Expanded CTG Trinucleotide Repeats Stimulate Homologous Recombination in *Escherichia coli*

John Kenneth Blackwood

Thesis presented for the degree of Doctor of Philosophy
Institute of Cell Biology
University of Edinburgh
June 2006
Declaration

I hereby declare that this thesis was composed by me, and the research presented is my own, except where otherwise stated.
Acknowledgements

I would like to thank David Leach for allowing me to do my PhD in his lab. For his patience, guidance and for the constantly open door. My PhD would have not taken shape the way it did without the efforts of Rabaab Zahra – thank you Rabaab! Although we have focused on different parts of TNR biology we have always worked well together and she has been a fantastic ‘TNR team’ member! John Eykelenboom and I began our PhD’s at the same time and I have enjoyed working opposite and collaborating with him for the past three and a half years. Our lunchtime walks down to Sainsbury’s, countless tea breaks, pints and interesting discussions about lab work and nonsense! Also most of my time in the lab has been with Elise Darmon. I would like to thank Elise for caring about the lab; making sure that group meetings went smoothly, for taking the time to double check things and for carefully proof-reading my thesis. Thanks also to Laura Wardrope, Martin White and Federica Andreoni for the endless supply of ‘good-chat’, laughs and microscope help. When I began in the lab, Emily Wilson took me under her wing and showed me the ropes until I began to feel comfortable in the lab. Finally, thanks to Phil Eastlake and Transgenomic for financial support. This thesis is dedicated to my mother, Anne Blackwood.
Abstract

Expanded trinucleotide repeats (TNRs) (e.g. CAG, CTG, CCG) cause 40 different human diseases, however the molecular mechanism underlying the expansion of TNRs is poorly understood. This work describes the integration, in the chromosome of the bacterium *Escherichia coli*, of differently sized CAG and CTG TNRs into the start of the *lacZ* gene and of a zeocin resistance recombination reporter substrate into the nearby gene, *cynX*. We show that TNRs stimulate recombination at *cynX* in a length dependant manner. Furthermore, stimulation of recombination is dependant on TNR orientation with respect to the origin of replication. Experiments indicate that zeocin recombination is reduced in *E. coli* mutants for double strand break repair (*recA* and *recB*); but not for gap repair (*recR*). TNR induced stimulation of recombination is shown to be independent of the DNA hairpin nuclease SbcCD arguing against any role of secondary structure formed by TNRs. The formation of DSBs by TNRs is investigated using pulse field gel electrophoresis (PFGE). Since reversed replication forks can initiate homologous recombination (HR), this thesis investigates the possibility that replication through TNRs lead to replication fork reversal (RFR). We test this hypothesis by assaying for HR using zeocin recombination reporter substrates positioned proximal and distal with respect to the origin of replication. I show that TNR induced HR is lower at the distal site. Furthermore the protein UvrD is known to be essential for RFR in certain replication mutants. TNR dependant stimulation of HR is lost in *uvrD* mutants. Both these data support the hypothesis that TNRs can cause RFR. I show that expanded CAG or CTG TNRs do not generate sufficient single stranded DNA to stimulate the SOS response in wild-type and but do so in CTG95 cells deficient for the accessory helicase Rep.
Finally two-dimensional neutral / neutral gel electrophoresis is setup to investigate replisome progression through TNRs.
Table of contents

Chapter One: Introduction 1
1.1 Trinucleotide repeats 2
1.2 DNA containing TNRs forms secondary structures 9
1.3 TNR instability is dependant on length, purity and orientation 12
1.4 TNRs cause double strand DNA breaks 14
1.5 TNRs impede DNA replication fork progression 16
1.6 Okazaki fragment processing may mediate TNR instability 19
1.7 Mismatch repair may mediate TNR instability 20
1.8 The effect of SbcCD, MRX complexes on TNR instability 22
1.9 Escherichia coli 23
1.10 Homologous recombination in Escherichia coli 23
1.11 Replication Fork Reversal (RFR) in Escherichia coli 27
1.12 UvrD and Rep helicases 31
1.13 This thesis 32

Chapter Two: Materials and Methods 34
2.1 Materials 35
2.1.1 General Reagents 35
2.1.1.1 Chemicals 35
2.1.1.2 Enzymes 35
2.1.1.3 Growth Reagents 35
2.1.2 Bacterial Growth Media 35
2.1.2.1 General Information 35
2.1.2.2 Bacterial Media 35
2.1.2.3 Antibiotics 37
2.1.3 Commonly Used Buffers 37
2.1.4 Escherichia coli Strains 42
2.1.5 Bacteriophage Strains 43
2.1.6 Oligonucleotides 44
2.1.7 Plasmids 45
2.2 Microbiological Methods 46
2.2.1 Growth of Bacteria 46
2.2.2 Preservation of Bacteria 46
2.2.3 P1 Transduction 46
2.2.3.1 Production of P1 plate lysates 47
2.2.4 UV irradiation 48
2.2.5 SbcCD phenotypic test 48
2.2.6 Heat shock transformation of Escherichia coli 49
2.2.7 pKO integration in Escherichia coli 49
2.2.8 Fluctuation analysis 51
2.3 Nucleic Acid Methods 52
2.3.1 Phenol/chloroform Extraction 52
2.3.2 Precipitation of Nucleic Acid 52
2.3.3 Mini Preparation of Plasmid DNA by Spin Column 52
2.3.4 Large Scale Preparation of Plasmid DNA by Filtration Column 53
2.3.5 Escherichia coli Genomic DNA preparation 53
2.3.6 Restriction Digestion of DNA 53
2.3.7 Agarose Gel Electrophoresis 53
2.3.8 Purification of DNA from Agarose Gels 54
2.3.9 Ligation of DNA Molecules 54
2.3.10 Dephosphorylation of oligonucleotides 54
2.3.11 Annealing of oligonucleotides 55
2.3.12 Amplification of DNA by the Polymerase Chain Reaction (PCR) 55
2.3.13 Escherichia coli Boiled Cells 56
2.3.14 Cross-over PCR 57
2.3.15 Site-Directed Mutagenesis (SDM) 58
2.3.16 Genemapper analysis 58
2.3.16.1 Sample preparation 59
2.3.16.2 Data analysis 59
2.3.17 Purification of PCR Products 60
2.3.18 DNA Sequencing 60
2.3.19 Pulse Field Gel Electrophoresis (PFGE) 61
Chapter Three: Expanded CTG TNRs stimulate HR in *E. coli* 67

3.1 Introduction 68
3.2 Introduction of TNRs into the start of the *lacZ* gene in the *E. coli* chromosome 68
3.2.1 Construction of plasmid, pLacD1 69
3.2.2 Collaboration with Rabaab Zahra 72
3.3 Construction of DL1786 73
3.4 Construction of the zeocin recombination reporter substrate 75
3.5 Zeocin recombination requires RecA and RecB but not RecR 79
3.6 TNRs stimulate recombination in *E. coli* in a length dependent manner 80
3.7 Recombination in different recombination mutants harbouring TNRs 82
3.8 TNR induced recombination is independent of SbcCD nuclease 83
3.9 TNRs mediated recombination may be orientation dependant 85
3.10 HR near *lacZ* does not stimulate CTG TNR instability 86
3.11 Conclusions 87

Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal in *E. coli*? 90

4.1 Introduction 91
4.2 Do expanded TNRs cause DNA double strand breaks? 91
4.3 Is zeocin HR stimulated by replication fork reversal? 94
4.4 Origin distal zeocin recombination (*mhpC*) 96
4.5 UvrD helicase 99
4.6 TNR induced recombination is independent of RecG helicase 101
4.7 Conclusions 102

Chapter Five: Expanded CTG and CAG TNRs do not stimulate the SOS response in *E. coli*? 104

5.1 Introduction 105
5.2 Expanded TNRs do not stimulate the SOS response 105
5.2.1 TNR induced SOS induction in a rep background 110
5.3 Two dimensional neutral / neutral gel electrophoresis 113
5.4 Conclusions 116

Chapter Six: Overview 118

6.1 Discussion 119
6.2 The molecular mechanism behind TNR induced HR 122
6.3 The future of TNR research 124

References 126

List of Figures

Figure 1.1 Trinucleotide Repeats (TNRs) 3, 4
Figure 1.2 Trinucleotide Repeat Expansion Disorders (TREDs) and anticipation 7, 8
Figure 1.3 Structures proposed to be formed by TNRs 10
Figure 1.4 Molecular mechanisms of TNR instability 18
Figure 1.5 Homologous recombination (HR) in *E. coli* 25
Figure 1.6 Replication fork reversal (RFR) models 29, 30
Figure 2.1 pKO integration and excision 51
Figure 2.2 Cross-over PCR 58
Figure 2.3 2D gel electrophoresis 63
Figure 3.1 Plasmids pMH9 and pTOF24
Figure 3.2 LacZ homology arms
Figure 3.3 Introduction of Bsal, Mfet and BbsI into the 5' region of lacZ
Figure 3.4 Plasmid pLacD1
Figure 3.5 Plasmid pDL2309
Figure 3.6 Plasmid pDL1828
Figure 3.7 A system to measure TNR stimulated HR in Escherichia coli
Figure 3.8 Plasmid pDL1998
Figure 3.9 Integration of trinucleotide repeats (TNRs) into DL1786
Figure 3.10 Recombination rates in recA, recB and recR mutants
Figure 3.11 Long CTG trinucleotide repeats stimulate HR
Figure 3.12 CTG TNR induced recombination rates in recA, recB and recR mutants
Figure 3.13 TNR induced recombination is independent of SbcCD
Figure 3.14 TNR induced recombination may be orientation dependent
Figure 3.15 System to determine TNR instability in zeocin resistant cells
Figure 4.1 A system to detect rare TNR induced DNA breaks
Figure 4.2 PFGE analysis of TNR induced chromosome breakage
Figure 4.3 Hypothesis for replication fork reversal leading to zeocin HR
Figure 4.4 Zeocin repeats at mhpC (origin distal)
Figure 4.5 Origin distal zeocin-recombination (mhpC)
Figure 4.6 Plasmids pDL2391 and pGBuvrD-2
Figure 4.7 TNR dependent HR is lost in an uvrD mutant
Figure 4.8 TNR induced recombination is dependent of RecG helicase
Figure 5.1 Plasmid pGB150
Figure 5.2 Naladixic acid stimulates the SOS response
Figure 5.3 Expanded TNRs do not stimulate the SOS response
Figure 5.4 Plasmid pDL2341
Figure 5.5 Cell length and GFP intensity in rep mutant strains
Figure 5.6 Plasmid pDL2908 and an example of a 2D-gel
Figure 5.7 - 5 Kbp region used for two dimensional gel analysis
Figure 5.8 Southern analysis to confirm EcoRI restriction site integration
Figure 5.9 Initial two dimensional gel analysis
Figure 5.10 Two dimensional gel analysis – the y-arc

List of Tables
Table 2.1 Antibiotics
Table 2.2 Escherichia coli Strains
Table 2.3 Bacteriophage λ Strains
Table 2.4 Oligonucleotides
Table 2.5 Plasmids
# Common abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>amino acid(s)</td>
</tr>
<tr>
<td>%</td>
<td>percent</td>
</tr>
<tr>
<td>ADP, ATP</td>
<td>adenosine 5'-diphosphate, adenosine 5'triphosphate</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
</tr>
<tr>
<td>Cm</td>
<td>chloramphenicol</td>
</tr>
<tr>
<td>cm</td>
<td>centimetre</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymine triphosphate</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DSBR</td>
<td>double-strand break repair</td>
</tr>
<tr>
<td>ds</td>
<td>double strand</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>diaminoethanetetra-acetic acid disodium salt</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>Hfr</td>
<td>bacterial strain carrying integrated F plasmid</td>
</tr>
<tr>
<td>k</td>
<td>kilo (10³)</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase pairs (10³ bp)</td>
</tr>
<tr>
<td>Km</td>
<td>kanamycin</td>
</tr>
<tr>
<td>l</td>
<td>litres</td>
</tr>
<tr>
<td>LB</td>
<td>Luria broth (medium)</td>
</tr>
<tr>
<td>-l</td>
<td>(fraction of a) litre</td>
</tr>
<tr>
<td>m</td>
<td>metre</td>
</tr>
<tr>
<td>μ</td>
<td>micro (10⁻⁶)</td>
</tr>
<tr>
<td>μM</td>
<td>microns (10⁻⁶)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>mol</td>
<td>moles</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pfu</td>
<td>plaque forming units</td>
</tr>
<tr>
<td>pfu pol</td>
<td>Pyrococcus furiosus polymerase</td>
</tr>
<tr>
<td>pH</td>
<td>power of hydrogen (-log(_{10}[H^+]))</td>
</tr>
<tr>
<td>phage</td>
<td>bacteriophage</td>
</tr>
<tr>
<td>superscript 'r' (')</td>
<td>resistant</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>sec</td>
<td>seconds</td>
</tr>
<tr>
<td>superscript 's' (')</td>
<td>sensitive</td>
</tr>
<tr>
<td>ss</td>
<td>single strand</td>
</tr>
<tr>
<td>SSB</td>
<td>single strand binding protein</td>
</tr>
<tr>
<td>SSGR</td>
<td>single strand gap repair</td>
</tr>
<tr>
<td>Str</td>
<td>streptomycin</td>
</tr>
<tr>
<td>Tc</td>
<td>tetracycline</td>
</tr>
<tr>
<td>Tris</td>
<td>hydroxymethyl aminomethane</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet radiation</td>
</tr>
<tr>
<td>wt</td>
<td>wild-type</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>V</td>
<td>volts</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Introduction

Richards and Sutherland, 1997
1.1 Trinucleotide repeats (TNRs)

Deoxyribonucleic acid (DNA) is the genetic material of most organisms. It has a major role in the determination of hereditary characteristics by encoding and controlling protein synthesis. For an organism to function as it should, its genome must remain stable. It must be copied and passed onto its offspring faithfully and any chemical or physical insults that threaten its integrity must be either prevented or dealt with.

A sudden change or mutation in the genome, which can be distinguished from “normal”, results in a mutant. Mutation causes genetic variation – sometimes this variation is of benefit to an organism, driving evolution; but more often than not it is detrimental. Mutations can happen in any part of the genome and DNA can undergo unique types of mutation.

DNA is made up of four nucleotides containing the bases – adenine, thymine, guanine, and cytosine. These nucleotides can exist in combinations of either one, two, three or more that can be repeated tandemly. One type, trinucleotide repeats (TNRs), can form 64 different types \(4^3\). Four of these are mono-nucleotide runs; the others are classed in 10 groups of 6 types of trinucleotide repeats that give the same sequence when repeated tandemly. Out of these 10, 3 are associated with human disease (Fig. 1.1a).
Chapter One: Introduction

(A)  
<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>A</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG</td>
<td>GGG</td>
<td>GGA</td>
<td>GGT</td>
<td>GGC</td>
</tr>
<tr>
<td>AG</td>
<td>AGG</td>
<td>AGA</td>
<td>AGT</td>
<td>AGC</td>
</tr>
<tr>
<td>TG</td>
<td>TGG</td>
<td>TGA</td>
<td>TGT</td>
<td>TGC</td>
</tr>
<tr>
<td>CG</td>
<td>CGG</td>
<td>CGA</td>
<td>CGT</td>
<td>CGC</td>
</tr>
<tr>
<td>GA</td>
<td>GAG</td>
<td>GAA</td>
<td>GAT</td>
<td>GAC</td>
</tr>
<tr>
<td>AA</td>
<td>AAG</td>
<td>AAA</td>
<td>AAT</td>
<td>AAC</td>
</tr>
<tr>
<td>TA</td>
<td>TAG</td>
<td>TAA</td>
<td>TAT</td>
<td>TAC</td>
</tr>
<tr>
<td>CA</td>
<td>CAG</td>
<td>CAA</td>
<td>CAT</td>
<td>CAC</td>
</tr>
<tr>
<td>GT</td>
<td>GTG</td>
<td>GTA</td>
<td>GTT</td>
<td>GTC</td>
</tr>
<tr>
<td>AT</td>
<td>ATG</td>
<td>ATA</td>
<td>ATT</td>
<td>ATC</td>
</tr>
<tr>
<td>TT</td>
<td>TTG</td>
<td>TTA</td>
<td>TTT</td>
<td>TTC</td>
</tr>
<tr>
<td>CT</td>
<td>CTG</td>
<td>CTA</td>
<td>CTT</td>
<td>CTC</td>
</tr>
<tr>
<td>GC</td>
<td>GCG</td>
<td>GCA</td>
<td>GCT</td>
<td>GCC</td>
</tr>
<tr>
<td>AC</td>
<td>ACG</td>
<td>ACA</td>
<td>ACT</td>
<td>ACC</td>
</tr>
<tr>
<td>TC</td>
<td>TCG</td>
<td>TCA</td>
<td>TCT</td>
<td>TCC</td>
</tr>
<tr>
<td>CC</td>
<td>CGG</td>
<td>CCA</td>
<td>CCT</td>
<td>CCC</td>
</tr>
</tbody>
</table>

(B)  

- Interrupted repeats
- Perfect (uninterrupted) repeats

![Graph showing frequency of interrupted and perfect repeats](image)

- **disease alleles**
- **repeat or copy number**
Figure 1.1 Trinucleotide repeats (TNRs)

(A) Table showing all 64 possible TNRs. Four groups are mono-nucleotide arrays (shown in italics). The rest are 10 groups of TNRs representing the same infinite double strand repeat sequence (shown in black). Out of the 10 groups, TNRs have been shown to be involved in human disease (each shown in a different colour) (B) Distribution and frequency of TNR array sizes in human population. Expansion of TNRs are shown using arrows. Longer TNR arrays are more unstable than short ones and are less likely to contain interruptions. (C) Schematic representation of the location of non-coding (top) and coding (bottom) disease-associated TNRs. These regions include: promoter, 5'-UTR, exons, introns, 3'-UTR or other chromosomal locations. Taken from Pearson et al., 2005.
Arrays of TNRs are not stable genetic elements. Poorly understood, rare initial events lead to arrays of TNRs changing in length (usually increase in humans) between generations and somatically. Furthermore, longer repeats have a tendency to expand more often than shorter ones and are less likely to contain interruptions (Fig 1.1b). This unique type of mutation in humans is termed “dynamic mutation” (Richards and Sutherland, 1992; Richards and Sutherland, 1997). Dynamic mutation is associated with the genetic phenomenon of anticipation. Originally observed in Huntington’s disease, it describes the decreasing age of onset and increased severity of disease observed in successive generations in an affected family (Fu et al., 1991; Sherman et al., 1985; Vegter-van der Vlis et al., 1976).

TNR instability has been separated into two groups. Small expansions and contractions in TNR copy number are classified as type one TNR instability. As the repeats increase in number towards a threshold, there is a sudden bias towards expansion over contraction – this is known as type two TNR instability. Type two TNR instability leads to repeats expanding to a much larger copy number (McMurray, 1995).

