



Studying and Improving Lambda Red Recombination for Genome Engineering in Escherichia coli

Citation

Mosberg, Joshua Adam Weintrob. 2013. Studying and Improving Lambda Red Recombination for Genome Engineering in Escherichia coli. Doctoral dissertation, Harvard University.

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11156813>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

Studying and Improving Lambda Red Recombination for
Genome Engineering in *Escherichia coli*

A dissertation presented

by

Joshua Adam Weintrob Mosberg

to

The Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University
Cambridge, Massachusetts

April 2013

© 2013 – *Joshua Adam Weintrob Mosberg*
All rights reserved.

Studying and Improving Lambda Red Recombination for Genome Engineering in *Escherichia coli*

Abstract

The phage-derived Lambda Red recombination system utilizes exogenous DNA in order to generate precise insertion, deletion, and point mutations in *Escherichia coli* and other bacteria. Due to its convenience, it is a frequently-used tool in genetics and molecular biology, as well as in larger-scale genome engineering projects. However, limited recombination frequency constrains the usefulness of Lambda Red for several important applications. In this work, I utilize a mechanism-guided approach in order to improve the power and utility of Lambda Red recombination.

In Chapter 1, I discuss the capabilities of Lambda Red recombination, and introduce its notable past uses, particularly in genome engineering. I also summarize the current mechanistic understanding of Lambda Red, describe several past improvements of the recombination process, and discuss our motivation for improving Lambda Red recombination further.

In Chapter 2, I advance and support a novel mechanism for the Lambda Red recombination of dsDNA, in which Lambda Exonuclease entirely degrades one strand, leaving the other strand intact as ssDNA. This single-stranded intermediate then anneals at the lagging strand of the replication fork and is incorporated into the newly synthesized strand. In Chapter 3, I use this mechanistic insight to investigate new methods for improving dsDNA recombination frequency, finding that recombination frequency can significantly be enhanced by using phosphorothioate bonds to protect the 5' end of the lagging-targeting strand, and by removing the endogenous nuclease ExoVII.

Chapters 4 and 5 detail my efforts to improve the multiplex recombination of short oligonucleotides. In collaboration with Marc Lajoie and Christopher Gregg, I find that removing a set of five exonucleases results in significantly higher multiplex oligonucleotide recombination frequencies, as well as the improved inheritance of mutations carried on the 3' ends of oligonucleotides. Additionally, multiplex recombination frequencies were further improved by modifying DnaG primase so as to increase the amount of accessible ssDNA at the lagging strand of the replication fork.

Finally, in Chapter 6, I suggest additional ways in which Lambda Red recombination may be improved in the future, and discuss a recent project that illustrates how the developments described in this thesis have enabled exciting new applications.

Table of Contents

<u>Section</u>	<u>Page</u>
Abstract	iii
Table of Contents	v
List of Figures	vi
List of Tables	viii
Acknowledgements	ix
Chapter 1 Introduction: Improving Lambda Red Recombination to Create a Powerful Tool for Genome Engineering	1
Chapter 2 Lambda Red Recombination of Double-Stranded DNA Proceeds through a Fully Single-Stranded Intermediate	22
Chapter 3 Studying and Improving Lambda Red Double-Stranded DNA Recombination via Phosphorothioate Placement and Nuclease Removal	55
Chapter 4 Studying and Improving Lambda Red Oligonucleotide Recombination via Phosphorothioate Placement and Nuclease Removal	80
Chapter 5 Improving Lambda Red Oligonucleotide Recombination via Primase Modification	105
Chapter 6 Conclusion: Lambda Red Recombination, Today and Going Forward	139
Appendix 1 Strains Used in this Thesis	151

List of Figures

<u>Figure</u>	<u>Page</u>
Figure 2.1: The Court Model for Lambda Red dsDNA Recombination	24
Figure 2.2: The Poteete Model for Lambda Red dsDNA Recombination	25
Figure 2.3: Our Model for Lambda Red dsDNA Recombination	27
Figure 2.4: Testing the Proposed Overhang Intermediate for Lambda Red dsDNA Recombination	30
Figure 2.5: Co-electroporation of the Proposed Overhang Intermediate with Annealing Oligonucleotides	31
Figure 2.6: Strand Bias in Lambda Red ssDNA Insertion Recombination	32
Figure 2.7: Using Designed Mismatches to Assess the Mechanism of Lambda Red dsDNA Recombination	34
Figure 3.1: Diagram of the Variably Phosphorothioated (VPT) Cassette Series	59
Figure 3.2: <i>In vitro</i> Lambda Exo Digestion	59
Figure 3.3: Recombination Frequencies of the VPT Cassette Series in EcNR2	60
Figure 3.4: Recombination Frequencies of the VPT Cassette Series in nuc4 ⁻	63
Figure 3.5: VPT4:VPT1 Ratios of Tested Strains	64
Figure 3.6: VPT4:VPT7 Ratios of Tested Strains	64
Figure 3.7: Recombination Frequencies of the VPT Cassette Series in EcNR2. <i>xseA</i> ⁻	65
Figure 3.8: Removal of ExoVII Improves dsDNA Mutation Inheritance	67
Figure 4.1: Removal of ExoVII Improves Oligonucleotide Mutation Inheritance	84
Figure 4.2: Orientation of the Tested Oligo Sets	85

<u>Figure</u>	<u>Page</u>
Figure 4.3: Effect of Nuclease Removal on CoS-MAGE Performance	86
Figure 4.4: Effect of Phosphorothioate Bonds on CoS-MAGE Performance	88
Figure 5.1: Effect of DnaG Attenuation on Replication Fork Dynamics	107
Figure 5.2: DnaG Q576A Mutation Improves MAGE Performance	109
Figure 5.3: DnaG Mutations Improve CoS-MAGE Performance	110
Figure 5.4: Enhanced Recombination Frequencies are not Achieved by Targeting a Single Putative Okazaki Fragment	114
Figure 5.5: Testing DnaG Variants with an Expanded CoS-MAGE Oligo Set	116
Figure 5.6: Effect of DnaG Modification on Oligo-mediated Deletion Frequency	117
Figure 5.7: Effect of Primase Modification on Leading-targeting CoS-MAGE	119

List of Tables

<u>Table</u>	<u>Page</u>
Table 2.1: Recombination Results for Tracking Designed Mismatches	35
Table 2.2: Oligonucleotides used in Chapter 2	42
Table 3.1: Identifying the Nuclease(s) which Degrade Phosphorothioated Cassette Ends	63
Table 3.2: Oligonucleotides used in Chapter 3	71
Table 4.1: Oligonucleotides used in Chapter 4	94
Table 5.1: Estimation of EcNR2.dnaG.K580A and EcNR2.dnaG.Q576A Okazaki Fragment Lengths	108
Table 5.2: CoS-MAGE Performance versus EcNR2	111
Table 5.3: Summary of Mean Number of Alleles Converted per Clone	112
Table 5.4: Oligonucleotides used in Chapter 5	127

Acknowledgments

No scientist is an island, and there are a number of people to whom I owe a tremendous debt of gratitude. First and foremost, I thank my advisor, Prof. George Church, for his mentorship and support, and for cultivating a truly singular laboratory. I also appreciate the striking amount of intellectual freedom that he has afforded me during graduate school – I hope that I have done well with it. I also thank (now-Prof.) Farren Isaacs for generously taking the time to train me during the beginning of my time in the Church lab. Profs. Daniel Kahne, Frederick Ausubel, and Pamela Silver served on my Dissertation Advisory Committee (Dan and Pam also served on my Preliminary Qualifying Exam Committee), and provided me with valuable feedback and advice throughout the research process. I would also be remiss not to thank Yveta Masarova, Jason Millberg, Samantha Reed, KeyAnna Wright, Laura Glass, and Meghan Radden for their help and administrative support.

I am also extraordinarily thankful for my collaborators – in particular, Marc Lajoie and Christopher Gregg. Marc and I worked together on essentially all of the research presented in this thesis. As well as being a good friend and one of the sharpest scientists that I've encountered, he is also a tireless worker who consistently goes the extra mile to help his colleagues. I will miss working with him. Chris joined the lab about two years ago, and quickly established himself as an equally indispensable collaborator. It has been a pleasure to work alongside him, and he has contributed enormously to the work described here.

In addition to Marc and Chris, there are a number of other collaborators and co-workers for whom I am grateful. Sriram Kosuri, Harris Wang, Gabriel Washington, and Di Zhang all collaborated on portions of this work, and a number of other people – John Aach, Sara Vassallo, Nikolai Eroshenko, Francois Vigneault, Uri Laserson, Julie Norville, Xavier Rios, Srivatsan

Raman, Jaron Mercer, and Michael Napolitano – helped by providing technical assistance, valuable insight, and/or useful feedback. I am also thankful for the NRB 238 “Mantown” crew and all of my other co-workers who made the lab an engaging and fun environment. Lastly, I want to thank and give remembrance to Tara Gianoulis; though her time with us was tragically far too short, she made our lab a warmer, brighter, and more creative place.

Last but far from least, I thank my family and friends for helping me stay (usually) happy and grounded during my time in graduate school. In particular, I thank my parents, Henry Mosberg and Deborah Weintrob, for their unwavering and unconditional love and support – I would not be where I am without their support and sacrifice. Finally, I thank my girlfriend, Emily Lau. We met about halfway through my time in graduate school; she has made the second half far brighter than the first, and has made me a better and happier person.

Chapter One

Introduction: Improving Lambda Red Recombination to Create
a Powerful Tool for Genome Engineering

Synthetic biology – broadly defined by a European Commission expert working group as “the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature” – has recently emerged as an important and rapidly evolving scientific field.¹ While a detailed review of synthetic biology is well beyond the scope of this thesis, recent accomplishments representative of the diversity of the field include the engineering of a strain of yeast to produce the antimalarial drug precursor artemisinic acid,² the use of DNA origami³⁻⁵ to generate a “nanorobot” capable of selectively delivering a molecular payload (e.g., an antibody) to specific cell types,⁶ the *de novo* design of enzymes with novel catalytic activities,⁷⁻⁹ and the engineering of programmed genetic circuits within bacterial^{10,11} and mammalian^{11,12} cells.

While synthetic biology has benefited from advancements in a number of different technologies and capabilities – such as DNA sequencing, computational modeling, and a variety of analytical techniques – the development of the field is perhaps most directly a result of breakthroughs in DNA synthesis and editing. Specifically, advancements in the ability to “write” DNA have enabled the emergence of genome engineering as an important subfield of synthetic biology. Indeed, the past three years have seen two particularly notable accomplishments in bacterial genome engineering. First, in 2010, scientists from the J. Craig Venter Institute reported the creation of a self-replicating bacterial cell with an entirely synthetic genome, generated by transplantation of an *in vitro* synthesized *Mycoplasma mycoides* genome into a *Mycoplasma capricolum* cell.¹³ This achievement built upon a large body of prior methodology development,¹⁴⁻²⁰ and stands as the crowning achievement of *de novo* genome synthesis and transplantation reported to date.

A second major achievement in bacterial genome engineering was recently reported by George Church and colleagues. In this work, our group generated a strain of *Escherichia coli* in which all instances of the UAG stop codon were changed to the UAA stop codon, and the function of the UAG codon was subsequently reassigned.^{21,22} This strain demonstrated steps toward genetic isolation and virus resistance, and enabled the dedicated use of the UAG codon for the incorporation of non-standard amino acids.²² Rather than generating this strain by *de novo* genome synthesis as above, we utilized an approach in which we started with a wild type *E. coli* bacterium and edited its genome *in vivo* until it had the desired sequence.

While these two projects both leveraged recent advancements in DNA writing to achieve impressive ends, they embody two starkly different approaches to synthetic biology and genome engineering. In Venter's work, a genome was designed *de novo*, synthesized with the desired sequence, and then transplanted into a suitable recipient cell.¹³ While this approach has considerable promise and will likely enable powerful applications in the future, it comes with several drawbacks. For one, synthesizing entire genomes remains a substantial technical challenge. The *M. mycoides* genome synthesized at the Venter Institute has a size of 1.08 million base pairs (1.08 Mb), and represented the culmination of many years of intense labor and methodology development. However, this genome is considerably smaller than those of *E. coli* (4.64 Mb)²³ and most other bacterial species commonly used for research and biotechnological applications. Synthesizing a larger bacterial genome such as that of *E. coli* appears to be beyond current capabilities, and will likely require additional technological development.

A second drawback to this approach is its requirement for designing and synthesizing a viable genome sequence *de novo*. Because the genome must be synthesized in its entirety prior to transplantation, any errors or changes that result in a non-viable genome are not likely to be

discovered until the final transplantation step. Indeed, a 1 bp deletion in essential gene *dnaA* resulted in the initial inability to generate a viable transplanted *M. mycoides* genome, as “one wrong base out of more than 1 million in an essential gene rendered the genome inactive.”¹³ Thus, copying a genome with perfect fidelity (or generating a genome with only non-functional “watermark” changes, as in the described work¹³) remains a significant technical challenge. However, a far greater challenge arises if one is trying to synthesize a genome with functional alterations, as in our work reassigning the UAG codon.^{21,22} Because *de novo* genome synthesis does not provide for a straightforward way to troubleshoot non-viable genomes, this method therefore necessitates designing a viable genome *a priori*. Given that even minor genetic changes can have significant and unpredictable effects,²⁴ this constitutes a significant challenge, and limits the degree to which *de novo* genome synthesis can be used for ambitious genome engineering projects.

Rather than synthesizing genomes *de novo*, our group is pioneering methods for editing existing genomes *in vivo* to generate desired sequences and functions. This approach has several advantages. For one, it circumvents the need to undertake a laborious and expensive whole genome synthesis project in order to make a limited number of changes to a genome. Additionally, because our approach involves making genetic changes *in vivo*, it provides a straightforward means of identifying and troubleshooting lethal or deleterious design elements. Changes that result in a non-viable genotype will fail to be generated *in vivo*, and can therefore be identified and redesigned. Similarly, otherwise serious deleterious changes will have an immediately detectable phenotype, and can be altered or reverted if necessary. Furthermore, this approach has the additional advantage of enabling the simultaneous testing of multiple different genome designs. Because our designed genetic changes are encoded as discrete units, they can

be generated combinatorially, thereby allowing billions of genome designs to be created at once.²⁵ Desired clones can subsequently be isolated by screening or selection; this approach therefore also enables the use of directed evolution to generate and identify genomes which confer a targeted phenotype. Thus, editing genomes *in vivo* is a powerful, versatile, and practical alternative to *de novo* genome synthesis. As such, it is likely to be a cornerstone strategy for future efforts in synthetic biology and genome engineering.

Our *in vivo* genome engineering approach was enabled by the development and advancement of Lambda Red recombination (“recombineering”). This recombination system utilizes components derived from the Lambda phage in order to carry out efficient genetic modification in *Escherichia coli*²⁶ and a variety of other bacteria.²⁷⁻³² The Lambda Red recombination system is RecA-independent,²⁶ and can precisely recombine exogenous oligonucleotides³³ or double-stranded DNA (dsDNA) cassettes²⁶ at genomic loci specified by flanking homology regions of as little as 35 bp.³⁴ Lambda Red recombination can be used to create insertion, deletion, and mismatch mutations.²⁵ To generate an insertion or mismatch mutation, a heterologous or mismatched sequence is placed in between the two homology regions of the recombinogenic DNA; in contrast, a deletion can be generated by using a recombineering oligonucleotide or cassette with homology regions that flank the sequence to be deleted.²⁵ A combination of these approaches can be used to simultaneously delete a genomic region and replace it with a heterologous sequence (e.g., a selectable marker).³⁵

Three phage-derived proteins – Gam, Exo, and Beta – are required in order to mediate the efficient Lambda Red recombination of exogenous dsDNA cassettes. The first protein, Gam, prevents the degradation of linear dsDNA by the endogenous nucleases RecBCD and SbcCD.³⁶ Exo then degrades the recombinogenic dsDNA in a 5′ to 3′ manner, leaving single-stranded

DNA (ssDNA) in the recessed regions. Finally, Beta binds to the single-stranded regions produced by Exo, and facilitates recombination by promoting annealing to the homologous genomic target sites.^{37,38} For the Red-mediated recombination of exogenous ssDNA oligonucleotides, only the Beta protein is required, and it is well-accepted that Beta mediates the annealing of the recombinogenic oligonucleotide to its homology targets on the lagging strand of the replication fork.^{33,37} Once annealed at the replication fork, the oligo is subsequently incorporated into the newly synthesized strand as an Okazaki Fragment.³⁷ In contrast, the precise mechanism by which Lambda Red recombines dsDNA was largely uncharacterized prior to the work described in this thesis; this topic is covered in detail in Chapter 2.

In addition to its more recent uses in large-scale genome engineering projects, Lambda Red recombineering has been a broadly used tool in genetics and molecular biology since its development about 15 years ago.^{39,40} Lambda Red recombination is frequently used to make precise genetic changes on a number of different targets, such as bacterial chromosomes,^{33,41,42} plasmids,⁴³ phages,⁴⁴ and BACs,^{45,46} including those used for downstream applications in eukaryotes.^{47,48} Both ssDNA oligonucleotides³³ and dsDNA cassettes²⁶ are commonly used for recombineering, and both strategies have been used for a broad array of powerful applications. Recombineering with dsDNA is frequently used to insert exogenous genes onto a recombination target; this is facilitated by the fact that insertion cassettes can easily be generated by PCR. Homology regions can be encoded on the 5' end of PCR primers, and subsequent use of such primers for amplification of the desired insertion sequence will yield a dsDNA cassette ready for Lambda Red recombineering. Thus, this strategy is commonly applied to insert heterologous genes⁴² and pathways⁴⁹ onto the *E. coli* chromosome for use in metabolic engineering and other efforts. A similar dsDNA recombineering strategy is also frequently used to replace endogenous

genes with selectable markers, thereby facilitating the one-step generation of knockout mutants.³⁵ Beyond its use as a standard molecular biology technique, this one-step knockout method has also been employed in larger-scale projects such as the creation of a complete library of single-gene knockout *E. coli* strains,⁵⁰ as well as a strain of *E. coli* with 15% of its genome removed.⁵¹ Additionally, dsDNA recombineering has been applied as a cloning tool, facilitating the incorporation of large regions of BAC or chromosomal DNA onto plasmids,^{45,52} as well as the development of entirely new cloning strategies.⁵³

Similar to dsDNA recombineering, Lambda Red recombination with ssDNA has also found a variety of applications. Like dsDNA recombineering, ssDNA recombineering has been used to make precise genetic changes on a number of different targets, including the chromosomes of *E. coli*^{33,54} and other bacteria,^{54,55} as well as plasmids,⁴³ BACs,^{56,57} and phages.^{44,58} Moreover, the power and versatility of ssDNA recombineering has recently been bolstered by the discovery that Lambda Red is capable of simultaneously recombining multiple oligonucleotides at once.²⁵ This technique – called Multiplex Automatable Genome Engineering (“MAGE”), due to its ability to be automated and applied in iterative cycles – can be used for both generating genetic diversity and creating a single strain with a desired genotype. For the first category of applications, a pool of several oligonucleotides is introduced into cells in one or multiple cycles; different cells recombine different combinations of oligos, thereby giving rise to a highly diverse population. The resulting population of cells can then be selected or screened for a desired property. This technique has been applied to rapidly optimize the pathway coding for the biosynthesis of the small molecule lycopene,²⁵ and to improve the production of indigo by combinatorially modifying promoters in the aromatic amino acid biosynthesis pathway.⁵⁹ Alternatively, similar methodology can be used to generate a single strain with a desired

genotype. In this case, cycling can be continued until an isogenic fully converted population is attained; more practically, the desired strain can also be identified by screening after the appropriate number of recombination cycles. This strategy was used to generate the UAG-recoded *E. coli* strain described above,^{21,22} and to create strains of *E. coli* with multiple components of the protein translation machinery His-tagged for convenient co-purification.⁶⁰

The broad utility of Lambda Red recombineering – particularly the ability to use MAGE for genome engineering, as described above – is largely a result of a great deal of work that has gone into improving the process. While the RecA-independent recombination capabilities of the Lambda phage have been known for almost 50 years,^{61,62} it was not until 1998 that the Lambda Red proteins (and the similar phage-derived RecET system) were utilized for targeted bacterial genetic engineering.^{39,40} The potential usefulness of these phage-derived recombination systems was quickly recognized, initiating effort to improve their power, versatility, and convenience. In 2000, researchers from Donald Court's group isolated a portion of the Lambda genome containing the Lambda Red genes. They then placed this Lambda fragment on the *E. coli* chromosome, under the control of the temperature-inducible *cI857* repressor.²⁶ This system – also used in the work described in this thesis – allows for the heat-controlled induction of the Lambda Red proteins, and prevents toxicity due to undesired expression of Lambda genes *kil*⁶³ and *gam*.²⁶ The researchers also optimized several parameters – including induction time, dsDNA concentration, and flanking homology lengths – in order to maximize dsDNA recombination frequency.²⁶ However, given that most dsDNA recombineering applications have involved the use of selectable markers, recombination frequency was generally not considered to be of critical importance. Thus, little additional work was done to improve the performance of dsDNA recombineering.

In contrast to the work with dsDNA recombineering, several discoveries have resulted in significant improvement of the recombination frequency of ssDNA oligonucleotides. Initial work by the Court group achieved a rough optimization of oligo length, and provided anecdotal evidence that mutations encoded near the ends of an oligonucleotide are inherited less frequently than those encoded nearer to the center – a result that has been confirmed in subsequent studies.^{33,64,65} However, most of the recombineering oligonucleotides tested in this study demonstrated low recombination frequencies – typically well under 1%.³³ Oligo recombination frequency was subsequently improved by way of a number of different strategies.

First, it was observed that recombineering oligos that anneal at the lagging strand of the replication fork have significantly higher recombination frequencies than their leading-targeting counterparts.⁶⁶ This reinforces the ssDNA recombination model involving annealing at the lagging strand and subsequent incorporation as an Okazaki fragment;³⁷ moreover, targeting the lagging strand represents a simple and predictable way to ensure improved recombination frequencies. It was also observed that the magnitude of this lagging strand bias was quite variable.⁶⁶ Subsequent experiments showed that this was largely due to differences in the ability of mismatch repair proteins to recognize and revert the respective mismatches conferred by the lagging- and leading-targeting oligos.^{41,66} By removing the cell's ability to carry out mismatch repair entirely (e.g., by knocking out *mutS*), recombination frequencies can reliably be improved.⁴¹ If wholesale removal of the mismatch repair system is contraindicated – for example, in genome engineering applications where off-target mutations are highly undesired – recombination frequency can alternatively be improved by evading mismatch repair through the use of C-C mismatches,⁴¹ stretches of six or more mismatched bases,⁶⁷ or chemically modified

oligo bases such as 2'-fluorouridine, 5-methyldeoxycytidine, 2,6-diaminopurine, and isodeoxyguanosine.⁶⁴

Subsequent work from the Church lab has achieved further improvement of single-stranded oligonucleotide recombineering. First, an optimization of oligo concentration found that singleplex recombination frequency is maximal at roughly 1 μ M, and a more rigorous optimization of oligonucleotide length determined that 90 bp oligos – rather than the 70 bp oligos previously used by the Court group – yield the highest recombination frequencies.²⁵ Additionally, it was discovered that recombination frequency can further be improved by protecting oligos from nuclease degradation through the use of phosphorothioate bonds, particularly on the 5' end of oligonucleotides.²⁵ A correlation was also observed between internal oligonucleotide secondary structure and recombination frequency. Secondary structure of $\Delta G < -12.5$ kcal/mol was found to be markedly detrimental to recombination frequency, presumably as a result of intramolecular interactions interfering with the ability of an oligo to anneal to its recombination target.²⁵ Finally, our group found that oligos with a high degree of off-target genomic homology typically have depressed recombination frequencies.²¹ Thus, by minimizing internal secondary structure and off-target genomic homology, recombination frequency can be augmented further. Through the combined use of these strategies, we are routinely able to attain singleplex oligo recombination frequencies of around 30%²⁵ – a stark improvement from the initial reported frequency of 0.2% obtained by the Court group.³³

These enhanced singleplex recombination frequencies paved the way for the development of MAGE. Our group's discovery that Lambda Red is capable of simultaneously recombining multiple oligonucleotides would not have had practical implications if singleplex recombination frequencies had remained below 1%, as very few cells would be expected to recombine more

than one oligo. However, given singleplex recombination frequencies of about 30%, recombination with a pool of multiple oligonucleotides should result in a population of cells with a broad distribution of mutations. The frequency of desired mutations can further be increased by performing iterative cycles of recombination – i.e., Lambda Red induction, cell washing, introduction of oligos by electroporation, and subsequent cell regrowth.²⁵ Nevertheless, even a single cycle of MAGE facilitates significant recombination; for a cycle of MAGE using ten oligonucleotides, we found that cells recombined an average of 0.4 oligos, and that several clones recombined as many as 5 of the 10 oligonucleotides.⁶⁸

Further improvement of MAGE was enabled by the recent discovery that multiplex recombination frequencies can be enhanced by directing an oligo to repair a defective selectable marker near the loci targeted by the other recombineering oligonucleotides.⁶⁹ Subsequent selection for the repaired marker results in roughly fourfold higher recombination frequencies at the other targeted loci, due to enrichment for cells with high levels of recombination in the vicinity of the selectable marker.⁶⁹ This technique, called co-selection MAGE (CoS-MAGE), substantially improves the power of MAGE, yielding an average of 1.4 non-selectable oligos recombined per cycle with the 10-plex sets described above.⁶⁸

This improvement of Lambda Red recombination stands as a significant achievement in technology development, and enabled the varied recombineering applications described above. However, even with these enhancements, the power of recombineering still has some significant limitations. For one, the recombination frequency for the insertion of gene-sized (i.e., ~1 kb) dsDNA fragments remains quite low – on the order of 10^{-4} recombinants per viable cell.^{26,70} As a result, selectable markers must be used to identify these rare insertion events. This requires Red-mediated gene insertions to be performed serially, precluding their convenient use in large-

scale genome engineering. Improving the recombination frequency of gene insertion would enable a variety of powerful new applications. For one, this would greatly simplify pathway transfer, and would allow for the seamless introduction of genes without needing to insert and then remove selectable markers. Perhaps more importantly, this would also provide for the multiplex insertion of genes, enabling the combinatorial introduction of a panel of exogenous genes thought to be involved in facilitating a property of interest (e.g., heat resistance). The resulting population of strains could then be screened or selected for that property. Finally, improved gene insertion technology could be used to introduce hundreds of heterologous genes of interest (e.g., genes coding for biosynthetic enzymes) into a single strain. Such a strain would serve as a versatile vessel for bioproduction, with genes being turned on or off by MAGE as needed to produce a desired compound.

Similarly, augmenting the power of MAGE and CoS-MAGE would also be highly beneficial. Despite the considerable strides made in improving these techniques – as well as the impressive applications that they have already enabled – some substantial limitations remain. Even with CoS-MAGE, the average cell recombines just slightly more than 1 non-selectable oligo per cycle, and relatively few cells recombine more than four oligos out of ten.⁶⁸ This constrains the degree of diversity that can be attained, as well as the extent to which a genome can feasibly be reengineered. As an example, the reassignment of all 321 instances of the UAG stop codon was a laborious task, and required over ten scientist-years of work.²² Additional codons must also be reassigned in order to create an organism with a truly orthogonal genetic code;²² this would require tens of thousands of targeted mutations genome-wide,²⁴ well outside of the reach of our current Red-mediated genome engineering capabilities. Improving the average and top number of mutations generated per cycle will make such ambitious recoding

efforts more feasible, and will broadly enhance the power of MAGE for a number of additional applications in diversity generation and genome engineering.

In this thesis, I describe the studies that my collaborators and I have performed in order to improve Lambda Red dsDNA and ssDNA recombination in *E. coli*. For this work, we used a mechanism-guided approach in which we first conducted experiments to better understand the recombineering process and the factors that constrain its efficiency. We then used these mechanistic insights to suggest and test means for improving recombination frequency. Thus, this work enabled both the improved understanding of Lambda Red recombination, as well as the significant enhancement of its performance.

In Chapter 2, I discuss our efforts to determine the mechanism by which Lambda Red recombines dsDNA. In contrast to previously proposed mechanisms, we advance a model in which Lambda Exo entirely degrades one strand, leaving the other strand intact as ssDNA. This single-stranded intermediate then recombines via Beta-catalyzed annealing at the replication fork. We support this mechanism by showing that single-stranded insertion cassettes are highly recombinogenic, and that these cassettes preferentially target the lagging strand. In contrast, we find that the recombination intermediate suggested by previously proposed mechanisms is minimally recombinogenic. Finally, we demonstrate that the recombination of a dsDNA cassette containing multiple internal mismatches results in strand-specific mutations cosegregating roughly 80% of the time – a result much more consistent with our model than with previous models. Elucidating the mechanism for dsDNA recombination enabled our subsequent work to improve the process, discussed in Chapter 3.

In Chapter 3, we first examine the effect of phosphorothioate (PT) bonds on the recombination frequency of dsDNA cassettes. We show that using PT bonds to protect the 5'

end of the lagging-targeting strand results in a significant improvement of recombination frequency. In contrast, PT bonds on the leading-targeting strand diminish recombination frequency, presumably by preventing Lambda Exo from degrading that strand to generate the lagging-targeting intermediate. We also observed that dsDNA with PT bonds on both 5' ends is markedly recombinogenic, despite PT bonds blocking the action of Lambda Exo. This surprising result led us to discover that the exonuclease ExoVII readily degrades these 5' PT bonds, thereby enabling Lambda Exo to generate the ssDNA recombination intermediate. We subsequently show that ExoVII also degrades non-phosphorothioated cassette ends, and that its removal results in significantly better dsDNA recombination frequency and inheritance of mutations near the ends of a cassette. Thus, this work further supports our mechanism for dsDNA recombination, and demonstrates the impact that endogenous nucleases can have on this process. Additionally, by removing ExoVII and protecting the lagging-targeting strand with PT bonds, we establish two straightforward means of improving dsDNA recombination frequency.

In Chapter 4, we study the impact of endogenous nucleases on CoS-MAGE performance. We show that ExoVII also acts on ssDNA oligonucleotides, and that its removal improves CoS-MAGE frequencies and the inheritance of mutations encoded near the 3' ends of oligos. We go on to show that removing a set of five exonucleases (RecJ, ExoI, ExoVII, ExoX, and Lambda Exo) further improves the performance of CoS-MAGE. In comparison with our standard recombineering strain EcNR2 (*Escherichia coli* MG1655 $\Delta mutS::cat \Delta(ybhB-bioAB)::[\lambda cl857 \Delta(cro-ea59)::tetR-bla]$), the resulting “nuc5” strain demonstrates 46% more alleles converted per average clone, 200% more clones with five or more allele conversions, and 35% fewer clones without any allele conversions in a given 10-plex cycle of CoS-MAGE. Finally, we use these nuclease knockout strains to investigate and clarify the effects of oligonucleotide

phosphorothioation on recombination frequency. We show that PT bonds can be detrimental as well as beneficial, and that the net effect depends on the nuclease background of the recombineering strain.

In Chapter 5, we further improve the performance of CoS-MAGE, this time by manipulating the *E. coli* replisome. We find that certain mutations in DnaG primase increase the amount of accessible ssDNA at the lagging strand of the replication fork, and that this significantly improves multiplex oligonucleotide recombination frequency. Moreover, we show that primase modification and nuclease removal have additive beneficial effects on CoS-MAGE. By combining these two strategies, we generated a strain that demonstrates 111% more alleles converted per cycle, 527% more clones with five or more allele conversions, and 71% fewer clones with zero allele conversions, in comparison with EcNR2. This indicates that the number of oligonucleotides within the cell and the amount of accessible ssDNA at the lagging strand are both limiting factors for Lambda Red recombination. Thus, this work provides further insight into multiplex oligonucleotide recombination, and yields a breakthrough improvement in CoS-MAGE performance.

Finally, in Chapter 6, I take stock of the current state of Lambda Red recombination technology, and suggest additional ways by which recombineering performance may be improved in the future. I conclude by briefly describing a recent project from the Church lab. In this work, we utilized dsDNA recombineering and CoS-MAGE to demonstrate that many essential genes are amenable to radical codon changes, and that at least 13 codons are likely dispensable across the *E. coli* genome.²⁴ Hence, in addition to markedly advancing our understanding of Lambda Red recombination and improving its performance, this thesis work has also enabled compelling new applications.

References

1. Serrano, L. Synthetic biology: promises and challenges. *Mol Syst Biol* **3**, 158 (2007).
2. Ro, D.K. et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-3 (2006).
3. Rothemund, P.W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).
4. Dietz, H., Douglas, S.M. & Shih, W.M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725-30 (2009).
5. Douglas, S.M. et al. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-8 (2009).
6. Douglas, S.M., Bachelet, I. & Church, G.M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831-4 (2012).
7. Rothlisberger, D. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-5 (2008).
8. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science* **319**, 1387-91 (2008).
9. Siegel, J.B. et al. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309-13 (2010).
10. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-8 (2000).
11. Fung, E. et al. A synthetic gene-metabolic oscillator. *Nature* **435**, 118-22 (2005).
12. Danino, T., Mondragon-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326-30 (2010).
13. Gibson, D.G. et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-6 (2010).

14. Gibson, D.G. et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215-20 (2008).
15. Lartigue, C. et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science* **325**, 1693-6 (2009).
16. Gibson, D.G. et al. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A* **105**, 20404-9 (2008).
17. Benders, G.A. et al. Cloning whole bacterial genomes in yeast. *Nucleic Acids Res* **38**, 2558-69 (2010).
18. Lartigue, C. et al. Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632-8 (2007).
19. Gibson, D.G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* **37**, 6984-90 (2009).
20. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-5 (2009).
21. Isaacs, F.J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-53 (2011).
22. Lajoie, M.J. et al. Genomically Recoded Organisms Impart New Biological Functions. *Science (in revision)* (2013).
23. Blattner, F.R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-62 (1997).
24. Lajoie, M.J. et al. Towards a Radically Reassigned Genetic Code. (*submitted*) (2013).
25. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).
26. Yu, D. et al. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 5978-83 (2000).

27. Yu, B. et al. A method to generate recombinant *Salmonella typhi* Ty21a strains expressing multiple heterologous genes using an improved recombineering strategy. *Appl Microbiol Biotechnol* **91**, 177-88 (2011).
28. Liang, R. & Liu, J. Scarless and sequential gene modification in *Pseudomonas* using PCR product flanked by short homology regions. *BMC Microbiol* **10**, 209 (2010).
29. Kang, Y. et al. Knockout and pullout recombineering for naturally transformable *Burkholderia thailandensis* and *Burkholderia pseudomallei*. *Nat Protoc* **6**, 1085-104 (2011).
30. Bryan, A., Abbott, Z.D. & Swanson, M.S. Constructing unmarked gene deletions in *Legionella pneumophila*. *Methods Mol Biol* **954**, 197-212 (2013).
31. Pontes, M.H. & Dale, C. Lambda red-mediated genetic modification of the insect endosymbiont *Sodalis glossinidius*. *Appl Environ Microbiol* **77**, 1918-20 (2011).
32. van Kessel, J.C. & Hatfull, G.F. Efficient point mutagenesis in mycobacteria using single-stranded DNA recombineering: characterization of antimycobacterial drug targets. *Mol Microbiol* **67**, 1094-107 (2008).
33. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6 (2001).
34. Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. & Court, D.L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**, 206-23 (2009).
35. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5 (2000).
36. Kulkarni, S.K. & Stahl, F.W. Interaction between the *sbcC* gene of *Escherichia coli* and the *gam* gene of phage lambda. *Genetics* **123**, 249-53 (1989).
37. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88 (2002).

38. Poteete, A.R. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* **68**, 66-74 (2008).
39. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* **20**, 123-8 (1998).
40. Murphy, K.C. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**, 2063-71 (1998).
41. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100**, 15748-53 (2003).
42. Wang, Y. & Pfeifer, B.A. 6-deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in *E. coli*. *Metab Eng* **10**, 33-8 (2008).
43. Thomason, L.C., Costantino, N., Shaw, D.V. & Court, D.L. Multicopy plasmid modification with phage lambda Red recombineering. *Plasmid* **58**, 148-58 (2007).
44. Oppenheim, A.B., Rattray, A.J., Bubunencko, M., Thomason, L.C. & Court, D.L. In vivo recombineering of bacteriophage lambda by PCR fragments and single-strand oligonucleotides. *Virology* **319**, 185-9 (2004).
45. Lee, E.C. et al. A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65 (2001).
46. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**, e36 (2005).
47. Bouvier, J. & Cheng, J.G. Recombineering-based procedure for creating Cre/loxP conditional knockouts in the mouse. *Curr Protoc Mol Biol* **Chapter 23**, Unit 23 13 (2009).
48. Chaverroche, M.K., Ghigo, J.M. & d'Enfert, C. A rapid method for efficient gene replacement in the filamentous fungus *Aspergillus nidulans*. *Nucleic Acids Res* **28**, E97 (2000).

49. Lemuth, K., Steuer, K. & Albermann, C. Engineering of a plasmid-free *Escherichia coli* strain for improved in vivo biosynthesis of astaxanthin. *Microb Cell Fact* **10**, 29 (2011).
50. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
51. Posfai, G. et al. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044-6 (2006).
52. Fu, J. et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* **30**, 440-6 (2012).
53. Li, M.Z. & Elledge, S.J. MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. *Nat Genet* **37**, 311-9 (2005).
54. Datta, S., Costantino, N. & Court, D.L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109-15 (2006).
55. Gerlach, R.G., Jackel, D., Holzer, S.U. & Hensel, M. Rapid oligonucleotide-based recombineering of the chromosome of *Salmonella enterica*. *Appl Environ Microbiol* **75**, 1575-80 (2009).
56. Yang, Y. & Sharan, S.K. A simple two-step, 'hit and fix' method to generate subtle mutations in BACs using short denatured PCR fragments. *Nucleic Acids Res* **31**, e80 (2003).
57. Bird, A.W. et al. High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat Methods* **9**, 103-9 (2012).
58. Thomason, L.C., Oppenheim, A.B. & Court, D.L. Modifying bacteriophage lambda with recombineering. *Methods Mol Biol* **501**, 239-51 (2009).
59. Wang, H.H. et al. Genome-scale promoter engineering by coselection MAGE. *Nat Methods* **9**, 591-3 (2012).
60. Wang, H.H. et al. Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multi-enzyme catalysis. *ACS Synth Biol* **1**, 43-52 (2012).

61. Brooks, K. & Clark, A.J. Behavior of lambda bacteriophage in a recombination deficient strain of *Escherichia coli*. *J Virol* **1**, 283-93 (1967).
62. van de Putte, P., Zwenk, H. & Rorsch, A. Properties of four mutants of *Escherichia coli* defective in genetic recombination. *Mutat Res* **3**, 381-92 (1966).
63. Greer, H. The kil gene of bacteriophage lambda. *Virology* **66**, 589-604 (1975).
64. Wang, H.H., Xu, G., Vonner, A.J. & Church, G. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res* **39**, 7336-47 (2011).
65. Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. & Church, G.M. Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *PLoS One* **7**, e44638 (2012).
66. Li, X.T. et al. Identification of factors influencing strand bias in oligonucleotide-mediated recombination in *Escherichia coli*. *Nucleic Acids Res* **31**, 6674-87 (2003).
67. Sawitzke, J.A. et al. Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol* **421**, 171-99 (2007).
68. Lajoie, M.J., Gregg, C.J., Mosberg, J.A., Washington, G.C. & Church, G.M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res* **40**, e170 (2012).
69. Carr, P.A. et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40**, e132 (2012).
70. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**, 791-9 (2010).

Chapter Two

Lambda Red Recombination of Double-Stranded DNA Proceeds through a Fully Single-Stranded Intermediate

Portions of this chapter are adapted from the following published paper:

Mosberg, J.A. *, Lajoie, M.J. * & Church, G.M. Lambda Red Recombineering in *Escherichia coli* Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186**, 791-9 (2010).

*Indicates co-first authorship

Research contributions are as follows:

J. Mosberg and M. Lajoie jointly devised the proposed mechanism. J. Mosberg, M. Lajoie, and G. Church planned the experiments to test the mechanistic hypothesis, and interpreted the results. J. Mosberg and M. Lajoie performed the experiments. J. Mosberg wrote the majority of the published paper, with writing and editing contributions from M. Lajoie and G. Church.

Introduction

As discussed in Chapter 1, the Lambda Red recombination system has emerged as a powerful and prevalently used tool in genetics and molecular biology. Double-stranded DNA (dsDNA) recombination in particular has been applied toward a number of different ends, including replacing chromosomal genes,¹⁻⁴ developing novel cloning methods,^{5,6} and inserting heterologous genes and pathways into the *E. coli* chromosome,^{3,7} plasmids,⁸ and BACs.^{5,9} Additionally, larger-scale dsDNA recombineering projects have included creating a complete library of single-gene knockout *E. coli* strains,¹⁰ and generating an *E. coli* strain in which 15% of its original genomic material was removed.¹¹ However, despite its widespread use over the past 15 years, the mechanism of Lambda Red dsDNA recombination has remained a significant unanswered question. By elucidating this mechanism, we hope to identify ways to improve the functionality, ease, and versatility of Lambda Red dsDNA recombineering.

