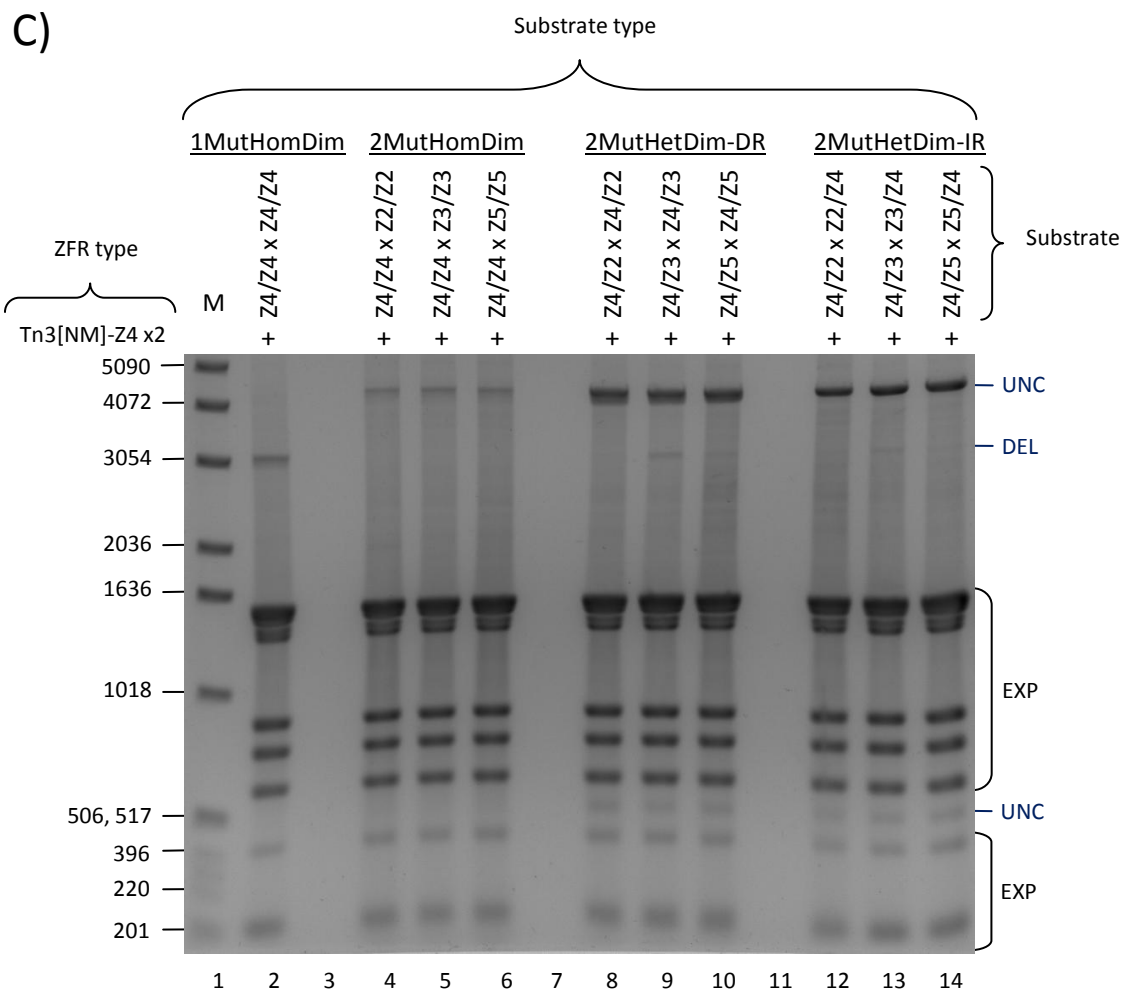
The results, shown in Figure 3-5, indicate that all ZFR binding domain variants appear to be highly specific for their binding sites, within the level of detection of the assay. With the exception of one faintly detectable band in lane 9 of Figure 3-5C, no other off-target activity was detectable.

Curiously, the product signal attenuation effect that was observable for Tn3[NY G70C]-Z4 in Figure 3-4, was again observable, not only in the control lane 2, containing Tn3[NM]-Z4 in conjunction with its cognate 1MutHomDim substrate, but also in lanes 4–6, containing Tn3[NM]-Z4 in conjunction with the 2MutHomDim substrate (Figure 3-5 C). This is interesting because the attenuated signals in lanes 4–6 represent the unchanged substrate products of the reaction, indicating that the plasmid level reduction might not be dependent on tetramer formation and recombination activity. Also interestingly, the signal attenuation effect does not appear as pronounced with 2MutHetDim-DR and 2MutHetDim-IR substrates, indicating that properly bound homodimers appear to be more facilitative to this effect.

The signal level in lanes 8–10 and 12–14 of Figure 3-5 C, might suggest an attenuation effect happening for Tn3[NM]-Z4 on the 2MutHetDim-DR and 2MutHetDim-IR substrates as well, to a lesser degree. However, the variability in the signal strength across the tracks in Figure 3-5 A, B and D, suggests that there is also some 'normal' variability in signal strength within this experiment. Presumably the normal range of variability in signal strength, which can be observed across the tracks of the gels, is simply the result of variability in the DNA yield from the 17 Hour Recombination Assay plasmid harvest. The bands representing the expression plasmids provide a good control to differentiate between differences in signal strength that arise from the interesting reduction in substrate plasmid concentration within the cells, and overall reductions in signal strength resulting from variations in DNA yield of the 17 Hour Recombination Assay.


A)
Substrate type
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
Substrate
ZFR type
M
Z2/Z2 x Z2/Z2
Z2/Z2 x Z3/Z3
Z2/Z2 x Z4/Z4
Z2/Z2 x Z5/Z5
Z2/Z3 x Z2/Z3
Z2/Z4 x Z2/Z4
Z2/Z5 x Z2/Z5
Z2/Z3 x Z3/Z2
Z2/Z4 x Z4/Z2
Z2/Z5 x Z5/Z2
Tn3[NM]-Z2 x2
+
5090
4072
3054
2036
1636
1018
506, 517
396
220
201
154
UNC
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14


B)
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
M
Z3/Z3 x Z3/Z3
Z3/Z3 x Z2/Z2
Z3/Z3 x Z4/Z4
Z3/Z3 x Z5/Z5
Z3/Z2 x Z3/Z2
Z3/Z4 x Z3/Z4
Z3/Z5 x Z3/Z5
Z3/Z2 x Z2/Z3
Z3/Z4 x Z4/Z3
Z3/Z5 x Z5/Z3
Tn3[NM]-Z3 x2
+
5090
4072
3054
2036
1636
1018
506, 517
396
220
201
154
UNC
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14


C)
Substrate type
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
ZFR type
Substrate
M
Z4/Z4 x Z4/Z4
Z4/Z4 x Z2/Z2
Z4/Z4 x Z3/Z3
Z4/Z4 x Z5/Z5
Z4/Z2 x Z4/Z2
Z4/Z3 x Z4/Z3
Z4/Z5 x Z4/Z5
Z4/Z2 x Z2/Z4
Z4/Z3 x Z3/Z4
Z4/Z5 x Z5/Z4
Tn3[NM]-Z4 x2
+
5090
4072
3054
2036
1636
1018
506, 517
396
220
201
UNC
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14

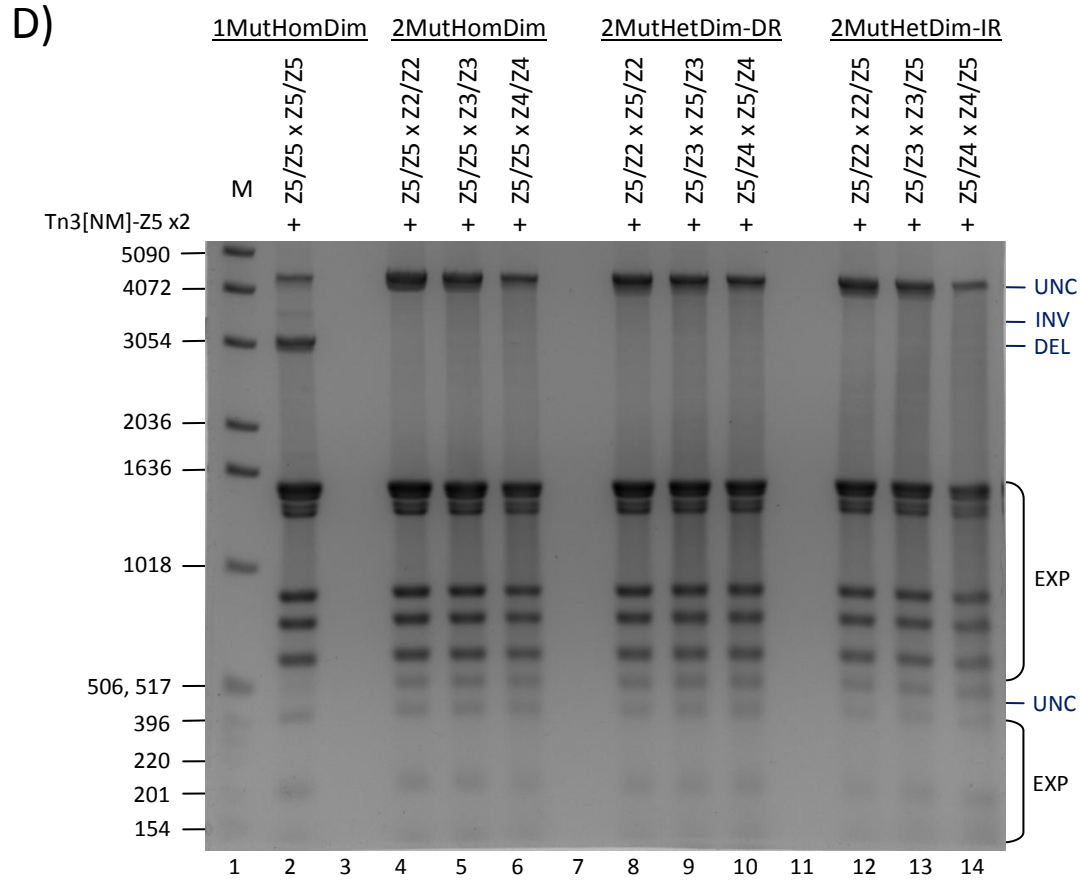

D)
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
M
Z5/Z5 x Z5/Z5
Z5/Z5 x Z2/Z2
Z5/Z5 x Z3/Z3
Z5/Z5 x Z4/Z4
Z5/Z2 x Z5/Z2
Z5/Z3 x Z5/Z3
Z5/Z4 x Z5/Z4
Z5/Z2 x Z2/Z5
Z5/Z3 x Z3/Z5
Z5/Z4 x Z4/Z5
Tn3[NM]-Z5 x2
+
5090
4072
3054
2036
1636
1018
506, 517
396
220
201
154
UNC
INV
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 3-5: New ZFRs: targeting specificity assay results. The images show the results of the *Fsp*I + *Kpn*I + *Pvu*II restriction digest on a 1.2% agarose gel. To the right of the images, indications are given for bands corresponding to the unchanged substrate (UNC), inverted substrate product (INV), deletion substrate product (DEL), and the expression plasmids (EXP). To the left of the images **A)** and **C)**, sizes are given for the 1 kb marker ladder which is contained in lane 1 of all gels. Lane 2 in all images displays a positive control which consists of a ZFR reaction on its cognate 1MutHomo substrate. Lanes 4–6, 8–10, and 12–14 in each gel image show the results of ZFRs in conjunction with 2MutHomDim, 2MutHet-DR, or 2MutHetDim-IR substrates, respectively, as indicated above the image. The experimental substrates contain combination binding sites consisting of cognate half-sites and the half-site indicated above the image. **A)** ZFR= Tn3[NM]-Z2. **B)** ZFR = Tn3[NM]-Z3. **C)** ZFR = Tn3[NM]-Z4. **D)** ZFR = Tn3[NM]-Z5. All ZFRs were used in dual expression mode, i.e. expressed from pMS140 and pSA3022 vectors simultaneously in each of the experiments shown.

3.7 Alteration of ZFR linker

3.7.1 Rationale and design for the experiment

The ZFAs (Z2, Z3, Z4 and Z5) were predicted from bacterial-2-hybrid (B2H) assay data in the Maeder *et al.* (2008) study (see Table 3-1) to impart ZFRs with similar levels of binding affinity, and thus recombination activity (Maeder et al., 2008). However, not only are the ZFR variants not as similar in activity as expected, but the activity also does not vary predictably with the variation in B2H activity scores of the ZFA domains. Additionally, the product signal attenuation effect observed in the last two experiments for reactions involving ZFRs with the Z4 ZFA, would make the use of Z4 in ZFRs for future experiments problematic.

One possible explanation for these variations was that not all ZFAs are equally compatible with the current ZFR architecture. The ZFR amino acid linker (named 'TS linker') that currently connects the Tn3 arm region of the catalytic module (ending at Arg148) to the ZFA, was optimized for attachment of the Zif268 ZFA, and is very short, consisting of only a Thr-Ser dipeptide (see Figure 3-1 A). The docking geometry of ZFAs may be somewhat variable depending on the residue-base contacts that are required for DNA sequence recognition. It is possible that the TS linker, though optimal for the Tn3-Zif268 ZFR, does not provide sufficient flexibility at the attachment point to permit proper residue-base contacts for all attached ZFAs.

The original experiments with Tn3[NM]-Zif268 performed by Akopian *et al.* (2003) utilized much longer peptide linkers that could be safely assumed to provide sufficient flexibility at the attachment point to allow proper binding of any ZFA (Akopian et al., 2003). Therefore, one of the top performing linkers from the Akopian study ('L6') was used to replace the current linker in the current ZFR architecture. This linker is composed of four Gly-Ser-Gly repeats followed by Thr-Ser. Compared with the TS linker architecture, the attachment point of L6 to the Tn3 catalytic domain replaces more of the Tn3 arm region. Linker L6 begins after Tn3 residue Arg144, whereas the TS linker, begins after residue Arg148. Therefore, the total difference in peptide length is nine residues, rather than 12. The Gly-Ser-Gly repeat sequence was originally used because it was shown in early ZFA engineering studies to be fit for making flexible peptide linkers connecting zinc fingers (Moore et al., 2001b, Moore et al., 2001a).

Four new ZFRs were constructed (Tn3[NY G70C]-L6-Z2, Tn3[NY G70C]-L6-Z3, Tn3[NY G70C]-L6-Z4, and Tn3[NY G70C]-L6-Z5) in order to replicate the activity level experiment of Section 3.5 using ZFRs incorporating the L6 linker. The purpose of this was simply to look for a normalization of the pattern of activity witnessed in Section 3.5, and only ZFRs utilizing linker L6 were used for this experiment.

3.7.2 ZFR linker is not responsible for the product attenuation anomaly

The results from this experiment, which are displayed in Figure 3-6, show a virtually identical pattern of activity using the L6 linker as was observed in Section 3.5 (Figure 3-4) using the TS linker. This demonstrates clearly, that neither the variability in activity between the ZFR binding domain variants, nor the product signal attenuation effect seen in reactions using ZFRs with Z4, are the result of the linker sequence. The only noteworthy feature in the results, is that compared with the previous activity test using the TS linker and the same ZFR catalytic and binding modules, also expressed from a pMS140-type vector (Figure 3-4, lanes 7–10), there does appear to be an overall reduction in activity when using L6 linker (Figure 3-6, lanes 2–5). This reduction in activity is indicated by some remaining unchanged substrate in the reaction with Tn3[NY G70C]-L6-Z3 (lane3), and almost no recombination activity from Tn3[NY G70C]-L6-Z5 (lane 5), relative to the comparable reactions in Figure 3-4 (lanes 8 and 10, respectively). This

result, indicating that ZFRs containing the L6 linker are less active than those containing the TS linker, is consistent with the results of an *in vitro* recombination assay reported in Prorocic et al. (2011), although difference is less pronounced in the *in vivo* assay reported here.

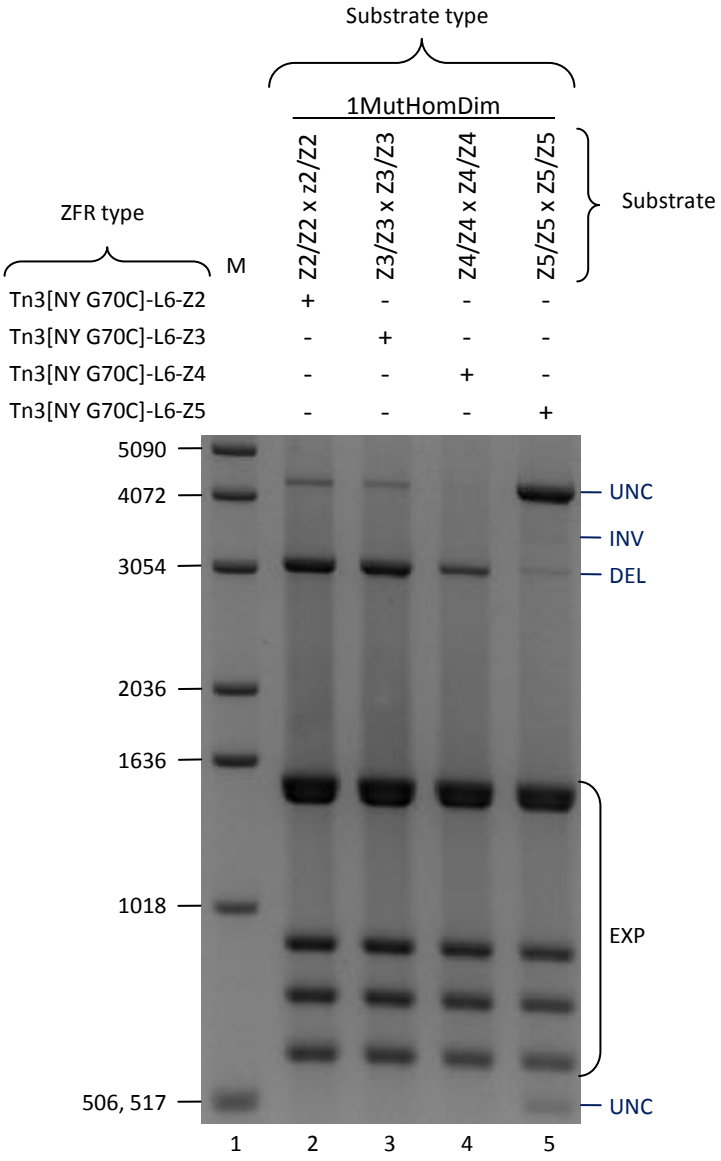


Figure 3-6: Results of Junction Modification Experiment. The image shows the results of the *Fsp*I + *Kpn*I + *Pvu*II restriction digest on a 1.2% agarose gel. To the right of the image, indications are given for bands corresponding to the unchanged substrate (UNC), inverted substrate product (INV), deletion/excision substrate product (DEL), and the expression plasmids (EXP). To the left of the image, sizes are given for the 1kb marker ladder, which is contained in lane 1. Lanes 2–5 show the results of the reactions of the junction modified ZFRs (Tn3[NY G70C]-L6-Z2–Z4) on their cognate 1MutHomo substrates, expressed from pMS140-based expression vectors.

3.8 Binding activity and specificity comparisons with the Tn3-Zif268 ZFR

3.8.1 Rationale and design of the experiment

Because of the product signal attenuation effect encountered when Tn3-Z4 ZFRs were expressed from the pMS140 and pSA3022 vectors, and because of the relatively low level of activity of the Tn3-Z5 ZFRs compared with the other binding domain variants, it was thought the Zif268 (hereafter referred to as Z1) ZFA might be required as a substitute for use in conjunction with the other binding domain variants in further experiments. It was therefore necessary to compare Tn3-Z1 ZFR activity and binding specificity to activity and specificity of the other ZFR binding domain variants. In addition, these experiments also provide a performance comparison between ZFRs utilizing OPEN engineered ZFAs and that utilizing the Z1 ZFA, which is a well known natural ZFA considered to have a good activity and specificity profile.

Substrates containing combinations of Z-sites based on Z1–Z5 ZFRs were constructed in the fashion described in Section 3.3. It should be noted, that the sequences for these Z-sites were derived from a pre-existing Z1 ZFR Z-site that does not have the correct base at the 3′ F1 overlap position (see Section 1.7.1.2) in the right half of the Z-site. This Z-site has a cytosine at the right side 3′ F1 overlap position, whereas a thymine or guanine base is specified by Zif268 (Swirnoff and Milbrandt, 1995, Wolfe et al., 1999). This was not noticed at the time of construction or it would have been corrected. It should also be noted, that the Z1 1MutHomDim substrate used in the following experiment was a pre-existing substrate that, additionally, does not have correct 5′ extended position bases. The 5′ extended position bases in the 1MutHomDim substrate are a cytosine in the left half of the Z-site and an adenine in the right half of the Z-site, instead of the thymines that should be at these positions (Swirnoff and Milbrandt, 1995, Wolfe et al., 1999). Therefore, in all but the case of the 1MutHomDim substrate, within the new substrates when a Z1 ZFR half-site occurs on the right half of the Z-site, this 3′ F1 overlap position base is incorrect, as described above. In all other cases, except the 1MutHomDim substrate, both the 5′ extended position (also see Section 1.7.1.2) and 3′ F1 overlap positions are correct.

The activity and specificity tests were carried out as described in Sections 3.5 and 3.6, respectively. For the activity test, a Tn3[NY G70C] catalytic domain was used in conjunction with the Z1 and Z3 ZFAs, which were paired with their cognate 1MutHomDim substrates. For the fidelity test, a Tn3[NM] catalytic domain was used in conjunction with the Z1, Z2, Z3, Z4, and Z5 ZFAs, which were paired with 2MutHomDim, 2MutHetDim-IR and 2MutHetDim-DR substrates (on which they bind half of the binding sites; see Table 3-2). All experiments were carried out using the 17 Hour Recombination Assay.

3.8.2 Results of a binding activity assay comparing ZFRs using the Z1 and Z3 ZFAs.

The results of the binding activity assay (displayed in Figure 3-7), show that the activity level of the ZFR using Z1 is virtually identical to the activity of the ZFR using Z3. One curious feature though, is the appearance of a strong inversion product signal in lane 2, versus its near absence in lane 3. Usually in reactions involving active ZFRs and 1MutHomDim substrates, the inversion product signal is weak, because the inversion product is subsequently converted to either excision product, or back into unchanged substrate (sees Figure 1-10). The Z1/Z1 x Z1/Z1 substrate was constructed using a different cloning strategy than the Z3/Z3 x Z3/Z3 substrate, which resulted in a small difference in plasmid size (responsible for the slight difference in the size of the small 'unchanged substrate' bands on the gel), but no obvious differences between the substrates exist that might lead to a difference in their ability to be converted to inversion product.

One explanation for the discrepancy between the inversion product signal strength in lane 2 and 3 of Figure 3-7 is that the inversion product from the Z1 ZFR reaction was more stable than the inversion product of the Z3 ZFR reaction. As described in Section 3.8.1, there is a difference in the 5′ extended positions between the right and left half sites of the 1MutHomDim substrate. Additionally, (also described in Section 3.8.1) because the 3′ F1 overlap position base was correct in the left half of the Z-sites of the 1MutHomDim substrate, but incorrect in the right half of the Z-sites (Z1/Z1 x Z1/Z1), inversion reactions would have produced a product substrate where one whole Z-site contained incorrect 3′ F1 overlap bases (Z1/Z1 x Z1/Z1), as well as the difference in 5′ extended positions that would also co-localize to the same Z-sites within the inversion

product. Having Z-sites with incorrect 3′ F1 overlap positions and the same incorrect 5′ extended position bases (adenine or cytosine) within the half-sites, might have slowed down subsequent recombination reactions on these substrates enough to produce the lane 2 and lane 3 discrepancy observed (since in both lanes 2 and 3 the reactions were slow enough that they did not go to completion). The experiment was not repeated to confirm the result due to time constraints. This result suggests, if the foregoing interpretation of it is correct, that differential binding affinity among ZFRs in a complementation reaction may be used to stabilize a recombination product that contains homodimer Z-sites (that pair the more weakly binding subunits).

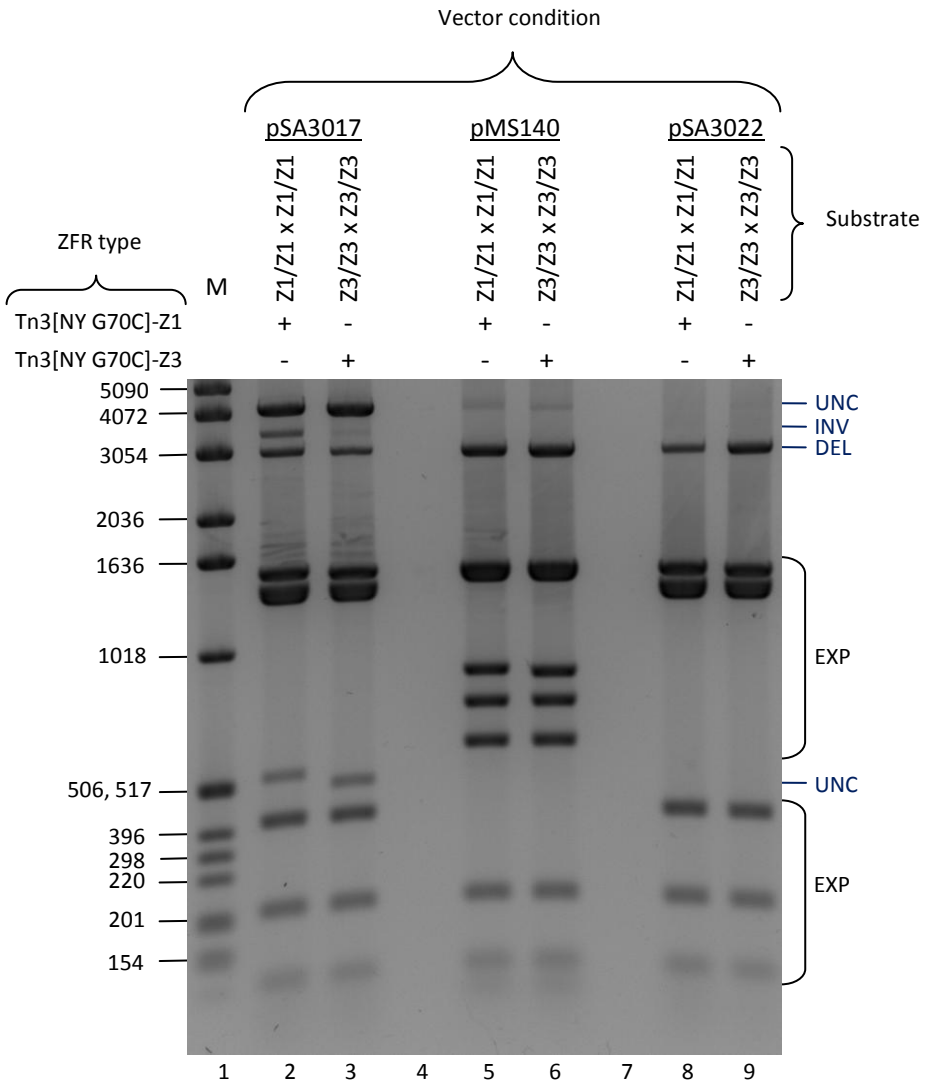


Figure 3-7: Comparative activity test of ZFRs with Z1 and Z3. The results of a 17 Hour Recombination Assay are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. Please note, the difference in the small unchanged substrate band size is due to the fact that the Z1/Z1 x Z1/Z1 substrate was constructed using a different cloning strategy from the Z3/Z3 x Z3/Z3 substrate, which resulted in a small difference in the size of the plasmids.

3.8.3 Results of a binding specificity test comparing the ZFR using the Z1 ZFA to ZFRs using the Z2, Z3, Z4 and Z5 ZFAs

An experiment was conducted that tested the binding specificity of the Tn3[NM]-Z1 ZFR on substrates that bore 50% Z1 site specificity, and 50% site specificity to Z2, Z3, Z4, and Z5. The Tn3[NM] catalytic domain was used because it was thought be more active than the Tn3[NY G70C] domain, and thus, could produce a more rigorous test (although, a later experiment in Chapter 4 suggests these catalytic domains may have similar activity; see Figure 4-20). The ZFR was expressed from two expression vectors (pMS140-based and pSA3022-based) simultaneously to increase the concentration of ZFR in the cells, again producing the most rigorous test possible. This experiment is similar to the previous binding activity experiment characterizing the Z2, Z3, Z4, and Z5 (Section 3.6)

The results of this experiment (shown in Figure 3-8) indicate that the Tn3[NM]-Z1 ZFR is slightly cross-reactive on all heterodimer binding sites (lanes 7–10, 12–15). Also worth noting, the unchanged substrate and recombination product signal attenuation effect appeared again on all 2MutHomDim substrates (although the effect was less pronounced on the Z5/Z5 x Z1/Z1 substrate; lane 5). Lane 5 might have a bit more DNA in it, as the slowest running expression plasmid signal in lane 5 also appears to be slightly stronger than the others in the 2MutHomDim set (lanes 2–5). It is curious that this substrate contains Z5 sites, and that it is the Tn3-Z5 ZFRs on Z5 containing substrates that produce the least amount of signal attenuation effect. However, only the Tn3-Z1 ZFR is used in this reaction, and if it bound to the Z5 homodimer Z-site, it would likely produce detectable recombination activity. Therefore, no further explanation of this anomaly can be provided.

There is also an anomalous weak signal at about 4.1 kb in lanes 2–5 and 7–9 (it cannot easily be see without the aid of a computer screen). It is distinct from the unchanged substrate signal and might indicate either, trace levels of undigested expression plasmid, or perhaps low-level inappropriate cleavage of the substrate.

Another experiment was conducted to test the binding specificity of ZFRs using the Z2, Z3, Z4, and Z5 ZFAs on substrates that were 50% cognate and 50% cognate for Z1. The Tn3[NM] catalytic domain was again used for the ZFRs, which were also expressed

from two expression vectors simultaneously (pMS140-based and pSA3022-based) in order to increase the rigorousness of the test. The results of these experiments (shown in Figure 3-9) indicate that there is no cross-reactivity of any of the ZFRs for the substrates containing 50% Z1 sites. It is noteworthy that the substrate attenuation effect again appeared for the Tn3[NM]-Z4 ZFR on all substrate types (lanes 4, 9, and 14), and also for the Tn3[NM]-Z2 and Tn3[NM]-Z3 ZFRs on the 2MutHomDim substrates (lanes 2 and 3).

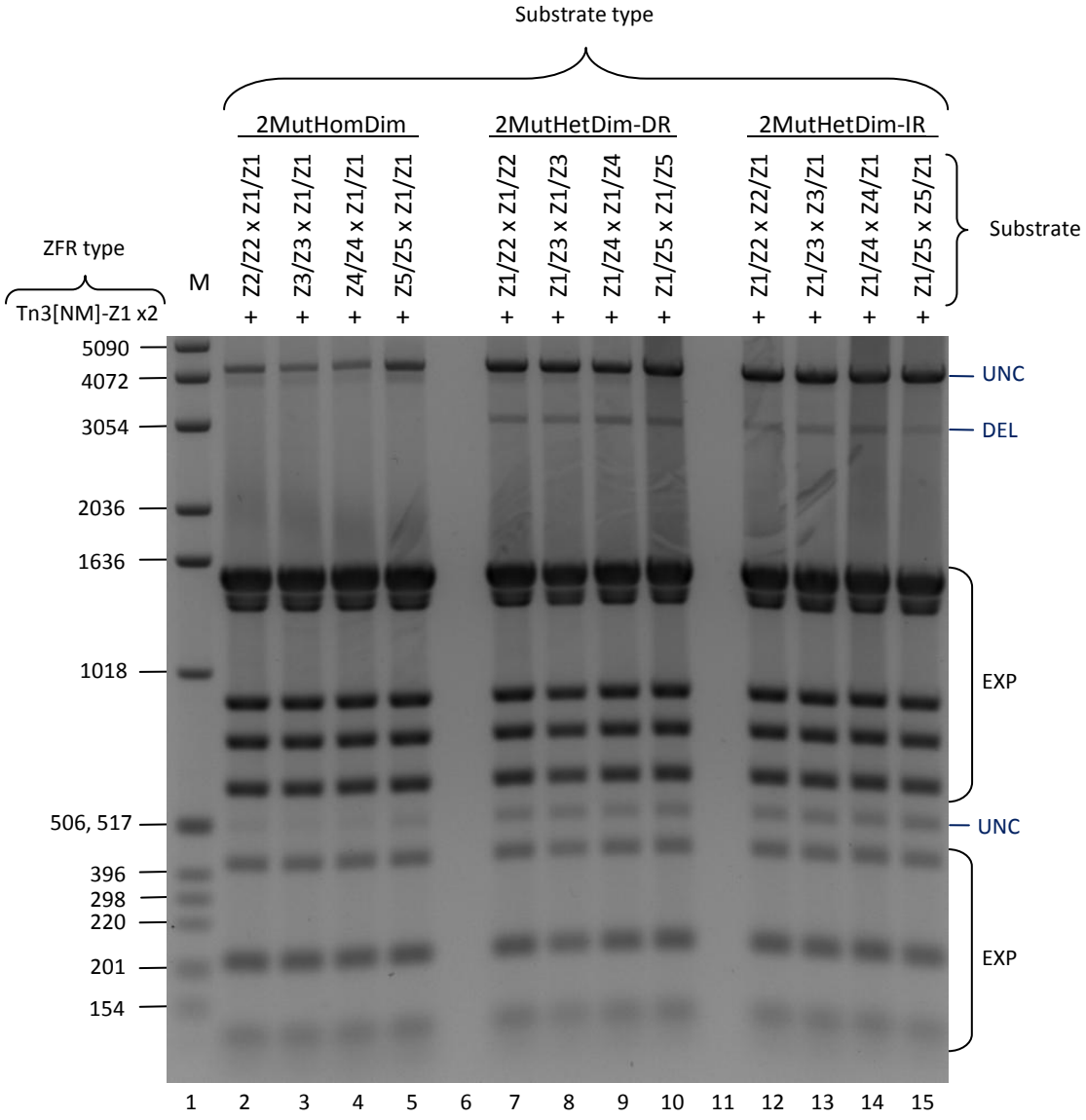


Figure 3-8: Binding specificity of a ZFR with the Z1 ZFA on substrates for other available binding domain variant ZFRs. The results of a 17 Hour Recombination Assay are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, DEL = deletion product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions.

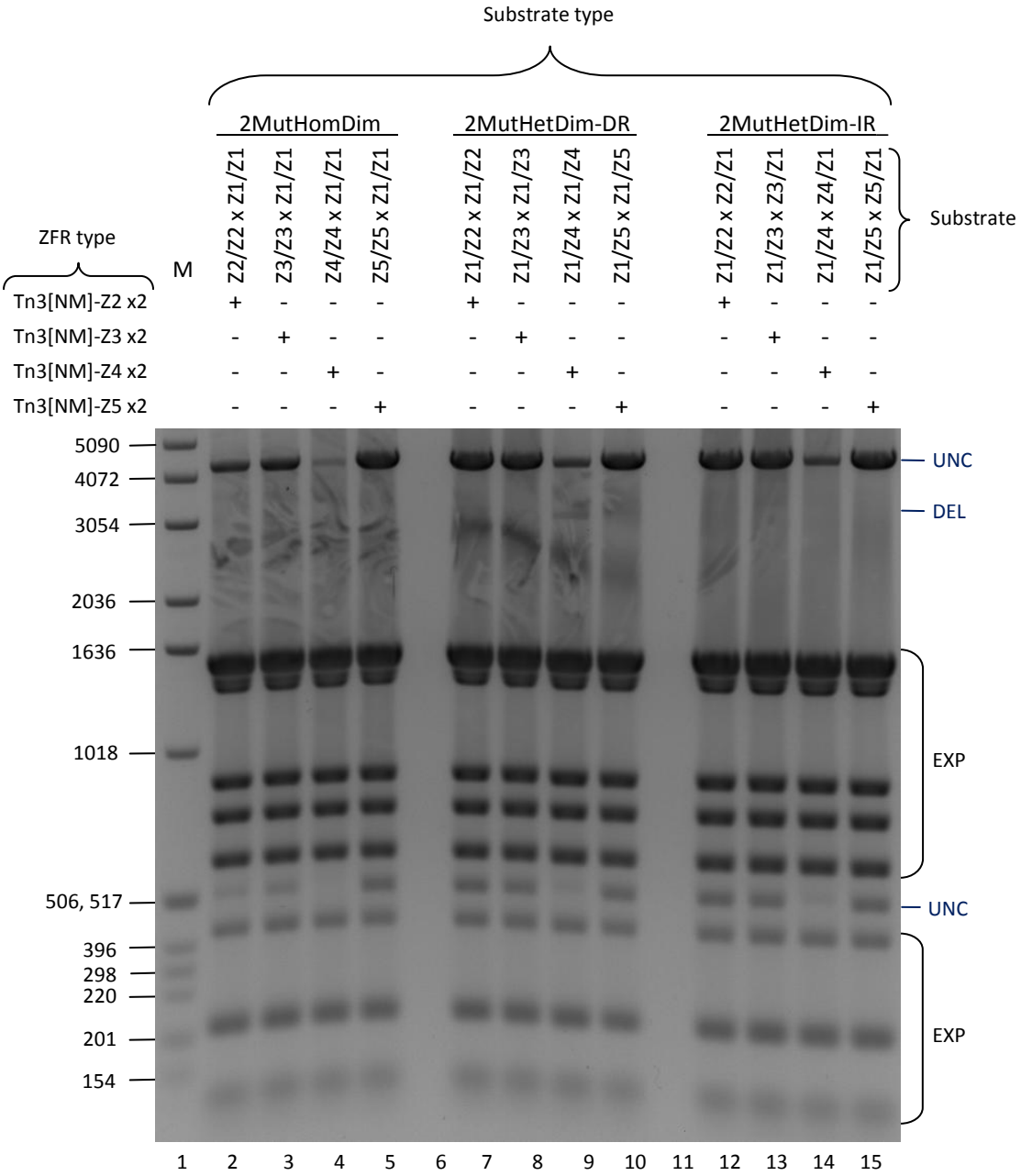


Figure 3-9: Binding specificity test of ZFRs on substrates containing Z1 ZFA binding sites. The results of a 17 Hour Recombination Assay are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions.

3.9 Construction and testing of double expression vectors

3.9.1 Rationale and design

Because later experiments in this work involve detecting the activity of tetramers composed of four unique ZFR subunits, a system for the simultaneous expression of four ZFRs in *E. coli* is required. Transformation of *E. coli* with four separate expression vectors plus one substrate plasmid is impractical. Simultaneous transformation of five plasmids would not be sufficiently efficient, sequential transformation of multiple plasmids to create multi-plasmid *E. coli* strains would be time-consuming, simultaneous use of five antibiotics might be problematic, and finding four compatible low-level expression vectors with comparable levels of expression would likely be very difficult.

A double-expression vector (DEV) based on pMS140 had been previously designed and constructed (Proudfoot et al., 2011). This design allowed two ZFR coding sequences, each with their upstream Shine-Dalgarno sequence, to be placed in tandem (separated by a short nucleotide sequence) downstream of a promoter within the ordinary ZFR expression region of pMS140. To accomplish this, the Tn3 resolvase reading frame within pMS140 was replaced with a short segment of DNA that allowed two ZFR coding sequences to be sequentially ligated into the plasmid using isocaudomer sites (restriction sites with different recognition sequences, but which produce identical overhangs).

Two new DEV frameworks for the pMS140 and pSA3022 expression vectors were designed based on a similar approach where two reading frames, each with their Shine-Dalgarno sequence, are placed downstream of one promoter and separated by a short sequence of DNA (see Figure 3-10). However, in order to speed up construction, this framework utilizes a short double-stranded oligonucleotide in a three-part ligation strategy to create the DEV in one step, instead of the sequential insertion of two coding regions. In this design, two single-ZFR expression vectors are cut at a backbone restriction site, and a restriction site at either the 3’ or 5’ end of their coding sequence, depending on whether the ZFR coding sequence is to be placed in the first or second expression ‘slot’, respectively. An oligonucleotide is then used to separate the 3’ end of the first reading frame, and the 5’ end of the second. The three pieces of DNA are then ligated together.

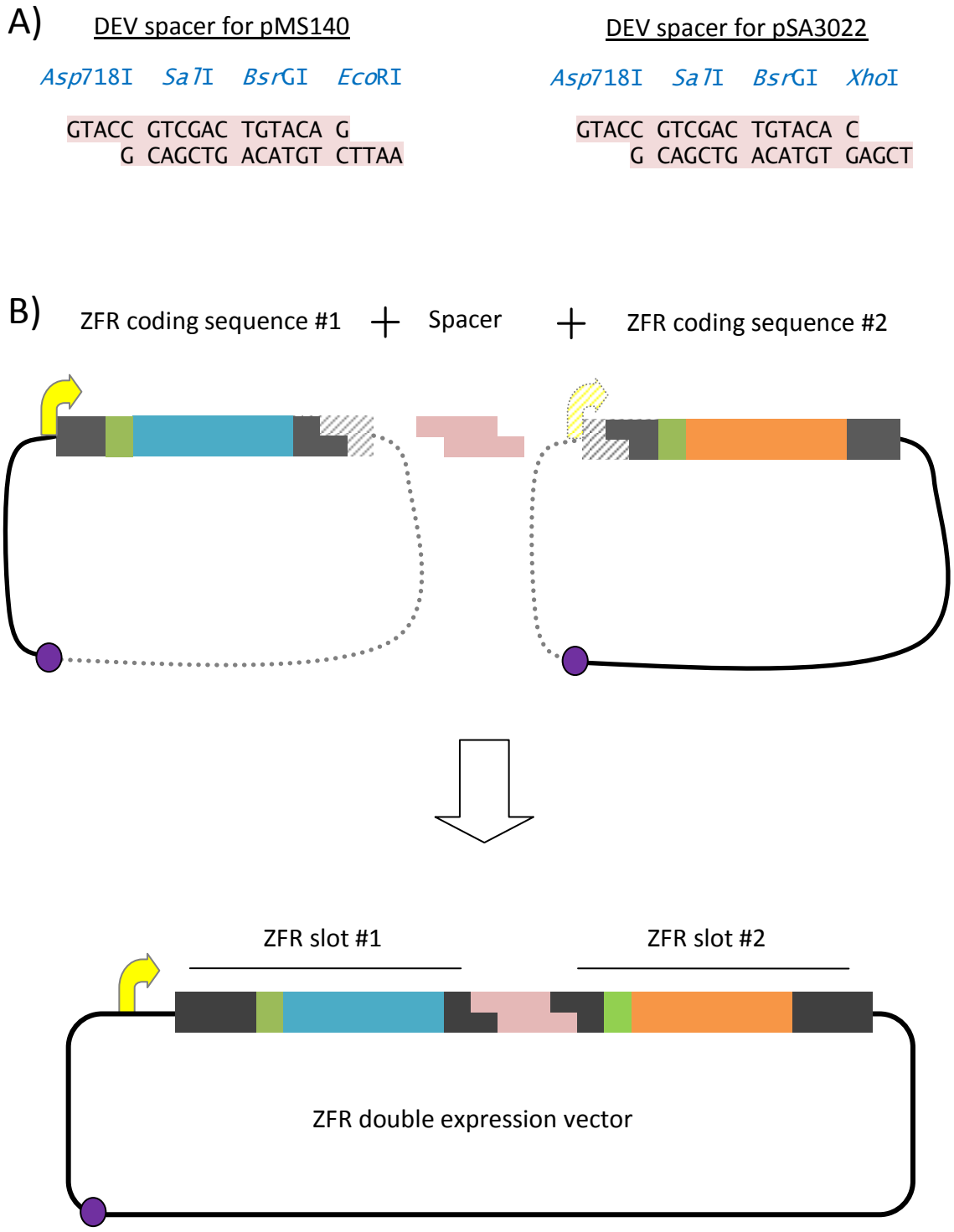


Figure 3-10: DEV construction strategy. A) The figure shows the sequences, including restrictions sites, of the DEV-enabling oligonucleotide spacers for pMS140 and pSA3022. **B)** The figure depicts the DEV construction framework. ZFR coding sequences are depicted as horizontal rectangles with overhangs. Blue and orange colours are used to indicate the coding regions, while green indicates the Shine-Dalgarno sequence which is included within the fragments. A purple circle is used to depict the restriction site used for cutting and ligating the backbones of the two expression vectors containing the ZFR coding sequences. Light grey dots and the faded out portions of the rectangles, are used to indicate the parts of the vector plasmids which are not retained for the assembly of the DEV. Dark grey is used to depict restriction site overhangs which are used for ligation. The spacer sequence is shown in pink. The expression vector backbone is shown with upstream promoter element indicated by a yellow arrow. Below, the product DEV is shown with ZFR 'slots' #1 and #2 labelled and the colour scheme depicting the organization of the DEV framework.

For the pMS140-based DEV, coding sequences to be placed in slot 1 are taken from *Bsa*I to *Asp*718, while coding regions to be placed in slot 2 are taken from *Eco*RI to *Bsa*I. The pMS140-based DEV oligonucleotide spacer has *Asp*718- and *Eco*RI-compatible overhangs. For the pSA3022-based DEV, coding sequences to be placed in slot 1 are taken from *Nco*I to *Asp*718, and coding sequences to be placed in slot 2 are taken from *Xho*I to *Nco*I sites. The pSA3022-based DEV oligonucleotide spacer has *Xho*I- and *Sal*I-compatible overhangs.

A *Sal*I restriction site was included in the oligonucleotide spacers in order to confirm their insertion by restriction digest after assembly. A *Bsr*GI site was also added to the oligonucleotides to allow assembled DEVs to be easily rearranged into new combinations (including the possibility for testing triple or even quadruple expression arrangements) should that have been deemed interesting.

Ideally, each ZFR encoded by the DEV would be expressed at approximately the same level, so a new experiment was constructed in order to test the individual expression level of each ZFR coding cassette. Each DEV (pMS140- or pSA3022-based) was loaded with both the arrangement 'Tn3[NY G70C]-Z2 followed by Tn3[NY G70C]-Z3', and the arrangement 'Tn3[NY G70C]-Z3 followed by Tn3[NY G70C]-Z2'. These DEVs were then transformed into *E. coli,* along with either the Z2 or Z3 1MutHomDim substrate for a 17 Hour Recombination Assay. Thus, for each of the four constructed DEVs, two activity tests were performed: one assaying the expression level of Tn3[NY G70C]-Z2, and the other the expression level of Tn3[NY G70C]-Z3, in both upstream and downstream cassette positions. This design also provides redundancy in the results which increases their reliability.

3.9.2 Results of the DEV expression tests

The results from the DEV experiment (shown in Figure 3-11) indicate that the DEVs express equal levels of ZFR from each expression slot, at least within the limit of detection of the assay. The primary concern was that the second slots might have markedly reduced levels of expression, and this is clearly not the case.

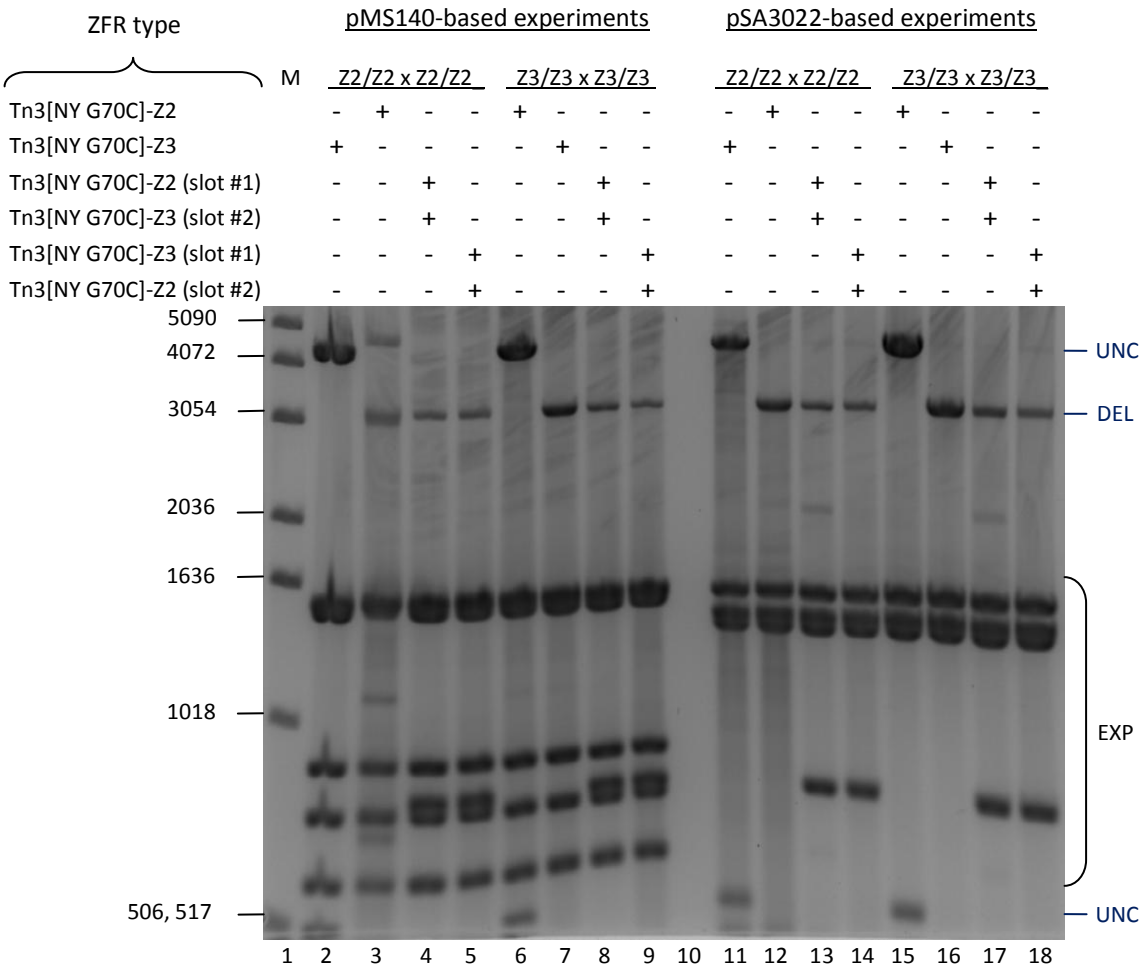


Figure 3-11: Experiment to check expression level from each DEV expression 'slot'. The results of a 17 Hour Recombination Assay are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion product, and EXP = expression vector plasmid). Above the gel image is a table that indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. Please note that the digest shown in lane 3 did not run to completion and there is some partially digested expression plasmid in this lane. The band between 4 kb and 5 kb in lane 3 is not substrate but expression plasmid.

Compared with the single ZFR expression, positive controls in lanes 3, 7, 12 and 16 there again appears to be a recombination product signal attenuation effect present in the DEV tracks (lanes 4, 5, 8, 9, 13, 14, 17 and 18). Importantly, it can be seen from this experiment that the product signal attenuation effect is enhanced by the co-expression of ZFRs that are non-cognate for any Z-sites. Compared with the reactions in which only the cognate ZFR was expressed (lanes 3, 7, 12 and 16), it can been seen that the signal level of the product is decreased in the reactions where a non-cognate ZFR was also expressed (lanes 4, 5, 8, 9, 13, 14, 17 and 18).

3.10 Conclusions

3.10.1 New ZFA binding domains: activity and specificity

New ZFAs were required for incorporation into the ZFR system. Four ZFAs (Z1, Z2, Z3, and Z4) were selected from published literature on engineered ZFAs. New ZFRs incorporating the new ZFAs were tested for recombination activity (Section 3.5) and binding specificity (Section 3.6). These experiments indicated that the Z2, Z3, Z4, and Z5 ZFAs bind with high specificity. However, the new ZFRs did not produce equal levels of recombination activity in the assays, which indicates that Z2, Z3, Z4, and Z5 ZFAs do not bind with equal binding affinity. The Z2 and Z3 ZFAs appeared ideal for use in further experiments, imparting their respective ZFRs with equal levels of recombination activity in the assays. The Z5 domain can be inferred to bind somewhat more weakly than the other three ZFAs since its ZFRs produced less recombination activity in the assays. ZFRs with the Z4 domain appeared to produce the highest level of recombination, and thus, the Z4 domain can be inferred to have the highest binding activity of the all the new ZFAs. It is worth noting that the mismatch in the target site at the Z4 ZFA F1 3′ overlap specificity position (see Section 3.3.2) did not appear to significantly impede the function of Z4.

In reactions involving the Z4 ZFR the amount of substrate and reaction products appeared noticeably attenuated when compared with reactions involving other the ZFRs. Data from an experiment using a vector with lower expression, suggested that the anomalous substrate and product attenuation effect result from Z4 binding more strongly than the other ZFAs, as the effect did not appear for the Z4 ZFR in the lower expression condition. See Section 3.10.5 for a full discussion on the product signal attenuation effect.

3.10.2 Linker experiment

In Section 3.7, an experiment, in which a longer linker was used to connect the Tn3 catalytic domain to the new ZFAs, tested whether the difference in the capabilities of the ZFA to impart ZFRs with recombination activity, or the substrate and product signal attenuation effect, were the result of using a linker that was too short. This experiment showed that linker length had no effect on the differences in ZFR activity, or product signal attenuation effect.

3.10.3 Comparisons of new ZFAs with Zif268

Recombination activity and binding specificity tests were also carried out comparing ZFRs with the new ZFAs, to ZFRs utilizing the Zif268 (Z1) ZFA. These experiments showed that Z3 (and by inference Z2) impart the ZFR with approximately the same level of recombination activity as Z1. However, ZFRs using Z2 and Z3 appear to bind their Z-sites with higher specificity than ZFRs using Z1, which showed a low level of cross-specificity for the other ZFA sites.

3.10.4 Double expression vectors

In Section 3.9, double expression vectors (DEVs) were designed, constructed, and tested in order to increase the number of ZFRs that could be expressed simultaneously within *E. coli*. These experiments demonstrated that the DEVs work well, and produce equal levels of enzyme activity for both Z2- and Z3-based ZFRs that were expressed from them.

3.10.5 The product signal attenuation effect

3.10.5.1 Description

A product signal attenuation effect was observed in several of the experiments described in this chapter. This signal attenuation effect is marked by a notable reduction in recombination product signal observed during agarose gel electrophoresis analysis, when compared to lanes where this effect is not present.

3.10.5.2 Instances of occurrence within the experiments

In Section 3.5, an experiment was conducted to observe the difference in activity level conferred to the ZFRs by the new ZFAs, and to observe the difference in activity level conferred by expression of ZFRs from three different expression vectors, which were expected to produce enzyme at different levels. The ZFRs were tested for activity on substrates with cognate Z-sites (1MutHomDim substrates). In this experiment, the signal attenuation effect was correlated with the ZFA (Z4), which appeared to confer the highest level of activity to its ZFR (Figure 3-4). The effect did not appear when the Z4 ZFR was expressed from the vector assessed to have the lowest level of enzyme production (pSA3017-based), and the effect appeared most pronounced when the Z4 ZFR was expressed from the vector assessed to have the highest level of enzyme production. An

experiment of Section 3.7 also demonstrated the attenuation effect using a Z4 ZFR (this time with altered linker) on its 1MutHomDim cognate substrate (Figure 3-6). The experiments in Section 3.5 and 3.7 both used NY G70C Tn3 catalytic domains.

The signal attenuation effect was again observed in the experiment of Section 3.6 that was designed to test the new ZFRs (and new ZFAs) on substrates for which they only bound to half of the binding sites. This experiment utilized simultaneous expression of the same ZFR from two expression vectors in order to approximately double the cellular enzyme concentration. In this experiment, the signal attenuation effect was even more pronounced in the reactions involving the Z4 ZFR, than it was in the previously mentioned experiments (Figure 3-5 C). The effect was most notable on the 1MutHomDim substrate and the 2MutHomDim substrate (which contains one full cognate Z-site and one full non-cognate Z-site). The attenuation effect may have also been present in the Z4 ZFR reactions on the 2MutHetDim substrates (containing Z-sites that were half-cognate to ZFRs used); however, the effect was more subtle. In this experiment the attenuation effect was less obvious for the ZFRs utilizing the other ZFAs (i.e. Z2, Z3, and Z5); although, it may have also been present in the Z2 and Z3 ZFR reaction on the 1MutHomDim and 2MutHomDim substrates. The experiments in Section 3.6 used NM Tn3 catalytic domains.

An experiment in Section 3.8 contains some of the same reactions as those just discussed for Section 3.6, but in a more useful comparative arrangement, allowing additional observations to be made. In Figure 3-9, the reactions of double expressed ZFRs with each binding domain type (i.e. Z2, Z3, Z4, and Z5) on both the 2MutHomDim and 2MutHetDim substrates, are shown side-by-side. It can be seen in Figure 3-9 that the Z4 ZFR has the most pronounced signal attenuation effect on the 2MutHomDim substrate, but the effect is also quite pronounced on the 2MutHetDim substrates. It can also be observed that the Z2 and Z3 ZFRs do, in fact, produce some signal attenuation effect when compared with the Z5 ZFR (the weakest binding activity ZFR, see Section 3.5 and Figure 3-4) on the 2MutHomDim substrates, but the effect does not appear to occur with the 2MutHetDim substrates. Another experiment, also in Section 3.8, produced the signal attenuation effect, this time using a Z1 (Zif268) ZFR on a 2MutHomDim substrate for which it was 50% cognate (Figure 3-8). These experiments from Section 3.8 used NM Tn3 catalytic domains.

Finally, in Section 3.9, an experiment was carried out in which different ZFRs were expressed from two DEVs producing approximately two-fold cellular concentration of each ZFR, and a four-fold cellular concentration of total ZFRs in the reaction. This experiment utilized 1MutHetDim substrates that had cognate Z-sites for one of the two ZFRs included in the reaction. This experiment was informative because it showed that the product signals were attenuated in reactions where both a ZFR cognate and a ZFR non-cognate for the substrate were co-expressed, compared with those reactions where only the cognate ZFR was expressed (Figure 3-11). This indicates that the signal attenuation effect is increased by non-DNA-bound enzyme. The experiment in Section 3.9 utilized ZFRs with Tn3[NY G70C] catalytic domains.

3.10.5.3 Key observations

Several general observations can be made from the data above: **1)** There is a relationship between the appearance of the signal attenuation effect and the binding activity of the enzyme, as the ZFR with what is presumed to be the high binding activity binding domain (Z4) is associated with this effect, even at single expression vector enzyme production levels (Figure 3-4 and Figure 3-5). Likewise, the ZFR with what is presumed to be the lowest binding activity binding domain (Z5) is not associated with this effect, even at two-fold enzyme production levels. **2)** There is a relationship between the cellular enzyme concentration of a ZFR with cognate sites on a substrate and the signal attenuation effect. Increase in cellular enzyme concentration, whether as the result of the use of different expression vectors, different enzyme production levels, the use of multiple expression vectors at once, or the use of DEVs, increases this attenuation effect. **3)** There is a relationship between the concentration of non-bound ZFRs and the signal attenuation effect, in reactions where bound ZFRs are also present. That is, in reactions in which one ZFR has cognate Z-sites on a substrate, co-expression of a ZFR for which there are no cognate Z-sites, appears to increase the signal attenuation effect. **4)** There is a relationship between the type of substrate used with respect to the Z-sites and the signal attenuation effect. Substrates that bind homodimers appear to be most permissive to this effect, with no detectable difference being observed between reactions involving 1MutHomDim substrates with two occupied Z-sites, and 2MutHomDim substrates with one occupied Z-site (Figure 3-5 C). In contrast, the effect is not as pronounced on 2MutHetDim substrates under the same conditions (Figure 3-5 C and Figure 3-9). **5)** The signal attenuation effect is produced by ZFRs utilizing either the Tn3[NM] or the Tn3[NY

G70C] catalytic domains, indicating that the effect is not a unique product of the Tn3[NY G70C] catalytic domain.

3.10.5.4 Potential explanation of the SA effect

An explanation for the signal attenuation effect that fits all of the above observations is that ZFRs are, at some level, capable of cleaving DNA outside of proper recombination reactions. This possibility might not be entirely unexpected as the mutations which activate the Tn3 catalytic domain may do so by mimicking some of the activating effects that the accessory subunits in the wild-type system have on the enzyme (see Section 1.6.2 and 1.6.4). This is to say, that while wild-type Tn3 resolvase must be activated by complex interactions involving tetramer formation and accessory subunit interactions, individual hyperactive Tn3 catalytic domains may be somewhat active by default. Additionally, this type of activity has been observed before using some combinations of activating mutations of Tn3 resolvase (unpublished work within the Stark lab). In this scenario, there would be a predicted relationship between having either a more active binding domain, or higher enzyme concentration, both of which would increase occupancy at the binding sites. If the ZFRs are capable of cleaving DNA outside of recombination reactions, then presumably this takes place at the Z-site where they are frequently bound.