Expanded TNRs are associated with around forty diseases in humans, known as Trinucleotide Repeat Expansion Disorders (TREDs) (Fig. 1.1c) (Ashley and Warren, 1995; Cummings and Zoghbi, 2000; Pearson et al., 2005). The most well known of these diseases, including myotonic dystrophy, fragile X syndrome, Huntington’s disease, several spinocerebellar ataxias and Friedreich’s ataxia, have been associated with expansion of CTG, CGG or GAA repeats. For ease of reading, CTG, CGG and GAA references refer to double stranded arrays of TNRs – i.e. (CTG)_n•(CAG)_m, (CGG)_n•(CCG)_m, or (GAA)_n•(TTC)_m.
TNIRs that cause disease reside in either coding regions (type I TREDs) (Fig 1.2a) or in non-coding regions (type II TREDs) (Fig 1.2b) (Reddy and Housman, 1997). In type I TREDs, expanded repeats affect genes at the protein level. For these TREDs, dynamic mutation has been, so far, found to be normally detrimental to the cell. The repeat copy number is usually < 100 even in affected individuals and expansion creates long tracts of the same amino acid, commonly glutamine. These mutations produce a toxic gain of function to the proteins they encode, often resulting in large protein aggregates. Although the toxic protein is synthesised ubiquitously, certain cells, such as neurons, seem to be especially susceptible (Grabcyk et al., 2001). One example, Huntington’s disease, is caused by the death of neurons that control movement. This causes involuntary muscle movement, cognitive impairment and behavioural abnormalities (Fig 1.2c) (McMurray and McMurray, 2001).

In type II TREDs repeats reside in untranslated regions and are able to expand to high copy numbers, often in excess of 2000. Rather than acting at the protein level, type II TREDs are often a result of modified, usually reduced, gene expression. Attenuated forms of disease occur when TNRs have expanded to a copy number of anywhere between 60 and 200 repeats. This is known as a proto-mutation or a pre-mutation allele.

An example of anticipation is shown in Fig 1.2(d) in a family suffering from Myotonic dystrophy. In this family the grandmother, carrying enough CTG repeats to cause a proto-mutation, developed some subtle characteristics of myotonic dystrophy at a late age. These repeats expanded on transmission to her daughter who developed some stronger symptoms at an earlier age. However, by this time the daughter had further transmitted her larger CTG array to her son. The CTG array by this stage had
(A) 

Chapter One: Introduction

(LOF) Fragile X syndrome Friedreich's ataxia Myotonic dystrophy

(B) 

Full mutation
Pre-mutation Normal
Chapter One: Introduction

Scott Roavorn
at ago 18 and 33
'oars
Huntington’s disease patient, McMurray et al., 2001

Figure 1.2 Trinucleotide Repeat Expansion Disorders (TREDs) and anticipation
(A) Type I TREDs are caused by transcribed expanded TNRs. TNR expansions in type I TREDs are small and rarely exceed 100. This expansion creates long tracts of the same amino acid, commonly glutamine (shown in coloured box). GOF represents the toxic gain of function proteins with expanded TNR arrays acquire. Abbreviations for TREDs are given in Fig 1.1. (B) Type II TREDs are caused by TNRs that are non-translated. The TNRs in this class are able to expand into much greater copy number compared to type II TREDs. Expansion of these TNRs often results in modification of gene expression, usually repression. This is also known as loss of function (LOF). (C) Huntington’s disease patient, Scott Redford, who developed the disease around the age of 30. Taken from McMurray and McMurray, 2001 (D) A family suffering from Myotonic dystrophy (MD), exhibiting the signs of anticipation. The grandmother, with small TNR size, developed MD at a late age. Two transmissions of the TNR array responsible for MD to her grandson caused large CTG expansions and resulted in early onset of a severe type of MD (Taken from Harper, 2002).
expanded to such a large size that her son, from birth, developed the most severe characteristics of the disease (Harper, 2002).

Despite the establishment of a large research field since 1991, the molecular basis of TNR expansion remains enigmatic. What is clear though, is that conversion of a DNA-repeat sequence from a harmless copy-number polymorphism to an unstable, disease causing genetic element is a complex process that is likely to involve multiple, perhaps discrete (certainly rare) changes. This introductory chapter will focus on general features of TNRs and the proposed molecular mechanisms of TNR instability.

1.2 DNA containing TNRs forms secondary structures

To date, the most commonly acknowledged factor that could contribute to TNR instability is their ability to form a variety of secondary structures such as hairpins (Fig. 1.3a), slipped strand intermediates (Fig. 1.3b), triplex (Fig. 1.3c) and quadruplex structures (Fig. 1.3d) (McMurray, 1999).

Single stranded (ss)(CAG)$_n$, (CTG)$_n$, (CGG)$_n$ and (CCG)$_n$ TNRs can form intramolecular hairpin structures (Fig. 1.3a) that are stabilised by base pairing in every 2 of the 3 bases of the triplet (Fig. 1.3e) (Darlow and Leach, 1998; Mitas, 1997; Pearson et al., 1998b; Yu et al., 1995). Due to their composition, (CGG)$_n$ and (CCG)$_n$ hairpins have the ability to fold into a hairpin structure with either a C•C or G•G mismatch; (CTG)$_n$ and (CAG)$_n$ hairpins can only form hairpins with either a T•T or A•A mismatch.
Figure 1.3 Structures proposed to be formed by TNRs
Schematic representations of structures (A) hairpins (B) slipped strand (C) triplex and (D) quadruplex formed by TNRs. TNRs known to form such structures are shown beside each diagram. (E) Hairpin diagram showing mis-pairs every second base pair of CTG or CAG repeats. In this example, U can represent either A or T. (F) Two types of base pairing can occur in polypurine • polypyrimidine TNR arrays: Watson-Crick and Hoogsteen base pairs. A Hoogsteen base pair allows hydrogen binding of two base pairs between the N7 position of the purine base and the C6 amino group of the pyrimidine base. This is shown by rectangles in Fig D. (G) The formation of triplex structures may account for the reduced frataxin mRNA levels in Friedreich's ataxia by RNA polymerase trapping and is described in Section 1.2. Black oval represents RNA polymerase (RNAP). Taken from Grabczyk and Usdin, 2000.
Of particular interest is that different TNR hairpins show differing degrees of stability. For example, (CTG)$_n$ hairpins are more stable than (CAG)$_n$ hairpins (Gacy and McMurray, 1998; Mitas, 1997; Petruska, 1996). Also, (CGG)$_n$ hairpins are more stable than (CCG)$_n$ hairpins. These observations are thought to be significant in determining (CAG)$_n$, (CTG)$_n$, (CGG)$_n$ and (CCG)$_n$ instability.

Petruska (1996) suggests that within a TNR array, rather than one large secondary structure being formed such as a hairpin, smaller hairpins may be formed but more frequently (Fig. 1.3b). Small hairpins formed on one DNA strand may result in similar sized hairpins being formed on the other, thereby preventing misalignment between the two DNA strands. An out-of-register misalignment could form structures known as slipped strand DNA (Pearson and Sinden, 1996; Pearson et al., 1998b).

(GAA)$_n$ TNRs can form a more complex triplex structure (Fig. 1.3c) (Potaman et al., 2004; Usdin and Grabczyk, 2000). Triplexes are three-stranded DNA structures formed at polypurine • polypyrimidine (R•Y) tracts (Fig. 1.3f). The third strand sits in the larger major DNA groove, forming Hoogsteen base pairs between the purine or pyrimidine bases with the purines of the Watson-Crick base pairs.

The reduced frataxin mRNA levels observed in Friedreich's ataxia patients may be due to the RNA polymerase becoming trapped within a triplex structure (Fig. 1.3g). During transcription, a region of negative super-coiling that follows RNA polymerase is believed to be relaxed as the untranscribed GAA strand begins to fold into the region of negative super-coiling, forming a triplex (Grabczyk and Usdin, 2000). It has also been suggested that (GAA)$_n$ TNR tracts can also form hairpin structures (LeProust et al., 2000).

(CGG)$_n$, (AGG)$_n$ and (TGG)$_n$ TNRs can form a variety of quadruplex structures (Fig. 1.3d) (Usdin, 1998). Resembling long hairpins that have folded in on
themselves (a uni-molecular quartet), quadruplex structures are held together by Hoogsteen base pairs between guanine (purine) bases (Fry and Loeb, 1994). Two individual hairpins could also form a quadruplex structure. Because of their complex nature, the formation of quadruplex structures would require a large amount of ss-TNR containing DNA.

Studies so far have documented an important amount of in vitro evidence suggesting that disease causing TNRs form secondary structures; however there is no in vivo data. Many observations of TNR length, orientation and purity are consistent with the formation of hairpins and frequently appear in models explaining TNR instability. Many of the more elaborate triplex and quadruplex structures may pose a significant threat to transcription (Grabczyk et al., 2001) or replication. However, it is unclear how a cell would tolerate such structures in its genome.

1.3 TNR instability is dependent on length, purity and orientation

To date, no model organism harbouring TNRs has successfully recapitulated the type of instability observed in humans. Regardless of repeat length integrated into the genome or plasmid, the most commonly observed type of somatic TNR instability is contraction (Freudenreich et al., 1998; Hirst and White, 1998; Kang et al., 1995a; Rolfsmeier et al., 2001; Schweitzer and Livingston, 1998). (Maurer et al., 1996; Schmidt, 2000)

However, some transgenic murine models do demonstrate a somatic age-dependent bias towards expansion over contraction (Fortune et al., 2000; Gomes-Pereira et al., 2001; Lia et al., 1998; Savouret et al., 2003; van den Broek et al., 2002). Interestingly, very few of the transgenic lines constructed demonstrate TNR instability. This suggests that chromosomal context, such as DNA flanking TNR
arrays, may be important in TNR instability. Furthermore, for unknown reasons, instability is often most pronounced in the kidney, small intestine and stomach out of all tissues analysed.

Experiments in bacteria and yeast have shown that the direction of replication through a TNR tract determines how stable the tract will be – TNR orientation with respect to the origin of replication. It is generally accepted that a CTG•CAG repeat tract is more unstable when (CTG), resides on the Okazaki fragment (for expansions) or on the lagging strand template (for deletions) (Fig. 4a) (Section 1.5) (Freudenreich et al., 1997; Kang et al., 1995a; Maurer et al., 1996; Miret et al., 1998).

This observation correlates with the ability of the TNR to form hairpin structures (Section 1.2) on the lagging strand. As the lagging strand is synthesised discontinuously, there is a greater chance for the formation of ss-hairpins. The difference in stability between CTG and CAG orientations is probably the consequence of CAG hairpins being less stable than CTG hairpins. Similar experiments have been carried out on CGG TNRs in bacteria and yeast. CGG TNRs deletions are more frequent when CGG resides on the lagging strand template compared to CCG (Balakumaran et al., 2000; Hirst and White, 1998; Rolfsmeier et al., 2001; Shimizu et al., 1996; White et al., 1999). The difference in stability between CGG and CCG TNRs is again thought to be related to experimental data suggesting that ss-CGG DNA forms more stable hairpins than CCG.

The conversion of type I TNR instability into type II instability is often accompanied by a founder effect – a loss of stabilising interruptions present in the TNR array. Founder effects leading to the loss of AGO, CAT, CAA and GAGGAA interruptions
in Fragile X syndrome, Spinocerebellar ataxias type 1, Spinocerebellar ataxia type 2 and Friedreich's Ataxia, respectively are well documented (Chung et al., 1993; Hirst et al., 1994; Imbert et al., 1996; Kunst and Warren, 1994; Montermini et al., 1997; Pulst et al., 1996)

The presence of interruptions in TNR arrays appears to strongly stabilise them (Kang et al., 1996; Maurer et al., 1998; Rolfsmeier and Lahue, 2000). The stabilisation caused by interruptions may be dependent on mismatch repair (Heale and Petes, 1995; Rolfsmeier et al., 2000). Physical evidence has shown that interrupted TNR arrays fold into less stable hairpins, slipped strand structures and quadruplexes compared to pure arrays (Gacy et al., 1995; Lyons-Darden and Topal, 1999; Pearson et al., 1998a; Weisman-Shomer et al., 2000).

The reason why interruptions stabilise TNR arrays is yet unclear. Perhaps interruptions provide an anchoring sequence, which helps the two strands of a TNR duplex to properly align, preventing replication slippage. Alternatively, interruptions could break up long unstable TNR arrays into a series of smaller stable arrays, preventing secondary structure formation. Finally, DNA mismatch repair may be more efficient when TNRs have interruptions (Section 1.8), providing a reference point in an otherwise homogenous stretch of DNA.

1.4 TNRs cause double strand DNA breaks

Large CTG and CGG TNRs tracts have previously been shown by genetic and physical analysis to be length-dependent sites of breakage in *Saccharomyces cerevisiae* during mitosis and meiosis (Balakumaran et al., 2000; Freudenreich et al., 1998; Jankowski et al., 2000; Richard et al., 2000). Interestingly, CGG TNRs appear to cause a greater level of chromosomal breakage (Balakumaran et al., 2000;
Freudenreich et al., 1998). This is probably a consequence of CGG TNRs forming more complex (or stable) secondary structures, perturbing transcription or replication, than CTG TNRs. Moreover, the use of drugs that slow down replication such as hydroxyurea, dramatically increased TNR induced rates of homologous recombination (HR). Stalled replication forks can be restarted via HR (Michel et al., 2001). This evidence strongly suggests a link between replication fork difficulties (Section 1.5) and double strand break formation when polymerases encounter TNR arrays. Perhaps the slower rate of polymerisation favours the formation of secondary structures.

However, it is unclear whether double strand breaks result in TNR instability. Unequal transfer or gene conversion between homologous DNA containing TNRs in human TRED patients has been observed using haplotype analysis (Brown et al., 1996; Losekoot et al., 1997; O'Hoy et al., 1993; Tsilfidis et al., 1992; van den Ouweland et al., 1994). In contrast, others disagree with unequal crossover as a mechanism for TNR expansions in Humans; observing no genetic exchange of markers between homologous chromosomes (Kovtun et al., 2001; Kremer et al., 1995; Leeflang et al., 1995; McMurray, 1995; Richards et al., 1991; Sutherland and Richards, 1992). All of these studies ignore sister-chromatid exchange as this is genetically silent.

Several studies have been carried out in order to address the role of recombination in TNR instability in different model organisms. Genetic exchange has been shown to destabilise long CTG tracts in mammalian cells (Meservy et al., 2003), meiotic yeast cells (Jankowski et al., 2000) and bacteria (Hebert et al., 2004; Napierala et al., 2002). In yeast, the 
\textit{SPO11} protein is responsible for meiosis-specific double strand DNA break formation. Most meiotic CAG repeat tract-length
alterations appeared to be SPO11 dependent, implicating that double strand breaks provoke TNR instability (Jankowski and Nag, 2002).

Many researchers have measured TNR instability in mutants for genes encoding key recombination proteins such as RecA in *E. coli* and RAD52 in *S. cerevisiae*. Short and expanded TNR arrays seem to be unaffected by RAD52 deletion (Freudenreich et al., 1998; Miret et al., 1997; Miret et al., 1998; Schweitzer and Livingston, 1999). In contrast, the absence of RecA, an essential protein in homologous recombination in *E. coli*, actually increases TNR instability (Hashem et al., 2002).

The formation of a double strand break (DSB) in or near a TNR array may lead to instability (Fig. 1.4c). This may result in the production of TNR containing 3’ recombinogenic ends, which go on to invade and anneal to a homologous chromosome or sister chromatid. However, accurate annealing to another TNR array may be difficult and could result in unfaithful annealing and TNR array expansion or contraction. Furthermore, the formation of secondary structures in the recombinogenic end may take place, leading to further instability.

**1.5 TNRs impede DNA replication fork progression**

It is clear from standard laboratory PCR analysis that DNA polymerases struggle traversing long TNR arrays. Indeed, two-dimensional gel analysis of replication fork progression through TNRs in plasmids, propagated in bacteria and yeast have shown that CGG, GAA and CTG TNRs hinder the progression of replication (Krasilnikova and Mirkin, 2004; Pelletier et al., 2003; Samadashwily et al., 1997). The efficiency of blockage depends on repeat length, purity and orientation relative to the plasmid origin. Furthermore, primer extension experiments using *Escherichia coli* DNA
polymerase I, T7 DNA polymerase and human DNA polymerase β have shown polymerase stalling in CTG, CGG and GTC TNRs (Kamath-Loeb et al., 2001; Kang et al., 1995b; Ohshima and Wells, 1997; Usdin and Woodford, 1995). Experiments in yeast have shown that defects in the replication machinery lead to an increase in TNR instability (Freudenreich et al., 1997; Gordenin et al., 1997; Kokoska et al., 1998; Maurer et al., 1996; Miret et al., 1998; Schweitzer and Livingston, 1998; Schweitzer and Livingston, 1999; White et al., 1999). Thus, polymerases traversing TNR arrays may pause or stall DNA replication. This could provide a molecular model for TNR instability (Fig. 1.4a). During pausing of DNA replication, dissociation of the primer from the template within a TNR array followed by improper realignment and formation of a loop structure may occur. This could result in polymerase slippage. Left unrepaired, this loop structure could result in either expansion if formed on the newly synthesised or deletion if formed on the template (parental) strand.

However there is though experimental evidence against this model. Firstly, if instability is a function of replication, cells that divide constantly, such as epithelial cells, may be expected to show elevated levels of instability. However, there is a lack of correlation between repeat instability and cell proliferation from murine and cell culture models (Gomes-Pereira et al., 2001; Lia et al., 1998; van den Broek et al., 2002). Secondly, if TNRs perturb replication fork progression, one might expect slower cell growth. It is known that cell lines harbouring expanded CTG repeats cultured from Myotonic dystrophy patients and the Dmt-D transgenic mouse model exhibit a selective growth advantage over isogenic repeat-free cell lines (Gomes-Pereira et al., 2001; Khajavi et al., 2001).
Figure 1.4 Molecular mechanisms of TNR instability

Individual TNRs are depicted by red rectangles. Any TNR expansions are shown in blue. (A) DNA polymerase slippage. During replication, DNA strands become separated and then re-anneal. In this time, small slipped strand structures may form resulting in misalignment. Depending on which strand these structures form (leading or lagging) will determine whether TNR expansion or contraction will occur. (B) Hairpin stabilisation by MMR. DNA gap formation can be initiated by a single-strand nick, which is then converted to a gap accompanied by the formation of a hairpin structure within the TNR array. The hairpin may be stabilised by Msh2 binding and expansion may occur via replication slippage and/or filling in of the gap and ligation. (C) DNA damage such as a double strand breaks or nicks are repaired by homologous recombination which could initiate TNR expansion. (D) Lagging strand Okazaki fragment processing. Synthesis of an upstream Okazaki fragment results in displacement of the 5' end of the downstream Okazaki fragment, generating a flap. The flap will be inefficiently excised if TNRs are present on this flap and fold into secondary structure such as a hairpin and incorporated into the newly synthesised strand, resulting in an expansion.
Chapter One: Introduction

1.6 Okazaki fragment processing may mediate TNR instability

During replication, TNRs may be more prone to folding into secondary structures on the lagging strand compared to the leading strand (Section 1.2, 1.5). Lagging strand DNA synthesis requires RNA-DNA containing Okazaki fragments that are subsequently processed and ligated together. Displacement synthesis generates a RNA-DNA flap that is removed by Flap Endonuclease 1 (FEN1) in mammals, Rad27 in yeast and the 5' → 3' exonuclease domain of DNA polymerase I in E. coli (Gordenin et al., 1997).