As noted previously, it is well-established that three Lambda Red proteins are necessary for carrying out efficient dsDNA recombination: Gam, Exo, and Beta.^{2,12} Gam prevents the degradation of linear dsDNA by the *E. coli* RecBCD and SbcCD nucleases; Lambda Exonuclease (Exo) degrades the dsDNA in a 5' to 3' manner, leaving single-stranded DNA in the recessed regions; and Beta binds to the single-stranded regions produced by Exo, facilitating recombination by promoting annealing to the homologous genomic target sites.¹³ Two mechanisms proposed prior to this work each claimed that Exo binds to both 5' ends of a dsDNA cassette, degrading in both directions simultaneously to produce a double-stranded region flanked on both sides by 3' overhangs.^{14,15} However, neither mechanism provided a comprehensive and convincing explanation of how this construct ultimately recombines with the chromosome.

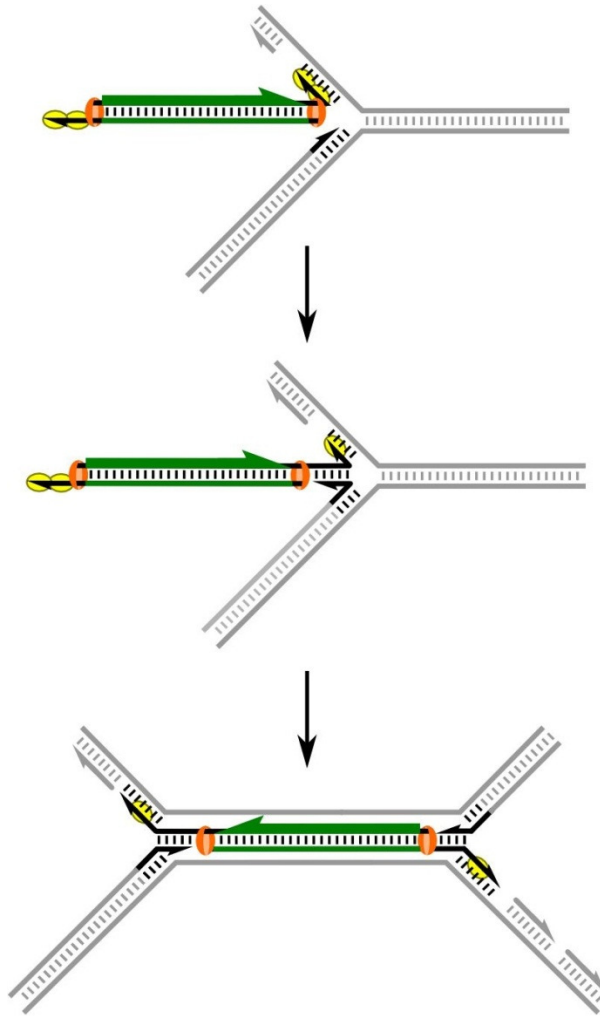


Figure 2.1: The Court Model for Lambda Red dsDNA Recombination. The Court mechanism posits that 1) Beta facilitates annealing of one 3' overhang to the lagging strand of the replication fork. 2) This replication fork then stalls and backtracks so that the leading strand can template switch onto the synthetic dsDNA. The heterologous dsDNA blocks further replication from this fork. 3) Once the second replication fork reaches the stalled fork, the other 3' end of the integration cassette is annealed to the lagging strand in the same manner as prior. Finally, the crossover junctions must be resolved by unspecified *E. coli* enzymes. For all diagrams in this chapter, heterologous dsDNA is shown in green; homology regions are shown in black; Exo is an orange oval, and Beta is a yellow oval. Figure adapted from Ref. 14.

Initially, it was proposed that Red-mediated recombination occurred via strand invasion.¹⁶ However, it has more recently been shown that Lambda Red dsDNA recombination remains highly proficient in a *recA*⁻ background.¹⁷ Thus, in the absence of the long regions of homology required for RecA-mediated recombination, strand invasion is unlikely to be a primary mechanism.

Furthermore, a detailed analysis of Lambda Red recombination products showed characteristics consistent with strand annealing rather than strand invasion.¹⁸ Finally, Lambda Red dsDNA recombination has been shown to preferentially target the lagging strand during DNA

replication, which suggests strand annealing rather than strand invasion.^{15,19}

To explain these results, Court *et al.* proposed a strand annealing model for insertional dsDNA recombination (Figure 2.1),¹⁴ in which one single-stranded 3' end anneals to its homologous target at the

replication fork. The replication fork then stalls due to the presence of a large dsDNA non-

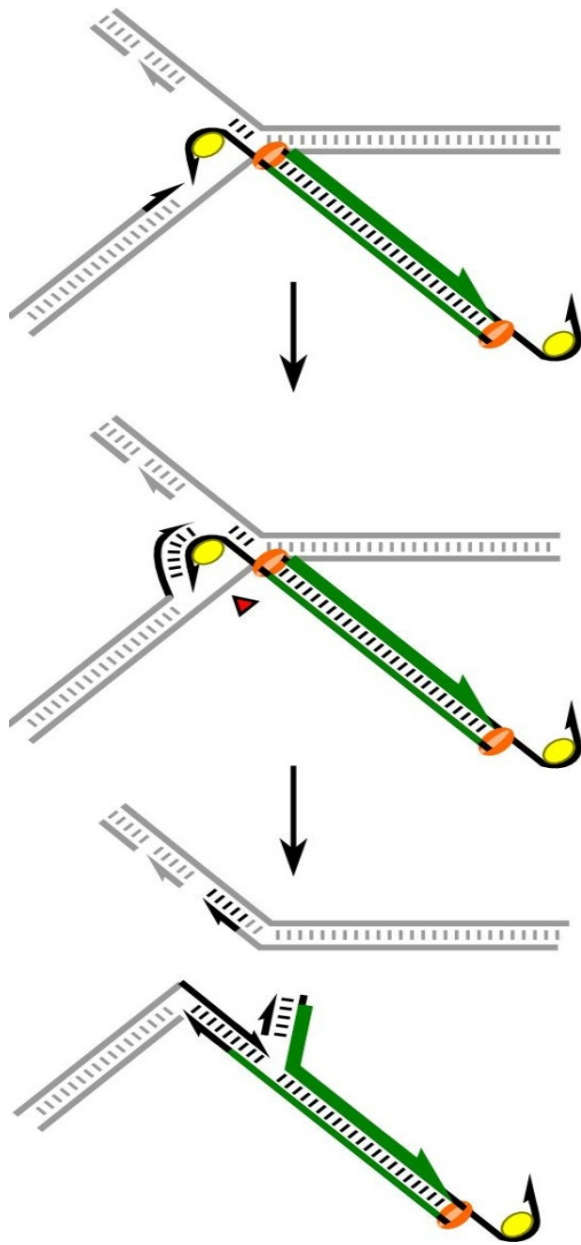


Figure 2.2: The Poteete Model for Lambda Red dsDNA Recombination. The Poteete mechanism suggests that 1) Beta facilitates 3' overhang annealing to the lagging strand of the replication fork, and 2) positions the invading strand to serve as the new template for leading strand synthesis. This structure is resolved by an unspecified host endonuclease (red triangle), and 3) the synthetic dsDNA becomes template for both lagging and leading strand synthesis. A second template switch must then occur at the other end of the synthetic dsDNA in the same manner as prior. Finally, the crossover junctions must be resolved by unspecified *E. coli* enzymes. Figure adapted from Ref. 15.

homology (i.e., the insertion cassette). The stalled replication fork is ultimately rescued by the second replication fork, traveling in the opposite direction around the circular bacterial chromosome. The other 3' end of the recombinogenic DNA anneals to the homology region exposed by the second replication fork, forming a crossover structure, which is then resolved by unspecified *E. coli* enzymes.¹⁴

The Court mechanism was challenged by Poteete,¹⁵ who showed that a dsDNA Lambda phage chromosome readily recombines onto a unidirectionally-replicating plasmid. Because such a plasmid lacks the second replication fork required by the Court mechanism, this result casts serious doubt on the Court model.¹⁴ Thus, Poteete proposed an alternate mechanism¹⁵ termed

“replisome invasion” (Figure 2.2), in which a 3' overhang of the Exo-processed dsDNA first anneals to its complementary sequence on the lagging strand of the replication fork. Subsequently, this overhang displaces the leading strand, thereby serving as the new template for leading strand

synthesis. The resulting structure is resolved by an unspecified endonuclease, after which the recombinogenic DNA becomes the template for the synthesis of both new strands. In the context of recombineering using a linear dsDNA cassette, a second strand switching event must then occur at the other end of the incoming recombinogenic dsDNA.

While Poteete's mechanism addresses some of the weaknesses of the Court mechanism, it is itself largely speculative. This mechanism does not identify the endonuclease responsible for resolving the structure after the first template switching event, nor does it explain how the recombinogenic DNA and replication machinery form a new replication fork. Additionally, the replisome would need to be transferred both on and off the incoming recombinogenic dsDNA in a well-coordinated manner, “like a train switching tracks.”²⁰ This may not be consistent with the relatively high recombination frequencies often observed for Lambda Red-mediated dsDNA insertion.² Finally, no direct experimental evidence has been advanced to support this model.

To address the deficiencies in the Court and Poteete models, we propose an alternative model in which Lambda Red dsDNA recombination proceeds via a fully single-stranded intermediate, in contrast with the dual-overhang intermediate proposed by Court and Poteete. According to our mechanism, Exo binds to one 5' end of the recombinogenic dsDNA cassette, and degrades that strand completely, leaving behind full-length ssDNA. This ssDNA then anneals to its homology targets at the lagging strand of the replication fork, and is incorporated into the newly-synthesized strand as an Okazaki fragment (Figure 2.3). This process is analogous to the accepted mechanism for the Lambda Red recombination of ssDNA oligonucleotides,¹⁴ and therefore unifies the mechanisms for ssDNA and dsDNA recombination. Notably, our mechanism also does not depend on the presence of multiple replication forks, thereby addressing Poteete's criticism of the Court mechanism.

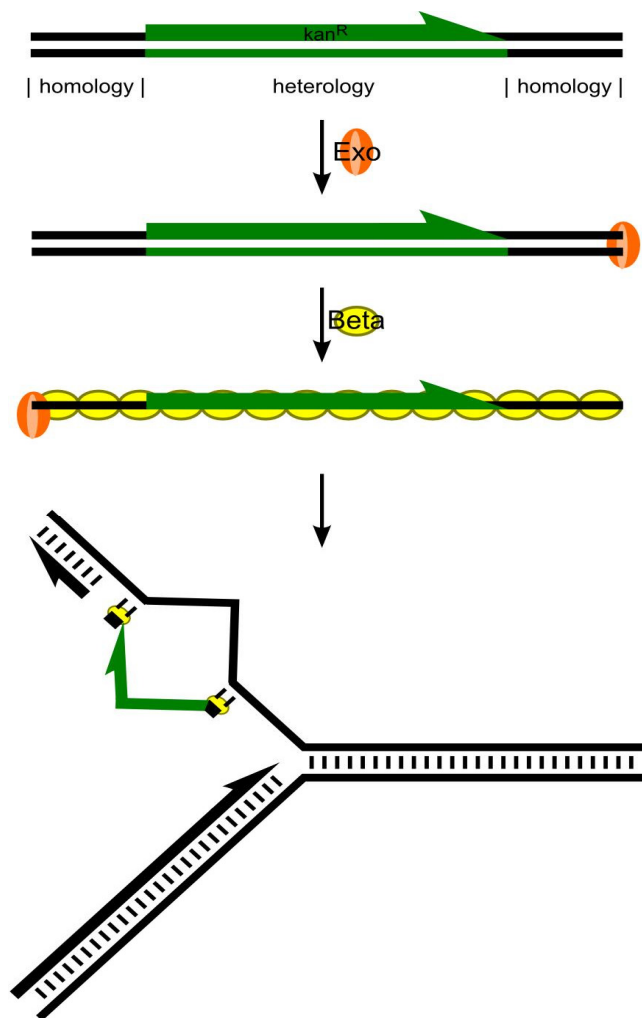


Figure 2.3: Our Model for Lambda Red dsDNA Recombination. Instead of a recombination intermediate involving dsDNA flanked by 3' ssDNA overhangs, we propose that one strand of linear dsDNA is entirely degraded by Exo. Beta then facilitates the annealing of the remaining strand to the lagging strand of the replication fork, with subsequent incorporation as an Okazaki fragment. The heterologous region does not anneal to the genomic sequence. This mechanism could account for gene replacement (as shown), or for insertions in which no genomic DNA is removed.

segregate during Lambda Red recombination. These results provide strong support of our proposed mechanism.

While it has not been previously suggested in this context, the degradation of an entire strand by Exo is feasible given the highly processive nature of the enzyme.²¹ Whereas previously proposed mechanisms assume that both dsDNA ends are degraded approximately simultaneously, our hypothesis implies that some dsDNA molecules will be entirely converted to ssDNA before a second Exo enzyme can bind to the other end. To support this mechanism, we demonstrate that single-stranded DNA is a highly recombinogenic intermediate and has a lagging strand bias. In contrast, we show that the intermediate suggested by the previously proposed mechanisms is minimally recombinogenic. Finally, we show that alleles carried on the ends of a given strand of dsDNA co-

Testing Predicted Recombination Intermediates

To test our mechanistic hypothesis, we first designed a *lacZ::kanR* cassette (~1.2 kb), consisting of a kanamycin resistance gene (*kanR*) flanked by 45 bp regions homologous to the *lacZ* gene on the *E. coli* chromosome. Successful Red-mediated recombination of the cassette at the *lacZ* chromosomal locus disrupts LacZ function, so proper targeting of the *lacZ::kanR* cassette could be verified by selecting on kanamycin and assaying for the inability to cleave X-Gal and release a blue chromophore (i.e., white colonies on X-Gal/IPTG). The *lacZ::kanR* dsDNA construct was generated by PCR and converted into lagging-targeting ssDNA using a biotin capture and DNA melting protocol.²² PAGE analysis confirmed the purity of the *lacZ::kanR* ssDNA construct, as no dsDNA band was observed. This construct was then recombined into the standard Church lab *E. coli* strain used for Lambda Red recombineering, EcNR2 (*Escherichia coli* MG1655 Δ mutS::cat Δ (ybhB-bioAB)::[λ CI857 Δ (cro-ea59)::tetR-bla]); see Appendix 1 for a list and description of strains used in this thesis).²³ The *lacZ::kanR* ssDNA construct was found to yield $1.3 \times 10^{-5} \pm 4.5 \times 10^{-6}$ recombinants per viable cell, in comparison with $1.9 \times 10^{-4} \pm 7.5 \times 10^{-5}$ for the corresponding dsDNA construct. Among the kanamycin resistant colonies, both ssDNA and dsDNA constructs gave over 99% white colonies, indicating correct targeting of the recombinogenic cassette.

This result demonstrates that a long ssDNA insertion cassette – the predicted intermediate for our mechanism – is considerably recombinogenic. It is, however, 14.8-fold less recombinogenic than the corresponding dsDNA. We hypothesize that this disparity is caused by ssDNA secondary structure, nuclease degradation of ssDNA, and/or the lack of Exo-Beta synergy. Previous work has demonstrated that ssDNA oligonucleotides longer than 90 bases and/or having secondary structure with $\Delta G < -12$ kcal/mol are likely to have substantially

reduced recombination frequency.²³ Given the length of the ~1.2 kb ssDNA cassette used in this experiment, we expect it to have significant secondary structure, and therefore, depressed recombination frequency. Secondly, it has been shown that several *E. coli* exonucleases readily degrade exogenous ssDNA, thereby compromising recombination frequency.^{24,25} Although our mechanism indicates that a dsDNA cassette must ultimately be processed into ssDNA, it is possible that this process occurs in a manner that mitigates the effects of secondary structure and/or nuclease degradation of the ssDNA intermediate. For instance, dsDNA may be converted to ssDNA shortly before recombination occurs, or this process may be coordinated with the binding of proteins that protect ssDNA and/or reduce internal secondary structure. Finally, it has previously been suggested²⁶ that Exo and Beta act synergistically, with Exo facilitating the binding of Beta to recessed regions of ssDNA. Since Exo does not bind to ssDNA, this synergistic action cannot occur on an exogenous ssDNA intermediate; therefore, recombination frequency may decrease in comparison with dsDNA. However, even in light of these considerations, it is clear that our predicted ssDNA intermediate is highly capable of recombination.

In order to confirm that the observed recombinants arose from *lacZ::kanR* ssDNA rather than from dsDNA contamination, this recombination experiment was repeated in SIMD90,²⁶ a strain of *E. coli* that expresses Beta, but lacks Exo and Gam. In the absence of Exo and Gam, dsDNA recombination frequency should decline significantly due to increased dsDNA degradation and inefficient processing into ssDNA. Indeed, *lacZ::kanR* dsDNA demonstrated a recombination frequency of only 8.7×10^{-7} in this strain, in comparison with a recombination frequency of 1.8×10^{-4} for the ssDNA (a 209-fold difference). This confirms that ssDNA – and

not contaminating dsDNA within the ssDNA sample – gives rise to recombinants in SIMD90.

Therefore, the observed recombinants in EcNR2 almost certainly also arose from ssDNA.

After the publication of the rest of the work described in this chapter,²⁷ we also performed experiments to test the dual-overhang intermediate suggested in the models proposed by Court and Poteete.^{14,15} This predicted intermediate was constructed by generating complementary individual strands of ssDNA in the same manner as above, then annealing the two strands together. We recombined this construct into EcNR2, side-by-side with the corresponding *lacZ::kanR* dsDNA and ssDNA. Results are shown in Figure 2.4.

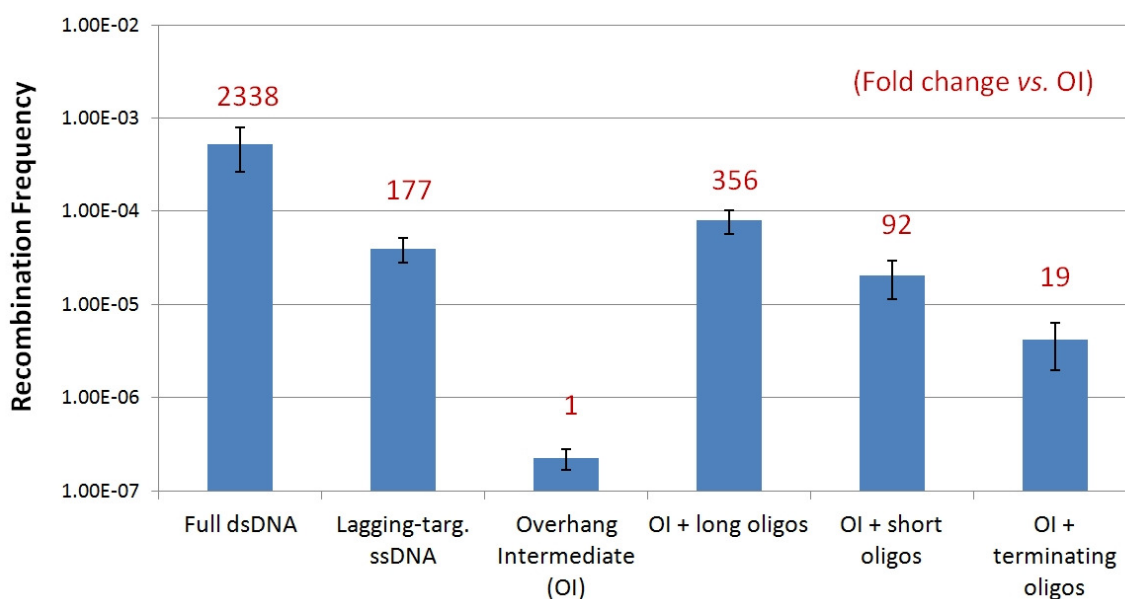


Figure 2.4: Testing the Proposed Overhang Intermediate for Lambda Red dsDNA Recombination. The overhang intermediate (OI) suggested by previous Lambda Red recombination mechanisms was generated by annealing together two ssDNA strands to give a dsDNA *kanR* insert, flanked on both sides by 45 bp 3' ssDNA *lacZ* homology overhangs. This construct was recombined into EcNR2 and found to have far lower recombination frequency than both the corresponding dsDNA as well as the full-length ssDNA intermediate suggested by our proposed mechanism. Co-electroporation of the overhang intermediate with oligonucleotides designed to anneal to its ssDNA homology regions (as shown in Figure 2.5) increased recombination frequency, indicating that extension of the overhang intermediate can readily occur within the cell. Thus, the recombinants observed for the overhang intermediate may arise due to the generation of one or both full-length strands via random priming and extension, followed by downstream recombination according to our proposed mechanism. Three independent replicates were performed for each sample, and error bars are given as the standard error of the mean.

We found that the overhang construct had 2340-fold lower recombination frequency than the corresponding dsDNA cassette, and 177-fold lower frequency than our predicted ssDNA

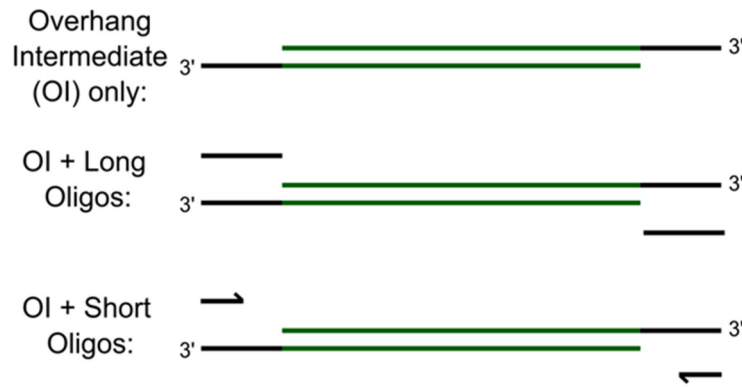


Figure 2.5: Co-electroporation of the Proposed Overhang Intermediate with Annealing Oligonucleotides. The overhang intermediate (OI) suggested by previous mechanisms was co-electroporated with either long (45 bp) oligonucleotides designed to anneal along the entire length of the homology overhang regions, or with short (20 bp) oligonucleotides designed to anneal to the 5' ends of the overhang regions. The short oligonucleotides could then be extended by an endogenous polymerase to generate one or both full-length strands. Short oligonucleotides with a chain-terminating dideoxy base on the 3' end were also tested. Recombination frequency results were as shown in Figure 2.4.

intermediate. Thus, it appears clear that this overhang construct is not a primary intermediate in Lambda Red dsDNA recombination. Indeed, the few recombinants observed with this construct may arise from the regeneration of a full-length double-stranded molecule via random priming and extension, rather than by recombination according to one of the

previously proposed mechanisms. To investigate this, we co-electroporated the overhang intermediate with long (45 bp) oligonucleotides designed to anneal along the entire length of the homology overhang regions, or short (20 bp) oligonucleotides designed to anneal to the 5' ends of the overhang regions, and then be extended by an endogenous polymerase to generate one or both full-length strands (Figure 2.5). Both co-electroporations gave substantial recombination frequency enhancements, with the short oligos providing a 92-fold improvement despite requiring the action of an endogenous polymerase. Capping the 3' ends of these oligos with dideoxy bases decreases this improvement 4.8-fold (Figure 2.4), consistent with this model. These results suggest that extension of the overhang intermediate can readily occur within the cell, and that the recombinants observed for the overhang intermediate may indeed arise due to random priming and extension, followed by downstream recombination according to our proposed mechanism.

Investigating the Strand Bias of the Recombination Intermediate

According to our proposed mechanism for Lambda Red dsDNA recombination, the full-length ssDNA intermediate recombines by annealing at the replication fork in the same manner as ssDNA oligonucleotides.¹⁴ It has been demonstrated that lagging-targeting oligonucleotides recombine with substantially greater frequency than the corresponding leading-targeting oligonucleotides, due to the greater accessibility of the lagging strand for annealing.²⁸ In order to test whether long ssDNA recombines in the same manner, we investigated whether several pairs of lagging-targeting and leading-targeting ssDNA insertion cassettes demonstrated a similar

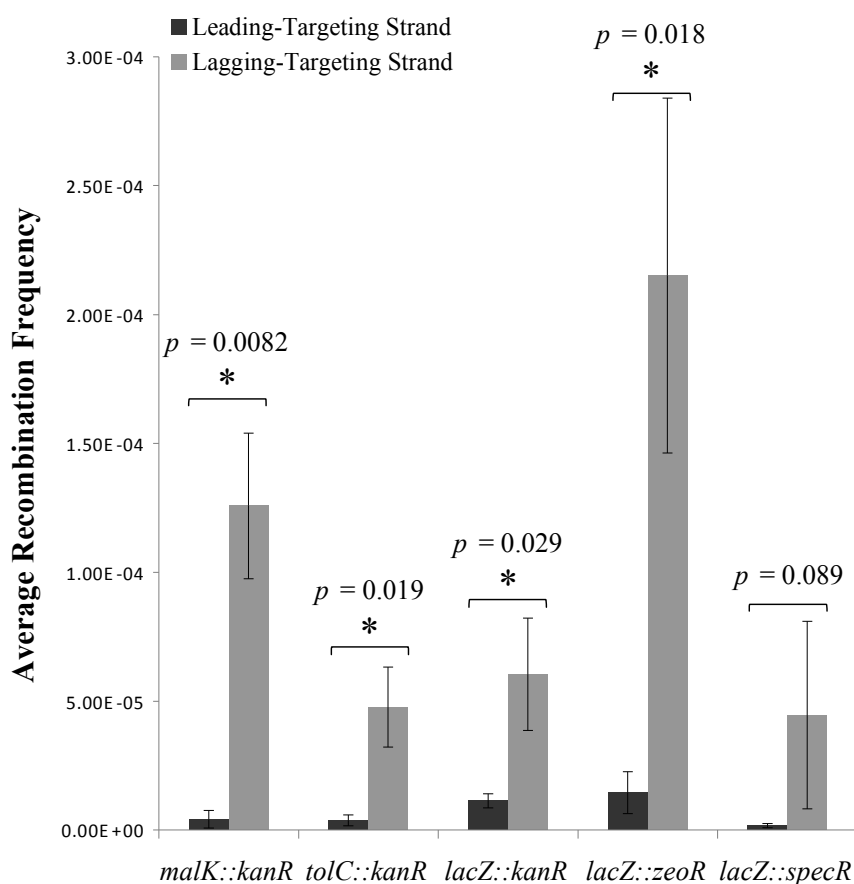


Figure 2.6: Strand Bias in Lambda Red ssDNA Insertion Recombination. Recombination frequencies were assessed for several leading-targeting and lagging-targeting complementary ssDNA pairs. Lagging-targeting strands were found to be uniformly more recombinogenic than leading-targeting strands. Error bars indicate standard deviation (n=3); * indicates $p < 0.05$.

strand bias. We controlled for differential secondary structure and other sequence-specific effects by testing the recombination of three different antibiotic resistance markers into *lacZ* – kanamycin (*lacZ::kanR*), zeocin (*lacZ::zeoR*), and spectinomycin (*lacZ::specR*).

Additionally, in order to demonstrate that any

strand bias was not caused by replicore-specific context or transcriptional direction, we constructed two additional *kanR* cassettes. To this end, *tolC::kanR* targets a gene located on the opposite replicore from *lacZ*, and *malK::kanR* targets a gene transcribed from the opposite strand of the chromosome as *lacZ*.

As shown in Figure 2.6, the lagging-targeting strand was substantially more recombinogenic than the leading-targeting strand for all five tested cassettes. As previously observed for oligonucleotides,²⁹ there appears to be a significant amount of locus-specific and sequence-specific variability in recombination frequency. Interestingly, a significant number of mistargeted recombinants (antibiotic-resistant colonies that retained LacZ function) were observed for both *lacZ::specR* strands. These mistargeted colonies were not scored as recombinants, and do not affect the broader interpretation of our results. Our results clearly indicate a robust lagging strand bias, likely arising from the greater accessibility of the lagging strand during DNA replication. This supports our claim that long ssDNA insertion constructs recombine by annealing at the replication fork in a manner similar to ssDNA oligonucleotides.

Testing Mechanistic Predictions by Tracking Designed Mutations

The preceding experiments provide strong indirect evidence supporting our proposed ssDNA annealing mechanism, and refuting the double overhang intermediate suggested by prior mechanisms. In order to more directly test the predictions of our mechanism, we designed a *lacZ::kanR* dsDNA cassette with internal mismatches (Figure 2.7), which enables us to empirically determine which strand provides genetic information during recombination. This construct was generated by annealing two strands of ssDNA and purifying the resulting dsDNA by agarose gel extraction. In each of the flanking *lacZ* homology regions, this construct contains

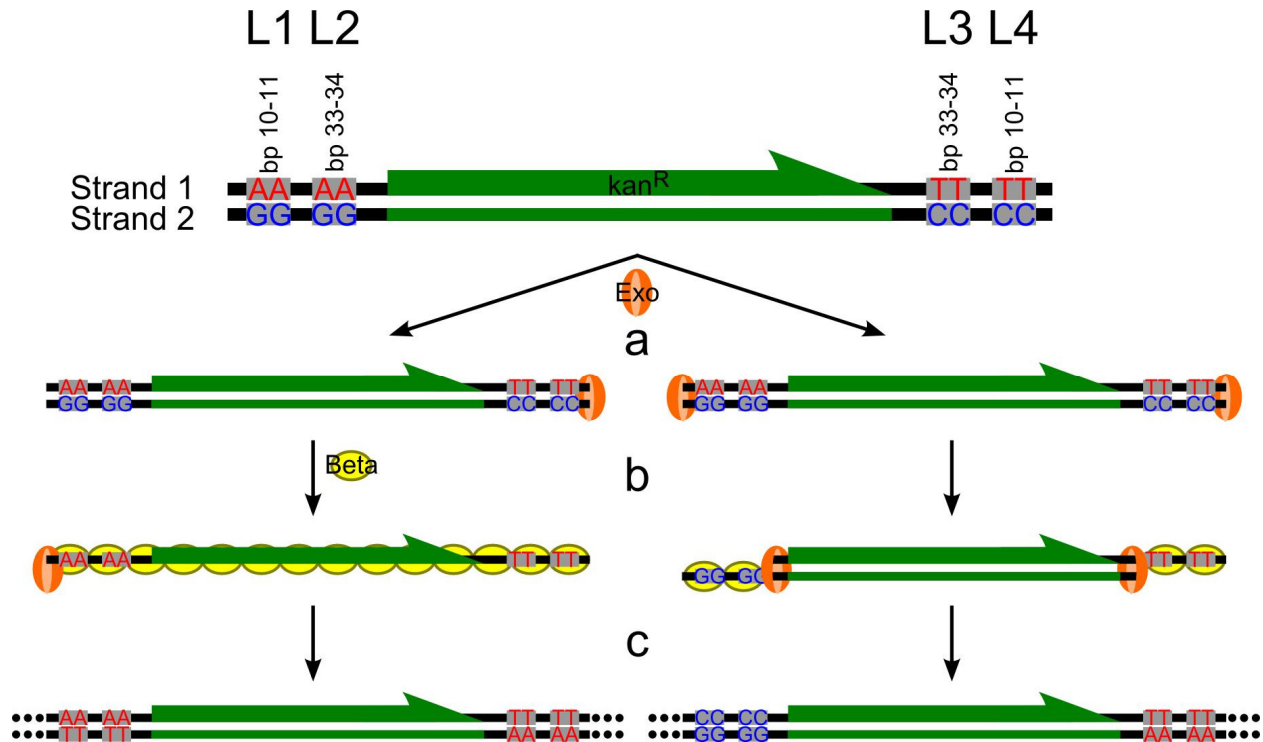


Figure 2.7: Using Designed Mismatches to Assess the Mechanism of Lambda Red dsDNA Recombination. Strand-specific mismatch alleles were used to identify the strand of origin for each recombined mutation. The mismatched *lacZ::kanR* cassette contained two consecutive internal mismatches at two loci in both flanking homology regions; at these loci, neither strand's sequence matched the targeted chromosomal copy of *lacZ*. Strand 1 was the lagging-targeting strand and strand 2 was the leading-targeting strand. If Lambda Red dsDNA recombination proceeds via a ssDNA intermediate (left), a) one Exo binds to a dsDNA end, b) Exo fully degrades one strand while helping to load Beta onto the remaining strand, and c) this strand provides all of the genetic information during recombination. This figure shows the case in which the lagging-targeting strand is recombined (coding strand genotypes: L1 = AA, L2 = AA, L3 = TT, L4 = TT), but the leading-targeting strand is also predicted to be observed (coding strand genotypes: L1 = CC, L2 = CC, L3 = GG, L4 = GG), albeit less frequently. If the Lambda Red recombination intermediate is a heterologous dsDNA core flanked by 3' ssDNA overhangs (right), a) one Exo binds to each dsDNA end, b) Exo recesses both strands while helping to load Beta onto both 3' overhangs, and c) both strands provide genetic information for each recombination. Since Exo always degrades 5'→3', the expected coding strand genotypes for the Court and Poteete mechanisms would be L1 = CC, L2 = CC, L3 = TT, L4 = TT.

two sets of adjacent dinucleotide mismatches that differentiate the two strands. At these loci, neither strand's sequence matches the targeted chromosomal copy of *lacZ*. Thus, one can infer which strand has recombined by observing which strand-specific alleles are present.

Our proposed ssDNA annealing mechanism can be distinguished from the prevailing recombination mechanisms based on the results of this experiment. Our mechanism predicts that the mutations located on either end of a single strand will be inherited together, and that the

mutations arising from the lagging-targeting strand will be observed more frequently than those arising from the opposite strand. Conversely, as detailed in Figure 2.7, the previously proposed mechanisms predict that the alleles on the 3' ends of both strands will be incorporated.

Table 2.1: Recombination Results for Tracking Designed Mismatches

Mutation Inheritance Pattern[§]	Count (Replicate 1)	Count (Replicate 2)	Total
1/1/1/WT	10	8	18
WT/1/1/1	0	3	3
WT/1/1/WT	24	23	47
Total – Inheritance from Strand 1 Only (Lagging-targeting)	34	34	68 (72%)
2/2/2/WT	0	1	1
WT/2/2/2	0	4	4
WT/2/2/WT	0	2	2
Total – Inheritance from Strand 2 Only (Leading-targeting)	0	7	7 (7%)
WT/1/1/2	4	1	5
WT/1/2/WT	4	0	4
WT/1/2/2	0	1	1
Total – Pattern Expected for 3'-to-5' Dual Resection	8	2	10 (11%)
WT/2/1/WT	5	3	8
WT/2/1/1	0	1	1
Total – Pattern Expected for 5'-to-3' Dual Resection	5	4	9 (10%)
Ambiguous	1	1	2
Sum	48	48	96

[§] Loci 1-4 (as in Figure 2.7) are listed in order. “1” indicates inheritance from strand 1, “2” indicates inheritance from strand 2, and “WT” indicates no mutation (*i.e.*, a wild type allele). Inheritance patterns are grouped based on the manner of Exo processing that is implied, as detailed at the bottom of each section.

This mismatched *lacZ::kanR* cassette was transformed into EcNR2, which is *mutS*⁻ and therefore deficient for mismatch repair. Recombinants were obtained by plating on kanamycin, and colonies were screened using mismatch amplification mutation assay (MAMA) PCR³⁰ in order to identify which strand-specific mutations were inherited in each colony. Two replicates were performed, and 48 colonies were screened for each replicate (Table 2.1). The accuracy of the assay was confirmed by sequencing the relevant alleles in several colonies, and by

performing a complementary MAMA PCR assay to detect unaltered wild type alleles at the targeted loci. In line with our predictions, we found that roughly 80% of the colonies inherited mismatch alleles from only one strand. Furthermore, of these colonies, 91% inherited mismatch alleles from only the lagging-targeting strand, strongly supporting our mechanism involving a full-length ssDNA intermediate.

Half of the remaining 20% of colonies showed an inheritance pattern consistent with resection from both 5' ends, and the other half was consistent with resection from both 3' ends. Resection from the 5' ends is predicted by the previously proposed mechanisms, and indicates that it is conceivable that one of these mechanisms also operates as a secondary process. However, Lambda Exo has not been shown to degrade dsDNA in a 3'→5' manner, and our results imply that this occurs equally as often as 5'→3' resection. Thus, we suggest two alternative explanations for the presence of colonies with apparent inheritance from both strands. First, another exonuclease may act on the dsDNA construct prior to Lambda Exo degrading one of the two strands. This nuclease could recess a portion of one of the two strands; subsequent re-extension by a polymerase would use the remaining strand as a template, and generate a dsDNA construct without internal mismatches in the affected region. Recombination could then occur according to our mechanism. Although only one of the two strands would be incorporated at the replication fork, the observed allele pattern would be consistent with inheritance from both strands, due to the previous nuclease degradation and repolymerization. This is particularly plausible for 3'→5' degradation, as ExoIII from *E. coli* is known to process dsDNA in this manner,³¹ and such a nuclease would leave 3' termini that could easily be re-extended by a polymerase. However, it is also conceivable that a similar process could occur for 5'→3' degradation. A second plausible explanation is that the colonies possessing alleles from both

strands may have undergone two sequential recombination events, each according to our proposed mechanism. The first recombination would proceed normally, and the second recombination would involve a partially degraded complementary strand. This second recombination event would be expected to occur quite frequently – after the first recombination event, the *kanR* gene is present in the genome, providing a large region of homology to which remaining fragments of *kanR* ssDNA can anneal in subsequent rounds of replication.

Interestingly, mutations arising from distal loci one and four (Figure 2.7) were observed only rarely in the studied recombinants (Table 2.1). This result suggests that a significant portion of DNA cassettes may be undergoing slight exonuclease degradation, or that annealed strands are processed at the replication fork in a manner that degrades or excludes the distal ends of the recombined DNA. This is consistent with a previous observation that mutations placed near the ends of a 90 bp oligonucleotide are inherited at a substantially lower frequency than mutations placed nearer to the center of the oligo.³² This observation and its implications will be discussed further in Chapters 3 and 4. Nevertheless, the results from this experiment provide strong direct evidence that our proposed single-stranded mechanism is the sole or dominant process by which Lambda Red dsDNA recombination occurs.

Discussion

This work provides strong empirical support for the proposed mechanism in which Lambda Red dsDNA recombination operates through a full-length ssDNA intermediate. This mechanism appears to be the dominant means of Lambda Red dsDNA recombination, although other mechanisms may still occur as minor processes. While our mechanism had not previously

been postulated as the manner by which the Lambda Red system recombines large dsDNA segments, it is consistent with numerous results observed by other groups.

By annealing two staggered oligonucleotides together, Yu *et al.* previously generated a 106 bp construct consisting of a dsDNA core flanked by 3' overhangs – the recombination intermediate predicted by the previous models for Lambda Red dsDNA recombination.³³ As the authors expected, recombination of this construct did not depend on the presence of Exo; however, even in the presence of Exo, the recombination frequency of this construct was roughly 4000-fold lower than that of its corresponding dsDNA. Given that the construct with 3' overhangs is postulated to be a downstream intermediate of this dsDNA, this result casts doubt upon the claim that the tested construct is indeed the predominant recombination intermediate. However, this result is explained by our proposed mechanism – only the intact dsDNA can generate the full-length ssDNA needed to undergo recombination, as neither individual strand of the construct containing 3' overhangs is sufficient for recombination.³³ We suggest that this 3' overhang construct recombines by a separate and disfavored process, or that one or both full-length strands are generated by random priming and extension prior to recombination. This line of reasoning is supported by the fact that the 3' overhang construct had no greater recombination frequency than the corresponding structure with 5' (rather than 3') overhangs.³³ It is unlikely that either of these structures represents the predominant intermediate in dsDNA recombination.

Muyrers *et al.*¹² have also provided evidence contrary to a dsDNA recombination intermediate with 3' overhangs. The authors created a dsDNA construct in which phosphorothioate linkages placed between an antibiotic resistance gene and its flanking genomic homology regions were used to prevent exonuclease degradation beyond these homology regions. Two 5'-to-3' exonucleases other than Exo were then used *in vitro* to resect the 5' ends of

this construct, in order to generate the putative intermediate for dsDNA recombination. However, it was found that none of the tested resection conditions could produce a construct that would recombine in the absence of Exo. Thus, the experiments by Yu *et al.* and Muyrers *et al.* reinforce our result showing that the previously proposed intermediate containing 3' overhangs is poorly recombinogenic and therefore highly unlikely to be the primary intermediate in Lambda Red dsDNA recombination.

Additionally, other prior work supports our proposed mechanism by reinforcing the processive nature of Exo. Hill *et al.* showed that non-replicating Lambda phage in *E. coli* is capable of converting linear phage dsDNA into ssDNA, creating single-stranded regions that span more than 1.4 kb.³⁴ They also demonstrated that *exo* is sufficient for generating these regions of ssDNA, which are similar in length to the ~1.2 kb constructs used in this experiment. An additional implication of this result is that a single-stranded intermediate is also present during crosses involving an intact Lambda chromosome. This suggests that our proposed mechanism may also be relevant for natural recombination between Lambda phage and bacterial chromosomes or plasmids.