The explanation that hyperactive Tn3 catalytic domains may be capable of inappropriate DNA cleavage might also be consistent with the observation that non-DNA-bound ZFRs increase the signal attenuation effect, as interaction with non-DNA-bound ZFRs might help activate the DNA-bound ZFRs through protein-protein interactions. Additionally, this explanation may also be consistent with the observation that the signal attenuation effect is more pronounced on homodimer-binding substrates (1MutHomDim and 2MutHomDim) than heterodimer substrates (2MutHetDim-DR and 2MutHetDim-IR). Dimers on the 2MutHomDim substrate would be capable of cleaving both strands, whereas a single ZFR subunit bound to only half Z-sites within the 2MutHetDim substrates would either only cleave one strand, or would have to co-opt the binding of another subunit to cleave both strands (see below).

It important to remember that cleavage, as it relates to ZFR reactions, implies not only that the ZFR has cut the DNA, but that the ZFR, itself, remains covalently attached to the DNA unless the DNA is re-ligated (see Section 1.6.2). Having a single ZFR subunit covalently attached to the DNA would likely interfere with plasmid replication, and this might explain why the signal attenuation effect is also possible in single ZFR reactions on 2MutHetDim-DR and 2MutHetDim-IR substrates. This activity may be analogous to the way that DNA topoisomerases I (topo I) remains covalently attached to DNA when the camptothecin class of chemotherapeutic drugs inhibit their re-ligation step, resulting in lethal collisions between topo-I DNA complexes and advancing replication forks (Lisby et al., 1998, Liu et al., 2000).

Results in the next chapter provide further support that the cause of the signal attenuation effect is caused by inappropriate cleavage of the DNA by ZFRs. The signal attenuation effect observations made in this chapter will be important when considering results in the following chapter.

Chapter 4: *Catalytic Domain Modifications*

4.1 Conspectus

4.1.1 Three types of predictions from structure-based analysis and mutagenesis studies

The general approach taken in this chapter to accomplish the project aims of achieving a dimer-dimer orientation specificity bias and a recombination reaction directionality bias (i.e. a non-reversible ZFR reaction), was to analyse crystallographic structures of γδ resolvase, and hypothesise explanations for the locations of Tn3 resolvase activating mutations, to derive mechanistic theories about resolvase catalysis that might be exploited to force new parameters within the reaction. Three models of Tn3 resolvase regulation were proposed, and were used as the bases of experiments that attempted to generate a dimer-dimer orientation specificity or recombination reaction directionality bias.

4.1.2 Interactions between counterpart residues at position 102 during tetramer formation

The first hypothesis postulated that the reason activating mutations at amino acid position Asp102 are so critical for Tn3 resolvase activation (see Chapter 1, Section 1.6.4), is because this residue is key to a negative regulation system within the protein. Aspartic acid is negatively charged, and modelling of γδ resolvase (a close Tn3 resolvase homologue; see Section 1.6.3) dimer-dimer synapsis suggested that counterpart 102 residues (Glu102 in γδ resolvase) might come within close proximity of one another at the point of initial contact (Li et al., 2005). It was hypothesized that this close proximity of negative charges might act as a barrier to synapsis, which could either be overcome by the energetic interactions of a properly formed resolvase regulatory synaptosome, or alleviated by amino acid substitutions at this position. Evidence from mutagenesis studies indicate that several uncharged amino acids will activate the protein when substituted at this position (Burke et al. (2004) (Burke et al., 2004), see Figure 1-14; and Stark, unpublished work), which is a pattern consistent with the alleviation of negative regulation. Therefore, an attempt was made to force new parameters in the reaction by

generating pairs of complementary subunits with oppositely charged residues at position 102. However, the experiments (described in Section 4.2) produced negative results and were unable to provide support for the theory that Asp102 acts in a negative regulatory interaction with its counterpart residues in opposing subunits at the initiation of synapsis.

4.1.3 ZFR activation via pairs of differentially disrupted 'locking interface' mutants

The second hypothesis tested was based on a model of Tn3 resolvase activation involving a sub-region of the catalytic tetramer dimer-dimer interface that is stabilized by interactions between the E-helix and α/β sub-domain of adjacent subunits. A model of Tn3 resolvase activation was proposed where this region of the dimer interface acts to lock dimers in an inactive conformation, until contact is made by the subunits of the regulatory module (see Chapter1, Section 1.6.2), which releases this 'lock' through allosteric regulation. The release of this dimer interface lock was proposed to allow tetramerization of the site I dimers that form the catalytic module (see Section 1.6.2). A key feature of this model of Tn3 resolvase activation, is that the two dimers of a catalytic tetramer are unlocked asymmetrically (i.e. on one side of the tetramer only), at or before the point of tetramer formation. The reasons the tetramer is proposed to be unlocked asymmetrically, is because the regulatory module of the synaptosome is located on one side of the tetramer only (see Section 4.3).

A hypothesis was generated that the asymmetric conformational change might facilitate a dimer-dimer orientation specificity bias during tetramer formation. The locations of many known secondary activating mutations suggest that they may produce their effect by disrupting the dimer locking interface. An attempt was made to test this hypothesis by generating differentially activated ZFR subunits, which when used in combination, mimic the postulated allosteric unlocking effect on the dimer interface to force dimer-dimer orientation specificity within the reaction.

The results of these experiments, described in Section 4.3, do not support the hypothesis that asymmetric dimer unlock can be used to produce a dimer-dimer orientation specificity bias. However, the experiments did result in two important discoveries. The first discovery was that catalytically active ZFRs are able to activate

catalytically inactive ZFRs when used in complementation reactions. This discovery may be valuable for increasing ZFR site-specificity by reducing the potential for off-target reactions and cytotoxicity observed when ZFRs are used in mammalian cell systems (Gaj et al., 2014). The second important discovery is that, unexpectedly, the pairs of differentially disrupted locking interface ZFR mutants did produce activity consistent with a recombination reaction directionality bias. Additionally, the recombination reaction directionality bias produced in the fashion described may be used to generate integration reactions with an orientation specificity bias.

4.1.4 Differentially mutated ZFR pairs based on residue 102 and E-helix 'landing pad' mutations

A model was proposed to explain the function of several activating hydrophobic amino acid substitutions at the N-terminal end of the E-helix, termed the 'landing pad'. In this model, these activating mutations promote the formation of the dimer-dimer complex by creating favourable hydrophobic interactions between the N-terminal ends of counterpart E-helices that interact when two dimers are brought together. This model was based on analysis of available γδ resolvase structures, the tetramer initiation trajectory model proposed by Li et al. (2005), and the locations of several Tn3 resolvase activating mutations. A hypothesis was generated from this model that differential inclusion of hydrophobic E-helix landing pad mutations in a ZFR pair might be used to produce complementation reactions that produce a dimer-dimer orientation specificity or recombination reaction directionality bias. Although only a limited set of experiments could be performed due to time constraints, the results displayed activity consistent with the potential for a significant recombination directionality bias (see Section 4.4.3). The E-helix landing pad model is functionally related to the dimer interface unlocking model, and therefore the recombination reaction directionality bias observed in Section 4.4 may be related to that observed in Section 4.3. Unfortunately, there was insufficient time available to construct experiments that could discriminate whether the mechanism behind the effect was actually related to the proposed E-helix landing pad model, or was a result of several other potential mechanistic explanations.

4.2 Ionic repulsion at initial dimer-dimer contact point

4.2.1 Hypothesis overview

The first model tested involves the hypothesis that the Asp102 residue is involved in negative regulation of Tn3 resolvase. This hypothesis is based on the possibility that the initial dimer-dimer contact point, which initiates active-site tetramer formation, occurs between counterpart Asp102 residues on opposing dimer subunits across the interface (see Figure 4-1, based on the 1GDT crystal structure of γδ resolvase). This interaction between Asp102 residues is what may take place if one were to assume that the pre-synaptic dimers are in approximately the conformation seen within the 1GDT crystal structure (Yang and Steitz, 1995), and are brought directly together during synapsis, similar to what was proposed by Li *et al.* (2005) (Li et al., 2005).

Apart from structural data, an additional piece of information that could support the hypothesis is the importance of mutations at residue 102 for creating hyperactive Tn3 resolvase mutants. Although many secondary activating mutations have been discovered in *conjunction* with primary mutations at position 102, amino acid substitution of Asp102 appears to be virtually prerequisite for hyperactivation (Burke et al., 2004). Furthermore, it appears that several different amino acid substitutions at this position are capable of contributing to activation—albeit in varying degrees ((Burke et al., 2004, Olorunniji et al., 2008); and Stark, unpublished work)—perhaps suggesting that it is the removal of aspartic acid at this position which is the key to activation. This observation could imply that Asp102 is participating in a negative regulatory barrier, which must first be relieved before further activation can take place. Because aspartic acid is a negatively charged amino acid, which is relatively short and has limited flexibility, ionic repulsion at the point of initial contact between dimers could be a good fit to the hypothesis that Asp102 is involved in negative regulation. Additionally, in the γδ resolvase system, which is virtually identical to the Tn3 resolvase system, a glutamic acid appears at position 102, further supporting the idea of negative regulation by negative charge at this position.

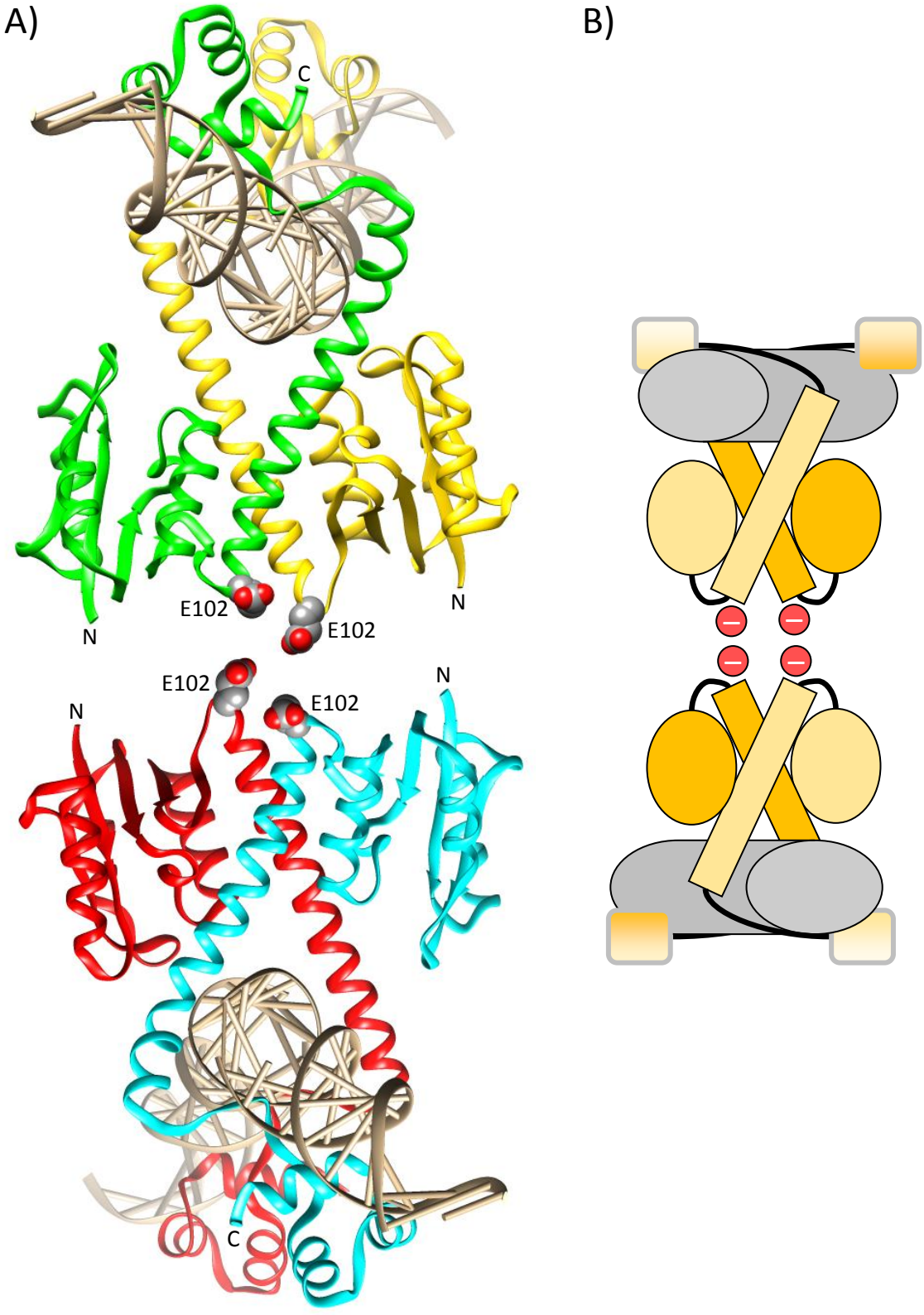


Figure 4-1: Glu102 hypothetical counterpart interactions at dimer-dimer synapsis. A) The panel shows the two 1GDT dimer structures in opposite orientation with one another as they may be during the initial dimer-dimer contact at synapsis. The subunits have been alternatively coloured green, yellow, cyan and red. Glu102 residues are shown in space-filling mode and appear at the N-terminal ends of the E-helices. **B)** The panel shows a simplified conceptual diagram of the structure in A). The representations are as follows: DNA, grey tubes; α/β sub-domains, ovals; E-helices, long rectangles; HTH domains, boxes with shade transitions; and loop regions, thick black lines. The subunits of the dimers are depicted in two shades of orange to aid with clarity of the image. At the N-terminal tips of the E-helices are red circles indicating the charge of the Glu102 residues at this location.

One approach to testing this hypothesis is to use ZFR complementation experiments where two ZFR mutants are alternately substituted at position 102 with positively and negatively charged residues (Figure 4-2). If the side chains of residues at position 102 do form the initial point of contact during tetramer formation, then one would expect that when paired in opposite polarity configurations there would be a dramatic increase in activity over the pairings in same polarity configurations. This differential activity would be clearly visible when using the 2MutHomDim substrates that test the ability of alternate homodimers to interact (and could be visible on the heterodimer-binding substrates as well, producing results consistent with dimer-dimer orientation specificity). Of course, it is important to consider that the side chains of the other various charged residues might not be adequately placed to produce an electrostatic interaction effect, as proposed for the aspartic acid, but by attempting all possible charged-residue combinations, a maximum effort can be made.

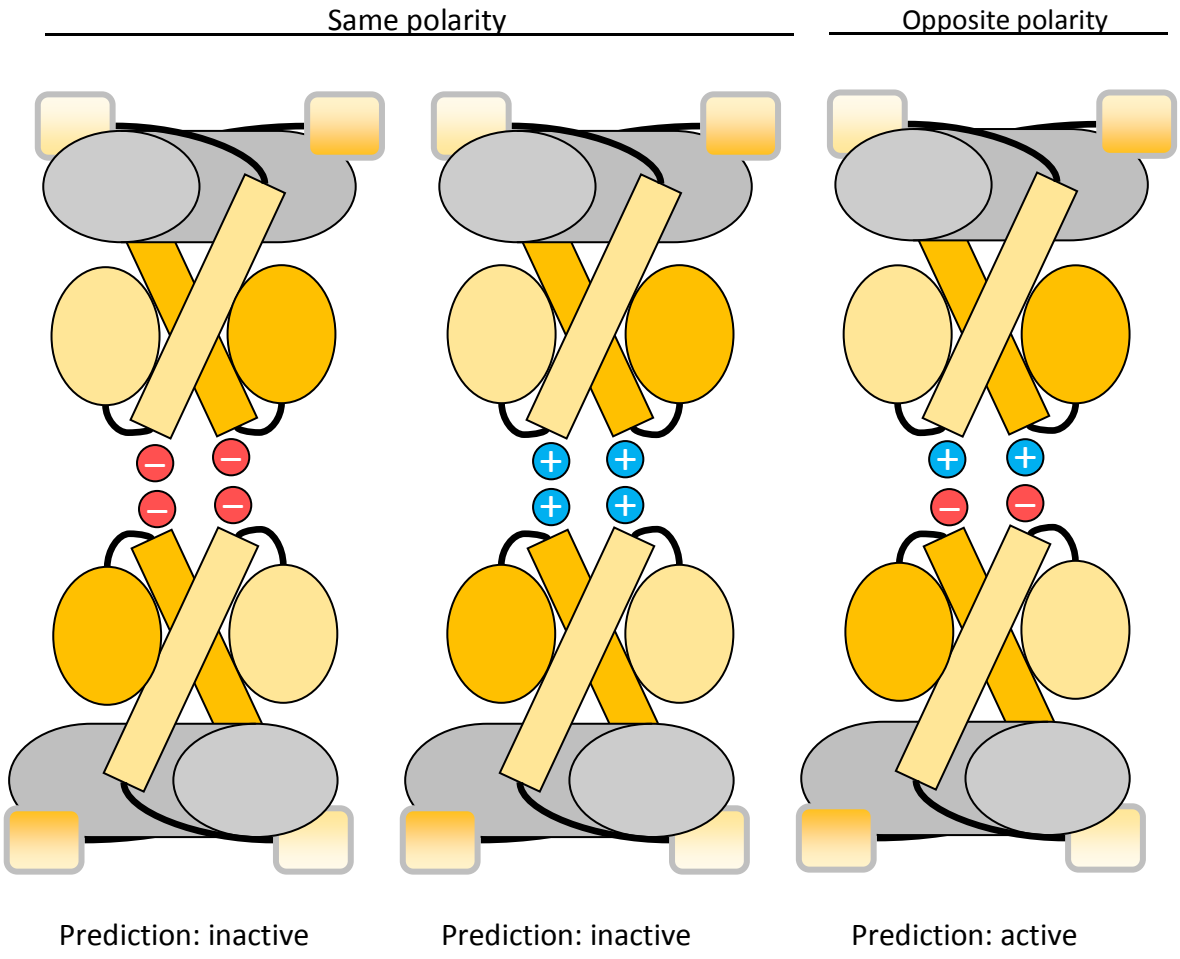


Figure 4-2: Charge complementarily arrangements and predictions. The figure shows the three arrangements of charge complementarity at residue 102 to be assayed in the experiment. All the image representations are the same as in **Figure 4-1 B**, with the exception that the colour transition boxes now represent ZFA binding domains.

4.2.2 Experimental design

A complementation experiment was designed in which a set of ZFR mutants, which differed only by the identity of their charged residue at position 102, were tested on a 2MutHomDim substrate (Z2/Z2 x Z3/Z3). The Z2 and Z3 ZFAs and binding sites were selected for site-specific targeting because they performed with the most comparable levels of binding activity within the experiments of Chapter 3, Section 3.5. The ZFRs with the Z2 ZFA were expressed from pSA3022-based vectors, while those with the Z3 ZFA were expressed from pMS140-based vectors, as this combination appeared to produce the most similar level of activity in the experiments of Section 3.5. The mutations at position 102 were placed in the Tn3[N G70C] catalytic domain background because it was demonstrated to be highly active in other experiments (such as those in Chapter 3) when used in combination with the primary mutation D102Y. If D102Y produces its activating effect by eliminating a regulatory barrier resulting from ionic repulsion by counterpart Asp102 residues, then the ionic substitutions and complementation arrangements used in this experiment, also designed to eliminate this barrier, should also result in the Tn3[N G70C] background providing an adequate level of activation. Tn3[NY G70C] was, in fact, used for many experiments throughout this project, and so this domain was employed where possible in order to maintain a level of a comparability between experimental results.

Three different types of complementations were carried out in order to test the hypothesis that counterpart Asp102 residues produce negative regulation via ionic repulsion. The first set of complementations tested the ionic mutation variants against themselves. Because the complementation experiments involve the expression of alternately targeted ZFRs from two vectors, the concentration of ZFR protein in the *E. coli* cells is approximately doubled in these complementations. Therefore, when making observations of the base level of activity of each ionic mutation variant ZFR for comparison with the opposite polarity complementations, it was appropriate to also express alternately targeted versions of the same ionic mutation variant ZFRs from two expression vectors. The second set of complementations were the primary experimental reactions, using opposite polarity ZFR variants. A third set of complementations, which was employed as an additional control, tested ionic variant ZFRs against other ionic

variant ZFRs of the same polarity. The complementations were performed in the 17 Hour Resolution Assay (described in Chapter 3, Section 3.4).

4.2.3 Results of ionic polarity complementations on 2MutHomDim substrates fail to support the hypothesis

The results of the complementation experiment (shown in Figure 4-3) fail to support the hypothesis that there is any interaction between 102 residue side chains on opposing subunits at tetramer formation. None of the complementations produced any significant level of recombination in the 17 Hour Recombination Assay. There was no difference in the recombination activity between reactions where ionic charge at position 102 would have been expected to product repulsion (lanes 2–6 and 15–18) and reactions where opposite charges were paired (lanes 8–13). Therefore, no ionic interaction between opposing dimers at residue 102 appears to exist. If Asp102 were operating in a negative regulator mechanism as hypothesized, then pairing oppositely charged residues at position 102 should have been expected to at least remove that negative regulation, if not produce an additional activating effect due to favourable ionic interactions. None of the substitutions used at position 102 were known to be substantially activating primary mutations when used in single mutant (i.e. not complementation) experiments, and therefore, only removal of the hypothesized ionic repulsion would have been expected to produce recombination using these mutants in complementation experiments. The lack of any significant level of activity among the opposite polarity complementations (lanes 8-13), means that there is no effect taking place that is suitable for use in either generating a dimer-dimer orientation specificity bias, or generating a recombination reaction directionality bias.

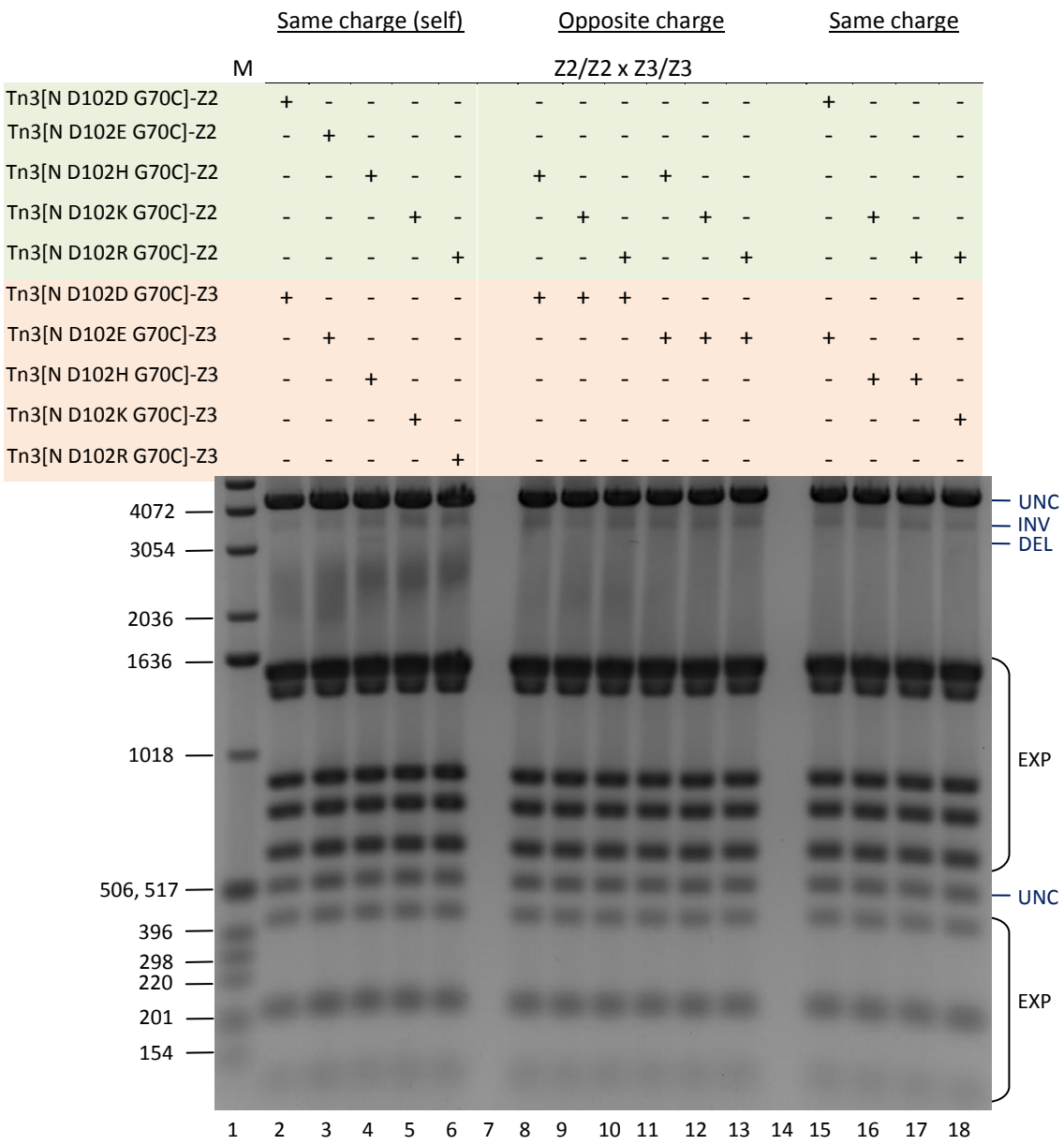


Figure 4-3: Ionic pairs complementation test. The results of a 17 Hour Recombination Assay are shown in the figure. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion product, and EXP = expression vector plasmid). Above the image is a table which ZFR pair and substrate is used in each lane. The tracks are grouped into same ionic charge, same residue identity at amino acid position complementations (lanes 2–6); opposite ionic charge at amino acid position 102 complementations (lanes 8–13); and same ionic charge, different residue identity at amino acid position 102 (lanes 15–18). Coloured shading is used within the table for clarity, which separates the ZFRs by ZFA type.

4.3 The dimer interface-unlocking activation model

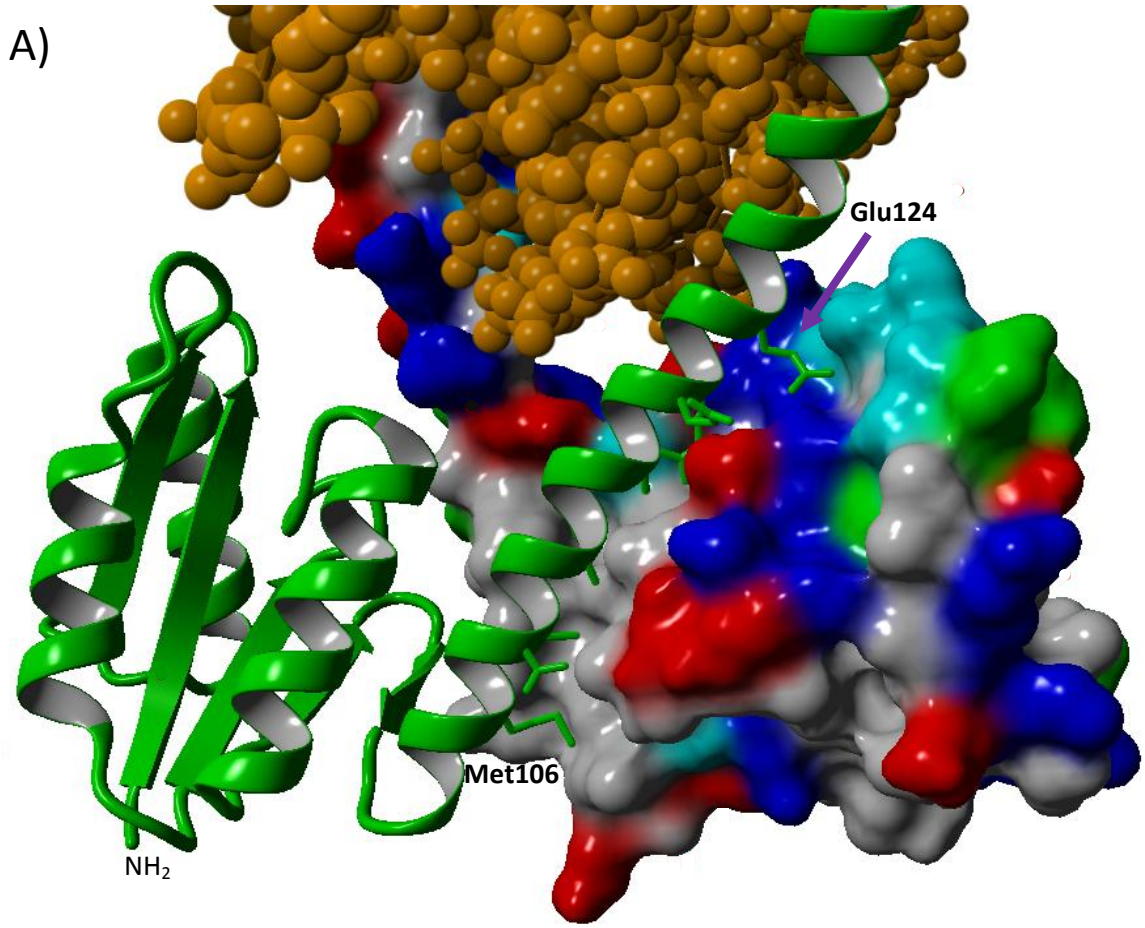
4.3.1 Conceptual overview

4.3.1.1 Dimer interface locking sub-region

The second hypothesis involves a sub-region of the dimer interface as it appears in the 1GDT crystal structure, which can be interpreted as displaying a 'locking' mechanism that must be disengaged in order for an active tetramer to form. This lock consists of the dimer interface contact between the E-helix of one subunit and the α/β sub-domain of the adjacent subunit (see Chapter 1, Section 1.6.2 for a description of the Tn3 resolvase catalytic domain). The locking region contains several potential hydrogen bonds, as well as hydrophobic packing, which must be disrupted before the inactive dimer conformation witnessed in 1GDT (Figure 1-12 A) can become the active tetramer conformation seen in 1ZR4 (Figure 1-12 B). The 1GDT dimer, which can be split into two halves about a two-fold axis, does not have perfect symmetry across both of those halves (Figure 4-4 A and B). One interpretation of that asymmetry could be that one side of the dimer has been captured in a partially unlocked conformation. One noteworthy example of the difference between dimer interface contacts between the right and left half the dimer is the absence of the trans-interaction seen between Gly70 and Arg121 in the unlocked half of the dimer (Figure 4-5).

4.3.1.2 Asymmetric dimer-unlocking

During wild-type Tn3 resolvase catalysis, it is believed that the dimers bound at site I (the catalytic site) asymmetrically interface (that is, the interaction takes place on one side of the dimers only) with the subunits of the regulatory module via the 2-3′ interface (see Chapter 1, Section 1.6.2). Interaction at the 2-3′ interface between site I-bound dimers and the subunits of the regulatory module is thought to facilitate the formation of the catalytic tetramer module of the synaptosome (see (Rowland et al., 2009) and (Rice et al., 2010) for a full description of a small serine recombinase synaptosome). The dimer interface-unlocking activation model suggests that for each dimer bound at site I, one of the two locking interfaces would be unlocked by the allosteric action of 2-3′ contacts to the subunits of the regulatory module (Figure 4-6). Because only one locking interface per dimer would be unlocked, this will be referred to as asymmetric dimer unlocking.


A)
Glu124
Met106
NH2

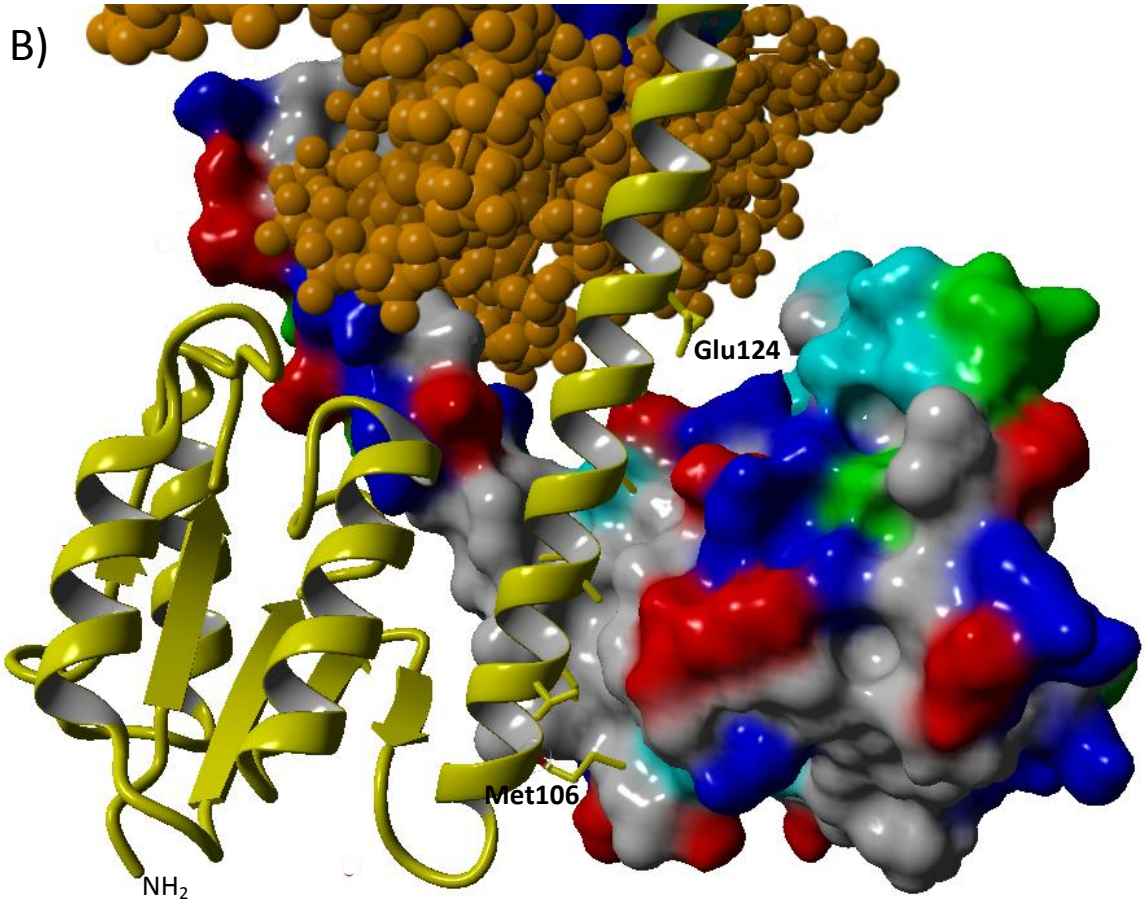
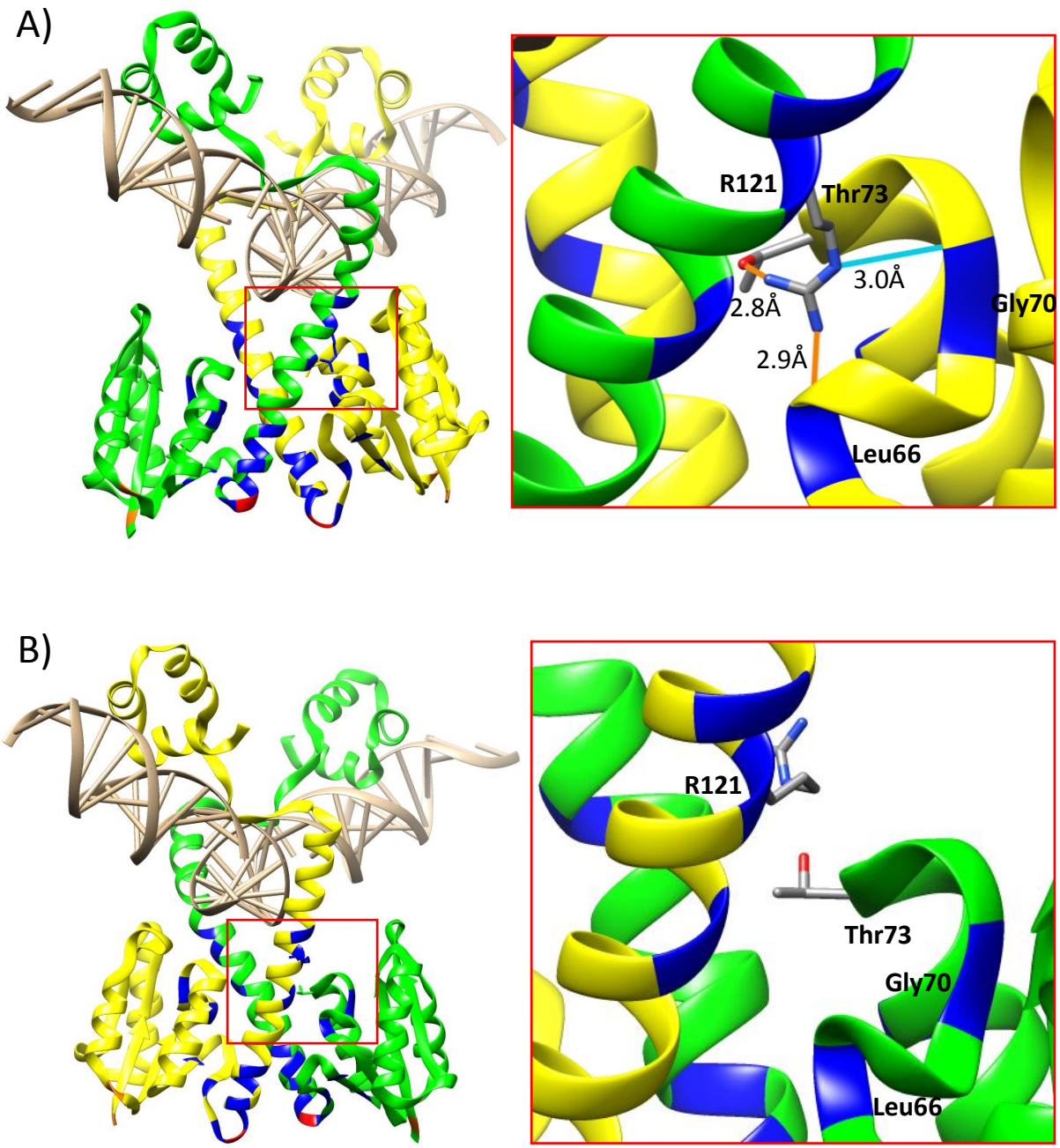

B)
Glu124
Met106
NH2

Figure 4-4: 1GDT dimer showing locked and semi-unlocked asymmetry. Dimer subunits are shown in either green or yellow, with the other subunit shown with a molecular surface in each image. The molecular surface been coloured by residue type (positively charged = blue, negatively charged = red, polar = turquoise, hydrophobic = grey)**.** The DNA at the top of the images is shown in space filling mode coloured brown-orange. **A)** This panel **s**hows the locked conformation. The E-helix of the green subunit is tightly bound to both the charged and hydrophobic portions of the subunit shown with a molecular surface. The residues shown in stick form in the figure are: Met106, Thr109, Ile110, Ala113, Val114, Ala117, Gln120, Arg121, and Glu124. **B)** This panel shows the pseudo-unlocked conformation. The E-helix of the yellow subunit is bound only to the hydrophobic region of the subunit shown with a molecular surface. The residues shown in stick form in the figure are: Met106, Thr109, Ile110, Ala113, Val114, Ala117, and Glu124. The residues shown in A) which no longer make dimer interface contacts have been omitted for clarity of the image, with the exception of the last residue, Glu124, which is displayed as a reference point for comparison between the figures. In A) and B), for clarity of the image, only the first and last residue of the set (Met106 and Glu124, respectively) have been labelled.



Primary (red)	Secondary (blue)	Tertiary (orange)
D102Y/I/F/V/T/W	L66I/F, G70A/C, I77T, A89T, F92S, T99S, G101S/C, M103I/V, Q105L, V107M, T109I, A117V, R121K/M/S, E124Q	R2A, E56K

Figure 4-5: Arg121 trans-interaction with Gly70 in the locked and unlocked states. The figures show the 1GDT crystal structure with the resolvase subunits coloured in yellow and green. The same colouring scheme as used in **Figure 1-14** is used to indicate the locations where activating mutations occur. On the left side of the figure a red square indicates the region of the Arg121 of one subunit and Gly70 of the other subunit. On the right side of the figure, the indicated area from the left side is shown magnified, and from a slightly altered angle which best displays the residues. **A)** The figure shows Arg121 trans-interaction with Gly70 in the locked state side of the 1GDT dimer. Arg121 makes a hydrogen bond to the backbone of Gly70 which is indicated with a blue line. Additionally, two hydrogen bonds from Arg121 to Thr73 and to the backbone Leu66, respectively, are shown in orange, which are detected when the hydrogen bond angle constraints are relaxed. There is also a potential hydrogen bond contact (3.1 Å) between Arg121 and Met76 which is not displayed in the figure **B)** The figure shows the same area as in A) but on the unlocked side of the 1GDT dimer. No hydrogen bonds are possible between the residues highlighted in A); although, Arg121 now makes hydrogen bond contact to the DNA backbone (not indicated in the figure).

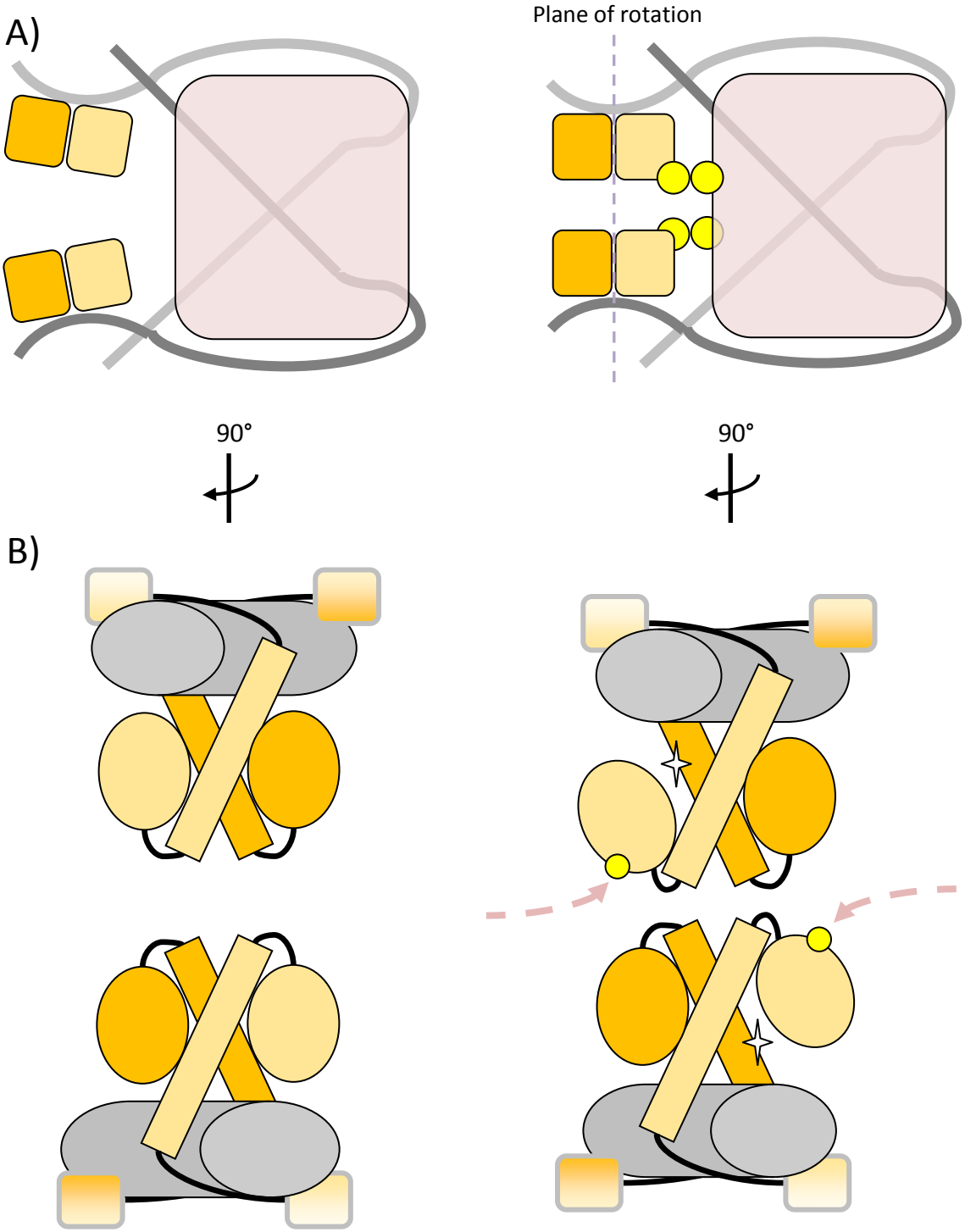


Figure 4-6: Asymmetric 2-3′ interface dimer unlocking model. A) The figure shows Th3 resolvase dimers (represented by orange squares) bound to site I (DNA is represented by thick grey lines) before and after interfacing with the regulator module of the synaptosome (large pink square). The 2-3′ interface contact interactions between the regulatory module and the catalytic site dimers are is indicated using yellow circles. The dimers, in both cases, are shown at a point before catalytic tetramer formation. **B)** The figure shows the same scene as in A), but with more detail of the subunits, and rotated 90° so that the regulatory module is beyond the foreground and out of view. The DNA is represented as a grey tube. The resolvase subunits are coloured in two shades of orange. The resolvase HTH DNA-binding domain is represented using squares with a shade transition in the colouring. The resolvase E-helix is shown using long rectangles, and the α/β sub-domain is depicted using ovals. Flexible amino acid regions connecting the domains (such as the arm region) are depicted using thick curved lines. The left side of the figure shows the two Tn3 resolvase dimers in locked conformation before interfacing with the regulatory module. The right side of the figure shows the dimers being asymmetrically unlocked via the 2-3′ interface by the regulatory module. Stars are used to indicate where locking interface contacts between the E-helices and α/β sub-domain have been disengaged, and pink dashed arrows indicate contact from the regulatory module.

4.3.1.3 E-helix and α/β sub-domain relationship

In order for the dimer shown in 1GDT (Figure 1-12 A) to enter into the active tetramer conformation seen in 1ZR4 (Figure 1-12 B), the E-helices of the resolvase subunits must form anti-parallel trans-interactions with one another across the tetramer interface (Figure 1-13). However, the interactions of the 1GDT dimer, between the E-helices and the α/β sub-domains, block the space that the anti-parallel E-helices interactions of the 1ZR4 tetramer must occupy. This blocking of space occurs through both a trans-interaction where the α/β sub-domain of one subunit occludes space on the E-helix of the adjacent subunit (Figure 4-7), and as a result of the orientation of the α/β sub-domain, occupying space relative to its own subunit, which is incompatible with tetramer formation (Figure 4-8). Therefore, in order for the resolvase dimers to form a tetramer, the α/β sub-domains must rotate away from their trans-interaction with the E-helices so that new anti-parallel trans-interactions can be formed between the E-helices of opposing subunits within the tetramer (Figure 4-8). Resolvase tetramer formation can be thought of in terms of two E-helix pairs that must come together, and four α/β sub-domains that must move out of the way for this to happen (one α/β sub-domain on each E-helix).

At the point right before tetramer formation between two 1GDT-like resolvase dimers, in both dimers, the E-helix of one subunit would be locked by the α/β sub-domain of the second subunit, and the E-helix of the second subunit would be left unlocked by the α/β sub-domain of the first subunit. This difference in subunit conformations gives rise to an asymmetry that differentiates one of the possible dimer-dimer orientations from the other. According to the dimer interface-unlocking model, one of these two orientations should occur in the wild-type system, while the other does not (Figure 4-9). In the wild-type dimer-dimer orientation, both E-helices of one partnering pair of anti-parallel E-helices are locked, while both E-helices in the other partnering pair of anti-parallel E-helices are unlocked. In the non-wild-type dimer-dimer orientation, both partnering pairs of anti-parallel E-helices consist of one locked and one unlocked E-helix. Two 1GDT resolvase dimers can be modelled together into a synaptic initiation complex, and this modelling produces unique geometric docking arrangements in the wild-type and non-wild-type dimer-dimer orientations (Figure 4-10 A and B, respectively). However, while these docking arrangements are interesting, conclusions cannot be drawn from them.

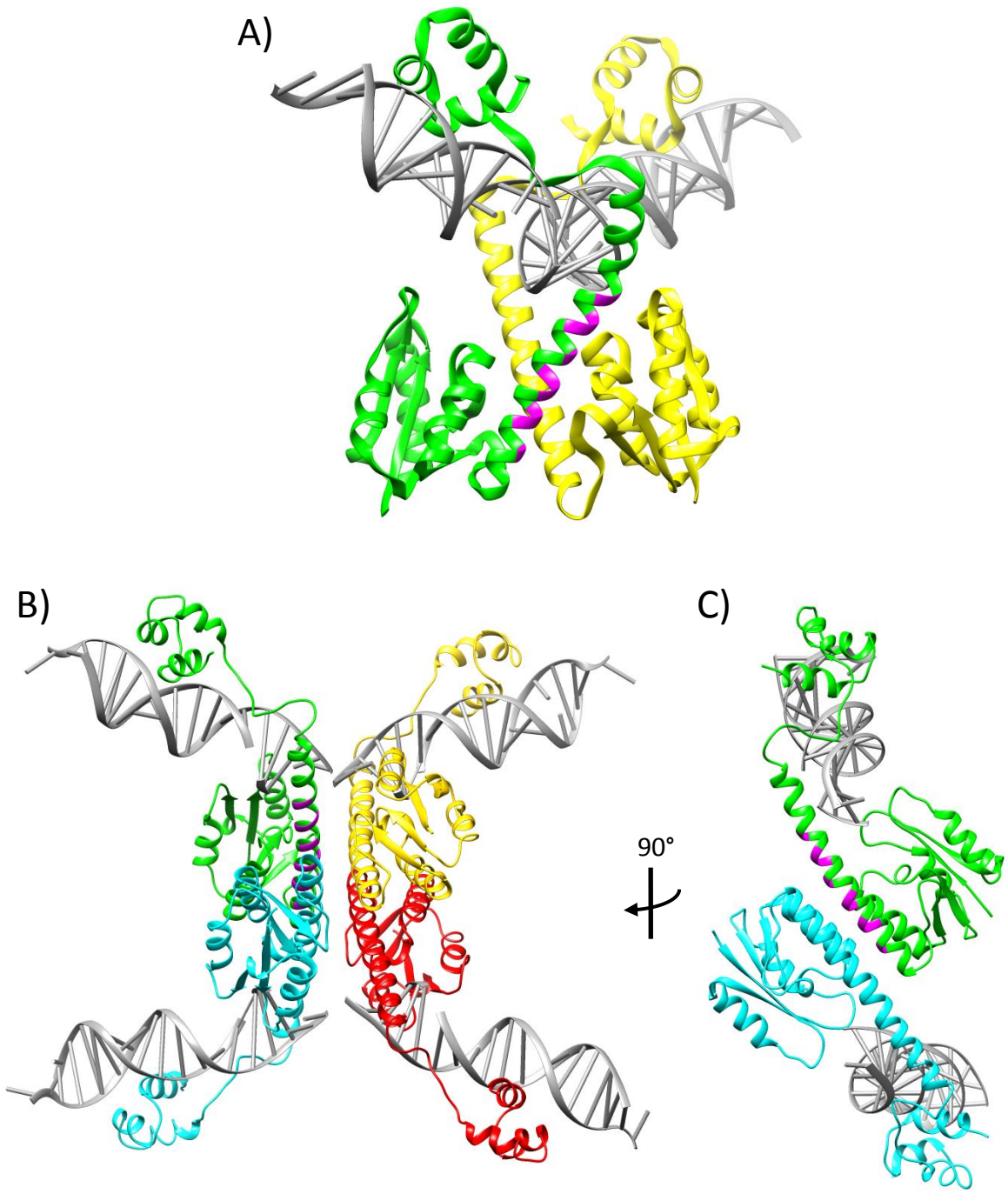


Figure 4-7: Trans-interaction occlusion of the E-helix by the α/β sub-domain. A) The figure shows the 1GDT resolvase dimer with amino acid positions in the E-helix of the green subunit which are occluded by the α/β sub-domain of the yellow subunit, coloured magenta (Met106, Thr109, Ile110, Ala113, Val114, Ala117, Gln120, Arg121, and Glu124; the same positions shown in **Figure 4-4 A**). **B)** The figure shows the 1ZR4 structure with the same amino acid positions as indicated in A) again coloured magenta within the green subunit, in order to demonstrate how this contact space must be unoccupied in order for the anti-parallel E-helix-E-helix interactions to take place. **C)** The figure shows the model as in B), but rotated 90° for a better view of the anti-parallel E-helix-E-helix interaction between the green and cyan resolvase subunits. The half of the structure containing the yellow and red subunits and the DNA they bind has been removed for clarity.

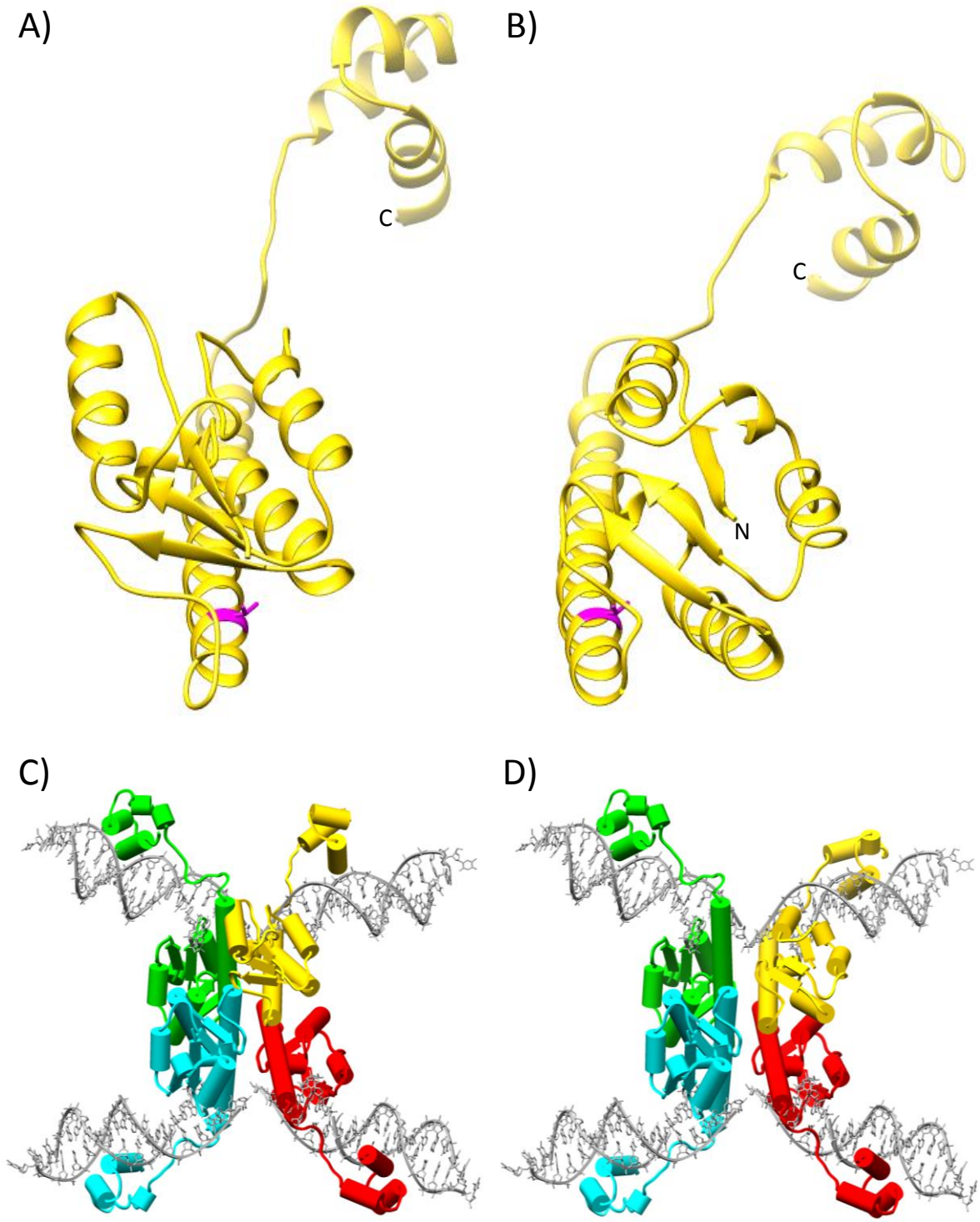


Figure 4-8: Translation of the α/β sub-domain about the E-helix.
Figure **A)** and **B)** show one resolvase subunit from 1GDT and 1ZR4, respectively, in identical orientation with respect to the E-helix. Residue Val108 is shown and coloured in magenta in order to provide a visual cue that the orientations of the subunits are matched. Comparison of A) and B) displays the translation of the α/β sub-domain about the E-helix which takes place during synapsis as inactive dimers form an active tetramer. In figure B) it may be noticed that the E-helix has been slightly extended by the ordering of some of the flexible loop. However, this may or may not take place in the wild-type system as the extension of the E-helix seen in 1ZR4 is stabilized by two hydrogen bonds, one of which is provided by the side chain of a G101S activating mutation (not shown). Figure **C)** shows the 1ZR4 tetramer structure with the yellow subunit replaced with the 1GDT subunit from figure A), in order to display the geometric incompatibility of the position of its α/β sub-domain with the tetramer structure. The 1ZR4 subunit was replaced by first superpositioning E-helices of the two subunits and then removing the 1ZR4 subunit. Figure **D)** shows the normal 1ZR4 structure. The subunits from A) and B) are in identical orientation to the yellow subunits in C) and D), respectively. It may be noted that position of the C-terminal HTH domain is also somewhat different between subunits of the 1GDT and 1ZR4 structures, due a change in position of the DNA (1GDT DNA not shown).