Deletion of the Rad27 in yeast results in an increase of CTG and CCG instability (Freudenreich et al., 1998; Schweitzer and Livingston, 1998; Spiro et al., 1999). Mutants of the 5’ to 3’ exonuclease domain of DNA polymerase I in E. coli show an increase in dinucleotide repeat instability (Morel et al., 1998). Biochemical evidence suggests that Okazaki fragments containing TNRs are resistant to FEN1 cleavage through the ability of the fragment to form secondary structures on the lagging daughter strand (Mirkin and Smirnova, 2002; Spiro et al., 1999).

From these data, it has been proposed that when TNRs happen to be present on the displaced 5’ flap of an Okazaki fragment, hairpin structures may form (Fig. 1.4d). These structures may be resistant to FEN-1 cleavage since FEN-1 only acts on ss-DNA and will stall at a hairpin or triplex structure. The displaced flap could then be processed in different ways. The hairpin in the Okazaki fragment could be unwound and allowed to ligate to the 3’ end of the next Okazaki fragment, resulting in an expansion. Alternatively, the stalling of FEN-1 on the lagging strand could result in other proposed mechanisms of TNR instability such as replication stalling (Section 1.11) and / or DSB formation (Section 1.4).
1.7 Mismatch repair may mediate TNR instability

The mismatch repair (MMR) system protects the genome from mismatched base-pairs, and repetitive DNA instability (Harfe and Jinks-Robertson, 2000). Also, inactivation of MMR results in increased instability of simple repetitive sequences such as mono- and di-nucleotides (Kolodner and Marsischky, 1999; Modrich and Lahue, 1996; Umar and Kunkel, 1996).

Mismatch repair in *E. coli* is encoded by the *mutSLH* system. Mismatches are bound by the mis-pair recognition protein, MutS. MutL and MutH are then recruited to the mismatched site initiating repair (Jiricny, 1998). In eukaryotes, Msh2 can form a heterodimer with either Msh3 or Msh6. Both Msh2/Msh3 and Msh2/Msh6 complexes are involved in the repair of single base insertion/deletions; whereas, Msh2/Msh6 repairs only mismatches.

From studies in bacteria, yeast and murine models, it is generally regarded that TNRs exhibit more small expansions or contractions when mismatch repair is deficient, somatically and between generations (Gomes-Pereira et al., 2004; Gomes-Pereira et al., 2001; Heale and Petes, 1995; Jaworski, 1995; Kovtun and McMurray, 2001; Miret et al., 1997; Rolfsmeier et al., 2000; Savouret et al., 2003; Savouret et al., 2004; Schmidt, 2000; Schweitzer and Livingston, 1997; van den Broek et al., 2002). This observation suggests that small slipped-strand structures are recognised and repaired by MMR. Indeed, biochemical evidence has shown that human Msh2 and murine Msh2-Msh3 complex bind to DNA containing TNRs in a structure specific manner (Owen et al., 2005; Pearson et al., 1997).

As mentioned previously, interruptions present in TNR arrays have a stabilising effect (Section 1.3). Some suggest that MMR prevents founder effects from occurring, preventing type II TNR instability. One study in yeast has shown that
interruptions do indeed stabilise small TNR arrays. However, this effect is lost in MMR deficient cells (Rolfsmeier et al., 2000). In another study, di-nucleotide poly(GT) tracts in yeast are dramatically stabilised by a single interruption and this is dependent on MMR (Heale and Petes, 1995).

However, although MMR deficiency may give rise to genome-wide microsatellite instability, a link between locus specific instability and MMR has not yet been observed in human cell lines with mutations in MMR (Kramer et al., 1996). Indeed, some murine models disagree with MMR stabilising TNR arrays, suggesting that the active repair of TNR arrays by MMR results in instability. Data from transgenic mice sperm cells, suggest that Msh2 actually exacerbates germline TNR expansions (Kovtun and McMurray, 2001). This is supported by two other studies that show wild-type Msh2 protein destabilises TNRs, resulting in small expansions (Manley et al., 1999; Wheeler et al., 2003).

A gap-repair model (Kovtun and McMurray, 2001) has been proposed to account for MMR mediated TNR instability (Fig. 1.4b). Although applicable to all types of cells, this model was initially used to explain the instability observed in post-meiotic germ cells from Huntington’s disease transgenic mice. As germ cells only have one copy of each gene, repair cannot proceed via genetic exchange with a homologous chromosome or a sister chromatid – unless that germ cell is undergoing DNA replication. Gap formation can be initiated by a single-stranded nick, which is then converted to a gap accompanied by the formation of a hairpin structure within the TNR array. The hairpin structure may be stabilised by Msh2 binding. Expansion may occur via replication slippage and / or filling in of the gap and ligation. If not cleaved or processed, the TNR array of one sister chromatid will increase in size after a round of replication.
1.8 The effect of SbcCD, MRX complexes on TNR instability

Closely spaced inverted repeat (or palindromic) DNA in organisms is either not recovered (inviability) or recovered with evidence of deletion in and around the palindrome (instability) (Leach, 1994). Instability is most likely due to the potential of palindromic DNA to adopt a hairpin or cruciform structure during or after DNA replication. *E. coli sbcC* or *sbcD* mutants can stably harbour long DNA palindromes (Connelly et al., 1997; Leach and Lindsey, 1986). The SbcCD complex is an ATP dependent exonuclease that cleaves hairpin DNA (Connelly et al., 1998) degrading it in a 3' to 5' direction (Connelly et al., 1999). SbcCD is proposed to generate double-strand breaks by cleaving a lagging strand hairpin structure. The break is then repaired by homologous recombination (Cromie et al., 2000; Leach, 1994).

Expansion-associated TNRs may have the ability to form secondary structures *in vivo* similar to those formed by palindromes, which could be cleaved by structure specific nucleases (Gacy et al., 1995). Therefore, proteins known to cleave palindrome hairpins may execute the same action on secondary structures formed by TNRs. In *E. coli*, Sarkar and collaborators document that long expansions of CTG repeats propagated in plasmids in a *sbcC* mutant strain grown at 25 °C. The observed "bimodal amplification" [two types of TNR expansion dependent on SbcCD] of repeats was lost upon plasmid complementation of *sbcC*. This observation has not been reproduced in the Leach laboratory (Rabaab Zahra, unpublished data) *S. cerevisiae* homologues of SbcC and SbcD, the Rad50 and Mre11 proteins, respectively, also modulate TNR instability during gene conversion (Richard et al., 2000). Double strand break repair (DSBR) was inhibited if long CTG TNRs were present on the homologous repair template. Over-expression of Mre11 alleviated this
inhibition of repair to that of a no TNR control. This suggests that the formation of secondary structures inhibits DSBR and that the excision or unwinding of such structures is needed to repair DSBs.

1.9 *Escherichia coli*

*E. coli* has been used as a model for studying TNR instability by several research groups. The genetics of basic cellular mechanisms such as replication, recombination and repair are especially well understood in *E. coli*. Also, precise genome alterations can be easily and quickly made. Therefore, the combination of bacterial genetics and recombinant DNA techniques provides a powerful set of tools for TNR research despite the evolutionary distance of *E. coli* from humans.

1.10 Homologous recombination in *Escherichia coli*

Homologous recombination (HR) is an important molecular system that ensures genome stability by permitting physical exchange between two double stranded DNA molecules that are homologous in sequence. This system repairs double strand DNA breaks (DSBs), single stranded gaps (SSGs) (Kowalczykowski, 2000); and ensures faithful DNA replication through the rescue of blocked, stalled or aborted DNA replication forks (Michel, 2000).

There are four steps involved in HR: initiation (presynapsis), homologous pairing (synapsis) and DNA strand exchange, DNA heteroduplex extension (branch migration) and resolution (Fig. 1.5a). HR is well understood in *E. coli* and will be the focus of this section.
Consider the formation of a double strand break (DSB) in the *E. coli* genome. RecBCD binds onto the blunt DNA end (Fig. 1.5b) and is required for almost 99% of recombination events at DSBs (Kowalczykowski, 2000). The RecB has 3' to 5' helicase and nuclease activity, RecD is a 5' to 3' helicase and RecD recognises the chi sequence (Singleton et al., 2004). RecBCD simultaneously unwinds and degrades the 3'-strand until it reaches a Chi site (χ) (5'-GCTGGTGG-3'). χ modifies the activity of the RecBCD complex, altering the nuclease activity of the enzyme, downregulating the degradation of the 3'-strand and up regulating the nuclease activity at the 5'-strand (Dixon and Kowalczykowski, 1993). RecBCD is highly processive and can, unwind dsDNA at 1 Kbp/sec, using 2 - 3 ATPs / bp (Roman and Kowalczykowski, 1989). RecBCD therefore produces a 3'-ssDNA tail on which protective single-stranded binding (SSB) proteins are removed and strand-exchange proteins, RecA, are actively loaded (Anderson and Kowalczykowski, 1997). The RecA-ssDNA nucleofilament finds and invades a homologous dsDNA molecule, a process not well characterised, producing a D-loop or heteroduplex DNA (hDNA) (Roca and Cox, 1997). The ss-DNA that is displaced
Chapter One: Introduction

Figure 1.5 Homologous recombination (HR) in E. coli.

(A) General double strand break (DSB) repair model of homologous recombination. Light yellow lines indicate newly synthesized DNA. The stages of homologous recombination are shown on the left; proteins involved in different stages are shown on the right (B) Double strand breaks are processed into a 3' single stranded recombinogenic end by the RecBCD / RecA pathway. RecBCD enzyme binds onto blunt double strand ends and degrades both strands until it encounters a chi site (x). Chi modifies RecBCD complex and results in attenuation of the 3' - 5' nuclease activity and activation of the 5' - 3' nuclease activity. RecBCD also facilitates the loading of RecA onto the 3' overhang. The resultant RecA-ssDNA nucleofilament is then able to invade homologous DNA (C) Single strandedaps repaired via the RecFOR / RecA pathway. The RecFOR complex encourages the loading of RecA onto ssDNA and limits its spread into dsDNA. RecA dependant strand exchange can then take place followed by branch migration and resolution of Holliday junctions.
in the formation of a D-loop is bound by single-stranded binding protein (SSB). This stabilises the D-loop and prevents the invading strand from being ejected. Invasion of the intact dsDNA homolog by each of the processed 3'- ssDNA ends need not occur simultaneously.

In a pathway alternative to RecBCD, the RecQ (3’ to 5’) helicase and RecJ (5’ to 3’) exonuclease can also stimulate RecA dependent recombination (Harmon and Kowalczykowski, 1998). In this pathway, RecJ is able to produce recombinogenic 3’ single stranded DNA ends together with the unwinding properties of RecQ. Although, in the presence of RecBCD, this pathway is unlikely to be the preferred one.

Alternatively, HR can be used in repairing single-stranded gaps (SSGs) (Fig. 1.5c). This repair mechanism takes place via the RecFOR pathway (Kowalczykowski, 2000). The RecFOR protein complex removes single-stranded binding protein which allows RecA to load, promoting strand exchange. RecF and RecR block the extension of the RecA nucleofilament from the ss-DNA region into ds-DNA. The recombinogenic DNA is then able to find homologous DNA and form a D-loop.

The next stage in either DSB or SSG repair is branch migration. Branch migration describes the physical exchange of one DNA strand of one molecule with that of the other. The Holliday junction (HJ) formed by the invading RecA-ssDNA filament is recognised by the protein complex RuvABC (West, 1996). A tetramer of RuvA binds to the HJ, holding the junction in a square planar conformation (Nishino et al., 1998). RuvA then allows the binding of two hexameric rings of RuvB that allows the ATP-dependent translocation of the HJ away from the point of initiation (West, 1997).
Chapter One: Introduction

After branch migration, the HJ can be cleaved in one of two ways at a specific DNA sequence (5' A/T TTTAGG/C 3') by the Holliday junction resolvase, RuvC (West, 1996). A RuvC dimer binds to the exposed face of the junction between the two RuvB rings and cleaves two DNA strands (Davies, 1992). Which two strands RuvC cleaves, will dictate whether cross-over (CO) or non cross-over (NCO) products are formed.

1.11 Replication Fork Reversal (RFR) in Escherichia coli

A number of pathways exist that allow the transformation of a stalled replication fork into a four-armed substrate. This process is known as replication fork reversal (RFR). RFR leads to restoration of the fork via degradation of the paired newly synthesised strands or recombination. The new fork is rescued by reassembly of the DNA replication machinery or 'replisome' and replication restart (Michel et al., 2004).

To date mutations in E. coli genes encoding replicative helicases, dnaB and rep (Michel et al., 1997); subunits of the replicative polymerase, dnaE and the β-clamp, dnaN (Grompone et al., 2002) and holD (Flores et al., 2001) have been used to study replication fork inactivation, which has been shown to occur by at least two different mechanisms: (i.) RecA dependent RFR in a dnaBts mutant at not permissive temperature (Seigneur et al., 2000) and (ii.) RecA independent RFR in all other replication mutants studied. All these mutants require the RecBC proteins for viability and undergo chromosome linearization in recBCts mutant backgrounds at high temperature suggesting the formation of double strand ends at blocked replication forks. RuvAB has been shown to act at arrested replication forks, cleaving the inactivated replication fork (in the presence of RuvC) resulting in DSBs and chromosome linearisation (Seigneur et al., 1998).
In the early stages of fork-reversal, newly synthesised strands anneal to one another along with annealing of the template strands. This reaction creates a four-way Holliday junction with a double-strand end, which is stabilised through RuvAB binding (Seigneur et al., 1998). The driving force of this early reaction is not clear although recent results implicate RuvAB itself in fork reversal in dnaEts, holD and rep mutants (Baharoglu et al., 2006) and UvrD is required for RFR to remove RecA protein which inhibits fork reversal in dnaEts and dnaNts mutants (Flores et al., 2005) but not in a dnaBts mutant (Seigneur et al., 2000).

Upon fork arrest in a dnaBts mutant, RecA may bind to the single-strand region of the lagging strand template and polymerise in the 5' to 3' direction (Fig. 1.6A RecA model). The RecA bound DNA is then able to perform strand exchange with the homologous leading strand resulting in a reversed fork. This is not the case in dnaEts, holD or rep mutants where replication fork reversal is unaffected by mutations in recA (Michel et al., 2004). It has also been suggested that RecG helicase following UV irradiation (McGlynn and Lloyd, 2000) and super-helical stress (Postow et al., 2001) may drive the annealing and re-annealing of nascent and parental DNA strands, respectively (Michel et al., 2001) (Fig. 1.6A RecG model and Supercoiling model).
Newly synthesised strands anneal at a block with re-annealing of template strands. This 4-way junction is stabilised by RuvAB binding.

RecBCD would normally degrade regressed DNA and reset reversed forks.

Four way Holliday junction formation with a DNA double-strand end

RecBCD degrades upto χ and loads RecA for homologous recombination

Target for replication restart

PriA-dependant replication restart

RuvC
Four models to account for the initial regression of the replication fork, initiating RFR: (I) RecA binds to the single ended region of the lagging strand template and polymerises in a 5' to 3' direction. This recombinogenic single ended region is then able to pair with the leading strand template allowing the newly synthesised strands to regress and anneal to one another, resulting in the formation of a Holliday junction that can be bound and stabilised by RuvAB. RecG binds to the replication fork and migrates towards the chromosomal replication origins, displacing the 5' end of the lagging strand, creating a four-way junction. (III) DNA supercoiling that has accumulated downstream of the fork on replication fork arrest is relaxed by the unwinding of the two newly synthesised strand from the template and their annealing. Purple circle represents an arbitrary protein complex that is inhibiting release of tensional stress. (III) RuvAB targets forks in vivo and directly converts replication forks into HJs. (Adapted with modifications from Michel et al. 2001). (B) Further processing of reversed forks. In (I) the replication fork is arrested causing it to form into a Holliday junction (drawn in two different ways, open X and parallel stacked X). (II) RecBCD initiates RecA-dependent RuvABC (again drawn in two ways). Alternatively (IV) if RecBCD encounters a HJ before it encounters a chi site, the stable strand end is degraded up to the HJ, restoring the fork structure. In both scenarios, replication restarts in a PriA-dependent manner. The absence of RecBCD (V) causes chromosome linearisation due to cleavage by RuvABC. Arrowheads represent the 3' end of a DNA strand. Continuous lines, parental chromosome; dashed lines, newly synthesised strands; small orange circle, RuvABC; outlined orange circle, RecBCD; blue circle, RecG.
From this point (Fig. 1.6B I), two distinct outcomes can occur. RecBCD will load onto the double stranded DNA end (Fig. 1.6B II) and degrade until it encounters either a chi site and initiates RecA-dependent homologous recombination (Fig. 1.6B III) or it will degrade up to the Holliday junction itself (Fig. 1.6B IV). Any remaining Holliday junction(s) are cleaved via RuvABC resolvase (Fig. 1.6b III). The fork structures created in both scenarios allow PriA-dependent replication restart. Alternatively in cells deficient for RecBCD, cleavage by RuvABC leads to chromosome linearization and cell death (Fig. 1.6B V).

1.12 UvrD and Rep helicases

DNA helicases are important enzymes involved in DNA metabolism such as replication, repair and recombination. Two important helicases in E. coli are UvrD and Rep, which are both implicated in RFR (Section 1.11). UvrD is a 3' to 5' helicase that is involved in the removal of DNA segments containing damaged or mutated bases in mismatch repair (Modrich, 1994) and nucleotide excision repair (Orren et al., 1992). UvrD may be closely associated with the replisome as it is regularly found in DNA Pol III protein preparations (Lahue et al., 1989). In vivo and in vitro studies have shown that UvrD functions as an anti-recombinase, disassembling RecA-ssDNA filaments (Veaute et al., 2005). This is in accordance with homologous recombination being elevated in a uvrd mutant and decreased in a strain over-expressing UvrD (Veaute et al., 2005). Furthermore, the expansion frequency of CTG, CAG and CGG TNRs in S. cerevisiae is elevated forty-fold in srs2 mutants, the closest orthologue to UvrD and Rep, suggesting that Srs2 is a potent and selective inhibitor of TNR expansions (Bhattacharyya and Lahue, 2004). Srs2 has been shown biochemically to selectively unwind duplex substrates containing CTG TNR arrays provided that Srs2
Chapter One: Introduction

encounters the TNR array immediately on entering the duplex (Bhattacharyya and Lahue, 2005).

Rep helicase also exhibits a 3' to 5' helicase activity and although less is known about its role, Rep is thought to act as an accessory helicase, removing obstacles in front of replication forks. Replication fork progression is two-fold slower in a rep helicase mutant compared to wild-type (Colasanti and Denhardt, 1987; Lane and Denhardt, 1975). This makes this mutant especially useful for studying the effects of slower replication progression through TNR arrays instead of using drugs such as hydroxyurea. Furthermore RFR is stimulated in rep mutants (Michel et al., 1997). Introduction of a recB or recC mutant into a rep background results in cell death (Uzest et al., 1995) and increased RuvABC dependent chromosome linearization in a recBCts mutant at high temperature (Michel et al., 1997). Together this suggests that replication stalling is more common in a rep mutant and/or the ability to remove protein blocks upstream of advancing replication forks is impaired.