The results of Lim *et al.*¹⁹ further reinforce that Exo generates long strands of ssDNA. These researchers created a dsDNA construct in which two antibiotic resistance genes were attached via a genome homology region and flanked with two additional regions of genome homology. Using this cassette, only about 10% of recombinants incorporated both resistance genes, while a majority of recombinants incorporated only one of the two. This implies that a majority of recombination events involved the central homology region, which is roughly 1 kb away from either end of the dsDNA construct. Given that strand annealing requires exposed

ssDNA, this result further supports that Exo is substantially processive *in vivo*, degrading large stretches of DNA rather than short flanking segments.

Finally, around the time this work was initially published, Maresca *et al.*³⁵ described complementary experiments in which strand-specific 5' phosphorylation and phosphorothioation were used to bias Exo degradation to each strand of a selectable cassette, separately. For both *in vitro* and *in vivo* Exo-mediated digestion, the authors observed a lagging-targeting strand bias in the subsequent recombination. Building upon these observations, the authors identified ssDNA as a recombinogenic species, and proposed the same mechanism as the one advanced in our work. These results further validate that Lambda Red dsDNA recombination proceeds through a fully single-stranded intermediate. Our experiment involving the mismatched dsDNA cassette extends this work by showing that information from a single strand co-segregates during Lambda Red dsDNA recombination, thereby providing direct evidence of a single-stranded intermediate.

This proposed mechanism may also describe other recombineering processes mediated by Lambda Red. One example is gap repair cloning, in which linear plasmid DNA is used to capture chromosomal DNA and generate a circularized plasmid.^{5,14} Notably, while a detailed mechanism has not yet been advanced for Lambda Red-facilitated gap repair, our model involving a single-stranded intermediate provides a plausible explanation. Given a full-length ssDNA intermediate, the linearized plasmid would anneal to the chromosomal target with its homology regions facing one another. The 3' end homology would then be elongated in the direction of the 5' end homology, thereby introducing the chromosomal DNA of interest onto the plasmid. The resulting linear single-stranded plasmid would be circularized by ligase in the same manner as Okazaki fragment joining. The circular ssDNA would then be liberated from the chromosome, possibly during chromosomal replication. Finally, the circular ssDNA would

be replicated to form a dsDNA plasmid, potentially via priming by residual ssDNA from the other strand of the linearized plasmid. Notably, this mechanism explains how gap repair cloning of large (> 80 kb) genomic sequences can occur,³⁶ since the two homology regions could anneal with multiple Okazaki fragments between them. These fragments would then be joined via the natural lagging strand replication mechanism.

The mechanism of Lambda Red dsDNA recombination has long been a matter of debate.³⁷ Here, we propose that Lambda Red dsDNA recombination proceeds via the annealing of a full-length ssDNA intermediate to the lagging strand of the replication fork. We support this mechanism with a body of evidence derived from both our work and from previous studies. Just as the mechanistic understanding of Red-mediated oligonucleotide recombination facilitated its profound optimization,²³ this work will likely indicate new strategies for improving the frequency and robustness of dsDNA recombineering. One such example of how this mechanistic knowledge was translated into a tangible improvement in recombination frequency is discussed in the following chapter.

Experimental Methods

Preparation of DNA Constructs

PCR primers were ordered from Integrated DNA Technologies and are listed and described in Table 2.2. Phosphorothioated primers contained four phosphorothioate linkages on the 5' end, dual-biotinylated primers contained a dual-biotin tag on the 5' end, and dideoxy-terminated primers contained a dideoxycytidine base terminating the 3' end. All primers were ordered with standard desalting, except dual-biotinylated primers, which were HPLC-purified.

Table 2.2: Oligonucleotides used in Chapter 2

Name	Use	Sequence	Versions used
LacZ::KanR.full-f	Forward strand for generation of the initial LacZ::KanR construct	TGACCATGATTACGGATTCA CTGGCCGTCGTTTTACAACG TCGTGCCTGTGACGGAAGAT CACTTCG	Unmodified
LacZ::KanR.full-r	Reverse strand for generation of the initial LacZ::KanR construct	GTGCTGCAAGGCGATTAAAGT TGGGTAACGCCAGGGTTTTTC CCAGTAACCAGCAATAGACA TAAGCGG	Unmodified
MalK::KanR.full-f	Forward strand for generation of the initial MalK::KanR construct	AATGTTGCTGTGATGACAG GTTGTTACAAAGGGAGAAGG GCATGCCTGTGACGGAAGAT CACTTCG	Unmodified
MalK::KanR.full-r	Reverse strand for generation of the initial MalK::KanR construct	GACCTCGCCCCAGGCTTTTCG TTACATTTTGCAGCTGTACGC TCGCAACCAGCAATAGACAT AAGCGG	Unmodified
TolC::KanR.full-f	Forward strand for generation of the initial TolC::KanR construct	AGTTTGATCGCGCTAAATAC TGCTTCACCACAAGGAATGC AAATGCCTGTGACGGAAGAT CACTTCG	Unmodified
TolC::KanR.full-r	Reverse strand for generation of the initial TolC::KanR construct	GAACCCAGAAAGGCTCAGGC CGATAAGAATGGGGAGCAAT TTCTTAACCAGCAATAGACA TAAGCGG	Unmodified
LacZ::ZeoR.full-f	Forward strand for generation of the initial LacZ::ZeoR construct	TGACCATGATTACGGATTCA CTGGCCGTCGTTTTACAACG TCGTGGGTGTTGACAATTAA TCATCGGC	Unmodified
LacZ::ZeoR.full-r	Reverse strand for generation of the initial LacZ::ZeoR construct	GTGCTGCAAGGCGATTAAAGT TGGGTAACGCCAGGGTTTTTC CCAGTAGCTTGCAAATTAAA GCCTTCG	Unmodified
LacZ::SpecR.full-f	Forward strand for generation of the initial LacZ::SpecR construct	TGACCATGATTACGGATTCA CTGGCCGTCGTTTTACAACG TCGTGCAGCCAGGACAGAAA TGC	Unmodified
LacZ::SpecR.full-r	Reverse strand for generation of the initial LacZ::SpecR construct	GTGCTGCAAGGCGATTAAAGT TGGGTAACGCCAGGGTTTTTC CCAGTTGCAGAAATAAAAAG GCCTGC	Unmodified
LacZ.short-f	Forward strand for generation of dual-biotinylated LacZ-targeting constructs	TGACCATGATTACGGATTCA CT	Unmodified, Phosphorothioated, Dual-biotinylated
LacZ.short-r	Reverse strand for generation of dual-biotinylated LacZ-targeting constructs	GTGCTGCAAGGCGATTAA	Phosphorothioated, Dual-biotinylated
LacZ.trunc-f	Forward primer for generation of 3' overhang intermediate precursor	CCTGTGACGGAAGATCACTT	Unmodified

Table 2.2 (Continued)

LacZ.trunc-r	Reverse primer for generation of 3' overhang intermediate precursor	AACCAGCAATAGACATAAGC G	Unmodified
Overhang.Anneal.S1	Short annealing oligo #1 for co-electroporation with 3' overhang intermediate	TGACCATGATTACGGATTCA CTGGCC	Unmodified, dideoxy-terminated
Overhang.Anneal.S2	Short annealing oligo #2 for co-electroporation with 3' overhang intermediate	GTGCTGCAAGGCGATTAAAGT TGGGTAAC	Unmodified, dideoxy-terminated
Overhang.Anneal.L1	Long annealing oligo #1 for co-electroporation with 3' overhang intermediate	TGACCATGATTACGGATTCA CTGGCCGTCGTTTACAACG TCGTG	Unmodified
Overhang.Anneal.L2	Long annealing oligo #2 for co-electroporation with 3' overhang intermediate	GTGCTGCAAGGCGATTAAAGT TGGGTAACGCCAGGGTTTTC CCAGT	Unmodified
MalK::KanR.short-f	Forward strand for generation of dual-biotinylated MalK::KanR constructs	AATGTTGCTGTCGATGACAG G	Phosphorothioated, Dual-biotinylated
MalK::KanR.short-r	Reverse strand for generation of dual-biotinylated MalK::KanR constructs	GACCTCGCCCCAGGC	Phosphorothioated, Dual-biotinylated
TolC::KanR.short-f	Forward strand for generation of dual-biotinylated TolC::KanR constructs	AGTTTGATCGCGCTAAATAC TG	Phosphorothioated, Dual-biotinylated
TolC::KanR.short-r	Reverse strand for generation of dual-biotinylated TolC::KanR constructs	GAACCCAGAAAGGCTCAGG	Phosphorothioated, Dual-biotinylated
MM.LacZ::KanR.AA-f	Forward strand for generation of Construct AA (For the creation of mismatched LacZ::KanR)	TGACCATGAAAACGGATTCA CTGGCCGTCGTTAAACAACG TCGTGCCTGTGACGGAAGAT CACTTCG	Unmodified
MM.LacZ::KanR.AA-r	Reverse strand for generation of Construct AA (For the creation of mismatched LacZ::KanR)	GTGCTGCAAAAACGATTAAAGT TGGGTAACGCCAAAGTTTTC CCAGTAACCAGCAATAGACA TAAGCGG	Unmodified
MM.LacZ::KanR.CC-f	Forward strand for generation of Construct CC (For the creation of mismatched LacZ::KanR)	TGACCATGACCACGGATTCA CTGGCCGTCGTTCCACAACG TCGTGCCTGTGACGGAAGAT CACTTCG	Unmodified
MM.LacZ::KanR.CC-r	Reverse strand for generation of Construct CC (For the creation of mismatched LacZ::KanR)	GTGCTGCAACCCGATTAAAGT TGGGTAACGCCACCGTTTTC CCAGTAACCAGCAATAGACA TAAGCGG	Unmodified
MM.AA.short-f	Forward strand for generation of dual-biotinylated Construct AA	TGACCATGAAAACGGATTCA C	Unmodified
MM.AA.short.DB-r	Reverse strand for generation of dual-biotinylated Construct AA	GTGCTGCAAAAACGATTAAAGT TG	Dual-biotinylated

Table 2.2 (Continued)

MM.CC.short.DB-f	Forward strand for generation of dual-biotinylated Construct CC	TGACCATGACCACGGATTC	Dual-biotinylated
MM.CC.short-r	Reverse strand for generation of dual-biotinylated Construct CC	GTGCTGCAACCCGATTAAG	Unmodified
Kan.L1.AA.set1	Forward MAMA PCR primer corresponding to the strand 1 specific mismatch at position 1 in MM.LacZ::KanR	CAGGAAACAGCTATGACCAT GAAA	Unmodified
Kan.L2.AA.set1	Forward MAMA PCR primer corresponding to the strand 1 specific mismatch at position 2 in MM.LacZ::KanR	GATTCAGTGGCCGTCGTAA	Unmodified
Kan.L3.TT.set1	Forward MAMA PCR primer corresponding to the strand 1 specific mismatch at position 3 in MM.LacZ::KanR	ATTGCTGGTTACTGGGAAAA CTT	Unmodified
Kan.L4.TT.set1	Forward MAMA PCR primer corresponding to the strand 1 specific mismatch at position 4 in MM.LacZ::KanR	GGCGTTACCCAACTTAATCG TT	Unmodified
Kan.L1.AA.set2	Forward MAMA PCR primer corresponding to the strand 2 specific mismatch at position 1 in MM.LacZ::KanR	GGAAACAGCTATGACCATGA CC	Unmodified
Kan.L2.AA.set2	Forward MAMA PCR primer corresponding to the strand 2 specific mismatch at position 2 in MM.LacZ::KanR	TCACTGGCCGTCGTTCC	Unmodified
Kan.L3.TT.set2	Forward MAMA PCR primer corresponding to the strand 2 specific mismatch at position 3 in MM.LacZ::KanR	GCTGGTTACTGGGAAAACGG	Unmodified
Kan.L4.TT.set2	Forward MAMA PCR primer corresponding to the strand 2 specific mismatch at position 4 in MM.LacZ::KanR	GCGTTACCCAACTTAATCGG G	Unmodified
Kan.L1.TT.setWT	Forward MAMA PCR primer corresponding to the wild type allele at position 1 in MM.LacZ::KanR	CAGGAAACAGCTATGACCAT GATT	Unmodified
Kan.L2.TT.setWT	Forward MAMA PCR primer corresponding to the wild type allele at position 2 in MM.LacZ::KanR	GATTCAGTGGCCGTCGTTTT	Unmodified

Table 2.2 (Continued)

Kan.L3.CC.setWT	Forward MAMA PCR primer corresponding to the wild type allele at position 3 in MM.LacZ::KanR	GCTGGTTACTGGGAAAACCC	Unmodified
Kan.L4.CC.setWT	Forward MAMA PCR primer corresponding to the wild type allele at position 4 in MM.LacZ::KanR	GCGTTACCCAACTTAATCGC C	Unmodified
Kan.L1.rev	Reverse MAMA PCR primer that is complementary to Kan.L1.AA.set1, Kan.L1.AA.set2, and Kan.L1.TT.setWT	ATGCATTTCTTTCCAGACTTG TTCA	Unmodified
Kan.L2.rev	Reverse MAMA PCR primer that is complementary to Kan.L2.AA.set1, Kan.L2.AA.set2, and Kan.L2.TT.setWT	GCATCAACAATATTTTCACC TGAATCA	Unmodified
Kan.L3.rev	Reverse MAMA PCR primer that is complementary to Kan.L3.TT.set1, Kan.L3.TT.set2, and Kan.L3.CC.setWT	CTGTAGCCAGCTTTCATCAA CA	Unmodified
Kan.L4.rev	Reverse MAMA PCR primer that is complementary to Kan.L4.TT.set1, Kan.L4.TT.set2, and Kan.L1.CC.setWT & Sequencing primer for MAMA PCR validation	AGGGGACGACGACAGTATC	Unmodified
Kan.L1.L2.seq	Sequencing primer for MAMA PCR validation	TAGCTCACTCATTAGGCACC	Unmodified

Antibiotic resistance insertion cassettes were generated using long PCR primers containing 45 bp genome homology regions on the 5' end, followed by roughly 20 bp of homology to the antibiotic resistance gene to be amplified. The insertion cassettes were designed such that the resistance gene was inserted 46 base pairs into the coding DNA sequence in the case of *lacZ*, and directly after the start codon in the cases of *malK* and *tolC*. This set of PCRs was performed using Qiagen HotStarTaq Plus Master Mix. Final primer concentrations were 0.4 μ M, and templates were resuspended bacterial colonies bearing the desired resistance gene (the *tn903 aphA1* gene for kanamycin resistance, the *Sh ble* gene for zeocin resistance, and

the *tn21 aadA1* gene for spectinomycin resistance; each cassette contained promoter and terminator sequences flanking the resistance gene). These PCRs were heat activated at 95 °C for 6:00, and then cycled 30 times using a denaturation step of 94 °C for 0:30, an annealing step of 56 °C for 0:30, and an extension step of 72 °C for 2:30. After a final 5:00 extension step at 72 °C, PCRs were held at 4 °C, then purified via 1% agarose gel extraction using the Qiagen gel extraction kit. DNA samples were quantitated using a NanoDrop ND1000 spectrophotometer.

These constructs were used as template for subsequent PCRs to generate dual-biotinylated dsDNA constructs. In each reaction, one primer contained a 5' dual-biotin tag. The other primer was unmodified, or contained four 5' phosphorothioate bonds. Phosphorothioate bonds were used in the experiment comparing leading-targeting and lagging-targeting ssDNA, with the rationale that this would increase recombination frequency by mitigating exonuclease degradation. PCR conditions were as above, but with 1 µM primers, a 1:30 extension step, and 0.1 ng of the relevant insertion construct used as template. PCR products were purified using the Qiagen PCR purification kit.

These dual-biotinylated dsDNA constructs were used to generate ssDNA via a biotin capture protocol. In this method, the dual-biotinylated DNA strand is bound by streptavidin-coated magnetic beads. Next, the dsDNA is chemically melted, allowing the non-biotinylated strand to be collected from the supernatant, while the biotinylated strand is retained by the beads. To this end, Invitrogen DynaBeads MyOne Streptavidin C1 beads were washed twice with 2x Bind and Wash buffer (10 mM Tris, 1 mM EDTA, 2 M NaCl, pH 7.5), then incubated in one initial bead volume of 1x Bind and Wash buffer, with 5 µg of dual-biotinylated dsDNA for every initial 100 µL of beads. This was rotated in a microcentrifuge tube at room temperature for 20 minutes, after which the beads were washed twice with 1x Bind and Wash buffer. Single-

stranded DNA was then released via incubation with one initial bead volume of chilled 0.125 M NaOH. Beads were vortexed for 30 seconds, incubated for 30 seconds, then placed on a magnet so that the supernatant could be collected. This process was repeated, and the two collected NaOH supernatants were pooled and neutralized using a 3 M solution of pH 5.0 sodium acetate. These samples were then cleaned using the Qiagen PCR purification kit; the standard protocol was used, with an additional rinse with Buffer PE. The purity of the resulting ssDNA was confirmed by PAGE. To this end, 10 ng of purified ssDNA was loaded onto a 6% TBE non-denaturing PAGE gel (Invitrogen), post-stained with SYBR Gold (Invitrogen), and examined under UV light.

A similar strategy was employed for creating the internally mismatched *lacZ::kanR* dsDNA cassette, and the dual 3' overhang intermediate suggested by previous mechanisms. For each, two dual-biotinylated dsDNA constructs were generated, each intended to give rise to one of the two strands of the final construct. The dsDNA constructs designed to give rise to the mismatched cassette were generated in the same manner as described above, with PCR primers as denoted in Table 2.2. For the dsDNA constructs designed to give rise to the overhang intermediate, PCRs were performed using Kapa HiFi Master Mix and primer concentrations of 0.5 μ M. Primers were as described in Table 2.2. These PCRs were heat activated at 95 °C for 5:00, and then cycled 30 times using a denaturation step of 98 °C for 0:20, an annealing step of 62°C for 0:15, and an extension step of 72 °C for 0:45. After a final 5:00 extension step at 72 °C, PCRs were held at 4 °C, then purified via 1% agarose gel extraction using the Qiagen gel extraction kit. For both sets of samples, the dual-biotinylated dsDNA constructs were used to produce ssDNA in the same manner as above. The dual-biotin tags were arranged such that complementary strands were purified, allowing them to be annealed together to form the desired

dsDNA constructs. Purified strands were annealed in equimolar amounts (25 nM for the strands used to generate the mismatched cassette, and 70 nM for the strands used to generate the overhang intermediate) in 5 mM Tris, 0.25 M NaCl, pH 8.0. Samples were annealed in a thermocycler by heating to 95 °C and then decreasing the temperature by 1 °C every two minutes, to a final temperature of 25 °C. The resulting annealed dsDNA was purified from a 1% agarose gel using the Qiagen gel extraction kit. Samples were desalted with Microcon Ultracel YM-100 columns (spinning thrice at $500 \times g$ for ~20 min, and bringing to 500 μ L in deionized H₂O before each spin).

Recombination of DNA Constructs

The above DNA constructs were recombined into EcNR2 cells (*Escherichia coli* MG1655 Δ mutS::cat Δ (ybhB-bioAB)::[λ cI857 Δ (cro-ea59)::tetR-bla]) in a similar manner as previously described.²³ Briefly, cells were grown in a rotator drum at 32 °C in LB Lennox media (10 g tryptone, 5 g yeast extract, 5 g sodium chloride per 1 L water, pH 7.4) until they reached an OD₆₀₀ of 0.4 – 0.6. At this time, the expression of the Lambda Red proteins was induced by vigorously shaking the cells in a 42 °C water bath for 15 minutes. Cells were then chilled on ice, washed twice with deionized water, and resuspended in 50 μ L of deionized water containing the desired DNA construct. For the experiment investigating strand bias, 20 ng of DNA was recombined. For the experiment comparing the 3' overhang construct with other recombinogenic species, 2 nM of overhang construct, ssDNA, or dsDNA was used (75 ng of dsDNA or the overhang construct; 37.5 ng of ssDNA). Where the overhang construct was recombined along with annealing oligos, each oligo was included at 200 nM. For all other experiments, 50 ng of DNA was used. DNA was introduced into the cells via electroporation (BioRad Gene Pulser; 0.1

cm cuvette, 1.78 kV, 25 μ F, 200 Ω). After electroporation, cells were recovered in 3 mL LB Lennox media for 3 hours in a rotator drum at 32 °C.

Analysis of Recombinants

Recombinants were identified by plating 50 μ L or 1 mL (concentrated to 50 μ L) of undiluted recovery culture onto selective media (LB Lennox agar plates with 30 μ g/mL kanamycin sulfate, 95 μ g/mL spectinomycin, or 10 μ g/mL Zeocin). The total viable cell count was determined by plating 50 μ L of a 10^{-4} dilution of the recovery culture (in LB Lennox) onto LB Lennox + 20 μ g/mL chloramphenicol plates (EcNR2 is chloramphenicol-resistant). For experiments involving *lacZ* gene disruption, the plates also contained Fisher ChromoMax X-Gal/IPTG solution at the manufacturer's recommended concentration. Recombination frequencies were determined by dividing the number of recombinants by the extrapolated total viable cell count. All experiments assessing recombination frequency were performed in triplicate, and the standard error of the mean was calculated. We tested our hypothesis that lagging strand recombination frequency is higher than leading strand recombination frequency by using a one-tailed *t*-test assuming unequal variances.

The mismatch amplification mutation assay (MAMA) PCR method³⁰ was used to analyze the genotypes of mismatched *lacZ::kanR* dsDNA recombinants. We used 2 bp mismatches in our mismatched *lacZ::kanR* cassette in order to increase the specificity of our MAMA primers and to decrease the chances of spontaneous point mutations confounding our results. We designed four primers for each mismatch locus: a forward primer corresponding to the strand 1 allele, a forward primer corresponding to the strand 2 allele, a forward primer corresponding to the wild type allele, and a universal reverse primer. Primers were designed so that the 2 bp

mismatch loci occurred at the 3' end of the primer, ensuring that amplification would only occur when these two nucleotides matched the recombinant colony's genotype. Primers were designed with a target T_m of 62 °C, and a subsequent gradient PCR (annealing temperature between 62 °C and 68 °C) determined that the optimal annealing temperature for maximum specificity and yield was 64° C. MAMA PCR reactions for loci 1&3 and loci 2&4 were each performed in a single mixture so as to minimize the number of reactions. Each KanR colony was interrogated using 4 MAMA PCR reactions: strand 1 L1&L3, strand 1 L2&L4, strand 2 L1&L3, and strand 2 L2&L4. For convenience, both strand 1 reactions and both strand 2 reactions were pooled prior to agarose gel analysis. PCR template was prepared by growing a monoclonal colony to stationary phase and performing a 1/100 dilution of this culture into PCR-grade water. Our 20 µL MAMA PCR reactions consisted of 10 µL Qiagen multiplex PCR master mix, 5 µL PCR grade water, 4 µL primer mix (1 µM each), and 1 µL template. PCRs were heat activated at 95 °C for 15:00, and then cycled 27 times using a denaturation step of 94 °C for 0:30, an annealing step of 64 °C for 0:30, and an extension step of 72 °C for 1:20. After a final 5:00 extension step at 72 °C, PCRs were held at 4 °C until they were analyzed on a 1.5% agarose gel stained using ethidium bromide. All 48 recombinants from replicate 1 were also screened using wild type MAMA PCR reactions, performed in an analogous manner as above. This experiment verified that all sites that were not detected as mutants were wild type alleles. The accuracy of the MAMA PCR method was also verified by Sanger sequencing the relevant loci in eight recombinant colonies (Genewiz).

References

1. Murphy, K.C. Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *J Bacteriol* **180**, 2063-71 (1998).
2. Murphy, K.C., Campellone, K.G. & Poteete, A.R. PCR-mediated gene replacement in *Escherichia coli*. *Gene* **246**, 321-30 (2000).
3. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* **20**, 123-8 (1998).
4. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5 (2000).
5. Lee, E.C. et al. A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65 (2001).
6. Li, M.Z. & Elledge, S.J. MAGIC, an in vivo genetic method for the rapid construction of recombinant DNA molecules. *Nat Genet* **37**, 311-9 (2005).
7. Wang, Y. & Pfeifer, B.A. 6-deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in *E. coli*. *Metab Eng* **10**, 33-8 (2008).
8. Thomason, L.C., Costantino, N., Shaw, D.V. & Court, D.L. Multicopy plasmid modification with phage lambda Red recombineering. *Plasmid* **58**, 148-58 (2007).
9. Warming, S., Costantino, N., Court, D.L., Jenkins, N.A. & Copeland, N.G. Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33**, e36 (2005).
10. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
11. Posfai, G. et al. Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044-6 (2006).

12. Muyrers, J.P., Zhang, Y., Buchholz, F. & Stewart, A.F. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* **14**, 1971-82 (2000).
13. Sawitzke, J.A. et al. Recombineering: in vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods Enzymol* **421**, 171-99 (2007).
14. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88 (2002).
15. Poteete, A.R. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* **68**, 66-74 (2008).
16. Thaler, D.S., Stahl, M.M. & Stahl, F.W. Double-chain-cut sites are recombination hotspots in the Red pathway of phage lambda. *J Mol Biol* **195**, 75-87 (1987).
17. Yu, D. et al. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 5978-83 (2000).
18. Stahl, M.M. et al. Annealing vs. invasion in phage lambda recombination. *Genetics* **147**, 961-77 (1997).
19. Lim, S.I., Min, B.E. & Jung, G.Y. Lagging strand-biased initiation of red recombination by linear double-stranded DNAs. *J Mol Biol* **384**, 1098-105 (2008).
20. Murphy, K.C. Phage recombinases and their applications. *Adv Virus Res* **83**, 367-414 (2012).
21. Subramanian, K., Rutvisuttinunt, W., Scott, W. & Myers, R.S. The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Res* **31**, 1585-96 (2003).
22. Pound, E., Ashton, J.R., Becerril, H.A. & Woolley, A.T. Polymerase chain reaction based scaffold preparation for the production of thin, branched DNA origami nanostructures of arbitrary sizes. *Nano Lett* **9**, 4302-5 (2009).
23. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).

24. Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. & Church, G.M. Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PLoS One* **7**, e44638 (2012).
25. Dutra, B.E., Suter, V.A., Jr. & Lovett, S.T. RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U S A* **104**, 216-21 (2007).
26. Datta, S., Costantino, N., Zhou, X. & Court, D.L. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* **105**, 1626-31 (2008).
27. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. *Genetics* **186**, 791-9 (2010).
28. Li, X.T. et al. Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli. *Nucleic Acids Res* **31**, 6674-87 (2003).
29. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6 (2001).
30. Qiang, Y.Z. et al. Use of a rapid mismatch PCR method to detect gyrA and parC mutations in ciprofloxacin-resistant clinical isolates of Escherichia coli. *J Antimicrob Chemother* **49**, 549-52 (2002).
31. Richardson, C.C., Lehman, I.R. & Kornberg, A. A Deoxyribonucleic Acid Phosphatase-Exonuclease from Escherichia Coli. II. Characterization of the Exonuclease Activity. *J Biol Chem* **239**, 251-8 (1964).
32. Wang, H.H., Xu, G., Vonner, A.J. & Church, G. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res* **39**, 7336-47 (2011).
33. Yu, D., Sawitzke, J.A., Ellis, H. & Court, D.L. Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci U S A* **100**, 7207-12 (2003).
34. Hill, S.A., Stahl, M.M. & Stahl, F.W. Single-strand DNA intermediates in phage lambda's Red recombination pathway. *Proc Natl Acad Sci U S A* **94**, 2951-6 (1997).

35. Maresca, M. et al. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* **11**, 54 (2010).
36. Zhang, Y., Muirers, J.P., Testa, G. & Stewart, A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat Biotechnol* **18**, 1314-7 (2000).
37. Szczepanska, A.K. Bacteriophage-encoded functions engaged in initiation of homologous recombination events. *Crit Rev Microbiol* **35**, 197-220 (2009).

Chapter Three

Studying and Improving Lambda Red Double-Stranded DNA Recombination via Phosphorothioate Placement and Nuclease Removal

This chapter is adapted from a portion of the following published paper:

Mosberg, J.A. *, Gregg, C.J. *, Lajoie, M.J. *, Wang, H.H. & Church, G.M. Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *PLoS One* 7, e44638 (2012)

*Indicates co-first authorship

Research contributions are as follows:

J. Mosberg designed, performed, and interpreted most of the experiments in this chapter, with contributions from C. Gregg, M. Lajoie, and G. Church. J. Mosberg wrote a majority of this portion of the published paper, with additional writing and editing contributions from C. Gregg, M. Lajoie, and G. Church.

Introduction

In Chapter 2, I propose and present data that support a novel mechanism detailing the process by which Lambda Red recombines dsDNA.¹ According to this mechanism, three phage Lambda-derived proteins are required to mediate efficient dsDNA recombination. The first, Gam, inhibits the endogenous RecBCD and SbcCD nucleases,² which would otherwise degrade the exogenous dsDNA recombineering cassettes. The second protein, Lambda Exonuclease (Exo), degrades one of the two strands in its entirety, leaving behind a full-length strand of ssDNA.^{1,3} The third protein, Beta, then binds to this ssDNA and catalyzes its annealing to the lagging strand of the replication fork, where it is incorporated into the newly synthesized strand as part of an Okazaki fragment.^{1,3} Given this mechanism, recombination predominantly occurs when Lambda Exo degrades the leading-targeting strand of dsDNA, leaving behind the lagging-targeting strand for recombination at the replication fork. The final step of Beta-catalyzed annealing and incorporation as an Okazaki fragment is analogous to the accepted mechanism for Red-mediated oligonucleotide recombination, which requires only the Beta recombinase.⁴

In addition to developing a greater understanding of the recombineering process, this work also pointed to ways in which the frequency of Lambda Red dsDNA recombination could be improved. Because our mechanism states that the lagging-targeting strand is the predominant intermediate in dsDNA recombination, it implies that Exo-mediated degradation of the leading-targeting strand promotes recombination, while Exo-mediated degradation of the lagging-targeting strand hinders recombination. Thus, it is possible that recombination frequencies could be improved by preventing Exo from degrading the lagging-targeting strand, and/or by promoting its digestion of the leading-targeting strand. Given that phosphorothioate (PT) bonds are known to impede degradation by Lambda Exo,⁵ placing these bonds on the 5' end of the

lagging-targeting strand may provide a straightforward means of biasing Exo toward the degradation of the undesired leading-targeting strand.

Beyond biasing Lambda Exo, PT bonds may also serve the function of protecting the recombinogenic lagging-targeting strand from deleterious endogenous exonuclease activity. Multiple lines of evidence suggest that endogenous nucleases can limit recombination. First, it has been shown that protecting oligonucleotides with PT bonds improves recombination frequency,⁶ implying that endogenous exonuclease degradation can render ssDNA non-recombinogenic. Additionally, mutations located near the ends of an oligonucleotide⁷ or dsDNA cassette (as discussed in Chapter 2) have been shown to be inherited less frequently than mutations located closer to the interior of the oligo or cassette, further implying the exonuclease degradation of both oligonucleotides and dsDNA. Thus, protecting the lagging-targeting dsDNA strand with PT bonds could improve both recombination frequency and the preservation of mutations encoded at the ends of cassettes.

Such phosphorothioate-mediated nuclease protection may be useful both before and after Lambda Exo digests the leading-targeting strand to yield the ssDNA intermediate. In addition to Lambda Exo and other dsDNA nucleases, *E. coli* also contains several endogenous nucleases which readily degrade ssDNA.⁸ Indeed, it has recently been shown that knocking out four of these ssDNA exonucleases improves Lambda Red oligo recombination frequency when low concentrations of oligos are used.⁹ Phosphorothioate protection of the 5' end of the lagging-targeting dsDNA strand may similarly diminish the effect of such nucleases on the recombinogenic ssDNA intermediate, thereby improving recombination frequency.

On the other hand, while using PT bonds for nuclease protection may be beneficial, it is unlikely to be a complete solution for preventing the deleterious effect of nuclease activity. For

one, 5' phosphorothioation does nothing to preclude the degradation of 3' ends of ssDNA or dsDNA, and while PT bonds can readily be incorporated onto the 5' ends of a cassette (via phosphorothioated PCR primers), they are much more difficult to install on a 3' end.

Additionally, it is possible that some endogenous nucleases are not effectively inhibited by PT bonds. Thus, it may be possible to further improve dsDNA recombination frequency through the targeted inactivation of endogenous nucleases, as in the aforementioned experiment involving oligonucleotide recombination.

Here, we describe a set of experiments utilizing phosphorothioate placement and nuclease removal in order to study Lambda Red dsDNA recombination and improve the frequency of gene insertion. We show that phosphorothioate protection can significantly improve dsDNA recombination frequency, and that the effects of phosphorothioate placement on recombination frequency follow a pattern consistent with the predictions of our proposed mechanism. These experiments also led to our identification of ExoVII as a key nuclease that degrades the ends of dsDNA cassettes. Removing this nuclease further improves dsDNA recombination frequency, and facilitates the preservation of mutations encoded on the ends of dsDNA cassettes. Thus, the work described in this chapter bolsters our understanding of Lambda Red dsDNA recombination, and also identifies concrete ways in which the process can be improved.

Investigating the Effect of Phosphorothioate Placement on dsDNA Recombination Frequency

In order to determine the effect of PT bonds on dsDNA recombination frequency, we generated a set of seven variably phosphorothioated (VPT) *lacZ::kanR* insertion cassettes (Figure 3.1). This VPT series of cassettes had four consecutive PT bonds located either

terminally (at the 5' end of the cassette), or internally (between the 5' homology and heterology regions). These PT bonds were placed on neither strand, on the lagging-targeting strand only, on the leading-targeting strand only, or on both strands.

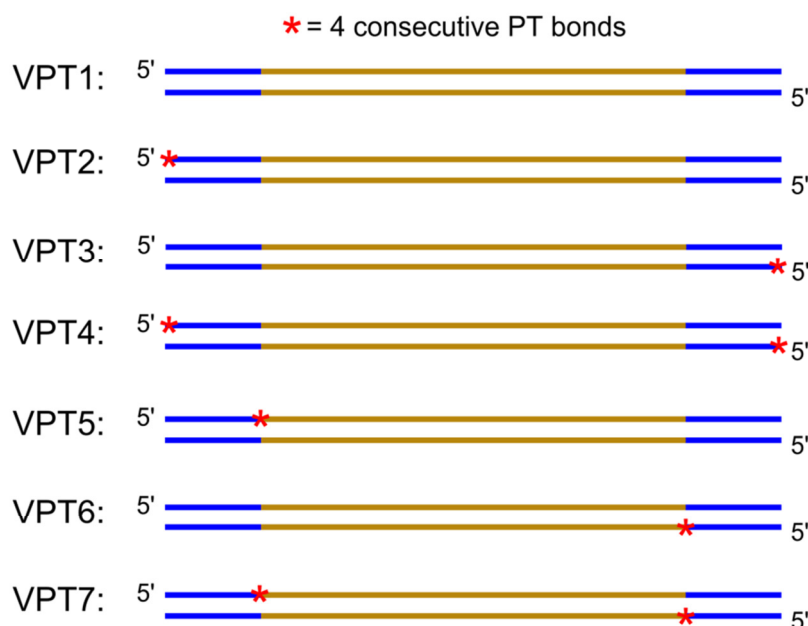


Figure 3.1: Diagram of the Variably Phosphorothioated (VPT) Cassette Series. Homology arms to *lacZ* are shown in blue, and the inserted heterologous *kanR* gene is shown in gold; the lagging-targeting strand is represented as the top strand, and the leading-targeting strand is represented as the bottom strand. The red asterisks indicate the presence of 4 consecutive phosphorothioate bonds.

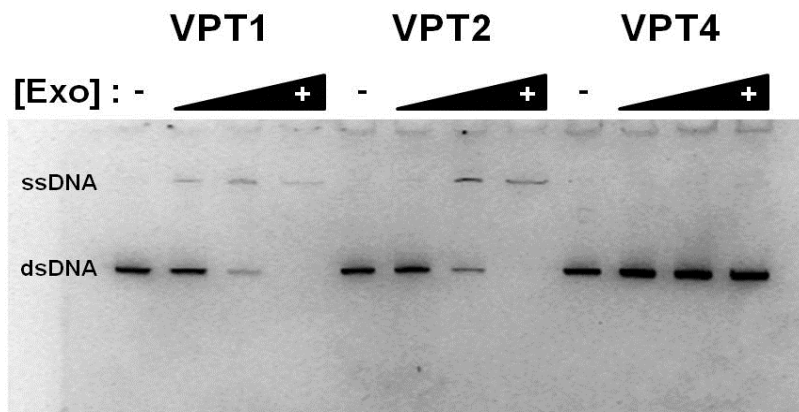


Figure 3.2: *In vitro* Lambda Exo Digestion. Non-phosphorothioated (VPT1, left), single-end phosphorothioated (VPT2, center), and dually phosphorothioated (VPT4, right) dsDNA cassettes were digested with no Lambda Exo (lanes 1, 5, & 9) and tenfold-increasing amounts of Lambda Exo, from left to right. The bottom gel band is dsDNA, and the top band is the ssDNA product after Lambda Exo degrades one of the two strands.

We then used three of these cassettes in an *in vitro* digestion experiment with purified Lambda Exo, in order to confirm that PT bonds block Lambda Exo degradation as previously reported⁵ (Figure 3.2). Indeed, we found that while the non-phosphorothioated dsDNA cassette (VPT1) is readily degraded by Lambda Exo, the cassette with phosphorothioate bonds on both 5' ends (VPT4) exhibits no apparent degradation. The tested cassette with phosphorothioate bonds on only one 5' end (VPT2) shows digestion of one strand, leaving behind the

remaining strand as ssDNA. Thus, it is clear that these phosphorothioated cassettes are highly effective at blocking the action of Lambda Exo in an *in vitro* context.

To determine how phosphorothioate placement affects recombination frequency *in vivo*, we recombined each of the VPT cassettes into the EcNR2 recombineering strain.⁶

Recombination frequencies were calculated by dividing the number of kanamycin-resistant recombinant cfu (colony forming units) by the number of total cfu, and results are shown in Figure 3.3.

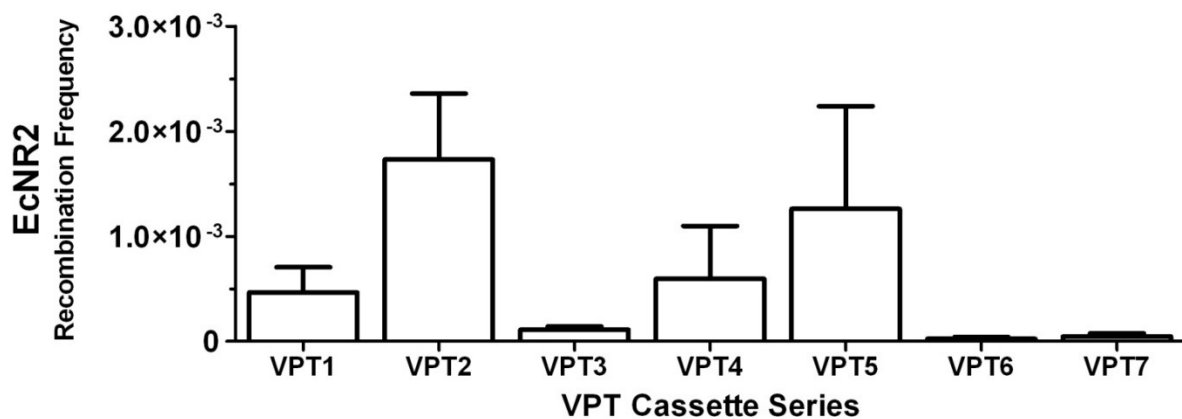


Figure 3.3: Recombination Frequencies of the VPT Cassette Series in EcNR2. Insertion frequencies were measured as the number of kanamycin resistant recombinants over the total number of viable cells (as plated on non-selective media). Three replicates were performed; data are presented as the mean with the error bars representing standard deviation.

Several interesting observations are apparent. First, the cassette with phosphorothioate bonds on the 5' end of the lagging-targeting strand (VPT2) had significantly greater recombination frequency than VPT1, the unmodified *lacZ::kanR* cassette ($p = 0.03$, by unpaired t-test). No other cassette had significantly greater recombination frequency than VPT1, and placing PT bonds on the 5' end of the leading-targeting strand (in VPT3) decreased frequency roughly four-fold. These results provide further support of our proposed mechanism for dsDNA recombination. For one, they argue against the mechanisms proposed by Court and Poteete, as alternating which of the two strands is protected by PT bonds would not be expected to have

differential effects if Exo resection occurred from both 5' ends. These results also confirm our hypothesis that protecting the 5' end of the lagging-targeting strand improves recombination frequency. Conversely, protecting the 5' end of the leading-targeting strand decreases recombination frequency, as it prevents Lambda Exo from degrading that strand to generate the lagging-targeting ssDNA intermediate. Consistent with these results, the cassette with terminal PTs placed on both 5' ends (VPT4) is roughly equivalent (1.23 times more recombinogenic) to unmodified dsDNA, presumably due to a cancellation of these two effects.