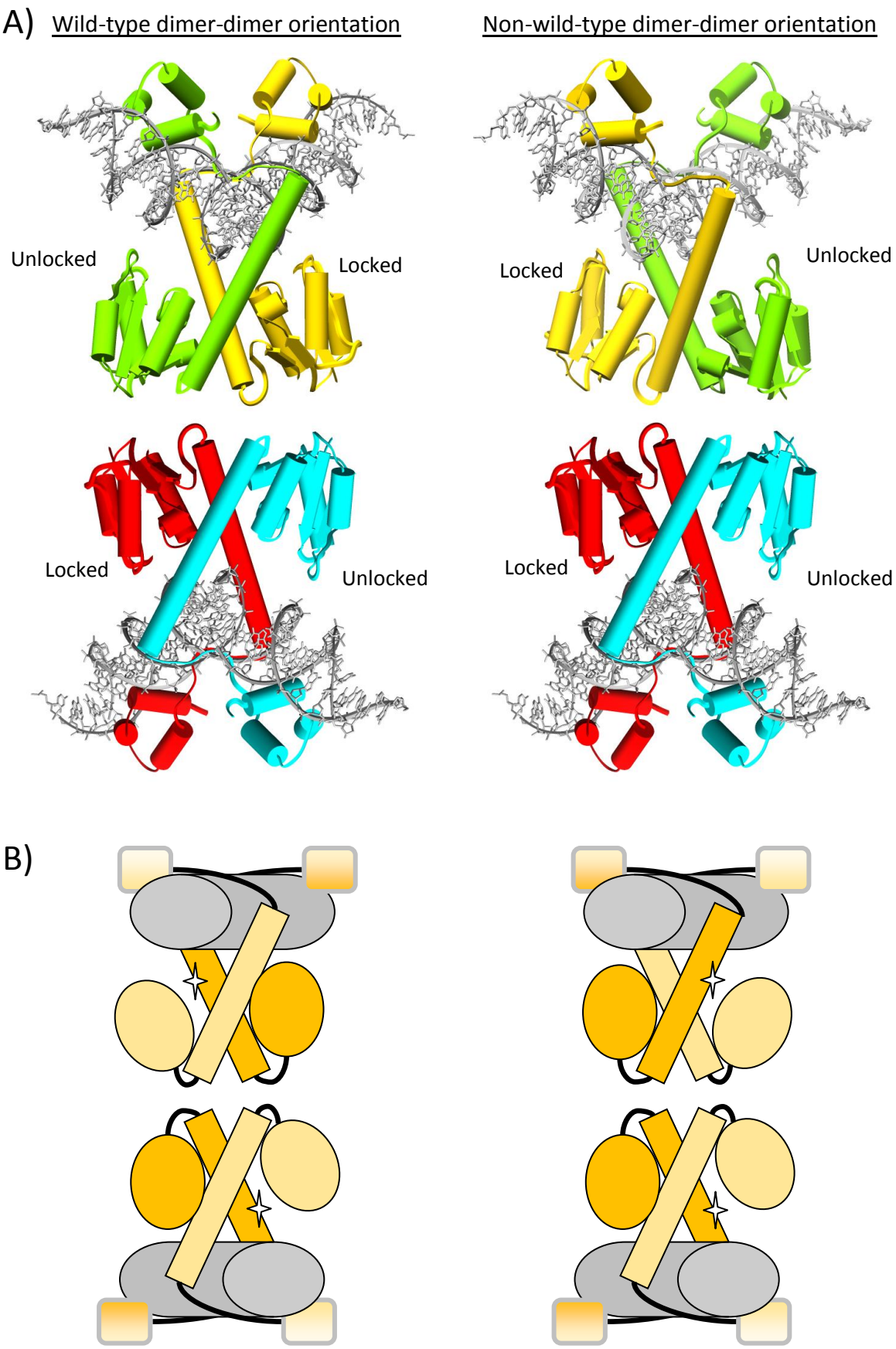
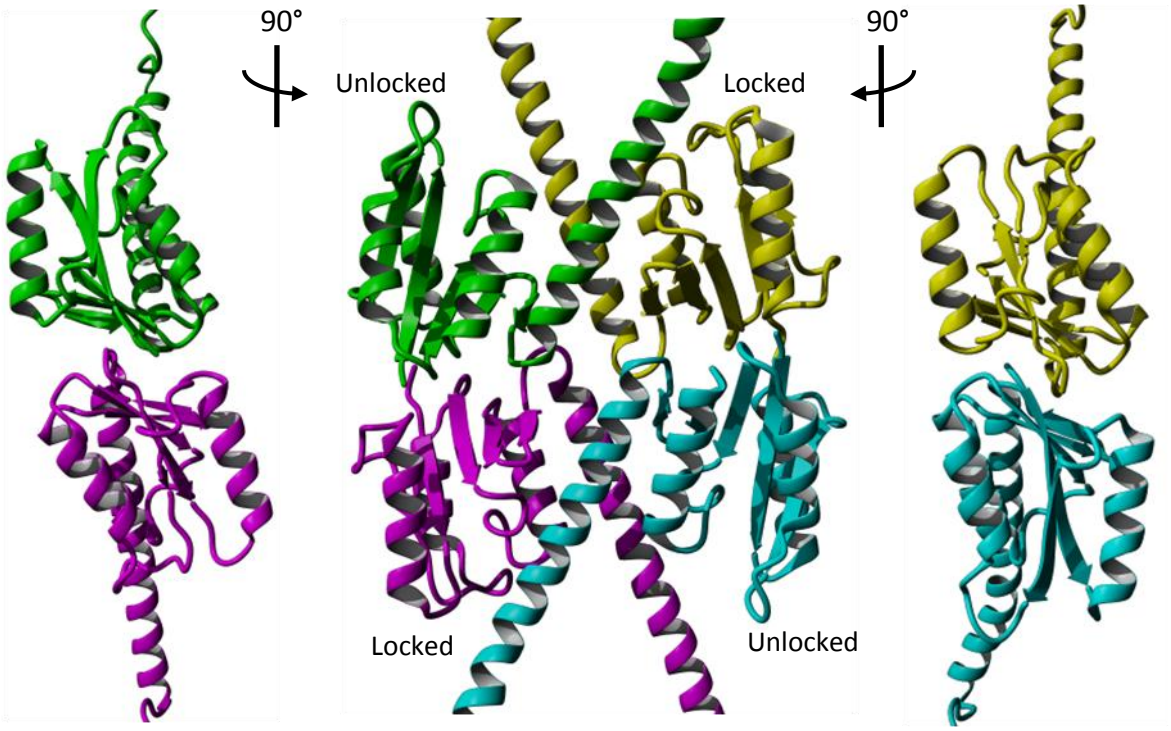


Figure 4-9: Wild-type and non-wild-type dimer-dimer orientation of asymmetrically unlocked resolvase dimers at synapsis. A) The panel shows the 1GDT dimer structure arranged in the proposed wild-type and non-wild-type dimer-dimer orientations of asymmetrically unlocked dimers. **B)** The panel shows a simplified diagram of the dimer-dimer orientations displayed in A). The structural components are represented as described in **Figure 4-6**.

A)

α/β sub-domain contact | Early-stage synaptic complex | α/β sub-domain contact



B)

α/β sub-domain contact | Early-stage synaptic complex | α/β sub-domain contact

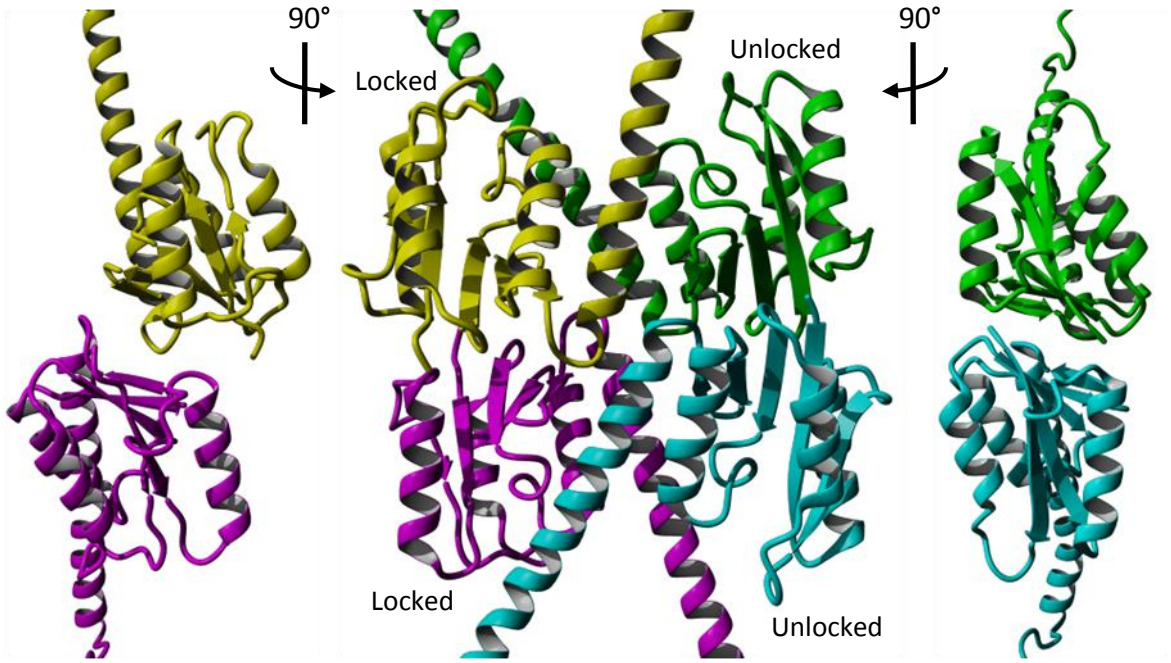


Figure 4-10: Modelling of tetramer initiation using 1GDT dimers. The tetramers are shown in the same orientation as in **Figure 4-6 B. T**he DNA and HTH domains are not shown. **A)** The figure shows two 1GDT dimers docked in the proposed wild-type dimer-dimer orientation, where the α/β sub-domains of both subunits in the foreground (green and cyan) provide an unlocked trans-interaction with the E-helices of their adjacent subunit partners within the dimers. Likewise, both subunits in the background (yellow and magenta) have α/β sub-domains which are involved in locking trans-interactions with the E-helices of their subunit partners within the dimers. **B)** The figure shows two 1GDT dimers docked in the proposed non-wild-type dimer-dimer orientation, where the α/β sub-domain of one subunit in the foreground (cyan) provides an unlocked trans-interaction with the E-helix of its adjacent subunit partner within the dimer, and the α/β sub-domain of the other subunit in the foreground (yellow) provides a locked trans-interaction with the E-helix of its adjacent subunit partner within the dimer. The subunits in the background share a similar arrangement with one (green) providing unlocked trans-interaction from its α/β sub-domain, and the other (magenta) providing locking trans-interaction from its α/β sub-domain. Docking of the 1GDT dimers was accomplished by manually moving the dimers together until steric clashes were observed between the side chains of opposing dimers (the sides chains are not displayed in the figure). Pairs of opposite-adjacent subunits are shown again at the sides of the image, rotated 90°, so that the interfacing between α/β sub-domains can be seen more clearly. One may also note that the dimer in B) have been docked in a less symmetric fashion, that those in A). This was done simply because it allowed the dimers to be moved closer together, which was not true for the dimer in A). This is justifiable because there is reason to believe the dimers in B) must synapse is a symmetric fashion.

4.3.1.4 Support from mutation studies

Some existing data that are consistent with the dimer interface-unlocking activation model can be found in the activation mutagenesis study of Burke *et al.* (2004) (Burke et al., 2004) (Burke et al., 2004). Several of the secondary activating mutations (mutations in addition to those at amino acid position 102 that are required for full activation of the enzyme; see Chapter 1, Section 1.6.5) change the amino acid identities of key residues that appear to stabilize the proposed locking interface (see Figure 1-14 for a depiction of the locations of resolvase activating mutations). Additionally, many of the remaining secondary activating mutation residues could also be proposed to destabilize that locking interface through disruptions of the N-terminal domain core. This disruption might mimic the allosteric action of 2-3′ interface contacts from the regulatory module to disengage locking contact between the E-helix and α/β sub-domain of adjacent subunits within the dimer. Of the 14 residue positions described by Burke *et al.* (2004) that produced secondary activation effects when mutated, most are located at positions that are consistent with the dimer interface-unlocking activation model.

4.3.1.5 Using the locking interface to specify dimer-dimer configurations

It was hypothesized that asymmetric dimer interface-unlocking activation might allow one to generate ZFR subunits that produce dimer-dimer interface orientation specificity during tetramer formation. This possibility followed from the observation that, in the wild-type context, the two dimers of the catalytic tetramer are unlocked by 2-3′

interface contacts on the same side, with respect to a plane that the separates subunit rotation pairs (Figure 4-6). It was hypothesised that since the asymmetric 2-3′ interface contacts to the dimers, may likewise produce an asymmetric conformational change *within* the dimers, this potential asymmetry of the dimers' conformation might make the dimer-dimer interface geometrically compatible in only one orientation during tetramer formation. This potential geometric incompatibility of one of the dimer-dimer synapsis orientations is not the mechanism by which the wild-type Tn3 resolvase achieves dimer-dimer orientation specificity, but could have arisen as an incidental consequence of that mechanism. Therefore, it was proposed that activating mutations that either directly destabilize the locked interaction, or disrupt the N-terminal domain core in a way that results in the destabilization of the locking interaction, might be used to produce a differentially mutated ZFR pair where one mutant provided the locking interaction and the other did not. This ZFR pair might then produce heterodimers that were only compatible with each other in one of the two possible dimer-dimer orientation configurations; the active configuration would mimic the activating interactions found within the wild-type system.

4.3.1.6 Dimer-dimer configurations

In order to consider the way in which pairs of differentially activated ZFR mutants may cooperate in different tetramer subunit configurations to produce orientation specificity of the dimer-dimer interface (or a non-reversible reaction), three tetramer subunit configuration states must be defined (see Figure 4-11). The first is that of the *alternate homodimers dimer-dimer* (AHD) configuration, which is comprised of two homodimers made of two different ZFR mutants. The second configuration is that of the *mirrored orientation heterodimers dimer-dimer* (MHD) configuration, which is comprised of two identical heterodimers, each containing two different ZFR mutants. The third configuration is that of the *reversed orientation heterodimers dimer-dimer* (RHD) configuration, which is also comprised of two identical heterodimers, each containing two different ZFR mutants. The difference between the MHD configuration and the RHD configuration is the orientation of the heterodimers with respect to one another. In the MHD configuration, same-type ZFR mutants are positioned directly across from one another in the dimer-dimer interface. In the RHD configuration, alternate-type ZFR mutants are positioned directly across from one another in the dimer-dimer interface.

Finally, after subunit rotation of an AHD configuration, the resulting tetramer state is the RHD configuration, and vice versa.

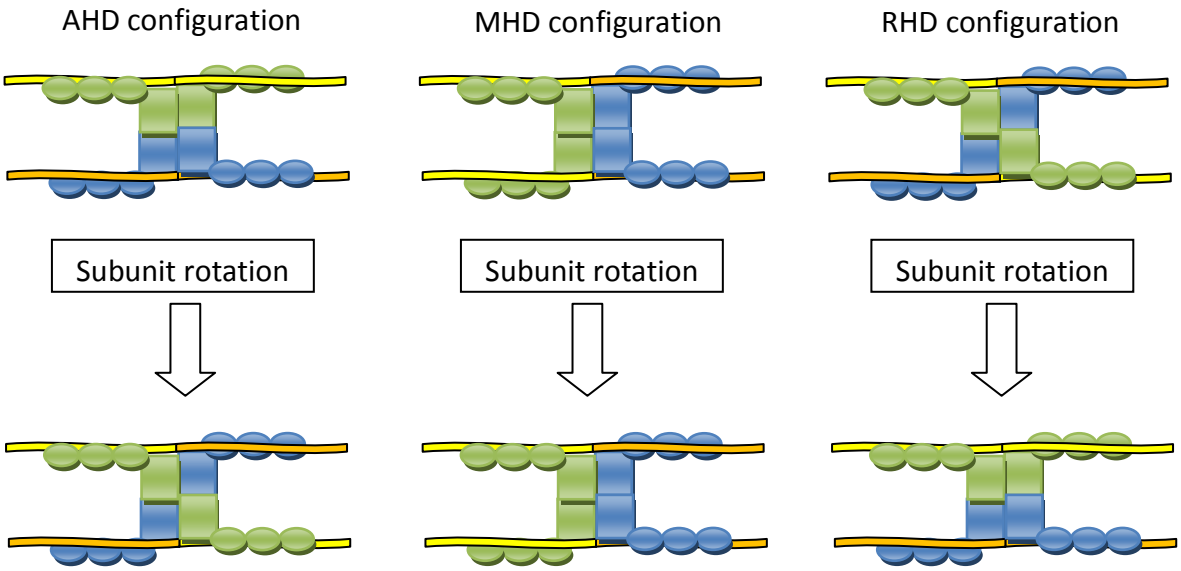


Figure 4-11: Dimer-dimer configuration states of two site-specific ZFR mutants within a tetramer. This figure depicts all three possible subunit configuration states of a tetramer comprised of two site-specific ZFR mutants. Double-stranded DNA molecules are represented by yellow and orange bars, while the two varieties of mutant ZFR subunits are depicted in green and blue. The green ZFR mutants bind the site represented by the yellow DNA, and the blue ZFR mutants binding the sites represented by the orange DNA. Arrows point to the post-subunit rotation state of the tetramer, which varies depending on the starting configuration.

4.3.1.7 Activity predictions based on E-helix and α/β sub-domain pairing mode

Working from the assumption that asymmetric dimer interface-unlocking might also give rise to compatible and non-compatible dimer-dimer configurations based on the E-helix and α/β sub-domain pairing mode (see Section 4.3.1.3), predictions were made about which tetramer configurations would be active (Figure 4-12). The MHD configuration was predicted to be active because its E-helix and α/β sub-domain pairing mode, with respect to the anti-parallel E-helix interactions during synapsis, would be wild-type-like. However, both the AHD and RHD configurations were predicted to be inactive because they share the same non-wild-type E-helix and α/β sub-domain pairing mode, with respect to the anti-parallel E-helix interactions during synapsis.

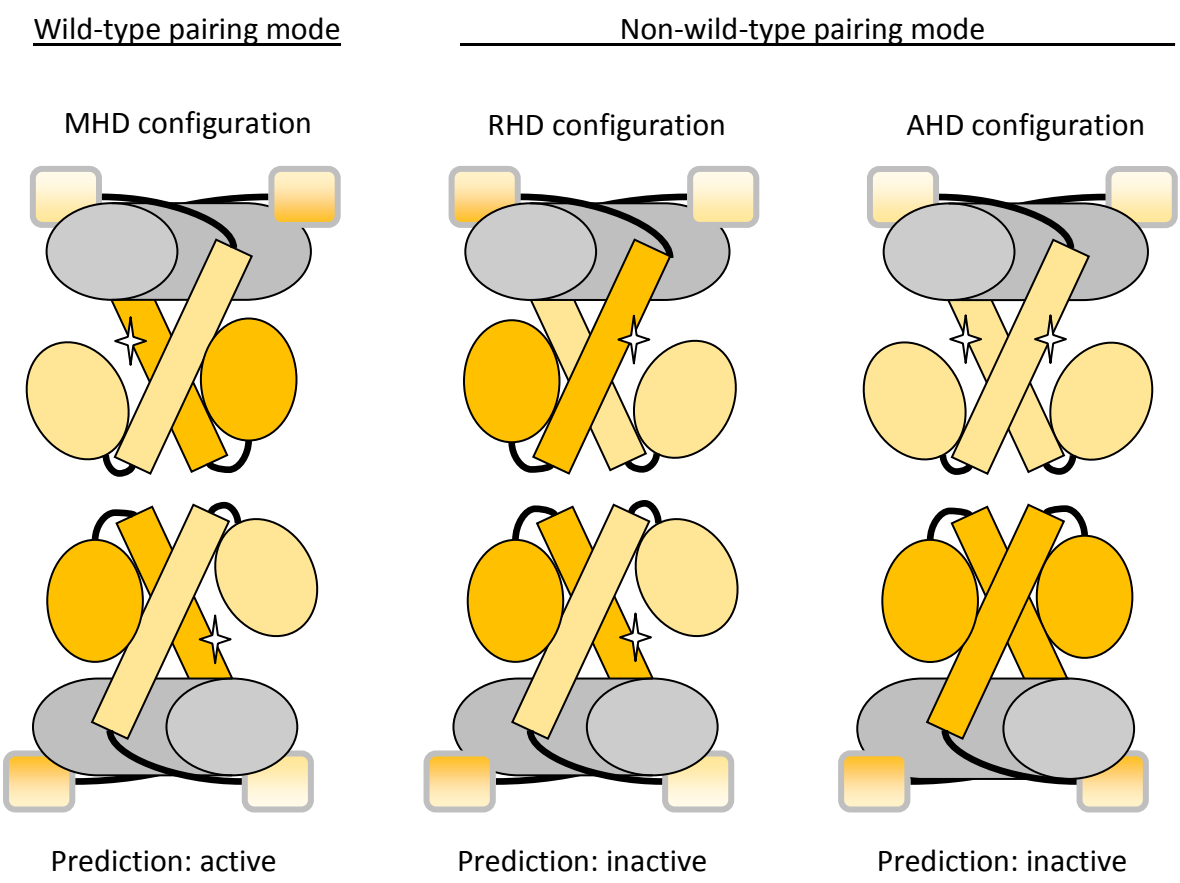


Figure 4-12: E-helix and α/β sub-domain dimer-dimer pairing modes. The figure shows the three possible E-helix and α/β sub-domain dimer-dimer pairing modes using two ZFRs where one contains a locking interface knockdown mutation (the light orange subunit). The MHD configuration is proposed to be the wild-type pairing mode, where the anti-parallel E-helices pair in the same mode, while the MHD and AHD configurations contain the non-wild-type pairing mode, where the anti-parallel E-helices pair in opposite modes. The wild-type pairing mode was predicted to be active while the non-wild-type pairing mode was predicted to be inactive. The resolvase subunits are coloured in two shades of orange. The ZFR ZFA DNA-binding domain is represented using squares with a shade transition in the colouring. The resolvase E-helix is shown using long rectangles, and the α/β sub-domain is depicted using ovals. Flexible amino acid regions connecting the domains (such as the arm region) are depicted using thick curved lines. Stars are used to indicate where the locking interface contact between the E-helices and α/β sub-domains has been disengaged.

4.3.2 Experiment design

4.3.2.1 Premise

In order to test whether the asymmetric dimer interface-unlocking activation could be used to generate a dimer-dimer orientation specificity bias, two ZFRs were constructed that were intended to mimic the asymmetric dimer unlocking activity of the wild-type system when used in combination with one another. One of these ZFRs possessed normal dimer interface residues that can participate in the locking interface, while the other ZFR possessed the secondary mutation, G70C, which was hypothesized to

disrupt the locking interaction of the dimer interface by disrupting contact from the α/β sub-domain side of the interaction with the E-helix (see Figure 4-5). When these two ZFRs form a heterodimer, the result is that only one of the two locking interfaces of the dimer is disrupted, which simulates the proposed asymmetric allosteric unlocking action from the regulatory module in a Tn3 resolvase synaptosome (see Figure 4-6).

Key to this experiment is that when the heterodimers comprised of differentially mutated ZFRs are in the MHD configuration, they are in the same E-helix and α/β subunit pairing mode as expected in the wild-type Tn3 resolvase synaptosome, with two unlocked locking regions being on the same side of the tetramer (Figure 4-12). Thus, the differentially mutated ZFR pairs were expected to produce recombination activity in the MHD configuration. However, if the differentially mutated ZFR pairs are in the RHD or AHD configurations, this produces a non-native E-helix and α/β subunit pairing mode, and these tetramer configurations were not expected to produce recombination activity.

4.3.2.2 Choice of locking-interface knockdown mutation

From the aforementioned candidate secondary mutations described in Burke et al. (2004) (see Section 4.3.1.4), G70C was selected for this experiment because it was predicted to have particularly robust disruptive effect on the locking region of the dimer interface. The carbonyl oxygen of the Gly70 backbone participates in a hydrogen bond across the locking interface with Arg121 (Figure 4-5 A). Residue Arg121, itself, when mutated, also acts as a secondary activating mutation (Burke et al., 2004). These mutations are, therefore, a good fit with the predicted model. Additionally, residue 70 is a flexible glycine that participates in a bend, so substitution for another amino acid at this position may disrupt the local geometry containing other residues participating in the locking interface. Finally, the glycine residue also does not possess a side chain, and so substitution with a cysteine residue may further distort the local conformation by introducing mass into an otherwise compacted space. Thus, the G70C mutation appeared to be both an ideal fit for the model being tested, and was predicted to have a particularly strong effect.

4.3.2.3 Substrates and detection of active tetramer configurations

A complementation experiment was designed that utilized the three substrate types: 2MutHomDim, 2MutHetDim-DR, and 2MutHetDim-IR (see Table 3-2), to discriminate between reactions that might produce either dimer-dimer orientation specificity (predicted), or a non-reversible reaction (not predicted). Dimer-dimer orientation specificity can be detected by using either the 2MutHetDim-IR or 2MutHetDim-DR substrates, which produce different recombination products (inversion product versus excision product) if only one of the two possible dimer-dimer orientations (MHD configuration or RHD configuration) is active (Figure 4-14 and Figure 4-15). However, because inversion products may themselves be subject to secondary reactions, analysis of the excision products from 2MutHetDim-IR and 2MutHetDim-DR substrates provides the most reliable information about a dimer-dimer orientation specificity bias.

In this experiment, a non-reversible reaction may be detected by comparing the outcomes of complementation reactions on the 2MutHomDim substrate with those on the 2MutHetDim-DR substrate or the 2MutHetDim-IR substrate (Figure 4-13, Figure 4-14, and Figure 4-15, respectively). Because the AHD configuration transforms into the RHD configuration after subunit rotation, and vice versa (see Figure 4-11), if one of these two tetramer configuration states is active while the other is not, this would likely indicate non-reversible reaction potential for that combination of ZFRs. That is, if the ZFRs that bind the *product* Z-sites are no longer able form an active tetramer configuration, then the reaction cannot be reversed.[4] Recombination activity in complementation reactions on the 2MutHomDim substrate, indicate that the AHD configuration is active. The production of inversion product from complementation reactions on the 2MutHetDim-DR substrate indicates that the RHD configuration is active. Likewise, the production of excision product from complementation reactions on the 2MutHetDim-IR substrate, also indicates that the RHD configuration is active. Table 4-1 provides a summary of the relationship between substrate type, reaction outcome, and active tetramer configurations.

[4] In the case that the AHD configuration was active, but the RHD configuration was not active, a secondary reaction might still be possible provided the MHD configuration was active, as the dimer-dimer interface could form in the other orientation. However, if the RHD configuration was active and the AHD configuration was not, then no secondary reaction would be possible as both dimer-dimer orientations produce the same AHD configuration.

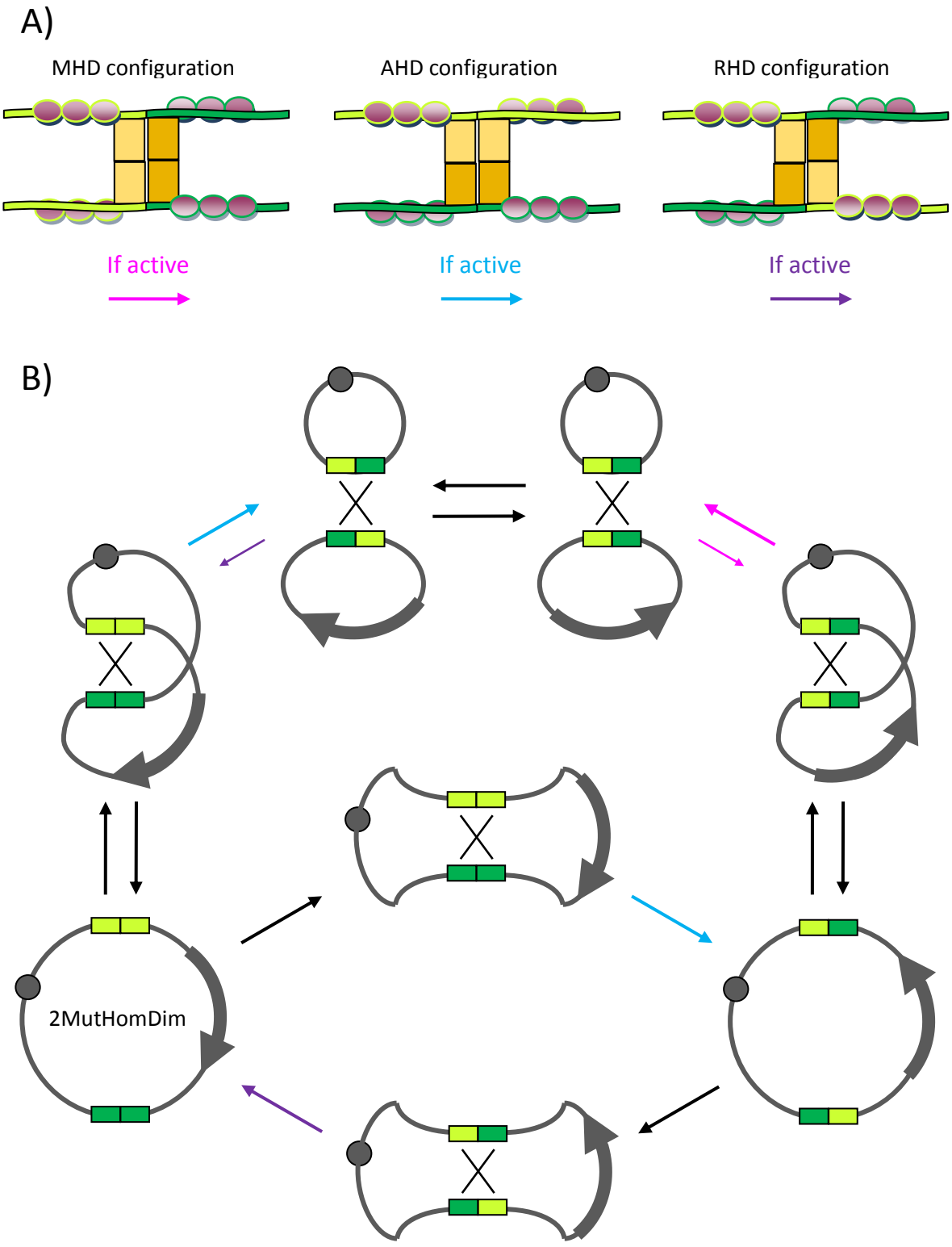


Figure 4-13: 2MutHomDim substrate and active tetramer configuration discrimination. A) The panel shows the three tetramer configurations with a colour code below to indicate which reactions in Panel B) involve the given configuration. **B)** The panel shows the possible reaction pathways for the 2MutHomDim substrate with the colour code from A) used to indicate which reactions are performed by a given tetramer configuration. The black arrows represent steps where the Z-sites are brought together, while the coloured arrows represent steps where recombination reactions take place. See **Figure 1-10** if a detailed description of the pictorial components of the figure is needed.

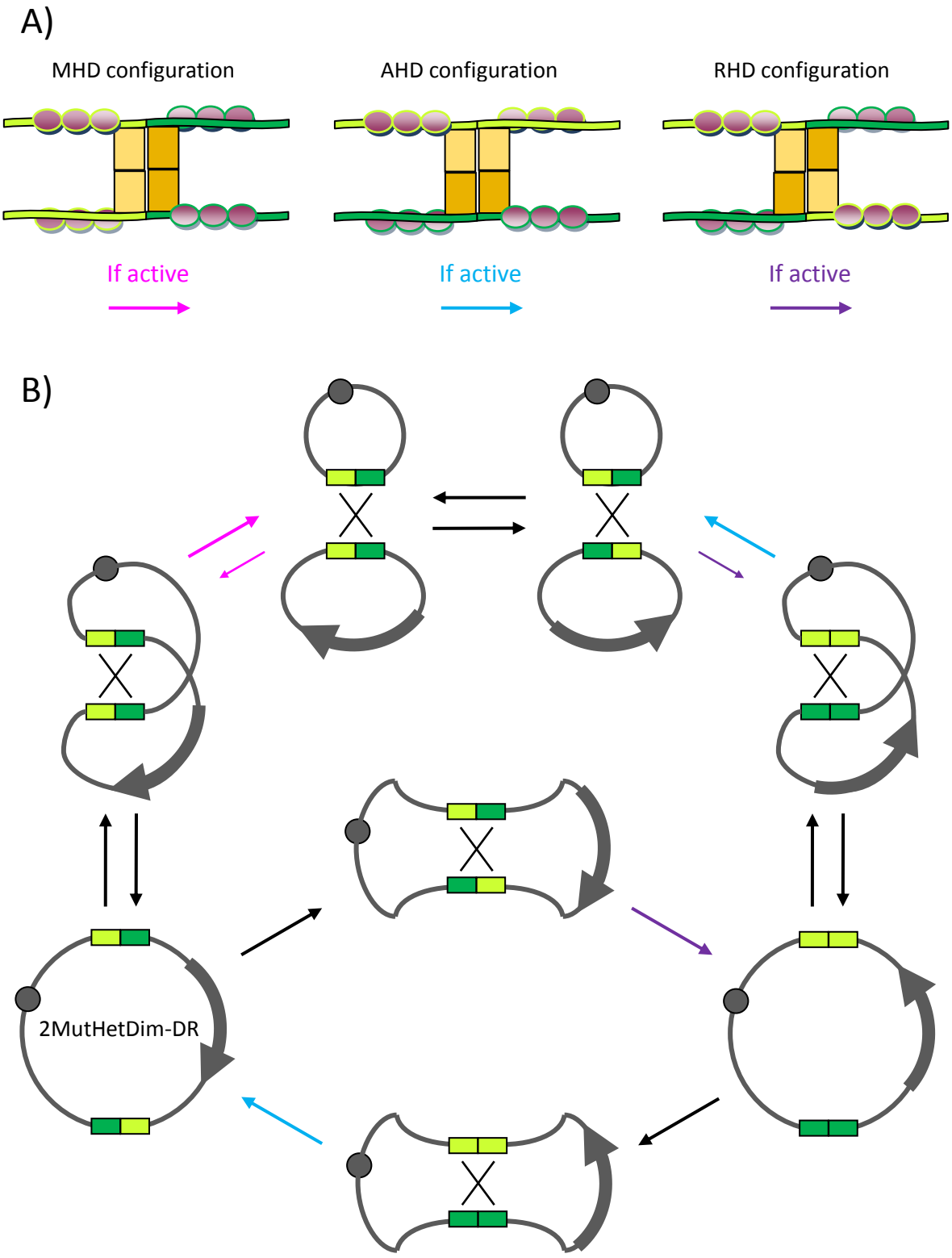


Figure 4-14: 2MutHetDim-DR substrate and active tetramer configuration discrimination. A) The panel shows the three tetramer configurations with a colour code below to indicate which reactions in panel B) involve the given configuration. **B)** The panel shows the possible reaction pathways for the 2MutHetDim-DR substrate with the colour code from A) used to indicate which reactions are performed by a given tetramer configuration. The black arrows represent steps where the Z-sites are brought together, while the coloured arrows represent steps where recombination reactions take place. See **Figure 1-10** if a detailed description of the pictorial components of the figure is needed.

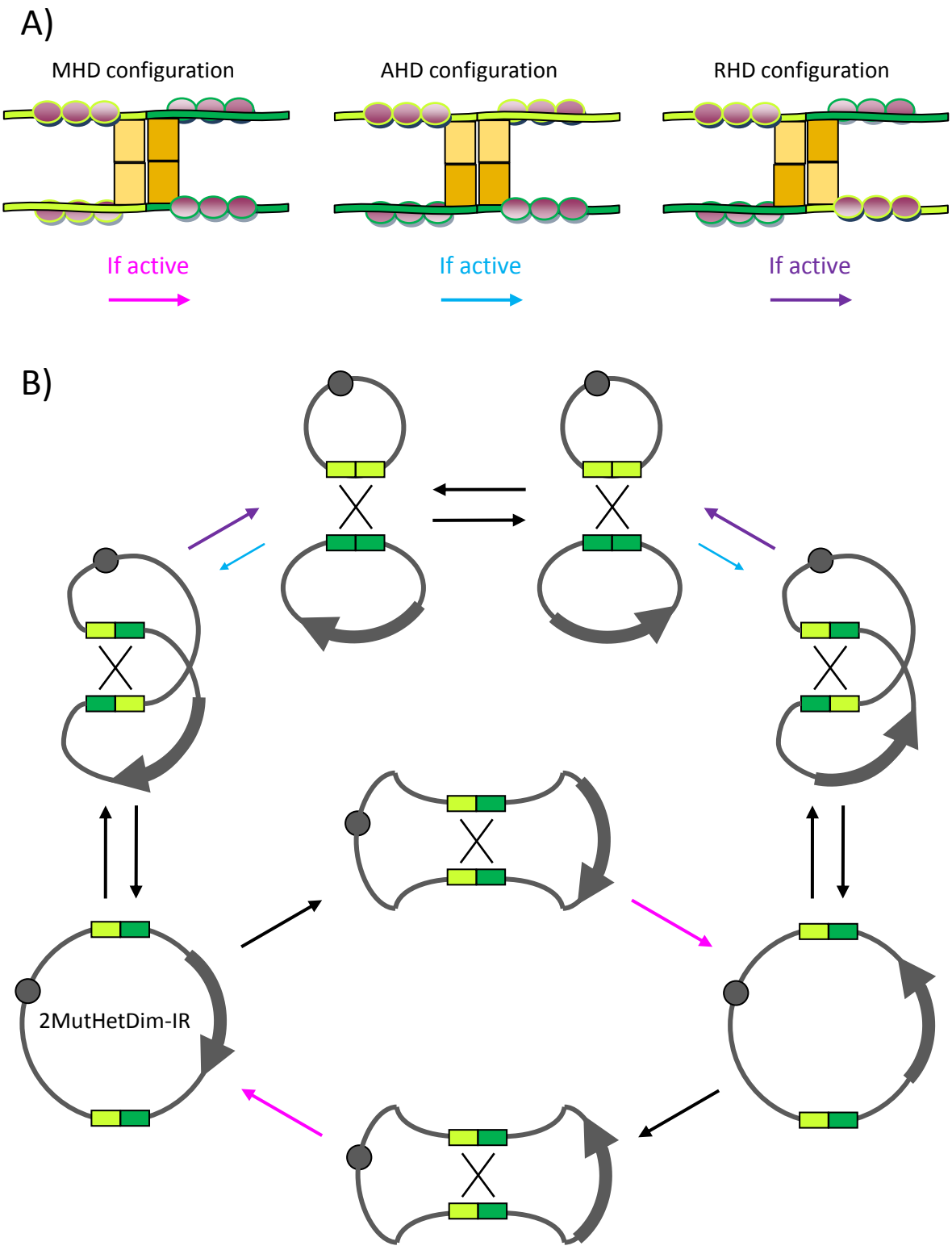


Figure 4-15: 2MutHetDim-IR substrate and active tetramer configuration discrimination. A) The panel shows the three tetramer configurations with a colour code below to indicate which reactions in panel B) involve the given configuration. **B)** The panel shows the possible reaction pathways for the 2MutHetDim-IR substrate with the colour code from A) used to indicate which reactions are performed by a given tetramer configuration. The black arrows represent steps where the Z-sites are brought together, while the coloured arrows represent steps where recombination reactions take place. See **Figure 1-10** if a detailed description of the pictorial components of the figure is needed.

Table 4-1: Summary of tetramer configurations and reaction outcomes on substrate types

	2MutHomDim		2MutHetDim-DR		2MutHetDim-IR	
	1°	2°	1°	2°	1°	2°
UNC		RHD		AHD		MHD
INV	AHD		RHD		MHD	
DEL	AHD	MHD	MHD	AHD	RHD	RHD

1° indicates the primary reactions, while 2° indicates secondary reactions that may take place only if the inversion product has been produced by a primary reaction.
UNC = unchanged substrate, INV = inversion product, DEL = deletion/excision product

4.3.2.4 ZFR pairs used for the experiment

For this experiment, two pairs of differentially mutated ZFRs were constructed. The catalytic domains used were the Tn3[NY] domain and the Tn3[NY G70C] domain. The 'N' and 'Y' designations represent the tertiary mutations R2A and E56K, and the primary activating mutation D102Y, respectively (see Chapter 1; Section 1.6.5 to see how 'primary', 'secondary', and 'tertiary' activating mutations are defined). Importantly, although an enzyme (either Tn3 resolvase or Tn3-based ZFR) with the NY mutation set contains activating mutations, this combination abolishes the activity of the protein when used on its own. Therefore, the ZFR with the Tn3[NY] domain is inactive by itself, and thus will be referred to as a 'sleepy' partner as it requires the unlocked partner to wake it up. The unlocked ZFR partner possesses the Tn3[NY G70C] catalytic domain, which contains a G70C secondary mutation in addition to NY. The G70C mutation, as mentioned in Section 4.3.2.2, disrupts the proposed locking interaction, and enzymes (either Tn3 resolvase or Tn3-based ZFR) with a Tn3[NY G70C] catalytic domain are active when used alone.

Although the binding activity of the Z2 and Z3 domains was shown to be approximately equal in the experiments of Chapter 3 (Section 3.5), both ZFAs were attached to both the sleepy and active catalytic domains (Tn3[NY] and Tn3[NY G70C], respectively), so that any minor differences in binding activity could be controlled for. Thus, the sleepy ZFRs, Tn3[NY]-Z2 and Tn3[NY]-Z3, and the active ZFRs, Tn3[NY G70C]-Z2 and Tn3[NY G70C]-Z3, were generated.

4.3.2.5 Substrates used for the experiment

As an additional control against any unexpected behaviour of the substrate plasmids, two substrates of each substrate type (2MutHomDim, 2MutHetDim-DR or 2MutHetDim-IR) with opposite binding sites arrangements were used (e.g. Z2/Z2 x Z3/Z3

and Z3/Z3 x Z2/Z2). Although the use of two varieties of each substrate, which should behave in an identical fashion, might seem redundant, there is a promoter somewhere upstream of one of the Z-sites, and so without having a lot of experience with these substrates it was thought prudent to control for variables of this nature. 1MutHomDim substrates were also used in order to confirm the expected activity levels of the sleepy and active ZFRs.

4.3.2.6 The experiment

A proof-of-principle experiment was conducted in which each set of sleepy + active ZFR pairs (Tn3[NY]-Z2 + Tn3[NY G70C]-Z3, and Tn3[NY]-Z3 + Tn3[NY G70C]-Z2) was tested on both varieties (see Section 4.3.2.5) of all substrate types (1MutHomDim, 2MutHomDim, 2MutHetDim-DR, and 2MutHetDim-IR), in order to test the asymmetric dimer interface-unlocking hypothesis. The experiment was carried out using the 17 Hour Recombination Assay.

The initial results from the experiment showed only very small amounts of recombination taking place and relevant signals observed by gel electrophoresis analysis were very faint. Therefore, the experiment was run again, and this time, instead of harvesting the cells at 17 hours, they were grown on in culture for another approximately 24 hour period to allow the apparently slow reaction to proceed further. It is the results from this cell passage that are described below. A second passage was also performed, and results from this passage showed even further conversion of substrate to reaction products; however, only results of the first passage are shown below. Catching the conversion of substrate to recombination products in an intermediate stage is most ideal for interpreting the results of the reaction because in some cases the inversion product of the reaction may disappear due to secondary conversion to excision product (see Table 4-1).

4.3.3 Results

4.3.3.1 Observations

The results of the experiment are shown in Figure 4-16 Panel A and B, and are summarized in Table 4-2. Panels A and B represent the same experiment but with the catalytic domain type and ZFA type combination, swapped. The controls on the 1MutHomDim substrates (lanes 2 and 3) confirm that the sleepy ZFR with the Tn3[NY] catalytic domain is inactive on its own, and that the active ZFR with the Tn3[NY G70C] catalytic domain produces a total conversion of substrate to recombination product, as was expected from the combinations of primary, secondary, and tertiary mutations used (see Chapter 1, Section 1.6.5).

Table 4-2: Summary of ostensible correlations between reaction results and configuration activity.

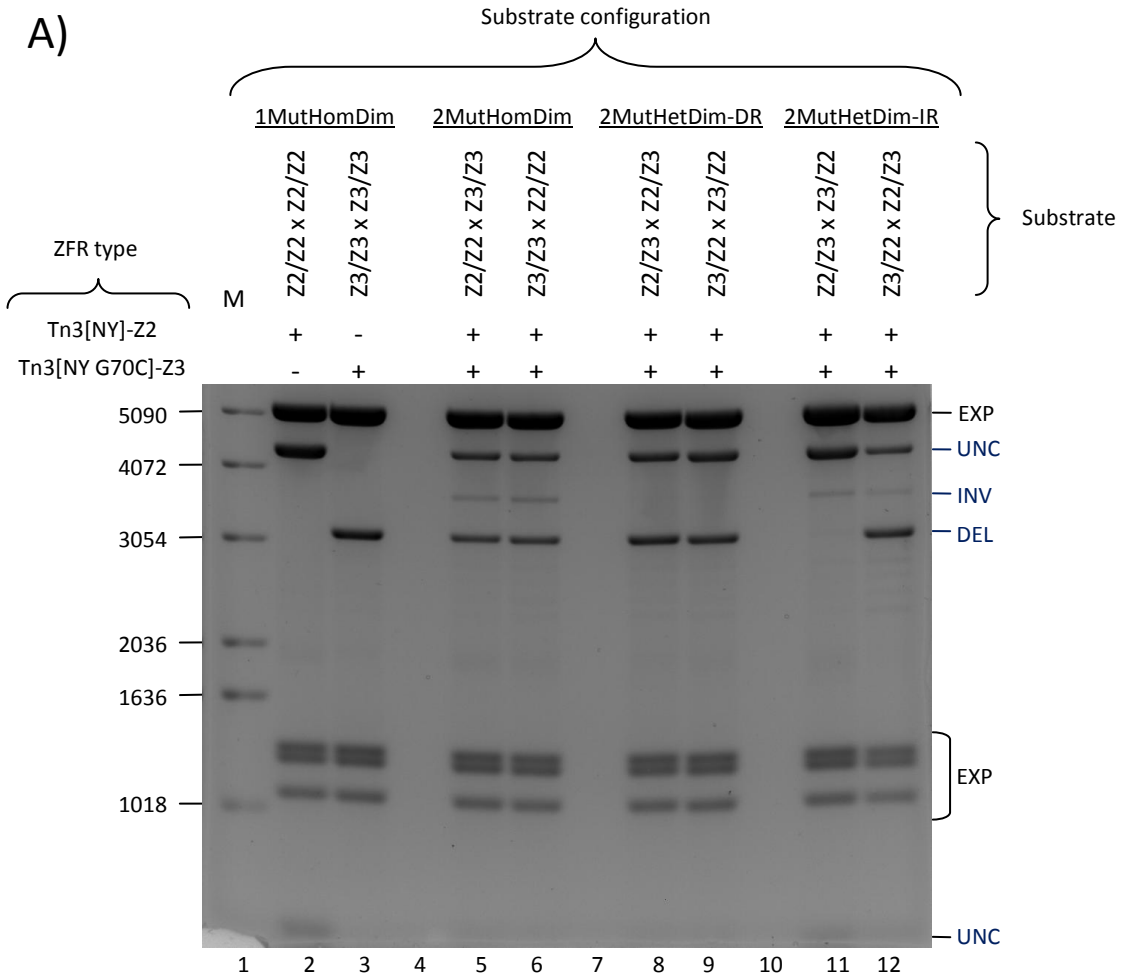
	2MutHomDim		2MutHetDim-DR		2MutHetDim-IR	
	1°	2°	1°	2°	1°	2°
UNC		RHD?		**AHD**		**MHD**
INV	**AHD**		~~RHD~~		**MHD**	
DEL	**AHD**	**MHD**	**MHD**	**AHD**	RHD	RHD?

1° indicates the primary reactions, while 2° indicates secondary reactions that may take place only if the inversion product has been produced by a primary reaction.

UNC = unchanged substrate, INV = inversion product, DEL = deletion/excision product

Bold is used to indicate configurations that appear to be active, or in the case of secondary reactions, can be deduced as active from the results of the primary reactions. Strikethrough is used to indicate configurations that clearly do not appear to be active in the assay. A question mark is used in secondary reactions when configuration activity cannot be clearly deduced from the primary reactions of the assay. The colour red is used to indicate disagreement between a complementary set of reactions that should have produce the same results.

The primary experimental data can be viewed in lanes 5 and 6, 8 and 9, and 11 and 12 (Figure 4-16 A and B), representing the sleepy + active ZFR pair complementation reactions on the 2MutHomDim, 2MutHetDim-DR, and 2MutHetDim-IR substrates, respectively. The first observation that can be made is that reactions on the 2MutHomDim substrates produced both inversion and excision product (lanes 5 and 6), indicating that the AHD tetramer configuration is active (see Table 4-1). This result falsifies the prediction that the AHD configuration would not be active due to E-helix and α/β sub-domain pairing modes (see Section 4.3.1.7).


A)
Substrate configuration
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
Substrate
Z2/Z2 x Z2/Z2
Z3/Z3 x Z3/Z3
Z2/Z2 x Z3/Z3
Z3/Z3 x Z2/Z2
Z2/Z3 x Z2/Z3
Z3/Z2 x Z3/Z2
Z2/Z3 x Z3/Z2
Z3/Z2 x Z2/Z3
ZFR type
M
Tn3[NY]-Z2
Tn3[NY G70C]-Z3
5090
4072
3054
2036
1636
1018
EXP
UNC
INV
DEL
EXP
UNC
1 2 3 4 5 6 7 8 9 10 11 12

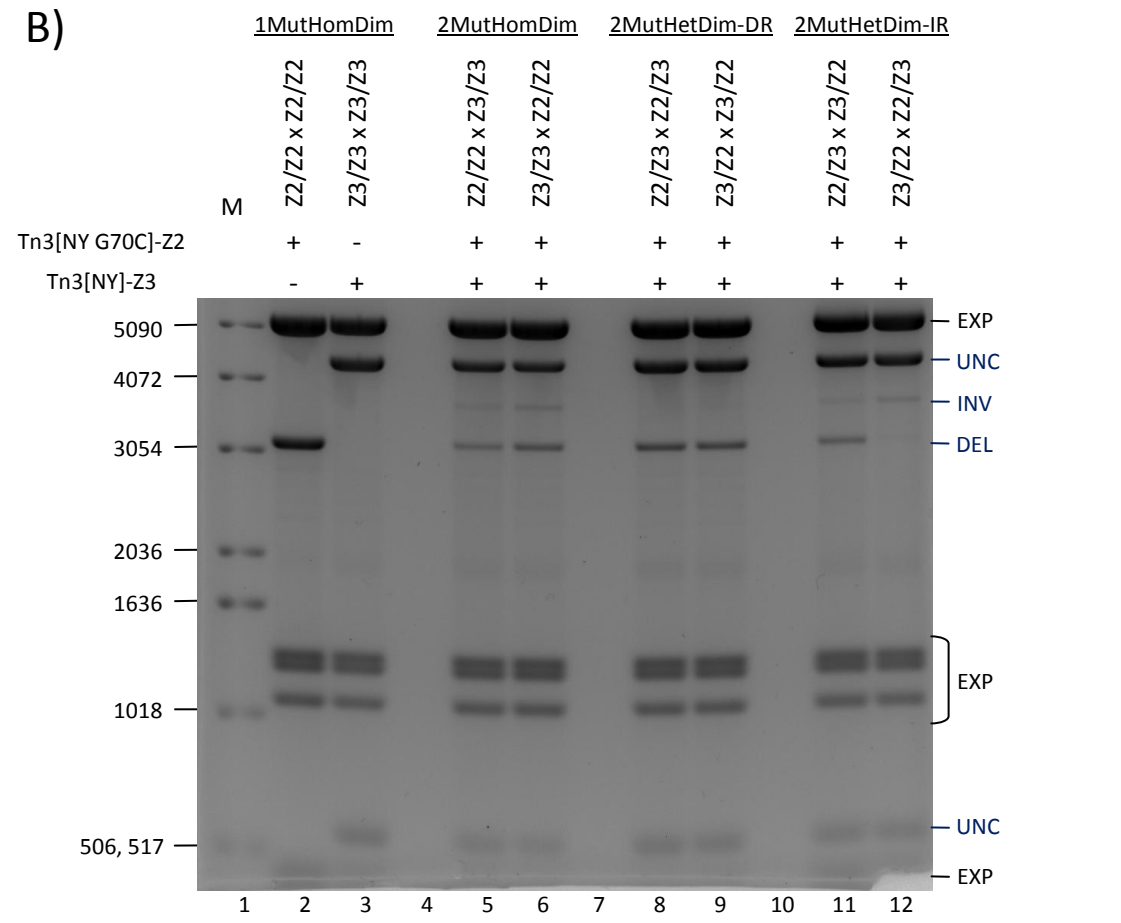

B)
1MutHomDim
2MutHomDim
2MutHetDim-DR
2MutHetDim-IR
Z2/Z2 x Z2/Z2
Z3/Z3 x Z3/Z3
Z2/Z2 x Z3/Z3
Z3/Z3 x Z2/Z2
Z2/Z3 x Z2/Z3
Z3/Z2 x Z3/Z2
Z2/Z3 x Z3/Z2
Z3/Z2 x Z2/Z3
M
Tn3[NY G70C]-Z2
Tn3[NY]-Z3
5090
4072
3054
2036
1636
1018
506, 517
EXP
UNC
INV
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12

Figure 4-16: Asymmetric dimer interface-unlocking activation proof-of-principle experiment. The results of a 17 Hour Recombination Assay which was further passaged in cell culture, are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. Above the table the substrate types are indicated. *Please note the restriction digest used for this experiment (AgeI + KpnI) was slightly different than that which is used for all the other experiments, and so the location of expression plasmid bands is somewhat different. This change was only made because of the availability of restriction enzymes on the day the digest was carried out. The bands of interest representing the substrate plasmid and its recombination products are still in their usual locations*.

The second observation that can be made is that reactions on the 2MutHetDim-DR substrates produced only excision product, but no inversion product (lanes 8 and 9). This result *appears* to indicate a dimer-dimer orientation specificity bias, as lack of inversion product from the 2MutHetDim-DR substrate indicates that the RHD configuration is inactive, while production of excision product indicates the MHD configuration is active (see Table 4-1). The RHD and MHD configurations represent the two possible orientations of the heterodimer dimer-dimer interactions (Figure 4-11), and so activity with one configuration and not the other, suggests a dimer-dimer orientation specificity bias. However, there is an alternative explanation for this result (see below).

The third observation that can be made is that there is disagreement between the two complementary 2MutHetDim-IR reactions in lanes 11 and 12. The appearance of inversion product in both lanes indicates that the MHD configuration is active (see Table 4-1), and this result is consistent with the previously observed results in lanes 8 and 9 that also indicate that the MHD configuration is active. However, in Panel A, an excision product signal is absent from lane 11, while a strong excision product signal appears in lane 12. In Panel B, this pattern is reversed with the excision product signal appearing in lane 11, but not in lane 12. In the lanes where the excision product band is absent, the result indicates the RHD configuration is inactive (consistent with the previously observed results in lanes 8 and 9), while its presence in the other lanes contradicts this result and indicates RHD configuration is, in fact, active (see Table 4-1).

The reactions indicated by lanes 11 and 12 in Panel A and B were expected to be functionally equivalent as they are all reactions on a 2MutHetDim-IR substrate. The reversed pattern of occurrence for the excision product between Panels A and B indicates

that the excision product is only produced when the sleepy ZFR subunits occupy the Z-site half-sites that are located on the same side as the origin of replication.[5] This correlation indicates that the excision product plasmids that contain sleepy ZFR homodimer Z-sites are present in the results, but those that contain active ZFR homodimer Z-sites are not.

4.3.3.2 Explanation based on the signal attenuation effect

One potential explanation that might explain all of the results is that the pattern of activity is caused by the product signal attenuation (SA) effect described in Chapter 3 (Section 3.10.5). This SA effect can occur in reactions involving the ZFRs utilizing the Z2 and Z3 ZFAs, especially when two *active* ZFRs are expressed simultaneously. In the current experiment, the cells had to be additionally passaged because the recombination product signals initially observed from the 17 Hour Recombination Assay experiment were weak, apart from the positive control (17 hour time point data not shown). The SA effect was proposed in Chapter 3 to result from unregulated cleavage of the DNA by ZFR dimers at a Z-site as the result of contact from non-DNA-bound ZFR subunits. Because the ZFR complementations of the current experiment produce slow reaction rates, as observed from the result before the additional cell passage, this is the type of scenario where the SA effect might significantly interfere with the results; i.e. the SA effect activity may outcompete the slow DNA recombination activity.

If the SA effect is caused by DNA cleavage from hyperactivated Tn3 catalytic domains as proposed, then the sleepy ZFR subunits used in this experiment might not produce the SA effect when bound to homodimer Z-sites. The sleepy ZFR subunits possess a Tn3 catalytic domain (NY) that is inactive without help from active ZFR subunits (and this is demonstrated by the negative controls in lanes 2 and 3 of Panel A and B, respectively, of Figure 4-16). If sleepy ZFR homodimers do not produce the SA effect, then what is expected is that substrates and substrate products that possess homodimer Z-sites cognate to the active ZFRs will be differentially prone to the SA effect. Under this assumption, the reaction products that would be most affected by the SA effect are the inversion products from reactions on the 2MutHetDim-DR substrates, and the excision products from the subset of reactions on 2MutHetDim-IR substrates that produce

[5] When looking at the substrates in Figure 4-16, one may recall from Table 3-2 that the format used to describe individual substrate plasmids is such that the half Z-sites which are on the same side as the 'x' are on the same side as the origin of replication.

product plasmids with active ZFR homodimer Z-site. The foregoing reaction products are the ones that are not observed in the results. Therefore, no conclusions from this experiment can be drawn with respect to the ability of the sleepy + active ZFR pair to produce dimer-dimer orientation specificity, in spite of the fact that some of the results ostensibly appear consistent with that possibility.

It can be additionally noted that it is unknown whether the single co-expression of sleepy ZFRs with active ZFRs would increase the SA effect in the same way that the dual expression active ZFRs does. However, since the signals observed after the 17 hour time point DNA harvest were weak, it is possible that signals at this level might be interfered with by even the SA effect produced by the single expression of the active ZFR in the reaction.

4.3.4 Increasing ZFR expression of the sleepy + active ZFR pair

4.3.4.1 Rationale for the experiment

Because it was not known how the sleepy ZFR subunits may interact with the SA effect, one way to attempt to determine this was to increase the expression of all the ZFRs in the reaction so that more obvious effects could be observed. Running the same experiment with double the expression level allows observations to be made about whether the sleepy ZFRs produce the SA effect when used alone at a higher expression level, and also whether they produce the SA effect when participating with active ZFRs in binding to heterodimer Z-sites. Additionally, the requirement of a cell passage and over 40 hours of total time to acquire the results from the sleepy + active ZFR pair complementations in the proof-of-principle experiment, implies the activity of this ZFR combination to be low. This low rate of recombination is not ideal for further experimental work. If the ZFR occupancy at the Z-sites is not saturated at the level of expression in the previous experiment, then increasing the expression of ZFRs would improve the rate of the reaction. Therefore, it was deemed worthwhile to perform the previous experiment again but with increased ZFR expression.

4.3.4.2 Experimental design

The sleepy ZFR (Tn3[NY]-Z2) and active ZFR (Tn3[NY G70C]-Z3) were loaded into double expression vectors (DEVs). Since there did not appear to be any confounding

effect produced by ZFA activity differences in the previous proof-of-principle experiment, only one sleepy and active ZFR pair were used for this DEV experiment. A complementation experiment using the sleepy + active ZFR pair was carried out in the precisely the same way as described for the previous proof-of-principle experiment, with the only change being that two copies of each ZFR were expressed from DEVs. For this experiment, select ZFR complementations in single expression mode were also performed again so that they could be placed side-by-side with the DEV-based complementations for comparison on the gel. Like the previous proof-of-principle experiment, results were collected at the initial 17 hour point, and again after one cell passage; the results collected at both of time points are shown bellow.

4.3.4.3 Results

The results of the sleepy + active ZFR complementation experiment using DEVs (displayed in Figure 4-17 and Figure 4-18) indicate that increasing ZFR expression does not substantially increase the rate of the reaction. As indicated in the table above the gels, for each substrate type, the single expression mode complementations are shown next to the DEV mode complementations. A comparison of the single expression mode and DEV mode complementations at the 17 hour time point (Figure 4-17), shows almost no difference between them; a larger difference in substrate recombination rate can be seen after one cell passage (Figure 4-18). The increase in the rate of the reaction, due to increasing the ZFR expression within the cells, is not sufficient to obviate the need for a cell passage, which was one of the goals of the DEV augmentation to the experiment.

The first observation that can be made regarding the SA effect is that the dual expression of the active ZFR cognate for the 1MutHomDim, in addition to dual expression of the sleepy ZFR that is not cognate for the substrate, produces a significant level of SA effect (Figure 4-17, lane 3). Conversely, dual expression of the sleepy ZFR cognate for the 1MutHomDim, in addition to dual expression of the active ZFR that is not cognate for the substrate, does not appear to produce the SA effect at all (Figure 4-17, lane 2). This provides a strong indication that the SA effect is the result of ZFR catalytic activity, rather than binding alone. Additionally, this result also indicates that the co-expression of active ZFRs along with sleepy ZFRs, will not activate those sleepy ZFRs to produce the SA effect when only sleepy ZFR Z-sites are contained in the substrate. This is a noteworthy observation given that in Chapter 3 it was observed that active ZFRs that were not

cognate for a substrate increased the SA effect when co-expressed with active ZFRs for that were cognate for the substrate (see Section 3.10.5).

The second observation that can be made regarding the SA effect is that compared with the lane 2 (Figure 4-17), where no SA effect was observed, the co-expression from two single expression vectors of the sleepy and active ZFR pair, produced a notable SA effect on the 2MutHomDim substrate (Figure 4-17, lane 5). This result confirms that the results of the previous proof-of-principle experiment may have indeed been affected by the SA effect under these expression conditions. Additionally, it can be observed that under the co-expression from two single-expression vectors condition, a detectable level of SA effect can also be observed on the reactions involving the 2MutHetDim substrates (Figure 4-17, lanes 9 and 13). The SA effect on the 2MutHetDim substrates is less pronounced than on the 2MutHomDim substrate.