1.13 This thesis

Long arrays of CTG and CCG TNRs have been shown to stimulate recombination in a length dependent manner in yeast (Balakumaran et al., 2000; Freudenreich et al., 1998). However, little is known about the nature of the event leading to stimulation of recombination. Furthermore the design of the yeast experiments has meant that selection for cells that undergo a recombination event results in loss of the TNR array. The subsequent growth of these recombinants in culture could give an inaccurate reading of TNR induced HR. To date, all the TNR research carried out in E. coli has been done in different types of plasmids, often of high copy number. Chapter three describes the insertion of different lengths of CAG
and CTG TNR arrays into the E. coli chromosome in the lacZ gene. Two homologous recombination substrate reporter cassettes, known as the zeocin repeats, were constructed and integrated approximately 6 Kbp on either side of the TNR arrays at lacZ. After I determined that recombination at the zeocin repeats was stimulated by double strand breaks, the system was used to address whether different lengths and orientations of CTG•CAG TNRs stimulate HR. To test whether secondary structure processing might generate recombinogenic DSBs, the role of the nuclease SbcCD on TNR induced HR was investigated. Finally I determine whether HR near lacZ modulates TNR instability. Pulse field gel electrophoresis (PFGE) was used in chapter four to attempt to detect TNR induced double strand breaks and the hypothesis that expanded TNRs stimulate HR via replication fork reversal (RFR) was tested genetically. Chapter four concludes with data that supports the hypothesis that the helicase RecG can abort RecA-mediated strand exchanges. Chapter five determines whether expanded TNRs stimulate the E. coli SOS response in wild-type and rep mutant cells. The technique, two dimensional neutral / neutral gel electrophoresis is set up and a replication y-arc is obtained.
CHAPTER TWO

Materials and Methods
2.1 Materials

2.1.1 General Reagents

2.1.1.1 Chemicals

Chemicals were purchased from the following suppliers, unless stated otherwise: Amersham, Fisher, Invitrogen, Melford Labs, Scotlab, Sigma.

2.1.1.2 Enzymes

Restriction enzymes were purchased from New England Biolabs, DNA polymerases were purchased from Roche, PFU and Herculase polymerases were purchased from Stratagene.

2.1.1.3 Growth reagents

Reagents for growth media were purchased from the following suppliers: Difco Laboratories, Gibco BRL, Oxiod and Sigma.

2.1.2 Bacterial Growth Media

2.1.2.1 General Information

All liquid growth media were autoclaved prior to use and stored at room temperature. For solid media, 1.5% (w/v) agar was added prior to autoclaving. After addition of drug/antibiotic and pouring, solid media were stored at 4°C in the dark.

2.1.2.2 Bacterial Media

Lauria-Bertani (LB) agar
10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 10 g NaCl and 15 g Bacto-agar (Difco). Made up to 1 L with distilled water and adjusted to a pH of 7.5 with NaOH and autoclaved.

**Low-Salt LB agar**

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 0.5 g NaCl and 15 g Bacto-agar (Difco). Made up to 1 L with distilled water and adjusted to a pH of 7.5 with NaOH and autoclaved.

**LC-agar**

10 g tryptone, 5 g yeast extract, 5 g NaCl and 10 g Difco-agar. Made up to 1 L with distilled water and adjusted to a pH of 7.2 with NaOH and autoclaved.

**LC top agar**

10 g tryptone, 5 g yeast extract, 5 g NaCl and 7 g Difco-agar. Made up to 1 L with distilled water and adjusted to a pH of 7.2 with NaOH and autoclaved.

**BBL-agar**

10 g trypticase-peptone, 5 g NaCl and 10 g Difco-agar. Made up to 1 L with distilled water and autoclaved.

**BBL top agar**

10 g trypticase-peptone, 5 g NaCl and 6.5 g Difco-agar. Made up to 1 L with distilled water and autoclaved.
**L broth**

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), and 10 g NaCl. Made up to 1 L with distilled water and adjusted to a pH of 7.5 with NaOH and autoclaved.

**Low-Salt LB agar**

10 g Bacto-tryptone (Difco), 5 g yeast extract (Difco), 0.5 g / 5 g NaCl. Made up to 1 L with distilled water and adjusted to a pH of 7.5 with NaOH and autoclaved.

### 2.1.2.3 Antibiotics

Antibiotics were added to liquid media immediately prior to use; whereas for solid media, antibiotics were added after autoclaving. Antibiotics or drugs were added to media with a temperature of ~55°C. All antibiotics were stored at -20°C.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Abbreviation</th>
<th>Solvent</th>
<th>Stock concentration (mg/ml)</th>
<th>Final concentration (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>Amp</td>
<td>H₂O</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Kan</td>
<td>H₂O</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>Tet</td>
<td>50% ethanol</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Zeocin</td>
<td>Zeo</td>
<td>H₂O</td>
<td>100</td>
<td>35</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Cm</td>
<td>100% ethanol</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Spectinomycin</td>
<td>Sp</td>
<td>H₂O</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

Ampicillin, GlaxoSmithKline; Chloramphenicol, Calibiochem; Kanamycin, Calibiochem; Spectinomycin, Sigma; Tetracycline, Calibochem; Zeocin, Invitrogen Corporation; Spectinomycin, Sigma

### 2.1.3 Commonly Used Buffers

All solutions were made up using distilled water and were either autoclaved (at 121 °C for 15 mins unless otherwise stated) or filter sterilised.

0.1 M or 0.5 M CaCl₂
Chapter Two: Materials and Methods

Made up fresh to 0.1 M or to 0.5 M in Milli-Q water and sterilised. Used for the introduction of DNA into bacterial cells or P1 transductions.

**1M Sodium Citrate**

Made up to 1 M in Milli-Q water and autoclaved. Stored at 4 °C.

**1M MgSO₄**

Made up to 1 M in Milli-Q water and autoclaved. Stored at room temperature.

**20 % (w/v) Glucose**

Made up to 20 % (w/v) using distilled water and autoclaved at 115 °C for 20 minutes.

**20 % (w/v) Sucrose**

Made up to 20 % (w/v) using distilled water and autoclaved at 115 °C for 20 minutes.

**1 M Tris**

Made up to 1 M using Milli-Q water and adjusted to desired pH using 10 M hydrochloric acid and autoclaved. Stored at room temperature.

**0.5 M EDTA (pH 8)**

Made up to 0.5 M using Milli-Q water and adjusted to desired pH 8.0 using 10 M NaOH and autoclaved. Stored at room temperature.

**5 M NaCl**

Made up to 5 M using Milli-Q water and autoclaved. Stored at room temperature.
1 mg/ml Ethidium bromide

Made up to 1 mg/ml using Milli-Q water and stored at room temperature in the dark. Diluted to 0.5 μg/ml in Milli-Q water for staining agarose gels.

80 % (v/v) Glycerol

Made up to 80 % (v/v) using Milli-Q water and autoclaved. Stored at room temperature.

2 % (w/v) 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactosidase (X-Gal)

Made up to 2 % (w/v) using dimethylformamide. Stored at –20 °C in the dark. Diluted to a working concentration of 2 x 10⁻³ % (w/v).

500 mg/ml Isopropyl-β-D-Thiogalactopyranoside (IPTG)

Made up to 500 mg/ml in sterile water. Stored at –20 °C. Diluted to a working concentration of 500 μg/ml.

50 mM Phenylmethanesulphonyl fluoride (PMSF)

Made up to 50 mM in isopropanol. Stored at –20 °C. Diluted to a working concentration of 1 mM.

Phage buffer

3 g KH₂PO₄, 7 g Na₂HPO₄, 5 g NaCl, 1 mM MgSO₄, 1 mM CaCl₂, 1 % (w/v) gelatine made up to 1 L with distilled water and autoclaved.
Chapter Two: Materials and Methods

50 x Tris-acetate (TAE)
242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8). Made up to 1 L with distilled water and stored at room temperature. Diluted to 1 x working concentration with distilled water.

5 x Tris-borate (TBE)
55 g Tris base, 27.6 g boric acid, 10 ml 0.5 M EDTA (pH 8). Made up to 1 L with distilled water and stored at room temperature. Diluted to 0.5 x working concentration with distilled water.

20 x SSC
350.64 g NaCl, 176.48 g Tri-sodium citrate (pH 7.5). Made up to 2 L with distilled water and autoclaved. Stored at room temperature.

TE buffer
2.5 ml 0.1 M EDTA and 1 ml 1 M Tris (pH 7.6). Made up to 500 ml with distilled water and autoclaved. Stored at room temperature.

TEE buffer
2.42 g Tris base, 74.4 g EDTA, 76.08 g EGTA (pH 8.5). Made up to 2 L with distilled water and autoclaved.

Depurination solution
12.5 ml 37 % HCl made up to 500 ml with distilled water. Stored at room temperature.
Chapter Two: Materials and Methods

**Denaturation solution**
40 g NaOH, 175.32 g NaCl made up to 2 L with distilled water. Stored at room temperature.

**Neutralisation solution**
121.1 g Tris, 175.32 g NaCl made up to 2 L with distilled water (pH 7.5). Stored at room temperature.

**Low stringency buffer**
100 ml 20 x SSC, 1 g SDS made up to 1 L with distilled water. Stored at room temperature. Any precipitated SDS was re-dissolved with gentle heating and mixing.

**High stringency buffer**
25 ml 20 x SSC, 1 g SDS made up to 1 L with distilled water. Stored at room temperature. Any precipitated SDS was re-dissolved with gentle heating and mixing.

**Maleic acid buffer**
23.2 g Maleic acid, 17.4 g NaCl made up to 2 L with distilled water. Adjust to pH 7.5 with NaOH. Stored at room temperature.

**Washing buffer**
Maleic acid buffer containing 0.3 % (w/v) Tween 20. Stored at room temperature.
Detection buffer

24.22 ml 0.1 M Tris-HCl, 11.68 ml 0.1 M NaCl made up to 2 L with distilled water. Adjusted to pH 9.5 with NaOH. Stored at room temperature.

Deoxynucleotide triphosphate (dNTP) mix

2 mM of dATP, dTTP, dGTP and dCTP (Roche) were mixed together in equal volumes in Milli-Q and aliquoted. Individual aliquots were stored at −20 °C.

2.1.4 Escherichia coli Strains

Strains that were used in this study are listed below. For all cloning procedures and propagation of plasmid DNA, XL1-Blue (Stratagene) cell was used. MG1655 was used for all other experiments. MG1655 is the E. coli strain sequenced by the Sanger Institute.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655</td>
<td>(DL)</td>
<td></td>
</tr>
<tr>
<td>1719</td>
<td>XL1-Blue - recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac, [F' proAB, lacP, lacZAM15, Tn10]</td>
<td>Stratagene</td>
</tr>
<tr>
<td>1675</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1751</td>
<td>MG1655 χKO</td>
<td></td>
</tr>
<tr>
<td>1777</td>
<td>MG1655 χKO, LacI</td>
<td></td>
</tr>
<tr>
<td>1786</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)50</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CAG)75</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95</td>
<td></td>
</tr>
<tr>
<td>2052</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)93</td>
<td></td>
</tr>
<tr>
<td>2079</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, recA::Cm</td>
<td></td>
</tr>
<tr>
<td>2080</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, recB::Km</td>
<td></td>
</tr>
<tr>
<td>2082</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, sbeC::Cm</td>
<td></td>
</tr>
<tr>
<td>2086</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, ΔsbeCD</td>
<td></td>
</tr>
<tr>
<td>2099</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, recG::Km</td>
<td></td>
</tr>
<tr>
<td>2147</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, ΔsbeCD</td>
<td></td>
</tr>
<tr>
<td>2205</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CAG)14</td>
<td></td>
</tr>
<tr>
<td>2250</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CAG)95</td>
<td></td>
</tr>
<tr>
<td>2252</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95 deletion event to (CTG)1</td>
<td></td>
</tr>
<tr>
<td>2258</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, lacZ::(CTG)95, recR::Km</td>
<td></td>
</tr>
<tr>
<td>2295</td>
<td>MG1655 χKO, LacI, cynX:zeo cassette χ+, recB::Km</td>
<td></td>
</tr>
</tbody>
</table>

Cold Spring Harbour
M. Masters
John Eykelenboom
John Eykelenboom
Rabaab Zahra
Rabaab Zahra
Rabaab Zahra
This work
This work
This work
This work
This work
This work
Rabaab Zahra
Rabaab Zahra
Rabaab Zahra
This work
This work
This work
This work
Rabaab Zahra
Rabaab Zahra
This work
This work
2.1.5 Bacteriophage λ Strains

The bacteriophage λ strains used in this study are listed below.
Table 2.3 Bacteriophage λ Strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>λDRL154</td>
<td>pal571, Asp6, cl857, χ⁺, 571 bp palindrome</td>
<td>This laboratory</td>
</tr>
<tr>
<td>λDRL152</td>
<td>pal571, Asp6, cl857, χ⁺</td>
<td>This laboratory</td>
</tr>
</tbody>
</table>

2.1.6 Oligonucleotides

Oligonucleotides that were used in this study are listed below. All oligonucleotides were synthesised and purchased from MWG Biotech.

Table 2.4 Oligonucleotides

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zeo-CF2a</td>
<td>TCG AGC TAG CTG GTG GTG CAG GGA ATC GGT TTT ATC</td>
</tr>
<tr>
<td>Zeo-CR2</td>
<td>AAA AAC TCG ACT ATG TAT TGG CCC GGG TAG TAT C</td>
</tr>
<tr>
<td>Zeo-Fla</td>
<td>TCC CTG CAC CAC CAG CTA GCT CGA GGG TGT GGA AAG</td>
</tr>
<tr>
<td>Zeo-R2</td>
<td>GTC GCT TCT GGG GTG CCC GGG GAA TTT AGA G</td>
</tr>
<tr>
<td>Zeo-CR1</td>
<td>TTG CCG GGG ATC CCC AGA CCT AAC CAC AC</td>
</tr>
<tr>
<td>Zeo-CF1</td>
<td>AAA AAG TCG ACT ATC AAA AAC TCG CCT GCT G</td>
</tr>
<tr>
<td>Zeo211</td>
<td>AAA AAG GGC CCT AAG TGA CTA GAA GGT GAC CAG ATG GCT TC</td>
</tr>
<tr>
<td>sdm-zeo_F</td>
<td>CGG GAC GCC TCC GGG CCC GCC ATG ACC GAG ATC G</td>
</tr>
<tr>
<td>sdm-zeo_R</td>
<td>CGA CCT CGG TCA TGG CGG CCC CGG AGG GGT CCT G</td>
</tr>
<tr>
<td>mhpC_8x⁵⁺</td>
<td>ACC AGC AGA TCT AAA AAT CTA GAT TTA ATC GAT GCG GAC GCC AGG G</td>
</tr>
<tr>
<td>mhpC_8x⁵⁺</td>
<td>AAA AAC TCG ACT TGG CCC GGG GAA GAT AAT GCC CTT GGC TAC G</td>
</tr>
<tr>
<td>LacIQ-CF</td>
<td>AAA AAG TCG ACC TTT TGG CCC GAA GAC G</td>
</tr>
<tr>
<td>LacIQ-CR</td>
<td>AAA AAC TCG ACT ATG AAT GAT CAC CCC ACT G</td>
</tr>
<tr>
<td>LacZ-chi-CF</td>
<td>AAA AAG TCG ACT CCC CAG TCC TGT ATG AAC G</td>
</tr>
<tr>
<td>LacZ-chi-IR</td>
<td>ATC CAT TTC ACT ATG GGT CAG ATG CAG GAT G</td>
</tr>
<tr>
<td>LacZ-chi-IF</td>
<td>CTG ACC ACT ATG GAA ATG GAT TTT TTC ATC G</td>
</tr>
<tr>
<td>LacZ-chi-OR</td>
<td>AAA AAC TCG ACT TTA CCC GCT CTC CTA CCT G</td>
</tr>
<tr>
<td>LacZ-chi-test-F</td>
<td>CGA AGT GAC CAG CAG ATA CC</td>
</tr>
<tr>
<td>LacZ-chi-test-R</td>
<td>ATC GCC ATT TGG CCA CCT TTA</td>
</tr>
<tr>
<td>Apal_SDM_R</td>
<td>CAC TAT ACC AAT TGA GAT GGG CCC GTC AAC GAT AAT TAC TAG TCC</td>
</tr>
<tr>
<td>Apal_SDM_F</td>
<td>GGA CTA GTA ATT ATC ATT GAC GGC CCC ATC TCA ATT GGT GTA GTC</td>
</tr>
<tr>
<td>SacI_SDM_R</td>
<td>GAA CTA GAT AGA CAA ATG GAG CTC CAT TGG TCT ATC TAG TCC</td>
</tr>
<tr>
<td>SacI_SDM_F</td>
<td>GAA CTA GAT AGA CAA ATG GAG CTC CAT TGG TCT ATC TAG TCC</td>
</tr>
<tr>
<td>MfeI_SDM_R</td>
<td>CAT CTC AAC TCG TCT TGG TGA TTT TAA TCA CTA TAC CAA CAG GTA G</td>
</tr>
<tr>
<td>MfeI_SDM_F</td>
<td>CCC ATC TCA GGT GGT GAT GAT GAT GAT AAG TAC ACC GGT TGA GAT G</td>
</tr>
<tr>
<td>BsaI_SDM_R</td>
<td>GTC TAT TGC TGG TAT CGG TAC CCG ACC TGC AGG</td>
</tr>
<tr>
<td>BsaI_SDM_F</td>
<td>GTC TAT TGC TGG TAT CGG TAC CCG ACC TGC AGG</td>
</tr>
<tr>
<td>BbsI_SDM_R</td>
<td>CCT GCA GGT CGG GTA CGG ATA CCC GCA CTA GAC</td>
</tr>
<tr>
<td>BbsI_SDM_F</td>
<td>CGA CCT CTG CTG CCC TTT CAT CCT CTA ATA AAT ACC</td>
</tr>
<tr>
<td>Lac1F</td>
<td>AAA AAC TCG ACT TGG TCC GGA TAT CTC GGT AGT GG</td>
</tr>
<tr>
<td>Lac1R</td>
<td>GAA GAC GCA ATT GTA GAC CAT GCT CAT ACG TGT TCC CTG TG</td>
</tr>
<tr>
<td>Lac2F</td>
<td>TTT TCC AGG TGA ATA ATT CCC GTC TGG CCT TCC TG</td>
</tr>
<tr>
<td>Lac2R</td>
<td>GGT CTC CTA TGG CGT CCT GGT CCT TTG ACA ACG TCG TGA TG</td>
</tr>
<tr>
<td>UvrD1F</td>
<td>AAA AAC TCG ACT GAC CTC GGT GAT ATA ATT A</td>
</tr>
<tr>
<td>UvrD1R</td>
<td>CGA GCA AGG GGC CGG ACC TCC GGG AGG AGT G</td>
</tr>
<tr>
<td>UvrD2F</td>
<td>AAA AAC TCG ACT GAT CGA TGG CAG GCT ATG C</td>
</tr>
<tr>
<td>UvrD2R</td>
<td>AAA AAC TCG ACT GAT CGA TGG CAG GCT ATG C</td>
</tr>
<tr>
<td>Rep1F</td>
<td>AAA AAC TCG ACT GAC CTC GGT GAT ATA TTG GCT A</td>
</tr>
</tbody>
</table>

44
2.1.7 Plasmids

Plasmids used during this work are listed below in table 2.5.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Features</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>(pDL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>327</td>
<td>pUC18</td>
<td>Leach lab</td>
</tr>
<tr>
<td>2911</td>
<td>pLacD2</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>2912</td>
<td>pLacD2 (CAG)_{98}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>2913</td>
<td>pLacD2 (CTG)_{98}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1603</td>
<td>pSV40/Zeo2</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>1604</td>
<td>pMH9</td>
<td>M. Masters</td>
</tr>
<tr>
<td>1605</td>
<td>pTOF24</td>
<td>Elise Darmon</td>
</tr>
<tr>
<td>1628</td>
<td>pTOF24 - sbcCD deletion vector</td>
<td>This work</td>
</tr>
<tr>
<td>1823</td>
<td>pLacD1 (L8 mutation)</td>
<td>This work</td>
</tr>
<tr>
<td>1828</td>
<td>pMH9-Lac{{P}}</td>
<td>This work</td>
</tr>
<tr>
<td>1830</td>
<td>pTOF24 - BAMBS</td>
<td>This work</td>
</tr>
<tr>
<td>1895</td>
<td>pLacD2 (CTG)_{6}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1900</td>
<td>pLacD2 (CAG)_{28}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1901</td>
<td>pLacD2 (CTG)_{28}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1911</td>
<td>pLacD2 (CAG)_{50}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1912</td>
<td>pLacD2 (CTG)_{50}</td>
<td>Rabaab Zahra</td>
</tr>
<tr>
<td>1998</td>
<td>pMH9 cynX::zeo\textsuperscript{cassette}</td>
<td>This work</td>
</tr>
<tr>
<td>2069</td>
<td>pMH9 mhpC::zeo\textsuperscript{cassette}</td>
<td>John Eykelenboom</td>
</tr>
<tr>
<td>2309</td>
<td>pScy\textsuperscript{K0}</td>
<td>John Eykelenboom</td>
</tr>
<tr>
<td>2341</td>
<td>pTOF24 - Rep deletion vector</td>
<td>This work</td>
</tr>
<tr>
<td>2342</td>
<td>pTOF24 - ΔcynX::zeo\textsuperscript{cassette}</td>
<td>This work</td>
</tr>
<tr>
<td>2391</td>
<td>pTOF24 - ΔavrD</td>
<td>This work</td>
</tr>
</tbody>
</table>
Chapter Two: Materials and Methods

2.2 Microbiological Methods

2.2.1 Growth of Bacteria

Overnight *E. coli* strains were created by streaking the appropriate strain from the -70°C glycerol stocks onto LB agar plates containing the appropriate antibiotic to obtain single colonies. A single colony was used to inoculate 5 ml of L Broth in a bijoux bottle. The culture was grown at the appropriate temperature overnight – usually 30 or 37 °C.