Nevertheless, it is surprising that VPT4 can so readily be processed by Lambda Exo and undergo recombination, given the degree to which this cassette resists Exo digestion *in vitro* (Figure 3.2). It is possible that Lambda Exo behaves differently *in vivo* than *in vitro* (e.g., due to the presence of protein partners or cofactors), and is able to degrade phosphorothioate bonds in a cellular context. Alternatively, an endogenous *E. coli* nuclease may degrade the 5' PT bonds on one or both strands, thereby allowing Lambda Exo to degrade the remainder of the strand and generate the ssDNA recombination intermediate.

The recombination frequencies of the VPT cassettes with internal PT bonds support the latter explanation. When placed on the lagging-targeting strand only (VPT5), internal PT bonds do not decrease recombination frequency. However, internal PT bonds on the leading-targeting strand (VPT6 and VPT7) are highly detrimental to recombination frequency. These cassettes have substantially lower recombination frequencies than the corresponding terminally phosphorothioated cassettes (VPT3 and VPT4, respectively), with a VPT4:VPT7 ratio of 12.3 and a VPT3:VPT6 ratio of 4.9. This suggests that internal PT bonds are significantly more effective than terminal PT bonds at blocking Lambda Exo from degrading the leading-targeting strand and generating the ssDNA recombination intermediate. Thus, it is unlikely that Lambda

Exo can independently degrade phosphorothioates *in vivo*; rather, it appears that an endogenous nuclease is able to cleave terminally located PT bonds, thereby allowing Lambda Exo to degrade the remainder of the strand. Internally located PT bonds are likely not directly accessible to this endogenous nuclease, and/or cannot be removed without rendering the cassette non-recombinogenic due to the degradation of one of its homology regions. Thus, blocking the leading-targeting strand with internal PT bonds has a uniquely detrimental effect on recombination frequency.

Utilizing the VPT Cassettes to Investigate Nuclease Processing of dsDNA

The above results suggest a strategy for identifying the nuclease(s) putatively responsible for degrading the phosphorothioated ends of dsDNA cassettes: inactivate candidate nucleases, and then compare the recombination frequencies of unmodified dsDNA (VPT1) and dsDNA with terminal 5' PT bonds on both strands (VPT4). If the postulated nuclease(s) responsible for pruning terminal PT bonds is/are still present, then VPT4 will have a recombination frequency that is roughly equal to that of VPT1 ($VPT4:VPT1 \approx 1.0$, as in EcNR2). However, if the nuclease(s) have been removed, the strain will no longer be able to degrade the phosphorothioate bonds on the 5' ends of the VPT4 dsDNA. Thus, the recombination frequency of VPT4 will drop well below that of VPT1 ($VPT4:VPT1 \ll 1$), as Lambda Exo can no longer generate the ssDNA recombination intermediate.

We first tested VPT1 and VPT4 in a strain lacking Endonuclease I (*endA*), a potent periplasmic endonuclease capable of degrading dsDNA.¹⁰ This resulted in a VPT4:VPT1 ratio of 1.55, similar to that observed for the EcNR2 strain (Table 3.1). Thus, Endonuclease I is not responsible for degrading the phosphorothioated ends of dsDNA cassettes.

Table 3.1: Identifying the Nuclease(s) which Degrade Phosphorothioated Cassette Ends

Strain Background	VPT4:VPT1
EcNR2	1.23
EcNR2. <i>endA</i> ⁻	1.55
EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> ⁻ , <i>xseA</i> ⁻ , <i>exoX</i> ⁻ (<i>nuc4</i> ⁻)	0.09
EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> ⁻ , <i>exoX</i> ⁻ (<i>xseA</i> ⁺)	0.77
EcNR2. <i>xonA</i> ⁻ , <i>xseA</i> ⁻ , <i>exoX</i> ⁻ (<i>recJ</i> ⁺)	0.06
EcNR2. <i>recJ</i> ⁻ , <i>xseA</i> ⁻ , <i>exoX</i> ⁻ (<i>xonA</i> ⁺)	0.09
EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> ⁻ , <i>xseA</i> ⁻ (<i>exoX</i> ⁺)	N/A [§]
EcNR2. <i>xseA</i> ⁻	0.07

[§] EcNR2.*xonA*⁻,*recJ*⁻,*xseA*⁻ did not give KanR recombinants for either cassette; this strain was recreated by restoring *exoX* function in *nuc4*⁻, whereupon it exhibited the same non-recombinogenic phenotype. The biological basis of this phenotype is unclear.

We next tested a strain which lacked the four primary⁸ *E. coli* ssDNA exonucleases – *recJ*/RecJ,¹¹ *xonA*/ExoI,¹² *xseA*/ExoVII,¹³ and *exoX*/ExoX.¹⁴ This strain (EcNR2.*recJ*⁻,*xonA*⁻,*xseA*⁻,*exoX*⁻, or “*nuc4*⁻”) demonstrated a VPT4:VPT1 ratio of 0.09, far lower than that of EcNR2 (Table 3.1). Thus, recombination frequency data for the full VPT series was collected in this strain (Figure 3.4).

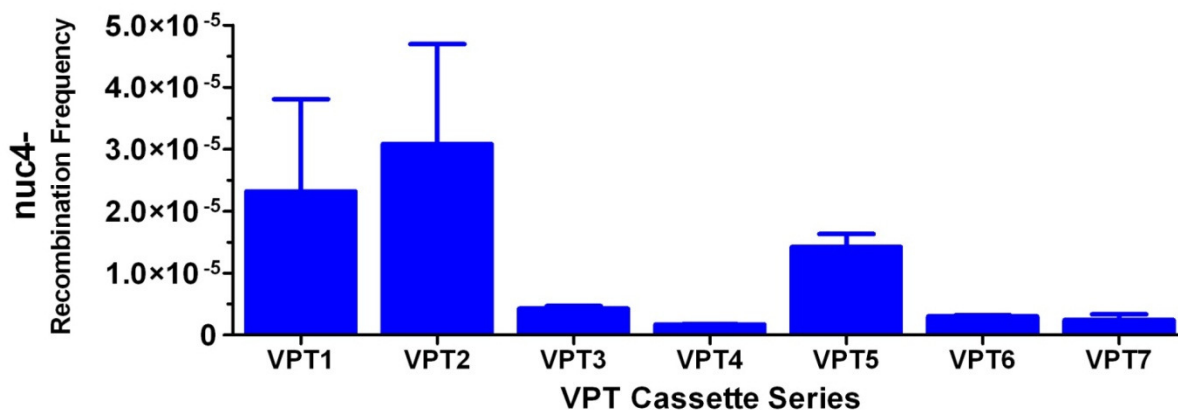


Figure 3.4: Recombination Frequencies of the VPT Cassette Series in *nuc4*⁻. Insertion frequencies were measured as the number of kanamycin resistant recombinants over the total number of viable cells (as plated on non-selective media). Two replicates were performed; data are presented as the mean with the error bars representing standard deviation.

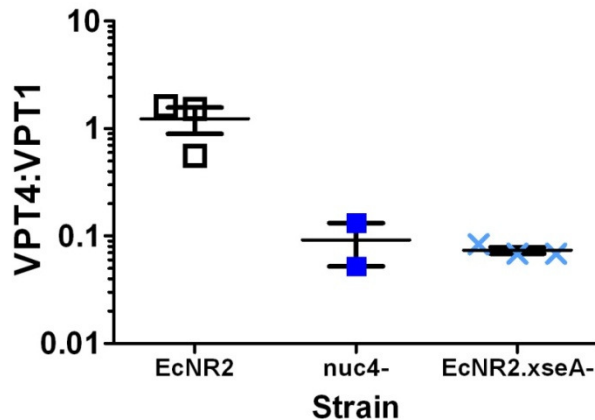


Figure 3.5: VPT4:VPT1 Ratios of Tested Strains. Cassette insertion frequencies were measured in technical replicates for strains EcNR2, *nuc4*⁻, and EcNR2.*xseA*⁻. Data are presented as the mean with the standard error of the mean. The ratio between VPT4 and VPT1 indicates a strain's ability to recombine cassettes with terminal 5' PT bonds preventing the direct action of Lambda Exo.

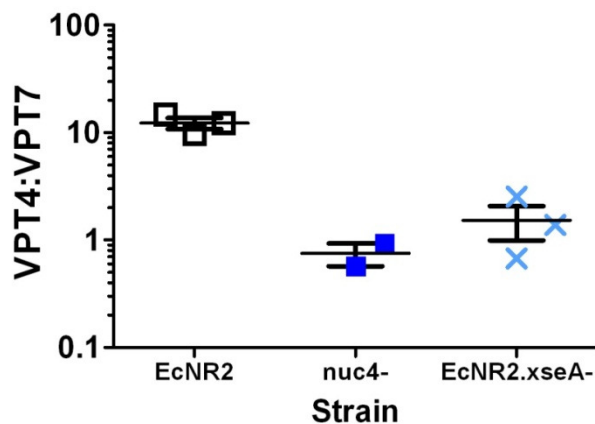


Figure 3.6: VPT4:VPT7 Ratios of Tested Strains. Cassette insertion frequencies were measured in technical replicates for strains EcNR2, *nuc4*⁻, and EcNR2.*xseA*⁻. Data are presented as the mean with the standard error of the mean. The ratio between VPT4 and VPT7 indicates whether a strain is more able to process terminal PT bonds (VPT4) in comparison with internal PT bonds (VPT7).

These data show a strikingly different trend than observed for EcNR2. The recombination frequency of VPT4 in *nuc4*⁻ is indeed far lower than that of VPT1 (Figure 3.5). Additionally, terminal PT bonds on the leading-targeting strand are just as detrimental to recombination frequency as internal PT bonds (VPT3:VPT6 = 1.41; VPT4:VPT7 = 0.75, Figure 3.6). Because constructs with terminal PT bonds on the leading-targeting strand are minimally recombinogenic in *nuc4*⁻, this suggests that these constructs can no longer be processed by Lambda Exo, and therefore that the nuclease(s) capable of cleaving off these PT bonds is/are absent in this strain. Residual recombinants for such cassettes may be due to a limited ability

of Lambda Exo to degrade PT bonds, or to slight endogenous nuclease activity that is still present.

To identify which of the four nucleases is/are responsible for cleaving terminal PT bonds, four strains were generated, each of which restored one of the nucleases removed in *nuc4*⁻ (i.e.,

three of the four nucleases were inactivated in each of these strains). VPT1 and VPT4 were recombined into these strains, and the VPT4:VPT1 ratio of recombinants was measured (Table 3.1). One of the four strains (EcNR2.*recJ*⁻,*xonA*⁻,*xseA*⁻) was surprisingly found to be non-recombinogenic. This was the case regardless of whether the strain was generated by inactivating the three nucleases from an EcNR2 background, or by reactivating ExoX in *nuc4*⁻; this suggests that the non-recombinogenic phenotype did not arise from an off-target mutation. Of the remaining strains, two showed a VPT4:VPT1 ratio similar to that of *nuc4*⁻, while the strain containing *xseA* (EcNR2.*recJ*⁻,*xonA*⁻,*exoX*⁻) showed a ratio similar to that of EcNR2. This suggests that ExoVII (*xseA*) may be the nuclease responsible for the ability of EcNR2 to degrade phosphorothioated ends of dsDNA cassettes. An ExoVII single knockout strain (EcNR2.*xseA*⁻) was therefore generated.

Recombineering Properties of the ExoVII Mutant Strain

The VPT series of cassettes was recombined into EcNR2.*xseA*⁻, and recombination frequencies were determined (Figure 3.7).

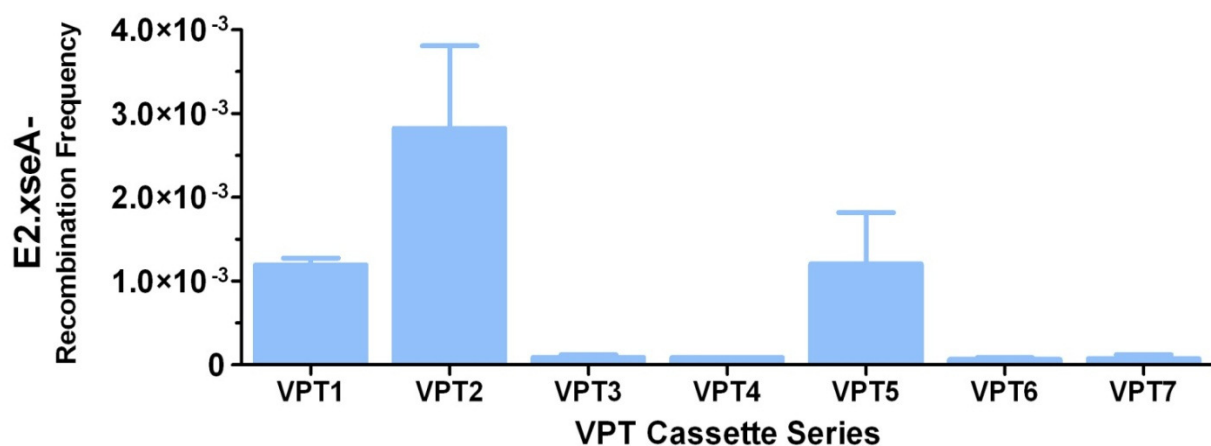


Figure 3.7: Recombination Frequencies of the VPT Cassette Series in EcNR2.*xseA*⁻. Insertion frequencies were measured as the number of kanamycin resistant recombinants over the total number of viable cells (as plated on non-selective media). Three replicates were performed; data are presented as the mean with the error bars representing standard deviation.

This strain shows a pattern of relative frequencies similar to that of *nuc4⁻*: terminal phosphorothioation on the leading-targeting strand is highly detrimental to recombination, and has impact equivalent to that of internal phosphorothioation (Figures 3.5 and 3.6). This suggests that these terminal PT bonds can no longer be cleaved off, and therefore that ExoVII is the primary *E. coli* nuclease responsible for degrading phosphorothioated 5' ends of dsDNA cassettes. Moreover, removing ExoVII significantly increases the recombination frequency of unmodified dsDNA (VPT1) above levels observed in EcNR2 (2.6-fold improvement; $p = 0.0076$ by unpaired t-test). This strongly suggests that the action of ExoVII on the ends of dsDNA cassettes may compromise recombination frequency. Furthermore, the recombination frequency of VPT2 also appears to be slightly improved in EcNR2.*xseA⁻*, although this result was not statistically significant (1.6-fold improvement; $p = 0.18$).

ExoVII is a nuclease encoded by the *xseA* and *xseB* genes.¹⁵ It is processive, and can degrade from either the 5' or 3' end of a DNA strand.¹⁶ While it is highly specific for ssDNA substrates, it is capable of degrading short overhangs and then continuing into duplex regions of DNA.¹⁶ Its observed ability to degrade dsDNA ends is likely due to strand “breathing,” possibly aided by endogenous helicase enzymes, or by PT bonds decreasing the strength of the annealing interaction between the two strands.¹⁷ Alternatively, an endogenous 3'-to-5' exonuclease may degrade part of the strand complementary to the 5' PT bonds, leaving behind an ssDNA overhang to which ExoVII can bind. While ExoVII is classified as an exonuclease due to its requirement for a free ssDNA end, it technically has endonucleolytic activity, given that its degradation products are 4-12 bp oligonucleotides.¹⁶ Thus, ExoVII likely cleaves to the 3' end of the four terminal PT bonds used in the VPT cassette series, rather than degrading the PT bonds directly. This explains the ability of ExoVII to readily process terminally phosphorothioated cassettes.

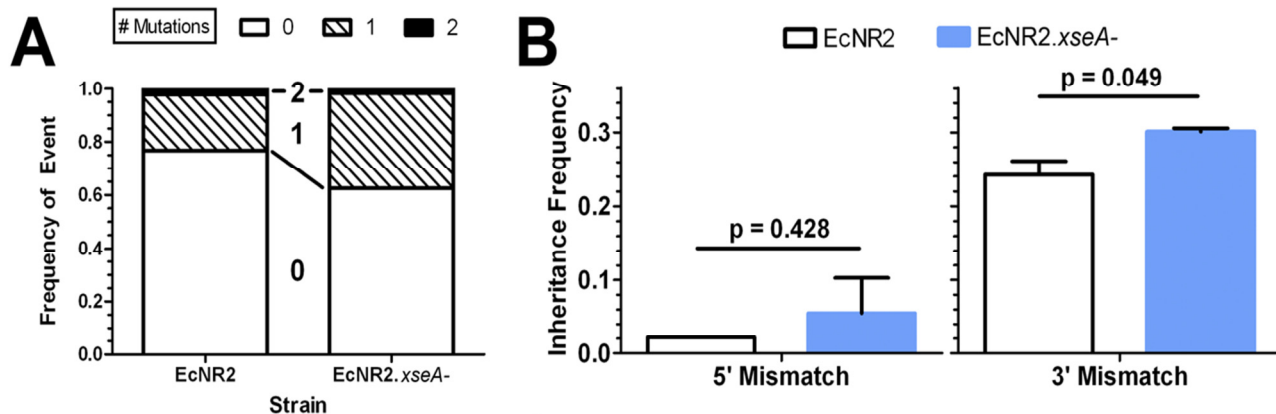


Figure 3.8: Removal of ExoVII Improves dsDNA Mutation Inheritance. **A)** A non-phosphorothioated *lacZ::kanR* dsDNA cassette encoding two 2-bp sets of mismatch mutations (placed 7&8 bp from each end of the cassette) was recombined into EcNR2 and EcNR2.xseA⁻ cells. Resulting kanamycin-resistant recombinants were then genotyped in order to determine whether they inherited these distally-located mutations. The distribution of mutations in EcNR2 (two independent experiments, n = 180) and EcNR2.xseA⁻ (two independent experiments, n = 177) was plotted as the frequency of clones inheriting 0 (empty), 1 (hatched black bars), or 2 (filled black bars) sets of mutations. **B)** When the inheritance of each 2-bp mutation is considered individually, both show increased preservation in EcNR2.xseA⁻, although this is only statistically significant for the 2-bp mutation on the 3' end of the cassette (as defined with respect to the lagging-targeting strand). Data are presented as the mean with the standard deviation from the mean.

Given that the action of ExoVII on the ends of dsDNA cassettes appears to compromise recombination frequency, we investigated whether its removal would increase the inheritance of mutations encoded near the ends of a dsDNA insertion cassette. Thus, we designed a non-phosphorothioated *lacZ::kanR* cassette with 2-bp mutations encoded in each homology region, 7 and 8 bp from the end of the cassette. This construct was then recombined into EcNR2 and EcNR2.xseA⁻. KanR colonies were selected, and their genotypes were analyzed at the mutation loci. The strain with ExoVII removed showed greater inheritance of these end-located mutations (Figure 3.8A), although the difference was only statistically significant for the mutation encoded at the 3' end of the cassette (as defined with respect to the lagging-targeting strand; Figure 3.8B, right panel). This confirms that ExoVII also degrades non-phosphorothioated dsDNA cassettes, and that it is partially responsible for the poor inheritance of mutations located at the ends of such cassettes. However, given that the EcNR2.xseA⁻ strain still demonstrates low levels of

inheritance of these mutations, it is likely that other endogenous nucleases are also responsible for degrading the ends of dsDNA cassettes.

Discussion

The results described in this chapter provide additional insight into the cellular processes occurring during recombineering. First, our surprising observation that dual 5'-phosphorothioated dsDNA is still highly recombinogenic – despite the fact that PT bonds block Lambda Exo – was explained by our discovery that ExoVII degrades the ends of dsDNA cassettes. Thus, dually-phosphorothioated dsDNA enters the cell, after which ExoVII degrades the 5' PT bonds of one or both strands. This step may also require the action of a helicase or another endogenous nuclease in order to generate ssDNA ends to which ExoVII can bind. After the action of ExoVII, Lambda Exo degrades the rest of the leading-targeting strand, leaving behind the lagging-targeting ssDNA recombination intermediate.^{1,3} Therefore, dually-phosphorothioated dsDNA recombines with a frequency equal to or exceeding that of unmodified dsDNA, despite requiring the action of one or more endogenous nucleases in addition to Lambda Exo. This suggests that there is a great deal of interaction between endogenous nucleases and recombineering cassettes. This notion is further supported by the variability of the dsDNA recombination frequencies exhibited by the tested nuclease knockout strains. For example, the *nuc4⁻* strain exhibits sharply decreased recombination frequency for all tested VPT cassettes. The origin of this phenotype is uncertain; it is possible that one of the removed exonucleases has a role in dsDNA recombination, or that the presence of exogenous linear dsDNA is somewhat toxic to cells with these four nucleases removed, thereby selecting against cells which take up the cassettes necessary for recombination. Along similar lines, strain

EcNR2.*recJ*,*xonA*,*xseA* appears to be non-recombinogenic, for reasons that are unclear, while EcNR2.*xseA* has slightly enhanced recombination frequency for cassettes without PT bonds blocking the leading-targeting strand. Taken together, this suggests that modifying endogenous nuclease activity may be a powerful lever for affecting Lambda Red recombination, and possibly a valuable strategy for engineering further improvements of dsDNA recombination frequency.

This work also further validates our mechanistic hypothesis for Lambda Red dsDNA recombineering, described in Chapter 2. As noted above, we found that EcNR2 recombination frequency could be improved by using 5' PT bonds to prevent Exo from degrading the lagging-targeting strand. In contrast, 5' PT bonds on the leading-targeting strand were detrimental, as they block Exo from processing that strand. This observation is predicted by our mechanism, but not by previously proposed mechanisms; those mechanisms posit resection occurring from both 5' ends, and therefore suggest that phosphorothioate bonds on the two strands should have equivalent effects. The behavior of the VPT7 cassette (Figure 3.1) in EcNR2 also helps differentiate our mechanism from the previously proposed mechanisms. According to our mechanism, VPT7 would be expected to have a recombination frequency significantly less than that of unmodified dsDNA, as placing PT bonds between the 5' homology and heterology regions of both strands would prevent Lambda Exo from degrading one strand to generate a full-length ssDNA intermediate. In contrast, the placement of PT bonds in VPT7 would be expected to facilitate Lambda Exo's generation of the 3' overhang intermediate suggested by the previously proposed mechanisms.^{4,18} Thus, if one of those mechanisms were correct, VPT7 would be expected to have equal or greater recombination frequency than unmodified dsDNA. As shown in Figure 3.3, VPT7 has 10.2-fold *lower* recombination frequency than unmodified

dsDNA (VPT1) in EcNR2. This result is consistent with those from similar experiments,^{3,19} and supports our mechanism while further refuting the previously proposed mechanisms.

Beyond providing additional mechanistic insight into Lambda Red recombination, this work also achieved several meaningful improvements of dsDNA recombineering. First, we established that recombination frequency can be improved by protecting the 5' end of the lagging-targeting strand of a dsDNA cassette. This is easily accomplished by using a phosphorothioated PCR primer, and represents a straightforward way to reliably improve gene insertion frequencies. Additionally, we found that removing ExoVII from *E. coli* improves gene insertion frequencies for unmodified cassettes, and likely also for cassettes with phosphorothioated lagging-targeting strands. Thus, this work establishes two simple and effective strategies for improving Lambda Red dsDNA insertion frequencies. Finally, we also showed that removing ExoVII improves the inheritance of mutations encoded near the 3' end of a dsDNA cassette. In sum, this work may enable new and more powerful applications of Lambda Red dsDNA recombineering technology, and also point to additional ways in which recombination frequency can be improved through nuclease modification and phosphorothioate placement.

Experimental

Oligonucleotides used in this Study

A full list of primers and recombineering oligonucleotides used in this work is presented in Table 3.2. Asterisks represent phosphorothioate bonds between the two indicated nucleotides. For these primers, “wt-f” refers to a forward allele-specific colony PCR (ascPCR) or multiplex allele-specific colony PCR (mascPCR) primer used to detect a wild type allele, “mut-f” refers to

a forward ascPCR/mascPCR primer used to detect a mutated allele, and “rev” refers to the reverse ascPCR/mascPCR primer used with both forward primers. All oligonucleotides were ordered from Integrated DNA Technologies with standard purification and desalting.

Table 3.2: Oligonucleotides used in Chapter 3

Name	Use	Sequence
endA.KO*	endA inactivating oligo	T*C*G*T*TTTAACACGGAGTAAGTGATGTACCGTTATTT GTCTATTGCTGCTGAGTGGTACTGAGCGCAGCATTTTCC GGCCCGGCGTTGGCC
exoX.KO*	exoX inactivating oligo	T*T*C*G*GCCTGGAGCATGCCATGTTGCGCATTATCGAT ACAGAAACTGATGCGGTTTGCAGGGAGGGATCGTTGAG ATTGCCTCTGTTGATG
xseA.KO*	xseA inactivating oligo	G*A*A*T*TTGATCTCGCTCACATGTTACCTTCTCAATCCC CTGCAATTGATTTACCGTTAGTCGCCTGAATCAAACGGT TCGCTGCTGCTTG
recJ.KO*	recJ inactivating oligo	G*G*A*G*GCAATTCAGCGGGCAAGTCTGCCGTTTCATCG ACTTCACGTCACGACGAAGTTGTATCTGTTGTTTCACGC GAATTATTTACCGCT
xonA.KO*	xonA inactivating oligo	A*A*T*A*ACGGATTTAACCTAATGATGAATGACGGTAAG CAACAATCTGAACCTTTTTGTTTCACGATTACGAAACCTT TGGCACGCACCCCG
endA.KO*-wt-f	endA wt-f mascPCR primer	CCGTTATTTGTCTATTGCTGCGG
exoX.KO*-wt-f	exoX wt-f mascPCR primer	GCGCATTATCGATACAGAAACCT
xseA.KO*-wt-f	xseA wt-f mascPCR primer	CTTCTCAATCCCCTGCAATTTTACC
recJ.KO*-wt-f	recJ wt-f mascPCR primer	CAACAGATACAACTTCGTCGCC
xonA.KO*-wt-f	xonA wt-f mascPCR primer	GAATGACGGTAAGCAACAATCTACC
endA.KO*-mut-f	endA mut-f mascPCR primer	CCGTTATTTGTCTATTGCTGCTGA
exoX.KO*-mut-f	exoX mut-f mascPCR primer	GCGCATTATCGATACAGAAACTGA
xseA.KO*-mut-f	xseA mut-f mascPCR primer	CTTCTCAATCCCCTGCAATTGA
recJ.KO*-mut-f	recJ mut-f mascPCR primer	CAACAGATACAACTTCGTCGTGA
xonA.KO*-mut-f	xonA mut-f mascPCR primer	GAATGACGGTAAGCAACAATCTGA

Table 3.2 (Continued)

endA.KO*-r	endA rev mascPCR primer	GCACGATTGCAGATCAACAACG
exoX.KO*-r	exoX rev mascPCR primer	GACCATGGCTTCGGTGATG
xseA.KO*-r	xseA rev mascPCR primer	GGTACGCTTAAGTTGATTTTCCAGC
recJ.KO*-r	recJ rev mascPCR primer	GGCCTGATCGACCACTTCC
xonA.KO*-r	xonA rev mascPCR primer	GAAATGTCTCCTGCCAAATCCAC
L:K.mut7-8.f	Forward primer for generating lacZ::kanR with distal mutations	TGACCATCCTTACGGATTCACTG
L:K.mut7-8.r	Reverse primer for generating lacZ::kanR with distal mutations	GTGCTGCTTGGCGATTAAAG
L:K.mut7-8.L1-mut	mut-f mascPCR primer for lacZ::kanR 5' distal mutation	AGGAAACAGCTATGACCATCC
L:K.mut7-8.L1-wt	wt-f mascPCR primer for lacZ::kanR 5' distal mutation	CAGGAAACAGCTATGACCATGA
L:K.mut7-8.L4-mut	mut-f mascPCR primer for lacZ::kanR 3' distal mutation	CGTTACCCAACTTAATCGCCAA
L:K.mut7-8.L4-wt	wt-f mascPCR primer for lacZ::kanR 3' distal mutation	CGTTACCCAACTTAATCGCCTT
Kan.L1.rev	rev mascPCR primer for lacZ::kanR 5' distal mutation	ATGCATTTCTTTCCAGACTTGTTCA
Kan.L4.rev	rev mascPCR primer for lacZ::kanR 3' distal mutation	AGGGGACGACGACAGTATC
Kan:LacZ.NoPT-f	Forward primer for generating VPT1, VPT3, VPT6	TGACCATGATTACGGATTCACTGGCCGTCGTTTTACAA
Kan:LacZ.NoPT-r	Reverse primer for generating VPT1, VPT2, VPT5	GTGCTGCAAGGCGATTAAAGTTGGGTAACGCCAGGGTTTT CCCAGT
Kan:LacZ.BlockPT-f	Forward primer for generating VPT5, VPT7	TGACCATGATTACGGATTCACTGGCCGTCGTTTTACAAC GT*C*G*T*G
Kan:LacZ.BlockPT-r	Reverse primer for generating VPT6, VPT7	GTGCTGCAAGGCGATTAAAGTTGGGTAACGCCAGGGTTTT CC*C*A*G*T

Table 3.2 (Continued)

Kan:LacZ.StartPT-f	Forward primer for generating VPT2, VPT4	T*G*A*C*CATGATTACGGATTCACTGGCCGTCGTTTTAC AA
Kan:LacZ.StartPT-r	Reverse primer for generating VPT3, VPT4	G*T*G*C*TGCAAGGCGATTAAGTTGGGTAACGCCAGGGT TTTCCCAGT

Generating dsDNA Recombineering Cassettes

Both the VPT series of dsDNA recombineering cassettes and the *lacZ::kanR* cassette with mutations encoded in its homology arms were generated by PCR using the primers denoted in Table 3.2. PCRs were performed using Kapa HiFi HotStart ReadyMix, with primer concentrations of 0.5 μ M, and 0.1 ng of *lacZ::kanR* dsDNA (as generated in previous work)¹ used as template. PCRs had a total volume of 50 μ L, and were heat activated at 95 °C for 5 min, then cycled 30 times with a denaturation temperature of 98 °C (20 sec), an annealing temperature of 62 °C (15 sec), and an extension temperature of 72 °C (45 sec). PCRs were brought to 72 °C for 5 min, and then held at 4 °C. PCR products were cleaned with the Qiagen PCR purification kit (elution in 50 μ L H₂O). Samples were desalted with Microcon Ultracel YM-100 columns (spinning twice at 500 $\times$ g for 20 min, and bringing to 500 μ L in H₂O before each spin). Resulting samples were then brought to 100 μ L in H₂O and quantitated on a NanoDrop ND1000 spectrophotometer. All samples were analyzed on a 1% agarose/ethidium bromide (EtBr) gel to confirm that the expected band was present and pure.

In vitro Digestion by Lambda Exo

lacZ::kanR dsDNA (20 ng) with neither, one, or both ends phosphorothioated (VPT1, VPT2, and VPT4, respectively; see Figure 3.1) was added to Lambda Exonuclease Buffer (New England Biolabs), such that the resulting volume was 9 μ L and the resulting buffer concentration

was 1X. Lambda Exonuclease (New England Biolabs) was serially diluted in 1X Lambda Exonuclease Buffer, and 1 μ L of the appropriate dilution was then added to the reaction. Reactions were incubated at 37 °C for 30 min, heat inactivated at 75 °C for 10 min, and then analyzed on an Invitrogen 6% TBE non-denaturing PAGE gel (180 V for 40 min, then post-stained in Invitrogen SYBR Gold for 15 min). This gel was then examined and imaged under UV light.

Strain Creation

EcNR2 (*Escherichia coli* MG1655 Δ mutS::cat Δ (ybhB-bioAB)::[λ cl857 N(*cro-ea59*)::tetR-bla])⁶ was used as the basis for all strains created in this work. From an EcNR2 background, nuclease genes *endA*, *xonA*, *recJ*, *xseA*, and *exoX* were inactivated singly or in combination, using oligo-mediated Lambda Red recombination as described below. Knockout oligos (Table 3.2) were designed to introduce a premature stop codon and a frameshift mutation at the beginning of the nuclease gene, thereby rendering the targeted nuclease inactive. Strain genotypes were verified using ascPCR as described below.

Performing Lambda Red Recombination

Lambda Red recombination was performed as previously described.^{1,6} In brief, cultures were grown in LB Lennox media (10 g tryptone, 5 g yeast extract, 5 g NaCl per 1 L water, pH 7.4), from a 1:100 dilution of an overnight culture. Cultures were placed in a rotator drum at 30 °C until they reached an OD₆₀₀ of 0.4-0.6 (typically 2.25-2.5 hrs). Lambda Red expression was then induced by shaking cultures at 300 rpm in a 42 °C water bath (15 min). Induced cultures were immediately cooled on ice, and 1 mL of cells were washed twice in ice cold deionized

water (dH₂O). The resulting cell pellet was resuspended in 50 µL of dH₂O containing the intended recombineering construct(s). For experiments in which dsDNA cassettes were recombined, 100 ng was used; for experiments in which oligonucleotide(s) were recombined, 1 µM of each oligo was used. Samples were electroporated (BioRad GenePulser; 0.1 cm cuvette, 1.78 kV, 200 Ω, 25 µF), and allowed to recover in 3 mL LB Lennox in a rotator drum at 30 °C. Cultures were recovered for at least 3 hours, and then were plated and analyzed as below.

Analyzing Recombination

In order to analyze the recombination of *lacZ::kanR* insertion cassettes, recovery cultures were plated onto LB Lennox agar plates with kanamycin sulfate (30 µg/mL). Plates were incubated at 30 °C overnight, and the number of resulting colonies was counted. To determine the total number of cells present, recovery cultures were diluted (in LB Lennox) and similarly plated onto LB Lennox plates with carbenicillin (50 µg/mL; EcNR2 and its derived strains are carbenicillin-resistant). These results were used to determine the recombination frequency of a given sample (# recombinants / # total viable cells), or the ratio of the number of kanR recombinants yielded by two different cassettes in a given strain. All platings were performed in duplicate. Recovery cultures from recombinations with nuclease knockout oligos were plated on non-selective LB Lennox media, and several clones were isolated and analyzed as described below.

Allele-specific colony PCR (ascPCR) or multiplex allele-specific colony PCR (mascPCR)^{20,21} was used to detect the 1-2 bp mutations generated in the inactivation of endogenous nucleases, and in the recombination of the *lacZ::kanR* cassette with mutations encoded in its homology arms. In these experiments, two PCRs are performed for each tested

clone – one with a forward primer designed to hybridize to the wild type sequence at a locus targeted by a recombineering cassette, and one with a forward primer designed to hybridize to the mutated sequence conferred by the recombineering cassette. The same reverse primer is used in both reactions. If the mutant detecting PCR gives an amplification band but the wild type detecting PCR does not, the clone is scored as a recombinant. In mascPCR, primer sets for interrogating several wild type or several mutant loci are combined in a single reaction, and each amplicon is designed to have a different size ranging from 100 bp to 850 bp.

Singleplex ascPCR reactions (used to determine whether a given nuclease had been inactivated) were performed with Kapa 2GFast HotStart ReadyMix including 10X Kapa dye. PCRs had a total volume of 20 μ L, with 0.5 μ M of each primer, and a template of 1 μ L of stationary phase culture derived from a given clone. These PCRs were carried out with an initial activation step at 95 °C for 2 min, then cycled 30 times with a denaturation temperature of 95 °C (15 sec), an annealing temperature of 63-67 °C (15 sec; temperature as optimized for a given pair of ascPCR reactions), an extension temperature of 72 °C (40 sec), and a final extension at 72 °C for 90 sec.

Multiplex allele-specific colony PCR (mascPCR) was used to detect the two mutations encoded in the homology arms of the *lacZ::kanR* cassette, and to simultaneously genotype several nuclease knockout mutations. To this end, monoclonal colonies were grown to stationary phase under proper antibiotic selection, and PCR template was prepared by diluting 2 μ L of culture into 100 μ L of H₂O. These mascPCRs used Kapa 2GFast Multiplex PCR ReadyMix with 10X Kapa dye, and had a total volume of 10 μ L, with 0.2 μ M of each primer and 2 μ L of template. PCRs were carried out with an initial activation step at 95 °C for 3 min, then cycled 26 times with a denaturation temperature of 95 °C (15 sec), an annealing temperature of 63-67 °C

(30 sec; temperature as optimized for a given pair of mascPCR reactions), an extension temperature of 72 °C (1 min), and a final extension at 72 °C for 5 min. All ascPCRs and mascPCRs were analyzed on 1.5% agarose/EtBr gels (180 V, 60 min).

We repeated recombination experiments with the mutation-encoding *lacZ::kanR* cassette twice for EcNR2 and EcNR2.*xseA*⁻. In each experiment, 96 individual colonies were genotyped for both strains. Only monoclonal and unambiguous mascPCR results were counted towards the final analysis presented here. 5' and 3' mutation inheritance data points were separately combined and analyzed for statistically significant differences between strains using the Mann-Whitney U-test with significance defined as $p < 0.05$.

References

1. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda Red Recombineering in *Escherichia coli* Occurs Through a Fully Single-Stranded Intermediate. *Genetics* **186**, 791-9 (2010).
2. Kulkarni, S.K. & Stahl, F.W. Interaction between the *sbcC* gene of *Escherichia coli* and the *gam* gene of phage lambda. *Genetics* **123**, 249-53 (1989).
3. Maresca, M. et al. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* **11**, 54 (2010).
4. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88 (2002).
5. Liu, X.P. & Liu, J.H. The terminal 5' phosphate and proximate phosphorothioate promote ligation-independent cloning. *Protein Sci* **19**, 967-73 (2010).
6. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).
7. Wang, H.H., Xu, G., Vonner, A.J. & Church, G. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res* **39**, 7336-47 (2011).
8. Dutra, B.E., Sutera, V.A., Jr. & Lovett, S.T. RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U S A* **104**, 216-21 (2007).
9. Sawitzke, J.A. et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* **407**, 45-59 (2011).
10. Jekel, M. & Wackernagel, W. The periplasmic endonuclease I of *Escherichia coli* has amino-acid sequence homology to the extracellular DNases of *Vibrio cholerae* and *Aeromonas hydrophila*. *Gene* **154**, 55-9 (1995).
11. Lovett, S.T. & Kolodner, R.D. Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*. *Proc Natl Acad Sci U S A* **86**, 2627-31 (1989).

12. Prasher, D.C., Conarro, L. & Kushner, S.R. Amplification and purification of exonuclease I from *Escherichia coli* K12. *J Biol Chem* **258**, 6340-3 (1983).
13. Chase, J.W. & Richardson, C.C. Exonuclease VII of *Escherichia coli*. Purification and properties. *J Biol Chem* **249**, 4545-52 (1974).
14. Viswanathan, M. & Lovett, S.T. Exonuclease X of *Escherichia coli*. A novel 3'-5' DNase and Dnaq superfamily member involved in DNA repair. *J Biol Chem* **274**, 30094-100 (1999).
15. Vales, L.D., Rabin, B.A. & Chase, J.W. Isolation and preliminary characterization of *Escherichia coli* mutants deficient in exonuclease VII. *J Bacteriol* **155**, 1116-22 (1983).
16. Chase, J.W. & Richardson, C.C. Exonuclease VII of *Escherichia coli*. Mechanism of action. *J Biol Chem* **249**, 4553-61 (1974).
17. Stein, C.A., Subasinghe, C., Shinozuka, K. & Cohen, J.S. Physicochemical properties of phosphorothioate oligodeoxynucleotides. *Nucleic Acids Res* **16**, 3209-21 (1988).
18. Poteete, A.R. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* **68**, 66-74 (2008).
19. Muyrers, J.P., Zhang, Y., Buchholz, F. & Stewart, A.F. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* **14**, 1971-82 (2000).
20. Isaacs, F.J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-53 (2011).
21. Wang, H.H. & Church, G.M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol* **498**, 409-26 (2011).

Chapter Four

Studying and Improving Lambda Red Oligonucleotide Recombination via Phosphorothioate Placement and Nuclease Removal

This chapter is adapted from a portion of the following published paper:

Mosberg, J.A. *, Gregg, C.J. *, Lajoie, M.J. *, Wang, H.H. & Church, G.M. Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *PLoS One* 7, e44638 (2012)

*Indicates co-first authorship

Research contributions are as follows:

J. Mosberg and M. Lajoie jointly came up with the idea for using nuclease removal to improve MAGE performance. J. Mosberg, C. Gregg, M. Lajoie, and G. Church designed the experiments and interpreted their results. J. Mosberg, C. Gregg, and M. Lajoie performed the experiments. J. Mosberg wrote a majority of this portion of the published paper, with additional writing and editing contributions from C. Gregg, M. Lajoie, and G. Church.