A final observation can be made that the SA effect occurs on the 2MutHetDim substrates when the sleepy + active ZFR pair are co-expressed (in both single expression vector and DEV expression conditions), but at reduced level compared with reaction on the 2MutHomDim substrates. This indicates that when the sleepy + active ZFR pair is co-expressed, sleepy ZFRs can participate in producing a modest level of SA effect when cooperating with active ZFRs at heterodimer Z-sites. Conversely, as previously mentioned, the lack of SA effect on the 1MutHomDim substrate when cognate sleepy ZFRs are co-expressed with active ZFRs in DEV mode, indicates that sleepy ZFRs cannot be stimulated to produce the SA effect by non-bound active ZFRs when bound to Z-sites as homodimers.

By making a comparison of the unchanged substrate signals across the lanes in Figure 4-17 a general characterization of the size of the SA effect under different conditions of expression and active ZFR binding can be made. Comparing the results in lanes 1 to lanes 9 and 13 it can be concluded that at single expression vector expression levels, when one Tn3[NY G70C] ZFR participates in heterodimer Z-site binding, the SA effect is modest (perhaps 25% reduction in signal strength). Comparing the results in lanes 1 to lane 5 it can be concluded that at single expression vector expression levels, when two Tn3[NY G70C] ZFRs participate in homodimer Z-site binding, the SA effect is intermediate (perhaps 50% reduction in signal strength). Comparing the results in lanes 1 to lanes 10, 11, 12, and 15 it can be concluded that at DEV expression levels, when one

Tn3[NY G70C] ZFR participates in heterodimer Z-site binding the SA effect is intermediate (perhaps 50% reduction in signal strength). Comparing the results in lanes 1 to lanes 6 and 7 it can be concluded that at DEV expression levels, when two Tn3[NY G70C] ZFRs participate in homodimer Z-site binding the SA effect is profound (perhaps 90% reduction in signal strength). These observations are summarized in Table 4-3.

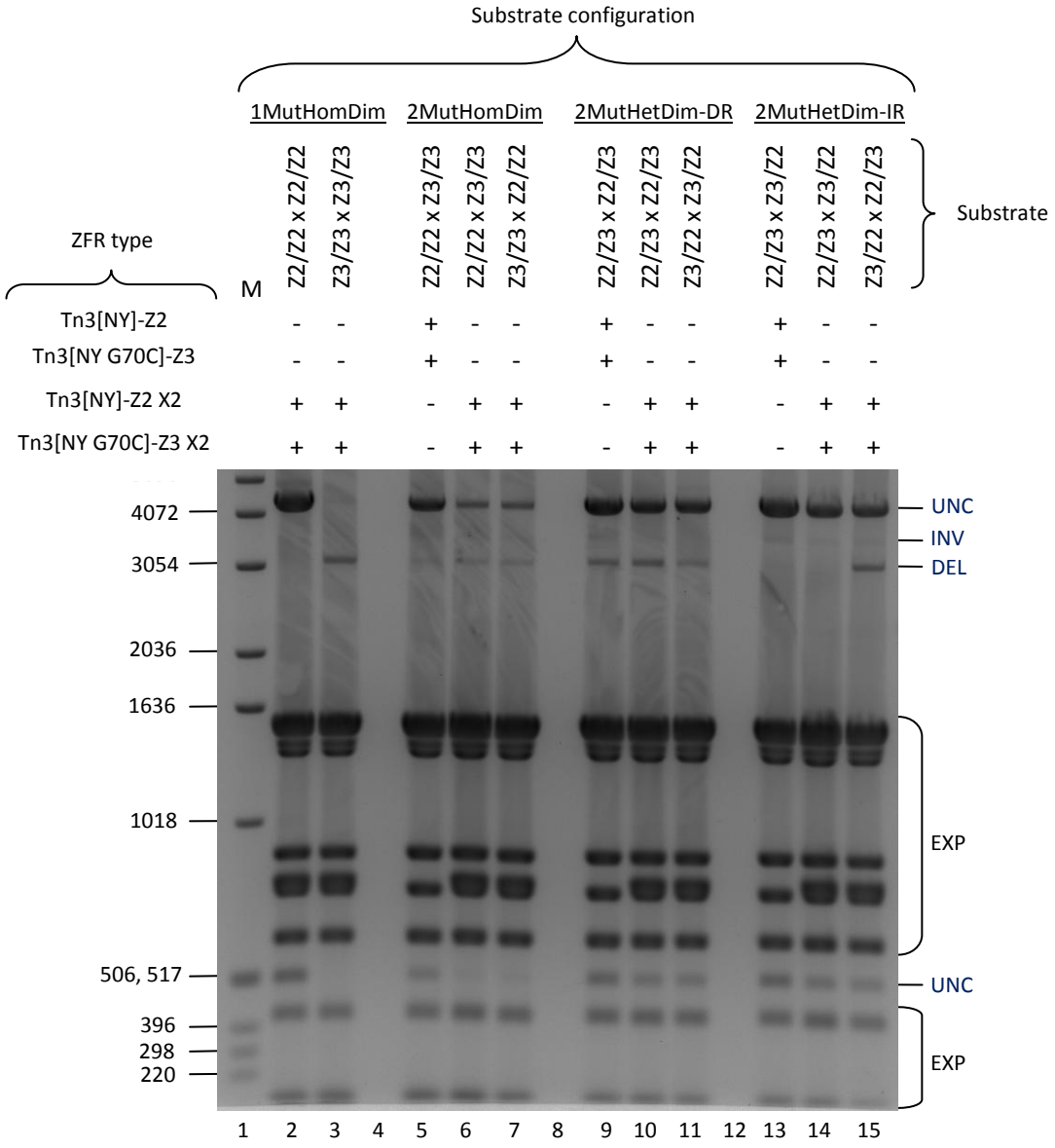


Figure 4-17: Sleepy and unlocked ZFRs complementation reaction using DEVs, at the 17 hour time point. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. Above the table the substrate types are indicated.

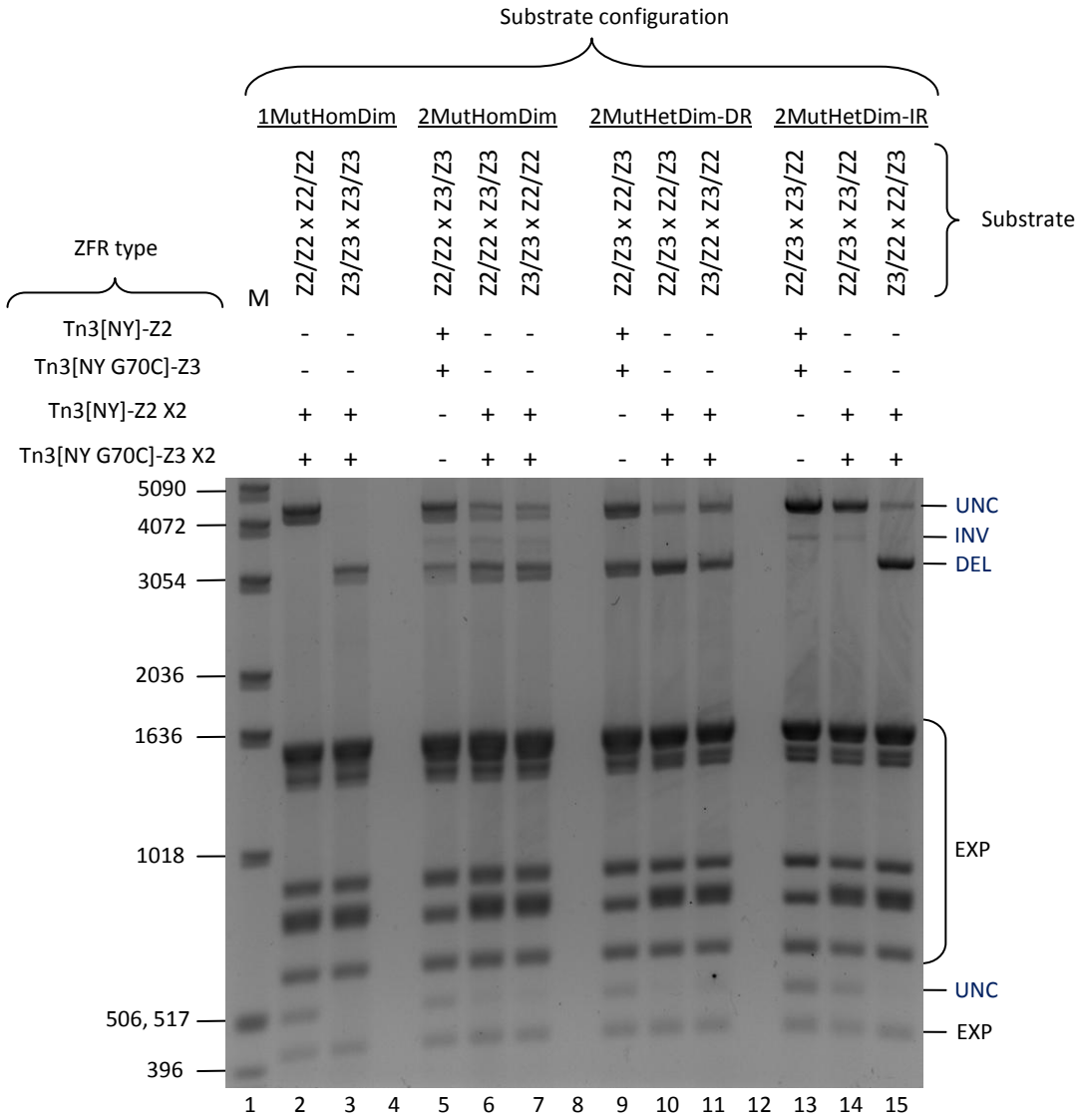


Figure 4-18: Sleepy and unlocked ZFRs complementation reaction using DEVs after one cell passage. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. Above the table the substrate types are indicated. *The doubling of the bands seen in the gel is an artefact resulting from sloping of the wells, and the fact that the DNA at the top of the sloped wells appears to have stained more strongly than the DNA throughout the rest of the sloped well.*

Table 4-3: Summary of the SA effect size relative to expression and binding conditions

	Tn3[NY G70C] Heterodimer	Tn3[NY G70C] Homodimer
SEV expression	Modest ~25% reduction	Intermediate ~50% reduction
DEV expression	Intermediate ~50% reduction	Profound ~90% reduction

SEV = single expression vector, DEV = double expression vector

The percentages given in the table are just visual estimates.

4.3.5 Activity screen for other potential interface-unlocking activation mutants

4.3.5.1 Rationale for the experiment

The results of the previous experiment were confounded by the SA effect both because the active ZFR mutant was capable of producing it, and because the reactions using the sleepy + active ZFR pair were so slow that the SA effect outcompeted the small amount of reaction products being produced. Additionally, this slow reaction, which required passaging the cells, is not ideal for further experimental work, and increasing the expression of ZFRs was also not successful in increasing the reaction rate. However, the experiment did demonstrate that an inactive ZFR mutant can be activated through cooperation with a more active ZFR mutant. This discovery is potentially valuable in and of itself, as gene therapy applications might benefit from such a ZFR set up in an integration scenario. The use of sleepy ZFRs at the Z-site(s) on the DNA to be integrated would reduce the possibilities for off-target reactions. Toxicity has been observed when ZFRs were used in mammalian cells, which appears to be due to off-target reactions (Gaj et al., 2014). Therefore, two desirable goals, in order to further the utility of the sleepy + active ZFR strategy, were to increase the reaction rate, and to reduce the SA effect from the active ZFR of the pair.

The sleepy ZFR used in the previous experiment did not produce the SA effect when bound as a homodimer to a Z-site, but it was able to participate in proper ZFR recombination reactions with the help of the active ZFR, albeit these reactions were slow. One approach taken to increase the activity in the sleepy + active ZFR reaction was to attempt to discover a new ZFR mutant to act as the sleepy ZFR which affords more activity to the reaction when used in complementation with the active ZFR, but yet still retains inactivity when operating on its own. Likewise, an attempt to discover a new ZFR mutant to act as the active ZFR, but which is less prone to producing the SA effect, was also made.

Although the hypothesis that E-helix and α/β sub-domain pairing modes might give rise to differential tetramer configuration activity was likely falsified by the results in Section 4.3.3, the basic premise that activation of the Tn3 catalytic domain occurs by an

E-helix and α/β sub-domain unlocking mechanism is still a viable model to work through. One way to test and exploit that model is to attempt to determine whether different levels of activation occur depending on whether two locking interface mutations, which act from opposite sides of the interface, are in a co-localized or distributed arrangement in a differentially mutated ZFR dimer pair (Figure 4-19). In this set up one ZFR would carry a mutation in the E-helix side of the interface, and the other ZFR would carry a mutation in the α/β sub-domain side of the interface. In a co-localized arrangement, heterodimers would be formed where both mutations would be paired in the same locking interface on one side of the dimer, while the other locking interface on the other side of the dimer would contain no mutations. In a distributed arrangement, two alternate homodimers would be formed that either contained the two E-helix locking interface mutations, one in each side of the dimer, or two α/β sub-domain locking interface mutations, one in each side of the dimer. Differential activity of the between the AHD and RHD tetramer configurations could lead to a desirable recombination reaction directionality bias.

4.3.5.2 Selection of mutations

Known Tn3 secondary activating mutations were categorized into those that might disrupt the locking interface from the E-helix side of the interaction, and those that might disrupt the interface from the α/β sub-domain side of the interaction (Table 4-4). The mutations that potentially interfere with the α/β sub-domain side of the locking interaction are either located at positions that directly participate in the interaction, positions that might destabilize the α/β sub-domain core, or positions that may help stabilize the unlocked conformation.[6] The mutations were selected primarily from Burke et al. (2004) (Burke et al., 2004), but also from a compendium of mutations compiled from unpublished work within the Stark lab. Because of the low activity produced by the sleepy + active ZFR pair in the proof-of-principle experiment, it was deemed prudent to first screen potential mutants for their base level of activity in the ZFR system prior to any complementation experiments.

[6] G101S was used in the γδ resolvase of the 1ZR4 structure and appears to stabilize the unlocked E-helix and α/β sub-domain conformation through a hydrogen bond from the serine side chain to the backbone nitrogen of Gly104. This is displayed later in Figure 4-21.

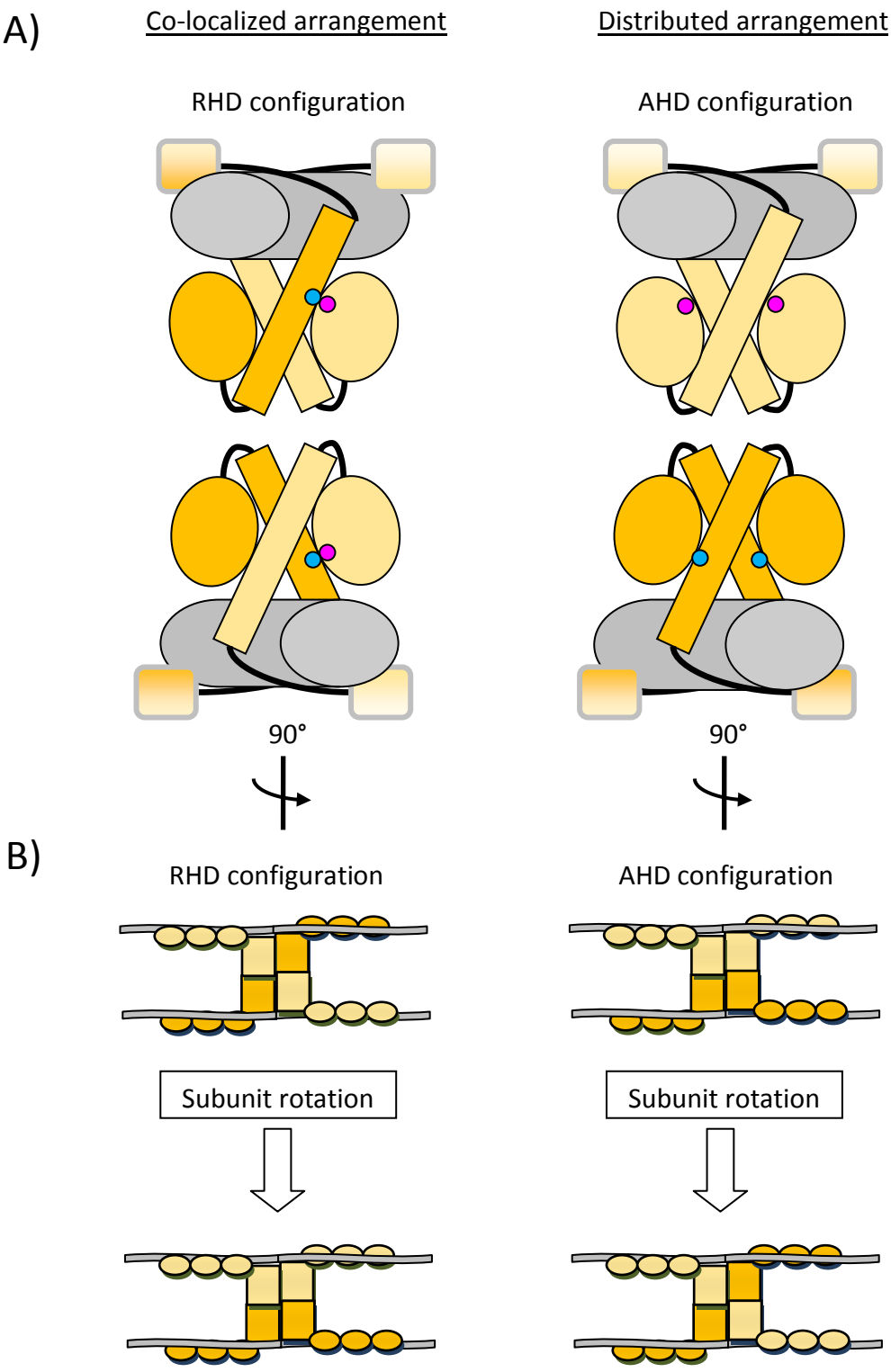


Figure 4-19: Differential locking interface mutation arrangements. A) The panel shows a co-localized and distributed arrangement of the locking interface mutations on the left and right, respectively. The locking interface mutation are indicated by small blue and magenta circles and are located in the E-helix or α/β sub-domain, respectively. The co-localized arrangement is shown in a RHD configuration, while the distributed arrangement must occur in the AHD configuration. **B)** The panel shows the same two dimer-dimer configurations as in A), but rotated 90°. This panel shows how during a recombination reaction the RHD configuration becomes the AHD configuration, and vice versa. Therefore, recombination reactions transform the co-localized locking interface mutation arrangement into the distributed arrangement (which produces a reaction directionality bias if one of the arrangements is more active than the other).

Table 4-4: Categorization of selected mutants

E-helix side	α/β sub-domain side
T109I, A117V, R121K/M/S, E124Q	L66I/F, G70A/C, I77T, A89T, F92S, I97M, T99S, G101S/C, M103I/V, Q105L, V107F/M, V108A

The mutations shown in red also have alternative or additional explanations for the activity they produce (see Section 4.4).

4.3.5.3 Experiment design

For this preliminary experiment, designed to measure the activity level of known Tn3 secondary mutations in the ZFR system, the mutations were incorporated into the Tn3[NY] (sleepy ZFR) background that was used in the previous experiments (see Section 4.3.2). The new catalytic domains were then attached to the Z3 ZFA, and the new ZFRs were tested on their cognate 1MutHomDim substrate in a 17 Hour Recombination Assay. Additionally, Tn3[N G70C]-Z3 and Tn3[NY]-Z3 were included in the experiment as negative controls, while Tn3[NY G70C]-Z3 and Tn3[NM]-Z3 acted as a positive controls.

4.3.5.4 Results

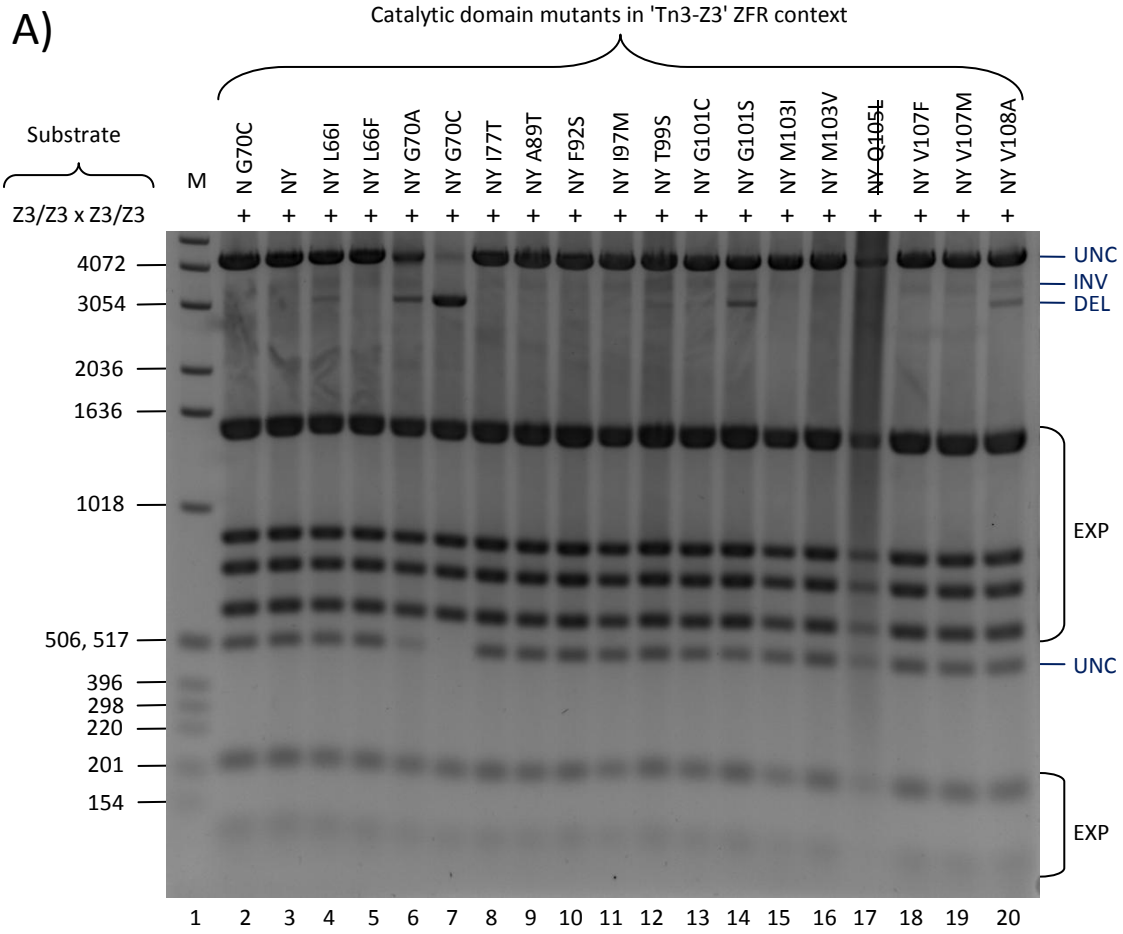
The results of the experiment (shown in Figure 4-20) were surprising in that almost all of the ZFRs created were not active. This result contrasts with the behaviour of these mutations in Tn3 resolvase, as the Burke et al. (2004) study and unpublished work within the Stark lab clearly demonstrated these mutant catalytic domains had surpassed the hyperactivity threshold (see Chapter 1, Section 1.6.5 for a description of hyperactivity). This indicates the Tn3-based ZFR system is intrinsically much less active in comparison with Tn3 resolvase.

Of the mutants that were notably active in the screen, Tn3[NY **L66I**]-Z3 produced a faint, but detectable, amount of recombination product (Figure 4-20 A, lane 4), Tn3[NY **G70A**]-Z3, Tn3[NY **G101S**]-Z3, and Tn3[NY **V108A**]-Z3 demonstrated a low level of recombination activity (Figure 4-20 A, lanes 6, 14, and 20, respectively), and, as expected, Tn3[NY **G70C**]-Z3 and Tn3[NM]-Z3 produced a total conversion of substrate to recombination product (Figure 4-20 A, lane 7 and Figure 4-20 B, lane 10, respectively). The results of this experiment suggested that G70A, G101S, and V108A might be suitable as replacement secondary mutations for G70C in the active ZFR of the sleepy + active ZFR pair, provided the sleepy ZFR has increased activation potential through the inclusion of its own secondary mutation. Likewise, the results also suggest that the remaining

mutations might be suitable for inclusion in the sleepy ZFR of a sleepy + active ZFR pair as these catalytic domains possess no substantial activity on their own in the ZFR system.

The mutations G101S and V108A might prove particularly useful if used in concert. It can be derived from analysing the 1ZR4 crystal structure of an activated resolvase tetramer that contains the G101S mutation, that G101S may stabilize the unlocked conformation producing a stabilizing hydrogen bond between the Ser101 hydroxyl and Gly104 backbone amide (Figure 4-21 B). Additionally, using the 'morph' tool of the UCSF Chimera molecular modelling software package (http://www.cgl.ucsf.edu/chimera) to interpolate the conformational states between a locked 1GDT subunit and an unlocked 1ZR4 subunit, it can be seen that the V108A mutation might produce a dimer interface-unlocking effect by removing a potential steric energy barrier between Val108, and Thr99 and Ile80 as the α/β sub-domain translocates across the E-helix in (Figure 4-22). The potential clash between Thr99 and Val108 is particularly sustained throughout the conformational transition, and this is interesting given that T99S is also a known activating mutation and T99S potentially removes this steric energy barrier as well. Because G101S and V108A may support the formation of the unlocked conformation without actually knocking out the locking interface, they might still require some stimulation from tetramer formation in order to engage in catalytic activity. Thus, these mutations might be less prone to the SA effect than G70C, which directly produces a locking interface disruption and thus may be more ready to produce unregulated cleavage when stimulated by unbound subunits (see Section 3.10.5 for a description of how the SA effect is proposed to operate).

One notable feature of the results is that they support the prediction that G70C would produce a disproportionately large locking interface disruption compared with other locking interface activating mutations (see Section 4.3.2.2). It appears in a comparison of the results from the Tn3[NY G70C]-Z3 ZFR and the Tn3[NM]-Z3 ZFR, that within this experiment the former may be as active as the latter; although, the reaction would have to be caught at an earlier stage to confirm that their activity level is actually equal, as all the substrate had already been converted to recombination product. This result is notable because the 'M' collection of mutations is considered highly active and requires several secondary mutations used in conjunction to produce this level of activity.


A)
Catalytic domain mutants in 'Tn3-Z3' ZFR context
Substrate
Z3/Z3 x Z3/Z3
M
N G70C
NY
NY L66I
NY L66F
NY G70A
NY G70C
NY I77T
NY A89T
NY F92S
NY I97M
NY T99S
NY G101C
NY G101S
NY M103I
NY M103V
NY Q105L
NY V107F
NY V107M
NY V108A
+
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNC
INV
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

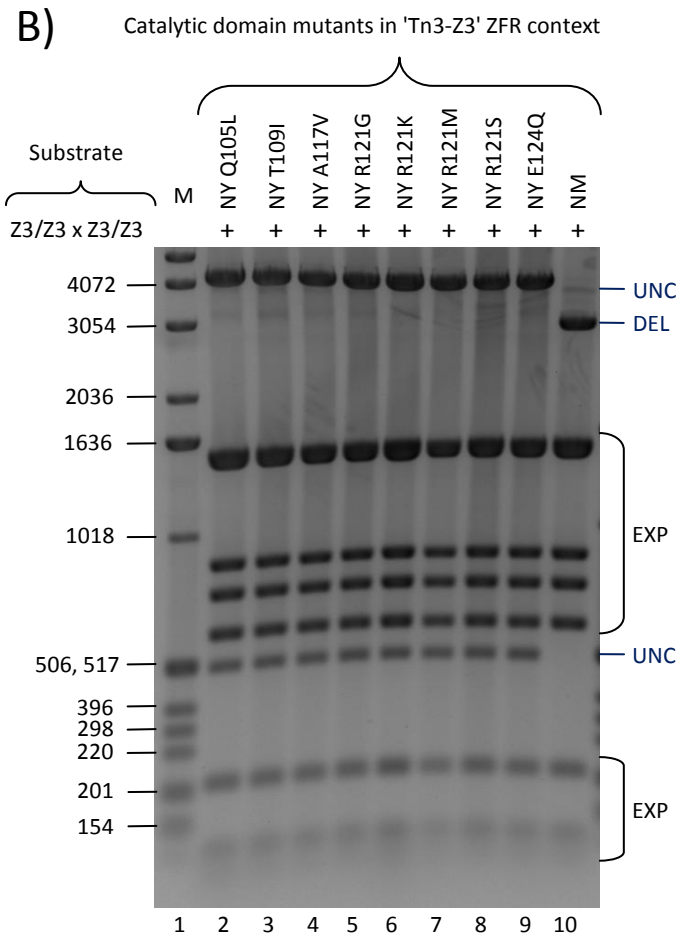

B)
Catalytic domain mutants in 'Tn3-Z3' ZFR context
Substrate
Z3/Z3 x Z3/Z3
M
NY Q105L
NY T109I
NY A117V
NY R121G
NY R121K
NY R121M
NY R121S
NY E124Q
NM
+
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNC
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10

Figure 4-20: Base activity level screen of other potential locking interaction knock-down mutants in the ZFR system. The results of a 17 Hour Recombination Assay are shown spread across two gels, A) and B). To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. The sample shown A) lane 17 appeared to be heavily contaminated with chromosomal DNA left over from the mini prep, and therefore that experiment was repeated and is displayed again in B) lane 2. Tn3[N G70C]-Z3 and Tn3[NY]-Z3 were included in the experiment as negative controls, and are shown in gel A) lanes 2 and 3, respectively. Tn3[NM]-Z3 was also included, as a positive control, and appears in gel B) lane 10.

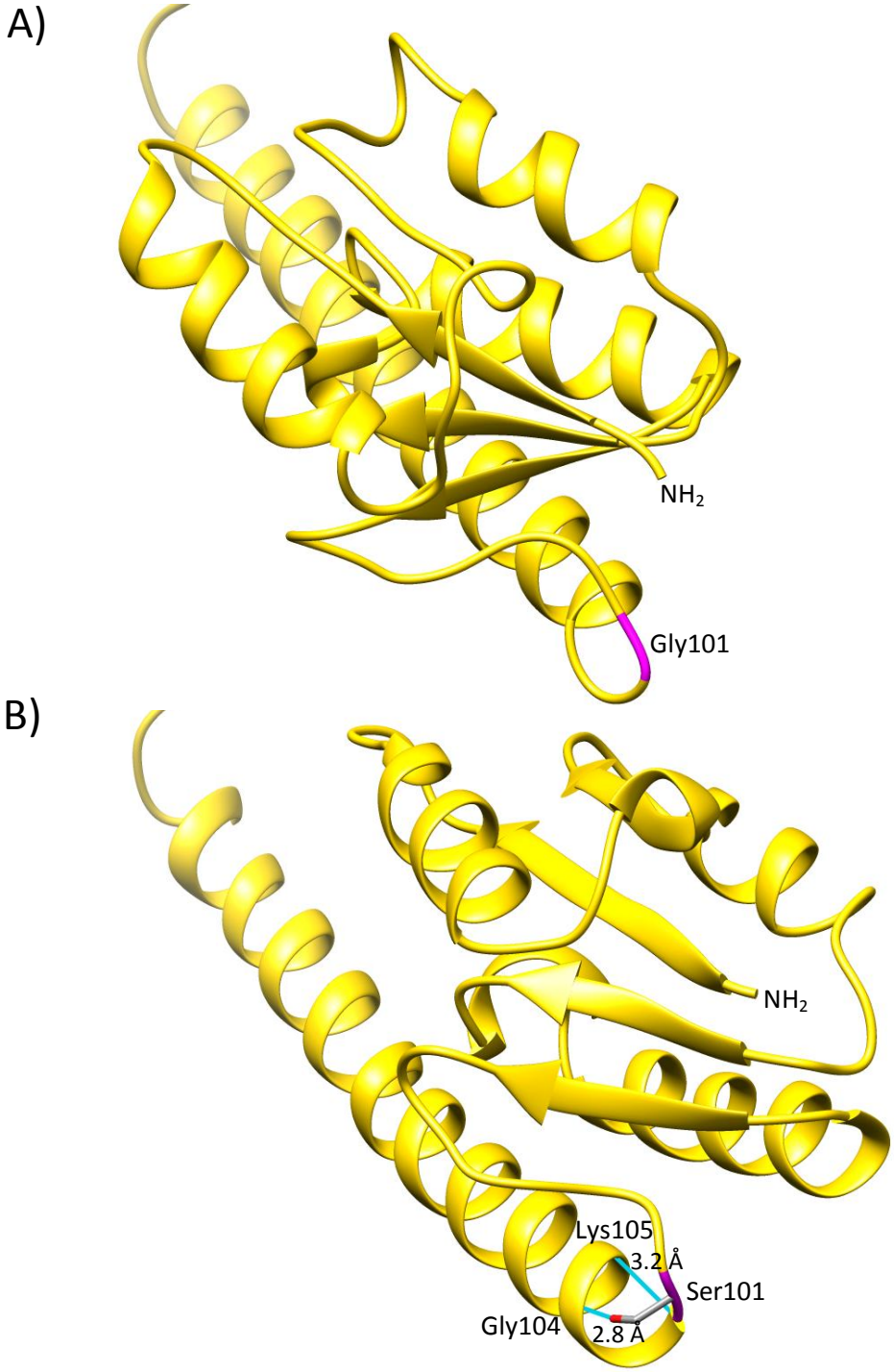


Figure 4-21: G101S stabilizes unlocked conformation. A) The panel shows a 1GDT subunit in the locked conformation, with Gly101 highlighted in magenta. **B)** The panel shows a 1ZR4 subunit in the unlocked conformation, with Ser101 (G101S) highlighted in magenta. Compared with the E-helix of the 1GDT subunit, the E-helix of the 1ZR4 subunit has been slightly extended by the hydrogen bond from Ser101 which stabilizes and extra loop of the helix. This conformational change in the E-helix appears to support the change in position of the α/β sub-domain relative to the E-helix seen between the 1GDT subunit and the 1ZR4 subunit. The Ser101 side chain makes a 2.8 Å hydrogen bond contact to the nitrogen in the backbone of Gly104, while the backbone oxygen of the Ser101 makes a 3.2 Å hydrogen bond contact to the nitrogen in the backbone of Lys105 (side chains and backbone atoms of these contacts are not displayed). The HTH domains are not shown in the figure.

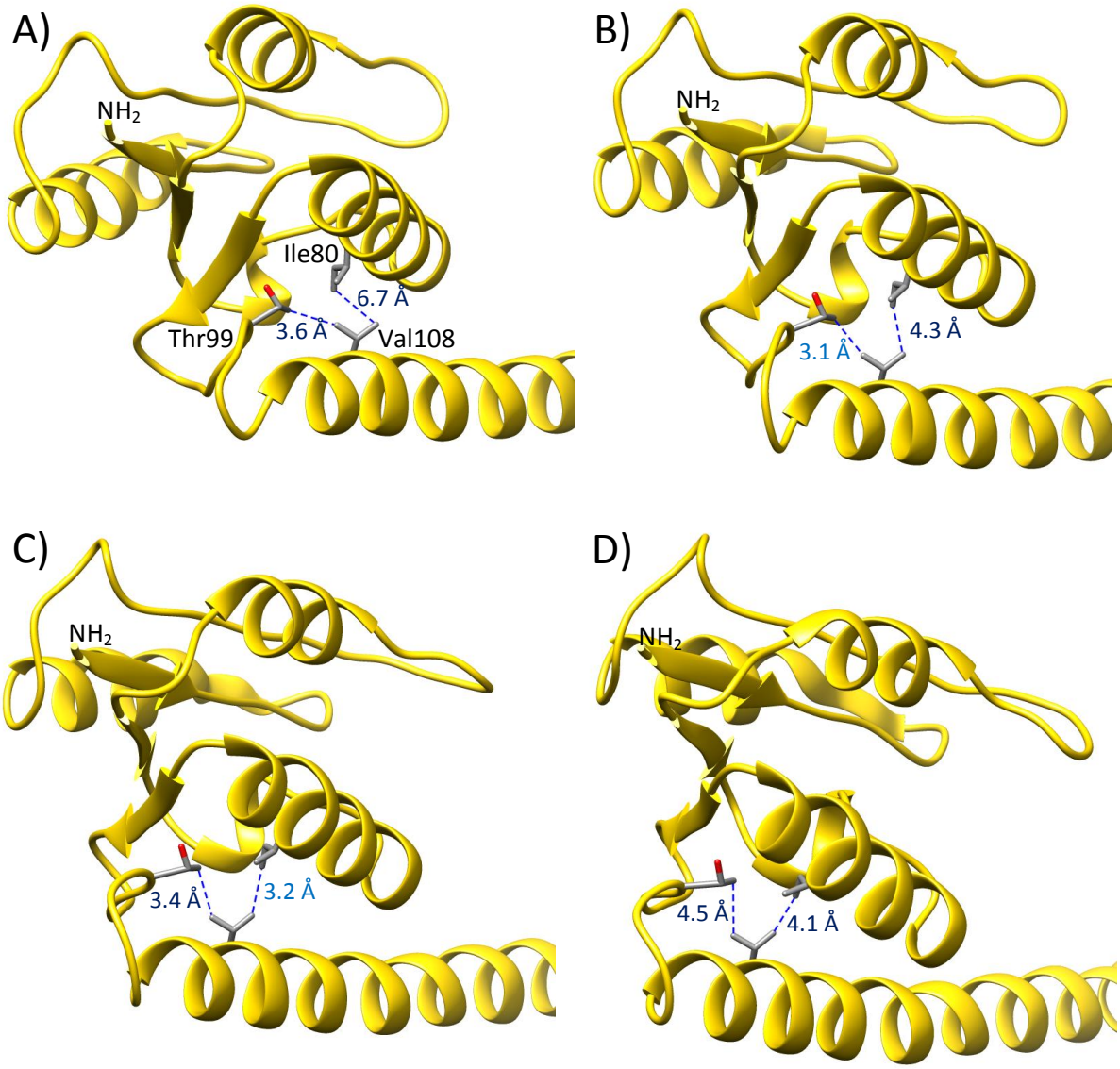


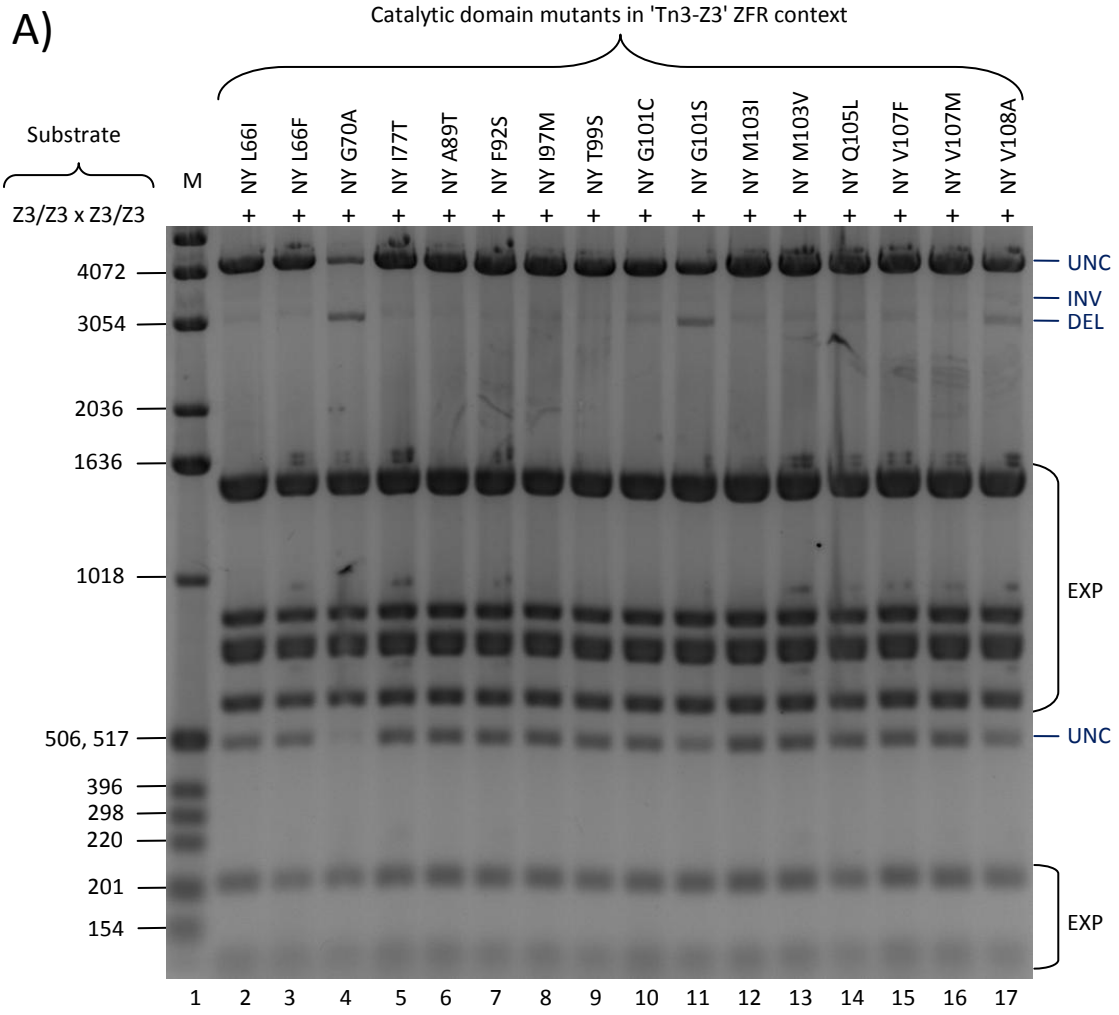
Figure 4-22: Potential clashes between Val108, and Ile80 and Thr99 during conformational transformation. The figure shows a morph conformational interpolation between a 1GDT subunit conformation and a 1ZR4 subunit conformation. All panels display the resolvase subunit in identical orientation relative to the position of the E-helix. The DNA and HTH domain, along with some of the E-helix, is not shown. **A)** The panel displays the 1GDT subunit, which is the beginning of the conformational transition from locked to unlocked state. The distances between the methyl groups of Ile80 and Val108, and Thr99 and Val108 are indicated, and these amino acid are labelled. **B)** The panel shows a transitional conformational state (modelled using Chimera's morph interpolation tool) between a 1GDT subunit conformation and a 1ZR4 subunit conformation, where the methyl group of Thr99 is 3.1 Å from a methyl group of Val108. **C)** The panel shows a transitional conformational between a 1GDT subunit conformation and a 1ZR4 subunit conformation, where the methyl group of Ile80 is 3.1 Å from a methyl group of Val108. **D)** The panel shows a 1ZR4 subunit which represents the end of the conformational transition.

4.3.5.5 DEV repeat of the experiment

The activity screen of the mutants was performed again using DEV-based double expression levels, in order to attempt to increase the recombination activity, and thus, increase the sensitivity of the experiment for detecting differences in activation level imparted by the various mutations. Tn3[NY G70C]-Z3 and Tn3[NM]-Z3 were not included in this experiment because they had already produced complete conversion of substrate to recombination product in the previous single expression vector assay. The negative controls, Tn3[N G70C]-Z3 and Tn3[NY]-Z3, were also dispensed with, as it was expected that, given the results of the previous experiment, there would be many mutants in the set that would effectively act as negative controls.

4.3.5.6 Results

There was almost no change observed in the results (shown in Figure 4-23) relative to the previous experiment, with only the Tn3[NY **G70A**]-Z3 ZFR showing an appreciably increased level of activity (Figure 4-23 A, lane 4). The results of the Tn3[NY **G70A**]-Z3 ZFR reaction also produced the familiar SA effect, in contrast to the inactive mutants that did not produce this effect. This result confirms the conclusion drawn from the results in Section 4.3.4.3 that the SA effect requires active catalytic domains in order to occur, likely indicating that the SA effect is the product of unregulated cleavage of the DNA, rather than being a product of ZFR binding itself.


A)
Catalytic domain mutants in 'Tn3-Z3' ZFR context
Substrate
Z3/Z3 x Z3/Z3
M
NY L66I
NY L66F
NY G70A
NY I77T
NY A89T
NY F92S
NY I97M
NY T99S
NY G101C
NY G101S
NY M103I
NY M103V
NY Q105L
NY V107F
NY V107M
NY V108A
+
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNC
INV
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

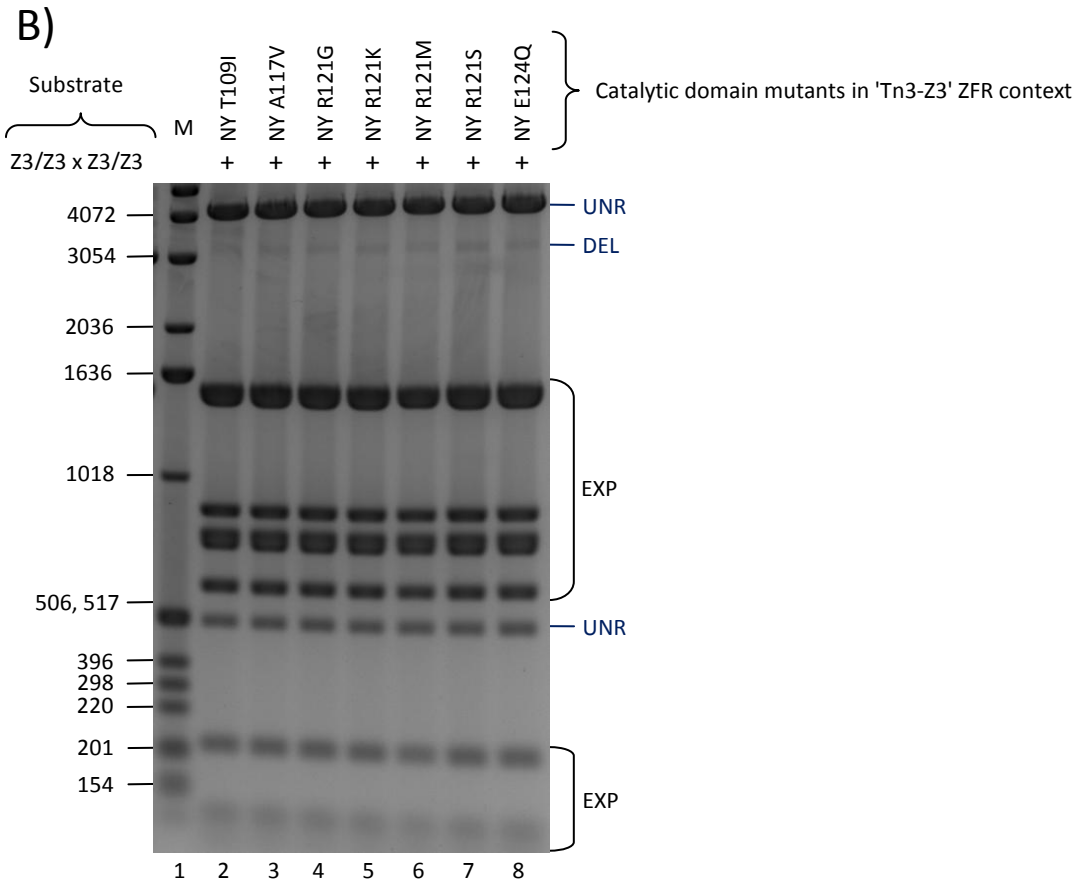

B)
Substrate
Z3/Z3 x Z3/Z3
M
NY T109I
NY A117V
NY R121G
NY R121K
NY R121M
NY R121S
NY E124Q
+
Catalytic domain mutants in 'Tn3-Z3' ZFR context
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNR
DEL
EXP
UNR
EXP
1 2 3 4 5 6 7 8

Figure 4-23: DEV-based activity level screen of other potential locking interaction knock-down mutants in the ZFR system. The results of a 17 Hour Recombination Assay are shown spread across two gels, A) and B). **DEVs were used to increase the concentration of ZFRs in the cells.** To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions.

4.3.6 Testing new sleepy + active ZFR pairs

4.3.6.1 Rationale for the experiment

As outlined in Section 4.3.5.1, it would be desirable to have a sleepy ZFR with increased activation in a sleepy + active ZFR complementation, but which is not active when used alone. It would also be desirable to have an active ZFR that had a lower tendency to produce the SA effect. Additionally, if two sleepy ZFRs could be used in concert to produce activation, this would also be desirable. Therefore, an experiment was designed to test whether the new ZFR mutants, which were previously assayed for base-level activity, could increase the reaction rate of sleepy + active ZFR complementations. In addition, because most the mutations in the previous base-level activity experiment are proposed to operate through a locking interface knockdown, some inferences can be made from the following experiment about that model. In particular, if a robust locking interface knockdown mutation, such as G70C, is paired in a co-localized arrangement (see Figure 4-19) with another locking interface knockdown mutation that operates in the same region of the interface, the second mutation might be expected to produce no additional activity due to redundancy.

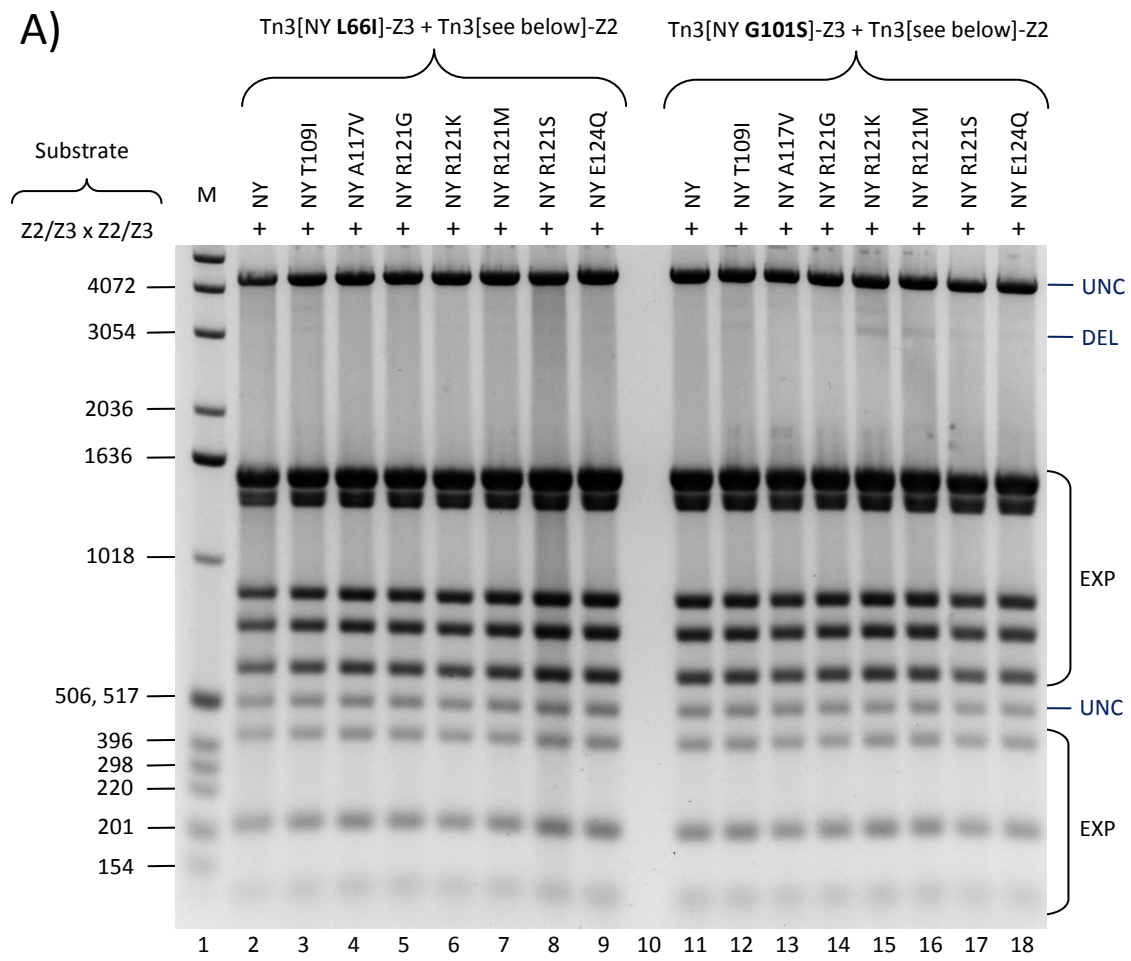
4.3.6.2 Experiment design

The secondary mutations used in the previous experiment can be categorized into those that may operate by interfering with the locking interface from the E-helix side or α/β sub-domain side of the interaction (see Table 4-4). For this experiment, *select* ZFRs with secondary mutations that may knockdown the locking interaction from the α/β sub-domain side of the interface, were paired with *all* of the ZFRs containing secondary mutations that may operate by knocking down the locking interaction from the E-helix side of the interface. In order to test whether two sleepy ZFRs can be used in concert to achieve activity, Tn3[NY **L66I**]-Z3 was used as the α/β sub-domain side knockdown ZFR. In

order to test whether a lower-activity active ZFR, which may have less tendency to produce the SA effect, can be used to achieve activity in a sleepy + active ZFR complementation, Tn3[NY **G101S**]-Z3 was selected as the α/β sub-domain side knockdown ZFR. In order to test whether sleepy mutants carrying a secondary mutation might be able to increase the activity in a sleepy + active ZFR complementation, Tn3[NY **G70C**]-Z3 was used as the as the α/β sub-domain side knockdown ZFR. In order to reduce the volume of work for this experiment, only a 2MutHetDim-DR substrate was used to determine whether complementations were significantly active. The experiment was carried out using the 17 Hour Recombination Assay.

4.3.6.3 Results

The results of this experiment are shown in Figure 4-24. The results of the complementations between the sleepy + unlocked ZFR pairs on the 2MutHetDim-DR substrate show: no noteworthy recombination activity for any of the complementations involving the Tn3[NY **L66I**]-Z3 sleepy ZFR (paired with other sleepy ZFRs); faint levels of recombination activity for complementations using the Tn3[NY **G101S**]-Z3 active ZFR + Tn3[NY **R121K**]-Z2, Tn3[NY **R121M**]-Z2, and Tn3[NY **R121S**]-Z2 sleepy ZFRs (Figure 4-24 A, lanes 15, 16, and 17, respectively); and, importantly, a substantial conversion of substrate to recombination product for the Tn3[NY **G70C**]-Z3 active ZFR + Tn3[NY **T109I**]-Z2 sleepy ZFR pair (Figure 4-24 B, lane 3). The Tn3[NY **A117V**]-Z2 and Tn3[NY **E124Q**]-Z2 sleepy ZFRs in complementation with the Tn3[NY **G70C**]-Z3 active ZFR (Figure 4-24 B, lanes 4 and 9, respectively), also showed a weak level of recombination activity. However, this recombination activity was not above the level of recombination seen for the complementation involving the Tn3[NY]-Z2 sleepy ZFR (Figure 4-24 B, lane 2), which does not possess any secondary unlocking mutation at all. Trace levels of activity can be seen in several additional lanes when the original image file is viewed on a computer screen.


A)
Tn3[NY **L66I**]-Z3 + Tn3[see below]-Z2
Tn3[NY **G101S**]-Z3 + Tn3[see below]-Z2
Substrate
Z2/Z3 x Z2/Z3
M
NY
NY T109I
NY A117V
NY R121G
NY R121K
NY R121M
NY R121S
NY E124Q
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNC
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

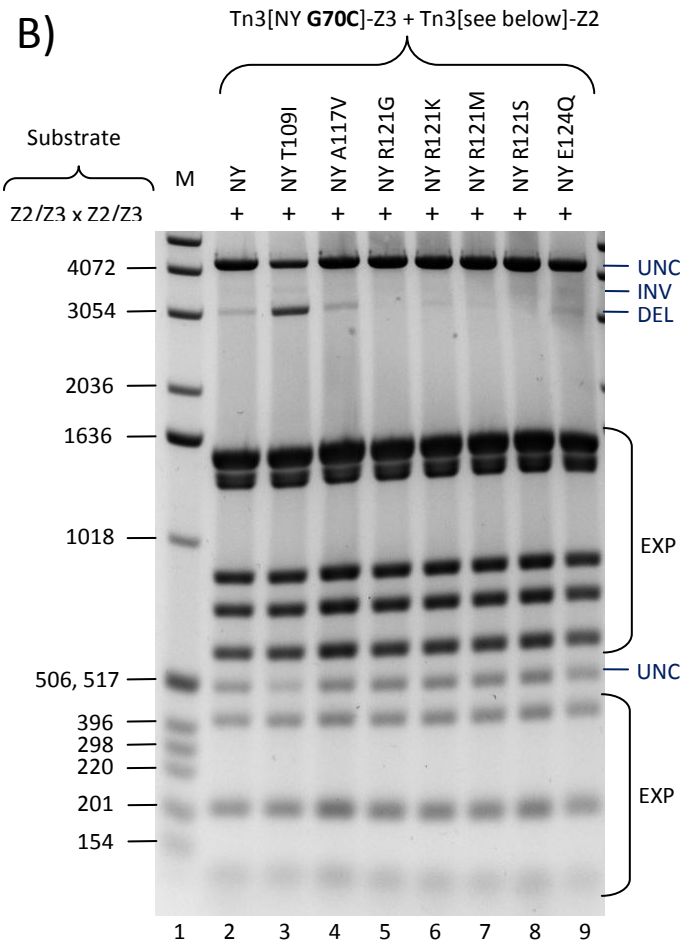

B)
Tn3[NY **G70C**]-Z3 + Tn3[see below]-Z2
Substrate
Z2/Z3 x Z2/Z3
M
NY
NY T109I
NY A117V
NY R121G
NY R121K
NY R121M
NY R121S
NY E124Q
4072
3054
2036
1636
1018
506, 517
396
298
220
201
154
UNC
INV
DEL
EXP
UNC
EXP
1 2 3 4 5 6 7 8 9

Figure 4-24: Synergistic knock-down of both sides of the locking interaction. The results of a 17 Hour Recombination Assay are shown spread across two gels, A) and B). To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions.

One interesting aspect of the results is that Tn3[NY **R121K**]-Z2 and Tn3[NY **R121M**]-Z2 had more activity when paired with Tn3[NY **G101S**]-Z3 than they did when paired with Tn3[NY **G70C**]-Z3. This result supports the locking interface activation model because the expectation was that if the model was correct, the co-localization of a robust knockdown mutation, such as G70C, with another interface knockdown mutation operating in the same region from the other side of the interface, would produce no additional activating effect because it is redundant. If these ZFRs pairs produced activity in a distributed arrangement, this would lead to a desirable bias in recombination reaction directionality, which might allow the stable integration reactions to be performed (see Figure 4-19). This differential activity profile can occur because co-localization pairs locking interface mutations redundantly on one side of the dimer, while leaving the other side of the dimer locked. Conversely, a distributed arrangement, using the same ZFR pair, places activating mutations in locking interfaces on both sides of the dimer, and thus this arrangement may be more active. Essentially, the co-localization arrangement may cancel out the activity contribution from two one of the ZFRs in the pair, while the distributed arrangement benefits from the activity contribution provided by both ZFRs.

Even more interesting is that the Tn3[NY **G70C**]-Z3 + Tn3[NY **R121G**]-Z2, Tn3[NY **R121K**]-Z2, Tn3[NY **R121M**]-Z2, Tn3[NY **R121S**]-Z2, and Tn3[NY **E124Q**]-Z2 complementations (Panel B, lanes 5–9), produce no activity, while the Tn3[NY **G70C**]-Z3 + Tn3[NY]-Z2 complementation produced minimal recombination activity. This implies that the inclusion of secondary activating mutation at position 121 within the sleepy ZFR actually reduced the recombination activity of the reaction. Since Tn3[NY **R121K**]-Z2 and Tn3[NY **R121M**]-Z2 did increase recombination activity over the use of Tn3[NY]-Z2, when paired with Tn3[NY **G101S**]-Z3, it can be concluded that the reduction in activity seen when paired with Tn3[NY **G70C**]-Z3 is unique to the complementation with Tn3[NY **G70C**]-Z3. One possible explanation for this effect, is that knocking out the locking interface

completely (as opposed to partial disruption), deactivates the enzyme. Since the locking interface comprises the major portion of the dimer interface, it may be the case that completely knocking out this region prevents dimers from forming, or prevent dimers from forming properly. Data from Burke et al. (2004) may also support this possibility. In Burke et al.'s experiments, several E-helix side locking interface mutations (A117V, R121K, and E124Q) that produced a high level of recombination activity when used as the sole secondary mutation, produced an inactive resolvase when combined together. If co-localization of locking interface mutations can be used to abolish the activity of the ZFR, then a high fidelity recombination reaction directionality bias should be possible by arranging ZFR mutants such that an active AHD configuration results in an inactive RHD configuration.