2.2.2 Preservation of Bacteria

*E. coli* strains were stored for short time periods on solid medium at 4 °C inverted and wrapped in Parafilm™. Strains were stored permanently at -70 °C in 40% (v/v) glycerol. Two glycerol stocks were created – one for ‘A’ stock and one for ‘B’ stock. 750 µl of cell culture was mixed with 750 µl of 80% sterile glycerol and mixed gently.

2.2.3 P1 Transduction

Upon infection of *E. coli* cells, phage P1 packages random fragments of the bacterial chromosome into phage particles (Margolin, 1987). Taking advantage of this behaviour, mutations in specific genes linked to antibiotic resistance genes were introduced into *E. coli* strains using P1 transduction.
Chapter Two: Materials and Methods

An overnight culture of the strain to be genetically altered was grown in LB broth containing 2.5 mM CaCl₂. 1 ml of this overnight was taken, spun down and resuspended in 100 µl of L Broth containing 250 mM CaCl₂. This was mixed with 100 µl of the P1 lysate containing the mutation of interest and incubated at 37 °C for 20 minutes. 800 µl of L Broth containing 500 mM Sodium citrate was added to stop further phage infection. Cells were allowed to recover at 37 °C for 1 hour. 100 µl of this culture was plated onto LB plates plus appropriate antibiotic. Colonies were purified away from phage by streaking onto full LB plates containing the appropriate antibiotic twice, sequentially. New mutants were tested phenotypically.

2.2.3.1 Production of P1 plate lysates

An overnight culture was diluted 1/10 into 10 ml of LB broth containing 2.5 mM CaCl₂ and grown at 37 °C for 2 hours with shaking. 200 µl of the freshly grown culture was mixed with 100 µl of diluted P1 lysate (10⁵ plaque forming units ml⁻¹) and incubated at 37 °C for 30 mins to allow for phage absorption. 2.5 ml of LC top agar containing 5 mM CaCl₂ was added and poured onto fresh LC plate(s) containing 5 mM CaCl₂. The plate(s) were incubated, not inverted, overnight at 37 °C. To practically achieve the phage titre required to make a useful P1 lysate, the P1 lysate used was diluted 10⁻⁴, 10⁻⁵ and 10⁻⁶ in phage buffer.

After incubation, the P1 dilution that just gives confluent lysis was taken. 5 ml of phage buffer was poured onto the plate and the top agar was scraped off and poured into a bijoux bottle containing 100 µl chloroform. The mixture was vortexed and then left to stand for 30 mins at room temperature. After centrifugation at full speed for 5 minutes in a bench-top centrifuge, the supernatant was poured into a sterile bijoux bottle. The P1 lysate was stored at 4 °C.
Sterile bottles used for storing P1 lysates should not be washed in detergent. This increases the ‘life’ of the lysate. The P1 lysate used to carry the new mutation of interest should not already harbour the same antibiotic resistance gene. A no phage control is useful when mixing cells and phage (cells only control).

2.2.4 UV irradiation

Ultra-violet (UV) radiation was used to phenotypically test mutations in genes involved in DNA repair such as recA, recBCD and recFOR. Although a Stratalinker™ is a useful tool for such tests, I preferred to use an UV lamp owned by Dr. Garry Blakely. The strain of interest was diluted $10^2$ to $10^6$ and 10μl of the dilutions were spotted onto LB plates in duplicate. One of the plates was irradiated for a period of time, typically 10 secs to 1 min and then wrapped in tin-foil immediately to prevent light activation of DNA repair pathways. Both plates were incubated over-night at 30 °C.

2.2.5 SbcCD phenotypic test

Bacteriophage lambda λDL154 harbours a palindrome, which is attacked by SbcCD when its genome is injected into the bacterium. λDL154 is therefore less infectious to E. coli that are wild-type for SbcCD. Indeed, compared to a no-palindrome phage, λDL152, λDL154 shows plaque cutback in a wild-type SbcCD. This serves as a useful phenotypic tool to test E. coli strains for SbcCD.

E. coli was grown overnight in 5 ml LB containing 0.2 % maltose and 5 mM MgSO₄. 250 μl of cells was mixed with 2.5 ml of BBL top and poured onto freshly poured BBL plates and allowed to set. The bacteriophage lambda strains, λDL154 and
\( \lambda \text{DL152} \) were diluted \( 10^{-2} \) to \( 10^{-8} \). 10 \( \mu l \) of these dilutions was spotted onto the BBL plates, allowed to dry and incubated overnight at 37 °C. Bacteriophage lambda is able to infect \( E. \text{coli} \) causing an area of cell lysis, known as a plaque.

2.2.6 Heat shock transformation of \( E. \text{coli} \)

An overnight culture of \( E. \text{coli} \) was diluted 1 ml into 50 ml of L-Broth and incubated for two hours at 37 °C. 1 ml of culture was transferred into 2 ml Eppendorf tubes (catalogue number: 0030 120.094) and spun on a bench top centrifuge for thirty seconds. The supernatant was discarded and the pellet was resuspended in 500 \( \mu l \) 0.1 M CaCl\(_2\) and left on ice for 30 minutes. The cell suspension was spun down at 4 °C and the supernatant removed. The presence of a hole in the cell pellet generally indicates that the cells have become competent. The pellet was then resuspended in 100 \( \mu l \) 0.1 M CaCl\(_2\). DNA to be introduced into the bacterium was added to the cell suspension and left for 30 minutes on ice then heat shocked for five minutes at 37 °C. 500 \( \mu l \) of LB was added and the cells were left for \(~1\) hour at the appropriate temperature before being plated on selective media. Super-competent XL1-Blue cells (Stratagene) were transformed according to the manufacturer's instructions.

2.2.7 pKO integration in \( E. \text{coli} \)

Precise alterations were made in the genome of \( E. \text{coli} \) using the method of Church and Link (Link et al., 1997). Integrations are based on the pKO plasmids, pMH9 and pTOF24 (Merlin et al., 2002). These plasmids have the following features: the chloramphenicol resistance gene, a temperature sensitive replication gene, \( \text{repA} \); that
is functional at 30 °C but not 37 – 42 °C and the sacB gene, which renders cells sensitive to the presence of sucrose in growth media (Fig. 2.1).

Into pKO plasmids, an 800 bp region of homology to the E. coli genome is cloned at PstI and SalI. This piece of DNA will contain any alterations that wish to be made. Such alteration(s) will have 400 bp of homology flanking either side. Often the DNA containing alterations will have been created using cross-over PCR (Section 2.3.14). The plasmid is transformed into recipient cells with chloramphenicol selection at 30 °C. Single colonies are streaked onto fresh chloramphenicol LB plates and grown overnight at 42 °C. This results in a mixture of very small colonies and larger healthier colonies. The larger colonies have integrated the plasmid into the genome. These integrants are purified onto new chloramphenicol LB plates and grown overnight at 42 °C. A colony is picked and grown overnight in 5 ml LB at 30 °C. This culture is diluted 10^-6 in fresh LB and plated onto LB agar containing 5 % sucrose. This selects for the excision of the plasmid from the chromosome. Sucrose resistant colonies are ‘patch-tested’ simultaneously onto LB containing 5 % sucrose and LB containing 5 % sucrose and chloramphenicol. Chloramphenicol sensitive colonies are tested for the required alteration via polymerase chain reaction (PCR) and sequencing.
Figure 2.1 pKO integration and excision
(A) pMH9 and pTOF24 have been used in this thesis to modify the *E. coli* genome. pMH9 carries the chloramphenicol resistance gene and pTOF24 carries both the chloramphenicol (CmR) and kanamycin resistance genes, although the kanamycin resistance gene is removed upon cleavage with *PstI* and *SalI* during cloning. Both vectors carry the temperature sensitive gene, *repA*, for DNA replication (ts ori) that is functional at 30 °C (indicated with 30 °C shaded in blue) but not at 42 °C (indicated with 42 °C shaded in red). After introducing the modified DNA region (in this instance the red striped DNA region flanked by two 400 bp homology arms shown in orange and blue. Not to scale) to be integrated into the *E. coli* genome into either pTOF24 or pMH9, the recombinant plasmid was introduced into *E. coli* cells. (B) Plating cells at 42 °C on chloramphenicol selects for pKO vectors that have integrated into the homologous DNA region (linear DNA shown below the circular vector) in the *E. coli* genome. Cells that have integrated the vector are now sensitive to the sugar sucrose. (C) Cells can undergo either one of two excision events (retaining the original DNA or acquiring the new, modified DNA) from the chromosome when plated on LB plates containing 5% sucrose. Cells whose DNA has been modified can be selected for by screening with PCR.

2.2.8 Fluctuation analysis

From -70 °C glycerol stock, strain(s) were streaked onto an LB plate and grown overnight at 37 °C to obtain single colonies. Twenty-four bijoux bottles containing 5 ml of low salt L Broth were inoculated with twenty-four different colonies of the same strain and grown overnight at 37 °C to stationary phase. Each overnight culture was diluted 10⁻² and 10⁻⁵ with 100 μl plated on low salt LB plates containing 35 μg/ml
zeocin and LB plates respectively. The plates were spread with glass beads. LB plates were incubated at 32 °C and LSL zeocin plates were incubated in a single layer at 37 °C in the dark overnight. Colonies were counted and recombination rates were calculated (Spell and Jinks-Robertson, 2004).

2.3 Nucleic Acid Methods

2.3.1 Phenol/Chloroform Extraction

Nucleic acids were purified away from protein in solution by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1), vortexing briefly and centrifuging at maximum speed on a bench top centrifuge for 4 minutes. The upper phase containing the nucleic acid was removed to a fresh tube.

2.3.2 Precipitation of Nucleic Acid

Nucleic acids were precipitated from solution by adding 2x volumes of 100% ethanol and 0.1x volumes of 3 M sodium acetate (pH 5.2). This was kept at -20 °C for 30 minutes. Nucleic acids were then ‘pelleted’ by centrifugation in a bench top centrifuge for 30 minutes at 4 °C. The pellet was washed in 70 % ethanol, then dried and resuspended in an appropriate volume of TE buffer.

2.3.3 Mini Preparation of Plasmid DNA by Spin Column

Plasmid DNA was prepared using the QIAprep spin miniprep kit (Qiagen) following the manufacturer’s instructions. DNA was routinely extracted from 5ml of E. coli culture grown to stationary phase and eluted in 25 – 50 μl of TE and stored at -20°C.
2.3.4 Large Scale Preparation of Plasmid DNA by Filtration Column

Plasmid DNA was prepared using the QIAfilter gigaprep kit (Qiagen), following the manufacturer's instructions. DNA was extracted from 500 ml of *E. coli* culture grown to stationary phase and eluted in 250 μl of TE and stored at -20°C.

2.3.5 *E. coli* Genomic DNA Preparation

Genomic DNA was prepared using the Wizard Kit (Promega), following the manufacturer's instructions. DNA was extracted from 5 ml of *E. coli* culture grown to stationary phase and eluted in 50 μl of TE and stored at -20°C.

2.3.6 Restriction Digestion of DNA

DNA was digested with restriction endonucleases (New England Biolabs or Roche) typically in volumes of 15 - 150 μl. Reactions contained the following: DNA, appropriate restriction digest buffer (X1) and 2 to 5 units of the required enzyme. Digestions were incubated at the appropriate temperature between 2 to 5 hours. Digested DNA was analysed on 1% agarose gel electrophoresis (section) or extracted (section) for further work.

2.3.7 Agarose gel electrophoresis

DNA was analysed for size, quantity or quality on 1 % (w/v) agarose gels. Gels were prepared by melting agarose in 500 ml of X1 TAE buffer in a microwave (Panasonic). This solution was kept at 60 °C for up to 3 weeks without loss of band separation resolution. Usually 5 μl of DNA to be analysed would be added to 2 μl of 60%
glycerol and was loaded directly into the gel wells. A potential difference was supplied across the gel so that the DNA fragments separated relative to their size (90 to 200 volts). The DNA was stained by immersing the gel in a solution containing ethidium bromide for 10 mins followed by rinsing in water to remove excess ethidium bromide. The DNA was then visualised using an UV trans-illuminator. DNA markers were purchased from NEB, allowing the size of DNA bands to be determined.

2.3.8 Purification of DNA from Agarose Gels

DNA fragments were purified from a population of differently sized DNA molecules by separating DNA fragments by agarose gel electrophoresis (section 2.3.2.5). The desired DNA band was isolated from the agarose gel using the QIAquick gel extraction kit (Qiagen) according to the manufacturer’s instructions. DNA was eluted in 30 – 50 μl of TE and stored at –20 °C.

2.3.9 Ligation of DNA molecules

Ligations were performed in a final volume of 10 μl, containing the following: ~20 ng of vector DNA and 3 times the number of moles of insert DNA, X1 ligation buffer and 200 units of T4 DNA ligase (NEB). Reactions were allowed to proceed at 16 °C overnight and stopped by heat inactivation at 70 °C for 15 minutes. Alternatively, ligations were performed using Quick Ligase Kit (NEB) according to the manufacturer’s instructions.

2.3.10 Dephosphorylation of DNA molecules
To prevent digested vector DNA from re-ligating, the 5' phosphates were removed using calf intestinal phosphatase (CIP) (NEB). This decreases the proportion of clones that have the original vector without insert. Reactions were performed in a final volume of 10 μl containing the following: ~20 ng of vector DNA, X1 CIP buffer and 200 units of CIP. Reactions were allowed to proceed at 37 °C for one hour exactly. CIP enzyme was removed either via DNA precipitation (section) or purification through a PCR purification column (section). Prolonged incubations with CIP can result in DNA ends being degraded.

2.3.11 Annealing of oligonucleotides

Oligonucleotides to be annealed were mixed in roughly equimolar amounts, around 40 μg, in 1x TE buffer containing 10 mM NaCl. The mixture was heated to 100 °C in a beaker containing 1L of water over a bunsen burner. The bunsen burner was switched off and the oligonucleotides were allowed to slowly cool down to room temperature. The annealed oligonucleotides were diluted 100 - fold in 1x TE buffer containing 10 mM NaCl and stored at 4 °C.

2.3.12 Amplification of DNA by the Polymerase Chain Reaction (PCR)

Regions of DNA were specifically amplified using the polymerase chain reaction (PCR). A typical PCR reaction using plasmid or genomic DNA as template was as follows:
DNA polymerases used were either Taq DNA polymerase (Roche) or Pfu DNA polymerase (Stratagene) or Herculase (Stratagene). PCRs were carried out in either a Hybaid PCR Express or an Applied Biosystems GeneAmp PCR System 2400 programmed in the following manner:

95°C for 5 minutes (1 cycle)
95°C for 15 seconds
X°C for 15 seconds (see text below)
72°C for 1 minute/1 Kbp of DNA to be amplified (30 cycles)
72°C for 10 minutes (1 cycle)
4°C hold.

The annealing temperature of the oligonucleotides was calculated in the following manner. The melting temperature of the oligonucleotides was supplied by MWGBiotech. The annealing temperature was calculated as lower of the two oligonucleotides melting temperatures less 5 °C.

### 2.3.13 E. coli Boiled Cells

Often boiling cells produced DNA of adequate quality for PCR and sequencing. A large colony was picked up with a sterile yellow tip, resuspended in a PCR tube containing 30 µl MilliQ™ and boiled in a PCR machine for 10 minutes. 1 µl typically would be used for PCR. The use of wooden toothpicks for boiled cell PCR is not recommended.
2.3.14 Cross-over PCR

Cross-over PCR is a useful technique for constructing homology arms for pKO integrative vectors such as pTOF24 and pMH9 (Ho et al., 1989) (Fig. 2.2). Often mutations can be introduced into DNA during construction. This prevents a round of site-directed mutagenesis after cloning.

Briefly, two 400 bp PCR products are created with two pairs of primers. The two sets of primers have the following features: the first forward and second reverse primer have a restriction site added to the end followed by five A nucleotides for cloning. The first reverse and the second forward primer have 20 bp of homology to one another.

After PCR amplification using these two pairs of primers the two PCR products are purified to remove any unused primers and are mixed together in a PCR reaction together with only the first forward and the second reverse primer. After initial denaturation and annealing, the two PCR products anneal at the 20 bp of homology. This fused piece of DNA is then amplified with the two flanking primers and cloned into either pTOF24 or pMH9 for pKO integration (Section 2.2.7)
Figure 2.2 Cross-over PCR
Cross-over PCR is a useful technique for creating homology arms for cloning into pKO integrative vectors such as pTOF24 and pMH9. This procedure effectively joins two PCR products together using overlap extension and negates the need for DNA ligation. (A) Two sets of primers are created for two individual PCRs. The first forward primer (‘forward’ in this context is from left to right across the page) and second reverse primer contain restriction endonuclease sites (shown in light blue) for cloning and the first reverse primer and second forward primer (shown in red and blue) have 20 bp of homology to one another. (C) After PCR amplification and removal of unincorporated primers these two PCR products are mixed together and in another PCR reaction with the first forward and second reverse primers. After denaturation and annealing the two PCR products anneal to each other at the 20 bp of homology and are subsequently amplified and cloned into a pKO vector (D).