Introduction

As discussed in Chapter 1, the recent discovery¹ that Lambda Red can simultaneously recombine multiple oligonucleotides has enabled several exciting new recombineering applications. This strategy, called “Multiplex Automatable Genome Engineering” (MAGE),¹ has been used to diversify and rapidly optimize the pathway coding for the biosynthesis of the small molecule lycopene,¹ to engineer promoters,² to simultaneously append hexa-histidine tag sequences onto a panel of genes,³ and to change all *E. coli* amber (TAG) stop codons into ochre (TAA) stop codons.⁴ However, despite the considerable power of this method, only a limited number of simultaneous mutations can reliably be generated in a given cycle of MAGE.⁴ This has recently been improved by the development of co-selection MAGE (CoS-MAGE),⁵ in which an oligo is directed to repair a defective selectable marker in the vicinity of the other targeted loci. Subsequent selection for the repaired marker significantly increases recombination frequencies. While CoS-MAGE represents the state of the art prior to the work described in this chapter and Chapter 5, it nevertheless yields only about 1 oligonucleotide recombined per average cell in a given cycle.⁶ This limitation constrains the degree of diversity that can be generated, as well as the extent to which a genome can feasibly be reengineered. By improving allele conversion frequencies in CoS-MAGE, we hope to develop this methodology into an even more powerful tool for ambitious diversification and genome engineering projects.

In Chapter 3, we showed that endogenous nucleases limit the frequency of dsDNA recombination. By protecting cassettes with phosphorothioate (PT) bonds and/or removing the nuclease ExoVII, we were able to improve dsDNA recombination frequency. Several lines of evidence suggest that these approaches could be similarly useful for improving the performance of CoS-MAGE. For one, it has been shown that the singleplex recombination frequency of

oligos can also be improved by protection with PT bonds,¹ and that mutations encoded near the ends of an oligo are inherited less frequently than mutations encoded closer to the center.⁷ Both of these results strongly imply exonuclease degradation of oligonucleotides. Secondly, ExoVII – the exonuclease which we identified to degrade the ends of dsDNA recombineering cassettes – is reportedly highly specific for ssDNA substrates.⁸ Thus, it may degrade recombineering oligos even more readily than dsDNA, and removing ExoVII may therefore substantially improve oligo recombination frequencies. Furthermore, it has been shown that while oligo-mediated recombination can occur in *E. coli* cells lacking Lambda Red, this recombination is severely limited by endogenous nucleases.⁹ Removing four potent ssDNA exonucleases improved the efficiency of this process by nearly 1000-fold, demonstrating that exonuclease degradation strongly hinders oligonucleotide recombination in this context.⁹ It has more recently been shown that this also holds true for Lambda Red singleplex oligo recombination, although only when low concentrations of oligos are used.¹⁰ However, because MAGE involves the simultaneous introduction of several different oligos, we hypothesize that the resulting intracellular concentration of any given oligo is low. Therefore, removing endogenous exonucleases may have a beneficial effect, even when normal oligo concentrations are used.

In this chapter, we show that removing ExoVII improves the inheritance of mutations encoded on the 3' ends of oligonucleotides, and slightly increases CoS-MAGE frequencies. We then extend this approach to show that removing a set of five exonucleases (RecJ, ExoI, ExoVII, ExoX, and Lambda Exo) further improves the performance of CoS-MAGE. In a given round of CoS-MAGE with ten ssDNA oligonucleotides, this “nuc5” strain yields on average 46% more alleles converted per clone, 200% more clones with five or more allele conversions, and 35% fewer clones without any allele conversions. Finally, we use these nuclease knockout strains to

investigate and clarify the effects of oligonucleotide phosphorothioation on recombination frequency. We show that PT bonds can be detrimental as well as beneficial, and that the net effect depends on the nuclease background of the recombineering strain. The results described in this chapter provide further mechanistic insight into Lambda Red oligonucleotide recombination, and achieve substantial improvement of recombineering performance.

Removal of ExoVII Improves Oligonucleotide Mutation Inheritance

In Chapter 3, we found that the removal of ExoVII improves the inheritance of mutations located on the ends of a dsDNA cassette. Given that ExoVII degrades ssDNA preferentially to dsDNA,⁸ we sought to assess whether ExoVII removal also improves the inheritance of mutations located on the ends of an oligonucleotide. Thus, we tested a 90mer oligo previously designed to disrupt the *lacZ* gene,⁷ with seven premature stop codons distributed along its sequence. This oligo was protected with four PT bonds on each end. This “*lacZ.7.stop*” oligo was recombined into three strains – EcNR2,¹ EcNR2.*xseA*⁻,⁶ and *nuc5*⁻ (EcNR2.*recJ*⁻*xonA*⁻*xseA*⁻*exoX*⁻*redA*⁻), a strain with Lambda Exo inactivated in addition to all four of the potent ssDNA exonucleases removed in the aforementioned prior work.^{9,10} *LacZ*⁻ colonies were identified by IPTG/X-Gal screening, and the relevant portion of *lacZ* was amplified and sequenced. From these sequences, it was determined which of the seven mutations were inherited in a given clone, and the total proportion of recombinants with each mutation was thereby calculated.

The results shown in Figure 4.1 clearly demonstrate that *nuc5*⁻ and EcNR2.*xseA*⁻ facilitate the inheritance of 3' mutations. *Nuc5*⁻ clones inherited significantly more of the mutations encoded by the *lacZ.7.stop* oligo (4.94 ± 0.20 , ***p* = 0.001, Figure 4.1A) compared to EcNR2 (3.92 ± 0.22). This was due to enhanced conferral of mutations at the 3' end of the 90mer

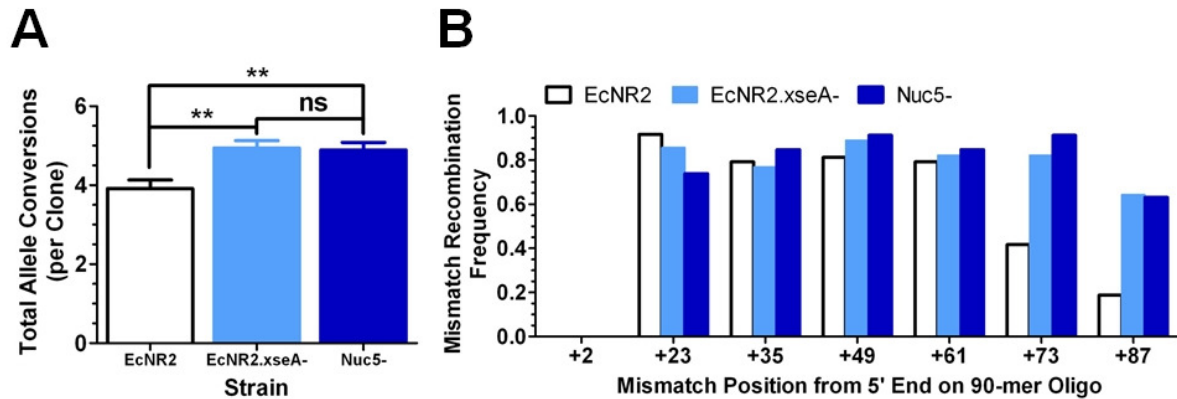


Figure 4.1: Removal of ExoVII Improves Oligonucleotide Mutation Inheritance A 90mer lagging strand-targeting oligonucleotide was designed to introduce 7 stop codons (placed at the +2, +23, +35, +49, +61, +73, +87 base pair positions with respect to the 5' end) into the *lacZ* gene. This oligo was transformed into EcNR2, EcNR2.xseA⁻, and nuc5⁻ cells. **A)** Average total number of stop codons introduced per clone for a given strain. Data are presented as the mean with the standard error of the mean. Both EcNR2.xseA⁻ (**p = 0.0004, n = 54) and nuc5⁻ (**p = 0.001, n = 46) exhibit similar, statistically significant increases in total mutation inheritance compared to EcNR2 controls (n = 45). “ns” = not significant. **B)** A breakdown of each strain’s inheritance of each mutation. Both EcNR2.xseA⁻ and nuc5⁻ show similarly increased inheritance of 3'-located mutations, but no increased inheritance of 5'-located mutations. No clones inherited the mutation encoded at position +2.

oligonucleotide (Figure 4.1B). Interestingly, removing ExoVII phenocopies the performance of nuc5⁻, leading to improved mean conversions (4.89 ± 0.18 , **p = 0.0004, Figure 4.1A) compared to EcNR2, and no significant difference from nuc5⁻ (p = 0.8416). Moreover, EcNR2.xseA⁻ also provides for significantly greater inheritance of mutations at the 3' end of the 90mer oligonucleotide (Figure 4.1B). Despite the dual polarity of ExoVII, its removal has no apparent effect on the inheritance of 5' mutations, nor does the removal of the other four exonucleases inactivated in nuc5⁻. This suggests that another unidentified nuclease may be responsible for degrading the 5' ends of oligonucleotides. Such degradation may be occurring postsynaptically, possibly mediated by the 5'-to-3' exonuclease domain of polymerase I.¹¹ The 3' protection effect of the ExoVII knockout strain is equivalent to that observed for the nuc5⁻ strain, suggesting that ExoVII is the only one of the five removed exonucleases which compromises the inheritance of mutations along the length of phosphorothioated oligos. Thus, the removal of ExoVII provides a simple solution for improving the conferral of mutations carried on the 3' ends of oligos.

Nuclease Knockouts Improve CoS-MAGE Performance

Given our reasoning described in the chapter introduction, as well as our above observation that ExoVII degrades oligonucleotides and may hinder their ability to recombine, we next tested whether the removal of ExoVII could improve multiplex oligonucleotide recombination frequency. To investigate this, we compared the MAGE performance of EcNR2 and EcNR2.*xseA*⁻. The *nuc5*⁻ strain discussed previously was also tested, in order to determine if other exonucleases impact MAGE performance; Lambda Exo was removed from this strain along with the four potent endogenous ssDNA exonucleases,⁹ as Lambda Exo has been shown¹² to have trace activity against ssDNA, and is not required for oligo recombination. In these experiments, we used co-selection MAGE (CoS-MAGE)⁵ in order to determine whether the nuclease knockout strains are able to improve upon the current best practices for MAGE. In CoS-MAGE, a co-selection oligo is directed to repair a mutated selectable marker near the loci

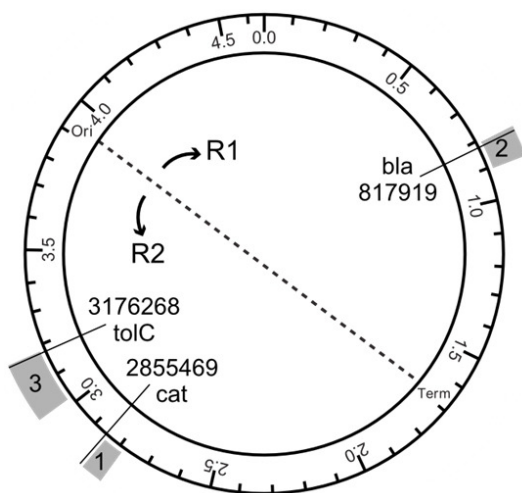


Figure 4.2: Orientation of the Tested Oligo Sets. The *E. coli* MG1655 genome is represented with the origin (Ori) and terminus (Term) of replication indicated, thus splitting the genome into Replichore 1 (R1) and Replichore 2 (R2). The genomic regions targeted by each of the three tested oligo sets are denoted in gray. Co-selection marker (*tolC*, *cat*, and *bla*) positions for each oligo set are indicated by radial lines.

targeted by the other recombineering oligos.⁵

Selection for the repaired marker thereby enriches for cells with high levels of recombination in the targeted vicinity, and increases recombination frequencies significantly.

We tested three different sets of 10 recombineering oligos (each specifying a single TAG→TAA mutation, as designed previously),⁴ so as to ascertain the CoS-MAGE performance of the strains at multiple loci, and in both replichores (Figure 4.2). This was done in order to confirm the

robustness of the results, as oligo recombination frequency can vary due to largely unelucidated oligo-specific and locus-specific effects.⁴ Each of the three oligo sets was paired with a co-selection oligonucleotide as described below. All recombineering oligos had two PT bonds on each end, as had previously been optimized for MAGE.⁴ Targeted loci were screened by mascPCR^{4,13} in order to determine which alleles were converted in a given clone.

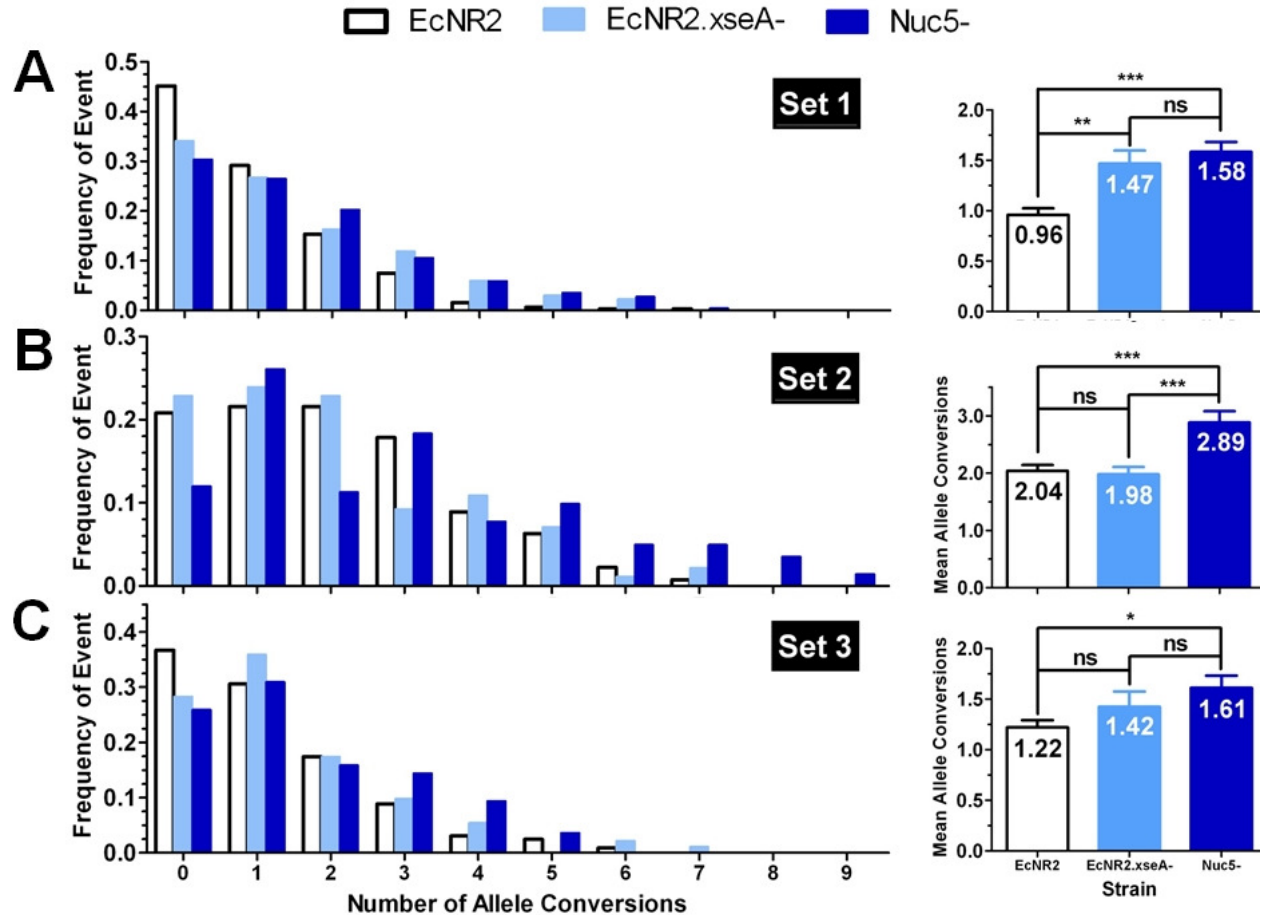


Figure 4.3: Effect of Nuclease Removal on CoS-MAGE Performance. CoS-MAGE was carried out in three strains (EcNR2, EcNR2.xseA⁻, and nuc5⁻), using three sets of ten oligos and a co-selection oligo, as shown in Figure 4.2. **A)** Set 1 was co-selected with *cat*, inserted at the *mutS* locus. In comparison with EcNR2 (n = 319), both EcNR2.xseA⁻ (**p = 0.0001, n = 135) and nuc5⁻ (***p < 0.0001, n = 257) show statistically significant increases in mean allele conversions, decreased proportions of clones exhibiting no allele conversions, and more clones with 5+ conversions. **B)** Set 2 was co-selected with *bla*, inserted with the Lambda prophage. Here, nuc5⁻ (n = 142) shows a statistically significant increase in recombineering performance compared to both EcNR2 (***p < 0.0001, n = 268) and EcNR2.xseA⁻ (***p < 0.0001, n = 184). **C)** Set 3 was co-selected with endogenous *tolC*. Here, nuc5⁻ (n = 139) shows a statistically significant increase in mean allele conversion compared to EcNR2 (*p = 0.002, n = 327). EcNR2.xseA⁻ (n = 92) shows an intermediate phenotype between EcNR2 (p = 0.2) and nuc5⁻ (p = 0.3). All oligos used in this experiment had two PT bonds on both ends. Data shown in the right-hand panels are presented as the mean with the standard error of the mean. Statistical significance is denoted as follows: ns indicates a non-significant variation, * indicates p < 0.003, ** indicates p < 0.001, and *** indicates p < 0.0001.

Results are shown in Figure 4.3. For all three recombineering oligo sets (Sets 1-3 in Figure 4.3A-C, respectively), *nuc5⁻* significantly outperformed EcNR2 (****p* < 0.0001, ****p* < 0.0001, **p* = 0.002, respectively). An average of 46% more alleles were converted per clone in *nuc5⁻*, and the frequency of clones with 5 or more conversions was increased by 200%. Furthermore, *nuc5⁻* reduced the frequency of clones with no conversions by 35%. All of these recombination improvements are particularly important given that MAGE can be performed in iterative cycles, thereby compounding these enhancements. Thus, removing all potent ssDNA exonucleases significantly improves the performance of CoS-MAGE.

The EcNR2.*xseA⁻* strain appears to have intermediate properties between those of EcNR2 and *nuc5⁻*. Although EcNR2.*xseA⁻* demonstrated a statistically significant increase in CoS-MAGE performance with Set 1 (1.47 ± 0.13) compared to EcNR2 (0.96 ± 0.07 , ***p* = 0.0001), this strain's performance with Sets 2 & 3 was not statistically different from that of EcNR2 (*p* = 0.7 & 0.2). Given that Set 1 also exhibited the largest difference in performance between EcNR2 and *nuc5⁻* (65% higher average allele conversion in *nuc5⁻*), it is possible that Set 1 is the most susceptible to nuclease repression; as such, the effect of removing ExoVII would be most apparent for this set. Overall, *nuc5⁻* was superior to EcNR2.*xseA⁻* for the three tested oligo sets. This suggests that the action of ExoVII somewhat compromises CoS-MAGE frequency, but that some or all of the other exonucleases removed in *nuc5⁻* also have roles in oligo degradation.

Examining the Effect of Phosphorothioate Bonds on CoS-MAGE Frequency

As noted, the above experiments were performed with recombineering oligonucleotides with two PT bonds on each end. We next sought to determine whether the optimal number of PT bonds was the same for each of our nuclease knockout strains, whether the benefits of nuclease

removal could be recapitulated simply by adding more PT bonds to the recombineering oligos, and whether the differences between the strains would be more pronounced if no PT bonds were used. Thus, we recombined EcNR2 (Figure 4.4A), EcNR2.*xseA*⁻ (Figure 4.4B), and nuc5⁻ (Figure 4.4C) with versions of recombineering oligo Set 1, co-selected with a restoration oligo for the nearby mutated *cat* gene. A non-phosphorothioated version of oligo Set 1 was tested, as was a version with 4 PT bonds on each end. The resulting allele conversion distributions were determined as above, and compared with those previously observed for the version of Set 1 with 2 PT bonds on each end.

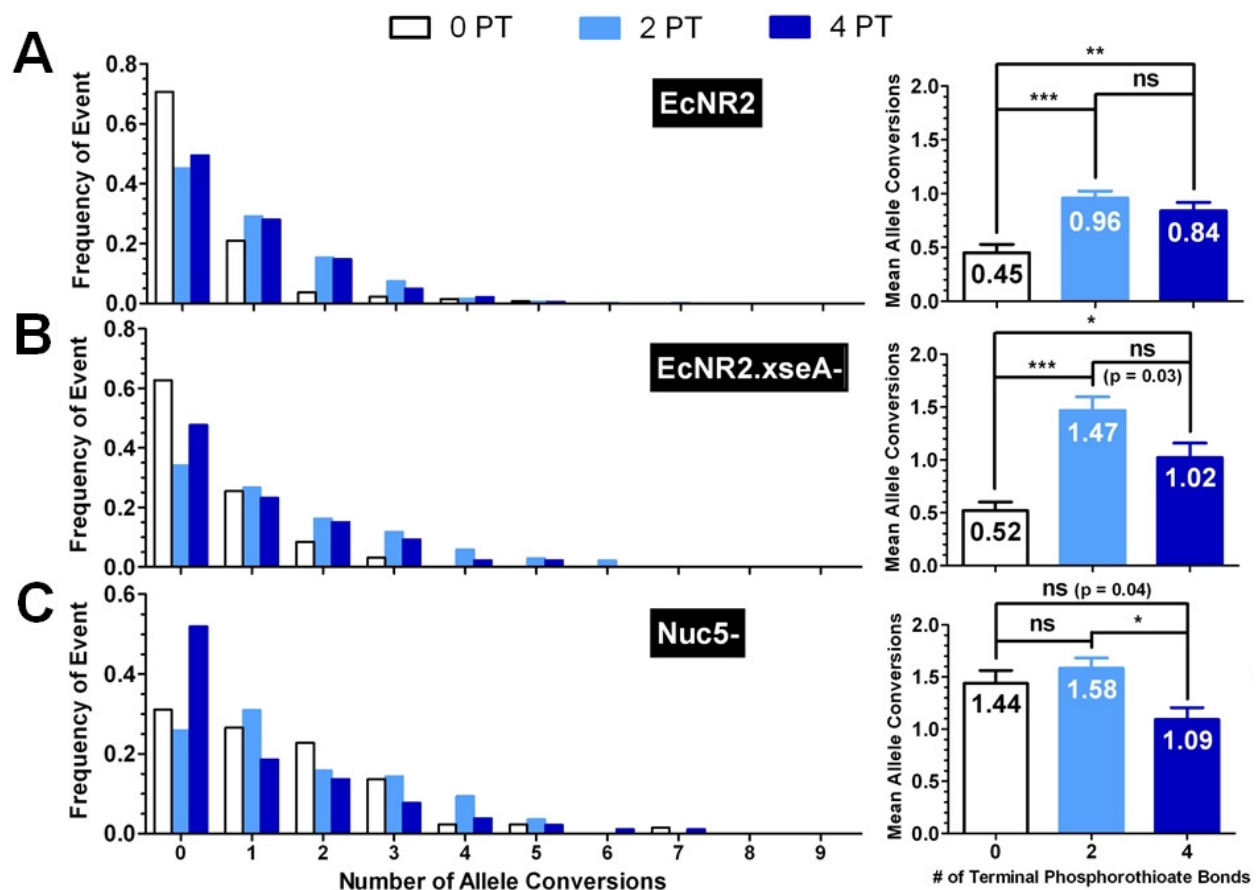


Figure 4.4: Effect of Phosphorothioate Bonds on CoS-MAGE Performance. We assessed the effect of oligo phosphorothioation on the CoS-MAGE performance of the various nuclease knockout strains. Variants of oligo Set 1 with no PT bonds (0 PT) and four PT bonds on both ends (4 PT) were compared with the initial version of Set 1, which had two PT bonds on both ends (2 PT). Recombination was performed with *cat* co-selection as prior. **A)** For EcNR2, n = 133 (0 PT), n = 319 (2 PT), n = 186 (4 PT). **B)** For EcNR2.*xseA*⁻, n = 94 (0 PT), n = 92 (2 PT), n = 86 (4 PT). **C)** For nuc5⁻, n = 132 (0 PT), n = 257 (2 PT), n = 183 (4 PT). Data shown in the right-hand panels are presented as the mean with the standard error of the mean. Statistical significance is as denoted in Figure 4.3.

The results are shown in Figure 4.4. Notably, for all tested strains, the oligo set with 2 PT bonds gave the greatest average number of allele conversions per clone. Comparatively, 4 PT bonds had a detrimental effect on recombination frequency that was most notable in *nuc5⁻* (Figure 4.4C), where the 2 PT set (1.58 ± 0.10) significantly outperformed the 4 PT set (1.09 ± 0.11 , $*p = 0.001$). Similarly, for *EcNR2.xseA⁻* (Figure 4.4B), the 2 PT set (1.47 ± 0.13) exhibited a local optimum compared to the 0 PT set (0.52 ± 0.08 , $***p < 0.0001$) and the 4 PT set (1.02 ± 0.14 , $p = 0.03$), corroborating the detrimental effect of too many PT bonds. Thus, the *EcNR2.xseA⁻* and *nuc5⁻* strains confer recombineering advantages that cannot be recapitulated simply by preventing nuclease degradation through the use of more PT bonds. The detrimental effect of high levels of phosphorothioation may be due to PT bonds decreasing the strength of the annealing interaction between the oligo and the lagging strand of the replication fork.¹⁴ Alternatively, oligos with PT bonds may be somewhat toxic, thereby killing cells that take up a large number of oligos and would otherwise yield many converted alleles.

The relative performance of the differently phosphorothioated oligo sets in the tested strains suggests that this detrimental effect of PT bonds is counterbalanced by the beneficial effect of nuclease protection. In *EcNR2* (Figure 4.4A), which has a full complement of endogenous nucleases, the 4 PT set (0.84 ± 0.08) greatly outperformed the 0 PT set (0.45 ± 0.08 , $**p = 0.0006$), suggesting that the effect of nuclease protection outweighs the detrimental impact of PT bonds on recombination frequency. Conversely, in *nuc5⁻* (Figure 4.4C), where nuclease degradation is mitigated by knockouts, the 0 PT set (1.44 ± 0.12) slightly outperformed the 4 PT set (1.09 ± 0.11 , $p = 0.04$), suggesting that the detrimental impact of the PT bonds outweighs any beneficial effect of nuclease protection. Interestingly, in this strain, the 2 PT set (1.58 ± 0.10) is statistically equivalent to the 0 PT set (1.44 ± 0.12 , $p = 0.38$), suggesting that most or all relevant

nuclease activity has been abrogated. However, some level of residual exonuclease activity in *nuc5⁻* is suggested by the strain's poor inheritance of mutations encoded at the 5' end of oligos (Figure 4.1B).

When no PT bonds were used to protect the recombineering oligonucleotides, the *nuc5⁻* strain yielded roughly threefold more allele conversions per average clone than EcNR2. This stands to reason, as non-PT oligos are likely to be particularly susceptible to the activity of endogenous ssDNA exonucleases. Similarly, EcNR2.*xseA⁻* also had low recombination frequency (roughly similar to that of EcNR2) when 0 PT oligos were used. This suggests that exonucleases other than ExoVII (i.e., RecJ, ExoI, ExoX, and/or Lambda Exo) are readily capable of degrading non-protected oligos. However, when PT bonds were used, EcNR2.*xseA⁻* notably outperformed EcNR2, and was only slightly less recombinogenic than *nuc5⁻* for such oligo sets. Thus, the exonuclease activity of ExoVII is relevant primarily for phosphorothioated oligonucleotides. Taken together, these results reinforce the importance of using phosphorothioated oligos when performing MAGE in a strain containing endogenous nucleases, but caution that the overuse of phosphorothioates can be detrimental.

Discussion

The work described in this chapter adds to our understanding of Lambda Red oligonucleotide recombination. We have shown that removing selected endogenous nucleases improves multiplex recombination frequency, despite the use of a high total concentration (5.2 μ M) of oligos. This suggests that the intracellular concentration of any given oligo is not enough to overcome the action of endogenous nucleases, and therefore that oligo entry into the cell is a limiting factor in MAGE. This observation initially appears to contradict the conclusion

presented in a recent study by Sawitzke *et al.*,¹⁰ which showed that the singleplex recombination frequency of a given oligonucleotide could be increased by the addition of non-specific carrier oligos. These carrier oligos saturate endogenous exonucleases, preventing them from degrading the recombinogenic oligonucleotide. However, this phenomenon was tested only for very low concentrations of the recombinogenic oligo (up to 0.01 μM). Even at this concentration, adding carrier oligos (at 0.1 μM) yielded less than a 2-fold enhancement of recombination frequency; a more pronounced enhancement was observed only for even lower concentrations of recombinogenic oligo (0.001 μM and below). In concentration regimes typical for MAGE (> 1 μM total oligos), it has been shown that adding a second oligonucleotide *decreases* the recombination frequency of the first oligonucleotide.⁵ Presumably, at these concentrations (which have previously been shown to be optimal for recombination frequency¹ and were therefore used in this work), any benefit conferred by the saturation of endogenous exonucleases is outweighed by competition for cellular entry. These findings therefore suggest that enhancing oligo uptake is likely to be a fruitful avenue for further improving MAGE.

Additionally, studying the consequences of different levels of oligonucleotide phosphorothioation in the EcNR2, EcNR2.*xseA*⁻, and nuc5⁻ strains enabled the deconvolution of the countervailing effects of phosphorothioation. Phosphorothioate bonds can improve recombination frequency by protecting oligos from nuclease degradation, but can also diminish recombination frequency – possibly by reducing the strength of the annealing interaction between the oligo and the lagging strand of the replication fork,¹⁴ or by causing toxicity to the cell. Placing two PT bonds on both ends of recombineering oligonucleotides was found to be optimal for all three strains tested in this work, but future nuclease-modified strains will need to

be optimized in order to determine the ideal number of PT bonds for oligos recombined into that strain.

Beyond providing additional insight into the recombination process, this work also achieves several important improvements of oligonucleotide recombineering. Firstly, removing ExoVII was shown to improve the inheritance of mutations at the 3' ends of oligonucleotides. This may be quite useful, as it allows more mutations to reliably be introduced by a single oligo. This could be leveraged for several applications, such as simultaneously modifying several residues near the active site of a protein, recoding a larger region with a given oligonucleotide,⁴ or modifying several genetic features (e.g., promoter strength, ribosome binding site strength, and the presence or absence of a premature stop codon) with a single oligo. Similarly, the improvement of CoS-MAGE recombination frequency via the use of EcNR2.*xseA*⁻ and nuc5⁻ will also have substantial utility. These strains facilitate greater modification of a population of cells, which will be useful for projects that seek rapid genomic diversification.¹ Such advancements are also expected to be useful for improving the Red-mediated diversification of BACs and plasmids. Similarly, the enhanced recombination frequency of these strains means that fewer cycles will be needed in order to achieve an isogenic recoded population of cells, or to identify a strain with all desired genetic changes among a set of screened clones. This will be highly useful for future genome engineering efforts.

Finally, this work provides guidelines as to the appropriate strain to use for a given recombineering application. It should be noted that the nuc5⁻ strain was observed to have poor regrowth after electroporation, taking roughly twice as long as EcNR2 or EcNR2.*xseA*⁻ to recover to confluence. The pre-electroporation growth rate of nuc5⁻ was only slightly less than those of EcNR2 and EcNR2.*xseA*⁻ (~150 minutes vs. ~125 minutes to reach mid-logarithmic

growth phase from a 1:100 dilution of overnight culture), likely due to the ability of *mutS* removal to suppress the known cold-sensitive growth phenotype of *recJ/xonA/xseA/exoX* quadruple mutants.¹⁵ Therefore, while *nuc5⁻* has somewhat better multiplex recombination frequency than *EcNR2.xseA⁻*, its poor regrowth properties cause each recombination cycle to take notably longer. Thus, using *EcNR2.xseA⁻* is likely optimal for applications in which multiple cycles are necessary; the *nuc5⁻* strain may be preferable for applications in which a single cycle is sufficient and fast regrowth is not necessary. Additionally, while quadruple mutants for *recJ*, *xonA*, *xseA*, and *exoX* have increased point mutation rates, this phenomenon is epistatic to *mutS*.¹⁵ Given that strains used for genome engineering are often *mutS⁻*, removing these nucleases (as in *nuc5⁻*) does not further exacerbate the mutator phenotype. However, the combined removal of *xonA*, *recJ*, and *exoX* has been shown to increase rates of rearrangement mutations involving repetitive sequences,¹⁶ and therefore this strain should not be used for applications in which genomic stability is paramount. Directed evolution and/or the restoration of selected nucleases may facilitate improved growth rates and genomic stability, without substantially compromising recombination frequency. Indeed, we have recently conducted experiments on *nuc5⁻* strains with either *xonA* or *recJ* reactivated. Preliminary results suggest that both of these strains have post-electroporation recovery rates equivalent to those of *EcNR2* and *EcNR2.xseA⁻*, and CoS-MAGE recombination frequencies equivalent to that of *nuc5⁻*. Additionally, given that these strains have active RecJ or ExoI, they would not be expected to have increased rates of rearrangement mutations. Thus, these strains may be ideal for recombineering.

In conclusion, the work presented in this chapter increases our understanding of the oligonucleotide recombineering process, confirming that nuclease degradation limits

recombination frequency. Additionally, our results indicate that oligo entry is a major limiting factor in recombination, and that phosphorothioate bonds can be detrimental as well as beneficial. This work also yielded strains with markedly improved mutation inheritance and allele conversion frequency. These strains will be highly useful as chassis in future recombineering efforts, and may enable new and powerful applications of Lambda Red technology.

Experimental

Oligonucleotides used in this Study

A full list of oligonucleotides used in this study is given in Table 4.1. For oligonucleotides described in this table, “wt-f” refers to a forward ascPCR/mascPCR primer used to detect a wild type allele, “mut-f” refers to a forward ascPCR/mascPCR primer used to detect a mutated allele, and “rev” refers to the reverse ascPCR/mascPCR primer used with both forward primers. Oligonucleotides from the three tested TAG→TAA recombineering sets are denoted “X.Y,” where X is the set that the oligo is from, and Y is the size of its corresponding mascPCR amplification band. Asterisks represent phosphorothioate bonds between the two indicated nucleotides. For oligo sequences denoted in the table with a §, variants with no PT bonds and 4 PT bonds at each end were also used. All oligonucleotides were ordered from Integrated DNA Technologies with standard purification and desalting.

Table 4.1: Oligonucleotides used in Chapter 4

Name	Use	Sequence
ygaR	Set 1.850 TAG → TAA	§G*C*GAAGATCAGTAAAGATATAGAAGGTGGTATCCCT GGCTATTAACAAGGTCAGGTTTTGATTCCATTCATTAAA GATCCAGTAACAA*A*A
yqaC	Set 1.700 TAG → TAA	§A*T*TAAAAATTATGATGGGTCCACGCGTGTCTGGCGGTG AGGCGTAACTTAATAAAGGTTGCTCTACCTATCAGCAGC TCTACAATGAAT*T*C

Table 4.1 (Continued)

gabT	Set 1.600 TAG → TAA	\$T^*C^*ACCATTGAAGACGCTCAGATCCGTCAGGGTCTGGA GATCATCAGCCAGTGTTTTGATGAGGCGAAGCAGTAACG CCGCTCCTATGC^*C^*G
ygaU	Set 1.500 TAG → TAA	\$T^*G^*ACGCCAATTCCCATTATCCAGCAGGCGATGGCTGG CAATTAATTACTCTTCCGGAATACGCAACACTTGCCCCG GATAAATTTTAT^*C^*C
ygaM	Set 1.400 TAG → TAA	\$G^*T^*AGGTATTTTTATCGGCGCACTGTAAAGCATGCGCA AATCGTAATGCAAAAATGATAATAAATACGCGTCTTTGA CCCCGAAGCCTG^*T^*C
luxS	Set 1.300 TAG → TAA	\$T^*T^*TGAACCTGGCTTTTTTCAATTAATTGTGAAGATAGTT TACTGATTAGATGTGCAGTTCCTGCAACTTCTCTTTCGGC AGTGCCAGTT^*C^*T
mltB	Set 1.250 TAG → TAA	\$A^*A^*TTTTACGAGGAGGATTCAGAAAAAAGCTGATTAG CCAGAGGGAAGCTCACGCCCCCTCTTGTAATAGTTAC TGTA CTGCGCCA^*G^*C
srlE	Set 1.200 TAG → TAA	\$A^*C^*TGTA CTGATCGCCTGGTTTGTCTCCGGTTTTATCTA TCAATAAAGGCTGAAACATGACCGTTATTTATCAGACCA CCATCACCCGT^*A^*T
norW	Set 1.150 TAG → TAA	\$A^*T^*CGGATGAAAGAGGCATTTGGATTGTTGAAAACATT GCCGATGTAAGTGGGCTACTGTGCCTAAAATGTCGGATG CGACGCTGGCGC^*G^*T
ascB	Set 1.100 TAG → TAA	\$A^*T^*CATTCTGGTGGTATAAAAAAGTGATTGCCAGTAAT GGGGAAGATTTAGAGTAAGTAACAGTGCCGGATGCGGC GTGAACGCCTTAT^*C^*C
bioD	Set 2.850 TAG → TAA	T^*C^*GAAGACGCGATCTCGCTCGCAATTTAACCAAATAC AGAATGGTTACAACAAGGCAAGGTTTATGTACTTTCCGG TTGCCGCATTTT^*C^*T
moaE	Set 2.700 TAG → TAA	C^*G^*TAAACGTATGTACTGAGCGGTGAAATTGCCGGACG CAGCGGTGCCTTATCCGGCTAACAAAAAATTACCAGCGT TTTGCCGCCTGC^*T^*G
ybhM	Set 2.600 TAG → TAA	G^*C^*GATGTGAAGTTTAGTTAAGTTCTTTAGTATGTGCAT TTACGGTTAATGAAAAAAACGCGTATGCCTTTGCCAGAC AAGCGTTATAG^*C^*T
ybhS	Set 2.500 TAG → TAA	T^*T^*TATCGGCCTGACGTGGCTGAAAACCAAACGTCGGC TGGATTAAAGGAGAAGCATGTTTCATCGCTTATGGACG TTAATCCGCAAA^*G^*A
ybiH	Set 2.400 TAG → TAA	C^*A^*TATCGACCTGATTTTGCAAGGATTATCGCAAAGGA GTTTGTAATGATGAAAAAACCTGTCGTGATCGGATTGGC GGTAGTGGTACT^*T^*G
ybiR	Set 2.300 TAG → TAA	T^*C^*TGAATTAATCTTCAAACTTAAAGCAAAAGGCGGA CTCATAATCCGCCTTTTTTATTTGCCAGACCTTAGTTGGC CGGGAGTATAA^*C^*T
yliD	Set 2.250 TAG → TAA	T^*T^*TCCTGTGAGGTGATTACCCTTTCAAGCAATATTCAA ACGTAATTATCCTTTAATTTTCGGATCCAGCGCATCGCGT AAACCATCGC^*C^*C
yliE	Set 2.200 TAG → TAA	G^*A^*CTGACTGTAAGTACGAACTTATTGATTCTGGACATA CGTAAATTACTCTTTTACTAATTTTCCACTTTTATCCCAG GCGGAGAATG^*G^*C
ybjK	Set 2.150 TAG → TAA	T^*C^*GGTTC AAGGTTGATGGGTTTTTTGTTATCTAAAAC TATCTATTACCCTGCAACCCTCTCAACCATCCTCAAAATC TCCTCGCGCG^*A^*T