Unfortunately, there was insufficient time available to carry out a follow-up experiment to further test the foregoing approaches to achieving a recombination reaction directionality bias. There was also insufficient time to carry out further experiments to test whether a combination of secondary mutations, such as G101S and V108A would be able to produce an active ZFR that is less prone to the SA effect (see Section 4.3.5.4).

4.4 Residue 102 and E-helix 'landing pad'

4.4.1 Introduction

4.4.1.1 The residue at position 102 is the lead contact during catalytic dimer-dimer synapsis

The results of Section 4.2 demonstrated that the activating effect provided by primary mutations (e.g. D102Y) is likely not caused by the removal of a critical ionic energy barrier between counterpart Asp102 residues across the initial dimer-dimer interface. Therefore, the question remains: why are the primary activating mutations so important for hyperactivation of Tn3 resolvase? All of the possible residue substitutions at position 102 have been attempted within the Stark lab, and the primary mutations capable of producing hyperactivation of Tn3 resolvase are as follows Y, F > I > V > T > W (ranked in order of the level of activating effect they provide; Burke *et al.* (2004) and unpublished work within the Stark lab). The first thing that can be noted is that all of the primary activating mutations are hydrophobic, with the exception of threonine—the only

polar residue with a methyl group in its side chain. Additionally, almost all of the possible hydrophobic amino acid substitutions are capable of acting as primary activating mutations. Therefore, it would appear that hydrophobicity at position 102 is correlated with activation.

During the initial stage of tetramer formation, the N-terminal ends of the E-helices of the dimers must make contact with each other in order to form the antiparallel interactions between the E-helices seen in the 1ZR4 activated tetramer structure (Figure 1-13 and Figure 1-12; (Li et al., 2005)). All of the activating mutations located on the face of the N-terminal end of the E-helix that presumably make this initial contact, are hydrophobic substitutions (e.g. D102Y, M103I, Q105L, T109I; see Figure 4-25). An explanation for the activating effect of these hydrophobic mutations is that they enable favourable hydrophobic contacts between opposing E-helices during the initial stage of dimer-dimer interaction. It should also be noted, that in spite of the results of Section 4.2, it is still possible that favourable hydrophobic interactions between counterpart 102 residues plays a role in activation.

The second thing that may be considered, is that in order for the tetramer to form, the E-helices, with residue 102 at the lead, must pass through a hydrophobic region that is occupied by the contacts of the locking interface in the pre-tetramer configuration of the dimer (Figure 4-4, Figure 4-8, and Figure 4-25 A). It is possible that mutating the lead residues of the E-helix to hydrophobic residues, allows the E-helix to more successfully displaces these hydrophobic locking region interactions during tetramer formation by replacing one set of hydrophobic contacts for another, perhaps stimulating the conformational change that leads to activation. This hypothesis of the E-helix interaction is functionally related to the locking interface unlocking model, as the locking interface knockdown mutations would also facilitate this same conformational transition, which allows the α/β sub-domain to move as the E-helix from an opposing subunit takes its place. It may also be noted that in the case of D102Y, modelling using the UCSF Chimera and the Dumbrack rotamer library (Dunbrack, 2002) indicated that the hydroxyl group of the tyrosine side chain at position 102 is most likely to face away from the hydrophobic contact region D102Y potentially makes contact with (model not shown).

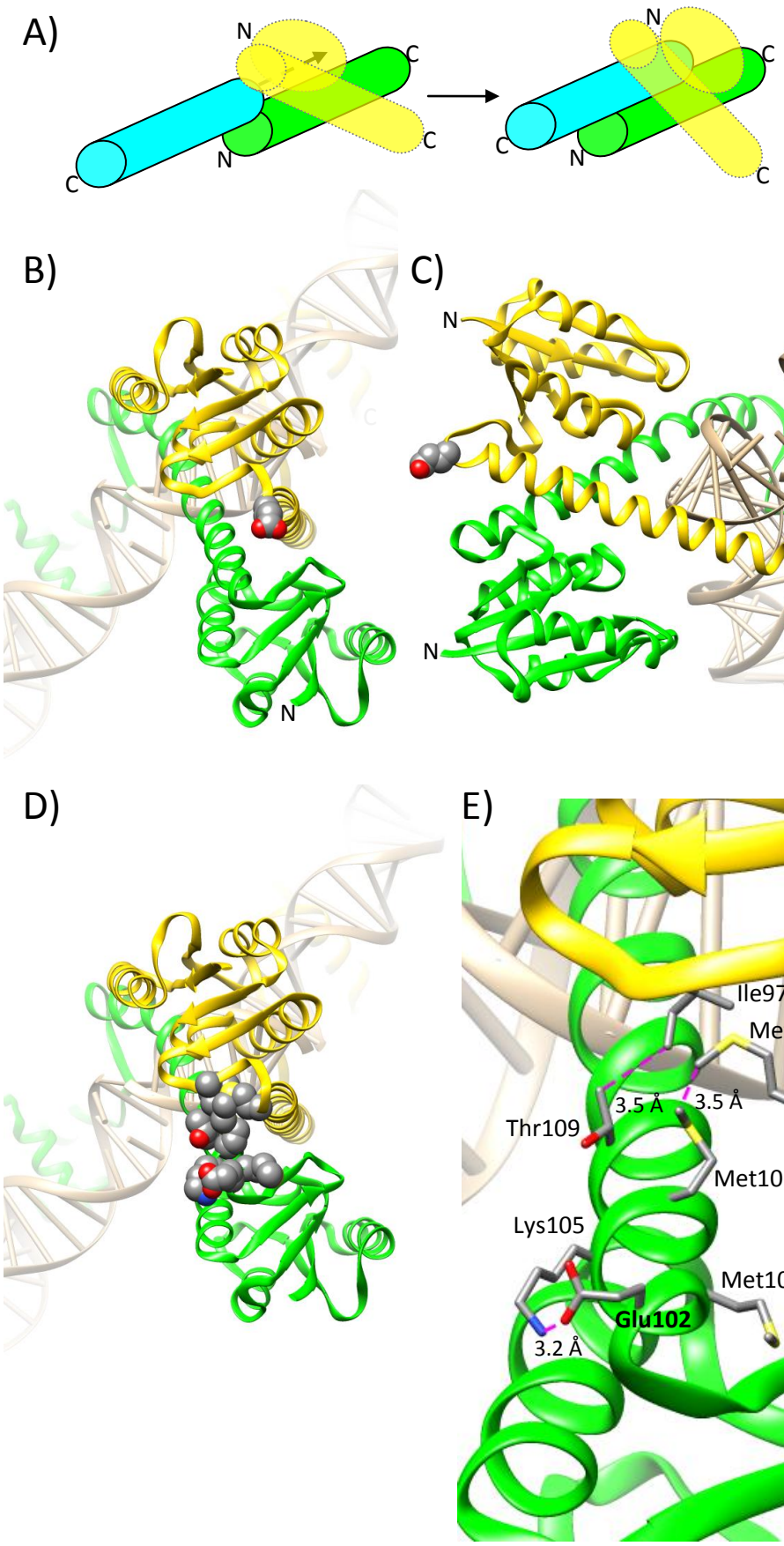

A)
N
C
C
N
C
N
C
C
N
B)
N
C)
N
N
C
D)
E)
Ile97
Met103
3.5 Å
3.5 Å
Thr109
Met106
Lys105
Met103
Glu102
3.2 Å

Figure 4-25: Early contacts during dimer-dimer synapsis. A) The panel shows the anti-parallel contacts being formed between two E-helices (cyan and green cylinders) of the catalytic tetramer during synapsis. Only the E-helices of the cyan and green subunits are shown. In addition, the adjacent dimer subunit partner of the green subunit is displayed (transparent yellow) with only its α/β sub-domain (oval) and E-helix shown. The adjacent dimer subunit partner of the cyan is not displayed at all for clarity of the image. The cyan E-helix makes contact with an opposing green E-helix in the dimer-dimer interface. The space the cyan E-helix must pass through is initially blocked by the α/β sub-domain of the dimer partner of the green subunit, the yellow subunit. As the initial dimer-dimer interface complex transitions to the catalytic tetramer complex, the α/β sub-domain of the yellow subunit translocates around the E-helix of the yellow subunit, freeing the space and allowing the E-helix of the cyan subunit to form its anti-parallel interaction with the E-helix of the green subunit. **B)** The panel shows the protrusion of the lead residue of the E-helix at position 102. In the 1GDT γδ resolvase structure shown the Glu102 of the yellow subunit is shown in space-filling mode. **C)** The panel shows the same image as in B) but rotated 90° for a different perspective. **D)** The panel shows the 'landing pad' region which the residue at position 102 of an opposing E-helix in the dimer-dimer interface must negotiate in order to achieve the anti-parallel E-helix-E-helix interface of the catalytic tetramer. The seven residues which make up the landing pad (Green subunit: Glu102, Met103, Lys105, Met106, and Thr109, and green subunit Ile97, and Met103) are shown in space-filling mode. **E)** This panel shows the same region as in D) but enlarged and with the seven landing pad residues labelled and shown in stick mode. It should be noted that regarding the image displayed in panels B) and C), that in the green subunit of the 1GDT structure, which has an unlocked α/β sub-domain, Glu102 is in a somewhat different position (not shown) than that of the yellow subunit (shown). However, even in the green subunit, Glu102 is still the leading residue on the face of the helix which must make the E-helix-E-helix dimer-dimer contacts.

The potentially unfavourable interaction between Tn3 resolvase wild-type residue, aspartic acid (or glutamic acid in the case of γδ resolvase), during initial contact with the hydrophobic region on the opposing E-helix (Figure 4-4, Figure 4-8, and Figure 4-25 A), might serve the wild-type enzyme as an energy barrier that prevents activation of the enzyme outside of proper synaptosome formation. Energy to overcome this potential energy barrier would presumably be provided by the synaptosome in the wild-type system. This explanation builds a connection between the activity of the primary mutations (at position 102) and the activity of the secondary activating mutations (most of which potentially affect the locking interaction), which may explain why they appear to work synergistically to activate the enzyme. The secondary mutations may make it easier for the locking interaction to be disrupted, but only if the lead residue of the E-helix is predisposed to favourable interaction with the hydrophobic portion of the locking region.

The activating mutation G101S should also be mentioned with respect to its potential effect on this hypothetical activity of D102Y negotiating the hydrophobic region of an opposing E-helix. Although G101S may produce its activating effect by stabilizing the unlocked conformation (see Figure 4-21), it is also true that this conformational change, which slightly extends the E-helix, also results in changing the position and orientation of the residue at position 102. Therefore, in addition to potentially stabilizing the unlocked

conformation, G101S might also promote more favourable contact between the D102Y and the hydrophobic region on the opposing E-helix.

Unfortunately, during this stage in the experimental work, there was insufficient time to meticulously test the hypothesis concerning the activity of the residue at position 102 and its contact with the opposing E-helix during the initial stage of tetramer formation. However, rudimentary experiment based on this hypothesis was possible. It was considered that this type of interaction might be exploited to produce the sought after biases in tetramer configuration activity, which might produce dimer-dimer orientation specificity and/or recombination reaction directionality. Therefore, this premise was considered when designing the current experiment.

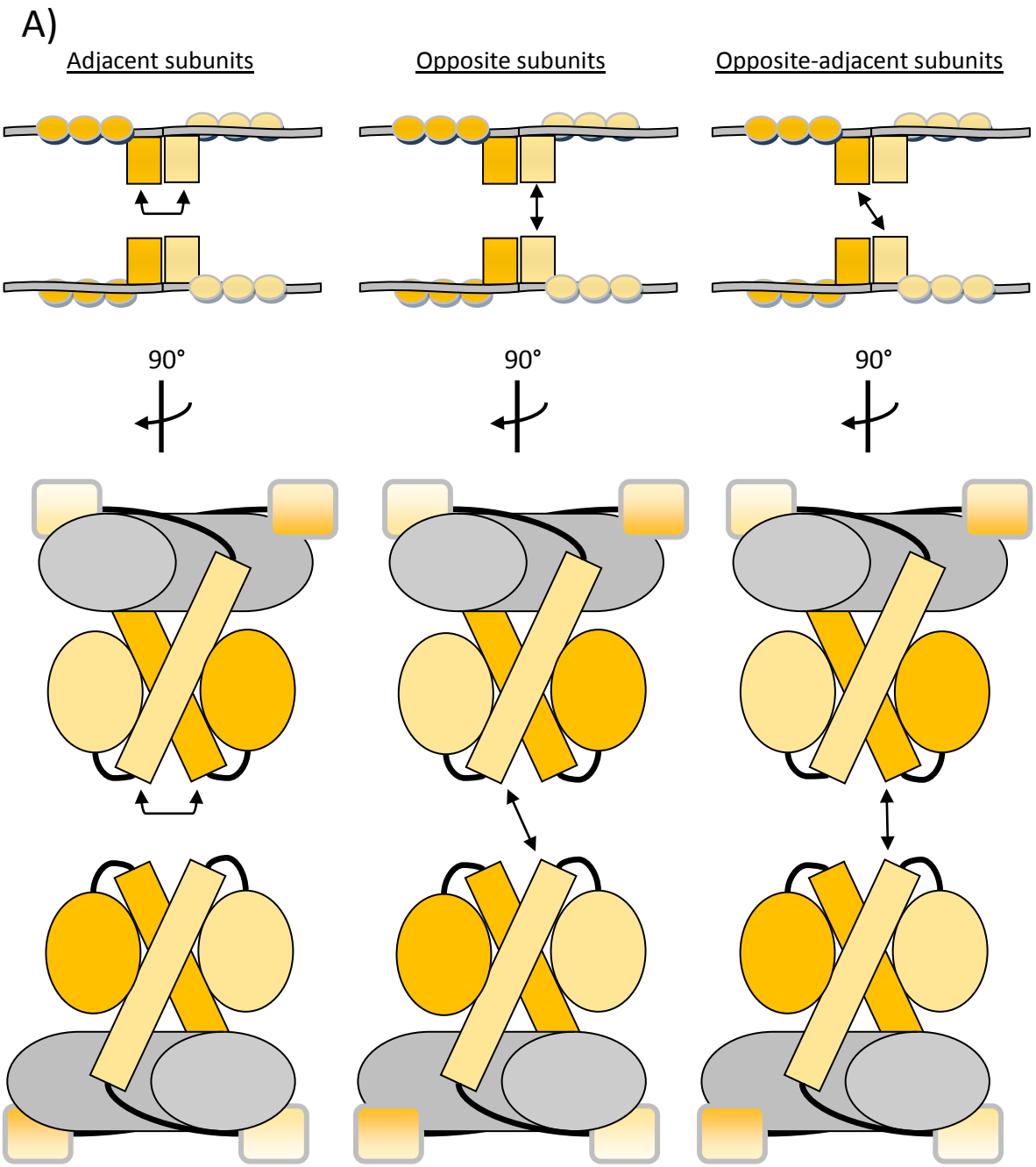
4.4.1.2 Residue 102 and the 'landing pad'

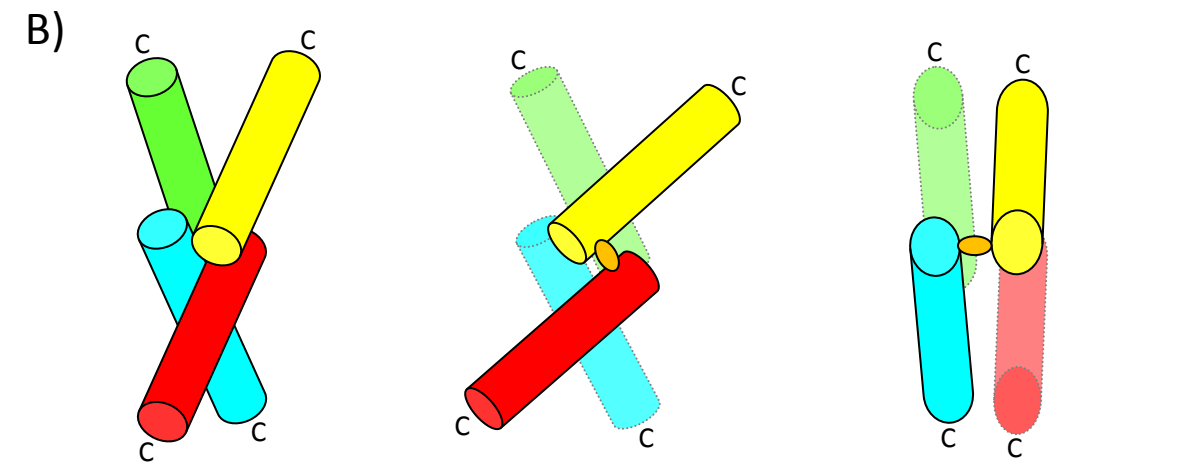
A region at the N-terminal of the E-helix, termed here as the 'landing pad', can be proposed as the initial contact interface between residue 102 and opposing E-helix it must contact during catalytic dimer-dimer synapsis (Figure 4-25 D and E). The proposal of this region as the initial contact point during synapsis is consistent with the modelling shown in Figure 4-10, and the dimer-dimer synapsis trajectory proposed by Li *et al.*, (2005). Of the seven residues that make up the landing pad, five are locations of known activating mutations (D102X,[7] M103I/V, Q105L, and T109I; note that two Met103 residues on adjacent subunits participate in the same landing pad, see Figure 4-25 D and E). All of these activating mutations replace polar residues with hydrophobic ones, with the exception of the substitutions at position 103, which can be said to increase the hydrophobicity at that position. Therefore, all of these secondary activating mutations could be proposed to produce their effect by creating a more favourable environment for the hydrophobic primary activating mutation residue to negotiate during the initial stage of catalytic dimer-dimer synapsis. It should also be considered that the dimers, and thus the E-helix N-terminal regions, likely do not initially contact with each in a perfectly symmetrical fashion. Therefore, increasing the general hydrophobicity at the N-terminal ends of the E-helices may help allow this initially imprecise contact to stabilise the interaction between the E-helices, facilitating a transition to a more bona fide initial synaptic state.

[7] 'X' represents the lists of activating substitutions at position 102 given in Section 4.4.1.1.

Aside from the potential contacts between residue 102 and the landing pad (Figure 4-25) and hydrophobic portion of the locking region (Figure 4-4), there are also two potential specific contacts that my take place between counterpart residues of the landing during dimer-dimer formation. Molecular modelling using the 1GDT dimer structure suggests that it is highly probable that opposite-adjacent counterpart residues at position 103 make contact with each other (the molecular model not shown, but a diagram depicting this interaction is displayed in Figure 4-26). Therefore, it is possible that the activating effect provided by mutations such as M103I is the result of either a favourable interaction between counterpart 103 residues, or the removal of an unfavourable interaction between counterpart 103 residues. Additionally, it is possible that opposite counterpart residues at position 105 also make contact with each other during an early stage of dimer-dimer interaction (Figure 4-26). Therefore, Q105L might also produce its effect by either generating a favourable counterpart interaction, or removing an unfavourable one. The tetramer configuration activity profile that would be predicted from favourable counterpart residue interaction between M103I or Q105L is as follows: the opposite-adjacent subunit interaction of M103I would be expected to increase the activity of only the RHD configuration, while the opposite subunit interaction of Q105L would be expected to increase the activity of only the MHD configuration.

It should also be noted that because of their participation in the locking interaction, alternative explanations exist for the activating effect of several secondary mutations of the landing pad. Mutations at position 103 might break a potential hydrophobic locking interface contact (Figure 4-25 E), which could destabilize the locked conformation. Q105L may help facilitate conformational transformation from the locked to unlocked dimer state, by making more favourable interactions during the transition (UCSF Chimera morph model not shown). Finally, T109I might produce its effect by destabilizing a potential hydrophobic contact between Thr109 and Ile97 that may help to stabilize the locking interaction (Figure 4-25 E).





Dimer-dimer E-helices pairing | Q105L-Q105L | M103I/V-M103I/V

Figure 4-26: E-helices dimer-dimer contact descriptions and mutations. A) This panel shows paired dimers, and provides a visual depiction of the subunit contact descriptions which are given above the diagrams. **B)** This panel shows hypothetical anti-parallel E-helix-E-helix contacts at the initial dimer-dimer interface of the catalytic tetramer. At the left is provided a reference diagram in order to make the images to the right easier to interpret. At the right and in the middle are diagrams showing potential self-complementary interactions (orange oval) between residues of the mutations given below.

Finally, it because the landing pad mutations may help stabilize or facilitate the E-helix–E-helix interactions that initiate the unlocking of the α/β sub-domain from the E-helix, the landing pad mutations located on the E-helix may be functionally equivalent to E-helix-side interface unlocking mutations that also promote the unlocking transformation from the E-helix side of the interaction. Q105L and G101S may be unique in this sense, as they might either act as α/β sub-domain-side activators or E-helix-side activators.

4.4.2 Experiment design

4.4.2.1 Experimental variables

In order to test whether the interaction between position of 102 of the E-helix and the landing pad of an opposing E-helix may be used to produce a dimer-dimer orientation specificity bias, or recombination reaction directionality bias, a new sleepy + active ZFR complementation experiment was designed. This experiment operated over three variable conditions: the presence or absence of a primary mutation in the absence of activating landing pad mutations, the presence or absence of a primary mutation in the presence of activating landing pad mutations, and the presence or absence of activating landing pad mutations in the presence of a primary mutation. In all cases D102Y was used as the primary mutation, while G70C was used as the secondary mutation under the experimental condition where no landing pad activating mutations were present. However the options for manipulating the landing pad mutation variable were limited by both the activity profiles of these mutations, and the time available to conduct the experiment, and thus, the mutation collection known as 'M' was used to test this variable (see below).

4.4.2.2 Rationale for using M as the landing pad variable

The data collected in Section 4.3.5 (Figure 4-20 and Figure 4-23) demonstrate that none of the landing pad activating mutations are sufficient to impart the ZFR with significant base-level activity when used as the sole secondary mutation. Therefore, in order to test the effect of the landing pad activating mutations, they must be incorporated into a background with other secondary activating mutations in order to produce ZFRs with sufficient activity. Ideally each landing pad activating mutation might have been paired alone with one or more secondary mutations sufficient to activate the ZFR. At the point where the work for this segment of the project was undertaken, there was insufficient time available to determine what secondary mutation background might be sufficient to activate mutants containing single landing pad activating mutations. Even G70C, for instance, does not produce a robust level of recombination activity when paired with sleepy Tn3[NY] ZFR. This experiment utilizes not only a Tn3[NY] sleepy ZFR but also a Tn3[N] sleepy ZFR, which was predicted to impart even less of an activity contribution than Tn3[NY] to the sleepy + active ZFR complementations. However, the mutation collection known as M (G101S, D102Y, M103I, and Q105L) was known to produce high-level baseline activity (Figure 4-20), and also contains all of the known landing pad activating mutations apart from T109I. Therefore, M was used as the independent variable in the experimental conditions where the presence or absence of the landing pad mutations was tested.

4.4.2.3 Substrate selections

In order to detect a dimer-dimer orientation specificity or reaction directionality bias, the 2MutHomDim, 2MutHetDim-DR, and 2MutHetDim-IR substrates were used as in the previous experiments. In order to reduce the volume of work for this experiment, only one of the two varieties of each substrate (e.g. Z2/Z2 x Z3/Z3 and Z3/Z3 x Z2/Z2) was used. It should be noted that the choice of 2MutHetDim-IR substrate variety, relative to the catalytic domain and ZFA combinations that are used, produces a bias when comparing the excision product signal from 2MutHetDim-DR and 2MutHetDim-IR substrates. This bias results because, while the excision product from either variety of 2MutHetDim-DR substrate is subject to an intermediate level of SA effect due to its heterodimer Z-site, the excision product from 2MutHetDim-IR substrates will be subject to ether a strong SA effect, or no SA effect at depending on whether its homodimer Z-site specifies the sleepy or active ZFR. If the homodimer Z-site in the 2MutHetDim-IR excision

product specifies the active ZFR, the SA effect will be strong, but if it specifies sleepy ZFR, then the SA effect will not be present. For this experiment, the Z2/Z3 x Z3/Z2 2MutHetDim-IR substrate was used, which produces an excision product with a homodimer Z-site that specifies the active ZFRs, and thus, is subject to a strong SA effect.

4.4.2.4 The experiment

The sleepy + active ZFR pairs, Tn3[N G70C]-Z2 + Tn3[NY G70C]-Z3, Tn3[N]-Z2 + Tn3[NM]-Z3, and Tn3[NY]-Z2 + Tn3[NM]-Z3, were assayed in the 17 Hour Recombination Assay on the 2MutHomDim (Z2/Z2 x Z3/Z3), 2MutHetDim-DR (Z2/Z3 x Z2/Z3), and 2MutHetDim-IR (Z2/Z3 x Z3/Z2) substrates. As in previous experiments, the production of the various reaction products (unchanged, inversion, and excision) from the various substrates indicates the activity level of the various possible tetramer configurations (AHD, MHD, and RHD). Differential activity of the AHD and RHD configurations indicates the potential for a recombination reaction directionality bias, while differential activity between of the MHD and RHD configurations indicates a dimer-dimer orientation specificity bias. A table is provided below the gel image in the results (Figure 4-27) that indicates the relationships between tetramer configuration activity and reaction products from the various substrates.

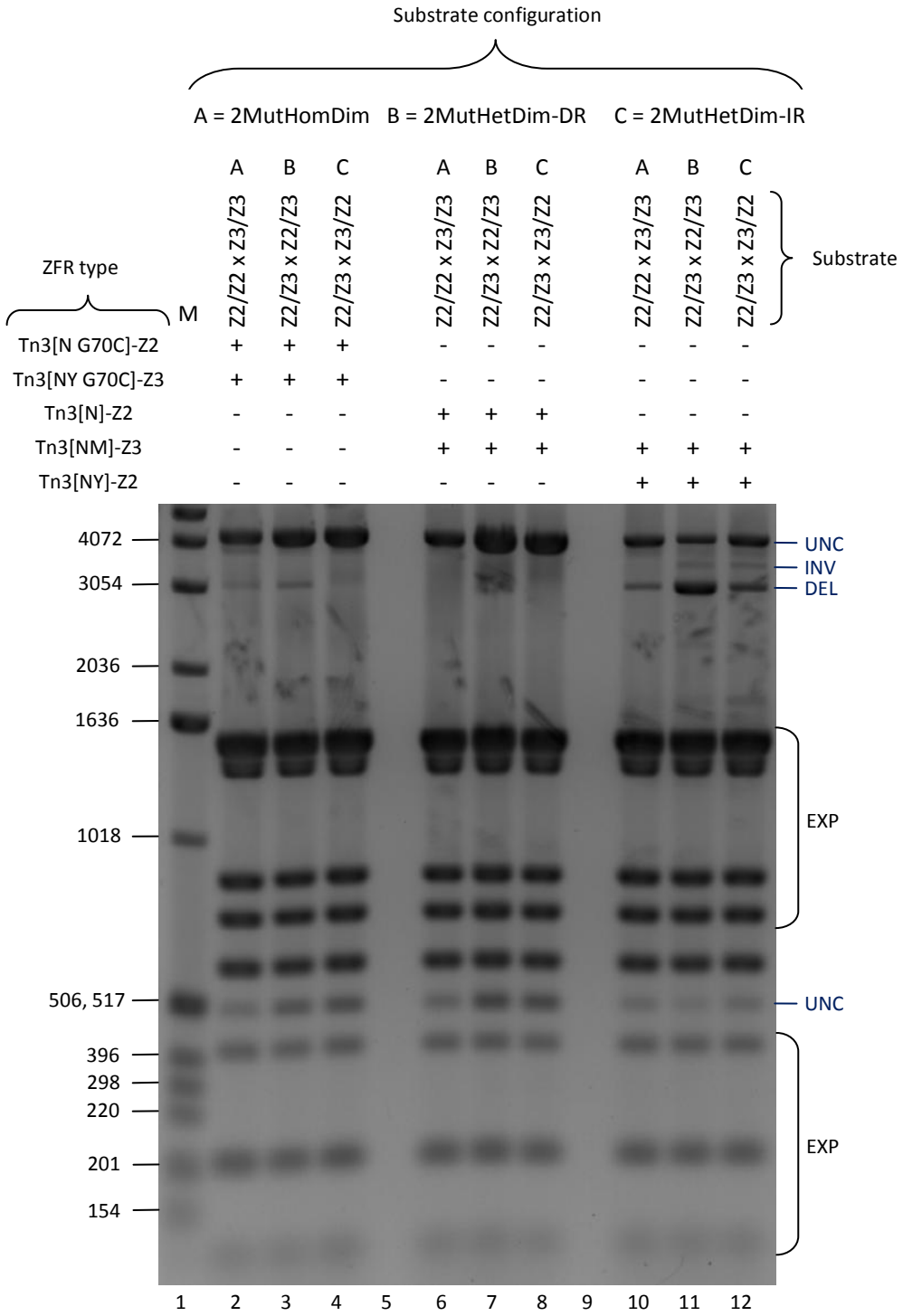
4.4.3 Results

4.4.3.1 Y independent variable in an N G70C background

The results from the complementations between the Tn3[N G70C]-Z2 + Tn3[NY G70C]-Z3 ZFRs (Figure 4-27, lanes 2–4) indicates that this pair is almost inactive. In fact, these results are very similar to those seen for the Tn3[NY]-Z2 + Tn3[NY G70C]-Z3 pair used in the Section 4.3.3 preliminary experiment (for which the results are not shown) and the comparable set of reactions in 4.3.4.3 (Figure 4-17, lanes 5, 9, and 13). Although the signals are very weak, it can be observed that the signal for the 2MutHetDim-DR excision product (lane 3) is stronger than that for the 2MutHomDim excision product (lane 2), indicating the MHD configuration may be slightly more active than the AHD configuration. There is no excision product visible for the 2MutHetDim-IR substrate (lane 4), but this is likely the result of the SA effect. The SA effect can also be observed when comparing the unchanged substrate bands between the three lanes, with the expected greater level SA effect occurring in lane 2.

4.4.3.2 M independent variable in a N background

The complementation between the Tn3[N]-Z2 + Tn3[NM]-Z3 ZFR pair (lanes 6–8) produced no recombination activity at all, but the SA effect can be seen for the unchanged substrate signal from the 2MutHomDim substrate (lane 6) when comparing the signal strength to the unchanged substrate signal from the 2MutHetDim reactions (lanes 7 and 8). This result provides important information for the next set of results by indicating the relative SA effect production by Tn3[NM]-Z3 on plasmids with a homodimer binding site (lane 6), compared to those with heterodimer binding sites (lanes 7 and 8). The signal strength for the unchanged substrate signal from the 2MutHetDim-DR (lane 7) substrate is also noticeably stronger than that from the 2MutHetDim-IR substrate (lane 8). However, this difference can be accounted for by an increased overall concentration of the DNA loaded into the lane 7 as deduced by comparing the signal strength of the digested expression plasmid bands (most obvious in the bands just below the 1.6 kb marker).



	2MutHomDim		2MutHetDim-DR		2MutHetDim-IR	
	1°	2°	1°	2°	1°	2°
UNC	RHD		AHD		MHD	
INV	AHD		RHD		MHD	
DEL	AHD	MHD	MHD	AHD	RHD	RHD

1° indicates the primary reactions, while 2° indicates secondary reactions that may take place only if the inversion product has been produced by a primary reaction.

UNC = unchanged substrate, INV = inversion product, DEL = deletion/excision product

Pink fields indicate reactions leading to a product that is most susceptible to the SA effect, while yellow fields indicate reactions leading to product that are only partially susceptible to the SA effect.

Figure 4-27: Differential substrate assay testing the primary mutation (D102Y) and landing pad mutations (M) variables using differentially mutated sleepy + active ZFR pairs. The results of a 17 Hour Recombination Assay are shown. To the left of the gel image, are given the sizes of each band of the molecular marker ladder shown in lane 1. To the right of the image, indicators are given to describe the origin of the bands on the gel (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression vector plasmid). Above the gel image is a table which indicates which ZFRs were used in conjunction with each substrate for the complementation reactions. A table provided underneath the gel image provides a key for understanding the tetramer configuration and reaction outcome relationships. This key includes information about which product bands are susceptible to the SA effect.

4.4.3.3 G101S + M103I + Q105L independent variable in an NY background

The results from the complementations involving the Tn3[NY]-Z2 + Tn3[NM]-Z3 ZFR pair (lanes 10–12) indicate a strong bias in the activity of tetramer conformations. The first thing that should be considered is that the difference in signal strength between the excision product from the 2MutHetDim-DR (lane 11) and 2MutHetDim-IR substrate (lane 12) is likely at least partially the result of the SA effect. The excision product in lane 12 contains a Tn3[NM]-Z3 homodimer binding site, which is expected to be more affected by the SA effect than the excision product in lane 11 that contains a heterodimer binding site. When looking at the size of the SA effect caused by Tn3[NM]-Z3 ZFR in lanes 6 and 8 of the previous set of reactions, it can be seen that this difference is somewhat comparable with the difference seen for the excision products in lanes 11 and 12 for this set of reactions. At the very least, the SA effect precludes drawing any conclusion about the differences in activity between the MHD and RHD configurations from observations of the excision products in lanes 11 and 12).

The largest difference in signal strength for excision products observed in this set of reactions is between those in lane 10 and 11. The difference in signal strength observed between these lanes is definitely not the result of the SA effect, as the excision products in these lanes both contain heterodimer binding sites, and thus, are subject to the same level of SA effect. The difference in excision product signals in lanes 10 and 11 indicates a significant difference in the activity of the AHD and MHD configurations, respectively. This difference is much larger than the difference observed between these configurations in previous sleepy + active ZFR complementations (Section 4.3.3 and 4.3.4.3). Furthermore, since at least some of the difference seen between excision products from the 2MutHetDim-DR (lane 11) and 2MutHetDim-IR (lane 12) substrate reactions is likely the result of the SA effect (and thus, the activities of the MHD and RHD

configuration are probably more equal than they appear from comparison of the excision products), it may be further concluded that there is likely an even more significant difference between the activity of the AHD and RHD configurations than appears from a comparison of the excision products in lanes 10 and 12. A significant difference in activity between the AHD and RHD configurations, such as is implied here, indicates that Tn3[NY]-Z2 + Tn3[NM]-Z3 ZFR pair might be used to produce a recombination reaction directionality bias, which could help stabilize integration reactions.

It should be noted, the conclusion that the RHD configuration is significantly more active than the AHD configuration could be reinforced by using the alternative 2MutHetDim-IR substrate (Z3/Z2 x Z2/Z3 as opposed to Z2/Z3 x Z3/Z2). Using the same sleepy + active ZFR pair, the Z3/Z2 x Z2/Z3 substrate will produce an excision product with a homodimer Z-site that specifies sleepy ZFRs, and thus, no SA effect will appear. However, there would still be a signal bias, now in the opposite direction, as the heterodimer Z-sites of the excision products of the 2MutHomDim and 2MutHetDim-DR substrates will be subject to a modest SA effect, while the excision product this 2MutHetDim-IR substrate will not.

The inversion products of the reactions also provide information about the activity of tetramer configurations, although they are somewhat more difficult to interpret due to the possibility of secondary reactions converting the substrate to other products. While the excision products accumulate as end products, the inversion products do not if secondary reaction can take place, and thus, may be more transient in the reaction. In the case of the 2MutHetDim-DR substrate (lane 11), the inversion products would be expected to accumulate to some degree due to the apparently low activity of the AHD configuration. However, the inversion product from the 2MutHetDim-DR substrate contains homodimer Z-sites, one of which specifies the active ZFR (Tn3[NM]-Z3). Thus, this inversion product is also subject to an intermediate SA effect. The intermediate SA effect might not be expected to completely eliminate this product signal if it were sufficiently strong, but the combination of the SA effect and recombination activity from the AHD configuration makes interpretations of the significance of the absence of this product signal difficult. As with the analysis of the excision product band from the 2MutHetDim-IR substrate (lane 12), while this reaction might suggest there is a difference in activity between the MHD and RHD configuration (implying a dimer-dimer orientation

specificity bias), no conclusions may drawn. In the case of the 2MutHetDim-IR substrate, the RHD and MHD configurations, which catalyse secondary reactions on this inversion product, both appear to be sufficiently active to greatly reduce the strength of this signal. Therefore, the absence of this signal is not surprising, and is consistent with the near absence of inversion products often observed in ZFR reactions that have not gone to completion.

Finally, it is tempting to speculate that the reason the difference between AHD and MHD or RHD configuration was enhanced in this reaction compared with previous sleepy + active ZFR complementations, might be because of the predicted interactions between counterpart M103I and Q105L residues (see Section 4.4.1.2). Neither of these interactions is predicted to an increase the activity of the AHD configuration, while M103I was predicted to increase the activity of the RHD configuration, and Q105L was predicted to increase the activity of the MHD configuration. However, because these mutations were used in combination, the pattern of activity is expected to resemble that already observed for sleepy + active ZFR pairs based on locking interface mutations. Additionally, because of the SA effect, no conclusion may be drawn about a potential difference between the RHD and MHD configurations. Furthermore, M includes a G101S, which is predicted to stabilize the unlocked conformation (Figure 4-21), and both M103I and Q105L, themselves, have alternative explanations for their activity suggesting they could facilitate locking interface disruption or locking interface conformational transformation, respectively (see Section 4.4.1.2) Therefore, it is impossible to know from this experiment whether the enhanced differential activity between the AHD and MHD or RHD configuration was the result of any unique interaction, or was simply the result of a more activated sleepy + active ZFR pair operating through the same mechanism as previous sleepy + active ZFR pair complementations.

4.5 Conclusions

4.5.1 Interaction between counterpart residues at position 102 across the dimer-dimer interface

In Section 4.2, the hypothesis that residues at position 102 interact with their counterparts across the dimer interface at the initial stage of tetramer formation was tested. This hypothesis followed from the observation that in the 1GDT γδ resolvase structure, Glu102 residues protrude from the resolvase. If two dimers are brought together during tetramer formation in a fashion similar to that described by Li et al. (2004), these residues at position 102 might come in contact with one another. Furthermore, mutation of residue 102 (e.g. D102Y in Tn3 resolvase) is key to the hyperactivation of the resolvase enzyme (see Section 1.6.4). It was hypothesised that mutation at residue 102 might act by removing a hypothetical ionic repulsion barrier between counterpart Asp102 (or Glu102 in the case of γ/δ resolvase) residues during dimer-dimer synapsis.

In order to test the hypothesis and determine whether it might be exploited to produce a dimer-dimer orientation specificity bias, ZFRs with all possible ionic substitutions at position 102 were generated and tested in complementation with one another. The results of this experiment (Section 4.2.3, Figure 4-3) showed no indication of any favourable interactions between oppositely charged residue 102 substitution mutants. Therefore, this approach to achieving the project aims was abandoned.

4.5.2 The interface-unlocking model of ZFR activation

In Section 4.3, a model of Tn3 catalytic tetramer activation was postulated based on analysis of 1GDT dimer and 1ZR4 tetramer γδ resolvase structures. In this model the catalytic dimers are activated by 2-3′ interface contacts from the regulatory module, which allosterically disengages a sub-region of the dimer interface seen in the 1GDT dimer structure (Figure 4-6). This region of the dimer interface was termed the 'locking interface' and it can be seen from comparison of the 1GDT and 1ZR4 structures that it must be disengaged in order for the resolvase to become active. Furthermore, the locations of several secondary activating mutations are also consistent with this dimer unlocking model. The hypothesis was generated from this model that since in the wild-

type resolvase system, the catalytic dimers would be unlocked on one side only, asymmetric unlocking of the dimers might produce a conformation that produces a dimer-dimer orientation specificity bias.

Due to its location, the secondary activating mutation G70C was postulated to produce a particularly robust disruptive effect on the locking interface. A differentially mutated ZFR pair, which either contained or did not contain G70C, in an NY background was then generated and termed the sleepy + active ZFR pair, as only a ZFR containing G70C is able to produce recombination activity when used alone. The sleepy + active ZFR pair was then tested in complementations reactions on three substrates that allow the detection of a dimer-dimer orientation specificity bias, and/or a recombination reaction directionality bias. The results of the experiments with the sleepy + active ZFR pair (Section 4.3.3 and Section 4.3.4.3) refuted the hypothesis that asymmetric dimer unlocking could be used to produce a dimer-dimer orientation specificity bias. However, the results also indicated that the sleepy + active ZFR pair produced a significant potential recombination reaction directionality bias as indicated by a difference in the reaction products of the AHD and RHD tetramer configurations (Figure 4-17, lanes 7 and 8). Also importantly, the experiments demonstrated that an active ZFR can be used to activate an inactive ZFR when used in complementation. However, the overall level of activity produced by the sleepy + active ZFR pair using these experiments was low.

4.5.3 Screening for new interface-unlocking mutants

Because the activity of the sleepy + active ZFR pair was low in the previous experiments, an attempt was made in Section 4.3.5 to look for new mutants that might be used in a sleepy + unlocked ZFR pair strategy. Twenty-four existing Tn3 resolvase secondary activating mutations in positions that fit the interface-unlocking model (including G70C from the previous experiment) were incorporated into ZFRs and screened for base-level activity. Surprisingly, in spite of these secondary mutations surpassing the hyperactivation threshold in Tn3 resolvase, all but four were found to be insufficient to significantly activate the ZFR (Figure 4-20 and Figure 4-23). Of those four, G70C (of the previous experiment) imparted the ZFR with sufficient recombination activity to completely convert the substrate to product. The remaining three mutations, G70A, G101S, and V108A, imparted their ZFRs with only modest recombination activity. The

results of this experiment indicated that new active mutants for sleepy + active ZFR pairs would have to be generated from combinations of secondary mutations, and demonstrated that the activation threshold for a Tn3-based ZFR is far higher than it is for Tn3 resolvase.

4.5.4 Testing for synergistic activity in new ZFR pair complementations

In Section 4.3.6, a selection of the secondary activating mutations screened in the previous experiment were used to test three hypotheses (see below). Mutant ZFRs with a secondary mutation in either the E-helix or α/β sub-domain were paired in these experiments such that the secondary mutations they possessed would synergistically disrupt the locking interface from both sides of the interaction.

The first hypothesis tested, was that two sleepy ZFRs might be able to synergistically activate each other when used in complementation. The results of this experiment (Figure 4-24 A, lanes 2–9) demonstrated that this was not accomplished using the mutations selected.

The second hypothesis tested, was whether a lower activity active ZFR might be able to activate sleepy ZFRs when used in complementation. The results of this experiment (Figure 4-24 A, lanes 11–18) demonstrated that no substantial recombination activity was produced using this strategy. However, a minimal amount of recombination product was seen for the complementations involving NY G101S + NY R121K and NY G101S + NY R121M (Figure 4-24 A, lanes 15 and 16, respectively).

The third hypothesis tested was that a sleepy ZFR containing a secondary mutation might produce a higher level of recombination activity when used in complementation with the original active ZFR (NY G70C), than the original sleepy ZFR (NY) which did not contain a secondary mutation. Two important results were gained from this experiment. First, a sleepy ZFR incorporating T109I was able to substantially increase the activity of the sleepy + active ZFR complementation (Figure 4-24 B, lanes 3), in spite of the sleepy ZFR being inactive on its own (Figure 4-20 B, lane 3). Second, sleepy ZFRs incorporating R121K and R121M produced less activity when paired with NY G70C (Figure

4-20 B, lane 6 and 7, respectively), than they did when paired with ZFRs incorporating NY G101S (Figure 4-24 A, lanes 15 and 16, respectively). Because G70C is proposed to have a robust disruption effect on the locking interface, and because R121K/M is located in the same region of that interface, it was predicted that the effect of R121K/M might be made redundant when co-localized across the interface with the G70C mutation. Indeed, this may appear to be the case (see Section 4.3.6.3). It should be noted, however, that the observed difference was small because even the complementations using G101S and R121K/M produced a very low level of recombination activity.

An additional and interesting observation from the foregoing set of reactions can be made. All of the complementations between NY G70C, and NY R121G/K/M/S and NY E124Q produced virtually no activity (Figure 4-24 B, lanes 5–9), while the complementation between NY G70C and NY had minimal activity (Figure 4-24 B, lanes 2). This result may suggest that pairing G70C, and R121G/K/M/S or NY E124Q within the same locking interface actually reduces the activity of the protein. Interestingly, Burke *et al.* (2004) reported that although A117V, R121K, and E124Q produced hyperactivity when used as lone secondary mutations, using a combination of these three mutations at once abolished the activity of Tn3 resolvase. Because the locations of A117V, R121K, and E124Q indicate they would all be expected to disrupt the slightly different portions of the locking interface (Figure 4-4), it is possible that results of the current experiment, and Burke *et al.* (2004), suggest that a complete knockout of this interface produces an unviable catalytic domain. The locking interface comprises a major portion of the dimer interface, and completely knocking out the interactions of this region may prevent the dimer from forming a proper quaternary conformation, or perhaps from forming at all. One consequence, were this the case, is that ZFR complementations might be designed that produce post-reaction tetramer configurations that are completely inactive due to the pairing of multiple locking interface mutations from both sides of the interaction. If such a strategy were successful it could lead to a high fidelity recombination reaction directionality bias.

4.5.5 Sleepy + active ZFR pairs based on position 102 and landing pad mutations

Experiments were carried out in Section 4.4 based on the model that the residue at position 102 of the E-helix must negotiate contact with a region of the N-terminal end of an opposing E-helix (termed here the landing pad) as the dimer-dimer interface forms. This model is suggested from analysis of the structural transformation that must take place in order to allow the 1GDT γδ resolvase structure to transform into the 1GDT tetramer structure, the tetramer formation trajectory proposed by Li *et al.* (2005), and manual docking 1GDT dimers together into a tetramer initiation structure (Figure 4-10). Additionally, this model explains the function of the primary mutations at residue 102 and several secondary activating mutations within the landing pad regions, all of which are hydrophobic resides, and thus are proposed to cause activation through mutual hydrophobic interaction during the initial stage of tetramer formation.

An experiment was designed to test whether using a mutation at 102 (Y), or a collection of landing pad mutations (M), as an independent variable, could produce a dimer-dimer orientation specificity bias or a recombination reaction directionality bias. The results of this experiment (Figure 4-27) demonstrated that all subunits in a ZFR sleepy + active ZFR complementation must possess a primary mutation in order to achieve significant activity. Impressively, the NY + NM complementations (lanes 10–12) appeared to produce a strong activity bias between the AHD, and MHD and RHD configurations. This activity bias between tetramer configurations indicates a strong potential for a recombination reaction directionality bias. Whether this differential configuration activity is a result of the interactions proposed by the landing pad model, is the product of alternative activities of a specific subset of mutations within M, or is simply a by product of any sleepy + active ZFR pair with sufficient activity, would require further experimentation to determine. However, the desired recombination reaction directionality bias that this project set out to generate does appear to have been achieved at some level. It can also be noted that a recombination reaction directionality bias based on these tetramer configurations may lead to an integration orientation specificity bias as well, as one integration orientation will be reversible and stochastically disfavoured, while the other integration orientation will be more stable; however, this topic will be covered in detail in the next chapter.

4.5.6 The sleepy + active complementation strategy may improve the targeting fidelity of the ZFR system.

Even apart from producing a recombination reaction directionality bias, the discovery that inactive ZFRs may be activated by active ZFRs in a sleepy + active complementation strategy is potentially valuable for increasing the targeting fidelity of the system. Cytotoxicity has been observed when using ZFRs in mammalians cells, which has been linked to off-target activity (Gaj et al., 2014). Because sleepy ZFRs are inactive when used alone or in combination with one another, the incorporation of sleepy ZFRs in a ZFR reaction reduces the number of possible aberrant recombination reaction that may take place as a result of unintended pairing of subunits at off-target sites.

4.5.7 Use of DEVs to increase activity levels in ZFR complementation experiments

In Sections 4.3.4 and 4.3.5.5, DEVs were used to increase the expression of ZFRs in an attempt to increase the rate of the reaction; however, this strategy was not successful. While it was observed in Chapter 3 that increasing expression of ZFRs with the Z5 ZFA (which have low activity) does produce increased recombination activity (Figure 3-4 and Figure 3-5 D), the results from Sections 4.3.4 and 4.3.5.5 indicate this effect plateaus somewhere between double expression and quadruple expression of the enzyme. This plateau may be the result of Z-site occupancy saturation.

4.5.8 Confirmation of SA effect hypothesis

A useful piece of information was regarding the SA effect was gained from an experiment in Section 4.3.4. In Figure 4-17 it can be observed that co-expression of sleepy + active ZFRs from two DEVs simultaneously, produces a profound SA effect for any plasmids that contain an active ZFR homodimer Z-site. Additionally, it may be observed that an intermediate SA effect is produced on any plasmid with a heterodimer Z-site. However, the reaction in lane 2 of Figure 4-17 clearly indicates that when the plasmid contains only homodimer Z-sites for the sleepy ZFR, no SA effect is produced. This result confirms that the SA effect results from the catalytic activity of the ZFR rather than simply from its binding activity. As outlined in Section 3.10.5.4, the SA effect is likely the result of

DNA-bound ZFR dimers being activated to cleave DNA outwith proper recombination reactions through interaction with non-DNA-bound ZFRs.

Chapter 5: *ZFR Binding Domain Modifications*

5.1 Introduction

5.1.1 Conspectus

Another avenue through which the action of the ZFR might be altered in order to achieve the project aims (i.e. integration orientation specificity, recombination reaction directionality bias, and loosening of the ZFR specificity requirements) is modification of the ZFA DNA-binding domain. Two novel approaches were developed in order to explore this idea: the use of ***subunits with non-specific ZFAs,*** and the use of ***subunits with reduced binding affinity***. Both of these approaches are based on the principle of generating a heterodimer comprised of one normal binding activity subunit and one modified binding activity subunit. The strategies used for both approaches, which will be discussed in the following sections, were designed to address all three project aims.

The experiments utilizing subunits with non-specific ZFAs produced only inactive ZFRs (with the exception of one potentially toxic ZFR) and were not successful in achieving any of the project aims. The experiments utilizing subunits with reduced binding affinity were successful in demonstrating active ZFR heterodimers, but the results suggest that if these enzymes were used for integration reactions one of the product Z-sites might not be re-ligated after recombination. Thus, this strategy does appear to have produced a useful modified ZFR system. However, a final experiment, which attempted to use a subunit with a Tn3 resolvase HTH domain as a non-specific and reduced binding affinity heterodimer partner, appears to have been completely successful in achieving all project aims. This strategy appears to produce an extremely robust recombination reaction directionality bias that will likely give rise to orientation specific integrations (see Section 5.3), and also reduces the binding recognition requirements for the ZFR dimer.

5.1.2 Binding specificity manipulation to reduce sequence-recognition limitations

5.1.2.1 Complementation with reduced binding specificity ZFRs

Non-specific DNA-binding domains could be used to loosen the DNA binding specificity requirements of the ZFR system. Heretofore there has been little use for a non-specific DNA binding domain in the ZFR system, as it would be expected to produce dimers with very low DNA-sequence recognition capability, leading to unwanted off-target recombination reactions. However, the work in Chapter 4 (Sections 4.3 and 4.4) has demonstrated the use of a ZFR complementation system where one subunit (the 'sleepy' subunit) has no activity when used alone, and requires an active heterodimer partner in order to function in recombination reactions. Were the sleepy subunits outfitted with the non-specific DNA-binding domain, homodimers comprised of them would be expected to produce no recombination activity on their own, thus obviating the problem of off-target reactions by the binding specificity reduced subunits.

5.1.2.2 Remaining sequence selectivity of reduced specificity ZFR heterodimers

In the reduced specificity ZFR heterodimer system, ZFR subunits with normal binding specificity are used in complementation with the subunits possessing a non-specific DNA-binding domain. Although the incorporation of a subunit with a non-specific DNA binding domain produces a ZFR heterodimer with reduced site-specificity, it still maintains enough sequence selection throughout to specify a theoretically unique target site within a human size genome. The human genome is approximately 3 billion bp in length; in order to achieve unique sequence selectivity within both strands of a 3 billion bp genome, a theoretical requirement of 17 bp of site-specific recognition is required (see Footnote 2 on page 42).

The Tn3 resolvase catalytic domain and its arm region, which are incorporated into the ZFR, retain DNA sequence recognition capability of 8 bp per subunit; although, in practice, this sequence recognition is reduced by a degree of non-specific binding tolerance ((Gaj et al., 2013b) and M. Prorocic, (2009), PhD thesis, Stark lab). The number of changes that are tolerated depends on the identity of the base changes and position of the changes within the recognition sequence (see Section 1.5.2.4). As a general rule, the

eighth and final base at the flank of the central portion of Z-site that has been carried over from Tn3 resolvase (see Chapter 3, Figure 3-2 A), is very permissive to base changes. In addition to changes at this permissive eighth base in a subunit recognition sequence, evidence from the work of M. Prorocic suggests that one to three additional base changes might be tolerated across the whole 16 bp portion of a ZFR dimer-binding recognition sequence contacting the Tn3 resolvase module of the ZFR enzyme. Whether one or three additional changes are tolerated depends on the identity of the changes. This consideration of site recognition tolerance of the Tn3 resolvase module of the ZFR indicates that perhaps 11 bp of reliable recognition specificity are contributed to the overall sequent-selective capability of a ZFR dimer.

While the ZFA module of the ZFR enzyme may contribute 9–11 bp (see Section 1.7.1.2 for discussion of the 5′ extended base recognition and 3′ overlap recognition of ZFAs) of sequence selectivity to the recognition capability of a ZFR subunit, some level of non-specific recognition tolerance is often present. A binding characterization of zinc finger nucleases based on the same ZFAs as the Z2 and Z3 domains used in this project, provides some idea of the non-specific recognition tolerance that might be expected when using ZFAs produced through the OPEN platform (Pattanayak et al., 2011). The data collected by Pattanayak *et al.* (2011) suggests that for each ZFA, one position within the 9 bp ZFA recognition site core will readily tolerate a variety of base changes. Additionally, the data from Pattanayak et al. suggests the 5′ extended base recognition and 3′ overlap recognition are not robustly specified. Correspondingly, the OPEN ZFAs might be expected to contribute 8 bp of reliable recognition specificity to each ZFR subunit binding at a recognition site.

It should be noted, that data from Pattanayak et al. (2011) also suggests that, at least in a ZFN context, a properly bound subunit may be able to compensate for incorrect binding of the other subunit in a dimer context, allowing for a wider array of specificity tolerance. Therefore, in a ZFR context, a subunit containing a high affinity non-specific binding domain might facilitate greater sequence recognition tolerance at the half-site of the other subunit with a sequence-specific ZFA. However, this problem, which would further reduce the recognition specificity of a ZFR dimer, would only present itself if the non-specific binding domain has sufficient binding affinity to compensate for the sequence changes to the half-site of its subunit partner.

The assessment of the non-specific sequence recognition tolerability of both the catalytic and binding modules of the ZFR, suggest that even if one non-specific binding domain were incorporated into a ZFR heterodimer, there is sufficient recognition capability to uniquely bind targets within a human-size genome. As mentioned above, statistically unique binding site recognition within the human genome requires at least 17 bp of sequence selectivity. The two Tn3 catalytic modules of a ZFR dimer have been estimated above to collectively contribute 11 bp of reliable sequence recognition. Even if the non-specific recognition tolerance of a sequence-specific ZFA used in one subunit were higher than discussed above—which in some cases it may be (Paillard et al., 2004, Lam et al., 2011)—unique genomic targeting by a ZFR heterodimer incorporating one non-specific binding domain, could still be achieved so long as the sequence-specific ZFA of the other subunit could reliably specify at least 6 bp.

5.1.3 Binding affinity manipulation for unidirectional recombination reactions

5.1.3.1 ZFR integration and excision reactions

The major barrier preventing the application of ZFRs to gene therapy is their inability to catalyse stable orientation-specific integration reactions (see Chapter 6 for a discussion of how ZFRs might be used in a therapeutic gene correction strategy provided they were able to achieve stable orientation-specific integration reactions). Currently, the orientation of integrations catalysed by ZFRs cannot be controlled. Furthermore, integration reactions catalysed by ZFRs are of low efficiency (0.13–1.6%) (Gordley et al., 2009, Gaj et al., 2011, Gersbach et al., 2011, Gaj et al., 2013b, Gaj et al., 2014). This low integration efficiency is the result of the bidirectional recombination activity of the ZFR. While ZFRs are capable of catalysing integration reactions, they are also capable of catalysing excision reactions. ZFR integration reactions involve two separate DNA molecules, each containing a Z-site, which must be brought into contact with one another so that the DNA may be recombined in order to produce one DNA molecule. Conversely, ZFR excision reactions involve one DNA molecule with two Z-sites, which must be brought into contact with each other so that they may be recombined in order to produce two separate DNA molecules. Thus, excision reactions are the reverse of integration reactions.

Within an *in vitro* system, the equilibrium between ZFR integration and excision reactions will be somewhat skewed towards excision reactions because the Z-sites on which the ZFR operates will be kept in closer proximity to one another, due to being located on the same DNA molecule. Conversely, the Z-sites involved in an integration reaction are able to travel much farther from each other, reducing the efficiency of insertion reactions relative to excision reactions. However, in an *in vivo* system within replicated cells, the equilibrium between ZFR integration and excision reactions is heavily skewed towards the excision reaction. This is because the Z-sites involved in the integration reactions are not just able to move away from each other, but are also capable of being distributed to different daughter cells during cellular replication, thus entirely preventing the possibility of an integration reaction. Additionally, in the case of integration reactions performed on cellular chromosomes, the DNA molecule to be integrated may not be capable of replicating with the cells, resulting in a decrease in the ratio of DNA donor molecules relative to chromosomal targets within the cell population.

5.1.3.2 Non-reversible recombination reactions in other systems

The most effective way of reducing the reaction equilibrium bias towards excision reactions is to make the excision reaction non-viable. Several integrase systems exist in nature that utilize site-specific recombinases, like the ZFR, but which possess unidirectional reaction capability. This unidirectional reaction capability is often based on the generation of post-recombination binding sites that are no longer able to facilitate the activation of the recombinases. In the case of the large serine integrases, this effect is achieved by utilizing a pair of recombinase binding sites (*attP* and *attB*), whose half-sites are reciprocally rearranged during DNA recombination to produce a new pair of recombinase binding sites (*attL* and *attR*) (Smith et al., 2010). These two pairs of binding sites produce different DNA-recombinase docking conformations resulting in two recombinase dimers (one bound at each site) either having, or not having, geometries that facilitate catalytic tetramer formation.

5.1.3.3 Exploiting reciprocal half-site rearrangements

Engineering the ZFR to produce unique DNA-subunit docking conformations is probably not a feasible approach to achieving a recombination reaction directionality bias. However, the basic principle of exploiting the reciprocal rearrangement of half-sites during recombination to generate recombinase dimers that are active versus inactive may

be feasible using a more rudimentary approach. The DNA binding domain of a ZFR subunit could be modified such that it lacked adequate binding affinity to allow homodimer reactions comprised of two of these subunits. These reduced affinity binding domain (RABD) subunits might then be used in conjunction with functional 'helper' subunits that could compensate for the reduced binding activity when participating in heterodimer interactions. Because the dimer interface of the ZFR is quite extensive, it was hypothesized that the negative energy it provides might be sufficient to allow a helper subunit to stabilize a RABD subunit. The key to achieving success with this strategy is to find a balance of binding activity attenuation that is sufficient to knock out binding activity of an RABD homodimer, but which still allows for binding activity compensation from a helper ZFR participating in a heterodimer.

5.1.3.4 Tetramer configurations and the RABD subunit

Another beneficial repercussion of using the RABD + helper ZFR strategy is that it may lead to orientation specific insertions. In order to understand why this would happen we must first revisit the concept of tetramer subunit configurations discussed in Chapter 4. As described in Section 4.3.1.6, tetramers comprised of an equal number of two types of ZFR subunits (i.e. two subunits of each type) may form in three different configurations: AHD, MHD, and RHD (Figure 4-11). The configuration that is formed depends on both the configuration of the Z-site (homodimer-binding versus heterodimer-binding), and on the orientation of the dimers relative to one another in the dimer-dimer complex. The AHD configuration forms when ZFR dimers are bound to homodimer Z-sites, while the MHD and RHD configurations may form when ZFR dimers are bound to heterodimer Z-sites. The MHD and RHD configurations differ by the orientation of the heterodimers relative to one another. Importantly, subunit rotation during recombination reactions starting from the AHD configuration results in its conversion into the RHD configuration, and vice versa. However, subunit rotation during recombination reactions starting from the MHD configuration results only in the regeneration of the MHD configuration.