2.3.15 Site-Directed Mutagenesis (SDM)

Small mutations or insertions or deletions were inserted into plasmids using site-directed mutagenesis (SDM) following Stratagene™ QuikChange© Site Directed Mutagenesis Kit recommendations. Although its recommendations were followed, the aforementioned kit was never actually purchased.

2.3.16 Genemapper analysis

Genemapper analysis was carried out the University of Edinburgh Sequencing Service using the ABI 3730 genetic analyser (Applied Biosystems). This system automatically
detects and determines the sizes of PCR products based on electrophoretic separation in a capillary. Known size standards are run in parallel enabling sizing of PCR fragments. The PCR products are labelled with a fluorescent label that is excited when passed through a laser beam and registered when fragments pass a certain point of the capillary.

2.3.16.1 Sample preparation

DNA containing TNRs was amplified for Genemapper analysis using two primers, one standard and one fluorescently labelled with 5'FAM. The PCR products were diluted 1 in 100 in MQ water and 1 µl of the diluted PCR was mixed with 9 µl of Hi-Di™ Formamide containing the GeneScan™-500 LIZ™ size standard. 1 µl of the size standard was added to 500 µl of Hi-Di™ Formamide. Samples were analysed in batches of either 48 reactions (half reactions) or 96 reactions (full reactions) by the University of Edinburgh sequencing service.

2.3.16.2 Data analysis

The data collected from the ABI 3730 Genetic Analyser was analysed using the GeneMapper® software version 3.7. The fragments are visualised as peaks on a graph. The abscissa is the time x-axis representing with labelled fragments have passed the laser. The size standard is used to convert the time x-axis to a size scale allowing the sizing of the PCR fragments.
2.3.17 Purification of PCR products

PCR generated DNA fragments were purified from primers, unincorporated nucleotides, polymerases and salts using the QIAquick PCR purification kit (Qiagen), following the manufacturer’s instructions. Purified DNA was eluted in 20 - 50μl of TE and stored at -20 °C.

2.3.18 DNA sequencing

Plasmid DNA or PCR products to be sequenced were prepared using QIAprep spin columns. DNA was roughly quantified by visualisation on an agarose gel. Sequencing reactions were performed using a dRhodamine terminator cycle sequencing kit (Perkin Elmer). A typical reaction mix was as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA</td>
<td>variable</td>
</tr>
<tr>
<td>Terminator Mix</td>
<td>2 μl</td>
</tr>
<tr>
<td>Primer</td>
<td>1.6 pmol</td>
</tr>
<tr>
<td>MilliQ water</td>
<td>upto 10 μl</td>
</tr>
</tbody>
</table>

Reaction DNA concentration was dependent on size and type of DNA being sequenced – manufacturer’s guidelines were followed. The thermocycler was programmed to carry out the following:

- 96 °C for 30 seconds
- 50 °C for 15 seconds
- 60 °C for 4 minutes for 35 cycles
- 4°C hold.

10 μl of MilliQ water was added to the reaction mix. Samples were run on an ABI PRISM 377 DNA sequencer and analysed using ABI View computer program.
2.3.19 Pulse Field Gel Electrophoresis (PFGE)

Pulsed Field Gel Electrophoresis (PFGE) is a technique used to separate especially long strands of DNA by length. The technique can be used to detect fragmentation of bacterial chromosomes as only linear chromosomes may enter the gel, with intact and σ-shaped molecules remaining in agarose plugs in the wells (Seigneur et al., 1998). To prevent random mechanical shearing of large DNA molecules, bacterial cells are embedded in agarose plugs, gently lysed and digested with a rare-cutting restriction enzyme.

Strains of interest were grown from a single colony to an OD₆₀₀ of 0.6 in 50 ml LB broth. 5 ml of culture was spun down for 10 mins at 3500 rpm and resuspended in TEE solution to give an OD₆₀₀ of 0.9. 350 μl of this was mixed with 350 μl of 2% ultra pure low melting point agarose (GIBCO) at 56 °C and pipetted into disposable Biorad CHEF plug moulds. The agarose was left to set, sitting on ice. The plugs were placed in 10 ml lysozyme solution to lyse the cells (0.05 % N-lauroylsarcosine, 5 mg/ml lyzosyme in TEE buffer) and incubated at 37 °C for 2 hrs with gentle shaking. The lysozyme solution is replaced with 5 ml proteinase K solution (1 mg/ml proteinase K, 1% SDS in TEE buffer) and incubated o/n at 55 °C with gentle shaking.

The plugs were rinsed in 10 ml TE buffer (2 mM Tris-HCl pH 8, 0.5 mM EDTA) three times for 1 hr. at room temperature with gentle shaking. After rinsing, the plugs were washed in PMSF solution (10 mM Tris-HCl, 1 mM EDTA, 1 mM PMSF) (1 ml/plug) for two, 1 hour washes. PMSF inhibits the proteinase K for the subsequent restriction digest. The plugs are then rinsed in 10 ml TE buffer twice for 30 mins. Afterwards the plugs were stored in fresh TE buffer at 4 °C. Plugs can be stored for around six months.
Single plugs were taken and equilibrated in 1 ml x 1 relevant restriction buffer and left for 1 hr at r/t for subsequent restriction digest. The buffer was replaced with 350 μl restriction digest mixture containing 2.5 μl restriction enzyme per plug and left for 4.5 hrs at the relevant temperature. The NEB catalogue contains useful information regarding the digestion of DNA embedded in agarose plugs.

The plugs were then inserted carefully into the wells of a 1 % ultra pure AquaPor™ ES ultra high gel strength agarose (National Diagnostics). Care was taken not to leave any air bubbles in the gel wells. The wells were then covered with extra 1 % agarose and ran in a Biorad PFGE apparatus with the following parameters: initial switch time, 5 secs; final switch time, 30 secs; run time, 17 hrs; volts, 160 V; temperature, 8 °C. The gel was stained with ethidium bromide and studied under an UV transilluminator.

2.3.20 Neutral / Neutral 2-D Agarose Gel Electrophoresis

Two-dimensional gel electrophoresis allows the visualisation of replication intermediates through a region of DNA of interest. This method was used to monitor progression of the replisome through TNR arrays integrated at lacZ. DNA is resolved by mass in the first dimension and by mass and shape in the second dimension (Fig. 2.3). This results in the separation of γ-arc intermediates that is detected via Southern blotting. Any pausing or stalling of the replication fork leads to the appearance of a bulge on the γ-arc.
Figure 2.3 2D gel electrophoresis
Schematic representation of electrophoretic separation of replication intermediates. (A) *E. coli* chromosome undergoes bi-directional DNA replication from the origin, through lacZ, towards the terminus where it ends. (B) Diagram of a γ-arc. The large dot (1x) shows the monomer DNA fragment and the arc extending upwards from the monomer fragment represents different replication intermediates. (C) Distribution of replication intermediates between two EcoRI restriction sites. Since the origin is not present within the restriction fragment, a γ-arc rather than a bubble arc should be observed.

A single colony was grown in 25 ml LB to an OD$_{600}$ of 1.0. 1 ml was taken and spun down in a bench-top centrifuge at full speed. The supernatant was discarded and the pellet resuspended in 200 μl PIV buffer (10 mM Tris (pH 8.0), 1 M NaCl). The cells were then spun down and resuspended in 100 μl PIV buffer. Cells were placed in a 50 °C water bath and 100 μl of 1% agarose dissolved in PIV buffer cooled to 50 °C was added.

The cell suspension was then pipetted into BIORAD plug moulds and left to set on ice. The plugs were then placed into Eppendorf tubes containing 1 ml EC-lysis buffer (6 mM Tris (pH 8.0), 1 M NaCl, 100 mM EDTA, 0.2 % deoxycholate, 0.5 % lauryl sarcosine freshly supplemented with 50 μg/ml RNaseA) and incubated at 37 °C for 2 hrs. The EC-lysis buffer was removed and 1 ml of ESP buffer (500 mM EDTA, 1 % lauryl sarcosine, freshly supplemented with 0.1 mg/ml proteinase K) was added and
incubated over-night at 50 °C. The ESP buffer was removed and the plugs were washed five times for 30 minutes with 10 ml TE buffer. Plugs were then stored at 4 °C until required.

Individual plugs were taken and equilibrated in 500 µl EcoRI restriction enzyme buffer for 30 mins at room temperature. The buffer was discarded and 150 µl of fresh enzyme buffer containing EcoRI restriction enzyme was added. The restriction digest lasted overnight. A 0.35 % agarose gel in 1x TBE was poured and allowed to set in a cold room with an appropriate gel comb. Digested plugs were placed into separate wells, topped up with any remaining melted agarose and ran at 2.2 V/cm for 16 hrs along with a size standard marker. Afterwards, the gel was stained with ethidium bromide and observed briefly under a longwave (not shortwave) transilluminator.

The position of the DNA fragment of interest was determined using the DNA standard marker. The gel lane was excised from the gel 1 cm below the DNA fragment and 1 cm above double the size of the DNA fragment. The gel slices were removed, positioned 90 ° to the original direction and ran in 0.95 % agarose containing 0.3 µg/ml ethidium bromide for 10 V/cm in 1x TBE + 0.3 µg/ml ethidium bromide at 4 °C for about 4 ½ to 5 hrs. Progression of the DNA through the gel was monitored using a hand-held long wave transilluminator. DNA was then transferred to Hybond N+ (Amerham) and blotted with an appropriate probe.

2.3.21 Southern Blotting of PFGE / 2D gels

DNA transfer to nylon membranes, construction / hybridisation of digoxigenin-labelled probes and chemiluminescent detection of probes on a blot was carried out using materials purchased from Roche. Manufactures instructions were followed. For
random priming the Amersham Rediprime™ random prime labelling system was used.

2.4 Microscopy Methods

2.4.1 GFP Microscopy

GFP microscopy was used to assay for the SOS response using the plasmid pGB150 (a gift of Dr. Garry Blakely). Cells were grown from single colony in 50 ml of LB media containing chloramphenicol to an OD_{600} = 0.6 (mid- to late-log phase). 1 ml of cell culture was spun down and re-suspended in 100 μl of LB media containing chloramphenicol and 10 μl was spotted onto an 1% agarose (dissolved in water) slab and examined under a Zeiss Axiovert 200 inverted microscope using the following settings: bright-field (exposure 802 ms), Z-sectioning ± 2 μm, Semrock GFP filter number 1 and a neutral density filter number 2 (exposure 500 ms). Images were processed on a Dell computer using Metamorph version 6.3r2 software.

2.5 Statistical Methods

Data obtained from fluctuation analysis were analysed statistically using two different methods. Mutation rates were calculated according to (Lea and Coulson, 1949). Rankings for interval calculations based on the number of cultures assayed was found using Table B11 from Practical Statistics for Medical Research (Altman, 1990). Statistical advice was provided by Jill Sales (Biomathematics and Statistics Scotland, University of Edinburgh).
2.6 Computer Analysis

*E. coli* gene sequences were obtained from Colibri. DNA sequences were manipulated using GeneJockeyII software (BioSoft) and VectorNTI.
CHAPTER THREE

Expanded CTG TNRs stimulate HR in *E. coli*
3.1 Introduction

As discussed in Chapter 1, CTG TNRs stimulate recombination in a length dependent manner in yeast (Freudenreich et al., 1998). Freudenreich and colleagues have concluded that recombination is stimulated by TNRs causing DSBs. However the mechanism of stimulation is not fully understood. In order to further elucidate this phenomenon, I have developed an assay for recombination rates in the *E. coli* genome in the presence of TNRs.

This chapter describes the integration, in the chromosome of the bacterium *Escherichia coli*, of differently sized CAG and CTG TNRs into the start of the lacZ gene and of a zeocin resistance recombination reporter substrate into the nearby gene, cynX. I show that TNRs stimulate recombination at cynX in a length dependent manner. Furthermore, stimulation of recombination is dependent on TNR orientation with respect to the origin of replication. Experiments indicate that recombination is reduced in *E. coli* mutants for double strand break repair (*recA* and *recB*); but not for gap repair (*recR*). TNR induced stimulation of recombination is shown to be independent of the DNA hairpin nuclease SbcCD arguing against any role of secondary structure formed by the TNRs. Since TNRs are unstable DNA elements and as cells that have undergone HR events near lacZ can easily be selected for using the zeocin resistance assay, I wished to investigate expanded CTG TNR instability in zeocin resistant and sensitive cells.

3.2 Introduction of TNRs into the start of lacZ gene in the *E. coli* chromosome

The development of a new technology has allowed precise genome engineering in *E. coli* (Link et al., 1997). We have used this system, to integrate CAG•CTG and
CTG•CAG TNR arrays into the beginning of the \textit{lacZ} gene in the \textit{E. coli} genome. Since a great deal is known about the biology of the \textit{lac} operon, of which \textit{lacZ} is a member, this system is especially well-defined.

3.2.1 Construction of plasmid, pLacD1

The plasmid pLacD1 was created to insert TNR arrays into the beginning of \textit{lacZ} in the \textit{E. coli} genome using a 800 bp region of \textit{lacZ} homology. pLacD1 is based on the plasmid pTOF24, related to pMH9 (Fig. 3.1), both of which are designed for \textit{E. coli} chromosome engineering (Section 2.2.7). Before inserting a 800 bp region of \textit{lacZ} homology (Fig. 3.2) into pTOF24, a series of restriction sites were removed using site directed mutagenesis.

\textbf{Figure 3.1 Plasmids pMH9 and pTOF24}

Plasmids pMH9 and pTOF24 used for \textit{E. coli} chromosome engineering. pMH9 and pTOF24 both contain a temperature sensitive replication gene (\textit{repA'}) that is non-functional at 42 °C, the \textit{sacB} gene that renders cells sensitive to the sugar sucrose and the chloramphenicol resistance gene (Cm\textsuperscript{R}). pTOF24 also contains the kanamycin resistance gene (Km\textsuperscript{R}) that is removed when DNA fragments are cloned at the restriction sites PstI and SalI.

The plasmid pTOF24 belongs to a family of pKO integration vectors and contains the following features: a temperature sensitive \textit{repA} gene whose product is functional at 30 °C but not 37-42 °C; the antibiotic resistance genes for kanamycin and chloramphenicol antibiotics and the \textit{sacB} gene, which renders cells sensitive to
sucrose. Into this plasmid, I intended to clone a DNA fragment with 800 bp region of lacZ homology using the restriction enzymes _PsyI_ and _SalI_ (Fig. 3.2). This cloning would remove the kanamycin resistance gene.

![Figure 3.2 LacZ homology arms](image_url)

(A) Sequence of 800 bp region at lacZ of wild-type _E. coli_. The start of the lacZ gene is shaded blue and the ATG start codon is underlined. The top and bottom red coloured text indicates the primer Lac1F and Lac1R respectively. _Lac_ operon CRP binding site in green. (B) Modified version of (A) containing _BsaI_, _MfeI_ and _BbsI_ restriction sites after ATG start codon of lacZ to be cloned into a modified version of pTOF24. This sequence contains the L8 mutation (green underlined A). Centrally coloured bases indicated regions where oligonucleotides anneal for cross-over PCR.

The 800 bp lacZ DNA fragment was constructed using cross-over PCR (Section 2.3.14) (Ho et al., 1989). Using CSH100 (DL844) as template DNA, primers Lac1F and Lac1R were used to create one homology arm; while primers Lac2F and Lac2R were used to create the other. Primers Lac1F and Lac2F were used to amplify the fusion of the two PCR products. In the middle of the homology arms is the start codon of the lacZ gene. The cross-over PCR introduced three restriction enzyme sites, _BsaI_, _MfeI_ and _BbsI_ 10 base pairs downstream to this start codon without altering the reading frame of lacZ (Fig. 3.3) (Section 3.2.2).

---

**Figure 3.2 LacZ homology arms**

(A) Sequence of 800 bp region at lacZ of wild-type _E. coli_. The start of the lacZ gene is shaded blue and the ATG start codon is underlined. The top and bottom red coloured text indicates the primer Lac1F and Lac1R respectively. _Lac_ operon CRP binding site in green. (B) Modified version of (A) containing _BsaI_, _MfeI_ and _BbsI_ restriction sites after ATG start codon of lacZ to be cloned into a modified version of pTOF24. This sequence contains the L8 mutation (green underlined A). Centrally coloured bases indicated regions where oligonucleotides anneal for cross-over PCR.
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Figure 3.3 Introduction of BsaI, MfeI and BbsI into the 5' region of lacZ
(A) First 50 nucleotides of the lacZ gene. LacZ start codon is underlined. BsaI, MfeI, and BbsI are shown in colour. Red coloured nucleotides indicates changed bases using site-directed mutagenesis.

CSH100 (DL844) contains the L8 mutation in the cyclic AMP (CAP) binding site upstream of the lacZ start codon. This mutation was therefore introduced into the lac homology arms during cloning. This mutation lowers maximal levels of the lac enzymes fifteen fold thereby reducing leakiness of transcription (Scaife and Beckwith, 1966). This mutation would be useful for another study investigating the effects of transcription on TNR instability.

The restriction enzyme sites for MfeI, BbsI and BsaI were present in the plasmid, pTOF24 (Fig. 3.1) and were removed using site-directed mutagenesis (SDM) before inserting the 800 bp of lacZ homology. MfeI was removed using primers, MfeI_SDM_F and MfeI_SDM_R; BsaI was removed with primers, BsaI_SDM_F and BsaI_SDM_R and BbsI was removed with primers, BbsI_SDM_F and BbsI_SDM_R.

For a collaboration involving single molecule analysis of TNR containing DNA, the restriction sites for ApaI and SacI were introduced into pTOF24. ApaI was introduced using primers, ApaI_SDM_F and ApaI_SDM_R; SacI was introduced using primers, SacI_SDM_F and SacI_SDM_R. These sites were introduced to linearise the plasmid with the TNRs positioned roughly in the middle of the restriction fragment. All alterations to pTOF24 were confirmed via restriction enzyme digestion and DNA sequencing. After five rounds of site-directed mutagenesis the 800 bp lacZ fragment was cloned using PstI and SalI. This modified plasmid was named, pLacD1 (Fig. 3.4).
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Figure 3.4 Plasmid pLacD1
Plasmid pLacD1 after five site directed mutagenesis (SDM). The DNA sequence shown in Fig. 3.2 B is cloned between the restriction sites PstI and SaI. After removing BbsI (1101 bp) from pLacD1 (Section 3.2.2) different lengths of CTG and CAG TNRs were introduced in pLacD2 by Rabaab Zahra.

3.2.2 Collaboration with Rabaab Zahra
At the beginning of my Ph.D. project, TNRs from human patients of Huntington’s disease were to be cloned into the MfeI site of pLacD1 and integrated into the E. coli genome. The MfeI restriction site present in the 800 bp of lacZ homology was used to allow cloning of TNR DNA fragments containing EcoRI sticky ends. Although MfeI and EcoRI sticky ends are compatible, after ligation to each other they cannot be re-cut with MfeI or EcoRI. Before transformation into E. coli, ligations were digested with MfeI or EcoRI to remove plasmids that were not recombinant. This strategy was used because initially a large range of TNRs available to us could be easily cloned using EcoRI. The BsaI and BbsI restriction sites could be used for cutting out cloned TNRs arrays, for instance into other vectors.