Table 4.1 (Continued)

rimK	Set 2.100 TAG → TAA	C*G*C*AAAAAGCGCAGGCAAAACCATGATCAGTAATGTG ATTGCGATTAACCACCCGTTTTTCAGGCAATATTCTGTCGT AGCGTGGCGTT*C*G
ygfJ	Set 3.850 TAG → TAA	C*C*GGACGACTTTATTACAGCGAAGGAAAGGTATACTG AAATTTAAAAAACGTAGTTAAACGATTGCGTTCAAATAT TTAATCCTTCCG*G*C
recJ	Set 3.700 TAG → TAA	G*G*GATTGTACCCAATCCACGCTCTTTTTTATAGAGAAG ATGACGTTAAATTGGCCAGATATTGTTCGATGATAATTTG CAGGCTGCGGT*T*G
argO	Set 3.600 TAG → TAA	C*T*CTGGAGGCAAGCTTAGCGCCTCTGTTTTATTTTTCC ATCAGATAGCGCTTAACTGAACAAGGCTTGTGCATGAGC AATACCGTCTC*T*C
yggU	Set 3.500 TAG → TAA	A*A*TCCGCAACAAATCCCGCCAGAAATCGCGGCGTTAA TTAATTAAGTATCCTATGCAAAAAGTTGTCCTCGCAACC GGCAATGTCCGT*A*A
mutY	Set 3.400 TAG → TAA	G*T*GGAGCGTTTGTACAGCAGTTACGCACTGGCGCGC CGGTTTAAACGCGTGAGTCGATAAAGAGGATGATTTATGA GCAGAACGATTT*T*T
glcC	Set 3.300 TAG → TAA	G*C*CACCATTTGATTCGCTCGGCGGTGCCGCTGGAGATG AACCTGAGTTAACTGGTATTAATCTGCTTTTCATACAAT CGGTAACGCT*T*G
yghQ	Set 3.250 TAG → TAA	A*C*TGAGTCAGCCGAGAAGAATTTCCCCGCTTATTCGCA CCTTCCTTAAATCAGGTCATACGCTTCGAGATACTTAAC GCCAAACACCA*G*C
yghT	Set 3.200 TAG → TAA	T*G*GTTGATGCAGAAAAAGCGATTACGGATTTTATGAC CGCGCGTGGTTATCACTAATCAAAAATGGAAATGCCCGA TCGCCAGGACCG*G*G
ygiZ	Set 3.150 TAG → TAA	T*T*CTCTGTCTATGAGAGCCGTTAAAACGACTCTCATAG ATTTTATTAATAGCAAAATATAAACCGTCCCCAAAAAAG CCACCAACCAC*A*A
yqiB	Set 3.100 TAG → TAA	A*G*GGTTAACAGGCTTTCCAAATGGTGTCTTAGGTTTC ACGACGTTAATAAACCGGAATCGCCATCGCTCCATGTGC TAAACAGTATC*G*C
ygaR_wt-f	Set 1.850_wt-f mascPCR	AAGGTGGTATCCCTGGCTATTAG
yqaC_wt-f	Set 1.700_wt-f mascPCR	CGGCGGTGAGGCGTAG
gabT_wt-f	Set 1.600_wt-f mascPCR	TTTTGATGAGGCGAAGCAGTAG
ygaU_wt-f	Set 1.500_wt-f mascPCR	GTTGCGTATTCCGGAAGAGTAG
ygaM_wt-f	Set 1.400_wt-f mascPCR	GTTAAGCATGCGCAAATCGTAG
luxS_wt-f	Set 1.300_wt-f mascPCR	GTTGCAGGAAGTGCACATCTAG
mltB_wt-f	Set 1.250_wt-f mascPCR	GCTGGCGCGAGTACAGTAG
srIE_wt-f	Set 1.200_wt-f mascPCR	GGTTTGTCTCCGGTTTTATCTATCAATAG
norW_wt-f	Set 1.150_wt-f mascPCR	GATTGTTGAAAACATTGCCGATGTAG
ascB_wt-f	Set 1.100_wt-f mascPCR	CCAGTAATGGGGAAGATTTAGAGTAG
bioD_wt-f	Set 2.850_wt-f mascPCR	AGTACATAAACCTTGCCCTGTTGTAG
moaE_wt-f	Set 2.700_wt-f mascPCR	GCGGCAAAACGCTGGTAG
ybhM_wt-f	Set 2.600_wt-f mascPCR	AAGGCATACGCGTTTTTTTCATTAG
ybhS_wt-f	Set 2.500_wt-f mascPCR	CCAAACGTCGGCTGGATTAG

Table 4.1 (Continued)

ybiH_wt-f	Set 2.400_wt-f mascPCR	AAGGATTATCGCAAAGGAGTTTGTAG
ybiR_wt-f	Set 2.300_wt-f mascPCR	TTAGTTATACTCCCGGCCAACTAG
yliD_wt-f	Set 2.250_wt-f mascPCR	CGCTGGATCCGAAAATTAAAGGATAG
yliE_wt-f	Set 2.200_wt-f mascPCR	TGGGATAAAAGTGGAAAATTAGTAAAAGAGTAG
ybjK_wt-f	Set 2.150_wt-f mascPCR	TTGAGAGGGTTGCAGGGTAG
rimK_wt-f	Set 2.100_wt-f mascPCR	GCCTGAAAACGGGTGGTTAG
ygfJ_wt-f	Set 3.850_wt-f mascPCR	AGCGAAGGAAAGGTATACTGAAATTTAG
recJ_wt-f	Set 3.700_wt-f mascPCR	TCATCGACAATATCTGGCCAATTTAG
argO_wt-f	Set 3.600_wt-f mascPCR	TGCACAAGCCTTGTTTCAGTTAG
yggU_wt-f	Set 3.500_wt-f mascPCR	CAGAAATCGCGGCGTTAATTAATTAG
mutY_wt-f	Set 3.400_wt-f mascPCR	GGCGCGCCGGTTTAG
glcC_wt-f	Set 3.300_wt-f mascPCR	GCTGGAGATGAACCTGAGTTAG
yghQ_wt-f	Set 3.250_wt-f mascPCR	CTCGAAGCGTATGACCTGATTTAG
yghT_wt-f	Set 3.200_wt-f mascPCR	CGCGCGTGGTTATCACTAG
ygiZ_wt-f	Set 3.150_wt-f mascPCR	TGGGGACGGTTTATATTTTGCTATTAG
yqiB_wt-f	Set 3.100_wt-f mascPCR	CGATGGCGATTCCGGTTTATTAG
ygaR_mut-f	Set 1.850_mut-f mascPCR	AAGGTGGTATCCCTGGCTATTAA
yqaC_mut-f	Set 1.700_mut-f mascPCR	CGGCGGTGAGGCGTAA
gabT_mut-f	Set 1.600_mut-f mascPCR	TTTTGATGAGGCGAAGCAGTAA
ygaU_mut-f	Set 1.500_mut-f mascPCR	GTTGCGTATTCCGGAAGAGTAA
ygaM_mut-f	Set 1.400_mut-f mascPCR	GTTAAGCATGCGCAAATCGTAA
luxS_mut-f	Set 1.300_mut-f mascPCR	GTTGCAGGAACTGCACATCTAA
mltB_mut-f	Set 1.250_mut-f mascPCR	GCTGGCGCGAGTACAGTAA
srIE_mut-f	Set 1.200_mut-f mascPCR	GGTTTGTCTCCGGTTTTATCTATCAATAA
norW_mut-f	Set 1.150_mut-f mascPCR	GATTGTTGAAAACATTGCCGATGTAA
ascB_mut-f	Set 1.100_mut-f mascPCR	CCAGTAATGGGGAAGATTTAGAGTAA
bioD_mut-f	Set 2.850_mut-f mascPCR	AGTACATAAACCTTGCCTTGTTGTAA
moaE_mut-f	Set 2.700_mut-f mascPCR	GCGGCAAAACGCTGGTAA
ybhM_mut-f	Set 2.600_mut-f mascPCR	AAGGCATACGCGTTTTTTTCATTAA
ybhS_mut-f	Set 2.500_mut-f mascPCR	CCAAACGTCGGCTGGATTAA
ybiH_mut-f	Set 2.400_mut-f mascPCR	AAGGATTATCGCAAAGGAGTTTGTAA
ybiR_mut-f	Set 2.300_mut-f mascPCR	TTAGTTATACTCCCGGCCAACTAA
yliD_mut-f	Set 2.250_mut-f mascPCR	CGCTGGATCCGAAAATTAAAGGATAA
yliE_mut-f	Set 2.200_mut-f mascPCR	TGGGATAAAAGTGGAAAATTAGTAAAAGAGTAA
ybjK_mut-f	Set 2.150_mut-f mascPCR	TTGAGAGGGTTGCAGGGTAA
rimK_mut-f	Set 2.100_mut-f mascPCR	GCCTGAAAACGGGTGGTTAA
ygfJ_mut-f	Set 3.850_mut-f mascPCR	AGCGAAGGAAAGGTATACTGAAATTTAA
recJ_mut-f	Set 3.700_mut-f mascPCR	TCATCGACAATATCTGGCCAATTTAA

Table 4.1 (Continued)

argO_mut-f	Set 3.600_mut-f mascPCR	TGCACAAGCCTTGTTTCAGTTAA
yggU_mut-f	Set 3.500_mut-f mascPCR	CAGAAATCGCGGCGTTAATTAATTAA
mutY_mut-f	Set 3.400_mut-f mascPCR	GGCGCGCCGGTTTAA
glcC_mut-f	Set 3.300_mut-f mascPCR	GCTGGAGATGAACCTGAGTTAA
yghQ_mut-f	Set 3.250_mut-f mascPCR	CTCGAAGCGTATGACCTGATTTAA
yghT_mut-f	Set 3.200_mut-f mascPCR	CGCGCGTGGTTATCACTAA
ygiZ_mut-f	Set 3.150_mut-f mascPCR	TGGGGACGGTTTATATTTTGCTATTAA
yqiB_mut-f	Set 3.100_mut-f mascPCR	CGATGGCGATTCCGGTTTATTAA
ygaR_rev	Set 1.850_rev mascPCR	TAGGTAGAGCAACCTTTATTAAGCTACG
yqaC_rev	Set 1.700_rev mascPCR	TAAAAATATCTACATTTCTGAAAAATGCGCA
gabT_rev	Set 1.600_rev mascPCR	GCGGCGATGTTGGCTT
ygaU_rev	Set 1.500_rev mascPCR	AGGGTATCGGGTGGCG
ygaM_rev	Set 1.400_rev mascPCR	CGCAACGCTTCTGCCG
luxS_rev	Set 1.300_rev mascPCR	ATGCCCAGGCGATGTACA
mltB_rev	Set 1.250_rev mascPCR	AGACTCGGCAGTTGTTACGG
srlE_rev	Set 1.200_rev mascPCR	GGATGGAGTGCACCTTTCAAC
norW_rev	Set 1.150_rev mascPCR	GTGTTGCATTTGGACACCATTG
ascB_rev	Set 1.100_rev mascPCR	CGCTTATCGGGCCTTCATG
bioD_rev	Set 2.850_rev mascPCR	CGGGAAGAACTCTTTTCATTTCGC
moaE_rev	Set 2.700_rev mascPCR	CGTCAATCCGACAAAGACAATCA
ybhM_rev	Set 2.600_rev mascPCR	TTACTGGCAGGGATTATCTTTACCG
ybhS_rev	Set 2.500_rev mascPCR	CTGTTGTTAGGTTTCGGTTTTCT
ybiH_rev	Set 2.400_rev mascPCR	GTCATAGGCGGCTTGCG
ybiR_rev	Set 2.300_rev mascPCR	ATGAGCCGGTAAAAGCGAC
yliD_rev	Set 2.250_rev mascPCR	AATAAAATTATCAGCCTTATCTTTATCTTTTCGTATAAA
yliE_rev	Set 2.200_rev mascPCR	CAGCAATATTTGCCACCGCA
ybjK_rev	Set 2.150_rev mascPCR	AACTTTTCCGCAGGGCATC
rimK_rev	Set 2.100_rev mascPCR	TACAACCTCTTTCGATAAAAAGACCG
ygfJ_rev	Set 3.850_rev mascPCR	GATGAACTGTTGCATCGGCG
recJ_rev	Set 3.700_rev mascPCR	CTGTACGCAGCCAGCC
argO_rev	Set 3.600_rev mascPCR	AATCGCTGCCTTACGCG
yggU_rev	Set 3.500_rev mascPCR	TAACCAAAGCCACCAGTGC
mutY_rev	Set 3.400_rev mascPCR	CGCGAGATATTTTTTCATCATTCGG
glcC_rev	Set 3.300_rev mascPCR	GGGCAAAATTGCTGTGGC
yghQ_rev	Set 3.250_rev mascPCR	ACCAACTGGCGATGTTATTCAC
yghT_rev	Set 3.200_rev mascPCR	GACGATGGTGGTGGACGG
ygiZ_rev	Set 3.150_rev mascPCR	ATCGCCAAATTGCATGGCA
yqiB_rev	Set 3.100_rev mascPCR	AAAATCCTGACTCTGGCCTCA

Table 4.1 (Continued)

Lexo.KO.MM*	Lambda Exo inactivating oligo	T*G*A*A*ACAGAAAGCCGCAGAGCAGAAGGTGGCAGCA TGACACCGTAACATTATCCTGCAGCGTACCGGGATCGAT GTGAGAGCTGTCGAAC
Lexo_WT-f	Lambda Exo wt-f mascPCR primer	GGCAGCATGACACCGGA
Lexo_MUT-f	Lambda Exo mut-f mascPCR primer	TGGCAGCATGACACCGTAA
Lexo-r	Lambda Exo rev mascPCR primer	CAAGGCCGTTGCCGTC
cat_mut*	cat inactivation oligo	G*C*ATCGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTG CTTAATGTACCTATAACCAGACCGTTCAGCTGGATATTA CGGCCTTTTTA*A*A
cat_restore*	cat reactivation oligo (for coselection)	G*C*ATCGTAAAGAACATTTTGAGGCATTTTCAGTCAGTTG CTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTA CGGCCTTTTTA*A*A
tolC-r_null_mut*	tolC inactivation oligo	A*G*CAAGCACGCCTTAGTAACCCGGAATTGCGTAAGTC TGCCGCTAAATCGTGATGCTGCCTTTGAAAAAATTAATG AAGCGCGCAGTCCA
tolC-r_null_revert*	tolC reactivation oligo (for coselection)	C*A*GCAAGCACGCCTTAGTAACCCGGAATTGCGTAAGT CTGCCGCCGATCGTGATGCTGCCTTTGAAAAAATTAATG AAGCGCGCAGTCCA
bla_mut*	bla inactivation oligo	G*C*C*A*CATAGCAGAACTTTAAAAGTGCTCATCATTGG AAAACGTTATTAGGGGCGAAAACTCTCAAGGATCTTACC GCTGTTGAGATCCAG
bla_restore*	bla reactivation oligo (for coselection)	G*C*C*A*CATAGCAGAACTTTAAAAGTGCTCATCATTGG AAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACC GCTGTTGAGATCCAG
lacZ.seq-f	Forward primer for amplifying/sequencing lacZ	CGCAATTAATGTGAGTTAGCTCACTC
lacZ.seq-2-r	Reverse primer for amplifying/sequencing lacZ	CGCCGAGTTAACGCCATCAA
LacZ7Stop.PT	Oligo for disrupting lacZ gene w/7 stop codons	T*A*G*C*GCAGCCTGAATGGCGAATAGCGCTTTGCCTAG TTTCCGGCACCATAAGCGGTGCCGTAAAGCTGGCTGTAG TGCGATCTTCC*T*T*A*G

Strain Creation

Strains were created as described in Chapter 3. The nuc5⁻ strain was generated from the nuc4⁻ strain, using the recombineering oligonucleotide denoted in Table 4.1 to inactivate Lambda Exo. Strains used for CoS-MAGE were generated by recombining an oligonucleotide designed to inactivate a chromosomal resistance marker (*cat*, *tolC*, or *bla*) and identifying resulting colonies with the appropriate sensitivity to antibiotic or SDS.¹⁷

Performing Lambda Red Recombination

Lambda Red recombination was performed as described in Chapter 3. For the *lacZ.7.stop* experiment, in which a single oligonucleotide was recombined, 1 μM of oligo was used. For experiments in which sets of ten recombineering oligos were recombined along with a co-selection oligo, 0.5 μM of each recombineering oligo was used, along with 0.2 μM of the co-selection oligo (5.2 μM total). CoS-MAGE recovery cultures were regrown for 5 or more hours at 30 °C, so as to eliminate polyclonal colonies.

Analyzing Recombination

CoS-MAGE recovery cultures were plated on media selective for the co-selected resistance marker (LB Lennox agar plates with 50 $\mu\text{g/mL}$ carbenicillin for *bla*, 20 $\mu\text{g/mL}$ chloramphenicol for *cat*, or 0.005% SDS for *tolC*, with 20 $\mu\text{g/mL}$ chloramphenicol added to enhance the robustness of selection). Targeted loci in the resulting clones were screened by mascPCR (as described in Chapter 3) using Kapa 2GFast Multiplex PCR ReadyMix with 10X Kapa dye. Monoclonal colonies were grown to stationary phase under proper antibiotic selection, and PCR template was prepared by diluting 2 μL of culture into 100 μL of H_2O . These mascPCR reactions had a total volume of 10 μL , with 0.2 μM of each primer and 2 μL of template. PCRs were carried out with an initial activation step at 95 °C for 3 min, then cycled 26 times with a denaturation temperature of 95 °C (15 sec), an annealing temperature of 63-67 °C (30 sec; temperature as optimized for a given pair of mascPCR reactions), an extension temperature of 72 °C (1 min), and a final extension at 72 °C for 5 min. All mascPCRs were analyzed on 1.5% agarose/EtBr gels (180 V, 60 min).

In CoS-MAGE experiments, all strains were recombined with all oligo sets at least twice. Replicates were combined to generate a complete data set for each strain's performance with each set of oligos. At least 96 total colonies were genotyped for each strain tested with each recombineering oligo set. Only monoclonal and unambiguous mascPCR results were counted toward the final analysis presented here. Given the large sample sizes tested ($n > 85$), we used parametric one way ANOVA to test for significant variance in the CoS-MAGE performance of the strains (EcNR2, EcNR2.*xseA*⁻, nuc5⁻) for a given oligo set.¹⁸ Subsequently, we used a Student's t-test to make pairwise comparisons, with significance defined as $p < 0.05/n$, where n is the number of pairwise comparisons. Here, $n = 15$, as these data were planned and collected as part of a larger set with 6 different strains (the remainder are discussed in Chapter 5), although only the results for EcNR2, EcNR2.*xseA*⁻, and nuc5⁻ are presented here. As such, significance was defined as $p < 0.003$ for the analyses presented in Figures 4.3 and 4.4. Statistical significance in Figures 4.3 and 4.4 is denoted using a system where * denotes $p < 0.003$, ** denotes $p < 0.001$, and *** denotes $p < 0.0001$. For the experiment in which oligo sets were tested with 0, 2, or 4 PT bonds on both ends, comparisons were made between EcNR2, EcNR2.*xseA*⁻, and nuc5⁻ (the only three tested strains in this experiment) for each of the three differently phosphorothioated oligo sets, separately. Additionally, comparisons were made between each oligo set for each of the three strains, also separately. Thus, 15 pairwise comparisons were performed, and significance thresholds are as above.

For the experiment in which a *lacZ*-inactivating oligo with 7 stop codons (*lacZ*.7.stop) was recombined, recombinants were identified as white colonies on plates containing Fisher ChromoMax IPTG/X-Gal solution, and recombination frequencies (# of white colonies / # of total colonies) were determined for every replicate. For the white colonies, the relevant portion

of the *lacZ* gene was amplified with primers lacZ.seq-f and lacZ.seq-2-r (Table 4.1) at 0.5 μ M, using Kapa HiFi HotStart ReadyMix. PCRs were heat activated at 95 °C for 5 min, then cycled 30 times with a denaturation temperature of 98 °C (20 sec), an annealing temperature of 62 °C (15 sec), and an extension temperature of 72 °C (45 sec). PCRs were brought to 72 °C for 5 min, and then held at 4 °C. Samples were then purified with the Qiagen PCR purification kit and quantitated on a NanoDrop ND1000 spectrophotometer. Purified DNA was submitted to Genewiz for Sanger sequencing (40 ng DNA, with 25 pmol of either lacZ.seq-f or lacZ.seq-2-r primer). Good quality sequence pairs were stitched together using SeqMan (Lasergene DNASTar) and exported as FASTA sequences. These sequences were then analyzed for their genotypes at the loci where the lacZ.7.stop oligo could impart inheritance of mutations. Recombinations and recombination frequency counts were repeated thrice so as to ensure consistency, but sequencing was only performed on one biological replicate. Mean allele conversion metrics were generated from the sequencing data by scoring each mutation locus as 1 for a mutant sequence or 0 for a wild type sequence. We tested for statistically significant variance in mean allele conversions using a parametric one way ANOVA. Subsequently, we used a Student's t-test to make pairwise comparisons, with significance defined as $p < 0.05 / 3$, i.e., $p < 0.01$. Statistical significance in Figure 4.1A is denoted using a system where * denotes $p < 0.01$, ** denotes $p < 0.001$, and *** denotes $p < 0.0001$.

References

1. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).
2. Wang, H.H. et al. Genome-scale promoter engineering by coselection MAGE. *Nat Methods* **9**, 591-3 (2012).
3. Wang, H.H. et al. Multiplexed in vivo His-tagging of enzyme pathways for in vitro single-pot multi-enzyme catalysis. *ACS Synth Biol* **1**, 43-52 (2012).
4. Isaacs, F.J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-53 (2011).
5. Carr, P.A. et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40**, e132 (2012).
6. Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. & Church, G.M. Improving lambda red genome engineering in Escherichia coli via rational removal of endogenous nucleases. *PLoS One* **7**, e44638 (2012).
7. Wang, H.H., Xu, G., Vonner, A.J. & Church, G. Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Res* **39**, 7336-47 (2011).
8. Chase, J.W. & Richardson, C.C. Exonuclease VII of Escherichia coli. Mechanism of action. *J Biol Chem* **249**, 4553-61 (1974).
9. Dutra, B.E., Sutter, V.A., Jr. & Lovett, S.T. RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U S A* **104**, 216-21 (2007).
10. Sawitzke, J.A. et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* **407**, 45-59 (2011).
11. Deutscher, M.P. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. XXIX. Hydrolysis of deoxyribonucleic acid from the 5' terminus by an exonuclease function of deoxyribonucleic acid polymerase. *J Biol Chem* **244**, 3029-37 (1969).

12. Little, J.W. An exonuclease induced by bacteriophage lambda. II. Nature of the enzymatic reaction. *J Biol Chem* **242**, 679-86 (1967).
13. Wang, H.H. & Church, G.M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol* **498**, 409-26 (2011).
14. Stein, C.A., Subasinghe, C., Shinozuka, K. & Cohen, J.S. Physicochemical properties of phosphorothioate oligodeoxynucleotides. *Nucleic Acids Res* **16**, 3209-21 (1988).
15. Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S.T. & Modrich, P. In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci U S A* **98**, 6765-70 (2001).
16. Bzymek, M. & Lovett, S.T. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc Natl Acad Sci U S A* **98**, 8319-25 (2001).
17. DeVito, J.A. Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. *Nucleic Acids Res* **36**, e4 (2008).
18. Jekel, J.F., Katz, D.L., Elmore, J.G. & Wild, D. *Epidemiology, Biostatistics, & Preventative Medicine* (W.B. Saunders, 2011).

Chapter Five

Improving Lambda Red Oligonucleotide Recombination via Primase Modification

This chapter is adapted from the following published paper:

Lajoie, M.J.*, Gregg, C.J.*, Mosberg, J.A.*, Washington, G.C. & Church, G.M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res* **40**, e170 (2012).

*Indicates co-first authorship

Research contributions are as follows:

M. Lajoie came up with the idea for using primase modification to improve MAGE performance. M. Lajoie, J. Mosberg, C. Gregg, and G. Church designed the experiments and interpreted their results. M. Lajoie, C. Gregg, J. Mosberg, and G. Washington performed the experiments. M. Lajoie wrote a majority of the published paper, with sections written by J. Mosberg and C. Gregg, and additional writing and editing contributions from G. Church.

Introduction

As discussed in Chapter 1 and Chapter 4, Multiplex Automatable Genome Engineering (MAGE) is a highly useful tool for *E. coli* diversity generation and genome engineering, and has enabled a variety of novel applications. However, the power of this technique is constrained by the limited number of simultaneous mutations that can be generated in a given recombination cycle. In Chapter 4, we established that MAGE recombination frequency can be improved by inactivating endogenous nucleases which would otherwise degrade the oligonucleotides used for recombineering.¹ Additionally, we introduced the technique of co-selection MAGE (CoS-MAGE), whereby a co-selection oligo is directed to repair a mutated antibiotic resistance gene (or another selectable marker) near the loci targeted by the other recombineering oligonucleotides.² CoS-MAGE thereby enhances the average multiplex allele conversion frequency approximately four-fold by selecting for cells that have high levels of recombination in the desired region of the genome.² Additionally, this approach selects against cells that do not take up oligos during electroporation, as it removes the population that does not revert the selectable allele.

As a result of these features, co-selection significantly augments the power of MAGE for diversity generation and genome engineering; furthermore, it points to additional means by which the process may be improved. The fact that CoS-MAGE is most effective for recombineering oligonucleotides targeted in close proximity to the selectable marker suggests that replication fork position and accessibility have a strong impact on Lambda Red recombination frequency.² Thus, we reasoned that we may be able to further improve recombination frequency by manipulating replication fork dynamics to increase the amount of accessible ssDNA on the lagging strand. To accomplish this, we took advantage of prior work

that showed that Okazaki Fragment (OF) size can be altered by modulating the frequency of OF primer synthesis by DnaG primase.³ This work also established that DnaB helicase is responsible for recruiting DnaG primase to the replicore, where it initiates primer synthesis. Thus, the frequency of primer synthesis can be controlled by altering the strength of the protein-

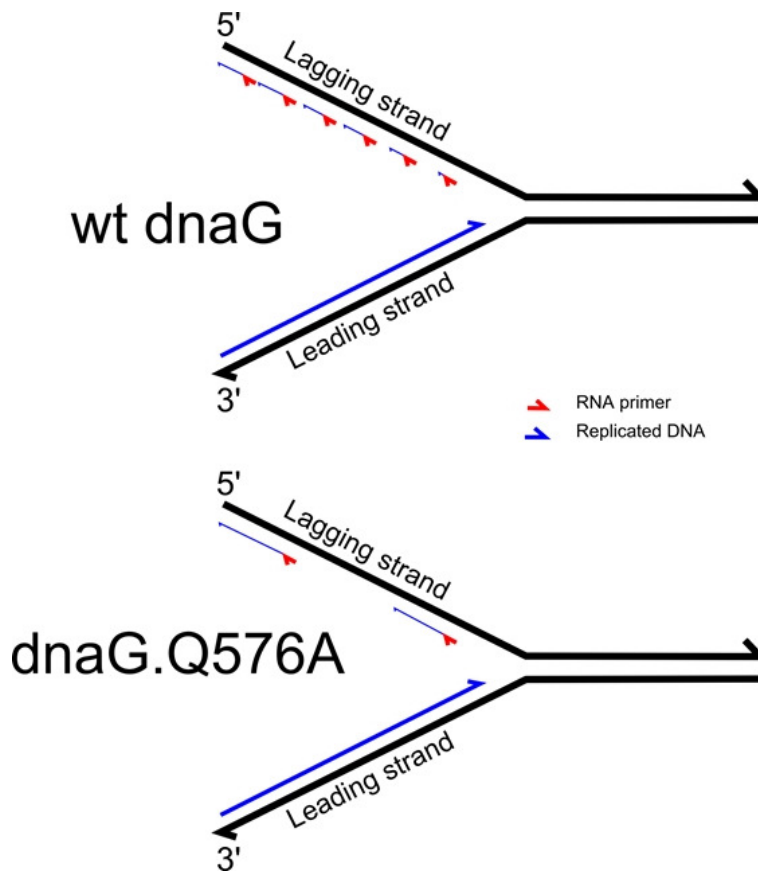


Figure 5.1: Effect of DnaG Attenuation on Replication Fork Dynamics. Mutations to DnaG primase such as Q576A and K580A disrupt the interaction between DnaG and DnaB helicase. This protein-protein interaction recruits DnaG primase to the replication fork and initiates the formation of an Okazaki fragment; thus, weakening the interaction decreases the number of Okazaki fragments formed. This, in turn, increases the average length of Okazaki fragments, as well as the amount of accessible ssDNA between nascent Okazaki fragments. Figure not to scale.

variants, DnaG K580A and Q576A, resulted in *in vitro* OFs that were significantly longer than those initiated by wild type DnaG.⁴ We therefore generated these DnaG mutations in our recombineering strain EcNR2,⁶ and used the resulting strains to explore whether increasing

protein interaction between DnaG primase and DnaB helicase.⁴ We hypothesized that attenuating this interaction would also increase the amount of accessible ssDNA on the lagging strand of the replication fork (as diagramed in Figure 5.1), and possibly enhance multiplex recombination frequencies as a result. To test this, we took advantage of previously reported *E. coli* primase variants with impaired DnaB helicase binding, but normal replication fork rate, priming efficiency, and primer utilization during *in vitro* replication.⁵ These

accessible ssDNA on the lagging strand could improve multiplex oligonucleotide recombination frequency.

In this chapter, we demonstrate that accessible ssDNA on the lagging strand of the replication fork is a limiting factor for multiplex oligonucleotide recombination, and show that weakening the interaction between DnaG primase and DnaB helicase significantly improves MAGE and CoS-MAGE frequencies. We further describe the creation of an optimized strain for CoS-MAGE. This strain builds upon the work presented in Chapter 4, combining approaches to mitigate the nuclease degradation of oligos and to expose more accessible ssDNA on the lagging strand of the replication fork. This strain demonstrates greatly improved CoS-MAGE performance, and provides a foundation for genome engineering projects of a much more ambitious scope.

Impaired Primase Activity Enhances MAGE and CoS-MAGE Frequency

It is generally accepted that the Beta protein from Lambda Red facilitates annealing of exogenous recombinogenic DNA to the lagging strand of the replication fork, prior to incorporation as an Okazaki Fragment.⁷⁻¹⁰ Therefore, we sought to increase the amount of

Table 5.1: Estimation of EcNR2.dnaG.K580A and EcNR2.dnaG.Q576A Okazaki Fragment Lengths

[Primase] (nM)	WT DnaG OF Size (kb) ^a	K580A OF Size (kb)	Q576A OF Size (kb)
80	2.5	5	23
160	1.5	2.5	18
320	1	1	8
640	0.8	N/A	3
Average (fold vs. WT)	1	1.6	8.2

^a All data in this table are from Tougu et al.⁴

accessible ssDNA on the lagging strand by disrupting the ability of DnaG primase to initiate OFs. Prior work,⁴ as described above, has indicated that the DnaG K580A and Q576A

mutations confer this effect, hindering OF formation and increasing OF lengths *in vitro*; Table 5.1 shows the *in vitro* OF lengths previously reported for these variants at a number of different primase concentrations. For each tested primase concentration, we compared both variants' OF size to that of wild type DnaG, and then determined the average fold difference. We thereby estimated that the K580A and Q576A mutations increase OF length by approximately 1.6-fold and 8.2-fold, respectively (Table 5.1). Given the reported ~1.5-2 kb OF lengths of wild type *E. coli* cells grown in rich media,¹¹⁻¹³ we therefore extrapolated that the K580A and Q576A variants have *in vivo* OF lengths of roughly 2.3-3.1 kb and 12-16 kb, respectively. However, these approximations may be imperfect since Tougu et al.⁴ performed their analysis solely *in vitro*. Other conditions and/or host factors not accounted for *in vitro* may affect priming efficiency, thereby rendering these calculations inaccurate.

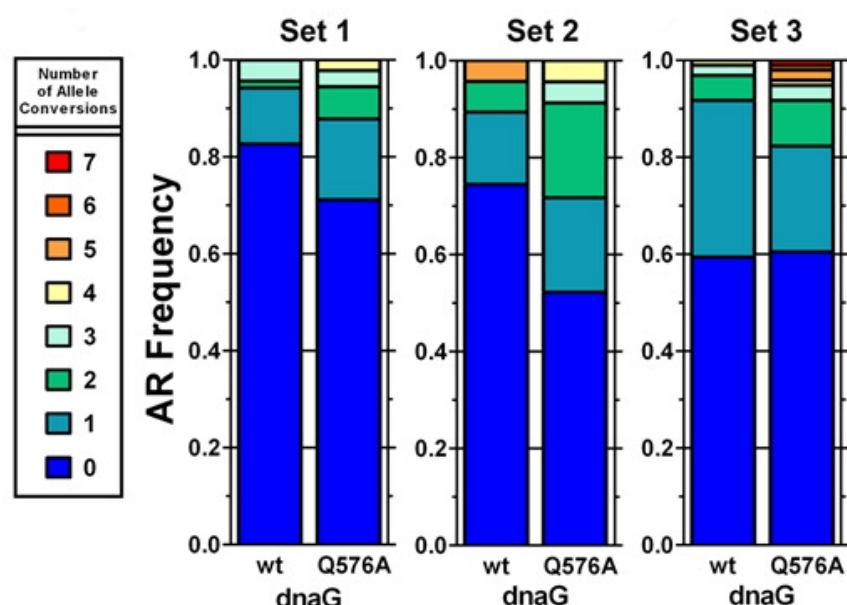


Figure 5.2: DnaG Q576A Mutation Improves MAGE Performance. EcNR2 (“wt”) and EcNR2.dnaG.Q576A (“Q576A”) were assessed for their MAGE performance using the three sets of oligos indicated in Figure 4.2, without co-selection. Data are presented using stacked allele replacement (AR) frequency plots, which show the distribution of clones exhibiting a given number of allele conversions. For all three oligo sets, EcNR2.dnaG.Q576A displayed a greater proportion of clones with 2 or more allele conversions. For oligo sets 1 and 2, EcNR2.dnaG.Q576A also had fewer clones with 0 allele conversions. For EcNR2, n = 69, 47, and 96 for Sets 1, 2, and 3, respectively. For EcNR2.dnaG.Q576A, n = 90, 46, and 96.

Nevertheless, to investigate whether longer OFs could improve oligonucleotide recombination frequency, we first compared the MAGE performance of EcNR2 and EcNR2.dnaG.Q576A, the more attenuated of the two primase variants. Three different sets of

recombineering oligos (the same as used in Chapter 4, but without co-selection) were tested, in order to control for potential oligo-, allele-, region-, and replichore-specific effects.¹⁴ Results are shown in Figure 5.2; for all three tested oligo sets, EcNR2.dnaG.Q576A exhibited a mean allele replacement (AR) frequency that was slightly higher than that of EcNR2.

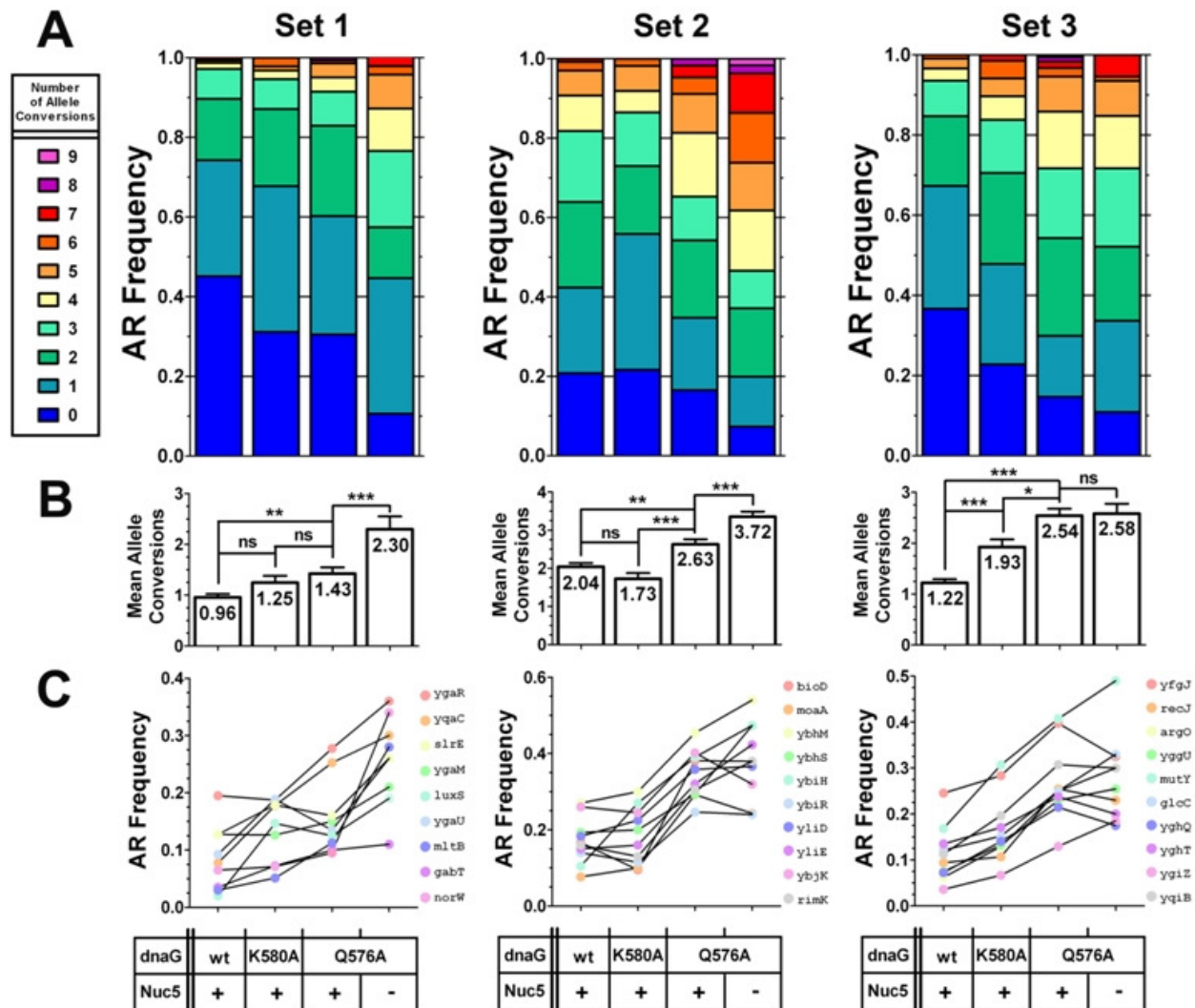


Figure 5.3: DnaG Mutations Improve CoS-MAGE Performance. EcNR2, EcNR2.dnaG.K580A, EcNR2.dnaG.Q576A, and nuc5⁻.dnaG.Q576A were tested for their CoS-MAGE performance with oligo Sets 1-3, co-selected with *cat*, *bla*, and *tolC*, respectively. **A)** Allele replacement data are presented using stacked frequency plots. **B)** The mean number of alleles converted for each strain is shown with p-values indicated above the bars; “ns” denotes non-significant variation, * denotes $p < 0.003$, ** denotes $p < 0.001$, and *** denotes $p < 0.0001$. Data are presented as the mean allele replacement frequency $\pm$ the standard error of the mean. **C)** Allele replacement frequencies for each individual allele are shown for all tested strains. For Set 1, $n = 319, 93, 141$, and 47 for EcNR2, EcNR2.dnaG.K580A, EcNR2.dnaG.Q576A, and nuc5⁻.dnaG.Q576A, respectively. For Set 2, $n = 269, 111, 236$, and 191 . For Set 3, $n = 327, 136, 184$, and 92 .

Encouraged by these results, we next used CoS-MAGE² to try to augment the observed effects. In this experiment, each of the three oligo sets was paired with a co-selection oligo which restored the function of a nearby mutated selectable marker (*cat* for Set 1, *bla* for Set 2, and *tolC* for Set 3). Here, we compared the performance of EcNR2, EcNR2.dnaG.K580A, and EcNR2.dnaG.Q576A. We also generated and tested the nuc5⁻.dnaG.Q576A strain, in order to determine whether additive beneficial effects could be attained by combining primase modification with nuclease removal (i.e., the nuc5⁻ strain described in Chapter 4). Results are shown in Figure 5.3.

Table 5.2: CoS-MAGE Performance versus EcNR2

Metric	Set	nuc5 ⁻	EcNR2.dnaG.Q576A	nuc5 ⁻ .dnaG.Q576A
Average	1	1.65 ^a	1.49	2.40
	2	1.41	1.29	1.82
	3	1.32	2.08	2.12
	Average	1.46	1.62	2.11
5+ Conversions	1	5.28	3.96	10.18
	2	2.65	2.01	4.11
	3	1.07	4.20	4.52
	Average	3.00	3.39	6.27
0 Conversions	1	0.67	0.68	0.24
	2	0.58	0.79	0.35
	3	0.71	0.40	0.30
	Average	0.65	0.62	0.29

^a Fold change was calculated as (strain performance)/(EcNR2 performance), where performance refers to the average number of allele conversions per clone, or the fraction of clones with 5+ or 0 conversions. Nuc5⁻ data are as described in Chapter 4.

EcNR2.dnaG.K580A demonstrated an intermediate phenotype between EcNR2 (shortest OF length) and EcNR2.dnaG.Q576A (longest OF length), and was only significantly more

For all three oligo sets, EcNR2.dnaG.Q576A robustly outperformed EcNR2, yielding a significantly increased mean number of alleles converted (**p = 0.0003, **p = 0.0003, and ***p < 0.0001, respectively, for Sets 1-3). On average, EcNR2.dnaG.Q576A yielded 62% more alleles converted per clone, 239% more clones with five or more allele conversions, and 38% fewer clones with zero allele conversions (Table 5.2).

recombinogenic than EcNR2 for one of the three tested oligo sets (Set 3; *** $p < 0.0001$). Thus, there appears to be a direct correlation between the degree of primase attenuation and the resulting recombination frequency. This result supports our hypothesis that exposing more ssDNA at the lagging strand of the replication fork enhances Lambda Red recombination frequency. Visualizing AR frequency for individual alleles in all three sets (Figure 5.3C) further reinforces the relationship between OF size and CoS-MAGE performance; compared to EcNR2, the K580A variant generally exhibits a modest increase in individual AR frequency, whereas the Q576A variant exhibits dramatically improved AR frequency.