In the RABD + helper ZFR strategy, a homodimer comprised of two RABD subunits will not be active because they have insufficient collective binding affinity to bind their Z-site. However, the RABD subunits are able to participate in heterodimer-based reactions because the binding affinity of the helper subunit and negative energy of the dimer

interface allows the helper subunit to stabilize the RABD subunit on the RABD subunit half-site. Therefore, in the RABD + helper ZFR strategy, the MHD and RHD configurations are active, while the AHD configuration cannot form, because RABD homodimers are unable to bind at their Z-sites, leading to a non-reversible reaction (Figure 5-1 A).

5.1.3.5 'Try and try again' orientation specific integration reactions

If the MHD and RHD configurations are active, while the AHD configuration is not, this can lead to the stochastic production of orientation specific insertions (Figure 5-1 B). In the scenario of a ZFR integration reaction targeting an endogenous chromosome, the two half-sites of the chromosomal Z-site are likely to require two different ZFAs to target them (i.e. the sequences of the two half-sites will not be the same). If the same Z-site was used on the DNA to be integrated (which would minimize the variety of ZFRs needed for the reaction), this would imply the tetramer configuration of the integration reaction would either be MHD or RHD (the two configurations possible by heterodimers).

If the ZFR tetramer carries out integration from the RHD configuration, then the integration will be irreversible, because the product Z-sites of the reaction are homodimer sites, one of which specifies the RABD homodimer, which cannot bind its Z-site. However, if the ZFR tetramer carries out integration from the MHD configuration, one of two subsequent reactions may follow. A subsequent inversion reaction, starting from an RHD tetramer, will again produce product homodimer Z-sites, one of which cannot be bound by its RABD homodimer. Alternatively, an excision reaction, starting from an MHD configuration, may reverse the integration reaction and return the system to its original state. If the system is returned to its original state, then the same set of reaction possibilities once again present themselves. Because only one orientation of the integrated DNA produces a product Z-site that can no longer be acted on by its ZFRs (whether by the initial integration reaction, or a secondary inversion reaction) the recombination equilibrium will ultimately favour that outcome. In short, if at first orientation specific integration does not succeed, the ZFR will *try and try again*.

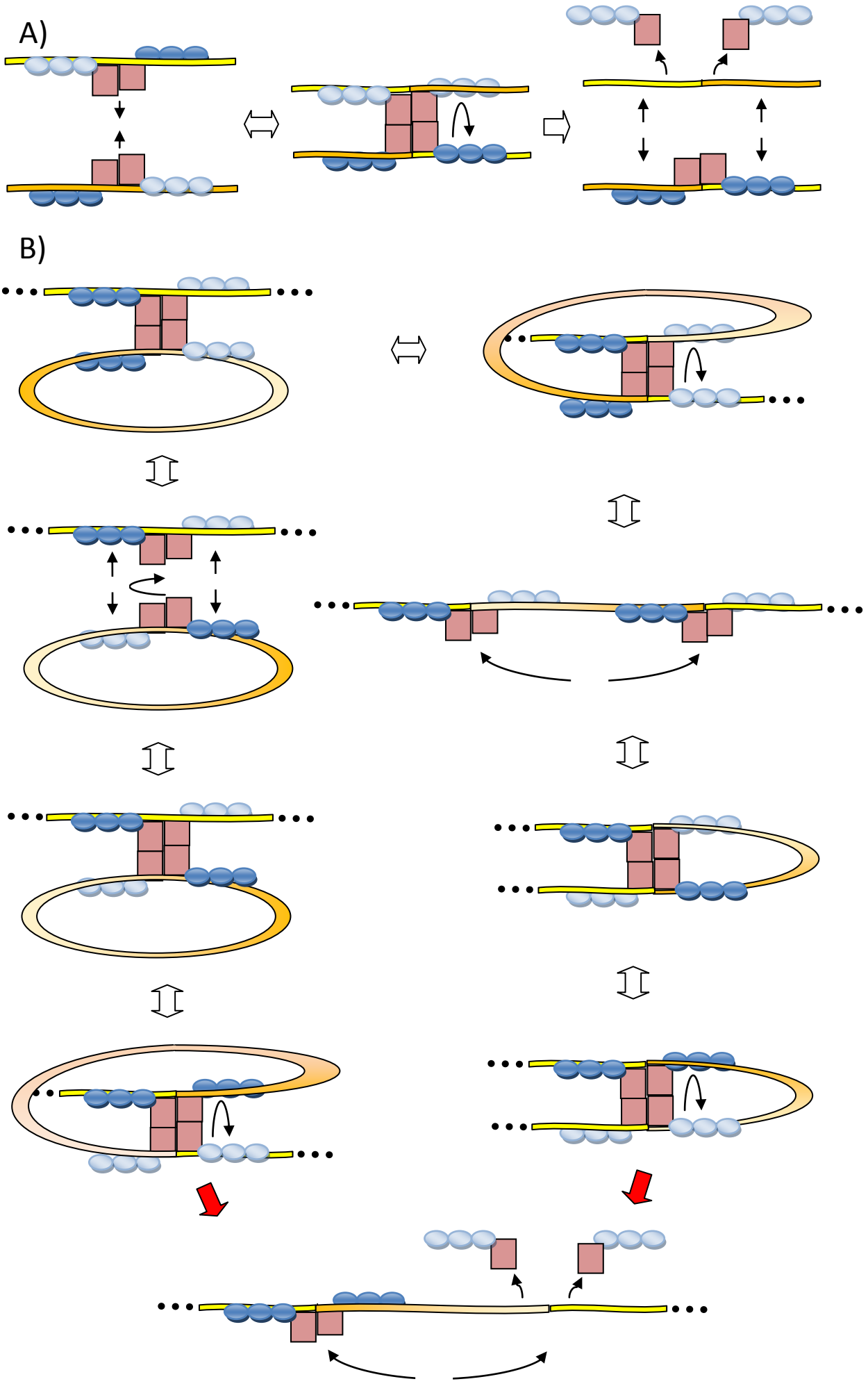

A)
B)

Figure 5-1: RABD + helper non-reversible reaction, and 'try and try again' integration orientation specificity. Two DNA molecules are depicted in yellow and orange with shading used to indicates directionality of the DNA, where relevant. The ZFR subunits are depicted with rose coloured boxes representing the catalytic domain, and blue ovals representing the ZFA binding domain. Dark blue is used to indicate a ZFA with strong binding affinity, while light blue is used to depict a ZFA with weak binding affinity. Black arrows are used to indicate spatial translocations of the ZFR subunits and DNA. **A)** At the left the figure depicts two DNA-bound heterodimers comprised of subunits with differing binding strength, which are moving into association. The middle drawing shows the two DNA-bound heterodimers in synapsis, about to undergo subunit rotation and DNA recombination. The image at the right depicts the subsequent post-recombination state, where the weak binding affinity ZFR subunits are no longer capable of associating with each other and the DNA. **B)** Depicts a pathway diagram showing how a circular molecule of DNA being integrated by differing binding affinity system will have a stochastic tendency towards insertion in a particular orientation. White arrows are used to indicate reaction directionality.

5.2 Non-specific ZFAs subunit + helper subunit complementations

5.2.1 Introduction

As described in Section 5.1.2, non-specific binding modules might be used to reduce the sequence recognition limitations of the ZFR system. This section describes experiments that were conducted using non-specific ZFAs constructed by modular assembly. The success rates of ZFAs constructed through modular assembly are reported to be quite low (Ramirez et al., 2008), and one study has indicated the factors influencing this low rate of success may be independent of site-specificity (Lam et al., 2011). Therefore, it was expected that the majority of these newly constructed non-site-specific ZFAs would have low binding affinity. Because the binding affinity of the modules was expected to be low, they also were intended to serve as a test of the RABD + helper ZFR strategy (see Section 5.1.3). However, the initial experiments to test their base-level activity showed they were not active, and thus, they did not make it to the stage where recombination reaction directionality bias could be assayed.

5.2.2 Modular assembly of non-specific ZFAs and their incorporation into ZFRs

5.2.2.1 Selection of non-specific ZFA monomers

Although any type of non-specific DNA binding domains might be incorporated into a chimeric recombinase, non-specific ZFAs were an obvious choice for this first experiment because the parameters for incorporating ZFAs into ZFRs are well understood.

Three engineered zinc finger monomers have been reported in the literature that are described as binding to DNA without any detectable sequence specificity ('SPG-H-LIE', 'RKWLRL'/'LZF25', and 'QRH-H-LVE') (Dreier et al., 2005, Joung et al., 2000, Lee et al., 2011). It should be noted, that although these ZFAs were shown to bind successfully to all substrate sequences on which they were tested (see Table 5-2), the experiments in which they were discovered were not designed to test the full range of sequence possibilities. The 'SPG-H-LIE' and 'QRH-H-LVE' zinc finger monomers were said to bind with 'high affinity' in addition to being non-specific, though no dissociation constant or other form of quantification was provided for them in the report by Dreier et al., (2005). To the best knowledge of this author, a completely non-specific ZFA has never been constructed, and these non-specific zinc finger monomers have never been assembled to each other.

Table 5-2: Site that the non-specific ZFAs were shown bind (Joung et al., 2000, Dreier et al., 2005).

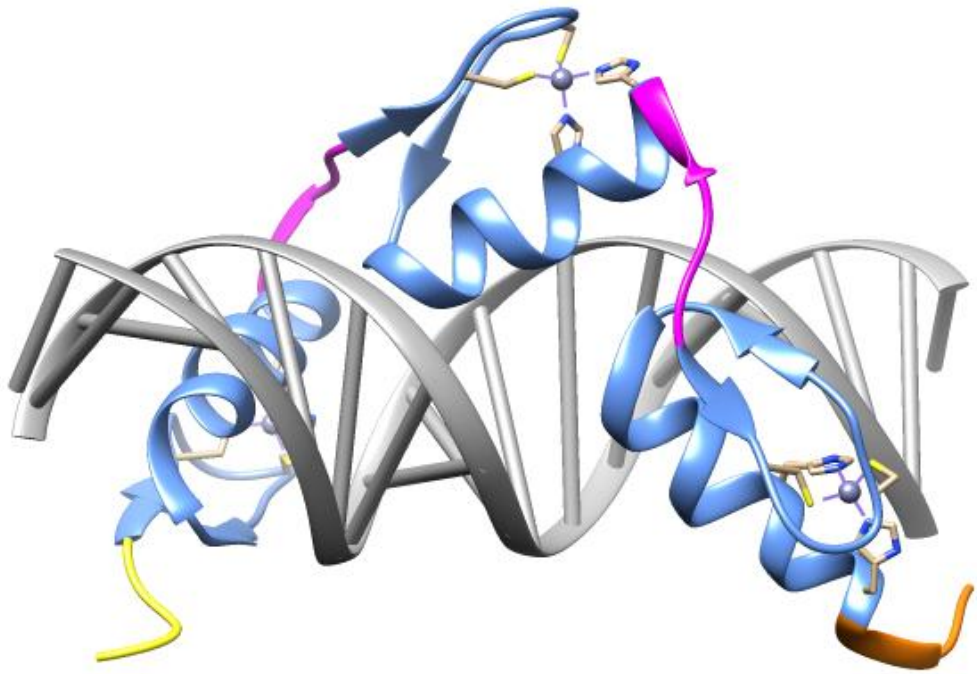
Z2 site	RKWLRL	SPG-H-LIE	QRH-H-LVE
GAC GCT GCT	AAA	CAA	CAA
	TCA	CAC	CAC
	TGT	CAT	CAG
		CCA	CAT
		CCC	CCA
		CCG	CCG
		CTT	CCT
			CGC
			CTC
			CTT

5.2.2.2 Rationale for construction strategy

The success of a ZFA comprised of randomly selected monomers may be affected by many variables including: the binding affinity contribution of each zinc finger monomer; whether the monomers have position specificity preference (e.g. finger 1, finger 2, and finger 3 in the array); and compatibility between adjacent monomers within the array. Using all combinations of the three non-specific zinc finger monomers, 27 three-fingered ZFAs and nine two-fingered ZFAs may be constructed. Because the ZFAs were not guaranteed to work, it was necessary to construct and test all 36 possible ZFAs in order to thoroughly test the principle.

The direct synthesis of 36 ZFAs is cost-prohibitive and so a strategy was devised to construct the 36 ZFAs from a minimal amount of starting material, and using a minimal number of construction steps. Architecturally, a Zif268-like ZFA amino acid sequence can

be broken down into two different types of components: the zinc finger monomers, which contain the recognition helices responsible for DNA-base recognition; and the monomer-flanking motifs (the N-terminal motif, inter-monomer linkers, and C-terminal motif; Figure 5-3). The zinc finger monomers selected for this experiment were all taken from projects based on a Zif268-like architecture, and so Zif268-like architecture motifs were used for the N-terminal motif, inter-monomer linkers, and C-terminal motif. The ZFA construction strategy used here was to synthesize the sequences for the monomers and monomer-flanking motifs separately, and then recombine the elements to produce all of the two- and three-fingered non-specific ZFAs. This avoided the need to synthesize whole ZFAs, or the need to synthesize the same monomer repeatedly in each position context (i.e. the same monomer flanked by the position-specific monomer-flanking motifs).



MERPYACPVESCDRRFS**Q**K**TN**LD**T**HIRIHTGQKPFQCRICMRNFS**QHTG**LN**Q**HIRTHTGEKPFACDICGRKFA**TLHT**RDRHTKIHLRQKD

Figure 5-3: Zif268-like ZFA architectural elements. The figure shows the crystal structure of $TATA_{ZF}$* (PDB: 1G2D), an engineered ZFA based on Zif268 architecture. Colours are used to delineate the structural elements. The zinc finger monomers are coloured blue, while the N-terminal motif, inter-monomer linkers, and C-terminal motif are coloured yellow, magenta, and orange, respectively. Below the structure is the amino acid sequence of the ZFA with the sequence highlighted to match the structure. The final Aspartic acid in the sequence is disordered in the crystal structure above and does not appear within it. Additionally, amino acids of the recognition helix have been underlined, and the residues involved in base recognition bolded (Wolfe et al., 2001).

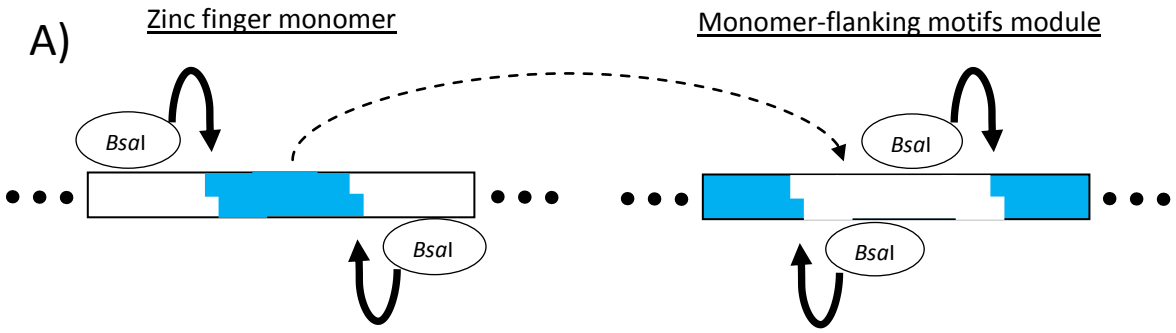
5.2.2.3 Construction strategy

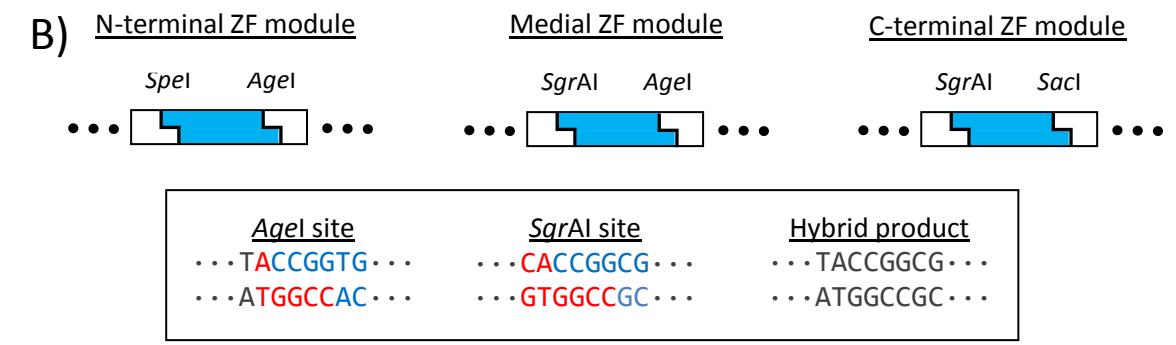
The amino acid sequences for 'SPG-H-LIE', 'RKWLRL'/'LZF25', and 'QRH-H-LVE' were extracted from the published literature, and then translated into DNA sequences appropriate for *E. coli.* These sequences were then flanked by *Bsa*I sites. *Bsa*I is a Type IIS restriction endonuclease that cleaves DNA at a distance from its recognition site, allowing DNA sequences flanked by *Bsa*I sites to be extracted with custom overhangs that are not dependent on *Bsa*I.

Sequences for the monomer-flanking motifs were arranged into three modules containing the flanking sequences for finger 1 (N-terminal motif and linker), finger 2 (linker and linker) and finger 3 (linker and C-terminal motif). Within each module, the two monomer-flanking motif sequences were separated by a filler sequence, but which contained *Bsa*I recognition sites. This arrangement allowed the non-specific monomers to be 'plugged' into their position context, which contained the appropriate monomer flanking sequences (Figure 5-4 A).

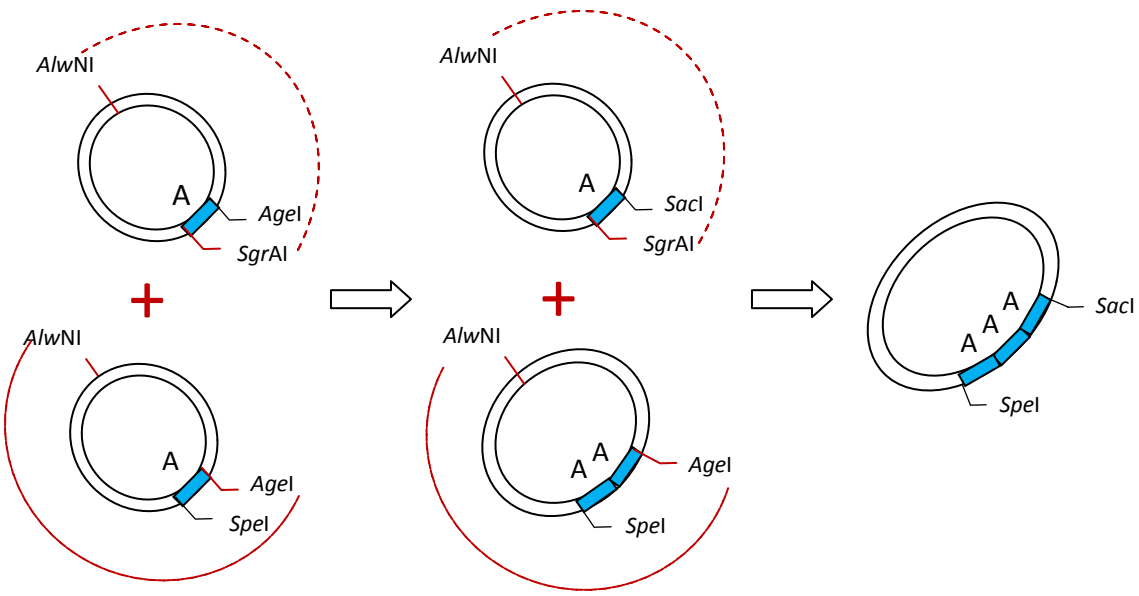
The inter-monomer linker sequences, themselves, were designed to contain *Age*I and *Sgr*AI isocaudomer sites, whose overhangs when ligated together produce a hybrid site on which neither of the isocaudomers are active (Figure 5-4 B). This allowed for a one-way assembly procedure where the zinc finger domains could be sequentially attached to one another, without the restriction enzymes that liberate the next attachment overhang, also re-cleaving zinc finger monomers that were already joined.

The DNA sequences for the N-terminal and C-terminal motifs were designed to contain restriction sites *Spe*I and *Sac*I, respectively, which allow fully constructed ZFAs to then be cloned into an expression plasmid containing the desired Tn3 catalytic domain mutant sequence.





AgeI site	SgrAI site	Hybrid product
···TACCGGTG···	···CACCGGCG···	···TACCGGCG···
···ATGGCCAC···	···GTGGCCGC···	···ATGGCCGC···



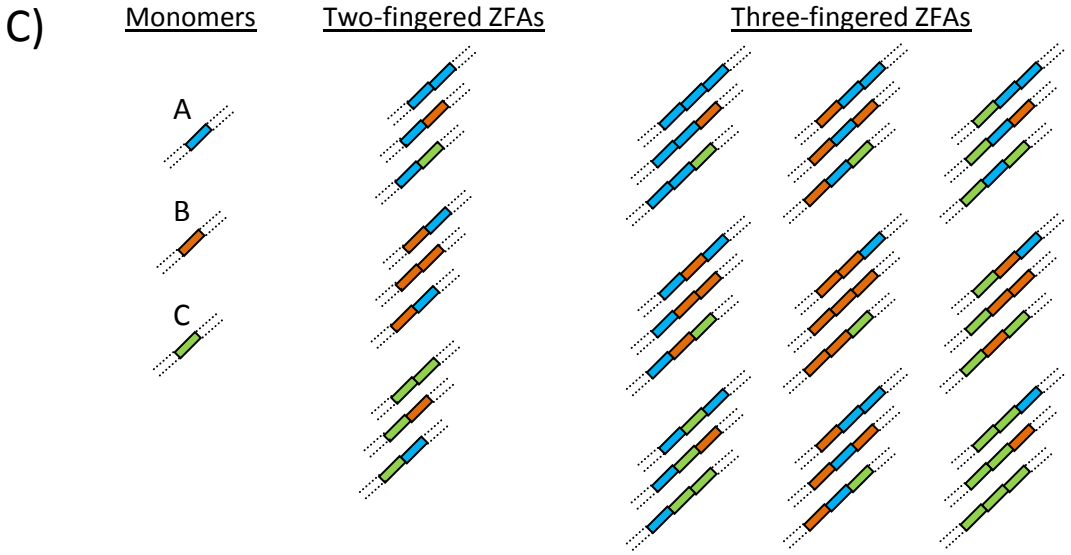


Figure 5-4: Minimal materials assembly strategy for 38 non-specific ZFAs. A) The panel depicts the initial assembly step where DNA sequences for zinc finger monomers are loaded into 'monomer-flanking motif modules' which provide DNA sequences for the linker and terminal amino acid motifs which determine zinc finger position within a ZFA. The relevant DNA sequences are represented by the blue portion of the rectangles. The *Bsa*I restriction enzyme is shown cutting at a distance from its own recognition sites, which are contained within the segments of DNA to be discarded. **B)** The panel depicts the one-way sequential monomer assembly process which relies on the use of isocaudomer restriction sites (*Sgr*AI and *Age*I) to produce ZFAs. The three types of position-specific zinc finger modules and their embedded restriction sites are shown at the top. Underneath, a box depicts the *Sgr*AI and *Age*I isocaudomer sites, along with a hybrid ligation product which is no longer recognized by either enzyme. Blue and red coloured bases are used to indicate the limit of the recognition sequences and overhangs while dark grey is used to depict relevant bases which are not recognized by the restriction enzymes. Bellow the box is a depiction of the one-way assembly of a three-fingered ZFA. As new zinc finger modules are added to build the ZFA (by digestion and relegation of plasmids containing the modules), only two restriction sites, flanking the total sequence, are left after each step. These restriction sites are used to add additional modules and extract ZFA when assembly is complete. **C)** The panel depicts the combinatorial possibilities for two- and three-fingered combinations based on three zinc finger monomers.

5.2.2.4 ZFA Assembly and incorporation into ZFRs

DNA sequences for the non-specific zinc finger monomers ('SPG-H-LIE', 'RKWLRL'/'LZF25', and 'QRH-H-LVE') and monomer-flanking motif modules were synthesized by GeneArt® (Life Technologies Ltd, Paisley, UK), and assembled in the fashion described above to create coding sequences for nine two-fingered, and 27 three-fingered ZFAs (Figure 5-4 C). At this point, the 'SPG-H-LIE', 'RKWLRL'/'LZF25', and 'QRH-H-LVE' zinc finger monomers were also renamed 'A', 'B', and 'C', respectively, to provide a simplified naming convention for describing the ZFAs. The success rates of ZFAs constructed through modular assembly are reported to be quite low (Ramirez et al., 2008), and one study has indicated the factors influencing this low rate of success may be independent of site-specificity (Lam et al., 2011). Therefore, it was expected that the majority of these newly constructed non-site-specific ZFAs would have low of binding activity. Because the Tn3[NY G70C] mutant catalytic domain is highly active (see Chapter 4, Section 4.3.5; see Figure 4-20), it was chosen for these experiments as its high level of activity was thought a good match for these potentially low binding affinity domains. Thus, the ZFA coding sequences were extracted from their parent plasmids and cloned into an expression vector downstream of a Tn3[NY G70C] catalytic domain.

5.2.2.5 The CAC ZFA may be toxic

During the final construction stage, when the ZFAs were joined to the activated Tn3[NY G70C] catalytic domain within an expression vector, it was particularly notable that no transformants were recovered from transformation of the plasmid construct containing the Tn3[NY G70C]-**CAC** ZFR, which appeared to be toxic to the *E. coli* cells that had been transformed with it. Cell toxicity is what one would expect if a ZFR with a highly affinity non-specific binding domain were to be expressed within cells, as it may allow the ZFR to perform random recombination reactions on both the expression vector and the circular *E. coli* chromosome (given sufficient homology to the central portion of the Z-site, which is bound by the Tn3 resolvase catalytic domain and arm region). However, *this result was not confirmed*, and also might have arisen as the result from an error in the transformation procedure. Nevertheless, the Tn3[NY G70C]-CAC ZFR is omitted from the experiment that follows.

It should be noted, that the possibility of toxicity for a few ZFRs with these domains was not entirely unexpected. However, the activity of most ZFRs produced by modularly assembled ZFRs is generally so poor (Ramirez et al., 2008) that it was expected that the majority of ZFRs with these domains would not have enough activity to produce cellular toxicity. It is important to remember that for the goal of this experiment was to produce non-specific ZFAs with *low binding activity*, so they could be used as RABD ZFR in RABD + helper complementations. Therefore, the appearance of toxicity in the construction stage essentially eliminates those ZFRs that will not function as RABD ZFRs.

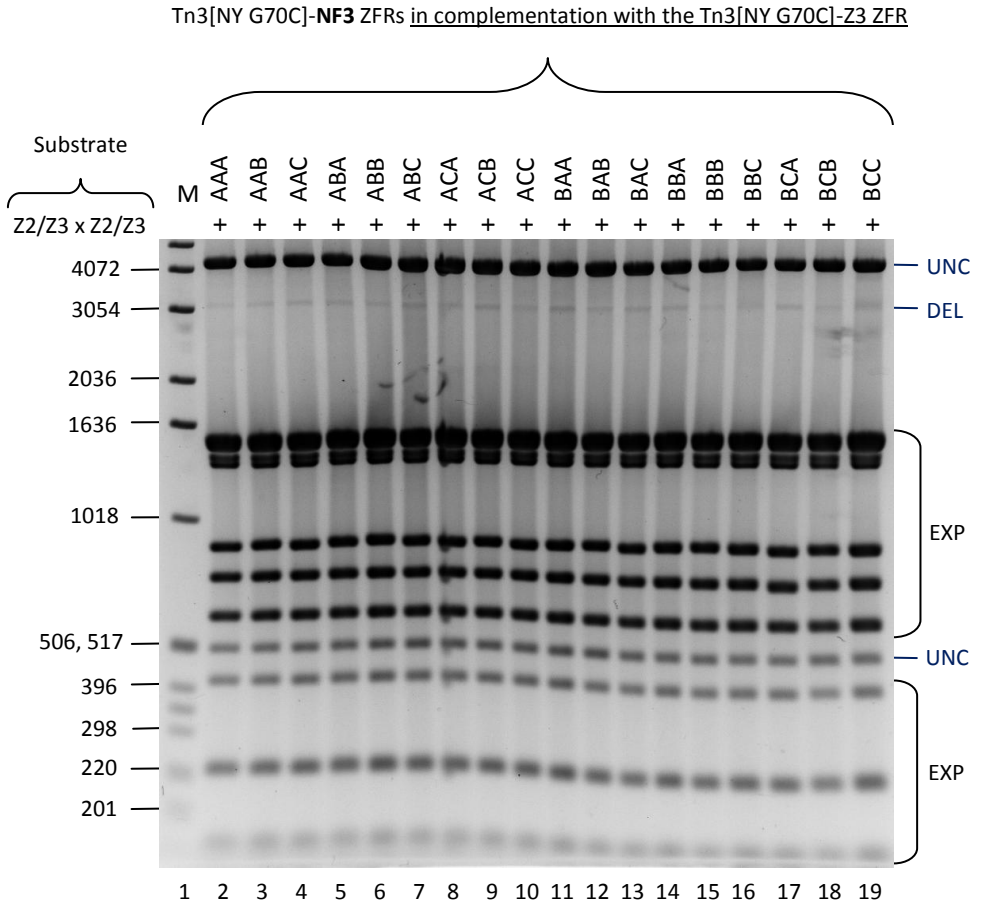
5.2.3 Experiment design

Experiments were designed to test the ability of the ZFRs with non-specific ZFAs to bind at half of a heterodimer-binding Z-site, in conjunction with a helper ZFR, bound site-specifically to the other half-site. The 26 available ZFRs containing three-fingered non-specific ZFAs comprised of all combinations of A, B, and C (see Section 5.2.2.4), with the exception of CAC (see Section 5.2.2.5), and nine available ZFR containing two-fingered non-specific ZFAs comprised of all combinations of A, B, and C (see Section 5.2.2.4), were used as the non-binding-specific ZFR subunits for this experiment. The Tn3[NY G70C]-Z3 ZFR was chosen as the site-specific helper ZFR for this experiment because the Tn3[NY G70C] catalytic domain mutation set imparts high activity (see Chapter4, Section 4.3.5)

and the Z3 ZFA imparts slightly more activity to the ZFR than the other available domains (see Chapter 3, Section 3.5). For these experiments, a 2MutHetDim-DR substrate was used, which contained two heterodimer-binding Z-sites (Z2/Z3 x Z2/Z3). Thus, a complementation experiment was carried out using the 17 Hour Recombination Assay.

5.2.4 Results

The results of both complementation experiments involving ZFRs with three-fingered non-specific ZFAs (Figure 5-5), and two-fingered non-specific ZFAs (Figure 5-6), and partnered with helper ZFRs, show only trace levels of recombination activity for all of the ZFRs pairs tested.



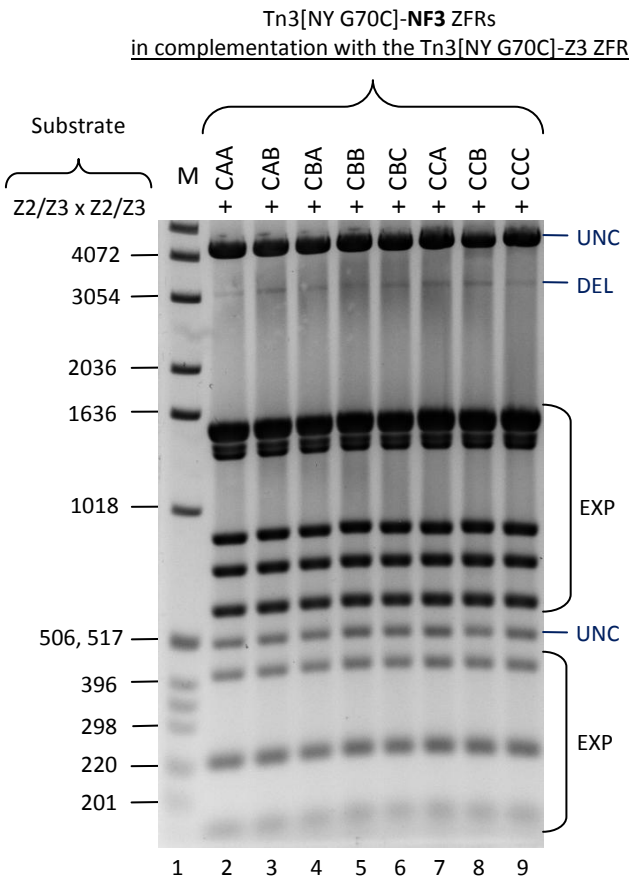


Figure 5-5: Results of complementation experiments involving ZFRs with non-specific ZFAs. The results are displayed across two gels. DNA ladder and fragment sizes are shown at the left of the images, while the origin of each of the experimental digests are indicated at the right (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression plasmids). Above the gel images, the substrate which was used for all experiments is indicated at the left, while directly above each gel track, the identity of each non-specific ZFA is given. **All experiments were carried out as complementations between the Tn3[NY G70C]-Z3 ZFR and Tn3[NY G70C]-NF3 ZFRs, where 'NF3' indicates a non-specific three-fingered ZFA.**

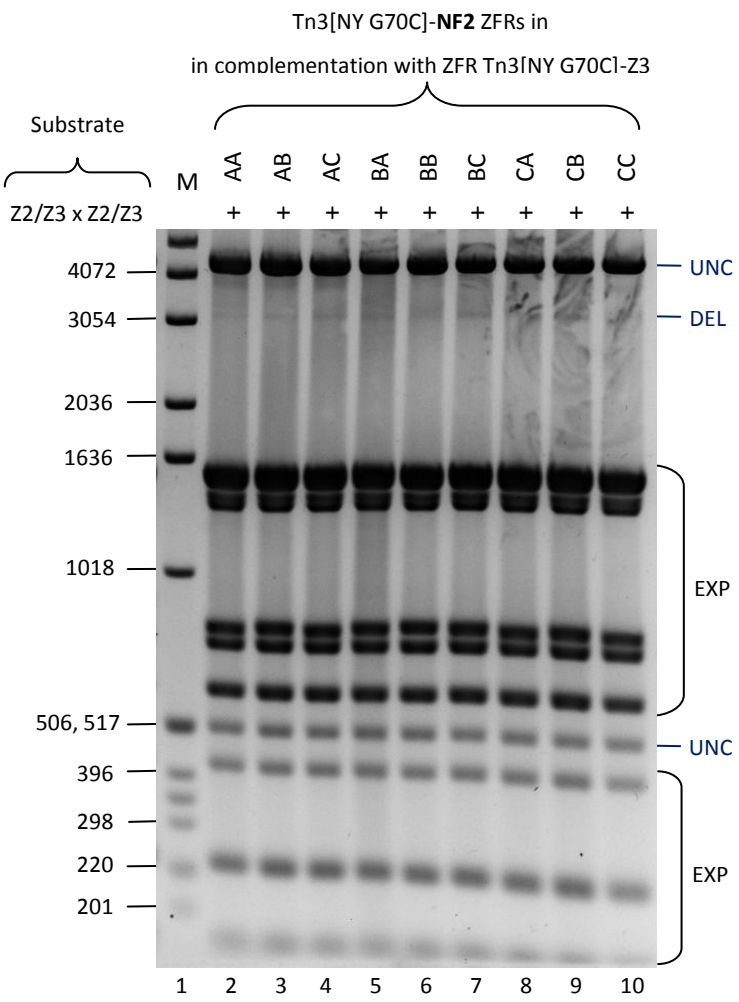


Figure 5-6: Results of complementation experiments involving two-fingered non-specific ZFAs. Marker ladder and fragment sizes are shown to the left of the gel image, while the origin of the plasmid fragments in each gel lane is indicated to the right of the image (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression plasmids). Above the image, and to left, is indicated the substrate which was used for all experiments. Directly above the gel image the identity of the non-specific ZFAs which were varied between experiments and represented by each gel track are indicated. **All experiments were complementations between ZFR Tn3[NY G70C]-Z3 and Tn3[NY G70C]-NF2, where 'NF2'indicates a non-specific two-fingered ZFA.**

5.3 RABD subunit + helper subunit complementations

5.3.1 Introduction

As discussed in Section 5.1.3, two ZFR subunits with differing binding affinity might be employed in a complementation system to produce a non-reversible reaction. The non-reversible reaction would take the form of two DNA binding-competent heterodimers bound to their Z-sites, which interact through a typical recombination reaction to produce two homodimers, one of which binds its Z-site competently, but the other of which does not. Because these product Z-sites are specific for their cognate homodimers, these product sites cannot be rearranged with one another to reverse the reaction, as one of the cognate homodimers would be non-functional. Additionally, this mode of non-reversible reaction would also likely enable an orientation specificity integration bias when used for integration reactions (see Section 5.1.3.5).

5.3.2 Recombination directionality bias recap

In the RABD + helper ZFR scheme, one of the subunits binds DNA with normal affinity, while the other subunit binds with much lower affinity. Because of the stabilizing negative energy of the extensive ZFR dimer interface during a heterodimeric subunit pairing, the subunits with normal binding affinity might act as 'helpers' to the subunits with a reduced affinity binding domain (RABD). However, *homodimeric* pairing of the subunits with RABD might not be possible because the total binding affinity of that homodimer might be insufficient to stabilize it on its cognate Z-site. Therefore, a significant difference between the binding capability of heterodimer and homodimer subunit pairs, might lead to a non-reversible reaction in the condition where the two heterodimers come together in the RHD tetramer configuration (as opposed to MHD tetramer configuration; see Figure 4-11), where recombination produces homodimers, one of which would be non-viable (Figure 5-1 A).

5.3.3 'Try and try again' integration orientation specificity recap

The alternative possible heterodimer-heterodimer arrangement to the RHD configuration, the MHD configuration, cannot produce a non-reversible reaction because the product of subunit rearrangement is the regeneration of the MHD configuration, identical to that which initiated the reaction. However, this may be a desirable property. Specifically, if one is interested in using the ZFR system to integrate DNA in an orientation-specific fashion (perhaps within a chromosome), this property of the RABD + helper ZFR scheme could allow the desired integration orientation to be stable, while the non-desired integration orientation would be highly unstable, producing a stochastic orientation-specific integration effect (see Figure 5-1 B).

5.3.4 Approaches to reducing ZFA binding affinity

Two straightforward approaches for identifying or generating ZFAs with reduced binding affinity are: screening of a large number of available ZFA for differences in binding affinity; and reduction of the number of fingers in the ZFA. Bacterial-2-hybrid (B2H) activity scores are provided by the Zinc Finger Consortium for all of the OPEN ZFAs (see Chapter 3, Section 3.2). These B2H scores are intended to indirectly indicate information about the binding activity of each domain. However, when the OPEN ZFAs used for this project were transferred into the ZFR system, the B2H activity scores (Table 3-1) did not

correlate entirely with the variability in ZFR activity levels observed in the 17 Hour Recombination Assay (Figure 3-4). It is certainly possible that by utilizing OPEN ZFAs with greater differences in their B2H activity scores, corresponding differences in ZFR activity would become apparent; however, within this current work, that approach was not explored.

The alternative method for producing ZFAs with reduced binding affinity is to reduce the number of finger modules within available functional arrays. There is a relationship between the number of fingers in a ZFA and its overall binding affinity, as each finger contributes to the strength of binding (though not necessarily to the same degree). Therefore, it is a straightforward assumption that reducing the number of fingers in a ZFA will reduce its binding affinity. The challenge to using this approach is that, without knowing the affinity contributions that each finger is making within the ZFA, it is impossible to predict how large a change in binding affinity will be produced by removing any given finger. Additionally, truncated ZFAs are not are not guaranteed to work at all, since there may be important synergistic interactions between the fingers within a ZFA that contribute to overall binding affinity and specificity. However, one potential advantage to this approach, within the context of this project's goals, is that reducing the number of fingers within a ZFA would also loosen the binding specificity requirements the ZFR system. This approach of reducing ZFA binding affinity, by truncating fingers from the array, is what was chosen for the work that follows.

5.3.5 Construction of truncated ZFAs

Zinc finger arrays rely on cooperative effects between adjacent fingers to produce their overall binding specificity and affinity. Because of these cooperative effects, it is sensible when constructing two-fingered ZFAs from known three-fingered ZFAs, to construct the two-fingered ZFA in such a way as to preserve adjacent pairs of fingers (that is, ZFAs consisting of finger 1 and finger 3 would be less likely work). In this work, a simple approach was taken to construct one-fingered zinc finger domains, and two-fingered ZFAs by truncating fingers from the Z1, Z2, Z3, and Z4 ZFAs (described in Chapter 3, Section 3.2).

During the initial design phase for the ZFAs, the DNA sequences coding for Z2, Z3, Z4, and Z5 had restriction sites incorporated into them, to allow finger 3 and finger 2 to be successively truncated from the three finger arrays, while preserving the ZFA C-terminal motif (see Section 3.2). This strategy allows for the production of zinc finger domains consisting of finger 1, or finger 1 and finger 2. The Tn3[NY G70C] catalytic domain was again chosen for this work because of its high activity and for comparability with other experiments in this work, which also use this domain. Thus, expression plasmids of ZFRs: Tn3[NY G70C]-Z2, Tn3[NY G70C]-Z3, Tn3[NY G70C]-Z4, Tn3[NY G70C]-Z5, were digested with *Ale*I to produce expression plasmids coding for Tn3[NY G70C]-Z2F2, Tn3[NY G70C]-Z3F2, Tn3[NY G70C]-Z4F2, Tn3[NY G70C]-Z5F2 (where F2 indicates a two finger array consisting of finger 1 and finger 2), and then subsequently digested with *Bsi*WI to produce expression plasmids coding for Tn3[NY G70C]-Z2F1, Tn3[NY G70C]-Z3F1, Tn3[NY G70C]-Z4F1, Tn3[NY G70C]-Z5F1 (where F1 indicates a binding domain containing only finger 1 of the parent array).

5.3.6 Experiment design

5.3.6.1 Overview

In order to test the capability of the new RABD ZFRs with one- and two-fingered ZFAs to produce non-reversible reactions, a complementation experiment was designed where the RABD ZFRs were used in conjunction with normal ZFRs, which act as 'helpers' to the weakly bound RABD subunits when heterodimers are formed. This experiment primarily relied upon 2MutHetDim-DR substrates, which are able to directly detect non-reversible reactions in the form of abnormal accumulation of inversion product if the AHD configuration (homodimers) is inactive. However, 2MutHomDim substrates were also employed to confirm that AHD configuration were inactive (more below).

5.3.6.2 Tetramer configurations tested

The two tetramer subunit configurations, which form the basis of this approach to a non-reversible reaction, are the RHD configuration, and the AHD configuration. The RHD configuration and the AHD configuration are interchangeable products of subunit rotation, which is to say, that after subunit rotation, the RHD configuration will transform into the AHD configuration, and vice versa (Figure 4-11). An additional tetramer subunit configuration is also possible, called the MHD configuration. The MHD configuration

differs from the two other tetramer subunit configurations in that subunit rotation results in the production of another MHD configuration tetramer (Figure 4-11). Because the starting configuration and product configurations are the same, the MHD configuration is not capable of producing a non-reversible reaction.

5.3.6.3 Substrate-mediated and enzyme-mediated reaction non-reversibility

Before continuing, it is important to lay out a fundamental aspect of ZFR recombination reaction reversibility. There are two basic ways in which a recombination reaction in the 17 Hour Recombination Assay could be non-reversible. The first type of non-reversibility is **substrate-mediated**. A non-reversible reaction that is substrate-mediated is non-reversible because of the form that the substrate product of the reaction takes, rather than being the result of any properties of the enzymes themselves. An inversion product is usually capable of being converted into either an excision product or back into the unchanged substrate form. Therefore, a reaction that produces inversion product allows for reversible reactions. Conversely, an excision product is not capable of being converted into anything else, because the segment of DNA that is excised from the substrate plasmid does not possess an origin of replication and is quickly lost as the *E. coli* hosts of the reaction divide.

The second type of non-reversibility is **enzyme-mediated**. A non-reversible reaction that is enzyme-mediated, is non-reversible because of intrinsic properties of the ZFRs used, and the non-reversibility of the reaction is independent of substrate product form. However, when considering the outcome when both substrate-mediated and enzyme-mediated reaction reversibility potentials might be in play, in either case, reaction non-reversibility dominates over reversibility.

5.3.6.4 Substrate selection

The 2MutHetDim-DR substrate type was chosen as the primary information production substrate type for this experiment, because (unlike 2MutHetDim-IR substrates) when using this substrate type, the recombination product of reactions involving an RDH configuration is *the inversion product*. Ordinarily, in the 17 Hour Recombination Assay, inversion product can be further converted into either excision product, or back into the unchanged substrate form. The excision product on the other hand, is a true end product that cannot be further converted to anything. Therefore,

when looking at the recombination activity results on a gel, there is usually a great preponderance of excision product over the inversion product. However, if the RABD ZFRs, in complementation with helper ZFRs, are able to produce a non-reversible reaction, then the inversion product would conspicuously accumulate. Likewise, when using 2MutHetDim-DR substrates, the product of a reaction involving an MHD configuration, which is not capable of producing a non-reversible reaction, would be the excision product, which is inherently non-reversible. Thus, the 2MutHetDim-DR substrate type matches the substrate-mediated reversible reaction with the enzyme-mediated non-reversible reaction, and the substrate-mediated non-reversible reaction with the enzyme-mediated reversible reaction, so that both reaction products are potentially stable and are not further converted to anything else.

In order to provide an additional layer of information about the activity of tetramer subunit configurations, the 2MutHomDim substrate was also employed for this experiment. The 2MutHomDim substrates bind two alternate homodimers, and are therefore able to provide information about whether the AHD configuration is active. Since the AHD configuration is the product configuration of a recombination reaction beginning with the RHD configuration, if the AHD configuration is determined to be inactive, while the RHD configuration is determined to be active, this strongly indicates potential for a non-reversible reaction. Therefore, if recombination activity—and particularly in the form of inversion product (see above)—is observed in complementation experiments utilizing the 2MutHetDim-DR substrates, while no activity is observed in complementation experiments utilizing the 2MutHomDim substrates, this would provide additional support for a conclusion that the ZFR complementation pair was capable of producing a non-reversible reaction.

1MutHomDim substrates were also used in this experiment in order to provide information about the base activity level of the RABD ZFRs. The 1MutHomDim substrates bind homodimers comprised of only one type of subunit, detecting the activity of a tetramers consisting of four identical subunits. Therefore, in addition to the complementation experiments, the RABD ZFRs were also tested on cognate 1MutHomDim substrates in order to test their activity level when used without the normal ZFR subunits acting as helpers.

It should be noted that the lack of inclusion of a 2MutHetDim-IR substrate proved to be a major deficiency in the experimental design. In spite of the order in which the experiments are laid out in this thesis, much of the work was being conducted in parallel, and at the point where this experiment was undertaken, insufficient data analysis and results collection had been carried out to accurately characterize the parameters for the SA effect. This experiment requires the detection of reactions where the RHD configuration is active. The SA effect is strongest on substrates that contain homodimer Z-sites for active ZFRs. All products of RHD reactions produce homodimer Z-sites. In light of the current understanding of the SA effect (Sections 3.10.5 and 4.5.8), it becomes apparent that only by using the variety of 2MutHetDim-IR substrate that place the RABD subunits on the homodimer Z-site of the excision product, can a result be produced for RHD reactions that is unaffected by the SA effect. The RABD subunits would presumably not cause the SA effect on a homodimer Z-site because they are unable to bind to it. Unfortunately, at the point where the implications of the SA effect for this experiment became fully appreciated, there was no longer sufficient time to re-run the experiment with the inclusion of the appropriate 2MutHetDim-IR substrate

5.3.6.5 Catalytic domain and ZFRs

The Tn3[NY G70C] mutant catalytic domain was again chosen for these experiments, both because it is highly active (see Chapter 4, Figure 4-20) and because it was used in many of the other experiments throughout this project, making the observed results comparable to others. Where the catalytic domain was attached to the four two-fingered ZFAs, they were named Tn3[NY G70C]-Z2F2, Tn3[NY G70C]-Z3F2, Tn3[NY G70C]-Z4F2, and Tn3[NY G70C]-Z5F2 , respectively. Where the catalytic domain was attached to the four one-fingered ZFAs, they were named Tn3[NY G70C]-Z2F1, Tn3[NY G70C]-Z3F1, Tn3[NY G70C]-Z4F1, and Tn3[NY G70C]-Z5F1 , respectively. The Tn3[NY G70C]-Z2 ZFR was used as the helper ZFR in these experiments, except in the complementations involving the Tn3[NY G70C]-Z2F2, and Tn3[NY G70C]-Z2F1 RABD ZFRs, where the Tn3[NY G70C]-Z3 ZFR was used instead to avoid the obvious problem of cross-specificity between the Z2 and Z2F2 or Z2F1 binding domains. Applicable subunit-specific varieties (Z2, Z3, Z4, and Z5) of the 1MutHomDim, 2MutHomDim, and 2MutHetDim-DR substrates were used for the experiment. Thus, a 17 Hour Resolution Assay was conducted using the above materials.

5.3.7 Results: two-fingered RABD + helper ZFRs

The results from the 17 Hour Resolution Assay measuring activity of two-fingered RABD + helper ZFR complementations on the 2MutHomDim and 2MutHetDim-IR substrates (shown in Figure 5-7) suggest that they MHD configurations are active, while AHD configuration are not. However, the RHD configuration either destroys DNA, or the SA effect has confounded the result.

The first thing that can be observed is that for complementations on the 1MutHomDim substrates (lanes 5–8), no recombination products are observed and the unchanged substrate signals are approximately equal to that of the negative control (lane 3), indicating that homodimers comprised of two-fingered RABD ZFRs are not able to form active tetramers with other homodimers comprised of two-fingered RABD ZFRs.

The second thing that can be observed is that there are no recombination products from the RABD + helper complementations on the 2MutHomDim substrates, indicating that the AHD configuration is inactive. While almost no recombination product can be seen in any of the complementations on the 2MutHomDim substrates (lanes 10–13), the unchanged substrate signals in these lanes are considerably attenuated when compared with that of either the negative control (lane 3), or the complementations on the 1MutHomDim substrates (lanes 5–8). These attenuated signals for the unchanged substrate product are consistent with the expected activity of the SA effect in plasmids with homodimer Z-sites for active ZFRs. The level of SA effect on the 2MutHomDim substrate compared with the 1MutHomDim substrates allows the observer to gauge the level of SA effect operating within the experiment on homodimer Z-sites binding the helper ZFRs. Since no recombination activity has taken place on the 2MutHomDim substrates it can be inferred that the RABD ZFRs do not bind their homodimer Z-sites well.

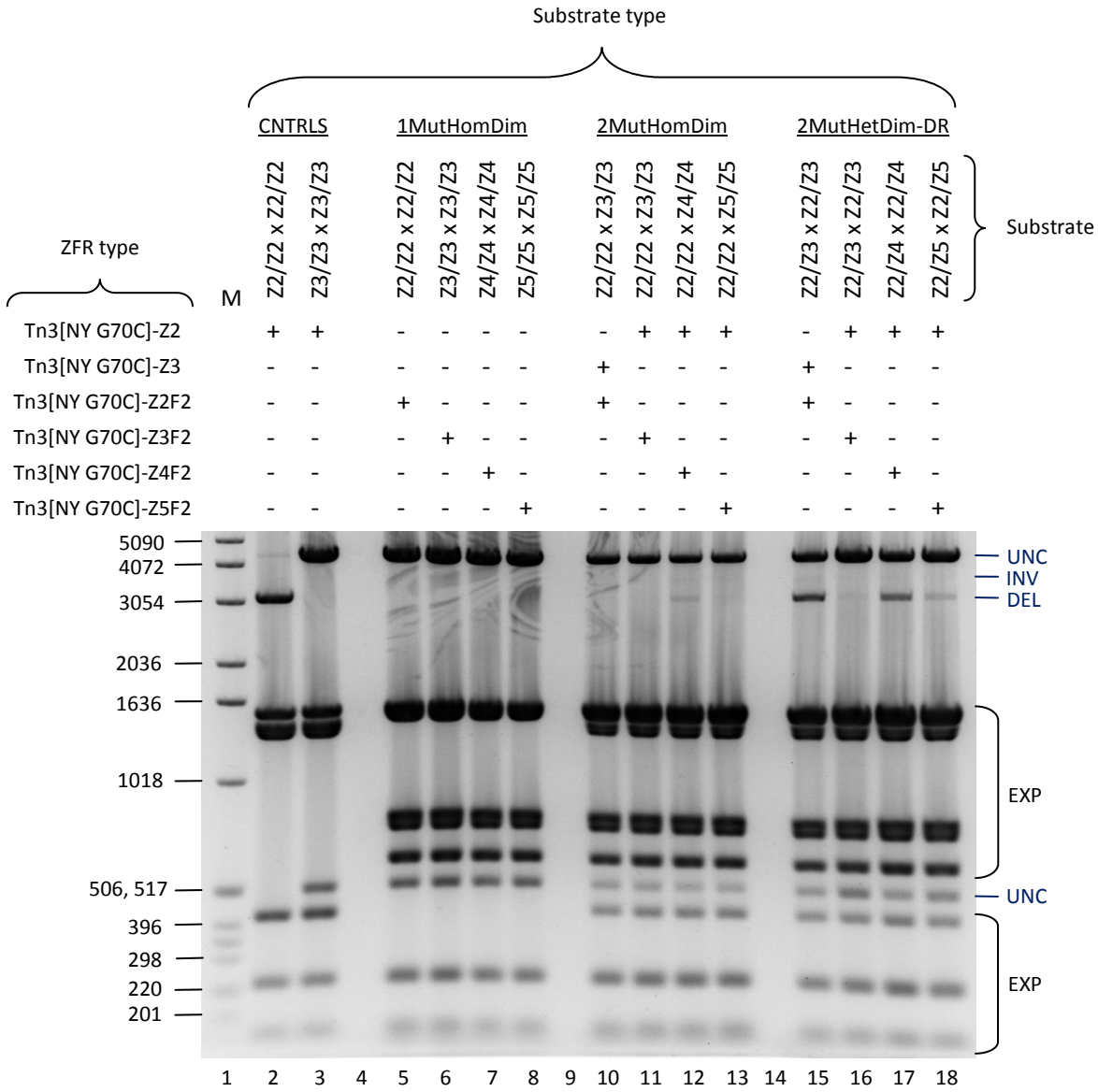


Figure 5-7: Complementation experiment to test ZFR-F2s for non-reversible reaction capability. The figure shows the results of a 17 Hour Resolution Assay involving complementation of ZFR-F2 with normal ZFRs on various substrate types where differential activity indirectly indicates non-reversible reaction potential. 1KB ladder markers are shown to the left of the image, while the origin of each restriction fragment is indicated to the right (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression plasmids). Above the gel image, a grid is shown which indicates ZFRs and substrates complementations which are shown in each track.

Finally, the results of the complementations on the 2MutHetDim-DR substrates (lanes 15–18) show that in three instances (lanes 15, 17 and 18) the excision product is being produced, while only trace levels of inversion product can be observed. The excision products from the 2MutHetDim-DR substrates indicate the activity of the MHD configuration, while inversion products indicate the activity of the RHD configuration (the MHD and RHD configurations are the two possible heterodimer-based tetramer configurations). The RABD + helper ZFR complementations are not expected to produce any dimer-dimer orientation specificity, and this is probably not responsible for the

difference observed between inversion and excision product signals. There are two explanations for the ostensible difference in the activity of the MHD and RHD configurations. The first explanation is that since the inversion product of the 2MutHetDim-DR substrates contain homodimer Z-sites, it is possible that the SA effect has destroyed these plasmids. The second explanation is that the RHD configuration is active, but the RABD subunits are unable to rejoin the DNA after subunit rotation, perhaps because their binding domains disassociate from the DNA. Although, it should be kept mind that if the DNA is not re-ligated after subunit rotation the ZFR would remain covalently attached to the DNA. There is no way to discriminate between these two possibilities by making deductions from the current results.

Since the results on the 2MutHomDim substrates clearly indicate that the AHD configuration is inactive, the appearance of excision product likely indicates that both heterodimer tetramer configurations (MHD and RHD) are active, as a dimer-dimer orientation specificity effect would be unlikely to arise using the RABD + helper ZFR system. Therefore, the question of whether the SA effect has destroyed the inversion product or whether the RABD subunits are unable to re-ligate the DNA is an important one. Fortunately, this could be easily determined by running the experiment again using the variety 2MutHetDim-IR substrates that place the homodimer Z-sites for the RABD subunits on the excision product (see Section 5.3.6.4). Unfortunately, however, there was no time available to carry out this additional experiment.

5.3.8 Result: one-fingered RABD + helper ZFRs

The results from the complementation experiment involving one-fingered + helper ZFRs (seen in Figure 5-8) show only trace amounts of recombination products across the lanes of the gel. Curiously, the highest level of trace activity appear in the complementations on the 1MutHomDim substrates (lanes 5–8), while the lowest level of trace activity appears in complementations on the 2MutHetDim-DR substrates (lanes 15–18). This pattern of activity is opposite to that observed for the two-fingered RABD + helper complementations in Section 5.3.7. However, the recombination activity from this experiment is so minimal that it is hard to observe unless the gel image is viewed on a computer screen. The one-fingered RABD + helper ZFR pairs do not appear capable of producing useful recombination reactions.

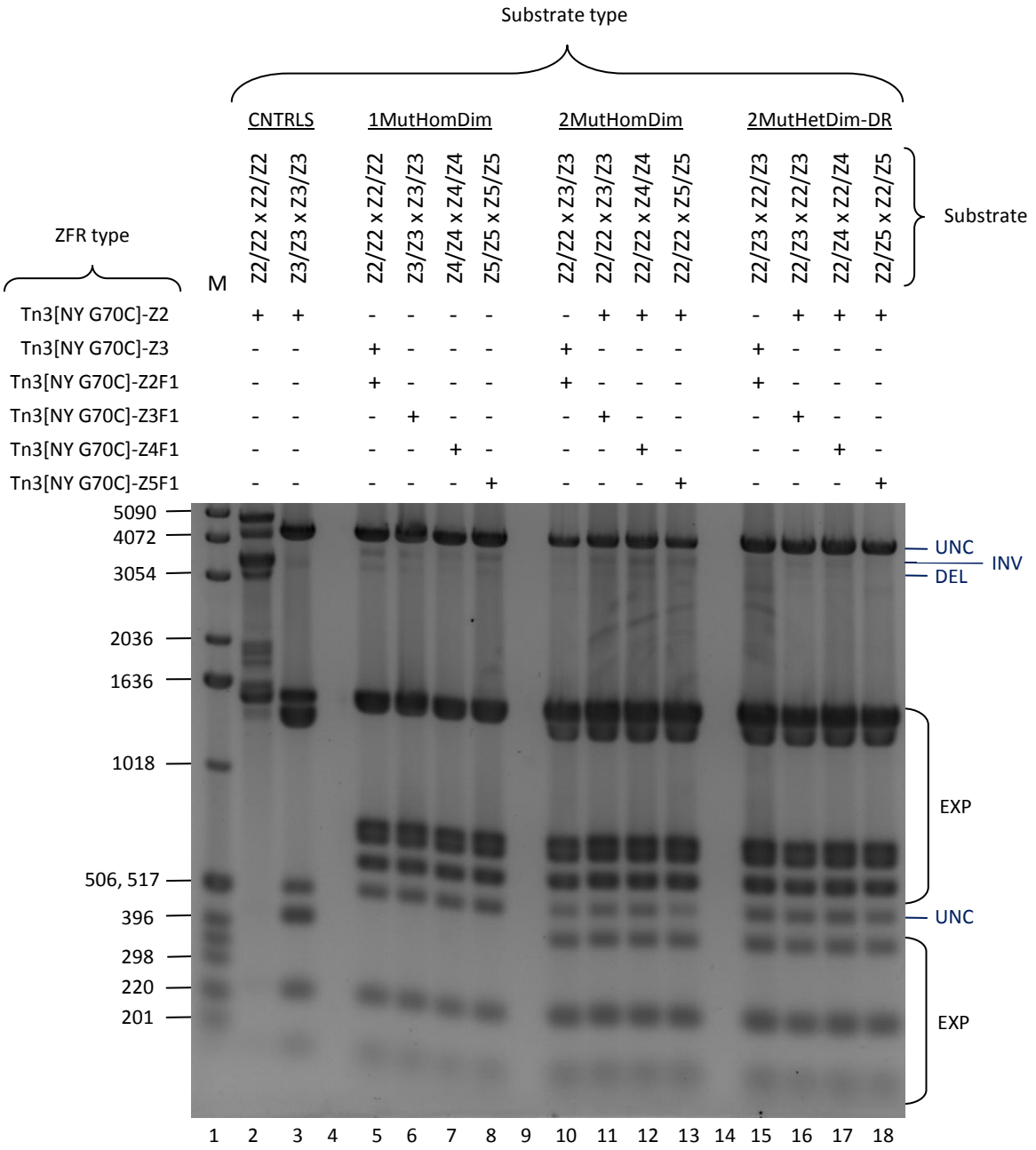


Figure 5-8: Complementation experiment to test ZFR-F1s for non-reversible reaction capability. The figure shows the results of a 17 Hour Resolution Assay involving complementation of ZFR-F1 with normal ZFRs on various substrate types where differential activity indirectly indicates non-reversible reaction potential. 1KB ladder markers are shown to the left of the image, while the origin of each restriction fragment is indicated to the right (UNC = unchanged substrate, INV = inversion product, DEL = deletion (excision) product, and EXP = expression plasmids). Above the gel image, a grid is shown which indicates ZFRs and substrates complementations which are shown in each track. The restriction digest which is designed to cleave the expression plasmids into small enough fragments that they do no occlude the substrate product signals, did not run to completion for the positive control (shown in lane 2). However, the band of interest, representing the deletion product of the substrate, is still visible in spite of the partial digest of the expression plasmids seen in that gel track.