I was unable to clone PCR fragments containing TNRs from the Huntington locus into the plasmid pLacD1. Rabaab Zahra started a project attempting to grow TNRs. She began by modifying the plasmid, pLacD1 replacing the MfeI site in pLacD1 with a small CTG or CAG TNR array using site-directed mutagenesis.
Rabaab Zahra also removed a BbsI site from one of the homology arms of pLacD1 via SDM. She named this plasmid, pLacD2.

The method used to expand the size of the TNRs in pLacD2 will be described briefly. Two pLacD2 plasmids containing a small TNR array were digested: one cut with BsaI and HindIII and the other with BbsI and HindIII. BbsI and BsaI digest within the TNRs. After gel extraction, she could ligate the two TNR containing halves of the plasmid together. After a period of time a collection of differently sized CAG and CTG TNR tracts existed in pLacD2. These were then subsequently integrated into the E. coli chromosome (DL1786) at lacZ (Table 2.2) (Section 3.4).

### 3.3 Construction of DL1786

Before integrating TNRs into MG1655 (DL1675) the following genome modifications were made: a naturally occurring chi site (χ) was removed (creating DL1751), the lacIq mutation was introduced (creating DL1777) and a reporter system designed to measure recombination levels, known as zeocin repeats (Section 3.4), was integrated at the cynX gene (creating DL1786).

Between the beginning of lacZ, zeocin and the zeocin repeats is a naturally occurring χ site (5'-GCTGGTGG-3'). To ensure that RecBCD paused at the χ site near the zeocin repeats and not at the χ site in lacZ, the latter site was removed from DL1675 using the plasmid, pDL2309 (Fig. 3.5) (constructed by John Eykelenboom) creating DL1751. This alteration would promote the creation of a recombinogenic end near the zeocin repeats integrated at the cynX gene allowing zeocin recombination (Section 3.4).
Figure 3.5 Plasmid pDL2309
Plasmid pDL2309 created to remove a naturally chi site present in the lacZ gene. Removal of this chi (χ₀) (1514 bp) site (replaced by an SpeI restriction endonuclease site) would promote zeocin recombination at cynX.

pDL2309 is a pMH9 derivative plasmid carrying an 800 bp lacZ fragment containing an altered χ sequence in the middle (5'-ACTAGTGG-3') cloned using PstI and SalI. This altered χ sequence introduced a SpeI restriction enzyme site. The two homology arms were amplified from DL1675 using cross-over PCR and the primer pairs LacZ-chi-OF, LacZ-chi-IR and LacZ-chi-IF, LacZ-chi-OR respectively. The two homology arms were fused using the primers LacZ-chi-OF and LacZ-chi-OR during cross-over PCR. A 20 bp region of homology used to fuse the two homology arms introduced the altered χ site. The removal of the χ site from DL1751 was checked via PCR using the primers, LacZ-chi-test-F and LacZ-chi-test-R and digestion with SpeI.

To tightly control transcription through lacZ the lacI₀ mutation was introduced. This mutation would be useful for other research on TNR instability. The lacI₀ mutation, causes ten-fold over expression of LacI, the lacZ repressor protein (Muller-Hill et al., 1968). This mutation was introduced into DL1751 using pDL1828 (Fig 3.6) creating DL1777. pDL1828 is a pMH9 derivative carrying an 800 bp fragment containing the LacI₀ mutation cloned using PstI and SalI. The 800 bp fragment was PCR amplified using primers LacIQ-CF and LacIQ-CR, from the strain
CSH100 (DL844). Introduction of the $lacI^2$ mutation was confirmed by DNA sequencing.

**Figure 3.6 Plasmid pDL1828**
Plasmid pDL1828 created to introduce the $lacI^2$ mutation which results in ten-fold over expression of the Lacl repressor protein for $lacZ$. This mutation would allow greater control of transcription through $lacZ$ and was used for other studies (R. Zahra) on the effect of transcription of TNR instability.

### 3.4 Construction of the zeocin recombination reporter substrate

In order to assay recombination rates, a zeocin recombination reporter substrate (known as zeocin repeats) was constructed and integrated in the $cytX$ gene in DL1777 (Fig. 3.7 A). The substrate consists of two incomplete copies of the gene that confers resistance to the drug zeocin. The first copy is 275 bp in length and is missing the last 100 nucleotides of the gene. The second copy is 369 bp and is missing the ATG start codon and three nucleotides (GCC) directly following the ATG (Fig. 3.7 B). Recombination between the two zeocin repeats can result in one functional copy of the zeocin drug resistance gene, allowing growth on zeocin containing media (Fig. 3.7 A / C). The replication fork progresses through the zeocin repeats at $cytX$ before reaching TNRs, 6.3 Kbp away at $lacZ$ (Fig. 3.9 A).
Chapter Three

Expanded CTG TNRs stimulate HR in *E. coli*

(A) The zeo recombination cassette consists of two incomplete copies of the zeocin resistance gene separated by three stop codons. The first copy is 275 bp in length and is missing 100 nucleotides at the 3' end of the gene (indicated by asterisk). The second copy is 369 bp in length and is missing an ATG start codon and three nucleotides directly after (indicated by asterisk). The replication fork progresses through the zeo repeats at cynX before reaching TNRs, 6.3 Kbp away at lacZ. Recombination during replication can result in zeocin resistance. (B) Sequence of the zeo repeats. An EM7 and a SV40 promoter allow RNA transcription in *E. coli* and eukaryotic cells, respectively. The first and second zeocin copies are termed zeocin α and zeocin β, respectively. The SV40 promoter and poly-adenylation signal allow the use of the zeocin repeats in eukaryotic cells. (C) Zeo_CF1 and Zeo_CR2 PCR across zeocin resistant (ZeoR) and zeocin sensitive (ZeoS) *E. coli* colonies. (D) The plasmid pDL1603 [pSV40/Zeo2 (Invitrogen)], from which the zeocin repeats were amplified, is a pUC based vector that is ampicillin resistant.

To prevent translation into the second zeocin repeat three stop codons (TAA, TGA and TAG) were introduced between the two copies. An EM7 and a SV40 promoter allow RNA transcription in *E. coli* and eukaryotic cells, respectively (Invitrogen). 129 bp after the second zeocin repeat is a SV40 poly-adenylation sequence (Fig. 3.7 D). The SV40 promoter and poly-adenylation signal will allow the use of the zeocin repeats in eukaryotic cells for future studies. Located directly 5' to the SV40 promoter is a chi (χ) sequence (Fig. 3.9 A). This was engineered to ensure that recombinogenic ends would be preferentially created at the zeocin repeats (Section 3.3).
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

The zeocin repeats were introduced into the gene \( \text{cynX} \) in DL1777 using the plasmid pDL1998 (Fig. 3.8) creating DL1786. pDL1998 consists of the zeocin repeats flanked by two 400 bp homology arms to the gene \( \text{cynX} \) cloned into the pKO integrative vector pMH9. The first \( \text{cynX} \) homology arm was amplified using the primers, Zeo-CF2a and Zeo-CR2; the second with Zeo-CF1 and Zeo-CR1. The first zeocin repeat (zeo \( \alpha \)) was amplified from the plasmid, pDL1603 (Fig. 3.7 D) using Zeo-F1a and Zeo-R2. The primer Zeo-F1a introduces a \( \chi \) site in the front of the first zeocin repeat.

![Figure 3.8 Plasmid pDL1998](image)

Figure 3.8 Plasmid pDL1998
Plasmid pDL1998 used to integrate the zeocin repeats into the origin of replication proximal \( \text{cynX} \) gene.

The primers Zeo-CF2a and Zeo-F1a have homology to one another as does Zeo-CR1 and Zeo-R2. All three PCR fragments were fused together using cross-over PCR and the primers Zeo-CF1 and Zeo-CR2. This longer PCR product was digested with the restriction enzymes \( \text{PstI} \) and \( \text{SalI} \) and cloned into pMH9. An \( \text{ApaI} \) restriction site was introduced into the first zeocin copy, at 276 bp, using primers sdm-zeo_F and sdm-zeo_R. This alteration was a silent mutation. The second zeocin copy (zeo \( \beta \))
was created using primers Zeo-F2ii and ZeoR2. After digestion with Apal, the second zeocin copy was cloned into the first using ApaI. The strategy for engineering an Apal site and the cloning of the second zeocin repeat was devised by John Eykelenboom and carried out by myself.

![Diagram of trinucleotide repeats integration](image)

**Figure 3.9 Integration of trinucleotide repeats (TNRs) into DL1786**

(A) Trinucleotide repeats were integrated into the 5' region of lacZ using pLacD2 plasmid derivatives containing different lengths of CAG and CTG TNRs. A Chi site was removed from the lacZ gene (strike-through symbol) and a zeocin recombination cassette containing a Chi site (χ symbol) was integrated upstream of cynX. (B) Ex_test_F and Ex_test_R PCR across the differently sized CTG TNRs in E. coli: (CNG)ₐ 148 bp; (CTG)₁₄, CTG₂₈, 232 bp; 190 bp; (CTG)₃₀, 232 bp; (CTG)₄₅, 442 bp; (CTG)₁₄₀, 568 bp. Strain numbers can be found in Chapter Two: Materials and Methods.

Into the strain DL1786 (Fig. 3.9 A) Rabaab Zahra integrated pLacD2 plasmid derivatives containing different lengths of CAG and CTG (Table 2.2 and 2.5) TNR arrays (Fig. 3.9 B). Integrating TNRs did not result in the transfer of the L8 mutation, from pLacD2, into the chromosome (Section 3.2.1).
3.5 Zeocin recombination requires RecA and RecB but not RecR

Initial experiments were carried out in order to characterise the effect of major recombination pathways on zeocin recombination. It was predicted that strains mutated in major recombination pathways would have decreased recombination. The TNR-free, wild-type control strain, DL1786, was assayed for zeocin recombination (Fig. 3.10). The frequencies of zeocin recombinants were determined (Section 2.2.8) and recombination rates were derived by fluctuation analysis (Section 2.5). The \textit{recA} mutant exhibited a 5.6-fold lower recombination rate compared to wild-type. This is in accordance with previous work reporting that \textit{recA} mutants are recombination defective (Birge and Low, 1974; Veaute et al., 2005). The residual level of zeocin recombination suggests that a \textit{recA} independent mechanism for generating zeocin resistant recombinants exists such as replication slippage, sister-chromosome exchange-associated slippage or single-strand annealing (Bzymek and Lovett, 2001). The \textit{recB} mutant exhibited a 2.2-fold lower recombination rate compared to wild-type. However, a \textit{recR} mutant, involved in gap repair, did not shown any defect in recombination. The greater reduction in zeocin recombination in \textit{recA} and \textit{recB} mutants compared to \textit{recR} suggests that zeocin recombination is induced by a DNA double strand break, producing an end that is repaired via the double break repair (DSBR) pathway (\textit{recA, recBCD}), rather than gap repair pathway (\textit{recA, recFOR}).
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Figure 3.10 Recombination rates in recA, recB and recR mutants.
Effect of recA, recB and recR mutations on zeocin recombination (TNR-free). The strains used were DL1786 (None cynX), DL2855 (1786 recA), DL2295 (1786 recB), DL2824 (1786 recR). The results are the means of three independent experiments. Error bars show 95% confidence intervals.

3.6 TNRs stimulate recombination in E. coli in a length dependent manner

To determine whether CTG TNRs stimulate recombination in E. coli, the recombination rate was assayed in a variety of strains harbouring different lengths of CTG TNRs on the leading strand template using fluctuation analysis (Fig. 3.11). The zeocin repeats were integrated into the gene cynX. Compared to the no repeat control, the rate was 2, 2.3, 2.4 and 2.5-fold greater for CTG\textsubscript{95}, CTG\textsubscript{140}, CTG\textsubscript{160} and CTG\textsubscript{184}, respectively revealing that expanded TNRs stimulate zeocin recombination. Furthermore, a deletion of CTG\textsubscript{95} to CTG\textsubscript{31} restored recombination to that of the no repeat control. In contrast the smaller CTG\textsubscript{14}, CTG\textsubscript{28} and CTG\textsubscript{50} did not stimulate recombination. This suggests a threshold length between CTG\textsubscript{50} and CTG\textsubscript{95} is
required to stimulate zeocin recombination. Compared to the no repeat control, the rate is equal for CTG_{50}; CTG_{14} and CTG_{28} were 1.6 and 1.5-fold lower, respectively. The reason for the slightly lower recombination rate in CTG_{14} and CTG_{28} is unclear.

In comparison to Freudenreich et al. 1998 we do not see as great a stimulation caused by the longer TNRs. They show that a CTG_{130} array stimulates recombination 9-fold over the TNR-free control; whereas in this study CTG_{184} only stimulates zeocin recombination 2.5-fold. Also in contrast to the yeast data, the data presented here suggest that although TNR-induced zeocin recombination is length-dependent, the rate of recombination is not directly proportional to TNR length. During this study, the effect of a 246 bp palindrome in \textit{lacZ} on zeocin recombination was investigated by Join Eykelenboom. The palindrome stimulated zeocin recombination 124-fold compared to a no palindrome control (e.g. DL1786) in a wild-type \textit{E. coli} strain (Data not shown). This would suggest that the stimulation of zeocin recombination by TNRs is not especially high and that whatever events that lead to zeocin resistant recombinants occur at a relatively low frequency.
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Recombination rate in the presence of tri-nucleotide repeats

Figure 3.11 Long CTG trinucleotide repeats stimulate HR
Effect of different lengths of CTG TNRs (CTG in the Okazaki fragment) on zeocin recombination. The strains used were DL1786 (None cytX), DL2347 (CTG14), DL2052 (CTG28), DL1984 (CTG30), DL2009 (CTG95), DL2305 (CTG140), DL2709 (CTG163), DL2728 (CTG184) and DL2252 (CTG95 deletion event to CTG31). The results shown are the means of three independent experiments (around 72 measurements). Error bars show 95% confidence intervals.

3.7 Recombination in different recombination mutants harbouring TNRs
Although I have established that maximal zeocin recombination depends on RecA and RecBCD rather than RecFOR, we wished to determine if any of these pathways would modify TNR-induced zeocin recombination. Since RecA and RecB are required for maximal zeocin recombination but not RecR, we expected that removing RecA and RecB would decrease CTG95 zeocin recombination and removing the RecFOR gap repair pathway would not. Strains harbouring CTG95 were assayed to determine recA, recB and recR independent recombination (Fig. 3.12). Although in the presence of CTG95 recA and recB mutants show a general increase in the number of zeocin resistant colonies compared to the no TNR control, they show 2.61 and 3.13-fold lower levels of recombination respectively compared to wild-type CTG95. In
the presence of CTG₉₅, recR mutants also exhibited an increase in the number of
zeocin resistant colonies compared to the no TNR, recR control and displayed a
modest 1.2-fold decrease in recombination compared to wild-type CTG₉₅. Again it is
clear, even in the presence of TNRs, that the rate of zeocin recombination is more
affected by RecA and RecBCD than RecR.

![Graph showing recombinant rates in recA, recB, and recR genetic backgrounds]

**Figure 3.12 Recombination rates in recA, recB and recR mutants.**
Effect of recA, recB and recR mutations on zeocin recombination with or without TNRs. The
strains used were DL1786 (None, cynX), DL2009 (CTG₉₅), DL2855 (1786 recA), DL2079
(CTG₉₅, recA), DL2295 (1786 recB), DL2080 (CTG₉₅, recB), DL2824 (1786 recR), DL2258
(CTG₉₅, recR). The results are the means of three independent experiments. Error bars show
95% confidence intervals.

### 3.8 TNR induced recombination is independent of SbcCD nuclease

The bacterial SbcCD and eukaryotic MRX complexes have been reported to be
implicated in TNR instability (Chapter 1). Furthermore, these complexes are known
to cleave secondary structures such as hairpins formed by palindromes. Therefore
proteins known to cleave palindrome hairpins may therefore execute the same action
on secondary structures proposed to be formed by TNRs. One prediction could be that
secondary structures such as hairpins may be cleaved by SbcCD in *E. coli* causing DNA breaks that stimulate zeocin recombination. To test this hypothesis, the SbcCD operon was deleted from DL1786 (No TNR) and DL2009 (CTG₉₅) using pDL1628 and assayed for zeocin recombination (Fig. 3.13). Introduction of the *sbcCD* mutation slightly decreases zeocin recombination in all strains assayed. However, TNR induced zeocin recombination appears to be no more influenced by SbcCD than recombination in the absence of TNRs. We conclude that the SbcCD complex is probably not able to cleave proposed secondary structures from by TNRs and that TNR.

![Figure 3.13 TNR induced recombination is independent of SbcCD](image)

*sbcCD* mutants were created in DL1786 (CNG₀) (creating DL2086) and DL2009 (CTG₉₅) (creating DL2147) and assayed for zeocin recombination. Introduction of the *sbcCD* mutation slightly decreases zeocin recombination in all strains assayed. However, TNR induced zeocin recombination appears to be independent of SbcCD. The results shown are the means of three independent experiments (around 72 measurements). Error bars show 95% confidence intervals.
3.9 TNR mediated recombination may be orientation dependent.

To determine whether TNR stimulation of recombination is orientation dependent, different lengths of CAG TNRs on the leading strand template were integrated into DL1786 and assayed (Fig. 3.14). In contrast to CTG repeats on the leading strand template stimulating recombination (Section 3.6); CAG repeats on the leading strand template appear not to. Unfortunately, I was unable to make a direct comparison between CTG and CAG orientations because the largest CAG TNR available was CAG\textsubscript{86}. Interestingly whilst the CTG orientation gives a higher recombination rate compared to the CAG orientation; investigations into TNR instability have shown that it is the CAG orientation (CAG on the leading strand template) that is more unstable (Rabaab Zahra – data not shown).
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Figure 3.14 TNR induced recombination may be orientation dependent
Effect of different lengths of CAG (CAG on the leading strand template) TNRs on zeocin recombination. The strains used were DL1786 (None cynX), DL2205 (CAG14), DL2347 (CTG14), DL2250 (CAG50), DL1994 (CTG50), DL1995 (CAG75), DL2639 (CAG94) and DL2009 (CTG95). The results shown are the means of three independent experiments (around 72 measurements). Error bars show 95% confidence intervals.

3.10 HR near lacZ does not stimulate CTG TNR instability.
I wished to determine whether CTG140 TNR instability differed in cells that had undergone a recombination event near lacZ compared to cells that had not. Taking advantage of the zeocin assay (Section 3.4), the instability of CTG140 TNRs was compared between cells plated on low-salt LB plates containing the drug zeocin (zeocin resistant colonies) and those plated on LB plates (zeocin sensitive colonies) using Genemapper™ analysis (Fig. 3.15). Previously (Freudenreich et al., 1998) analysis of TNR tracts known to stimulate HR was not possible as the HR event involved the deletion of the TNR tract. Using the zeocin assay, deletion of the TNR tract does not take place during HR, permitting TNR length analysis.
Chapter Three Expanded CTG TNRs stimulate HR in E. coli

Figure 3.15 System to determine TNR instability in zeocin resistant cells
(A) Location of primers Ex_test_F and Ex_test_R used to PCR amplify 148 bp of flanking lacZ DNA and (CTG)\textsubscript{140} (567 bp). The Ex_test_R primer is fluorescently labelled with FAM. (B) A standard zeocin assay was performed (Section 2.2.8) and zeocin resistant and sensitive cells were picked from zeocin and LB plates respectively and assayed for TNR length using Genemapper analysis (Section 2.3.16). This diagram was originally drawn by R. Zahra and modified by myself.