Table 5.3: Summary of Mean Number of Alleles Converted per Clone

Set	EcNR2	nuc5 ⁻	EcNR2.dnaG.Q576A	nuc5 ⁻ .dnaG.Q576A
	Mean $\pm$ SEM (n)	Mean $\pm$ SEM (n)	Mean $\pm$ SEM (n)	Mean $\pm$ SEM (n)
1	0.96 $\pm$ 0.07 (319)	1.58 $\pm$ 0.10 (257)	1.43 $\pm$ 0.12 (141)	2.30 $\pm$ 0.25 (92)
2	2.04 $\pm$ 0.10 (269)	2.89 $\pm$ 0.19 (142)	2.63 $\pm$ 0.13 (236)	3.72 $\pm$ 0.17 (191)
3	1.22 $\pm$ 0.07 (327)	1.61 $\pm$ 0.12 (139)	2.54 $\pm$ 0.14 (184)	2.59 $\pm$ 0.19 (92)

Interestingly, as shown in Table 5.3, EcNR2.dnaG.Q576A strongly outperformed nuc5⁻ for oligo Set 3 (*** $p < 0.0001$), but had slightly lower AR frequency than nuc5⁻ for Sets 1 and 2 ($p = 0.33$ and 0.26 , respectively). This suggests that the relative importance of oligo protection and the availability of accessible lagging strand ssDNA can vary, possibly due to oligo- and/or locus-specific effects that have not yet been elucidated. Since both factors are clearly important, combining impaired primase mutants with nuclease knockouts would be expected to improve recombineering performance. Indeed, the nuc5⁻.dnaG.Q576A strain exhibited markedly better multiplex recombination frequency than either the EcNR2.dnaG.Q576A strain or the nuc5⁻ strain, and yielded 111% more alleles converted per clone, 527% more clones with five or more

allele conversions, and 71% fewer clones with zero allele conversions in comparison with EcNR2 (Table 5.2). This demonstrates that primase modification and nuclease removal have additive beneficial effects on recombination frequency, and confirms that the number of oligonucleotides within the cell and the amount of accessible ssDNA at the lagging strand are both limiting factors for Lambda Red recombination.

Targeting Oligonucleotides to a Single Putative Okazaki Fragment

Given the significant enhancement of CoS-MAGE performance in EcNR2.dnaG.Q576A, we next sought to determine whether localizing all 10 targeted alleles to a single putative OF would result in frequent "jackpot" recombinants with nearly all 10 alleles converted. We hypothesized that nascent Okazaki Fragments sometimes obstruct target alleles, leading to a non-accessible lagging strand. According to this hypothesis, successful replacement of one allele would indicate permissive OF localization, greatly increasing the chance that other alleles occurring within the same OF could be replaced. Therefore, we speculated that the weakened primase-helicase interaction in EcNR2.dnaG.Q576A might allow many changes to occur within 12-16 kb, the predicted OF size in this strain. To test this, we designed 10 MAGE oligos that introduce premature stop codons into an 1829 bp region of *lacZ*. Despite their close proximity, all 10 alleles were spaced far enough apart so that their corresponding MAGE oligos would not overlap. Given the difference in average OF sizes between strains, we predict it to be fairly unlikely for all 10 alleles to be located within the same OF in EcNR2, but quite likely that all 10 alleles would be located within the same OF in EcNR2.dnaG.Q576A.

To carry out this experiment, a *tolC* cassette was installed ~50 kb upstream of *lacZ*, and then inactivated to facilitate co-selection. Since the *lacZ* mutations were too densely clustered to

enable mascPCR¹⁵ analysis, we used Sanger sequencing to analyze allele inheritance in white (*LacZ*⁻) colonies, while blue colonies were scored as having zero conferred mutations. Results for this experiment are shown in Figure 5.4.

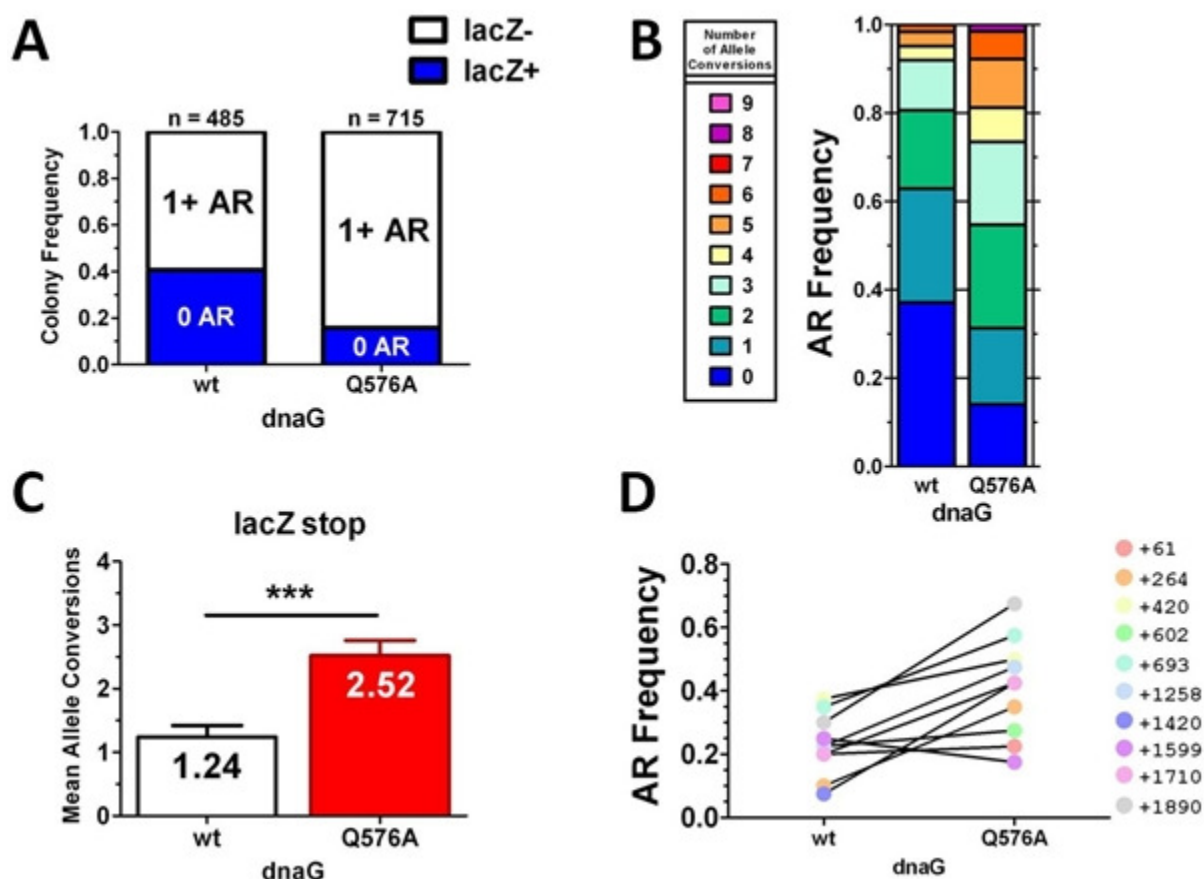


Figure 5.4: Enhanced Recombination Frequencies are not Achieved by Targeting a Single Putative Okazaki Fragment. EcNR2 and EcNR2.dnaG.Q576A were tested for their CoS-MAGE performance with 10 non-overlapping oligos that each introduce a premature stop codon in the first 1,890 bp of *lacZ* – a distance predicted to be within a single OF for EcNR2.dnaG.Q576A. **A**) EcNR2.dnaG.Q576A (5.33:1) exhibited a substantial increase in *lacZ*⁻:*lacZ*⁺ ratio compared to EcNR2 (1.46:1), consistent with the improved allele replacement frequency previously observed for this strain. **B**) EcNR2 and EcNR2.dnaG.Q576A both exhibited similar AR distributions with this oligo set compared with those observed for Sets 1-3 (which span 70 kb, 85 kb, and 162 kb, respectively). **C**) Compared to EcNR2, EcNR2.dnaG.Q576A exhibited a higher mean number of alleles converted (***p < 0.0001). For EcNR2, n = 39; for EcNR2.dnaG.Q576A, n = 55. **D**) Compared to EcNR2, AR frequencies increased for 9 out of 10 individual alleles in EcNR2.dnaG.Q576A. Taken together, these results show improved CoS-MAGE performance in EcNR2.dnaG.Q576A, but not to a significantly greater extent than observed with the other oligo sets. Thus, no significant additional enhancement is attained by targeting all oligos to a single putative OF.

For EcNR2, 59% of the clones were white, with 1.24 ± 0.23 (mean $\pm$ standard error of the mean) conversions per clone; in contrast, 84% of the EcNR2.dnaG.Q576A clones were white,

with 2.52 ± 0.25 allele conversions per clone (Figure 5.4A,C). While EcNR2.dnaG.Q576A exhibited more mean allele conversions than EcNR2 ($***p < 0.0001$), the magnitude of this improvement was not notably greater than those observed for Sets 1-3, where recombineering oligos were spread across 70, 85, and 162 kb, respectively. Moreover, “jackpot” clones with 7+ converted alleles were not frequently observed for EcNR2.dnaG.Q576A using this oligo set. Thus, multiplex AR frequencies cannot markedly be enhanced by targeting oligos to a single putative OF. This may be due to limited oligo entry, target sites being occluded by host proteins, OF extension occurring more quickly than oligo annealing, or multiple synthetic oligonucleotides destabilizing a nascent OF.

Testing a Larger Oligonucleotide Pool Size

A MAGE pool size of approximately 10 distinct oligos was found to be most effective in prior studies.¹⁴ However, given the improved performance of the strains discussed in this chapter, we next tested whether using a larger pool of oligos could lead to more alleles being converted in average and top clones. Thus, we designed an expansion set of ten oligonucleotides (Set 3X) to be recombined along with Set 3 and *tolC* co-selection. CoS-MAGE was then performed using this expanded 20-plex set, and results are shown in Figure 5.5.

For all strains, the new set of 20 oligonucleotides yielded more alleles converted per average clone than the 10-plex Set 3 alone. The fold increase in average number of alleles converted was 1.74 for nuc5⁻.dnaG.Q576A, which was greater than the fold increases for EcNR2, EcNR2.dnaG.K580A, and EcNR2.dnaG.Q576A (1.35, 1.02, and 1.17, respectively). The markedly beneficial effect of nuclease removal on the performance of this 20-plex set supports that a limited number of oligonucleotides are internalized during electroporation, as

suggested in Chapter 4. This results in low intracellular concentrations of each individual oligonucleotide in a large set, therefore augmenting the beneficial impact of nuclease protection. Notably, in this experiment, *nuc5⁻.dnaG.Q576A* yielded the greatest number of simultaneous allele conversions observed to date (12, plus *tolC* reversion). However, given that this is substantially less than the total possible number of conversions (20, plus *tolC* reversion), this suggests that further increases in oligo pool size may not markedly increase the mean or maximum number of alleles converted.

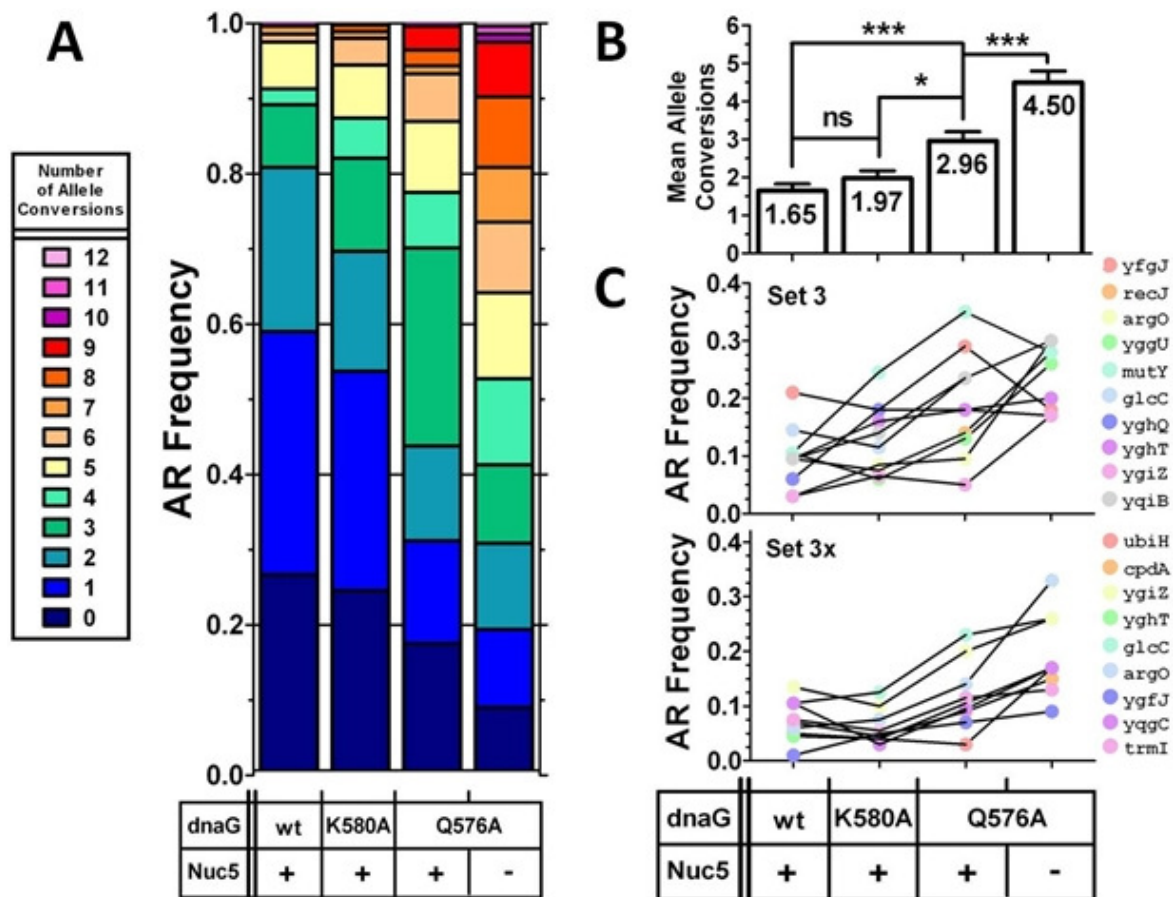


Figure 5.5: Testing DnaG Variants with an Expanded CoS-MAGE Oligo Set. EcNR2, EcNR2.dnaG.K580A, EcNR2.dnaG.Q576A, and *nuc5⁻.dnaG.Q576A* (n = 96, 113, 95, and 96, respectively) were tested for their CoS-MAGE performance using an expanded set of 20 oligos (Set 3+3X, with *tolC* co-selection). **A**) AR frequency distributions. **B**) Mean number of alleles converted $\pm$ standard error of the mean, with p-values indicated above the bars; “ns” denotes a non-significant variation, * denotes $p < 0.003$, ** denotes $p < 0.001$, and *** denotes $p < 0.0001$. **C**) Individual AR frequencies; allele *ygfT* from Set 3X could not be assayed by mascPCR. Overall, *nuc5⁻.dnaG.Q576A* had strongly improved performance with the 20-plex set (compared with Set 3 alone), while the other strains did not. This suggests that nuclease removal allows larger oligo sets to be used more effectively.

Disrupting DnaG Primase Activity Enhances Deletions but not Insertions

MAGE is most effective at introducing very short mismatches, insertions, and deletions, as these can efficiently be generated using Red-mediated recombination without direct selection.⁶ However, large deletions and gene-sized insertions are also important classes of mutations that could increase the scope of applications for MAGE. Thus, we investigated whether decreasing DnaG primase function could also enhance recombination frequency for large deletions and/or insertions.

Based on the accepted model for Lambda Red recombination,^{7,8} we expected enhanced deletion frequency in EcNR2.dnaG.Q576A, given the increased amount of accessible lagging

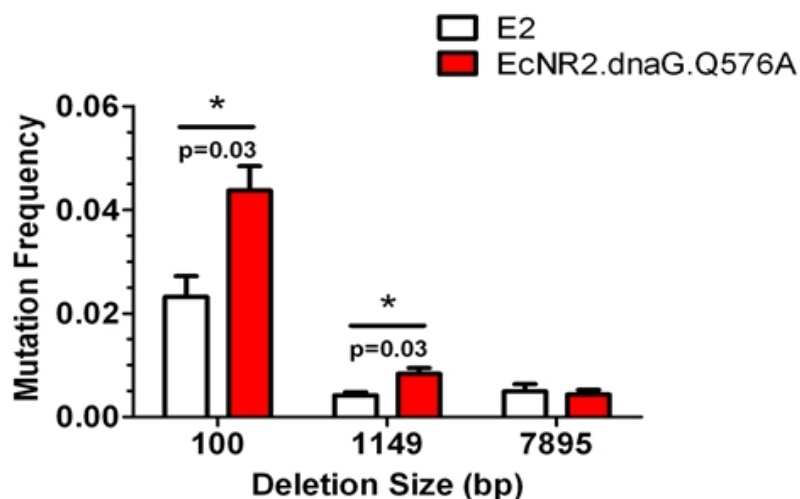


Figure 5.6: Effect of DnaG Modification on Oligo-mediated Deletion Frequency. Oligonucleotides were designed to delete regions of various sizes within and around the *galK* gene. These oligos were recombined into EcNR2 (“E2”; white) and EcNR2.dnaG.Q576A (red). Three independent replicates were performed for each oligo; error bars are shown as the standard error of the mean. EcNR2.dnaG.Q576A demonstrates a significant advantage over EcNR2 for the 100 bp and 1149 bp deletions, but not for the 7895 bp deletion.

strand ssDNA to which the deletion oligos could anneal. We further predicted that this effect might be particularly pronounced for intermediate-sized deletions (e.g., 1 – 10 kb), since less frequent priming would also increase the probability of both homology targets for such oligos being located within the same OF. In contrast, the two

annealing events required for larger deletions may span multiple OFs. Thus, to test the effect of OF size on oligo-mediated deletion frequency, we designed three oligos that deleted 100 bp,

1149 bp, or 7895 bp of the genome, including a portion of *galK*. The oligo that deleted 7895 bp also removed several nonessential genes in addition to *galK*. Recombined populations were screened for the GalK⁻ phenotype (white colonies) on MacConkey agar plates supplemented with galactose, and recombination frequencies were thereby determined. As shown in Figure 5.6, EcNR2.dnaG.Q576A significantly outperformed EcNR2 for the 100 bp (*p = 0.03) and 1149 bp (*p = 0.03) deletions, but there was no difference detected between the two strains for the 7895 bp deletion (p = 0.74). This lack of a primase-mediated improvement for the 7895 bp deletion oligo may be due to the two homology sites frequently being split across multiple OFs even in EcNR2.dnaG.Q576A.

To determine whether primase modification also results in improved insertion frequencies, we quantified the recombination frequency of a kanamycin resistance insertion cassette targeted to *lacZ* (*lacZ::kanR*,⁸ ~1.2 kb). Recombination frequencies of $1.81\text{E-}04 \pm 6.24\text{E-}05$ and $1.28\text{E-}04 \pm 4.52\text{E-}05$, respectively, were observed for EcNR2 and EcNR2.dnaG.Q576A (n = 3, p = 0.30 by unpaired t-test). Thus, primase modification does not improve insertion frequency, suggesting that gene insertion is constrained by factors other than the accessibility of ssDNA at the replication fork. Such factors may include cassette entry into cells, a low propensity for the heterologous portion of the cassette to loop out and allow the flanking homologies to anneal to their adjacent targets,^{7,8} or a limited probability that Lambda Exo degrades the entire leading-targeting strand before degradation of the lagging-targeting strand occurs.⁷

Disrupting DnaG Primase Activity Enhances Leading-targeting CoS-MAGE

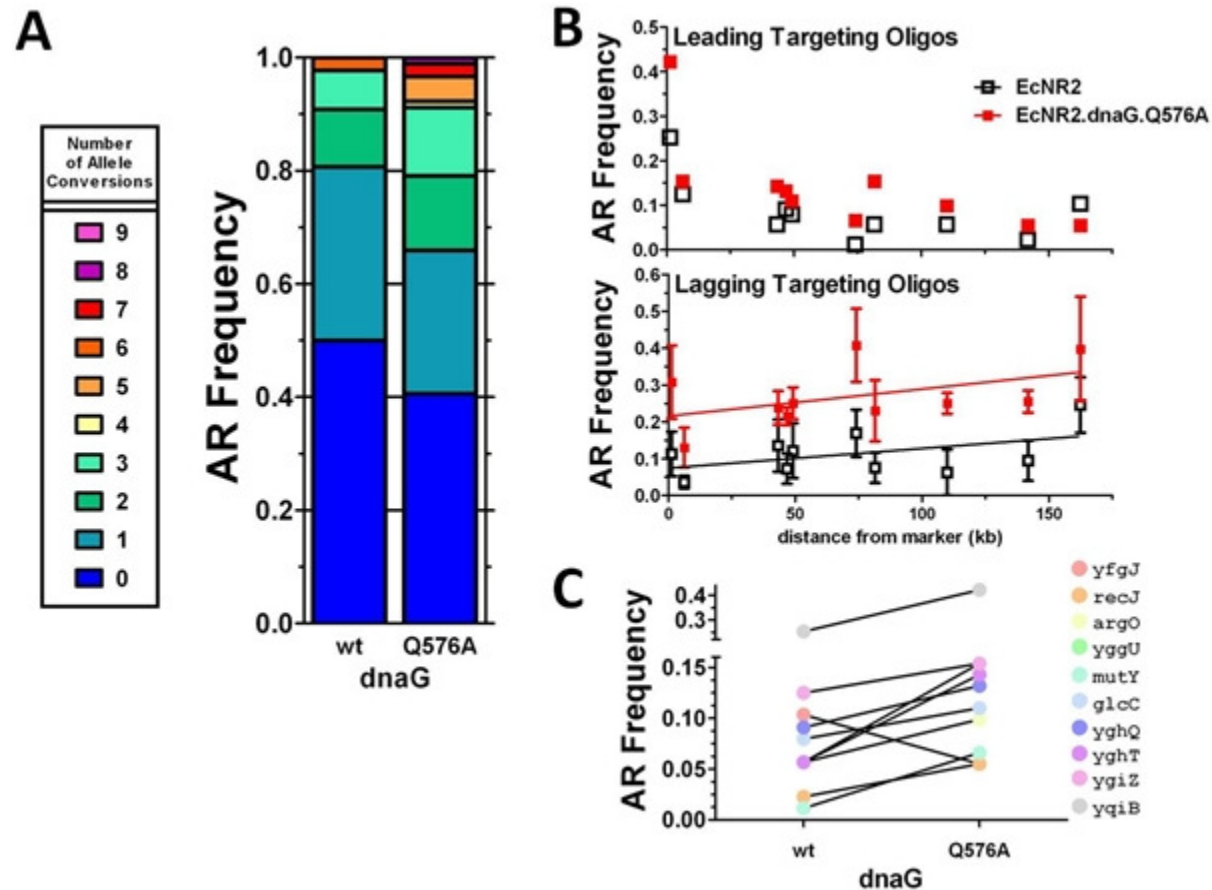


Figure 5.7: Effect of Primase Modification on Leading-targeting CoS-MAGE. **A)** EcNR2.dnaG.Q576A ($n = 91$) outperformed EcNR2 ($n = 88$) in leading-targeting Set 3 CoS-MAGE, with fewer clones exhibiting zero allele conversions, a greater proportion of clones exhibiting two or more conversions, and a greater average number of alleles converted per clone. **B)** For leading-targeting Set 3 oligos, recombination frequency decayed rapidly with increasing distance from the selectable marker (top panel). In contrast, the corresponding set of lagging-targeting oligos (bottom panel) demonstrated robust co-selection; linear regression analyses (solid trendlines) show that for both strains, recombination frequency did not decrease with distance over this 162 kb genomic region. **C)** Individual AR frequencies. AR frequency was improved for 9/10 alleles in EcNR2.dnaG.Q576A. Please note the broken y-axis, reinforcing that the recombination frequency of the most proximal allele to the selectable marker was much higher than those of the other alleles.

Finally, we tested whether primase modification results in improved recombination frequency for oligos that target the leading strand. To examine this, we used the reverse complements of the Set 3 oligos; the *tolC*-reverting co-selection oligo was also redesigned to target the leading strand. We anticipated that leading strand recombination frequencies would not significantly be impacted by primase modification, as DnaG predominantly acts on the

lagging strand. Surprisingly, however, we found (Figure 5.7) that the leading-targeting oligos yielded significantly more average allele conversions for EcNR2.dnaG.Q576A (1.39 ± 0.18) than for EcNR2 (0.85 ± 0.13 ; $*p = 0.018$). Similar to lagging-targeting Set 3, we also observed a reduction in zero conversion events for EcNR2.dnaG.Q576A, as well as a greater maximum number of alleles converted. Also contrary to our expectations, EcNR2.dnaG.Q576A exhibited greater AR frequency than EcNR2 at 9 out of 10 alleles targeted on the leading strand (Figure 5.7C). Interestingly, we further observed that for leading-targeting oligos, recombination frequency diminished quickly with increasing distance from the co-selected marker (Figure 5.7B, top panel). The co-selection advantage for lagging-targeting oligos, in contrast, typically persists over a large genomic distance (~ 0.5 Mb).² Indeed, this distance-robust co-selection advantage is clearly demonstrated by the lagging-targeting Set 3 oligos (Figure 5.7B, bottom panel). We discuss possible explanations for these surprising results in the section below.

Discussion

In order to increase the amount of accessible ssDNA on the lagging strand of the replication fork, we separately introduced two mutations into *E. coli* DnaG primase: K580A and Q576A. These mutations have been shown *in vitro* to increase OF size by interrupting the primase-helicase interaction on the replisome.⁴ Based on the measurements of Tougu *et al.*,⁴ we estimate that the K580A mutation increases OF length by about 1.6-fold, and the Q576A mutation increases OF length by about 8.2-fold (Table 5.1). Strains EcNR2.dnaG.K580A and EcNR2.dnaG.Q576A exhibited significant increases in the mean number of alleles converted during CoS-MAGE, as well as decreases in the proportion of clones with zero non-selectable alleles converted. Furthermore, the strongest recombination frequency enhancement was

observed in EcNR2.dnaG.Q576A (the variant with the longest OFs of the strains tested in this chapter), with an intermediate enhancement observed in EcNR2.dnaG.K580A (the variant with intermediate-sized OFs). This relationship between recombination frequency and OF length further supports the model in which Beta mediates annealing at the lagging strand of the replication fork,⁷⁻¹⁰ as well as our hypothesis that accessible ssDNA on the lagging strand is a limiting factor for this process. With this in mind, we unsuccessfully attempted to generate a DnaG Q576A/K580A double mutant; this failure suggests that such extensive manipulation of the DnaG C-terminal helicase interaction domain¹⁶ is lethal.

The results described here clearly establish that the intracellular concentration of MAGE oligos and the accessibility of their genomic targets are both limiting factors in multiplex oligonucleotide recombination. This is evidenced by the fact that the nuclease-depleted strains discussed in Chapter 4 and the primase-modified strains described in this chapter both displayed robust CoS-MAGE recombination frequency enhancements. Interestingly, the *nuc5⁻* strain¹ slightly outperformed the EcNR2.dnaG.Q576A strain for oligo Sets 1 and 2, while EcNR2.dnaG.Q576A strongly outperformed the *nuc5⁻* strain for Set 3 (Table 5.3). While oligo design parameters such as the type of designed mutation,⁶ oligo length,⁶ oligo secondary structure,⁶ and the amount of off-target genomic homology¹⁴ are major determinants of recombination frequency, these results also highlight the relevance of genomic context. For instance, one possible explanation for the variable effect of primase modification is that different genomic regions may have different replication fork speed or priming efficiency. These factors could locally modulate OF length (although replication fork speed did not appear to be a major determinant of OF length *in vitro*),⁴ thereby affecting the impact of further OF modulation on recombination frequency. Alternatively, certain oligos may be more susceptible to nuclease

degradation; therefore, removing the responsible nucleases would disproportionately improve recombination frequency for sets containing several of such oligos. With this in mind, we tested whether combining primase modification and nuclease removal would enhance CoS-MAGE performance more than either strategy used individually. Indeed, *nuc5⁻.dnaG.Q576A* consistently performed the best of all tested strains (Figures 5.3 & 5.5). Therefore, these two disparate strategies can be combined to yield a larger and more robust recombination frequency enhancement, reinforcing that both oligo availability and lagging strand ssDNA accessibility are limiting factors in MAGE and CoS-MAGE.

To explore the extent to which OF localization impacts CoS-MAGE performance, we tested whether targeting 10 oligos to a single putative OF would yield a bimodal AR frequency distribution, with subpopulations of little-modified (few or no alleles converted) and "jackpot" (most or all alleles converted) recombinants. However, CoS-MAGE in *EcNR2.dnaG.Q576A* using the densely-clustered *lacZ* oligos (Figure 5.4) produced a similar AR distribution to the ones observed for Sets 1-3 (Figure 5.3), which targeted regions of the genome spanning several putative OFs. Since mutations targeted to a single putative OF yielded an inheritance pattern similar to those observed when mutations were spread across many OFs, nascent OF localization does not appear to be a critical determinant of multiplex oligonucleotide recombination frequency. A number of hypotheses could explain why the expected "jackpots" are not observed. For one, it is likely that MAGE oligos are limiting due to degradation and/or lack of uptake. Thus, it is possible that most cells lack the oligos necessary for generating a majority of the desired mutations. Additionally, OF extension may occur too fast for all of the MAGE oligos to anneal before the nascent OF occludes their targets. Still another explanation could be that host factors such as ssDNA binding proteins occupy portions of the lagging strand, rendering these

regions inaccessible for Beta-mediated annealing. Finally, it is also possible that several MAGE oligos annealing within a single OF could destabilize lagging strand synthesis, leading to selection against highly-modified "jackpot" clones. Indeed, $\text{polIII}_{\text{lag}}$ dissociates from the replisome after completing an OF,¹³ and the repeated dissociation of $\text{polIII}_{\text{lag}}$ due to multiple nearby MAGE oligos could cause the replisome to proceed beyond the target region before lagging strand synthesis is completed. In the absence of the rest of the replisome, a cytosolic PolIII holoenzyme demonstrates considerably diminished activity.¹⁷ Therefore, if OFs are not completed while the replisome is in close proximity, this could result in persisting ssDNA lesions that lethally destabilize the chromosome.

We also investigated whether targeting a greater number of alleles would increase the resulting number of conversions in our enhanced strains (Figure 5.5). Although nuc5^- .dnaG.Q576A demonstrated 1.74-fold greater average allele conversion frequency for 20-plex Set 3+3X versus 10-plex Set 3, a difference of only 1.17-fold was observed for EcNR2.dnaG.Q576A. The superior enhancement for the nuclease-depleted nuc5^- .dnaG.Q576A strain suggests that intracellular oligo concentration is a particularly limiting factor for highly multiplexed MAGE (i.e., more than 10 alleles targeted). Therefore, enhancing DNA uptake and/or preservation may be a fruitful means of further improving MAGE. However, this improved multiplexability of nuc5^- .dnaG.Q576A could also be due to the Set 3X oligos being more responsive to decreased exonuclease degradation than to increased lagging strand ssDNA availability. Additionally, there may be other limiting factors such as insufficient Beta or unidentified host proteins with a role in Lambda Red recombination. Although there is no known precedent for the quantity of Lambda Red proteins being a limiting factor for recombination,¹⁸ our novel ability to attain 12 simultaneous non-selectable allele replacements

indicates that our improved strains are in uncharted territory for probing the limits of Lambda Red recombination.

Given that DnaG primase acts predominantly on the lagging strand of the replication fork, we expected that the tested primase modifications would enhance only lagging-targeting recombination. Therefore, the performance of leading-targeting CoS-MAGE in our strains was surprising, as EcNR2.dnaG.Q576A significantly outperformed EcNR2 (*p = 0.018). Furthermore, while the total number of co-selected *tolC*⁺ recombinants was considerably (~100-fold) lower for leading-targeting CoS-MAGE versus lagging-targeting, the AR frequency of non-selectable alleles in these recombinants was still quite impressive, especially for alleles in close proximity to the selectable marker. This suggests that one leading strand recombination event correlates strongly with additional recombination events. Two possible explanations for the superior performance of EcNR2.dnaG.Q576A in leading-targeting CoS-MAGE are that 1) an impaired primase-helicase interaction increases accessible leading strand ssDNA as well as lagging strand ssDNA, or 2) infrequent Beta-mediated strand invasion initiates a new replication fork that travels in the opposite direction and swaps which strands are leading and lagging.

There is some support for primase function affecting the dynamics of replication on both the leading and lagging strands.^{13,17,19} Lia *et al.*¹³ previously observed phases of replication in which OF synthesis was faster than helicase progression at the replication fork, alternating with phases in which the opposite was true. These results demonstrate that DnaB-PolIII_{lead} does not always progress at the same speed as PolIII_{lag}.¹³ Furthermore, Yao *et al.*¹⁹ showed that the velocity of leading-strand synthesis decreases during lagging strand synthesis, while its processivity increases. Thus, as a result of its effect on lagging strand synthesis, altering primase-helicase binding could bring about periods of transiently increased fork rate and

decreased PolIII_{lead} processivity. Given that PolIII tends to release from the replication fork more readily than DnaB helicase,¹⁹ this could exacerbate this effect, leading to helicase progression durably outstripping leading strand synthesis. This would increase the amount of accessible ssDNA on the leading strand, and likely improve leading-targeting CoS-MAGE frequencies.

Alternatively, the Beta protein from Lambda Red has been reported to facilitate strand invasion *in vitro*.²⁰ Although there is little evidence to suggest that such strand invasion also occurs *in vivo*, if it does, it could produce a D-Loop that acts as a new origin of replication.²¹ Thus, invasion of one leading-targeting MAGE oligo could initiate a new replication fork traveling in the opposite direction. In the reverse orientation, the leading strand would become the lagging strand, and oligos would thereby become lagging-targeting and much more likely to recombine. If this is the case, the non-selectable alleles would then be upstream of the co-selection marker. Since co-selection is most effective downstream of the selectable marker,² this may explain why co-selection enhancements were observed to decay rapidly with distance on the leading strand (Figure 5.7B).

In this chapter, we have identified accessible lagging strand ssDNA as a limiting factor in multiplex oligonucleotide recombination, and shown that such ssDNA can be increased via manipulation of the primase-helicase interaction. Compared to a standard recombineering strain (EcNR2), primase-modified strain EcNR2.dnaG.Q576A displayed on average 62% more alleles converted per clone, 239% more clones with five or more allele conversions, and 38% fewer clones with zero allele conversions in a given round of CoS-MAGE with ten oligonucleotides (Table 5.2). We used this strategy to build on the advances described in Chapter 4, generating the nuc5⁻.dnaG.Q576A strain, which has increased accessible lagging strand ssDNA and also lacks five potent exonucleases. These modifications exploited two distinct mechanisms that

together increased the robustness and potency of CoS-MAGE, enabling an average of 4.50 and a maximum of 12 non-selectable allele replacements in cells exposed to a pool of 20 different recombineering oligonucleotides. After only a single cycle with this oligo set, 48% of recombinants had five or more allele replacements and just 8% lacked any modified non-selectable alleles. Furthermore, in a given round of CoS-MAGE with ten oligos, nuc5⁺.dnaG.Q576A displayed on average 111% more alleles converted per clone, 527% more clones with five or more allele conversions, and 71% fewer clones with zero allele conversions, in comparison with EcNR2. This improvement in MAGE performance will be highly valuable for increasing the diversity explored during the directed evolution of biosynthetic pathways,⁶ and for enabling the rapid generation of desired genotypes involving tens to hundreds of allele replacements.¹⁴

Experimental

Oligonucleotides used in this Study

A full list of oligonucleotides used in this study is given in Table 5.4. In this table, “wt-f” refers to a forward ascPCR/mascPCR primer used to detect a wild type allele, “mut-f” refers to a forward ascPCR/mascPCR primer used to detect a mutated allele, and “rev” refers to the reverse ascPCR/mascPCR primer used with both forward primers. Sizes of mascPCR bands corresponding to a given locus are denoted in the table as 100, 150, 200, 250, 300, 400, 500, 600, 700, or 850. Asterisks represent phosphorothioate bonds between the indicated nucleotides. Recombination and screening oligonucleotides for Sets 1-3 are given in Table 4.1, and are not reproduced here. All oligonucleotides were ordered from Integrated DNA Technologies with standard purification and desalting.