5.4 Using the Tn3 resolvase HTH domain a non-site-specific RABD

5.4.1 Background

It was noticed during an initial set of complementation experiments (data not shown) that relied on the Zif268 ZFA and the Tn3 resolvase HTH (H1, so named to give it a comparable nomenclature to the ZFA) domain for targeting—that is, complementation experiments that paired a ZFR with a mutant Tn3 resolvase (MTR)—that the MTR was cross-reactive for the ZFR half of the hybrid binding site. This cross-reactivity was particularly apparent on heterodimer binding sites, where a properly bound Tn3[NY G70C]-HTH MTR could effectively co-opt another MTR subunit to incorrectly bind the ZFR half of the heterodimer binding site, forming a homodimer. In a control experiment where only the MTR was expressed, using a MRT/ZFR heterodimer-binding substrate, a total conversion of substrate to recombination products was observed. Although this cross-reactivity of the MTR for the Tn3-Zif268 Z-site revealed that complementation experiment design to be unreliable, it also raised the question of whether the H1 domain could be used intentionally as a non-site-specific binding domain. It was also considered that the H1 domain binding in a non-site-specific fashion would probably making low affinity contacts, and thus, may also simultaneously function as an RABD. As discussed in Section 5.3.2, RABD + helper complementations may be used to produce non-reversible reactions.

One question that arises is whether a system such as this could have utility in a mammalian cells system. It has been demonstrated that a properly bound MTR is capable of co-opting other MTR subunits to bind to sites that the H1 domain of that subunit does not bind site-specifically (see above). This implies that if the MTR + ZFR complementation were performed in a mammalian cell system, there might be some risk of the MTRs performing unintended recombination reactions at unintended binding sites. However, the discovery in Chapter 4 that inactive ZFRs can be activated by complementation with active ZFRs (the sleepy + active ZFR strategy), provides a way to eliminate that potential problem. The MTRs could be made into sleepy subunits that would have no activity on their own. Therefore, it was deemed worthwhile to explore the possibility of using MTRs to act as non-site-specific RABDs in complementation with helper ZFRs.

5.4.2 Experiment design

5.4.2.1 Overview

An RABD + helper subunit complementation experiment was designed to test the capability of the H1 domain to function as a non-site-specific RABD. The experiment design, where the H1 domain is used intentionally in this fashion, differs from the aforementioned preliminary experiment where the MTRs non-specific activity was first noticed, described in the previous section. Instead of a properly bound MTR co-opting another MTR to improperly form a homodimer at an MTR/ZFR hybrid binding site, in this experiment a ZFR bound at a Z-site attempts to co-opt the MTR to bind the other half of the full Z-site. The experiment was designed to test both non-specific binding capability of the H1 RABD, and the ability of the RABD + helper complementation to produce a non-reversible reaction.

5.4.2.2 Catalytic and binding domains for RABD + helper

The Tn3[NY G70C] catalytic domain was used for both the RABD and helper subunits because it is highly active, and also for comparability with the results of other experiments that utilized this domain. The Z3 ZFA was used for the helper subunit, because the Z3 domain imparts higher activity to ZFRs than any of the other available ZFAs (see Chapter 3, Section 3.5), giving the complementation experiment the best possible chance of success. The H1 domain was used for the RABD subunit.

5.4.2.3 Substrates selection

Non-reversible reactions may be generated as the result of differential activity between the RHD and AHD configurations in a RABD + helper complementation (Section 5.3.2; Figure 4-11; and Figure 5-1 A). In order to demonstrate a non-reversible reaction, 2MutHetDim-DR substrates were selected for this experiment. If the RHD and MHD configurations are active then the complementations on the 2MutHetDim-DR substrate will produce inversion and excision products, respectively (see Table 4-1). However, if the AHD configuration is also active, then the inversion products will be readily converted by secondary reactions back into unchanged substrates or into excision products. In reactions that go to completion, if all three tetramer configurations are active, the substrate will be completely converted to the stochastically favoured excision end product (see Figure 4-14). However, if the RHD and MHD configurations are active, while

the AHD configuration is inactive, the inversion product will accumulate and should equal the excision product at the end of the reaction. However, the SA effect must also be taken into consideration when interpreting the results as the inversion product from 2MutHetDim-DR substrate contains two homodimer Z-sites, and thus, its signal strength may be reduced.

2MutHomDim substrates were also used to confirm that the AHD configuration is not active. If the ADH configuration is active, then reactions on 2MutHomDim substrates will produce both inversion and excision products (see Table 4-1 and Figure 4-13). Conversely, if the AHD configuration is not active, then no recombination products will be produced from the 2MutHomDim substrates.

In order to demonstrate the H1 domain could be used on a non-site-specific fashion, this experiment utilized a set of 2MutHetDim-DR and 2MutHomDim substrates that contained binding sites for the Tn3-Z3 ZFR in conjunction with Tn3-Z1, Tn3-Z2, Tn3-Z4, or Tn3-Z5 ZFR binding sites. All four Z-site varieties, which were available in addition to the Tn3-Z3 Z-site, were used to test the maximum range of non-specific binding tolerance possible with available materials. It should be noted, that all of the Z-sites retain 2 bp of the 6 bp H1 binding site in their spacer sequences that separate the Tn3 site I core sequence from the 11 bp (including the 3′ F1 overlap position and the 5′ extended position, see Section 1.7.1.2) sequence of the ZFA sites (Figure 5-9). Although, it should be noted that the 5′ extended position of the ZFA sites, in all cases, of has the same base pair as either the left or right half-site of site I, at this position. However, even within this 3 bp region (the 2 bp spacer + the 1 bp that incidentally overlaps) that is common to site I, the left and right half-site versions of the site I sequences over these positions, are only the same at one position within the 3 bp sequence (Figure 5-9)—indicating a degree of binding tolerance by the H1 domain.

For the controls, the relevant portion of the preliminary experiment descried in Section 5.4.1 (data not shown) was essentially replicated. A Tn3-Z1 1MutHomDim substrate was used as a negative control, while a MTR/Tn3-Z1 2MutHetDim-DR substrate was used as the positive control. The experiment was conducted using the 17 Hour Recombination Assay.

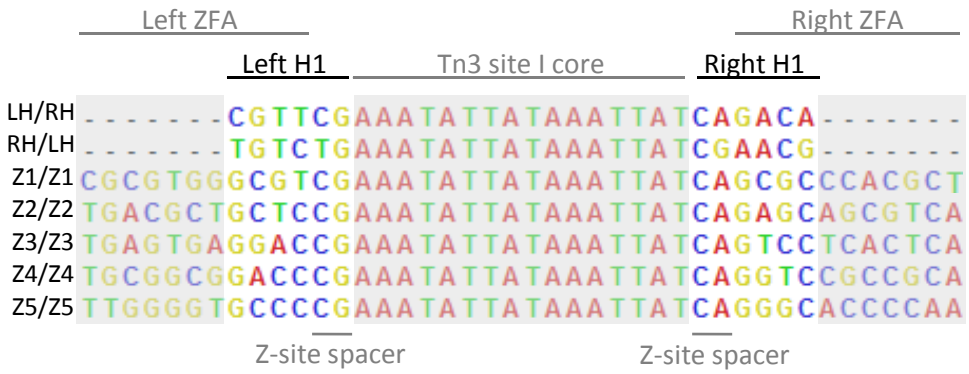


Figure 5-9: Alignment of binding sites sequences. The figure shows an alignment of binding sites in order to display the sequences which the H1 domain tolerates and is being tested on. The 11 bp ZFA sites, 6 bp H1 sites, 16 bp Tn3 site I core, and 2 bp Z-site spacers are labelled above and below the alignment. LH/RH site is Tn3 site I. The RH/LH is site I with the left and right H1 binding sites swapped. The RH/LH site is not included within the experiment, but is provided in the alignment to show additional sequence variations tolerated by the H1 domain. Within the experiment, the H1 domain is only tested against the left half of the Z-sites.

5.4.3 Results

5.4.3.1 Overview

The results of this experiment, which are displayed in Figure 5-10, show the H1 domain is able to act as a non-specific, or at least reduced specificity, binding domain, and that these RABD + helper complementations indeed produce an extremely robust non-reversible reaction effect. The negative control, shown in lane 2, shows that the Tn3[NY G70C]-H1 MTR, alone, is completely inactive on a Tn3-Z1 Z-site. The positive control, shown in lane 3, tested the ability of the Tn3[NY G70C]-H1 MTR, alone, to recombine H1/Z1 x H1/Z1 2MutHetDim-DR substrate. The MTR in the positive control is provided one cognate half-site, and it is co-opting another MTR subunit to bind the other (Tn3-Z1) half-site. This positive control replicates the result from a preliminary experiment described in Section 5.4.1 (data not shown), and demonstrates that the MRT, alone, can recombine the substrate when provided with only one cognate half-site. The positive control also demonstrates complete conversion of the substrate plasmid into recombination products (inversion product and excision product).

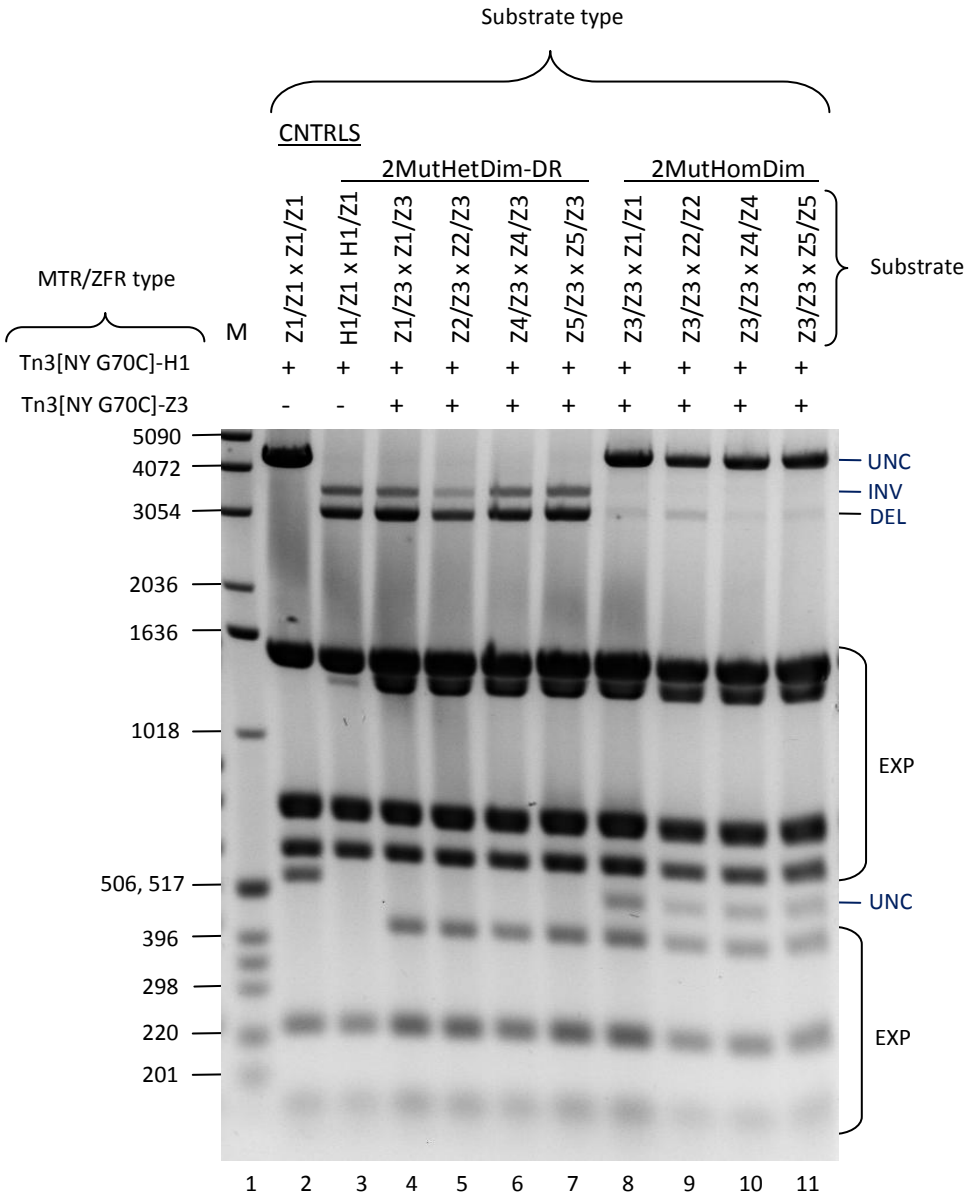


Figure 5-10: Complementation experiments with the ZFR acting as a helper subunit to an MTR bound to non-cognate sites. The figure shows the results of the complementation experiments using the MTR and ZFR on 2MutHetDim-DR and 2MutHomDim substrates. In gel track 1, the 1 kb marker ladder is show with its fragment sizes indicated to the left of the image. At the right of the image are indicators which describe the origin of the plasmid fragments, where 'UNC' indicates un-recombined substrate, 'INV' indicates substrate inversion product, 'DEL' indicates substrate deletion (excision) product, and 'EXP' indicates expression plasmid fragments. Above the gel image and to the left, the MTR and ZFR types used in the experiment are shown, while directly above the gel the substrate and substrate type used in each experiment are shown. Tracks 2 and 3 contain negative and positive controls, respectively, for which the results were known from a previous experiment (data not shown). Tracks 3–7 show the complementation experiments using the 2MutHetDim-DR substrates which test the ability of a ZFR subunit bound to its cognate Z-half-site to cooperate with a MRT bound to a non-cognate Z-half-site. (While the control experiment is shown in track 2 also uses a 2MutHetDim-DR substrate, this experiment is fundamentally different from the experiments shown in tracks 3–7 in that its 2MutHetDim-DR substrate contains cognate binding sites for the MTR, while the experiments in tracks 3–7 contain no MTR cognate sites.) Tracks 8–11 show the complementation experiments using the 2MutHomDim substrates, which test the ability of the ZFR homodimer bound to a cognate Z-site to cooperate with a MTR homodimer bound non-cognately to a Z-site. The Z1 designation indicates the Zif268 ZFA and its binding site, while the H1 designation indicates the Tn3 resolvase HTH domain and its binding site.

5.4.3.2 Non-site-specific capability of the HT domain

The tracks showing the RABD + helper complementation on the 2MutHetDim-DR substrates containing only Z-sites (lanes 4–7) show full levels of recombination. These results indicate that the MTR is able to act as a highly efficient non-specific, or reduced specificity, heterodimer partner to a ZFR. Total conversion of substrate to products takes place even when the MTR is acting on sites that contain all possible base changes at two of the three positions in the H1 binding site that do not vary between the left and right sides of site I (lanes 4, 6, and 7; and Figure 5-9). Additionally, total conversion of substrate plasmid to products is also observed when the H1 domain is binding at the Z5 site, which contains only 2 of 6 bp in common with *either* the left or right H1 binding sites of site I (lane 7; and Figure 5-9). However, it should be noted that all of the Z-sites contain 3 bp in common with the HT sites of site I when considering both the HT sites at once (i.e. when considering the variability from HT sites in a non-contiguous fashion; Figure 5-9).

5.4.3.3 Non-reversible reaction capability of the RABD + helper subunit complementations

The inversion product signal in experiments utilizing the 2MutHetDim-DR (including the positive control) is conspicuously strong (lanes 4–7). In no other experiments that the author has witnessed, has the inversion product band been so pronounced relative to the excision product band. The fact that the inversion band is even still present after the recombination reaction has gone to competition (i.e. after total conversion of substrate to products) is, in itself, a completely unique result. Inversion products are subject to secondary reactions, which results in the inversion products existing only transiently during reactions that go to completion. The only stable end product of recombination reactions is the excision product. Thus, when a typical reaction goes to completion, all of the substrate is inevitably converted to excision product, both directly, and via secondary reactions on the inversion product. However, if an RABD + helper complementation is successful in producing an RHD configuration that is active, and an AHD configuration that is inactive, then such reactions carried out on 2MutHetDim-DR substrates will produce a stable inversion product that is no longer subject to secondary reactions. The presence of a robust inversion product signal in the absence of the unchanged substrate signal (indicating the reaction has gone to completion) in lanes 3–7, directly demonstrates that a non-reversible reaction has been achieved using the RABD + helper subunit strategy.

The presence of excision product in lanes 3–7, indicates the MHD configuration is also active. Since no dimer-dimer orientation specificity effect is expected from RABD + helper complementations, the RHD and MHD configuration would be expected to product equal levels of activity. Equal activity of the RHD and MHD configurations on a 2MutHetDim-DR substrate would be expected to produce equal rates of production of the inversion product and excision product, respectively. The excision product signals in lanes 3–7 are stronger than the inversion product signals. However, the inversion product from the 2MutHetDim-DR substrate contains two homodimer Z-sites, while the excision product contains a heterodimer Z-site, making the inversion product subject to a much stronger SA effect (see Table 4-3). The difference in signal strength observed between the inversion product and excision product in lanes 3–7, is almost certainly the result of the SA effect, indicating there is probably no difference between the RHD and MHD recombination rates.

The results of the complementation experiments that were carried out on the 2MutHomDim substrates (lanes 8–11) also fully support the conclusion that an enzyme-mediated non-reversible reaction has been achieved in these complementation experiments. The 2MutHomDim substrates test the ability of a substrate-bound MTR homodimer to cooperate with a substrate-bound ZFR homodimer in the AHD configuration. The lack of significant reaction products from the reaction on the 2MutHomDim substrates confirms the AHD configuration in these RABD + helper complementations is inactive. Because the AHD configuration is the subunit configuration state produced after subunit rotation of the RHD configuration (see Figure 4-11), the presence of the inversion product from the 2MutHetDim-DR substrates (lanes 3–7), and the lack of significant reaction products from the 2MutHomDim (lanes 8–11) provides additional evidence of a non-reversible reaction using the RABD + helper subunit pair.

Small amounts of excision product can be seen in the in the lanes containing the 2MutHomDim substrates. This indicates the non-reversibility of the RABD + helper reactions are not perfectly tight, at least not on the 2MutHomDim substrates where the HT domain binds the Z1 and Z2 sites. If the MTR homodimer is still able to interact productively with the ZFR homodimers at a low level, this problem might be overcome by incorporating one of the sleepy subunit mutation sets described in Chapter 4 (such as those from Section 4.4), which are also capable of producing a recombination reaction

directionality bias. Additionally, transforming the RABD subunits into sleepy RABD subunits would reduce the potential for off-target reactions in a genomic integration context.

5.4.3.4 SA effect and H1 binding on Z2

It might also be noted that the SA effect appears strongest on the 2MutHetDim-DR substrate containing Z2, while the unchanged substrate signal of the 2MutHomDim substrate also appears to be relatively more affected. There are three bases conserved between the H1 binding sites on the left and right side of Tn3 site I. The Z2 site contains a thymine that is common with one of these conserved positions and that none of the other ZFA sites contain. This likely indicates that the SA effect for substrates containing Z2 is increased because of increased binding of the H1 domain.

5.4.3.5 Stochastic integration orientation specificity

Finally, although a conclusive demonstration would require an intermolecular integration reaction, it can be assumed that this RABD + helper complementation would be capable of achieving an integration orientation specificity effect via the stochastic 'try and try gain' pathway described in Section 5.1.3.5.

5.5 Conclusions

5.5.1 Complementations using ZFRs with non-specific ZFAs

The experiments utilizing the two- and three-fingered non-specific RABDs in complementation with normal ZFRs were not successful, and no significant amount of recombination was observed in any of the reactions. Among the 36 ZFAs that were produced, one ZFA (CAC) may have been active. No transformants were recovered after it was ligated to an active ZFR catalytic domain (Tn3[NY G70C]), indicating the ZFR might have been toxic to the cells. Further attempts to recover this ZFR were not pursued because it was assumed that if the ZFA did impart the ZFR with enough binding activity to produce toxicity, it would not be useful as an RABD. However, given that the other ZFAs did not function even in RABD + helper ZFR complementations, it might be worth exploring whether the CAC ZFA could be used purely as a non-specific binding domain (i.e. not as a RABD domain as well). If the CAC domain does have enough binding affinity to make active ZFRs toxic, then combining it with a sleepy catalytic domain (Chapter 4),

would obviate that problem. The sleepy catalytic domain would allow the CAC ZFA to be used in sleepy + active ZFR complementations, which could test whether the CAC could be used to relax the site-specificity parameters of the ZFR system.

A better approach may have been to attempt a similar type of experiment but setting for previously characterized ZFAs with *reduced* specificity ZFAs instead of attempting to generate entirely new non-specific ZFAs. Additionally, because the zinc finger monomers that were used had never been proven to be completely non-specific (see Table 5-2), it may have been interesting to test the non-specific ZFAs at sites that the monomers had previously been demonstrated to bind to, in order to determine whether these ZFA were, in fact, reduced specificity ZFAs.

5.5.2 Truncated ZFAs for RABD + helper complementations

Some of the experiments utilizing truncated versions of the Z2, Z3, Z4, and Z4 ZFAs as RABDs for RABD + helper complementations produced interesting results. While all of the one-fingered RABDs showed no activity on any substrates tested, the Z2F2, Z4F2, and Z5F2 domains did impart their ZFRs with significant recombination activity on the 2MutHetDim-DR substrates. The experiments were designed to test the RABD + helper ZFR pairs for their ability to produce a non-reversible reaction. Non-reversible reactions on from RABD + helper complementations would arise as the result of RHD configurations being active while the AHD configuration is not. The AHD configuration is the tetramer configuration produced after RHD subunit rotation. Thus, if the RHD configuration is active while the AHD configuration is not, then the reaction is irreversible.

The lack of recombination products produced by RABD + helper complementations on the 2MutHomDim substrates clearly indicates that the AHD configuration is inactive in these reactions. The presence of excision product on the 2MutHetDim-DR substrates indicates that the MHD configuration was active. Because no dimer-dimer orientation specific effect was expected from RABD + helper complementations, the RHD configuration should also have been active. However, an active RHD configuration should produce inversion products on the 2MutHetDim-DR substrate, and there were no inversion products present in any of the reactions that produced excision products. One explanation for this is that the SA effect destroyed the

inversion product signals. The inversion product from the 2MutHetDim-DR substrates would be expected to carry two homodimer Z-sites making them subject to the strongest level of SA effect. However, another possible explanation is that these RABD subunits are unable to rejoin the DNA, possibly because the binding domains of the RABD homodimers that are generated disassociate from the DNA before the DNA is re-ligated (in spite of the fact the catalytic modules are still covalently attached to the DNA).

A simple experiment could be conducted to test whether the lack of inversion products from the 2MutHetDim-DR substrates was caused by the SA effect or was the result of the RABD ZFRs being unable to re-ligate the DNA. Active RHD configurations produce excision products from reactions on the 2MutHetDim-IR substrates. If the RABD + helper complementations were performed on the 2MutHetDim-IR substrate variety that place the homodimer Z-site for the RABD subunits on the excision product, then the production of excision product would indicate that the RHD configuration was active and the RABD subunits are able to re-ligate the DNA. Additionally, the SA effect should be minimal under this condition as the RABD subunits should not be able to bind the excision product well. If the RABD subunits are able to re-ligate the DNA, this would indicate that the RABD + helper strategy does indeed produce a non-reversible reaction. The addition of the 2MutHetDim-IR substrates to the experiment was not considered because, at the time the experiment was conducted, the SA effect had not been fully characterized. A further experiment was not carried out because of insufficient available time.

5.5.3 H1 domain as a non-site-specific RABD for RABD + helper subunit complementations

5.5.3.1 RABD + helper complementations produced a non-reversible reaction

The experiment utilizing the MTR as non-site-specific RABD, in RABD + helper complementations, appears to have been entirely successful in producing a non-reversible reaction. The experiment demonstrated an accumulation of inversion products from a 2MutHomDim-DR substrate in a reaction that had gone to completion, which would not be possible without a non-reversible reaction. Ordinarily, the inversion product in ZFR reactions only exist transiently as they are further converted by secondary reaction either back into unchanged substrate, or into excision products. In reactions that go to competition, inevitably all substrate is converted to excision products, as it is the only

stable end product. The production of the both inversion product and excision product from the RABD + helper reactions on the 2MutHetDim-DR substrate in this experiment demonstrated the RHD and MHD configurations were active. Additionally, the lack of activity on the 2MutHomDim substrates, and inability of the RABD + helper pair to catalyse secondary reaction on the inversion products from the 2MutHetDim-DR substrates, indicates the AHD configuration is inactive. The non-reversibility of the recombination reaction is the result of the RHD configuration transforming the heterodimer-binding substrate into a substrate with homodimer binding sites that cannot be bound by the RABD subunits.

5.5.3.2 Differences in strength of the inversion and excision product signals

The inversion and excision product signal from the 2MutHetDim-DR reactions were not of the same strength, in spite of both products being formed as the result of heterodimer reactions (MHD and RHD configurations). However, the reduction in signal strength of the inversion product was the result SA effect, which is expected to reduce the signal strength of products, such as the 2MutHetDim-DR inversion product, which contain homodimer binding sites for active recombinases. The reduction in signal strength of the inversion product is not significant enough to challenge the conclusion that a non-reversible reaction has been produced.

The MHD and RHD configurations could be demonstrated as being comparably active by running the RABD + helper complementation on both the 2MutHetDim-DR and 2MutHetDim-IR substrates, and comparing the excision products. The excision products from the 2MutHetDim-DR substrate would indicate MHD activity, while excision products from the 2MutHetDim-IR substrates would indicate RHD activity. A 2MutHetDim-IR substrate could be used that would place the RABD subunits on the homodimer site of the excision product. Because the RABD subunits are not expected to bind a homodimer site well, the excision product from a 2MutHetDim-IR substrate that specified them would not be affected by the SA effect. This would make the excision product from the 2MutHetDim-IR substrate more comparable with the excision product from the 2MutHetDim-DR substrate, which contains a heterodimer site, and is subject to only a low level of SA effect.

5.5.3.3 The H1 domain as a non-specific RABD

The H1 domain was also demonstrated to act as a non-specific binding domain within the limits of detection of the experiment. The sites to which the H1 domain was applied were quite divergent, and only 2 of 6 bp were conserved among all the Z-sites sequence (Figure 5-9). Additionally, there is some variability in the sites that the H1 domain binds to in the left and right half-sites of Tn3 site I, also indicating a level of site-specific recognition tolerance of the domain. The base pairs at the two foregoing conserved positions among the Z-sites, can be found in either the right or left half-sites of site I. Additionally, of these 2 bp, 1 bp is at one of the three positions that are non-variable among the H1 binding sites in the left and right half-sites of site I. Overall, all of the Z-sites shared 3–4 bp with either the left or right H1 binding sites of site I, some of which were at positions that varied between the left and right half H1 binding sites of site I. Therefore, it is currently unknown whether the H1 domain would work as effectively at sites that altered these positions, with the one position conserved between both H1 sites of site I being of special interest.

5.5.3.4 The utility of non-specific RABDs

Although the repertoire of available engineered ZFAs is increasing versatile, they are still a major limiting factor in the application of ZFR system to genomic contexts, and so the reduction of the recognition specificity constraint is significant. Although the arrival of TALERs may circumvent the problem of ZFA availability (Mercer et al., 2012), TALE are still difficult to construct and their size may present a barrier to certain modes of delivery (e.g. using AAV vectors). The ZFR system possesses an important difference with the similar ZFN system, in that the catalytic domain in the ZFR system brings with it its own site-specific capability. The site-specific capability of the ZFR catalytic domain, although imposing its own set of constraints when locating a target site, obviates the need to have two ZFAs in the system to produce a unique binding site in a human-size genomic context. The primary contention with using a non-specific RABD in a gene therapy recombination strategy would likely be that the non-specific RABD might product undesirable off-target activity. However, work on the sleepy + active partners described in Chapter 4 (Section 4.3), may be able to address that concern by making RABD subunits catalytically inactive when binding sites on their own.

5.5.3.5 The experiment succeeded in achieve all of the project aims

The data indicates that complementation reactions between RABD + helper subunits can be orchestrated to produce a non-reversible reaction. This is the first ever demonstration of a non-reversible reaction using the ZFR system. Furthermore, this setup, when utilized for integrations reactions, should be able to produce an orientation specific integration outcome through stochastic selection (see Figure 5-1). Finally, because the RABD can be used in a non-specific or reduced specificity fashion, it significantly reduces the specificity constraints of the ZFR system that requires locating a Z-site that will accommodate two available ZFAs. Thus, with this experiment, all three objectives of this project (*to impart the ZFR system with non-reversible reaction capability, to impart the ZFR system with orientation-specific integration capability, and to reduce the specificity constraints imposed by limitations in the repertoire of available ZFAs*) would appear to have been achieved, within the limits of detection of the experimental system used.

Chapter 6: *Discussion*

6.1 Introduction

6.1.1 A case for the ZFR system

The ZFR is an enzyme system for site-specific genome editing which features alongside its better known cousins, the ZFN, TALEN, and CRISPR-Cas9. Based on a recombinase platform, the ZFR has the potential to overcome several of the disadvantages of the nuclease-based systems that currently dominate the field of targeted genome editing. The reliance of nuclease-based systems on HDR to perform gene correction, prevents their use in non-dividing cell types. Additionally, commonly occurring off-target activity by the nuclease-based systems generates unintended genomic rearrangements, such as translocations, inversions, and large deletions (Frock et al., 2015, Tsai et al., 2015), which may present a danger to their use as an *in vivo* gene therapy system.

However, the greatest barrier to consideration of nuclease-based systems as a high fidelity mode of gene correction, suitable for *in vivo* therapeutic applications, is that the rate of indel production at the target site outcompetes the rate of gene correction (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012). Greater indels production at a therapeutic target site has the potential to damage functional alleles in situations where correction of a heterozygous allele is desired (e.g. familial hypercholesterolemia, or p53 dysfunction). Additionally, the dominance of NHEJ over HDR may result in translocations between target sites and naturally occurring DSB hotspots (Frock et al., 2015, Tsai et al., 2015), and also between both homologous chromosomes containing the target site, potentially resulting in the production of oncogenic breakage-fusion-bridge cycles (Frock et al., 2015). Nickases-based strategies may greatly reduce the production of indels at target sites, but suffer from a significant reduction in gene correction efficiency (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012), and do not prevent translocations between homologous chromosome containing the target site (Frock et al., 2015).

Because the ZFR relies on a recombinase-mediated DNA recombination, it is better suited to driving high-fidelity reactions at the target site than nuclease-based systems, and is also not limited to use in dividing cells. Additionally, although the off-target activity of the ZFR in a genomic context has not been well characterized, the reliance of the ZFR on two dimers, bound to both target sites of the reaction, may present opportunities to limit the possibility of unwanted recombination reactions by modifying the subunits such that they will only interact in specified tetramer configurations. This strategy might also prevent on-target homologous translocation events, which while not yet reported for the ZFR system, are hypothetically possible. However, before such advancements to the ZFR system can be deemed worthwhile, basic control over parameters of the ZFR recombination reaction must first be achieved. To this end, this project set out to accomplish three goals: the generation of non-reversible reaction, integration orientation specificity control, and reduction of the sequence recognition specificity requirements of the ZFR system.

6.1.2 Addressing the challenges

The experiments in this work explored several lines of reasoning in order to address the challenges in making the ZFR a more viable system for gene therapy and genomic editing. The approaches taken primarily revolved around attempting to alter the residue or geometric contact between two dimers forming a ZFR tetramer, exploiting a model of ZFR activation called the 'dimer interface unlocking' hypothesis, and altering the activity parameters of the binding domains. Of these approaches, the attempts to modify the dimer-dimer interaction contacts or geometry were unsuccessful (Sections 4.2 and 4.3). Conversely, working through the dimer interface unlocking model, sleepy + active ZFR pairs were produce that have the capacity to limit off-target activity, produce a significant recombination reaction directionality bias, and as a consequence of that directionality bias, should be capable of producing some level of integration orientation specificity bias (Sections 4.3 and 4.4). The attempts to modify the activity parameters of binding domains produced RABD + helper ZFR pairs capable of an extremely robust non-reversible recombination reaction effect (Section Chapter 5: 5.4). This non-reversible reaction potential of the RABD + helper pair system is expected to likewise produce a robust integration orientation specificity effect through the stochastic reaction pathway processes described as 'try and try again' integration orientation specificity (Section

5.1.3.5). Additionally, RABD + helper strategy succeeded in reducing some of the sequence recognition specificity limitations of the system.

The 'sleepy + active' strategy and 'RABD + helper' strategy might be combined to produce sleepy-RABD + active-helper subunits that unite the benefits of both strategies, preventing possible off-target activity, which the RABD subunits might carry out due to the reduce specificity of their binding domains. Importantly, this project also characterized the parameters of the SA effect, which illustrates the need to reduce the number of recombinase subunits that are active on their own within a reaction. The SA effect results from cleavage of DNA by active subunits outside of recombination reactions, but is not produced by sleepy subunits (Sections 3.10.5 and 4.5.8).

6.2 Disproved hypotheses

6.2.1 Ionic energy barrier between counterpart 102 residues at dimer-dimer formation

Based on the observation that mutations at position 102 in Tn3 resolvase play the most important role in the hyperactivation of the enzyme, and models of the dimer-dimer synapsis based on existing crystal structures, it was suspected that the counterpart residues at position 102 might come in direct contact with one another during the initial stage of synapsis (Section 4.2). It was hypothesized that the reason mutations at 102 may play such an important role in Tn3 resolvase hyperactivation, is that the negative charges at the wild-type residue Asp102 may play a role in negative regulation of the resolvase by creating ionic repulsion between opposing dimers.

The hypothesis above was tested by generating ZFR mutants with oppositely charged residues at position 102, and arranging these ZFRs in complementation reactions such that opposing dimers, within a dimer-dimer interaction, carried either similar or oppositely charged 102 residues. The expectation was that, if the primary activating mutations at 102 exert their effect by removing a negative ionic energy barrier between opposing dimer at tetramer formation, then placing oppositely charged residues at position 102 between dimers, should produce at least as strong an activating effect as activating mutations such as D102Y.

All combinations of opposition charged 102 residues were tested in these complementation reactions, and none showed any activity. Thus, it was concluded that the ionic charge of Asp102 does not produce a crucial ionic charge barrier that regulates the activity of the resolvase, and that placing alternately charged residues at this position in complementation reactions is not a strategy that can be used to further the goals of this project.

6.2.2 Conformational dimer asymmetric to produce dimer-dimer orientation specificity

It was hypothesized that conformational asymmetry might be generated within a ZFR dimer to allow a dimer-dimer orientation specificity bias (Section 4.3). According to existing models of wild-type small serine recombinase activation, the dimers of the catalytic module are contacted on one side by the subunits of the regulatory module, and this contact appears to be important for activation of the enzyme (Rowland et al., 2009, Rice et al., 2010). Almost all of the secondary activating mutations can be proposed to produce their effect by destabilizing a region of the dimer-interface that was termed the 'locking interface' in Section 4.3. Additionally, it can be seen from the 1GDT γδ resolvase dimer structure (a close homologue of Tn3 resolvase) that one of these locking regions is in an unlocked configuration, similar to what might be produced by the secondary activating mutations. It can also been seen in the 1GDT resolvase dimer structure that asymmetric unlocking of the dimer leads to a conformational asymmetry that might affect the dimer-dimer interaction during the initial stage of tetramer formation. Thus, it was proposed that differential inclusion of activating mutations in a ZFR heterodimer could produce a conformational change that might lead to a dimer-dimer orientation specificity bias. However, the experiments of Section 4.3 demonstrated that this was not the case.

6.3 Discoveries and their potential applications

6.3.1 The sleepy + active complementation strategy

One of the most important discoveries made in this work, was that activated ZFR subunits can be used to activate inactive subunits when used in complementation reactions. This discovery is important because within a genomic context the use of

inactive ZFRs limits the potential for off-target reactions by reducing the number of active tetramer configuration that may be produced from the ZFR set used. This strategy appears especially valuable in light of the SA effect, which was commonly observed and has been linked to cleavage of DNA, by activated ZFRs, outside of recombination reactions. Such cleavage could pose a problem for gene therapy applications of the ZFR. Ideally, the ZFR system could be further modified such that all ZFR subunits within a reaction only produced activity when used in combination with one another.

Another important discovery that was made using the sleepy + active ZFR pairs was that they can produce a strong recombination reaction directionality bias (Figure 4-27). The heterodimer reactions are significantly more active than reactions where the same set of ZFRs are paired as homodimers.

6.3.2 The RABD + helper complementation strategy

Perhaps the most important discovery that was made during the course of this work was that RABD + helper subunit pairs can be used to produce a robust non-reversible reaction effect. Chapter 5 set out to both reduce the sequence recognition specificity limitations of the ZFR system and generate RABDs with the goal of producing a non-reversible reaction based differential binding capability of subunits. These RABD subunits are inactive on their own because they possess insufficient binding affinity for their target sites as homodimers. However, RABDs still retain enough binding affinity that they can bind target sites when paired in heterodimers with helper ZFRs that possesses normal binding capability. Three approaches were used to generate RABDs, and each had varying degrees of sequence specificity reduction as well.

The first strategy attempted to generate non-specific ZFAs through modular assembly, with the expectation that many of the ZFAs would possess weak binding affinity as well. All of the ZFRs produced through this strategy were inactive, presumably because the binding affinity of the ZFAs was too weak.

A second strategy attempted to determine whether the number of fingers in active ZFAs could be reduced, in order to reduce both their sequence specificity and binding affinity. Reducing the number of zinc fingers required for one ZFR in a

heterodimer would allow a greater range of sites to be targeted, as shorter ZFA binding sites are easier to target with the existent repertoire of zinc finger modules. These RABDs imparted their ZFRs with an intermediate level of activity when participating in heterodimers, and showed no activity when forming homodimers. However, the product from the key heterodimer configuration that could directly indicate a non-reversible reaction was not present in the result, possibly due to being destroyed by the SA effect.

The third strategy proved to be the most successful (Section 5.4). The H1 domain was used as the non-specific RABD, and these RABD + helper complementations were tested for both the ability to produce a non-reversible reaction, as well as the ability of the RABD to act as a non-specific binding domain. These reactions demonstrated a robust non-reversible reaction, and the RABDs were able to bind non-specifically to all of the sites they were tested on (although these sites did share, at minimum, 2 of 6 bp in common with either of the H1 domains natural binding sites within Tn3 site I). Additionally, the non-reversible reaction generated by the RABD + helper pair should also be capable of producing an orientation-specific integration reaction via the 'try and try again' stochastic pathway (Section 5.1.3.5).

It is possible that the RABD approach based on truncated versions of active ZFAs, might have worked just as well as the approach based on the H1 domain. The overall activity level of RABD + helper complementation reactions involving the truncated active ZFAs was much lower than it was for those involving the H1 domain. Since the inversion product band signalling success in the reaction involving the H1 domain was obviously reduced by the SA effect, it is possible that the lower overall level of activity in the reactions involving truncated active ZFA domains, allowed the SA effect to hide the success of the experiment. In any case, the success of the H1-based RABD + helper complementations demonstrates the success of the RABD + helper approach, and represents the first ever demonstration of a non-reversible reaction utilizing ZFRs. This result represents a major advance for the ZFR system and should allow for a dramatic increase in the efficiency of integration reactions, with the added benefit of producing orientation-specific integrations.

6.3.3 The SA effect

Through the use of ZFRs with variably active binding domains (Chapter 3 and 5), and the use of sleepy + helper complementations (Chapter 4), the experiments within this project were able to characterize the SA effect and definitively determine that it is caused by DNA cleavage, rather than simply being a product of tight ZFR DNA binding. Although cleavage by ZFRs outside of recombination reactions has been observed or suspected in previous experiments conducted within the Stark lab, the work in this volume contains the first characterization of the SA effect parameters within ZFR complementations reactions (Sections 3.10.5 and 4.5.8). The data and observations collected here provide useful guidance for interpreting results from experiments where ZFRs with differential activity are co-expressed and tested on the variety of substrates capable of detecting the activity of the various possible tetramer configurations comprised of given ZFR variants. Consideration of the derived rule-sets for the appearance of the SA effect (Table 4-3 and Figure 4-27) will be important for effective design of experiments like these in the future.

The fact that active ZFRs are able to cleave DNA outside of recombination reactions may present a problem for their use as high fidelity genome editing tools, in their current state. However, it should also be considered that the nuclease-based genome editing tools are unable to produce greater levels of gene correction than indels at the sites they target. The structural complexity of the ZFR system, may provide opportunities to operate over the problem of non-recombination reaction DNA cleavage. Indeed, the use of sleepy + active subunits, demonstrated in this work, may be an important step towards that goal.

6.3.4 Activity level of the ZFR system compared with the Tn3 resolvase system

Finally, the experiments of Section 4.3.5 demonstrated that the intrinsic activity level of the ZFR system is well below that of the hyperactive Tn3 resolvase system. An experiment was conducted where 24 secondary mutations that were sufficient to produce hyperactivity in the Tn3 resolvase system in conjunction with Y, were tested in the ZFR system in conjunction with NY. Only four of these secondary mutations imparted their ZFRs with any significant recombination activity. The NY background used for the

ZFR experiments is more activated than the Y background in which these secondary mutations had previously been characterized in. Thus, the inactivity of 20 of the 24 ZFRs, each carrying one of these secondary mutations, indicates the use of a ZFA binding module over the natural H1 binding module significantly reduces the activity of the system. This difference in activity might be a due to superior binding by the H1 domain relative to ZFAs, or it might be that the natural H1 domain affords some conformational flexibility to the arm region that better enables the enzyme. This result should be taken into consideration when attempting to transport activating mutations from the Tn3 resolvase system into the Tn3-based ZFR system.

6.4 Experiments to further this work

6.4.1 Sleepy + sleepy subunit complementations

In vivo gene therapy applications of the ZFR system will require that off-target reactions do not happen. Although this might be accomplished by optimizing the binding domains, there is an obvious benefit to having a ZFR system where the catalytic domains are only active when used in combination with one another. Therefore, if a sleepy + sleepy strategy could be made to work, it would represent one way to achieve this goal.

Although an experiment in Section 4.3.6 tested the ability of sleepy ZFR carrying the L66I secondary mutation, to produce recombination with other sleepy ZFRs carrying various secondary mutation on the other side of the locking interface, this was not a robust test of the sleepy + sleepy complementation strategy. L66I was one of the 20 of 24 secondary mutations that showed no activity in the base-level ZFR activity screen (Figure 4-20 A). Because these mutants all produced no recombination activity, it is impossible to know how far below the activation threshold they actually are. Additionally, it can be seen from the sleepy + active ZFR reactions where a G70C mutant was tested against several other sleepy ZFRs that have locking interface mutation on the other side of the interface, that one (T109I) was easily woken up while the other did not produce activity (Figure 4-20 B).

Therefore, since some sleepy mutants appear to be closer to the activation threshold then others, it would be sensible to attempt other combinations of sleepy + sleepy ZFR, in order to look for a synergistic activating effect. The logical place to start

would be testing T109I against all of the other sleepy ZFRs with secondary mutations operating from the other side of the locking interface. Additionally, secondary mutations from sleepy ZFRs might be combined in order to look for combinations that remain sleepy. Sleepy ZFRs with combinations of secondary mutations would likely be closer to the activation threshold than the sleepy ZFRs in which the individual mutations produced no effect. It is perhaps worthwhile to remind the reader that, in spite of their activity in the ZFR system, all of the secondary mutation used for the experiments of this project were known to have a robust activating effect on the Tn3 resolvase catalytic domain from experiments generating hyperactive Tn3 resolvase mutants (i.e. these mutation were all sufficient to hyperactivate the resolvase when used as the sole secondary mutation in combination with D102Y).

6.4.2 Co-localized versus distributed locking interface mutation aggregation

In Section 4.3.5.1, the idea was put forward that two ZFR mutants that possessed secondary mutations on opposite sides of the locking interface, might display differential levels of activity depending on whether the locations of the mutations within a tetramer were orchestrated in a co-localized or distributed arrangement (see Figure 4-19). Within the RHD and MHD configurations the mutations would co-localize and the two mutations would exist on opposite sides of one locking interface per dimer, with the other locking interface of the dimer containing no mutations. Within the AHD configuration, the mutations would be distributed, one mutation only, per locking interface.

There are two ways the co-localized versus distributed arrangement strategy might be employed. The first is that mutations on opposite sides of the locking interface could be chosen so they would amplify the disruption of the interface when co-localized. This strategy would involve selecting secondary mutations in slightly different regions of the locking interface so they would each disrupt a different set of specific contacts. In this scenario, mutations that were too weak to sufficiently disrupt the locking interface when used in distributed arrangement might create a synergistic activating effect when used in the co-localized arrangement. If differential levels of activity arose between the RHD and AHD configurations, this would give rise to a recombination reaction directionality bias.

This scenario, in effect, demonstrates how a successful sleepy + sleepy ZFR complementation might operate.

In Section 4.3.6, sleepy + active ZFR complementations were carried out that tested a ZFR with G70C, against a variety of ZFRs with secondary mutations operating from the other side of the locking interface (Figure 4-24 B). The complementation between a ZFR carrying G70C and a ZFR carrying T109I was observed to produce more activity than any of the other reactions in this set. This might have resulted from the fact that G70C and T109I would be expected to disrupt different portions of the locking interface, due to their locations (Figure 6-1). It should be noted that the term 'disruption', as used here, also includes the potential unlocking effect mediated by E-helix 'landing pad' interactions, described in Section 4.4. The experiment was able to detect activity from the MHD configuration, and thus, this complementation represented a test of the co-localized arrangement of interface unlocking mutations. Unfortunately, there was insufficient time to carry out further experiments to test difference between the activity of the RHD and AHD configuration, which might have revealed recombination reaction directionality bias, and, less likely, a dimer-dimer orientation specificity bias. Therefore, the experiment using this sleepy + active ZFR pair should be carried out again on the full complement of substrate types designed to detect the activities of all tetramer configurations.

The other type of co-localized versus distributed arrangement strategy involves pairing locking interface mutations for redundancy. In this strategy, interface locking knockdown mutations that affect the same contacts from opposite sides of the interface, would be paired such that they would produce an additive effect when co-localized. Success of this strategy would represent the ability to activate AHD configuration, instead of the RHD and MHD configurations, providing an additional option for control over the ZFR reaction parameters.

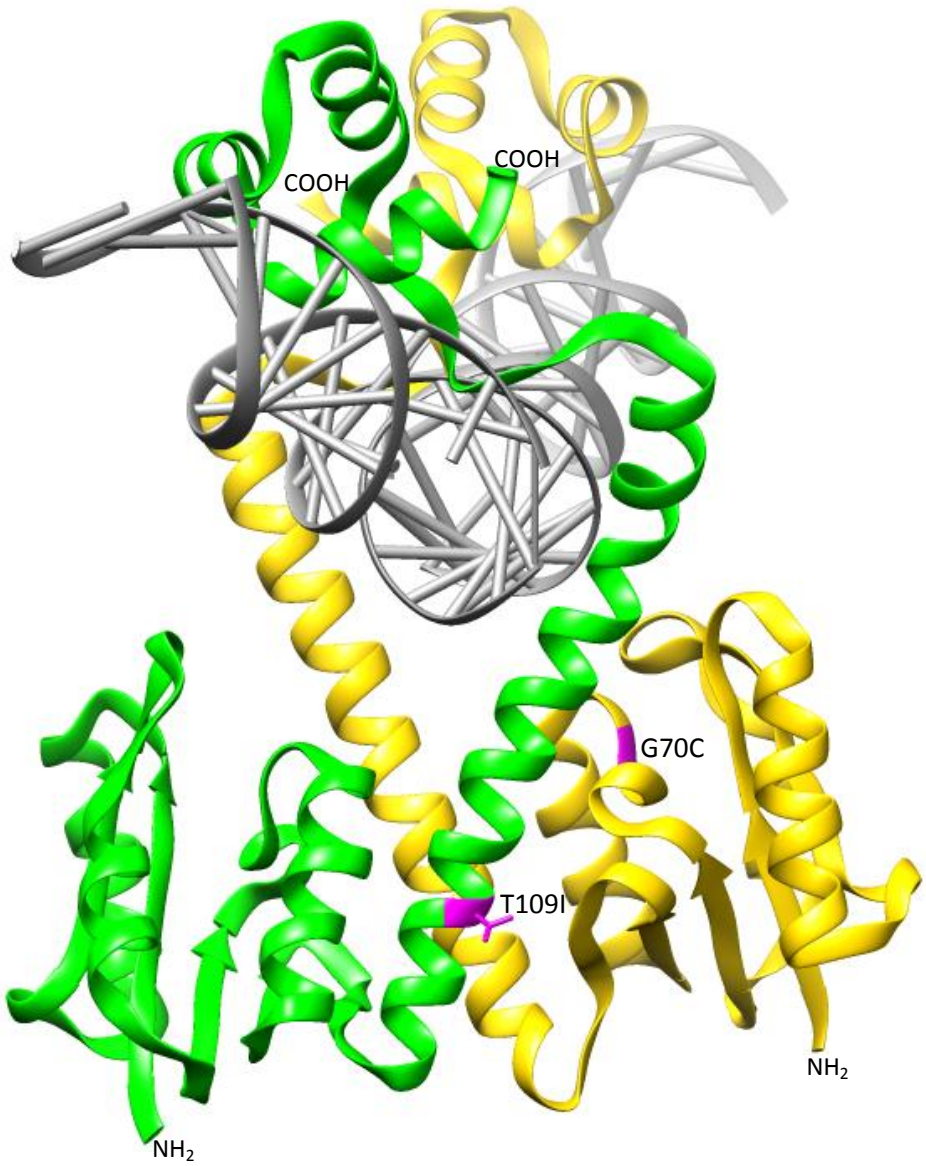


Figure 6-1: Locations of G70C and T109I. The figures shows the 1GDT crystal structure with the location of G70C indicated in magenta and labelled for the gold subunit, and the locations of T109I indicated in magenta and labelled for the green subunit.

6.4.3 Complete locking interface disruption

There is evidence from Burke et al. (2004) that placing too many locking interface knockdown mutations in combination can abolish the activity of Tn3 resolvase. The mutations A117V, R121K, and E124Q (collectively termed 'C') all produce an hyperactivation when used as the sole secondary mutation, but when all three are used in concert, Tn3 resolvase is inactive (Burke et al., 2004). One explanation for this activity is that this combination of mutation completely knocks out the locking interface. The locking interface comprises the majority of the dimer interface, and so completely knocking out the locking interface may disrupt the dimer interface to such a degree that

dimers are no longer able to form. There is support for this hypothesis in the form of evidence that some activated resolvase mutants have an increased monomeric character when studied *in vitro* (Arnold et al., 1999). Further support for this hypothesis can be found in Burke *et al.* (2004) where it was shown that D102Y + C is active only on a *res* x *res* substrate, but not on either *res* x site I, or site I x site I substrates. The fact that D102Y + C is only active on the *res* x *res* substrate may suggest that the accessory module of the synaptosome is helping to stabilize the catalytic dimer subunits, compensating for the disruption of their dimer interface.

One way this hypothesis might be tested and exploited would be to generate a pair of ZFRs that are designed to differentially knock out the locking interface, entirely, when either used as homodimers or heterodimers. This strategy could take two forms. In the first strategy, both ZFR subunits could contain redundant mutations operating from both sides of the locking interface, but located in different regions of the locking interface. For example, consider a pair of mutants (both with an NY background) where one subunit might have L66I + A117V, while the other might have G70A + R121K. These mutation sets operate within two discrete and separate locations within the locking interface. Within homodimers, the effect might be redundancy, but when used in a heterodimer context, the effect might be additive. If the additive effect was strong enough to completely knock out the locking interface, then homodimers might be active while the heterodimers were inactive.

However, the most impressive strategy, if it worked, would be to use the pair NY and NYC. Both of these resolvase mutants are inactive on their own. If the NYC mutant is inactive for the reasons stated above, then perhaps it can be brought back to life by combining it in a heterodimer with NY. If NYC is inactive because its dimer interface so sufficiently disrupted that dimers may not form, then perhaps it can be brought back to life in a NY/NYC heterodimer. In this heterodimer configuration, one locking interface of the dimer would be preserved. If this strategy were to work, it would represent the creation of a functional sleepy + sleepy ZFR pair, which would abolish the SA effect, and produce a ZFR pair where the subunits were only active when used in concert, greatly reducing opportunities for off-target reactions in a genomic context.

6.4.4 Dissection of M

It was postulated in Section 4.4 that the collection of mutations known as 'M' (G101S, D102Y, M103I, and Q105L), may produce their effect via the E-helix landing pad interaction aspect of the dimer interface unlocking model. Additionally, specific contacts between counterpart landing pad residues were proposed. It was proposed that M103V and Q105L might produce their activating effect by generating more favourable interactions between these residues and their counterparts during tetramer formation (Figure 4-26 B). Both of these interactions, involving M103V and Q105L, could have implications for directly controlling the compatibility of the two possible orientations at the dimer-dimer interface. The possibility of counterpart contact between M103V residues seems particularly plausible. The sleepy + active complementation utilizing M appeared to produce a robust recombination reaction directionality bias (Figure 4-27). However, it is not yet clear whether this recombination reaction directionality bias is of the same nature as that observed for previous sleepy + active ZFR pairs, but with an overall increase in activity level due to the increased activity level of M, or whether there were different fundamental mechanisms underpinning the outcome, such as counterpart residue interaction between M103V and Q105L.

The secondary mutations included in M cannot be tested independently in sleepy + active ZFR pairs because they do not possess enough activity to activate a ZFR (Figure 4-20 A). However, from the three secondary mutations included in M (G101S, M103I, and Q105L), three pairs of two could be tested in a differential substrate assay. The assay would determine whether tetramer configuration activities differ from the typical pattern of activity observed locking interface mutants (i.e. RHD and MHD are equally active while AHD is less active). Conversely, the alternative hypotheses for M103V and Q105L would both be predicted to produce a difference in the activity of the RHD and MHD configurations.

6.4.5 Landing pad-based dimer-dimer compatibility

The residues involved in the E-helix landing pad interaction may represent the best opportunity for controlling dimer-dimer interaction compatibility. This region is where the first contact between opposing E-helices of the dimer-dimer interaction are made, and thus, this is where compatibility or incompatibility between opposing dimers

may be specified. Generating selectively compatible dimers would not only provide a direct route to generating a dimer-dimer orientation specific reaction, but might also provide a strategy to prevent possible recombination between Z-sites on homologous chromosomes, and allow the generation of sets of ZFRs that could be used for recombinase-mediated cassette exchange (RMCE). RMCE utilizes two separate recombination reactions in order to facilitate the exchange of a segment of DNA between two DNA molecules (similar to the replacement vector strategy depicted in Figure 1-1). In order for ZFR-based RMCE to be performed efficiently, two sets of ZFRs would need to be generated that are not compatible with each other, in order that only the intended pairs of recombination sites would react with one another. It should be noted, that another obvious strategy for RMCE using ZFRs is simply to generate two sets of ZFRs with incompatible catalytic domains from two different serine recombinases.

Because some of the contacts of the E-helix landing pad are likely made between opposite subunits (e.g. 102-109 or105-105), while other are likely made between opposite-adjacent subunits (e.g. 103-103) of the dimer-dimer interface (see Figure 4-26), there may be potential for generating selective compatibility between both type of interactions. Three general approaches for attempting to generate selective compatibility between subunits are: (1) Place large hydrophobic residues on one side of the interaction and small hydrophobic residues on the other. (2) Place hydrophobic residues on one side of the interaction and polar residues on the other. (3) Place charged residues on one side of the interaction and oppositely charged residues on the other.

6.4.6 Truncated ZFA RABD + helper re-test

The experiment utilizing truncated ZFAs as RABDs for a RABD + helper complementation was essentially left unfinished (Section 5.3). It is possible that this experiment was successful in generating a non-reversible reaction, but because of the intermediate overall activity of this RABD + helper system, a key result (inversion on the 2MutHetDim-DR substrate indicating RHD configuration activity) may have been hidden by the SA effect. However, an alternative explanation is that the RABD subunits might not have been able to re-ligate the DNA. The SA effect had not been fully characterized at the time of this experiment and the activity information for the RHD configuration could have been easily obtained through the inclusion of the variety of 2MutHetDim-IR substrates

that place the RABD ZFRs on the excision product. The inclusion of these substrates would have provided an alternative method of detecting RHD activity and demonstrated whether the RABD subunits are able to re-ligate the DNA. Because the excision product from this variety of 2MutHetDim-IR substrates would contain homodimer binding sites for the RABD subunits (which do not bind well) it would also be less affected by the SA effect, making the result clear. The experiment should be repeated with the inclusion of this variety of 2MutHetDim-IR substrates.

6.5 Conclusion

Currently, the most popular tools for site-specific genome editing are the site-specific nuclease and nickase systems. The site-specific nuclease systems have the drawback that they produce unwanted indel mutations at their target sites at a rate that is equal or higher than that of desired gene correction (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012). This drawback alone, likely makes them poor candidates for *in vivo* gene therapy applications. Additionally, they are capable of producing wide-spread off-target reactions, leading to genomic rearrangements (Tsai et al., 2015, Frock et al., 2015). The nickase systems largely ameliorate the problem of on-target indel production, and have reduced off-target activity leading to genomic rearrangements, but suffer from a considerable loss in efficiency (Ramirez et al., 2012, Kim et al., 2012, Wang et al., 2012). It is possible that the gene correction efficiency of nuclease and nickase systems cannot be easily improved because they are limited by the level of endogenous cellular HDR activity. Additionally, neither site-specific nuclease nor nickase systems may be used for gene correction in dividing cell types because they rely on HDR which is only active during S phase and G2 phase of the cell cycle. The site-specific recombinase ZFR system may have the potential overcome the limitations of the nuclease and nickases systems, but a basic level of control over its recombination reactions must first be achieved in order for the tool to become useful.

The work described in this thesis attempted to generate control over fundamental ZFR reaction parameters such as recombination reaction directionality, and integration orientation control. The RABD + helper ZFR strategy using the H1 domain appears to have been completely successful in achieving the goal of producing a non-reversible reaction (Section 5.5.3.5). Additionally, this strategy is predicted to produce orientation-specific

integrations via a stochastic selection pathway (Section 5.1.3.5). The sleepy + active ZFR strategy, based on the locking interface model, also appears to have been successful in producing a recombination reaction directionality bias (Section 6.3.1). In addition, the sleepy + active strategy also represents the generation of a ZFR with reduced potential for off-target activity, as one ZFR of the pair is inactive without the activating effect of the other. By uniting the RABD + helper ZFR strategy and sleepy + active ZFR strategy, it should be possible to generate a robust system for stable orientation-specific ZFR-mediated integration reactions, with reduced potential for off-target reactions. Ultimately, it is the level of efficiency of these reactions in a mammalian cell system that will determine whether the ZFR can offer advantages over the nuclease and nickase systems as a tool for gene therapy.

References

AARTS, M. & TE RIELE, H. 2011. Progress and prospects: oligonucleotide-directed gene modification in mouse embryonic stem cells: a route to therapeutic application. *Gene Ther,* 18**,** 213-9.

AKOPIAN, A., HE, J., BOOCOCK, M. R. & STARK, W. M. 2003. Chimeric recombinases with designed DNA sequence recognition. *Proc Natl Acad Sci U S A,* 100**,** 8688-91.