Table 5.4: Oligonucleotides used in Chapter 5

Name	Use	Sequence
ygfJ_AGR	Set 3X.850 recombineering oligo	C*C*ACTATGTCAGCCATCGACTGTATAATTACCG CTGCCGGATTATCATCAAGGATGGGGCAATGGAA AATGATGTTACCCTGGGAACA*G*G
ygfT_AGR	Set 3X.700 recombineering oligo	G*A*TGCCTTCGTATCAAACAGAGTTAACATATCG CGCGCCGCCTGTCTTCTGCGGCCATTGCAGTGAC AACCAGATCCGCGCCATGAA*C*T
ubiH_AGR	Set 3X.600 recombineering oligo	G*T*GCAGAGTTTGCGCCGATTGCCACCAGCAC GGTACGATGGGTAATAGACCTGGCGGCGTGGGTT AACGCCAGCGGATAAGCACTG*C*G
argO_AGR	Set 3X.500 recombineering oligo	G*G*ATTCAGCCAGGTCACTGCCAACATGGTGGCG ATAATTTTCCACCTGCCTTGCTTCATGACTTCGGC GCTGGCTAACTCAATATTAC*T*G
yqgC_AGR	Set 3X.400 recombineering oligo	G*A*ATCCTGAGAAGCGCCGAGATGGGTATAACA TCGGCAGGTATGCAAAGCAGGGATGCAGAGTGCG GGGAACGAATCTTCACCAGAAC*G*G
trmI_AGR	Set 3X.300 recombineering oligo	T*T*TTTTACGCAGACGACGGCTACGGTCTTTGC CATTATTTCACTCTCTCGAACATTAAGTCCCATAC TCCGTGACCAAGACGATGAC*C*A
glcC_AGR	Set 3X.250 recombineering oligo	A*C*GATCTGCTCGACGTTTCGCGCATTACTGGAGG GCGAATCGGCAAGACTGGCGGCAACGCTGGGAA CGCAGGCTGATTTTGTGTGAT*A*A
yghT_AGR	Set 3X.200 recombineering oligo	G*T*GAACATCTTATTACCGTTGTCGAAAAATATG GTGCTGCCGAAAGGGTTCATTTAGGAAAACAGGC CGGAAATGTCGGTCGTGCAGT*G*A
ygiZ_AGR	Set 3X.150 recombineering oligo	A*A*TACATATACCCAAAACCTCGAACATTTCCCGC ATAAAGAGTTTCTTAAGATAAGAATAATAAGTG GCGTAAGAAGAAAAAATGCTG*C*A
cpdA_AGR	Set 3X.100 recombineering oligo	C*T*TCGTGCTTTTGTGCAAACAGGTGAGTGTCGG TAATTTGTAAAATCCTGACCCTGGCCTCACCAGCC AGAGGAAGGGTTAACAGGCT*T*T
lacZ_KO1	Set lacZ TAG → TAA +61	T*C*ACTGGCCGTCGTTTTACAACGTCGTGACTGG GAAAACCCTTGAGTTACCCAACCTTAATCGCCTTGC AGCACATCCCCCTTTCGCCA*G*C
lacZ_KO2	Set lacZ TAG → TAA +264	G*C*TGGAGTGCGATCTTCTGAGGCCGATACTGT CGTCGTCCCCTCATAATGGCAGATGCACGGTTAC GATGCGCCCATCTACACCAAC*G*T
lacZ_KO3	Set lacZ TAG → TAA +420	C*A*CATTTAATGTTGATGAAAGCTGGCTACAGGA AGGCCAGACGTAAATTATTTTTGATGGCGTTAACT CGGCGTTTCATCTGTGGTGC*A*A
lacZ_KO4	Set lacZ TAG → TAA +602	T*G*ATGGTGCTGCGCTGGAGTGACGGCAGTTATC TGGAAGATCAGTAGATGTGGCGGATGAGCGGCAT TTCCGTGACGTCTCGTTGCT*G*C
lacZ_KO5	Set lacZ TAG → TAA +693	T*A*AACCGACTACACAAATCAGCGATTTCCATGT TGCCACTCGCTAAAATGATGATTTTCAGCCGCGCT GTACTGGAGGCTGAAGTTCAG*A*T
lacZ_KO6	Set lacZ TAG → TAA +1258	T*A*CGGCCTGTATGTGGTGGATGAAGCCAATATT GAAACCCACTGAATGGTGCCAATGAATCGTCTGA CCGATGATCCGCGCTGGCTAC*C*G
lacZ_KO7	Set lacZ TAG → TAA +1420	G*G*GAATGAATCAGGCCACGGCGCTAATCACGA CGCGCTGTATTGATGGATCAAATCTGTGCATCCTT CCCGCCCGGTGCAGTATGAAG*G*C

Table 5.4 (Continued)

lacZ_KO8	Set lacZ TAG → TAA +1599	G*T*CCATCAAAAAAATGGCTTTCGCTACCTGGAGA GACGCGCCCGTAGATCCTTTGCGAATACGCCAC GCGATGGGTAACAGTCTTGGC*G*G
lacZ_KO9	Set lacZ TAG → TAA +1710	G*T*TTTCGTCAGTATCCCCGTTTACAGGGCGGCTT CGTCTGGGACTAAGTGGATCAGTCGCTGATTAAA TATGATGAAAACGGCAACCCG*T*G
lacZ_KO10	Set lacZ TAG → TAA +1890	A*G*CGCTGACGGAAGCAAAACACCAGCAGCAGT TTTTCCAGTTCTGATTATCCGGGCAAAACCATCGAA GTGACCAGCGAATACCTGTTC*C*G
ygfJ_2*:2*_lead	Set 3.850_lead TAG → TAA oligo	G*C*CGGAAGGATTAAATATTTGAACGCAATCGTT TAACTACGTTTTTTTAAATTTTCAGTATACCTTTCCTT CGCTGTAATAAAGTCGTCC*G*G
recJ_2*:2*_lead	Set 3.700_lead TAG → TAA oligo	C*A*ACCGCAGCCTGCAAATTATCATCGACAATAT CTGGCCAATTTAACGTCATCTTCTCTATAAAAAAG AGCGTGGATTGGGTACAATC*C*C
argO_2*:2*_lead	Set 3.600_lead TAG → TAA oligo	G*A*GAGACGGTATTGCTCATGCACAAGCCTTGTT CAGTTAAGCGCTATCTGATGGAAAAATAAAACAG AGGCGCTAAGCTTGCCTCCAG*A*G
yggU_2*:2*_lead	Set 3.500_lead TAG → TAA oligo	T*T*ACCGACATTGCCGTTGCGAGGACAACCTTTT TGCATAGGATACTTAATTAATTAACGCCGCGATT CTGGCGGGATTGTTGCGGA*T*T
mutY_2*:2*_lead	Set 3.400_lead TAG → TAA oligo	A*A*AAATCGTTCTGCTCATAAATCATCCTCTTTAT CGACTCACGCGTTAAACCGGCGCGCCAGTGCGTA ACTGCTGTAACAAACGCTCC*A*C
glcC_2*:2*_lead	Set 3.300_lead TAG → TAA oligo	C*A*AGCGTTACCGATTGTATGAAAAGCAGATTTA ATACCAGTTAACTCAGGTTTCATCTCCAGCGGCAC CGCCGAGCGAATCAAATGGTG*G*C
yghQ_2*:2*_lead	Set 3.250_lead TAG → TAA oligo	G*C*TGGTGTGTTGGCGTTAAGTATCTCGAAGCGTA TGACCTGATTAAAGGAAGGTGCGAATAAGCGGGG AAATTCTTCTCGGCTGACTCA*G*T
yghT_2*:2*_lead	Set 3.200_lead TAG → TAA oligo	C*C*CGGTCCTGGCGATCGGGCATTTCCATTTTTG ATTAGTGATAACCACGCGCGGTCATAAAATCCGT AATCGCTTTTCTGCATCAAC*C*A
ygiZ_2*:2*_lead	Set 3.150_lead TAG → TAA oligo	T*T*GTGGTTGGTGGCTTTTTTGGGGACGGTTTATA TTTTGCTATTAATAAAATCTATGAGAGTCGTTTTA ACGGCTCTCATAGACAGAG*A*A
yqiB_2*:2*_lead	Set 3.100_lead TAG → TAA oligo	G*C*GATACTGTTTAGCACATGGAGCGATGGCGAT TCCGGTTTATTAACGTCGTGAAACCTAAGGACAC CATTTGGAAGCCTGTTAACC*C*T
dnaG_Q576A	Oligo used to make DnaG Q576A mutation	GCACGCATGGTTTAAAGCAACGAAGAACGCCTGGA GCTCTGGACATTAAACGCGGAACTGGCGAAAAAG TGATTTAACGGCTTAAGTGCCG
dnaG_K580A	Oligo used to make DnaG K580A mutation	CGCACGCATGGTTTAAAGCAACGAAGAACGCCTGG AGCTCTGGACATTAAACCAGGAACTGGCGGCAAA GTGATTTAACGGCTTAAGTGCC
tolC.90.del	Oligo that deletes endogenous tolC	GAATTTTCAGCGACGTTTGACTGCCGTTTGAGCAGT CATGTGTAAAGCTTCGGCCCCGTCTGAACGTAA GGCAACGTAAAGATACGGGTTAT
galK_KO1.100	Oligo used to delete 100 bp including a portion of galK	C*G*CGCAGTCAGCGATATCCATTTTCGCGAATCC GGAGTGTAAGAAAACACACCGACTACAACGACG GTTTCGTTCTGCCCTGCGCGAT*T*G

Table 5.4 (Continued)

galK_KO1.1149	Oligo used to delete 1149 bp including a portion of galK	C*G*CGCAGTCAGCGATATCCATTTTCGCGAATCC GGAGTGTAAAGAAACGAAACTCCCGCACTGGCACC CGATGGTCAGCCGTACCGACT*G*T
galK_KO1.7895	Oligo used to delete 7895 bp including a portion of galK	C*G*CGCAGTCAGCGATATCCATTTTCGCGAATCC GGAGTGTAAAGAACTTACCATCTCGTTTTACAGGCT TAACGTTAAAACCGACATTA*G*C
ygfJ_WT	Set 3X.850_wt-f mascPCR	GCTGCCGGATTATCATCAAGA
ygfT_WT	Set 3X.700_wt-f mascPCR	GCAATGGCCGCAGGAAGG
ubiH_WT	Set 3X.600_wt-f mascPCR	GCACGGTACGATGGGTAATAGAT
argO_WT	Set 3X.500_wt-f mascPCR	GAAGTCATGAAGCAAGGCAGA
yqgC_WT	Set 3X.400_wt-f mascPCR	CGGCAGGTATGCAAAGCAGA
trmI_WT	Set 3X.300_wt-f mascPCR	AGTATGGGACTTAATGTTTCGAGAGG
glcC_WT	Set 3X.250_wt-f mascPCR	AGGGCGAATCGGCAAGG
yghT_WT	Set 3X.200_wt-f mascPCR	GAAAAATATGGTGCTGCCGAAAGA
ygiZ_WT	Set 3X.150_wt-f mascPCR	CTTCTTACGCCACTTATTATTCTTATCTTAAGA
cpdA_WT	Set 3X.100_wt-f mascPCR	TGGCTGGTGAGGCCAGA
dnaG_Q576A_wt-f	dnaG_Q576A wt-f ascPCR primer	TGGAGCTCTGGACATTAAACCA
dnaG_K580A_wt-f	dnaG_K580A wt-f ascPCR primer	CATTAAACCAGGAAGTGGCGAA
ygfJ_MUT	Set 3X.850_mut-f mascPCR	GCTGCCGGATTATCATCAAGG
ygfT_MUT	Set 3X.700_mut-f mascPCR	GCAATGGCCGCAGGAAGA
ubiH_MUT	Set 3X.600_mut-f mascPCR	GCACGGTACGATGGGTAATAGAC
argO_MUT	Set 3X.500_mut-f mascPCR	GAAGTCATGAAGCAAGGCAGG
yqgC_MUT	Set 3X.400_mut-f mascPCR	GGCAGGTATGCAAAGCAGG
trmI_MUT	Set 3X.300_mut-f mascPCR	GAGTATGGGACTTAATGTTTCGAGAGA
glcC_MUT	Set 3X.250_mut-f mascPCR	GAGGGCGAATCGGCAAGA
yghT_MUT	Set 3X.200_mut-f mascPCR	AAAATATGGTGCTGCCGAAAGG
ygiZ_MUT	Set 3X.150_mut-f mascPCR	CTTCTTACGCCACTTATTATTCTTATCTTAAGG
cpdA_MUT	Set 3X.100_mut-f mascPCR	GGCTGGTGAGGCCAGG
dnaG_Q576A_mut-f	dnaG_Q576A mut-f ascPCR primer	GGAGCTCTGGACATTAAACGC
dnaG_K580A_mut-f	dnaG_K580A mut-f ascPCR primer	ACCAGGAAGTGGCGGC
ygfJ_rev	Set 3X.850_rev mascPCR	TCTGTTTGCACTGCGGGTAC
ygfT_rev	Set 3X.700_rev mascPCR	TGGTTGGGCAATCTAATAGATTCTCC
ubiH_rev	Set 3X.600_rev mascPCR	ATGAGCGTAATCATCGTCGGTG
argO_rev	Set 3X.500_rev mascPCR	CCGTCTCTCGCCAGCTG
yqgC_rev	Set 3X.400_rev mascPCR	AGCACACGACGTTTCTTTTCG
trmI_rev	Set 3X.300_rev mascPCR	ATCTGTTCTTCCGATGTACCTTCC
glcC_rev	Set 3X.250_rev mascPCR	CTCCAGCTCGATATCGTGGAG

Table 5.4 (Continued)

yghT_rev	Set 3X.200_rev mascPCR	CACCACCAAAGGTAACTGTGG
ygiZ_rev	Set 3X.150_rev mascPCR	CACAAACCAGACAAATACCGAGC
cpdA_rev	Set 3X.100_rev mascPCR	CGATGGTATCCAGCGTAAAGTTG
dnaG_seq-r	dnaG rev ascPCR primer for both Q576A and K580A	GCTCCATAAGACGGTATCCACA
Rx-P19	forward screening primer for wt tolC deletion	GTTTCTCGTGCAATAATTTCTACATC
Rx-P20	reverse screening primer for wt tolC deletion	CGTATGGATTTTGTCCGTTTCA
lacZ_jackpot_seq-f	forward sequencing primer for lacZ alleles	GAATTGTGAGCGGATAACAATTTTC
lacZ_jackpot_seq-r	reverse sequencing primer for lacZ alleles	CCAGCGGCTTACCATCC
cat_mut*	cat inactivation oligo	G*C*ATCGTAAAGAACATTTTGAGGCATTTTCAGTC AGTTGCTTAATGTACCTATAACCAGACCGTTCAGC TGGATATTACGGCCTTTTTTA*A*A
cat_restore*	cat reactivation oligo	G*C*ATCGTAAAGAACATTTTGAGGCATTTTCAGTC AGTTGCTCAATGTACCTATAACCAGACCGTTCAG CTGGATATTACGGCCTTTTTTA*A*A
tolC-r_null_mut*	tolC inactivation oligo	A*G*CAAGCACGCCTTAGTAACCCGGAATTGCGTA AGTCTGCCGCTAAATCGTGATGCTGCCTTTGAAA AAATTAATGAAGCGCGCAGTCCA
tolC-r_null_revert*	tolC reactivation oligo	C*A*GCAAGCACGCCTTAGTAACCCGGAATTGCGT AAGTCTGCCGCCGATCGTGATGCTGCCTTTGAAA AAATTAATGAAGCGCGCAGTCCA
tolC_null_revert*	tolC reactivation oligo (leading-targeting)	T*G*GACTGCGCGCTTCATTAATTTTTTCAAAGGC AGCATCACGATCGGCGGCAGACTTACGCAATTCC GGGTTACTAAGGCGTGCTTGCTG
bla_mut*	bla inactivation oligo	G*C*C*A*CATAGCAGAACTTTAAAAGTGCTCATC ATTGGAAAACGTTATTAGGGGCGAAAACCTCTCAA GGATCTTACCGCTGTTGAGATCCAG
bla_restore*	bla reactivation oligo	G*C*C*A*CATAGCAGAACTTTAAAAGTGCTCATC ATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAA GGATCTTACCGCTGTTGAGATCCAG
313000.T.lacZ.coMA GE-f	Forward cassette primer for generating T.co-lacZ (inserted for lacZ co-selection)	TGCTTCTCATGAACGATAACACAACCTGTTTCATGA ATTAACCATTCGGATTGAGGCACATTAACGCC
313001.T.lacZ.coMA GE-r	Reverse cassette primer for generating T.co-lacZ (inserted for lacZ co-selection)	ACGGAAACCAGCCAGTTCCTTTTCGATGCCTGAAT TTGATCCCATAGTTTATCTAGGGCGGCGGATT
312869.seq-f	Forward screening primer for T.co-lacZ	GAACTTGCACTACCCATCG
313126.seq-r	Reverse screening primer for T.co-lacZ	AGTGACGGGTTAATTATCTGAAAG
LacZ::KanR.full-f	Forward cassette primer for generating lacZ::kanR	TGACCATGATTACGGATTCACTGGCCGTCGTTTTC CAACGTCGTGCCTGTGACGGAAGATCACTTCG
LacZ::KanR.full-r	Reverse cassette primer for generating lacZ::kanR	GTGCTGCAAGGCGATTAAAGTTGGGTAAACGCCAGG GTTTTCCAGTAACCAGCAATAGACATAAGCGG

Strain Creation

Lambda Red recombination was used to generate all mutations, and was performed as described below and in previous chapters. All of the strains used in this work were derived from EcNR2 (*Escherichia coli* MG1655 Δ mutS::cat Δ (ybhB-bioAB)::[λ cl857 N(*cro-ea59*)::tetR-bla]).⁶ Strain nuc5⁻.dnaG.Q576A was generated by recombining oligo dnaG_Q576A into nuc5⁻, an EcNR2-derived strain described in Chapter 4.¹ EcNR2.DT was created by deleting the endogenous *tolC* gene using the tolC.90.del recombineering oligo.¹⁴ EcNR2.T.co-lacZ was created by recombining a *tolC* cassette (T.co-lacZ) into the genome of EcNR2.DT, upstream of the lac operon, as described below. CoS-MAGE strains were prepared by inactivating a chromosomal selectable marker (*cat*, *tolC*, or *bla*) using an oligonucleotide. Clones with sensitivity to the appropriate antibiotic or SDS²² were identified by replica plating. The growth rates of strains EcNR2, EcNR2.dnaG.K580A, and EcNR2.dnaG.Q576A were approximately equivalent, while nuc5⁻.dnaG.Q576A had a doubling time that was roughly 7% longer than those of the other strains.

Generating the T.co-lacZ dsDNA Recombineering Cassette

The T.co-lacZ dsDNA recombineering cassette (for inserting *tolC* in the vicinity of the *lacZ* gene) was generated by PCR, using primers 313000.T.lacZ.coMAGE-f and 313001.T.lacZ.coMAGE-r (Table 5.4). PCR was performed using Kapa HiFi HotStart ReadyMix, with primer concentrations of 0.5 μ M, and T.5.6¹⁴ used as template (wherein a terminator was present downstream of the *tolC* stop codon). The PCR (50 μ L total) was heat activated at 95 °C for 5 min, then cycled 30 times at 98 °C (20 sec), 62 °C (15 sec), and 72 °C (45 sec). The final extension was at 72 °C for 5 min. The Qiagen PCR purification kit was used

to isolate the PCR product (with elution in 30 μ L H₂O). Purified PCR product was quantitated on a NanoDrop ND1000 spectrophotometer, and analyzed on a 1% agarose gel with ethidium bromide staining.

Performing Lambda Red Recombination

Lambda Red recombination was performed as described in Chapter 3. Here, for recombination of dsDNA PCR products, 50 ng of PCR product was used. For experiments in which a single oligo was recombined, 1 μ M of oligo was used. For experiments in which sets of ten or twenty recombineering oligos were recombined along with a co-selection oligo, 0.5 μ M of each recombineering oligo and 0.2 μ M of the co-selection oligo were used (5.2 μ M total for 10-plex sets and 10.2 μ M total for the 20-plex set).

Electroporated cells were allowed to recover in 3 mL LB Lennox media, in a rotator drum at 30 °C. For MAGE and CoS-MAGE experiments, cultures were recovered to apparent saturation (5 or more hours) to minimize polyclonal colonies. MAGE recovery cultures were diluted to $\sim 5 \times 10^3$ cfu/mL, and 50 μ L of this dilution was plated on non-selective LB Lennox agar plates. To compensate for fewer recombinants surviving co-selection, CoS-MAGE recovery cultures were diluted only to 10^5 cfu/mL, and 50 μ L of this dilution was plated on appropriate selective media for the co-selected resistance marker (LB Lennox with 50 μ g/mL carbenicillin for *bla*, 20 μ g/mL chloramphenicol for *cat*, or 0.005% w/v SDS for *tolC*, with 20 μ g/mL chloramphenicol added to enhance the robustness of selection). Leading-targeting CoS-MAGE recovery cultures were diluted to $\sim 5 \times 10^6$ cfu/mL before plating.

Analyzing Recombination

GalK activity was assayed by plating recovered recombination cultures onto MacConkey agar supplemented with 1% galactose as a carbon source. Red colonies were scored as *galK*⁺, and white colonies were scored as *galK*⁻. LacZ activity was assayed by plating recovery cultures onto LB Lennox + X-Gal/IPTG (Fisher ChromoMax). Blue colonies were scored as *lacZ*⁺ and white colonies were scored as *lacZ*⁻.

Kapa 2GFast ReadyMix was used in colony PCRs to screen for the correct insertion of the *tolC* selectable marker that was placed near *lacZ*. PCRs had a total volume of 20 μ L, with 0.5 μ M of each primer. These PCRs were carried out with an initial activation step at 95 °C for 2 min, then cycled 30 times at 95 °C (15 sec), 56 °C (15 sec), and 72 °C (40 sec), followed by a final extension at 72 °C (90 sec). PCR products were analyzed on a 1% agarose gel with ethidium bromide, to confirm amplification of the expected band.

Allele-specific colony PCR (ascPCR) was used to detect the DnaG K580A and Q576A mutations, and was performed as described in Chapter 3. Multiplex allele-specific colony PCR (mascPCR)¹⁵ was used to detect the 1-2 bp mutations generated in the MAGE and CoS-MAGE experiments, and was performed as described in Chapter 4. Annealing temperatures for ascPCR and mascPCR were optimized for each pair of reactions, and ranged between 63 °C and 67 °C.

In CoS-MAGE experiments, all strains were recombined with all tested oligo sets at least twice. All replicates for a given strain and oligo set were combined to generate a complete data set. Polyclonal or ambiguous mascPCR results were discarded from our analysis. The mean number of alleles replaced per clone was determined by scoring each allele as 1 for converted or 0 for unmodified. Data for EcNR2 and nuc5⁻ were as collected and described in Chapter 4. Given the sample sizes tested in the CoS-MAGE experiments ($n > 47$), we used parametric

statistical analyses, as these analyses are preferable to their non-parametric equivalents for large sample sizes.²³ We used a one-way ANOVA to test for significant variance in the CoS-MAGE performance of the strains for a given oligo set. Subsequently, we used a Student's t-test to make pairwise comparisons, with significance defined as $p < 0.05/n$, where n is the number of pairwise comparisons. Here, $n = 15$, as this data set was planned and collected as part of a larger set with 6 different strains, although only EcNR2, EcNR2.dnaG.K580A, EcNR2.dnaG.Q576A, and nuc5⁻.dnaG.Q576A are presented here (the others, EcNR2.xseA⁻ and nuc5⁻, are discussed in Chapter 4). As such, significance was defined as $p < 0.003$ for the analyses presented in Figures 5.3 and 5.5. Statistical significance in these figures is denoted using a system where * denotes $p < 0.003$, ** denotes $p < 0.001$, and *** denotes $p < 0.0001$. For the experiment comparing the CoS-MAGE performance of EcNR2 and EcNR2.dnaG.Q576A with leading-targeting oligos, we tested for statistical significance using a single t-test with significance defined as $p < 0.05$.

For the experiment in which 10 loci were targeted within *lacZ*, recombinants were identified by blue/white screening. The frequency of clones with 1 or more alleles replaced (# of white colonies / # of total colonies) was determined for every replicate. For white colonies only, a portion of the *lacZ* gene was amplified with primers lacZ_jackpot_seq-f and lacZ_jackpot_seq-r, using Kapa HiFi HotStart ReadyMix as described above. Purified (Qiagen PCR purification kit) amplicons were submitted to Genewiz for Sanger sequencing in both directions using lacZ_jackpot_seq-f and lacZ_jackpot_seq-r. Combined, the two sequencing reads interrogated all 10 targeted alleles. Three replicates of recombinations and blue/white analyses were performed to ensure consistency, but only one replicate was sequenced ($n = 39$ for EcNR2 and $n = 55$ for EcNR2.dnaG.Q576A). The mean number of alleles converted per clone was determined as above. We tested for statistically significant differences in mean allele conversion between

the strains using a Student's t-test with significance defined as $p < 0.05$. Statistical significance in Figure 5.4C is denoted using a system where *** denotes $p < 0.0001$.

References

1. Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. & Church, G.M. Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *PLoS One* **7**, e44638 (2012).
2. Carr, P.A. et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40**, e132 (2012).
3. Zechner, E.L., Wu, C.A. & Marians, K.J. Coordinated leading- and lagging-strand synthesis at the *Escherichia coli* DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. *J Biol Chem* **267**, 4045-53 (1992).
4. Tougu, K. & Marians, K.J. The interaction between helicase and primase sets the replication fork clock. *J Biol Chem* **271**, 21398-405 (1996).
5. Tougu, K. & Marians, K.J. The extreme C terminus of primase is required for interaction with DnaB at the replication fork. *J Biol Chem* **271**, 21391-7 (1996).
6. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-8 (2009).
7. Maresca, M. et al. Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* **11**, 54 (2010).
8. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**, 791-9 (2010).
9. Ellis, H.M., Yu, D., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6 (2001).
10. Erler, A. et al. Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J Mol Biol* **391**, 586-98 (2009).
11. Corn, J.E. & Berger, J.M. Regulation of bacterial priming and daughter strand synthesis through helicase-primase interactions. *Nucleic Acids Res* **34**, 4082-8 (2006).

12. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. & Sugino, A. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A* **59**, 598-605 (1968).
13. Lia, G., Michel, B. & Allemand, J.F. Polymerase exchange during Okazaki fragment synthesis observed in living cells. *Science* **335**, 328-31 (2012).
14. Isaacs, F.J. et al. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-53 (2011).
15. Wang, H.H. & Church, G.M. Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. *Methods Enzymol* **498**, 409-26 (2011).
16. Oakley, A.J. et al. Crystal and solution structures of the helicase-binding domain of Escherichia coli primase. *J Biol Chem* **280**, 11495-504 (2005).
17. Tanner, N.A. et al. Single-molecule studies of fork dynamics in Escherichia coli DNA replication. *Nat Struct Mol Biol* **15**, 998 (2008).
18. Nakayama, M. & Ohara, O. Improvement of recombination efficiency by mutation of red proteins. *Biotechniques* **38**, 917-24 (2005).
19. Yao, N.Y., Georgescu, R.E., Finkelstein, J. & O'Donnell, M.E. Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. *Proc Natl Acad Sci U S A* **106**, 13236-41 (2009).
20. Rybalchenko, N., Golub, E.I., Bi, B. & Radding, C.M. Strand invasion promoted by recombination protein beta of coliphage lambda. *Proc Natl Acad Sci U S A* **101**, 17056-60 (2004).
21. Asai, T. & Kogoma, T. D-loops and R-loops: alternative mechanisms for the initiation of chromosome replication in Escherichia coli. *J Bacteriol* **176**, 1807-12 (1994).
22. DeVito, J.A. Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. *Nucleic Acids Res* **36**, e4 (2008).

23. Jekel, J.F., Katz, D.L., Elmore, J.G. & Wild, D. *Epidemiology, Biostatistics, & Preventative Medicine* (W.B. Saunders, 2011).

Chapter Six

Conclusion: Lambda Red Recombination,
Today and Going Forward

The work described in this thesis has substantially advanced our understanding of Lambda Red recombination. In Chapter 2, we proposed a novel mechanism for Lambda Red dsDNA recombination in which Lambda Exo degrades the leading-targeting strand to generate a full-length ssDNA intermediate.¹ This intermediate then anneals at the lagging strand of the replication fork, where it is incorporated into the newly-synthesized strand as an Okazaki Fragment; this final step is analogous to the accepted mechanism for Lambda Red ssDNA recombination.² Thus, our mechanism unifies these two models and presents a parsimonious and logical explanation of how dsDNA recombination occurs. Moreover, we supported our model with a large body of experimental evidence showing that a full-length ssDNA intermediate is much more likely than the intermediate proposed in prior mechanisms for dsDNA recombination.¹

In Chapter 3, we further supported this mechanism through our recombineering experiments with phosphorothioated dsDNA cassettes. These results also led us to discover that ExoVII degrades dsDNA cassette ends, and that this has a negative impact on recombination frequency. Indeed, the variable dsDNA recombination frequencies of the nuclease knockout strains tested in this chapter indicated that endogenous nucleases routinely act on dsDNA cassettes, and that this can profoundly impact recombination.³ In Chapter 4, we went on to show that endogenous nucleases also degrade the ssDNA oligonucleotides used for MAGE and CoS-MAGE. By extension, this established that the concentration of oligonucleotides within the cell – and hence, oligo uptake – is a limiting factor for multiplex oligonucleotide recombination. We also showed that oligonucleotide phosphorothioation has countervailing effects on recombination frequency. While phosphorothioation can increase recombination frequency by protecting oligos from nuclease degradation, it can also hinder recombination, presumably by reducing the

strength of the annealing interaction between the oligo and the lagging strand of the replication fork. The net effect of oligo phosphorothioation depends on the nuclease background of the strain used for recombineering.³ Finally, in Chapter 5, we showed that the amount of accessible ssDNA on the lagging strand limits multiplex oligonucleotide recombination frequency. These results thereby established that intracellular oligo concentration and accessible lagging strand ssDNA are separate factors that both significantly influence MAGE and CoS-MAGE frequency.⁴

As a result of this advancement in our understanding of Lambda Red recombination, we were able to markedly improve the performance of both dsDNA and ssDNA recombineering. For dsDNA recombineering, we showed that recombination frequency could significantly be enhanced in two different ways – by protecting the 5' end of the lagging-targeting strand with phosphorothioate (PT) bonds, and by inactivating the exonuclease ExoVII.³ By combining these two strategies, we were able to achieve a roughly six-fold improvement of gene insertion frequency.³ Moreover, removing ExoVII enabled mutations encoded near the ends of a dsDNA cassette to more reliably be conferred.

In addition to its impact on dsDNA recombination, we found that ExoVII removal also improved the inheritance of mutations encoded near the 3' ends of ssDNA oligonucleotides, and slightly increased CoS-MAGE frequencies. Building upon these results, we discovered that CoS-MAGE performance could further be enhanced through the deletion of four additional nucleases (RecJ, ExoI, ExoX, and Lambda Exo). In comparison with our standard recombineering strain EcNR2, the resulting “nuc5” strain yielded 46% more alleles converted per average clone, 200% more clones with five or more allele conversions, and 35% fewer clones without any allele conversions in a cycle of 10-plex CoS-MAGE.³ Additionally, we found that CoS-MAGE allele conversion frequency could be improved by attenuating the

interaction between DnaG primase and replisome-associated DnaB helicase, which increases the amount of accessible ssDNA on the lagging strand. The more attenuated of the two primase variants tested in our study displayed 62% more alleles converted per average clone, 239% more clones with five or more allele conversions, and 38% fewer clones without any allele conversions.⁴ Moreover, combining primase modification and nuclease removal had additive beneficial effects on CoS-MAGE performance. In comparison with EcNR2, the resulting “nuc5⁻.dnaG.Q576A” strain demonstrated 111% more alleles converted per average clone, 527% more clones with five or more allele conversions, and 71% fewer clones without any allele conversions in a cycle of 10-plex CoS-MAGE.⁴ This strain also displayed improved performance for larger CoS-MAGE oligo sets; in a single cycle of 20-plex CoS-MAGE, nuc5⁻.dnaG.Q576A yielded 4.50 non-selectable alleles converted per average clone.⁴ This represents a nearly three-fold improvement over EcNR2, and significantly augments the power of CoS-MAGE for diversity generation and genome engineering.

These advancements represent a significant step forward in the development of Lambda Red technology; however, further work will be necessary in order to maximize the utility of recombineering. While we have managed to increase Lambda Red dsDNA gene insertion frequency approximately six-fold,³ the resulting recombination frequency of ~0.3% is still such that hundreds or thousands of colonies would need to be screened by PCR in order to ensure the identification of a recombinant colony. As a result, selectable markers are still necessary in order to identify recombinants in a practical fashion, complicating pathway transfer and preventing the multiplex and/or combinatorial insertion of exogenous genes. Similarly, although we significantly increased both the average and top number of alleles converted in a given CoS-MAGE cycle, the power of the technique is still limited by the number of mutations that it can

generate. While our work will certainly facilitate novel CoS-MAGE applications, genome engineering efforts involving several thousand or more genetic changes probably remain beyond the reach of current recombineering capabilities. Thus, the development of Lambda Red technology must continue. Here, I briefly discuss what I believe to be the most fruitful avenue for the future improvement of dsDNA and ssDNA recombineering, respectively.

For several reasons, I believe that the performance of Lambda Red dsDNA-mediated gene insertion could significantly be improved through directed evolution. For one, the recombination of antibiotic resistance genes provides for an extremely straightforward selection for gene insertion. Successive rounds of recombination and selection should enrich for highly recombinogenic cells, and insertion cassettes can be designed so that they replace the cassette(s) used in prior rounds,⁵ thereby allowing those selectable markers to be reused. It should therefore be fairly simple to carry out a directed evolution project using such a strategy. Another reason for optimism regarding this approach is that a previous attempt at a similar experiment yielded positive results. In this work,⁶ the authors mutagenized the Lambda Red genes, then performed two rounds of selection for the ability to recombine a dsDNA antibiotic resistance insertion cassette. Despite performing only one round of diversification and two rounds of selection, the authors were nevertheless able to isolate mutants of Exo and Beta that together provided for a roughly four-fold increase in recombination frequency.⁶ Additional rounds of diversification and selection are likely to yield even more impressive results. Finally, there are several promising targets for diversification. First, the Lambda Red genes could themselves be altered (as in the above-described study), to isolate variants with improved recombineering properties. Additionally, the strong impact of nuclease activity on dsDNA recombination frequency³ suggests that the inactivation, downregulation, and/or upregulation of endogenous nucleases

could significantly improve recombineering performance. For either of these sets of targets, diversification could easily be facilitated by MAGE. Lastly, genome-wide diversification could be accomplished through the use of untargeted mutagenesis (e.g., via *mutS* removal, chemical mutagens, and/or UV radiation). Any recombination-enhanced variants arising from genome-wide modification could then be characterized by next-generation sequencing. Such an unbiased approach could reveal host factors that impact recombination, for example by interacting with Lambda Red proteins, altering replisome dynamics, or affecting DNA uptake. Thus, directed evolution is likely to be a versatile and powerful platform for further improving Red-mediated dsDNA gene insertion; efforts to utilize this strategy are currently underway in our lab.

Directed evolution may also be useful for improving ssDNA recombineering; for example, cells could be selected for the ability to carry out the simultaneous oligo-mediated repair of several broken antibiotic resistance genes. However, I believe that the most promising near-term avenue for improving ssDNA recombineering is through making CoS-MAGE amenable to cycling. As discussed previously, CoS-MAGE involves the use of a co-selection oligonucleotide which repairs a chromosomally-located defective selectable marker.⁷ After selection for the repaired marker, that marker (or a newly placed one) must then be inactivated before the next round of CoS-MAGE can proceed. This complication prevents CoS-MAGE from easily being applied in iterative cycles.

The use of bidirectionally selectable markers presents a possible solution to this problem. Such markers could both be broken and repaired with concomitant selection, thereby facilitating the seamless cycling of CoS-MAGE. Our lab typically uses *tolC* as a bidirectionally selectable marker, as alternative markers such as *galK* and *thyA* require selection on minimal media, which is time-consuming and inconvenient.⁸ SDS can be used to select for functional *tolC* (positive

selection), and colicin E1, a toxin that depends on *tolC* for cellular entry, can be used to select for non-functional *tolC* (negative selection).⁸ However, we have found that *tolC*⁺ cells can frequently survive negative selection, thereby limiting the usefulness of *tolC* as a bidirectionally selectable marker. To determine the origin of this problem, members of our lab sequenced the full genomes of 96 *tolC*⁺ cells that survived colicin selection. Subsequent analysis revealed that nearly all of these cells had spontaneous inactivating mutations in the *tolQRA* operon; it is likely that defective *tolQ*, *tolR*, or *tolA* proteins prevent colicin from entering the cell, even when functional *tolC* is present.⁹ We subsequently found that duplicating the *tolQRA* operon results in substantially fewer *tolC*⁺ escapees evading negative selection. By duplicating this operon, using vancomycin along with colicin for negative selection, and making other minor adjustments to the selection protocol, we succeeded in decreasing the dysfunction rate of *tolC* negative selection by over 2000-fold. This improvement enables the use of *tolC* for iterative CoS-MAGE. We are currently in the process of fine-tuning this technique and demonstrating its utility in a proof of concept experiment involving the recoding of AGG and AGA codons. When combined with the improvements described in Chapters 4 and 5, this iterative CoS-MAGE technology should facilitate a breakthrough in MAGE performance, enabling genome engineering of a heretofore unprecedented scope.

While future work will likely increase the power of recombineering, significant efforts are now achievable with current technology – due in part to the developments described in this thesis. To illustrate, I will briefly discuss a recent project in which we utilized Lambda Red recombineering to assess the feasibility of radically altering the genetic code of *E. coli* – a longstanding goal of the Church lab. Such a recoded organism would be immune to viruses (given that viral proteins would be mistranslated by recoded host cells), and unable to exchange

functional genetic material with the environment. Furthermore, codon reassignment would allow for the creation of organisms with novel chemical functionalities, through the dedicated incorporation of one or more non-standard amino acids (NSAAs). As discussed in Chapter 1, we have recently completed the creation of a strain with all instances of the UAG stop codon changed to UAA. This enabled the subsequent deletion of Release Factor 1 (RF1), which would otherwise be necessary for terminating genes ending in a UAG stop codon. Removing RF1 allowed the UAG codon to serve as a dedicated channel for the site-specific incorporation of NSAAs. Moreover, because RF1 was no longer present to terminate viral genes ending in UAG, this strain also demonstrated slight phage resistance.¹⁰ The UAG-reassigned strain therefore serves as a powerful proof of concept for the utility of recoding. However, this work also confirmed that several additional codons will likely need to be reassigned in order to achieve complete virus resistance and genetic isolation, and to enable the dedicated incorporation of multiple NSAAs. It was therefore imperative to assess the feasibility of more significantly altering the *E. coli* genetic code, and to test whether several additional codons could possibly be removed genome-wide.

To do this, we attempted to remove all instances of 13 rare codons (hereafter referred to as “forbidden” codons) from a panel of 42 essential genes, chosen to serve as a stringent test bed for codon essentiality. This panel included *prfB*, which relies on a programmed frameshift for proper translation,¹¹ as well as all 41 essential protein-coding ribosomal genes,¹² which are both highly expressed and tightly regulated. We first assessed the feasibility of radically recoding these genes: eliminating all instances of the 13 forbidden codons and shuffling all other possible codons to synonymous alternatives, changing all start codons to AUG, separating overlapping genes, removing frameshifts, and eliminating several restriction sites.¹³ Thus, our resulting

recoded gene designs had only 65.4% average nucleotide identity with the corresponding wild type genes, despite amino acid sequences being unchanged.¹³

We utilized Lambda Red dsDNA recombination to attempt to replace the 42 targeted wild type genes with their respective recoded genes. To do so, we synthesized the recoded genes from chip-generated oligonucleotides,¹⁴ and attached *kanR* genes to their C-termini via isothermal assembly.¹⁵ We then PCR-amplified these constructs with primers containing homology regions flanking the corresponding wild type genes, and separately recombined the resulting cassettes into EcNR2. Kanamycin-resistant recombinants were assayed by PCR and Sanger sequencing in order to determine whether the targeted gene had correctly and completely been replaced. Somewhat surprisingly, we found that 26/42 tested genes could successfully be converted to their radically recoded analogs, with varying effects on strain fitness.¹³ Encouraged by this, we next set out to answer the question of whether all instances of the 13 forbidden codons could be removed from the panel of 42 genes. Through recombination with the radically recoded dsDNA cassettes, we successfully removed 294 of the 405 targeted instances of forbidden codons; we next utilized CoS-MAGE to try to remove the remaining forbidden codons in small groups across several strains. We performed these CoS-MAGE recombinations in an EcNR2.*xseA*⁻ background, taking advantage of the fact that the removal of ExoVII (*xseA*) provides for greater inheritance of mutations encoded near the 3' end of an oligo, as discussed in Chapter 4.³ This allowed a greater number of codon changes to be conferred by a single oligonucleotide. Ultimately, we were able to remove all 111 remaining forbidden codons using this CoS-MAGE strategy, although one forbidden codon in *rplQ* could not be converted to the first attempted synonymous replacement codon. In subsequent attempts, this codon was successfully converted to a number of other synonymous and non-synonymous codons.¹³

Taken together, these results indicate that the *E. coli* genome is surprisingly malleable, and that the genome-wide removal of the 13 forbidden codons appears feasible. However, given the 16 failed recoded gene replacements and the case of the codon in *rplQ*, it is clear that many recoded genome designs will be non-viable for reasons that may not be apparent *a priori*. This reinforces the importance of using a genome engineering strategy that rapidly prototypes and tests many designs in small pieces, such as our approaches based on MAGE and dsDNA recombineering. As a result, it is highly likely that Lambda Red recombination will be a cornerstone technique for future efforts in genome engineering. The advancements described in this thesis have significantly increased our understanding of Lambda Red recombination, and substantially broadened its power. Thus, it is my hope that this work will enable new and exciting applications in synthetic biology and genome engineering.

References

1. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**, 791-9 (2010).
2. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu Rev Genet* **36**, 361-88 (2002).
3. Mosberg, J.A., Gregg, C.J., Lajoie, M.J., Wang, H.H. & Church, G.M. Improving lambda red genome engineering in *Escherichia coli* via rational removal of endogenous nucleases. *PLoS One* **7**, e44638 (2012).
4. Lajoie, M.J., Gregg, C.J., Mosberg, J.A., Washington, G.C. & Church, G.M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res* **40**, e170 (2012).
5. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5 (2000).
6. Nakayama, M. & Ohara, O. Improvement of recombination efficiency by mutation of red proteins. *Biotechniques* **38**, 917-24 (2005).
7. Carr, P.A. et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. *Nucleic Acids Res* **40**, e132 (2012).
8. DeVito, J.A. Recombineering with *tolC* as a selectable/counter-selectable marker: remodeling the rRNA operons of *Escherichia coli*. *Nucleic Acids Res* **36**, e4 (2008).
9. Lazzaroni, J.C., Dubuisson, J.F. & Vianney, A. The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins. *Biochimie* **84**, 391-7 (2002).
10. Lajoie, M.J. et al. Genomically Recoded Organisms Impart New Biological Functions. *Science (in revision)* (2013).
11. Higashi, K. et al. Enhancement of +1 frameshift by polyamines during translation of polypeptide release factor 2 in *Escherichia coli*. *J Biol Chem* **281**, 9527-37 (2006).

12. Yamazaki, Y., Niki, H. & Kato, J. Profiling of Escherichia coli Chromosome database. *Methods Mol Biol* **416**, 385-9 (2008).
13. Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D., Church, G.M. Towards a radically reassigned genetic code. (*submitted*) (2012).
14. Kosuri, S. et al. Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat Biotechnol* **28**, 1295-9 (2010).
15. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-5 (2009).

Appendix 1: Strains Used in this Thesis

Strain Name	Genotype/Description	Chapter(s) used in
EcNR2	<i>Escherichia coli</i> MG1655 $\Delta mutS::cat \Delta(ybhB-bioAB)::[\lambda cI857 \Delta(cro-ea59)::tetR-bla]$	2-5*
SIMD90	<i>Escherichia coli</i> W3110 $\Delta mutS::cat lac\Delta U169 galK_{TYR145UAG}[\lambda cI857 \Delta(cro-bioA), (int-cIII)<>bet]$	2§
EcNR2. <i>endA</i> ⁻	N/A (see left)	3
nuc4 ⁻	EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> , <i>xseA</i> ⁻ , <i>exoX</i>	3
EcNR2. <i>recJ</i> , <i>xseA</i> ⁻ , <i>exoX</i>	N/A (see left)	3
EcNR2. <i>xonA</i> ⁻ , <i>xseA</i> ⁻ , <i>exoX</i>	N/A (see left)	3
EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> , <i>exoX</i>	N/A (see left)	3
EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> , <i>xseA</i> ⁻	N/A (see left)	3
EcNR2. <i>xseA</i> ⁻	N/A (see left)	3, 4*
nuc5 ⁻	EcNR2. <i>xonA</i> ⁻ , <i>recJ</i> , <i>xseA</i> ⁻ , <i>exoX</i> , <i>redA</i> ⁻	4, 5*
EcNR2.dnaG.K580A	N/A (see left)	5*
EcNR2.dnaG.Q576A	N/A (see left)	5*
nuc5 ⁻ .dnaG.Q576A	N/A (see left)	5*
EcNR2.DT	EcNR2 $\Delta tolC$	5¶
EcNR2.T.co-lacZ	EcNR2.DT with <i>tolC</i> reinserted at nucleotide position 313000, and subsequently inactivated	5¶

* Versions of these strains were also made with *cat*, *bla*, and *tolC* inactivated (separately).

§ Strain SIMD90 was a generous gift from Dr. Donald Court, and is described in the following paper: Datta, S., Costantino, N., Zhou, X. & Court, D.L. Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* **105**, 1626-31 (2008).

¶ Versions of these strains were also made with the DnaG Q576A mutation.