ALLERS, K., HUTTER, G., HOFMANN, J., LODDENKEMPER, C., RIEGER, K., THIEL, E. & SCHNEIDER, T. 2011. Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. *Blood,* 117**,** 2791-9.

ARNOLD, P. H., BLAKE, D. G., GRINDLEY, N. D., BOOCOCK, M. R. & STARK, W. M. 1999. Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *EMBO J,* 18**,** 1407-14.

ARNOULD, S., CHAMES, P., PEREZ, C., LACROIX, E., DUCLERT, A., EPINAT, J. C., STRICHER, F., PETIT, A. S., PATIN, A., GUILLIER, S., ROLLAND, S., PRIETO, J., BLANCO, F. J., BRAVO, J., MONTOYA, G., SERRANO, L., DUCHATEAU, P. & PAQUES, F. 2006. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol,* 355**,** 443-58.

ARNOULD, S., PEREZ, C., CABANIOLS, J. P., SMITH, J., GOUBLE, A., GRIZOT, S., EPINAT, J. C., DUCLERT, A., DUCHATEAU, P. & PAQUES, F. 2007. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J Mol Biol,* 371**,** 49-65.

ASHWORTH, J., HAVRANEK, J. J., DUARTE, C. M., SUSSMAN, D., MONNAT, R. J., JR., STODDARD, B. L. & BAKER, D. 2006. Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature,* 441**,** 656-9.

ASHWORTH, J., TAYLOR, G. K., HAVRANEK, J. J., QUADRI, S. A., STODDARD, B. L. & BAKER, D. 2010. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res,* 38**,** 5601-8.

BAE, K. H., KWON, Y. D., SHIN, H. C., HWANG, M. S., RYU, E. H., PARK, K. S., YANG, H. Y., LEE, D. K., LEE, Y., PARK, J., KWON, H. S., KIM, H. W., YEH, B. I., LEE, H. W., SOHN, S. H., YOON, J., SEOL, W. & KIM, J. S. 2003. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat Biotechnol,* 21**,** 275-80.

BEDELL, V. M., WANG, Y., CAMPBELL, J. M., POSHUSTA, T. L., STARKER, C. G., KRUG, R. G., 2ND, TAN, W., PENHEITER, S. G., MA, A. C., LEUNG, A. Y., FAHRENKRUG, S. C., CARLSON, D. F., VOYTAS, D. F., CLARK, K. J., ESSNER, J. J. & EKKER, S. C. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature,* 491**,** 114-8.

BEERLI, R. R., SEGAL, D. J., DREIER, B. & BARBAS, C. F., 3RD 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A,* 95**,** 14628-33.

BHAKTA, M. S., HENRY, I. M., OUSTEROUT, D. G., DAS, K. T., LOCKWOOD, S. H., MECKLER, J. F., WALLEN, M. C., ZYKOVICH, A., YU, Y., LEO, H., XU, L., GERSBACH, C. A. & SEGAL, D. J. 2013. Highly active zinc-finger nucleases by extended modular assembly. *Genome Res,* 23**,** 530-8.

BHAKTA, M. S. & SEGAL, D. J. 2010. The generation of zinc finger proteins by modular assembly. *Methods Mol Biol,* 649**,** 3-30.

BITINAITE, J., WAH, D. A., AGGARWAL, A. K. & SCHILDKRAUT, I. 1998. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A,* 95**,** 10570-5.

BLANCAFORT, P., MAGNENAT, L. & BARBAS, C. F., 3RD 2003. Scanning the human genome with combinatorial transcription factor libraries. *Nat Biotechnol,* 21**,** 269-74.
BOCH, J. & BONAS, U. 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol,* 48**,** 419-36.
BOCH, J., SCHOLZE, H., SCHORNACK, S., LANDGRAF, A., HAHN, S., KAY, S., LAHAYE, T., NICKSTADT, A. & BONAS, U. 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science,* 326**,** 1509-12.
BOGDANOVE, A. J., SCHORNACK, S. & LAHAYE, T. 2010. TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol,* 13**,** 394-401.
BOLIVAR, F., RODRIGUEZ, R. L., GREENE, P. J., BETLACH, M. C., HEYNEKER, H. L., BOYER, H. W., CROSA, J. H. & FALKOW, S. 1977. Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene,* 2**,** 95-113.
BRIGGS, A. W., RIOS, X., CHARI, R., YANG, L., ZHANG, F., MALI, P. & CHURCH, G. M. 2012. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res,* 40**,** e117.
BUNING, H. 2013. Gene therapy enters the pharma market: the short story of a long journey. *EMBO Mol Med,* 5**,** 1-3.
BURKE, M. E., ARNOLD, P. H., HE, J., WENWIESER, S. V., ROWLAND, S. J., BOOCOCK, M. R. & STARK, W. M. 2004. Activating mutations of Tn3 resolvase marking interfaces important in recombination catalysis and its regulation. *Mol Microbiol,* 51**,** 937-48.
CARBERY, I. D., JI, D., HARRINGTON, A., BROWN, V., WEINSTEIN, E. J., LIAW, L. & CUI, X. 2010. Targeted genome modification in mice using zinc-finger nucleases. *Genetics,* 186**,** 451-9.
CARLSON, D. F., TAN, W., LILLICO, S. G., STVERAKOVA, D., PROUDFOOT, C., CHRISTIAN, M., VOYTAS, D. F., LONG, C. R., WHITELAW, C. B. & FAHRENKRUG, S. C. 2012. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A,* 109**,** 17382-7.
CARROLL, D. 2014. Genome engineering with targetable nucleases. *Annu Rev Biochem,* 83**,** 409-39.
CARROLL, D., MORTON, J. J., BEUMER, K. J. & SEGAL, D. J. 2006. Design, construction and in vitro testing of zinc finger nucleases. *Nat Protoc,* 1**,** 1329-41.
CATHOMEN, T. & JOUNG, J. K. 2008. Zinc-finger nucleases: the next generation emerges. *Mol Ther,* 16**,** 1200-7.
CERMAK, T., DOYLE, E. L., CHRISTIAN, M., WANG, L., ZHANG, Y., SCHMIDT, C., BALLER, J. A., SOMIA, N. V., BOGDANOVE, A. J. & VOYTAS, D. F. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res,* 39**,** e82.
CHANG, A. C. & COHEN, S. N. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of Bacteriology,* 134**,** 1141-1156.
CHEN, F., PRUETT-MILLER, S. M., HUANG, Y., GJOKA, M., DUDA, K., TAUNTON, J., COLLINGWOOD, T. N., FRODIN, M. & DAVIS, G. D. 2011. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods,* 8**,** 753-5.
CHEN, Z., WEN, F., SUN, N. & ZHAO, H. 2009. Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng Des Sel,* 22**,** 249-56.
CHEN, Z. & ZHAO, H. 2005. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res,* 33**,** e154.

CHEVALIER, B. S., KORTEMME, T., CHADSEY, M. S., BAKER, D., MONNAT, R. J. & STODDARD, B. L. 2002. Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell,* 10**,** 895-905.

CHIN, J. Y., KUAN, J. Y., LONKAR, P. S., KRAUSE, D. S., SEIDMAN, M. M., PETERSON, K. R., NIELSEN, P. E., KOLE, R. & GLAZER, P. M. 2008. Correction of a splice-site mutation in the beta-globin gene stimulated by triplex-forming peptide nucleic acids. *Proc Natl Acad Sci U S A,* 105**,** 13514-9.

CHO, S. W., KIM, S., KIM, Y., KWEON, J., KIM, H. S., BAE, S. & KIM, J. S. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res,* 24**,** 132-41.

CHOO, Y., SANCHEZ-GARCIA, I. & KLUG, A. 1994. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature,* 372**,** 642-5.

CHRISTIAN, M., CERMAK, T., DOYLE, E. L., SCHMIDT, C., ZHANG, F., HUMMEL, A., BOGDANOVE, A. J. & VOYTAS, D. F. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics,* 186**,** 757-61.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339**,** 819-23.

CONG, L., ZHOU, R., KUO, Y. C., CUNNIFF, M. & ZHANG, F. 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Commun,* 3**,** 968.

CONNELLY, J. P., BARKER, J. C., PRUETT-MILLER, S. & PORTEUS, M. H. 2010. Gene Correction by Homologous Recombination With Zinc Finger Nucleases in Primary Cells From a Mouse Model of a Generic Recessive Genetic Disease. *Mol Ther,* 18**,** 1103-1110.

CRADICK, T. J., FINE, E. J., ANTICO, C. J. & BAO, G. 2013. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research,* 41**,** 9584-9592.

CUI, X., JI, D., FISHER, D. A., WU, Y., BRINER, D. M. & WEINSTEIN, E. J. 2011. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol,* 29**,** 64-7.

CULVER, K. W., HSIEH, W. T., HUYEN, Y., CHEN, V., LIU, J., KHRIPINE, Y. & KHORLIN, A. 1999. Correction of chromosomal point mutations in human cells with bifunctional oligonucleotides. *Nat Biotechnol,* 17**,** 989-93.

DAYA, S. & BERNS, K. I. 2008. Gene therapy using adeno-associated virus vectors. *Clin Microbiol Rev,* 21**,** 583-93.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471**,** 602-7.

DENG, D., YAN, C., PAN, X., MAHFOUZ, M., WANG, J., ZHU, J. K., SHI, Y. & YAN, N. 2012a. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science,* 335**,** 720-3.

DENG, D., YIN, P., YAN, C., PAN, X., GONG, X., QI, S., XIE, T., MAHFOUZ, M., ZHU, J. K., YAN, N. & SHI, Y. 2012b. Recognition of methylated DNA by TAL effectors. *Cell Res,* 22**,** 1502-4.

DESJARLAIS, J. R. & BERG, J. M. 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc Natl Acad Sci U S A,* 89**,** 7345-9.

DIANOV, G. L. & HUBSCHER, U. 2013. Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res,* 41**,** 3483-90.

DIFILIPPANTONIO, M. J., PETERSEN, S., CHEN, H. T., JOHNSON, R., JASIN, M., KANAAR, R., RIED, T. & NUSSENZWEIG, A. 2002. Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J Exp Med,* 196**,** 469-80.

DOUGLAS, K. L. 2008. Toward development of artificial viruses for gene therapy: a comparative evaluation of viral and non-viral transfection. *Biotechnol Prog,* 24**,** 871-83.

DOYLE, E. L., HUMMEL, A. W., DEMOREST, Z. L., STARKER, C. G., VOYTAS, D. F., BRADLEY, P. & BOGDANOVE, A. J. 2013. TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1. *PLoS One,* 8**,** e82120.

DOYON, J. B., PATTANAYAK, V., MEYER, C. B. & LIU, D. R. 2006. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J Am Chem Soc,* 128**,** 2477-84.

DOYON, Y., VO, T. D., MENDEL, M. C., GREENBERG, S. G., WANG, J., XIA, D. F., MILLER, J. C., URNOV, F. D., GREGORY, P. D. & HOLMES, M. C. 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods,* 8**,** 74-9.

DREIER, B., BEERLI, R. R., SEGAL, D. J., FLIPPIN, J. D. & 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**,** 29466-78.

DREIER, B., FULLER, R. P., SEGAL, D. J., LUND, C. V., BLANCAFORT, P., HUBER, A., KOKSCH, B. & BARBAS, C. F., 3RD 2005. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem,* 280**,** 35588-97.

DREIER, B., SEGAL, D. J. & BARBAS, C. F., 3RD 2000. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol,* 303**,** 489-502.

DUNBRACK, R. L., JR. 2002. Rotamer libraries in the 21st century. *Curr Opin Struct Biol,* 12**,** 431-40.

EASTBERG, J. H., MCCONNELL SMITH, A., ZHAO, L., ASHWORTH, J., SHEN, B. W. & STODDARD, B. L. 2007. Thermodynamics of DNA target site recognition by homing endonucleases. *Nucleic Acids Res,* 35**,** 7209-21.

ELLIOTT, B., RICHARDSON, C., WINDERBAUM, J., NICKOLOFF, J. A. & JASIN, M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol Cell Biol,* 18**,** 93-101.

ELLIS, B. L., HIRSCH, M. L., PORTER, S. N., SAMULSKI, R. J. & PORTEUS, M. H. 2013. Zinc-finger nuclease-mediated gene correction using single AAV vector transduction and enhancement by Food and Drug Administration-approved drugs. *Gene Ther,* 20**,** 35-42.

ELROD-ERICKSON, M., ROULD, M. A., NEKLUDOVA, L. & PABO, C. O. 1996. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure,* 4**,** 1171-80.

EPINAT, J. C., ARNOULD, S., CHAMES, P., ROCHAIX, P., DESFONTAINES, D., PUZIN, C., PATIN, A., ZANGHELLINI, A., PAQUES, F. & LACROIX, E. 2003. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res,* 31**,** 2952-62.

FAJARDO-SANCHEZ, E., STRICHER, F., PAQUES, F., ISALAN, M. & SERRANO, L. 2008. Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res,* 36**,** 2163-73.

FENG, X., BEDNARZ, A. L. & COLLOMS, S. D. 2010. Precise targeted integration by a chimaeric transposase zinc-finger fusion protein. *Nucleic Acids Res,* 38**,** 1204-16.

FROCK, R. L., HU, J., MEYERS, R. M., HO, Y. J., KII, E. & ALT, F. W. 2015. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol,* 33**,** 179-86.

FU, F., SANDER, J. D., MAEDER, M., THIBODEAU-BEGANNY, S., JOUNG, J. K., DOBBS, D., MILLER, L. & VOYTAS, D. F. 2009. Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. *Nucleic Acids Res,* 37**,** D279-83.

FU, Y., FODEN, J. A., KHAYTER, C., MAEDER, M. L., REYON, D., JOUNG, J. K. & SANDER, J. D. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotech,* 31**,** 822-826.

FU, Y., SANDER, J. D., REYON, D., CASCIO, V. M. & JOUNG, J. K. 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotech,* 32**,** 279-284.

GAJ, T., GERSBACH, C. A. & BARBAS, C. F., 3RD 2013a. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol,* 31**,** 397-405.

GAJ, T., GUO, J., KATO, Y., SIRK, S. J. & BARBAS, C. F., 3RD 2012. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat Methods,* 9**,** 805-7.

GAJ, T., MERCER, A. C., GERSBACH, C. A., GORDLEY, R. M. & BARBAS, C. F., 3RD 2011. Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc Natl Acad Sci U S A,* 108**,** 498-503.

GAJ, T., MERCER, A. C., SIRK, S. J., SMITH, H. L. & BARBAS, C. F., 3RD 2013b. A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res,* 41**,** 3937-46.

GAJ, T., SIRK, S. J., TINGLE, R. D., MERCER, A. C., WALLEN, M. C. & BARBAS, C. F., 3RD 2014. Enhancing the specificity of recombinase-mediated genome engineering through dimer interface redesign. *J Am Chem Soc,* 136**,** 5047-56.

GARDLIK, R., PALFFY, R., HODOSY, J., LUKACS, J., TURNA, J. & CELEC, P. 2005. Vectors and delivery systems in gene therapy. *Med Sci Monit,* 11**,** RA110-21.

GELLHAUS, K., CORNU, T. I., HEILBRONN, R. & CATHOMEN, T. 2010. Fate of recombinant adeno-associated viral vector genomes during DNA double-strand break-induced gene targeting in human cells. *Hum Gene Ther,* 21**,** 543-53.

GERSBACH, C. A., GAJ, T. & BARBAS, C. F., 3RD 2014. Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Acc Chem Res,* 47**,** 2309-18.

GERSBACH, C. A., GAJ, T., GORDLEY, R. M. & BARBAS, C. F., 3RD 2010. Directed evolution of recombinase specificity by split gene reassembly. *Nucleic Acids Res,* 38**,** 4198-206.

GERSBACH, C. A., GAJ, T., GORDLEY, R. M., MERCER, A. C. & BARBAS, C. F., 3RD 2011. Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. *Nucleic Acids Res,* 39**,** 7868-78.

GEURTS, A. M., COST, G. J., FREYVERT, Y., ZEITLER, B., MILLER, J. C., CHOI, V. M., JENKINS, S. S., WOOD, A., CUI, X., MENG, X., VINCENT, A., LAM, S., MICHALKIEWICZ, M., SCHILLING, R., FOECKLER, J., KALLOWAY, S., WEILER, H., MENORET, S., ANEGON, I., DAVIS, G. D., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D., JACOB, H. J. & BUELOW, R. 2009. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science,* 325**,** 433.

GIOVANNANGELI, C. & HELENE, C. 1997. Progress in developments of triplex-based strategies. *Antisense Nucleic Acid Drug Dev,* 7**,** 413-21.

GORDLEY, R. M., GERSBACH, C. A. & BARBAS, C. F., 3RD 2009. Synthesis of programmable integrases. *Proc Natl Acad Sci U S A,* 106**,** 5053-8.

GORDLEY, R. M., SMITH, J. D., GRASLUND, T. & BARBAS, C. F., 3RD 2007. Evolution of programmable zinc finger-recombinases with activity in human cells. *J Mol Biol,* 367**,** 802-13.

GREISMAN, H. A. & PABO, C. O. 1997. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science,* 275**,** 657-61.

GRIZOT, S., SMITH, J., DABOUSSI, F., PRIETO, J., REDONDO, P., MERINO, N., VILLATE, M., THOMAS, S., LEMAIRE, L., MONTOYA, G., BLANCO, F. J., PAQUES, F. & DUCHATEAU, P. 2009. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res,* 37**,** 5405-19.

GUILINGER, J. P., THOMPSON, D. B. & LIU, D. R. 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol,* 32**,** 577-82.

GUO, J., GAJ, T. & BARBAS, C. F., 3RD 2010. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J Mol Biol,* 400**,** 96-107.

GUPTA, A., CHRISTENSEN, R. G., RAYLA, A. L., LAKSHMANAN, A., STORMO, G. D. & WOLFE, S. A. 2012. An optimized two-finger archive for ZFN-mediated gene targeting. *Nat Methods,* 9**,** 588-90.

HANDEL, E. M., GELLHAUS, K., KHAN, K., BEDNARSKI, C., CORNU, T. I., MULLER-LERCH, F., KOTIN, R. M., HEILBRONN, R. & CATHOMEN, T. 2012. Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors. *Hum Gene Ther,* 23**,** 321-9.

HENDRIE, P. C. & RUSSELL, D. W. 2005. Gene targeting with viral vectors. *Mol Ther,* 12**,** 9-17.

HERRMANN, F., GARRIGA-CANUT, M., BAUMSTARK, R., FAJARDO-SANCHEZ, E., COTTERELL, J., MINOCHE, A., HIMMELBAUER, H. & ISALAN, M. 2011. p53 Gene repair with zinc finger nucleases optimised by yeast 1-hybrid and validated by Solexa sequencing. *PLoS One,* 6**,** e20913.

HOCKEMEYER, D., SOLDNER, F., BEARD, C., GAO, Q., MITALIPOVA, M., DEKELVER, R. C., KATIBAH, G. E., AMORA, R., BOYDSTON, E. A., ZEITLER, B., MENG, X., MILLER, J. C., ZHANG, L., REBAR, E. J., GREGORY, P. D., URNOV, F. D. & JAENISCH, R. 2009. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat Biotechnol,* 27**,** 851-7.

HOLKERS, M., MAGGIO, I., LIU, J., JANSSEN, J. M., MISELLI, F., MUSSOLINO, C., RECCHIA, A., CATHOMEN, T. & GONCALVES, M. A. 2013. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res,* 41**,** e63.

HSU, P. D., LANDER, E. S. & ZHANG, F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell,* 157**,** 1262-78.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013a. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech,* 31**,** 827-832.

HSU, P. D., SCOTT, D. A., WEINSTEIN, J. A., RAN, F. A., KONERMANN, S., AGARWALA, V., LI, Y., FINE, E. J., WU, X., SHALEM, O., CRADICK, T. J., MARRAFFINI, L. A., BAO, G. & ZHANG, F. 2013b. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol,* 31**,** 827-32.

HU, J., TEPSUPORN, S., MEYERS, R. M., GOSTISSA, M. & ALT, F. W. 2014. Developmental propagation of V(D)J recombination-associated DNA breaks and translocations in mature B cells via dicentric chromosomes. *Proceedings of the National Academy of Sciences,* 111**,** 10269-10274.

HUANG, P., XIAO, A., ZHOU, M., ZHU, Z., LIN, S. & ZHANG, B. 2011. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotech,* 29**,** 699-700.

HUGHES, R. E., HATFULL, G. F., RICE, P., STEITZ, T. A. & GRINDLEY, N. D. 1990. Cooperativity mutants of the gamma delta resolvase identify an essential interdimer interaction. *Cell,* 63**,** 1331-8.

HURT, J. A., THIBODEAU, S. A., HIRSH, A. S., PABO, C. O. & JOUNG, J. K. 2003. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci U S A,* 100**,** 12271-6.

HUTTER, G., NOWAK, D., MOSSNER, M., GANEPOLA, S., MUSSIG, A., ALLERS, K., SCHNEIDER, T., HOFMANN, J., KUCHERER, C., BLAU, O., BLAU, I. W., HOFMANN, W. K. & THIEL, E. 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med,* 360**,** 692-8.

ISALAN, M. & CHOO, Y. 2001. Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol,* 340**,** 593-609.

ISALAN, M., KLUG, A. & CHOO, Y. 2001. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol,* 19**,** 656-60.

JAMIESON, A. C., MILLER, J. C. & PABO, C. O. 2003. Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov,* 2**,** 361-8.

JENSEN, N. M., DALSGAARD, T., JAKOBSEN, M., NIELSEN, R. R., SORENSEN, C. B., BOLUND, L. & JENSEN, T. G. 2011. An update on targeted gene repair in mammalian cells: methods and mechanisms. *J Biomed Sci,* 18**,** 10.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337**,** 816-21.

JOSHI, R., HO, K. K., TENNEY, K., CHEN, J. H., GOLDEN, B. L. & GIMBLE, F. S. 2011. Evolution of I-SceI homing endonucleases with increased DNA recognition site specificity. *J Mol Biol,* 405**,** 185-200.

JOUNG, J. K., RAMM, E. I. & PABO, C. O. 2000. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A,* 97**,** 7382-7.

JOUNG, J. K. & SANDER, J. D. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol,* 14**,** 49-55.

JOUNG, J. K., VOYTAS, D. F. & CATHOMEN, T. 2010. Reply to [ldquo]Genome editing with modularly assembled zinc-finger nucleases[rdquo]. *Nat Meth,* 7**,** 91-92.

KACZOROWSKI, T., SKOWRON, P. & PODHAJSKA, A. J. 1989. Purification and characterization of the FokI restriction endonuclease. *Gene,* 80**,** 209-16.

KAUFMANN, K. B., BUNING, H., GALY, A., SCHAMBACH, A. & GREZ, M. 2013. Gene therapy on the move. *EMBO Mol Med,* 5**,** 1642-61.

KAY, S. & BONAS, U. 2009. How Xanthomonas type III effectors manipulate the host plant. *Curr Opin Microbiol,* 12**,** 37-43.

KAY, S., HAHN, S., MAROIS, E., HAUSE, G. & BONAS, U. 2007. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science,* 318**,** 648-51.

KHAN, I. F., HIRATA, R. K., WANG, P. R., LI, Y., KHO, J., NELSON, A., HUO, Y., ZAVALJEVSKI, M., WARE, C. & RUSSELL, D. W. 2010. Engineering of human pluripotent stem cells by AAV-mediated gene targeting. *Mol Ther,* 18**,** 1192-9.

KIM, E., KIM, S., KIM, D. H., CHOI, B. S., CHOI, I. Y. & KIM, J. S. 2012. Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res,* 22**,** 1327-33.

KIM, H. & KIM, J. S. 2014. A guide to genome engineering with programmable nucleases. *Nat Rev Genet,* 15**,** 321-34.

KIM, H. J., LEE, H. J., KIM, H., CHO, S. W. & KIM, J. S. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res,* 19**,** 1279-88.

KIM, S., LEE, M. J., KIM, H., KANG, M. & KIM, J. S. 2011. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat Methods,* 8**,** 7.

KIM, Y., KWEON, J., KIM, A., CHON, J. K., YOO, J. Y., KIM, H. J., KIM, S., LEE, C., JEONG, E., CHUNG, E., KIM, D., LEE, M. S., GO, E. M., SONG, H. J., KIM, H., CHO, N., BANG, D. & KIM, J. S. 2013. A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol,* 31**,** 251-8.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A,* 93**,** 1156-60.

KIM, Y. G. & CHANDRASEGARAN, S. 1994. Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A,* 91**,** 883-7.

KLUG, A. 2010. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem,* 79**,** 213-31.

KNAUERT, M. P. & GLAZER, P. M. 2001. Triplex forming oligonucleotides: sequence-specific tools for gene targeting. *Hum Mol Genet,* 10**,** 2243-51.

KNAUERT, M. P., KALISH, J. M., HEGAN, D. C. & GLAZER, P. M. 2006. Triplex-stimulated intermolecular recombination at a single-copy genomic target. *Mol Ther,* 14**,** 392-400.

LAM, K. N., VAN BAKEL, H., COTE, A. G., VAN DER VEN, A. & HUGHES, T. R. 2011. Sequence specificity is obtained from the majority of modular C2H2 zinc-finger arrays. *Nucleic Acids Res,* 39**,** 4680-90.

LAMB, B. M., MERCER, A. C. & BARBAS, C. F., 3RD 2013. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res,* 41**,** 9779-85.

LEE, J., HIRSH, A. S., WITTNER, B. S., MAEDER, M. L., SINGAVARAPU, R., LANG, M., JANARTHANAN, S., MCDERMOTT, U., YAJNIK, V., RAMASWAMY, S., JOUNG, J. K. & SGROI, D. C. 2011. Induction of stable drug resistance in human breast cancer cells using a combinatorial zinc finger transcription factor library. *PLoS One,* 6**,** e21112.

LETT, M. C. 1988. Tn3-like elements: molecular structure, evolution. *Biochimie,* 70**,** 167-76.

LI, H., HAURIGOT, V., DOYON, Y., LI, T., WONG, S. Y., BHAGWAT, A. S., MALANI, N., ANGUELA, X. M., SHARMA, R., IVANCIU, L., MURPHY, S. L., FINN, J. D., KHAZI, F. R., ZHOU, S., PASCHON, D. E., REBAR, E. J., BUSHMAN, F. D., GREGORY, P. D., HOLMES, M. C. & HIGH, K. A. 2011a. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature,* 475**,** 217-21.

LI, H., PELLENZ, S., ULGE, U., STODDARD, B. L. & MONNAT, R. J., JR. 2009. Generation of single-chain LAGLIDADG homing endonucleases from native homodimeric precursor proteins. *Nucleic Acids Res,* 37**,** 1650-62.

LI, L., WU, L. P. & CHANDRASEGARAN, S. 1992. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A,* 89**,** 4275-9.

LI, T., HUANG, S., JIANG, W. Z., WRIGHT, D., SPALDING, M. H., WEEKS, D. P. & YANG, B. 2011b. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res,* 39**,** 359-72.

LI, T., HUANG, S., ZHAO, X., WRIGHT, D. A., CARPENTER, S., SPALDING, M. H., WEEKS, D. P. & YANG, B. 2011c. Modularly assembled designer TAL effector nucleases for

targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res,* 39**,** 6315-25.

LI, W., KAMTEKAR, S., XIONG, Y., SARKIS, G. J., GRINDLEY, N. D. & STEITZ, T. A. 2005. Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs. *Science,* 309**,** 1210-5.

LIANG, J., CHAO, R., ABIL, Z., BAO, Z. & ZHAO, H. 2014. FairyTALE: a high-throughput TAL effector synthesis platform. *ACS Synth Biol,* 3**,** 67-73.

LIANG, P., XU, Y., ZHANG, X., DING, C., HUANG, R., ZHANG, Z., LV, J., XIE, X., CHEN, Y., LI, Y., SUN, Y., BAI, Y., SONGYANG, Z., MA, W., ZHOU, C. & HUANG, J. 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell***,** 1-10.

LISBY, M., KROGH, B. O., BOEGE, F., WESTERGAARD, O. & KNUDSEN, B. R. 1998. Camptothecins Inhibit the Utilization of Hydrogen Peroxide in the Ligation Step of Topoisomerase I Catalysis. *Biochemistry,* 37**,** 10815-10827.

LIU, J., GAJ, T., PATTERSON, J. T., SIRK, S. J. & BARBAS, C. F., 3RD 2014. Cell-penetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. *PLoS One,* 9**,** e85755.

LIU, L. F., DESAI, S. D., LI, T.-K., MAO, Y., SUN, M. E. I. & SIM, S.-P. 2000. Mechanism of Action of Camptothecin. *Annals of the New York Academy of Sciences,* 922**,** 1-10.

LIU, Q., SEGAL, D. J., GHIARA, J. B. & BARBAS, C. F., 3RD 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A,* 94**,** 5525-30.

LIU, Q., XIA, Z., ZHONG, X. & CASE, C. C. 2002. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem,* 277**,** 3850-6.

LOMBARDO, A., GENOVESE, P., BEAUSEJOUR, C. M., COLLEONI, S., LEE, Y. L., KIM, K. A., ANDO, D., URNOV, F. D., GALLI, C., GREGORY, P. D., HOLMES, M. C. & NALDINI, L. 2007. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol,* 25**,** 1298-306.

LONKAR, P., KIM, K. H., KUAN, J. Y., CHIN, J. Y., ROGERS, F. A., KNAUERT, M. P., KOLE, R., NIELSEN, P. E. & GLAZER, P. M. 2009. Targeted correction of a thalassemia-associated beta-globin mutation induced by pseudo-complementary peptide nucleic acids. *Nucleic Acids Res,* 37**,** 3635-44.

MAEDER, M. L., THIBODEAU-BEGANNY, S., OSIAK, A., WRIGHT, D. A., ANTHONY, R. M., EICHTINGER, M., JIANG, T., FOLEY, J. E., WINFREY, R. J., TOWNSEND, J. A., UNGER-WALLACE, E., SANDER, J. D., MULLER-LERCH, F., FU, F., PEARLBERG, J., GOBEL, C., DASSIE, J. P., PRUETT-MILLER, S. M., PORTEUS, M. H., SGROI, D. C., IAFRATE, A. J., DOBBS, D., MCCRAY, P. B., JR., CATHOMEN, T., VOYTAS, D. F. & JOUNG, J. K. 2008. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell,* 31**,** 294-301.

MAEDER, M. L., THIBODEAU-BEGANNY, S., SANDER, J. D., VOYTAS, D. F. & JOUNG, J. K. 2009. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc,* 4**,** 1471-501.

MAGGIO, I., HOLKERS, M., LIU, J., JANSSEN, J. M., CHEN, X. & GONCALVES, M. A. 2014. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Sci Rep,* 4**,** 5105.

MAHER, L. J., 3RD 1996. Prospects for the therapeutic use of antigene oligonucleotides. *Cancer Invest,* 14**,** 66-82.

MAIER, D. A., BRENNAN, A. L., JIANG, S., BINDER-SCHOLL, G. K., LEE, G., PLESA, G., ZHENG, Z., COTTE, J., CARPENITO, C., WOOD, T., SPRATT, S. K., ANDO, D., GREGORY, P., HOLMES, M. C., PEREZ, E. E., RILEY, J. L., CARROLL, R. G., JUNE, C. H. & LEVINE, B. L.

2013. Efficient clinical scale gene modification via zinc finger nuclease-targeted disruption of the HIV co-receptor CCR5. *Hum Gene Ther,* 24**,** 245-58.

MAK, A. N., BRADLEY, P., CERNADAS, R. A., BOGDANOVE, A. J. & STODDARD, B. L. 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science,* 335**,** 716-9.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science,* 339**,** 823-6.

MANDELL, J. G. & BARBAS, C. F., 3RD 2006. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res,* 34**,** W516-23.

MARESCA, M., LIN, V. G., GUO, N. & YANG, Y. 2013. Obligate ligation-gated recombination (ObLiGaRe): custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res,* 23**,** 539-46.

MCCONNELL SMITH, A., TAKEUCHI, R., PELLENZ, S., DAVIS, L., MAIZELS, N., MONNAT, R. J., JR. & STODDARD, B. L. 2009. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc Natl Acad Sci U S A,* 106**,** 5099-104.

MENG, X., NOYES, M. B., ZHU, L. J., LAWSON, N. D. & WOLFE, S. A. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol,* 26**,** 695-701.

MERCER, A. C., GAJ, T., FULLER, R. P. & BARBAS, C. F., 3RD 2012. Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Res,* 40**,** 11163-72.

MEYER, M., DE ANGELIS, M. H., WURST, W. & KUHN, R. 2010. Gene targeting by homologous recombination in mouse zygotes mediated by zinc-finger nucleases. *Proc Natl Acad Sci U S A,* 107**,** 15022-6.

MILLER, J. C., HOLMES, M. C., WANG, J., GUSCHIN, D. Y., LEE, Y. L., RUPNIEWSKI, I., BEAUSEJOUR, C. M., WAITE, A. J., WANG, N. S., KIM, K. A., GREGORY, P. D., PABO, C. O. & REBAR, E. J. 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol,* 25**,** 778-85.

MILLER, J. C., TAN, S., QIAO, G., BARLOW, K. A., WANG, J., XIA, D. F., MENG, X., PASCHON, D. E., LEUNG, E., HINKLEY, S. J., DULAY, G. P., HUA, K. L., ANKOUDINOVA, I., COST, G. J., URNOV, F. D., ZHANG, H. S., HOLMES, M. C., ZHANG, L., GREGORY, P. D. & REBAR, E. J. 2011. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol,* 29**,** 143-8.

MILLER, J. C., ZHANG, L., XIA, D. F., CAMPO, J. J., ANKOUDINOVA, I. V., GUSCHIN, D. Y., BABIARZ, J. E., MENG, X., HINKLEY, S. J., LAM, S. C., PASCHON, D. E., VINCENT, A. I., DULAY, G. P., BARLOW, K. A., SHIVAK, D. A., LEUNG, E., KIM, J. D., AMORA, R., URNOV, F. D., GREGORY, P. D. & REBAR, E. J. 2015. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Methods,* 12**,** 465-71.

MINETA, Y., OKAMOTO, T., TAKENAKA, K., DOI, N., AOYAMA, Y. & SERA, T. 2008. Enhanced cleavage of double-stranded DNA by artificial zinc-finger nuclease sandwiched between two zinc-finger proteins. *Biochemistry,* 47**,** 12257-9.

MISIURA, A., PIGLI, Y. Z., BOYLE-VAVRA, S., DAUM, R. S., BOOCOCK, M. R. & RICE, P. A. 2013. Roles of two large serine recombinases in mobilizing the methicillin-resistance cassette SCCmec. *Molecular Microbiology,* 88**,** 1218-1229.

MOEHLE, E. A., ROCK, J. M., LEE, Y. L., JOUVENOT, Y., DEKELVER, R. C., GREGORY, P. D., URNOV, F. D. & HOLMES, M. C. 2007. Targeted gene addition into a specified

location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A,* 104**,** 3055-60.
MOORE, M., CHOO, Y. & KLUG, A. 2001a. Design of polyzinc finger peptides with structured linkers. *Proc Natl Acad Sci U S A,* 98**,** 1432-6.
MOORE, M., KLUG, A. & CHOO, Y. 2001b. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A,* 98**,** 1437-41.
MOSCOU, M. J. & BOGDANOVE, A. J. 2009. A simple cipher governs DNA recognition by TAL effectors. *Science,* 326**,** 1501.
MULLER, U. 1999. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev,* 82**,** 3-21.
MURLEY, L. L. & GRINDLEY, N. D. 1998. Architecture of the gamma delta resolvase synaptosome: oriented heterodimers identity interactions essential for synapsis and recombination. *Cell,* 95**,** 553-62.
MUSSOLINO, C., MORBITZER, R., LUTGE, F., DANNEMANN, N., LAHAYE, T. & CATHOMEN, T. 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res,* 39**,** 9283-93.
NEKLUDOVA, L. & PABO, C. O. 1994. Distinctive DNA conformation with enlarged major groove is found in Zn-finger-DNA and other protein-DNA complexes. *Proc Natl Acad Sci U S A,* 91**,** 6948-52.
NIELSEN, P. E. 2004. PNA Technology. *Mol Biotechnol,* 26**,** 233-48.
NIELSEN, P. E., EGHOLM, M. & BUCHARDT, O. 1994. Evidence for (PNA)2/DNA triplex structure upon binding of PNA to dsDNA by strand displacement. *J Mol Recognit,* 7**,** 165-70.
NISHIMASU, H., RAN, F. A., HSU, P. D., KONERMANN, S., SHEHATA, S. I., DOHMAE, N., ISHITANI, R., ZHANG, F. & NUREKI, O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell,* 156**,** 935-49.
NIU, Y., SHEN, B., CUI, Y., CHEN, Y., WANG, J., WANG, L., KANG, Y., ZHAO, X., SI, W., LI, W., XIANG, A. P., ZHOU, J., GUO, X., BI, Y., SI, C., HU, B., DONG, G., WANG, H., ZHOU, Z., LI, T., TAN, T., PU, X., WANG, F., JI, S., ZHOU, Q., HUANG, X., JI, W. & SHA, J. 2014. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell,* 156**,** 836-43.
NIU, Y., TENNEY, K., LI, H. & GIMBLE, F. S. 2008. Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J Mol Biol,* 382**,** 188-202.
NOMURA, W., MASUDA, A., OHBA, K., URABE, A., ITO, N., RYO, A., YAMAMOTO, N. & TAMAMURA, H. 2012. Effects of DNA binding of the zinc finger and linkers for domain fusion on the catalytic activity of sequence-specific chimeric recombinases determined by a facile fluorescent system. *Biochemistry,* 51**,** 1510-7.
OCHIAI, H. & YAMAMOTO, T. 2015. Genome Editing Using Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs). *In:* YAMAMOTO, T. (ed.) *Targeted Genome Editing Using Site-Specific Nucleases.* Springer Japan.
OLORUNNIJI, F. J., HE, J., WENWIESER, S. V. C. T., BOOCOCK, M. R. & STARK, W. M. 2008. Synapsis and catalysis by activated Tn3 resolvase mutants. *Nucleic Acids Research,* 36**,** 7181-7191.
OWENS, J. B., MAURO, D., STOYTCHEV, I., BHAKTA, M. S., KIM, M. S., SEGAL, D. J. & MOISYADI, S. 2013. Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. *Nucleic Acids Res,* 41**,** 9197-207.
PAILLARD, G., DEREMBLE, C. & LAVERY, R. 2004. Looking into DNA recognition: zinc finger binding specificity. *Nucleic Acids Res,* 32**,** 6673-82.

PATTANAYAK, V., RAMIREZ, C. L., JOUNG, J. K. & LIU, D. R. 2011. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods,* 8**,** 765-70.

PAVLETICH, N. P. & PABO, C. O. 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science,* 252**,** 809-17.

PEARSON, S., JIA, H. & KANDACHI, K. 2004. China approves first gene therapy. *Nat Biotech,* 22**,** 3-4.

PEREZ-PINERA, P., OUSTEROUT, D. G. & GERSBACH, C. A. 2012. Advances in targeted genome editing. *Curr Opin Chem Biol,* 16**,** 268-77.

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. & JUNE, C. H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

PORTEUS, M. H. 2006. Mammalian Gene Targeting with Designed Zinc Finger Nucleases. *Mol Ther,* 13**,** 438-446.

PRASEUTH, D., GUIEYSSE, A. L. & HELENE, C. 1999. Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim Biophys Acta,* 1489**,** 181-206.

PRIETO, J., REDONDO, P., PADRO, D., ARNOULD, S., EPINAT, J. C., PAQUES, F., BLANCO, F. J. & MONTOYA, G. 2007. The C-terminal loop of the homing endonuclease I-CreI is essential for site recognition, DNA binding and cleavage. *Nucleic Acids Res,* 35**,** 3262-71.

PROROCIC, M. M., WENLONG, D., OLORUNNIJI, F. J., AKOPIAN, A., SCHLOETEL, J. G., HANNIGAN, A., MCPHERSON, A. L. & STARK, W. M. 2011. Zinc-finger recombinase activities in vitro. *Nucleic Acids Res,* 39**,** 9316-28.

PROUDFOOT, C., MCPHERSON, A. L., KOLB, A. F. & STARK, W. M. 2011. Zinc finger recombinases with adaptable DNA sequence specificity. *PLoS One,* 6**,** e19537.

PRUETT-MILLER, S. M., CONNELLY, J. P., MAEDER, M. L., JOUNG, J. K. & PORTEUS, M. H. 2008a. Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Mol Ther,* 16**,** 707-17.

PRUETT-MILLER, S. M., CONNELLY, J. P., MAEDER, M. L., JOUNG, J. K. & PORTEUS, M. H. 2008b. Comparison of Zinc Finger Nucleases for Use in Gene Targeting in Mammalian Cells. *Mol Ther,* 16**,** 707-717.

RAMAKRISHNA, S., KWAKU DAD, A. B., BELOOR, J., GOPALAPPA, R., LEE, S. K. & KIM, H. 2014. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res,* 24**,** 1020-7.

RAMIREZ, C. & JOUNG, J. K. 2013. Engineered Zinc Finger Nucleases for Targeted Genome Editing. *In:* RENAULT, S. & DUCHATEAU, P. (eds.) *Site-directed insertion of transgenes.* Springer Netherlands.

RAMIREZ, C. L., CERTO, M. T., MUSSOLINO, C., GOODWIN, M. J., CRADICK, T. J., MCCAFFREY, A. P., CATHOMEN, T., SCHARENBERG, A. M. & JOUNG, J. K. 2012. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res,* 40**,** 5560-8.

RAMIREZ, C. L., FOLEY, J. E., WRIGHT, D. A., MULLER-LERCH, F., RAHMAN, S. H., CORNU, T. I., WINFREY, R. J., SANDER, J. D., FU, F., TOWNSEND, J. A., CATHOMEN, T., VOYTAS, D. F. & JOUNG, J. K. 2008. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat Methods,* 5**,** 374-5.

RAN, F. A., HSU, P. D., LIN, C. Y., GOOTENBERG, J. S., KONERMANN, S., TREVINO, A. E., SCOTT, D. A., INOUE, A., MATOBA, S., ZHANG, Y. & ZHANG, F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154**,** 1380-9.

REYON, D., TSAI, S. Q., KHAYTER, C., FODEN, J. A., SANDER, J. D. & JOUNG, J. K. 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol,* 30**,** 460-5.

RICCIARDI, A. S., MCNEER, N. A., ANANDALINGAM, K. K., SALTZMAN, W. M. & GLAZER, P. M. 2014. Targeted genome modification via triple helix formation. *Methods Mol Biol,* 1176**,** 89-106.

RICE, P. A., MOUW, K. W., MONTANO, S. P., BOOCOCK, M. R., ROWLAND, S. J. & STARK, W. M. 2010. Orchestrating serine resolvases. *Biochem Soc Trans,* 38**,** 384-7.

RICHARDSON, P. D., AUGUSTIN, L. B., KREN, B. T. & STEER, C. J. 2002. Gene repair and transposon-mediated gene therapy. *Stem Cells,* 20**,** 105-18.

RITACCO, C. J., KAMTEKAR, S., WANG, J. & STEITZ, T. A. 2013. Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase. *Nucleic Acids Res,* 41**,** 2673-82.

ROMER, P., HAHN, S., JORDAN, T., STRAUSS, T., BONAS, U. & LAHAYE, T. 2007. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science,* 318**,** 645-8.

ROSEN, L. E., MORRISON, H. A., MASRI, S., BROWN, M. J., SPRINGSTUBB, B., SUSSMAN, D., STODDARD, B. L. & SELIGMAN, L. M. 2006. Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res,* 34**,** 4791-800.

ROWLAND, S. J., BOOCOCK, M. R., MCPHERSON, A. L., MOUW, K. W., RICE, P. A. & STARK, W. M. 2009. Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome. *Mol Microbiol,* 74**,** 282-98.

SADELAIN, M. 2004. Insertional oncogenesis in gene therapy: how much of a risk? *Gene Ther,* 11**,** 569-73.

SANDER, J. D., CADE, L., KHAYTER, C., REYON, D., PETERSON, R. T., JOUNG, J. K. & YEH, J.-R. J. 2011a. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotech,* 29**,** 697-698.

SANDER, J. D., DAHLBORG, E. J., GOODWIN, M. J., CADE, L., ZHANG, F., CIFUENTES, D., CURTIN, S. J., BLACKBURN, J. S., THIBODEAU-BEGANNY, S., QI, Y., PIERICK, C. J., HOFFMAN, E., MAEDER, M. L., KHAYTER, C., REYON, D., DOBBS, D., LANGENAU, D. M., STUPAR, R. M., GIRALDEZ, A. J., VOYTAS, D. F., PETERSON, R. T., YEH, J. R. & JOUNG, J. K. 2011b. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods,* 8**,** 67-9.

SANDER, J. D. & JOUNG, J. K. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol,* 32**,** 347-55.

SANDER, J. D., MAEDER, M. L., REYON, D., VOYTAS, D. F., JOUNG, J. K. & DOBBS, D. 2010. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res,* 38**,** W462-8.

SANDER, J. D., ZABACK, P., JOUNG, J. K., VOYTAS, D. F. & DOBBS, D. 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res,* 35**,** W599-605.

SANDER, J. D., ZABACK, P., JOUNG, J. K., VOYTAS, D. F. & DOBBS, D. 2009. An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res,* 37**,** 506-15.

SANJANA, N. E., CONG, L., ZHOU, Y., CUNNIFF, M. M., FENG, G. & ZHANG, F. 2012. A transcription activator-like effector toolbox for genome engineering. *Nat Protoc,* 7**,** 171-92.

SCHIERLING, B., DANNEMANN, N., GABSALILOW, L., WENDE, W., CATHOMEN, T. & PINGOUD, A. 2012. A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res,* 40**,** 2623-38.

SCHLEIFMAN, E. B., BINDRA, R., LEIF, J., DEL CAMPO, J., ROGERS, F. A., UCHIL, P., KUTSCH, O., SHULTZ, L. D., KUMAR, P., GREINER, D. L. & GLAZER, P. M. 2011. Targeted disruption of the CCR5 gene in human hematopoietic stem cells stimulated by peptide nucleic acids. *Chem Biol,* 18**,** 1189-98.

SCHMID-BURGK, J. L., SCHMIDT, T., KAISER, V., HONING, K. & HORNUNG, V. 2013. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol,* 31**,** 76-81.

SEGAL, D. J., BEERLI, R. R., BLANCAFORT, P., DREIER, B., EFFERTZ, K., HUBER, A., KOKSCH, B., LUND, C. V., MAGNENAT, L., VALENTE, D. & BARBAS, C. F., 3RD 2003. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry,* 42**,** 2137-48.

SEGAL, D. J., DREIER, B., BEERLI, R. R. & BARBAS, C. F., 3RD 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A,* 96**,** 2758-63.

SEGAL, D. J. & MECKLER, J. F. 2013. Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet,* 14**,** 135-58.

SELIGMAN, L. M., CHISHOLM, K. M., CHEVALIER, B. S., CHADSEY, M. S., EDWARDS, S. T., SAVAGE, J. H. & VEILLET, A. L. 2002. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res,* 30**,** 3870-9.

SERA, T. & URANGA, C. 2002. Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry,* 41**,** 7074-81.

SHAH, S. A., ERDMANN, S., MOJICA, F. J. & GARRETT, R. A. 2013. Protospacer recognition motifs: mixed identities and functional diversity. *RNA Biol,* 10**,** 891-9.

SHALEM, O., SANJANA, N. E., HARTENIAN, E., SHI, X., SCOTT, D. A., MIKKELSEN, T. S., HECKL, D., EBERT, B. L., ROOT, D. E., DOENCH, J. G. & ZHANG, F. 2014. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science,* 343**,** 84-7.

SILVA, G., POIROT, L., GALETTO, R., SMITH, J., MONTOYA, G., DUCHATEAU, P. & PAQUES, F. 2011. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther,* 11**,** 11-27.

SILVA, G. H., BELFORT, M., WENDE, W. & PINGOUD, A. 2006. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J Mol Biol,* 361**,** 744-54.

SIRK, S. J., GAJ, T., JONSSON, A., MERCER, A. C. & BARBAS, C. F., 3RD 2014. Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants. *Nucleic Acids Res,* 42**,** 4755-66.

SMIH, F., ROUET, P., ROMANIENKO, P. J. & JASIN, M. 1995. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res,* 23**,** 5012-9.

SMITH, J., BIBIKOVA, M., WHITBY, F. G., REDDY, A. R., CHANDRASEGARAN, S. & CARROLL, D. 2000. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res,* 28**,** 3361-9.

SMITH, J., GRIZOT, S., ARNOULD, S., DUCLERT, A., EPINAT, J. C., CHAMES, P., PRIETO, J., REDONDO, P., BLANCO, F. J., BRAVO, J., MONTOYA, G., PAQUES, F. & DUCHATEAU,

P. 2006. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res,* 34**,** e149.

SMITH, M. C., BROWN, W. R., MCEWAN, A. R. & ROWLEY, P. A. 2010. Site-specific recombination by phiC31 integrase and other large serine recombinases. *Biochem Soc Trans,* 38**,** 388-94.

SMITHIES, O., GREGG, R. G., BOGGS, S. S., KORALEWSKI, M. A. & KUCHERLAPATI, R. S. 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature,* 317**,** 230-4.

SOLDNER, F., LAGANIERE, J., CHENG, A. W., HOCKEMEYER, D., GAO, Q., ALAGAPPAN, R., KHURANA, V., GOLBE, L. I., MYERS, R. H., LINDQUIST, S., ZHANG, L., GUSCHIN, D., FONG, L. K., VU, B. J., MENG, X., URNOV, F. D., REBAR, E. J., GREGORY, P. D., ZHANG, H. S. & JAENISCH, R. 2011. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell,* 146**,** 318-31.

SOOD, R., CARRINGTON, B., BISHOP, K., JONES, M., RISSONE, A., CANDOTTI, F., CHANDRASEKHARAPPA, S. C. & LIU, P. 2013. Efficient methods for targeted mutagenesis in zebrafish using zinc-finger nucleases: data from targeting of nine genes using CompoZr or CoDA ZFNs. *PLoS One,* 8**,** e57239.

SPIEGEL, P. C., CHEVALIER, B., SUSSMAN, D., TURMEL, M., LEMIEUX, C. & STODDARD, B. L. 2006. The structure of I-CeuI homing endonuclease: Evolving asymmetric DNA recognition from a symmetric protein scaffold. *Structure,* 14**,** 869-80.

STREUBEL, J., BLUCHER, C., LANDGRAF, A. & BOCH, J. 2012. TAL effector RVD specificities and efficiencies. *Nat Biotechnol,* 30**,** 593-5.

SUKHATME, V. P., CAO, X. M., CHANG, L. C., TSAI-MORRIS, C. H., STAMENKOVICH, D., FERREIRA, P. C., COHEN, D. R., EDWARDS, S. A., SHOWS, T. B., CURRAN, T. & ET AL. 1988. A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. *Cell,* 53**,** 37-43.

SUKHATME, V. P., KARTHA, S., TOBACK, F. G., TAUB, R., HOOVER, R. G. & TSAI-MORRIS, C. H. 1987. A novel early growth response gene rapidly induced by fibroblast, epithelial cell and lymphocyte mitogens. *Oncogene Res,* 1**,** 343-55.

SUSSMAN, D., CHADSEY, M., FAUCE, S., ENGEL, A., BRUETT, A., MONNAT, R., JR., STODDARD, B. L. & SELIGMAN, L. M. 2004. Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J Mol Biol,* 342**,** 31-41.

SWIRNOFF, A. H. & MILBRANDT, J. 1995. DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol Cell Biol,* 15**,** 2275-87.

SZCZEPEK, M., BRONDANI, V., BUCHEL, J., SERRANO, L., SEGAL, D. J. & CATHOMEN, T. 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol,* 25**,** 786-93.

TAN, W., ZHU, K., SEGAL, D. J., BARBAS, C. F., 3RD & CHOW, S. A. 2004. Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. *J Virol,* 78**,** 1301-13.

TEBAS, P., STEIN, D., TANG, W. W., FRANK, I., WANG, S. Q., LEE, G., SPRATT, S. K., SUROSKY, R. T., GIEDLIN, M. A., NICHOL, G., HOLMES, M. C., GREGORY, P. D., ANDO, D. G., KALOS, M., COLLMAN, R. G., BINDER-SCHOLL, G., PLESA, G., HWANG, W. T., LEVINE, B. L. & JUNE, C. H. 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med,* 370**,** 901-10.

THYAGARAJAN, B., OLIVARES, E. C., HOLLIS, R. P., GINSBURG, D. S. & CALOS, M. P. 2001. Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol,* 21**,** 3926-34.

THYME, S. B., BOISSEL, S. J., ARSHIYA QUADRI, S., NOLAN, T., BAKER, D. A., PARK, R. U., KUSAK, L., ASHWORTH, J. & BAKER, D. 2014. Reprogramming homing endonuclease specificity through computational design and directed evolution. *Nucleic Acids Res,* 42**,** 2564-76.

THYME, S. B., JARJOUR, J., TAKEUCHI, R., HAVRANEK, J. J., ASHWORTH, J., SCHARENBERG, A. M., STODDARD, B. L. & BAKER, D. 2009. Exploitation of binding energy for catalysis and design. *Nature,* 461**,** 1300-4.

TSAI, S. Q., WYVEKENS, N., KHAYTER, C., FODEN, J. A., THAPAR, V., REYON, D., GOODWIN, M. J., ARYEE, M. J. & JOUNG, J. K. 2014a. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol,* 32**,** 569-76.

TSAI, S. Q., WYVEKENS, N., KHAYTER, C., FODEN, J. A., THAPAR, V., REYON, D., GOODWIN, M. J., ARYEE, M. J. & JOUNG, J. K. 2014b. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotech,* 32**,** 569-576.

TSAI, S. Q., ZHENG, Z., NGUYEN, N. T., LIEBERS, M., TOPKAR, V. V., THAPAR, V., WYVEKENS, N., KHAYTER, C., IAFRATE, A. J., LE, L. P., ARYEE, M. J. & JOUNG, J. K. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol,* 33**,** 187-97.

URNOV, F. D., MILLER, J. C., LEE, Y. L., BEAUSEJOUR, C. M., ROCK, J. M., AUGUSTUS, S., JAMIESON, A. C., PORTEUS, M. H., GREGORY, P. D. & HOLMES, M. C. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature,* 435**,** 646-51.

URNOV, F. D., REBAR, E. J., HOLMES, M. C., ZHANG, H. S. & GREGORY, P. D. 2010. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet,* 11**,** 636-46.

VALTON, J., DUPUY, A., DABOUSSI, F., THOMAS, S., MARECHAL, A., MACMASTER, R., MELLIAND, K., JUILLERAT, A. & DUCHATEAU, P. 2012. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J Biol Chem,* 287**,** 38427-32.

VANAMEE, E. S., SANTAGATA, S. & AGGARWAL, A. K. 2001. FokI requires two specific DNA sites for cleavage. *J Mol Biol,* 309**,** 69-78.

VASQUEZ, K. M. & WILSON, J. H. 1998. Triplex-directed modification of genes and gene activity. *Trends Biochem Sci,* 23**,** 4-9.

WAEHLER, R., RUSSELL, S. J. & CURIEL, D. T. 2007. Engineering targeted viral vectors for gene therapy. *Nat Rev Genet,* 8**,** 573-87.

WANG, G., LEVY, D. D., SEIDMAN, M. M. & GLAZER, P. M. 1995. Targeted mutagenesis in mammalian cells mediated by intracellular triple helix formation. *Mol Cell Biol,* 15**,** 1759-68.

WANG, G., SEIDMAN, M. M. & GLAZER, P. M. 1996. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science,* 271**,** 802-5.

WANG, J., FRIEDMAN, G., DOYON, Y., WANG, N. S., LI, C. J., MILLER, J. C., HUA, K. L., YAN, J. J., BABIARZ, J. E., GREGORY, P. D. & HOLMES, M. C. 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res,* 22**,** 1316-26.

WANG, T., WEI, J. J., SABATINI, D. M. & LANDER, E. S. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science,* 343**,** 80-4.

WILSON, J. M. 2005. Gendicine: the first commercial gene therapy product. *Hum Gene Ther,* 16**,** 1014-5.

WILSON, M. H., KAMINSKI, J. M. & GEORGE, A. L., JR. 2005. Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. *FEBS Lett,* 579**,** 6205-9.

WIRTH, T., PARKER, N. & YLA-HERTTUALA, S. 2013. History of gene therapy. *Gene,* 525**,** 162-9.

WOLFE, S. A., GRANT, R. A., ELROD-ERICKSON, M. & PABO, C. O. 2001. Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes. *Structure,* 9**,** 717-23.

WOLFE, S. A., GREISMAN, H. A., RAMM, E. I. & PABO, C. O. 1999. Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J Mol Biol,* 285**,** 1917-34.

WU, S. C., MEIR, Y. J., COATES, C. J., HANDLER, A. M., PELCZAR, P., MOISYADI, S. & KAMINSKI, J. M. 2006. piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci U S A,* 103**,** 15008-13.

YANG, J., YUAN, P., WEN, D., SHENG, Y., ZHU, S., YU, Y., GAO, X. & WEI, W. 2013. ULtiMATE system for rapid assembly of customized TAL effectors. *PLoS One,* 8**,** e75649.

YANG, J., ZHANG, Y., YUAN, P., ZHOU, Y., CAI, C., REN, Q., WEN, D., CHU, C., QI, H. & WEI, W. 2014. Complete decoding of TAL effectors for DNA recognition. *Cell Res,* 24**,** 628-31.

YANG, W. & STEITZ, T. A. 1995. Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell,* 82**,** 193-207.

YANIK, M., ALZUBI, J., LAHAYE, T., CATHOMEN, T., PINGOUD, A. & WENDE, W. 2013. TALE-PvuII fusion proteins--novel tools for gene targeting. *PLoS One,* 8**,** e82539.

YIN, H., XUE, W., CHEN, S., BOGORAD, R. L., BENEDETTI, E., GROMPE, M., KOTELIANSKY, V., SHARP, P. A., JACKS, T. & ANDERSON, D. G. 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol,* 32**,** 551-3.

YLA-HERTTUALA, S. 2012. Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. *Mol Ther,* 20**,** 1831-2.

YLA-HERTTUALA, S. 2015. Glybera/'s Second Act: The Curtain Rises on the High Cost of Therapy. *Mol Ther,* 23**,** 217-218.

YUSA, K., RASHID, S. T., STRICK-MARCHAND, H., VARELA, I., LIU, P. Q., PASCHON, D. E., MIRANDA, E., ORDONEZ, A., HANNAN, N. R., ROUHANI, F. J., DARCHE, S., ALEXANDER, G., MARCINIAK, S. J., FUSAKI, N., HASEGAWA, M., HOLMES, M. C., DI SANTO, J. P., LOMAS, D. A., BRADLEY, A. & VALLIER, L. 2011. Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature,* 478**,** 391-4.

ZHANG, Z., LI, D., XU, H., XIN, Y., ZHANG, T., MA, L., WANG, X. & CHEN, Z. 2013. A simple and efficient method for assembling TALE protein based on plasmid library. *PLoS One,* 8**,** e66459.

ZHU, C., GUPTA, A., HALL, V. L., RAYLA, A. L., CHRISTENSEN, R. G., DAKE, B., LAKSHMANAN, A., KUPERWASSER, C., STORMO, G. D. & WOLFE, S. A. 2013. Using defined finger–finger interfaces as units of assembly for constructing zinc-finger nucleases. *Nucleic Acids Research,* 41**,** 2455-2465.

ZHU, C., MILLS, K. D., FERGUSON, D. O., LEE, C., MANIS, J., FLEMING, J., GAO, Y., MORTON, C. C. & ALT, F. W. 2002. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell,* 109**,** 811-21.

ZHU, C., SMITH, T., MCNULTY, J., RAYLA, A. L., LAKSHMANAN, A., SIEKMANN, A. F., BUFFARDI, M., MENG, X., SHIN, J., PADMANABHAN, A., CIFUENTES, D., GIRALDEZ,

A. J., LOOK, A. T., EPSTEIN, J. A., LAWSON, N. D. & WOLFE, S. A. 2011. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development,* 138**,** 4555-64.

ZURIS, J. A., THOMPSON, D. B., SHU, Y., GUILINGER, J. P., BESSEN, J. L., HU, J. H., MAEDER, M. L., JOUNG, J. K., CHEN, Z. Y. & LIU, D. R. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol,* 33**,** 73-80.