Work previously carried out by Rabaab Zahra identified that the CAG TNR orientation was, interestingly, more unstable compared to the CTG orientation. In agreement with this observation the level of instability observed in both zeocin resistant and sensitive CTG\textsubscript{140} cells was very low. Data from two individual assays measuring TNR length in 120 colonies was pooled and out of 240 zeocin sensitive, CTG\textsubscript{140} colonies analysed 9 had deleted in TNR size from CTG\textsubscript{140} to CTG\textsubscript{108}, CTG\textsubscript{103}, CTG\textsubscript{52}, CTG\textsubscript{97}, CTG\textsubscript{86}, CTG\textsubscript{42}, CTG\textsubscript{120} and CTG\textsubscript{85} and from 240 zeocin resistant CTG\textsubscript{140} colonies analysed 4 had deleted in TNR size to CTG\textsubscript{131}, CTG\textsubscript{130}, CTG\textsubscript{95} and CTG\textsubscript{114}. From this it was concluded that HR near TNR did not increase levels of CTG instability.

3.11 Conclusions

A collaborative effort between myself and Rabaab Zahra has resulted in the integration of differently sized CAG (CAG on the leading strand template) and CTG (CTG on the leading strand template) TNRs into the beginning of the lacZ gene in E. coli. To investigate the effects differently sized CAG and CTG TNRs have on HR, a
zeocin recombination substrate reporter was integrated into the gene 6.3 Kbp upstream (relative to direction of replication fork) at the gene cynX. Analysis of HR in different recombination mutants suggests that zeocin recombination is stimulated by a double strand break repair pathway, involving RecA and RecBCD, rather than a gap repair pathway, involving RecR. Strains harbouring differently sized CAG and CTG TNRs were assayed for rates of HR using the zeocin assay. Expanded CTG TNRs (CTG in the Okazaki fragment) were found to increase the rate of HR compared to a no repeat control in agreement with a previous study (Freudenreich et al., 1998). However expanded CTG TNRs did not increase HR as much as previously reported in yeast (Freudenreich et al., 1998). Indeed the rate of HR, although increasing with CTG length, seemed to plateau around a two-fold increase. This raises the issue whether the rate of TNR induced HR is proportional to TNR length. TNR instability (Sarkar et al., 1998) and the efficiency of a TNR containing DNA template to repair a broken homologous DNA region (Richard et al., 2000) have been suggested to involve the bacterial SbcCD complex and eukaryotic Mre11/Rad50/Xrs2 (MRX) complex, respectively. Moreover, TNRs are proposed to form secondary hairpin structures similar to those known to be formed by palindromes in E. coli. In E. coli these secondary structures are cleaved by the structure specific endonuclease, SbcCD. If secondary structures such as hairpins are formed by CTG TNRs, they may be cleaved by SbcCD in E. coli causing DNA breaks that stimulate zeocin HR. However the sbcCD mutation slightly decreased zeocin HR in an TNR independent manner. It seems probable that the SbcCD complex is not able to cleave secondary structures formed by TNRs to stimulate recombination, if they are formed at all. The opposite CAG TNR orientation (CAG in the leading strand template) was assayed for zeocin HR. Surprisingly expanded CAG TNRs stimulated zeocin recombination to a lesser
extent than CTG TNRs. In this context the observation is that the CAG orientation (CAG on the leading strand template) is more unstable than the CTG orientation (CTG on the leading strand template) (Rabaab Zahra, data not shown). Finally, since deletion of the TNR tract is not a necessary consequence of zeocin HR, I compared the frequency of CTG\textsubscript{140} instability (deletion) in zeocin sensitive and resistant cells. Despite the low level of TNR instability in a CTG\textsubscript{140} it is clear that TNR instability is not stimulated amongst the zeocin resistant recombinants.
CHAPTER FOUR

Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

4.1 Introduction

This chapter begins by investigating the formation of DSBs by TNRs using pulse field gel electrophoresis (PFGE). Also since the rescue of reversed replication forks can use homologous recombination (HR), this chapter investigates the possibility that TNRs lead to replication fork reversal (RFR). We test this hypothesis by assaying for HR using zeocin recombination reporter substrates positioned proximal and distal with respect to the origin of replication. We show that TNR induced HR is lower at the distal site. Furthermore the protein UvrD is known to be essential for RFR in certain replication mutants. TNR dependent stimulation of HR is lost in *uvrD* mutants. Both these data support the hypothesis that TNRs can cause RFR. Finally I present data, not related to TNRs and generated from zeocin assays, that supports the hypothesis that RecG, implicated in RFR, can abort RecA mediated genetic exchanges.

4.2 Do expanded TNRs cause DNA double strand breaks?

So far expanded CTG TNRs have been shown to stimulate zeocin recombination and recombination mutants suggest that zeocin recombination is a result of double strand break repair rather than single strand gap repair. We used pulse field gel electrophoresis (PFGE) in an attempt to physically detect double strand breaks caused by expanded repeats. As DNA breaks are rapidly processed by RecBCD, strains under investigation were rendered *recB*−. Also, as the level of TNR stimulated zeocin recombination was approximately 50-fold lower than that stimulated by a palindrome, it was likely that DNA breaks would be rare and subsequently hard to detect using PFGE. I therefore devised a system that would allow as sensitive detection of DSBs as practically achievable.
Initial investigations to detect TNR induced DSB involved digesting the *E. coli* genome using the rare cutting restriction enzyme, *NotI* (Section 2.3.19) and probing for appropriate fragments using Southern blotting (Section 2.3.21) were undertaken. However no breaks could be detected in strains harbouring CTG TNRs known to stimulate zeocin recombination (data not shown). *NotI* cleaved the *E. coli* chromosome 23 times producing background noise at the same location as the expected broken fragment. To reduce background noise we wished to cleave the *E. coli* genome only twice allowing for over-exposure of the DNA-bound membrane to x-ray film without any danger of cross-hybridisation to other DNA fragments. This was achieved by integrating the *S. cerevisiae* IScel restriction site (IScel<sub>cs</sub>) into the genes *proA* and *tsx* (Fig 4.1 A).

**Figure 4.1 A system to detect rare TNR induced DNA breaks**

(A) Two IScel restriction sites were integrated into *E. coli* strains harbouring either no repeats (DL2868) expanded CAG<sub>64</sub> (DL2869) and CTG<sub>100</sub> (DL2870) TNRs, a 246 bp imperfect palindrome (DL2798) or an IScel restriction site (DL2849) at the beginning of lacZ at *tsx* and *proA* using plasmids (B) pDL2755 and (C) pDL2736 respectively. Digestion of genomic DNA with IScel results in the liberation of a ~174 Kbp fragment containing the lacZ and TNRs.
Integration of the ISceI restriction site into the genes proA and tsx was carried out using pDL2736 and pDL2755, respectively (Fig. 4.1 C / B). Both plasmids contained two 400 bp homology arms to either proA (ProA1F, ProA1R.1 and ProA2F.1, ProA2R) or tsx (TsxA1F, TsxA1R.2 and TsxA2F.2, TsxA2R) amplified from DL1675 using primer pairs indicated in parenthesis. An ISceI restriction site was engineered into the 24 bp of homology used to fuse the two arms using primers TsxA1F, TsxA2R and ProA1F, ProA2R during cross-over PCR. Plasmids pDL2736 and pDL2755 were used to integrate the ISceI restriction sites into the strains DL1786 (CNG0), DL2728 (CTG184), DL2639 (CAG84) and DL2006 (246 bp palindrome at lacZ) creating the strains DL2868, DL2870, DL2869 and DL2798 respectively. Unfortunately the strain DL2870 deleted from CTG184 to CTG100 during construction. An ISceI restriction site was integrated into the lacZ gene in strain DL2792, which already contained both the ISceI restriction sites at proA and tsx creating DL2849. Strains containing a palindrome or an ISceI restriction site at lacZ were created by John Eykelenboom. Despite best efforts, no TNR (CAG84 or CTG104) induced chromosome breakage was observed using this new system above background hybridisation (Fig 4.2A and B, lanes 3 to 10, fragment D). The two positive controls (lane 12: lacZ::ISceI and Lane 13 lacZ::246 bp palindrome) yielded the expected broken chromosome bands. Previously Freudenreich et al., 1998 and Jankowski et al., 2000 detected both mitotic and meiotic chromosome breakage induced by CTG250 and CAG7964 (diploid S. cerevisiae strain), respectively. Interestingly Jankowski et al., 2000 observed no chromosome breakage during mitosis. Given that the rate of zeocin recombination induced by TNRs was low it is not surprising that breaks were difficult to detect physically and these experiments do not rule out the existence of a low level of TNR induced breaks.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

4.3 Is zeocin HR stimulated by replication fork reversal?

So far the data collected indicate that expanded CTG TNRs stimulate zeocin recombination. However the mechanism of stimulation is unclear. Sections 1.4 and 1.5 describe how TNRs appear to cause double strand breaks and perturb DNA replication respectively. Another hypothesis for the TNR dependent stimulation of
recombination is explored in this thesis. Here we propose that the reason for the stimulation of recombination is that as the DNA replisome encounters and begins to traverse an expanded TNR array it stalls or pauses. This stall or pause results in a process of events resulting in replication fork reversal (Section 1.11) (Fig. 1.6). This process of events might occur as follows: as the replisome encounters a TNR array (Fig. 4.3 A) it stalls or pauses causing the replisome to disassemble. As the template and newly synthesised strands begin to anneal a fourth strand emerges forming a chicken-foot like structure (Fig. 4.3 B). By removing RecA from the fork, the helicase UvrD is required for the formation of this four-way junction, which is stabilised by RuvAB binding. The branch migration activity of RuvAB allows the further re-annealing of template and newly synthesised strand, pushing out the fourth strand further (Fig. 4.3 C) containing the newly synthesized copy of the zeocin repeats (direct repeats shaded green). These can go on to recombine with the original template zeocin repeats (direct repeats shaded blue) (Fig. 4.3 D). This recombination event would normally lead to replication restart in a PriA dependent manner.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

Figure 4.3 Hypothesis for replication fork reversal leading to zeocin HR
Proposed hypothesis of replication fork reversal resulting in zeocin recombination and zeocin resistant colonies. (A) In this diagram, the replisome is moving right to left through the zeocin repeats at cynX (template zeocin repeats in blue and newly synthesised zeocin repeats in green) where it pauses or stalls on encountering a TNR array (red bars), which leads to the template and newly synthesised DNA strands re-annealing to one another in a UvrD and RuvAB dependent reaction (B). As the template and newly synthesised strands continue to re-anneal a chicken-foot structure forms (C). In the majority of cases, this structure would be acted upon by RecBCD, allowing RecA dependent HR or replication restart or degradation and restart. In a minority of RecBCD cells the newly synthesised zeocin repeats (green) may be able to recombine unequally with the original template copy (D) resulting in zeocin resistance.

4.4 Origin distal zeocin recombination (mhpC)
If zeocin recombination were due to RFR then a prediction would be that the level of TNR induced zeocin recombination would be lower on the origin distal side than the origin proximal side of TNRs. This is because the distal zeocin repeats would not be part of a reversed replication fork stimulated by the TNRs. In order to measure TNR mediated recombination after the replication fork had progressed through TNRs, the zeocin repeats were removed from cynX and positioned into the gene region at mhpC/mhpB (Fig. 4.4 A).

Due to the difficulty of integrating TNRs into the E. coli genome, it was technically easier to delete the zeocin repeats from cynX in strains harbouring TNRs and re-introduce zeocin repeats at mhpC/mhpB. Zeocin repeats were removed from cynX in the strains DL1786 and DL2009 using the plasmid pDL2342 (Fig. 4.4 B) which only contains the cynX homology arms and not the zeocin repeats creating the
strains DL2369 and DL2374 respectively. pDL2342 contains a 800 bp region of cynX that was amplified from wild-type E. coli strain MG1655 (DL1675) with the primers, Zeo-CF1 and Zeo-CR2. This PCR product was cloned into pMH9 at PstI and SalI. After excision of the zeocin repeats from cynX, strains were tested for the inability to give rise to zeocin resistant colonies.

The zeocin repeats were integrated into mhpC / mhpB using the plasmid, pDL2069, created by John Eykelenboom creating the strains DL2393 (CNG0) and DL2394 (CTG95) (Fig. 4.4 C). The two homology arms for mhpC and mhpB were amplified from DL1803 using the primer pairs: mhpC-F, mhpCRf and mhpB-R and rnhpBFf, respectively. These two homology arms were joined using the primers, mhpCF and mhpBR and cloned into the plasmid pMH9 at PstI and NotI. The zeocin repeats were amplified from pDL1998 using the primers, zeo_cass_F and zeo_cass_R. After digestion with BglII and XbaI, the zeocin repeats were cloned into the mhpC / mhpB homology arms.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

Figure 4.4 Zeocin repeats at mhpC (origin distal)
(A) An assay to measure origin distal levels of TNR induced zeocin recombination. The zeocin repeats were removed from DL1786 and DL2009 using pDL2342 (B). After testing for loss of the zeocin repeats via PCR and sensitivity to the drug zeocin, the zeocin repeats were introduced into mhpC/mhpB region using pDL2069 (C).

In a no repeat strain, recombination rates measured at cynX and mhpC are approximately equal (Fig. 4.5). Interestingly, zeocin recombination induced by CTG₉₅ at the origin distal mhpC site was 6-fold lower compared to the origin proximal cynX site. This evidence suggests that TNR stimulated recombination is one sided and supports the hypothesis that TNRs cause RFR at a low frequency resulting in recombination between zeocin repeats and zeocin resistance.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in *E. coli*?

Origin distal zeocin recombination rates in the absence and presence of TNRs was measured. The strains used were DL1786 (None *cynX*), DL2393 (None *mhpC*), DL2009 (CTG95 *cynX*), DL2394 (CTG95 *mhpC*).

### 4.5 UvrD helicase

Considering that the helicase, UvrD, is essential for replication fork reversal in *dnaE* and *dnaN* mutants (Flores et al., 2004), we wished to test whether TNR-stimulated zeocin recombination was dependent on UvrD. A pKO vector (pDL2391) (Fig. 4.6) for deleting a 1.65 Kbp central region of the *uvrD* gene was constructed. After integration and excision of pDL2391 the *uvrD* mutants had an in-frame deletion, 513 bp in length. pDL2391 contains two 400 bp regions of homology to the beginning (UvrD1F, UvrD1R) and end (UvrD2F, UvrD2R) of the *uvrD* gene amplified from DL1675 using the primers indicated in parenthesis. The two homology arms were fused using primers UvrD1F, UvrD2R during cross-over PCR and cloned into pTOF24 using the restriction sites *PstI* and *SalI*.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

Figure 4.6 Plasmids pDL2391 and pGBuvrD-2
Plasmids pDL2391 used to delete the uvrD gene and pGBuvrD-2 used to complement the uvrD chromosomal deletion.

As previously reported, uvrD mutants show a two-fold increase in recombination (Vcaute et al., 2005) (Fig. 4.7). Introduction of pGBuvrD-2 (a gift from B. Michel), which expresses UvrD into uvrD mutants (Fig. 4.6), reduces recombination to wild-type, indicating that the increase in recombination in the uvrD mutant results from the absence of a functional UvrD protein. To ensure that pGBuvrD-2 was still present after assaying for zeocin recombination cells were plated onto the LB-plates containing the antibiotic spectinomycin (pGBuvrD-2 confers resistance to the antibiotic spectinomycin). On the RFR hypothesis, we predicted that TNRs would not stimulate recombination beyond the 2-fold seen in the no repeat uvrD control. Consistent with this, stimulation of zeocin recombination in strains containing CTG95 and CTG140 is lost in a uvrD background (Fig. 4.7). CTG95 and CTG140 uvrD mutants give approximately equal zeocin recombination rates compared to a no-repeat control. Since UvrD is essential for RFR in certain replication mutants (Flores et al., 2004) these data support the hypothesis that TNR mediated zeocin recombination is the result of replication stalling at TNRs leading to fork reversal.
Chapter Four: Do expanded CTG TNRs stimulate replication fork reversal (RFR) in E. coli?

Figure 4.7 TNR dependent HR is lost in an uvrD mutant

uvrD mutants were created and recombination rates were assayed in the absence and presence of CTG TNRs. The strains used were DL1786 (CNG0), DL2649 (CNG0 uvrD::CmR), DL2539 (CNG3 ΔuvrD using pDL2391), DL2651 (CNG3 uvrD::CmR + pGBuvrD-2), DL2009 (CTG95), DL2629 (CTG95 ΔuvrD using pDL2391), DL2305 (CTG140) and DL2477 (CTG140 ΔuvrD using pDL2391). Complementation assays were carried out using pGBuvrD-2 which expresses the UvrD protein (a gift from B. Michel).

4.6 TNR induced recombination is independent of RecG helicase

RecG is a helicase that unwinds branched DNA structures such as Holliday junctions, promotes repair of UV-damaged DNA in a RuvABC mutant (Mahdi et al., 1997; McGlynn et al., 1997) and has been implicated in RFR (Section 1.11). RecG previously has been shown to reduce the efficiency of recombination in conjugational crosses with Hfr donors and in plasmid recombination assays (Lloyd and Buckman, 1991) however little is still known about what roles this helicase plays in HR. Surprisingly, using the zeocin repeat assay, a no-TNR recG mutant exhibited an elevated level of recombination (2.3-fold increase) compared to a no-TNR wild-type control (DL1786) (Fig. 4.8). A CTG95 recG mutants gave a 1.6-fold greater level of recombination compared to a no-TNR, recG control. It is clear from Fig. 4.8 that a
recG mutation increases the basal level of recombination. Any additional TNR mediated stimulation is independent of recG. However, it is unclear why introduction of a recG mutation should stimulate zeocin recombination. Biochemical evidence suggests that RecG antagonises RecA by driving branch migration in the reverse direction to that catalysed by RecA-mediated strand exchange (Whitby and Lloyd, 1995). Data presented here supports the idea that RecG can abort exchanges mediated by RecA.

**Figure 4.8 TNR induced recombination is independent of RecG helicase**

Effect of recG mutation on zeocin recombination with or without TNRs. The strains used were DL1786 (None cynX), DL2361 (1786 recG), DL2009 (CTG95), DL2099 (CTG95, recG). The results are the means of three independent experiments. Error bars show 95 % confidence intervals.

### 4.7 Conclusions

In an attempt to confirm the observation that expanded CTG TNRs stimulate HR, I used pulse field gel electrophoresis (PFGE) to physically investigate the formation of DSBs. Initial observations using the rare cutting enzyme NotI did not reveal any TNR
induced DBS and also produced significant amounts of background signal. At this stage it was clear that a system would have to be devised that would allow as sensitive and as stringent detection of a DSB as practically achievable. This would allow over-exposure of the membrane to the phosphoimager, whilst keeping background levels to a minimum. I devised a system using the S. cerevisiae restriction endonuclease site ISceI that, when engineered twice into the E. coli chromosome, would cleave a ~174 Kbp fragment, roughly centred on the TNRs at lacZ when digested with ISceI restriction endonuclease. I was still unable to detect a break above background, using this more sensitive method. This chapter continues with exploring the possibility that TNRs stimulate replication fork reversal (RFR) using the zeocin assay and the following two predictions based on the model presented in Fig. 4.3: (i.) TNR induced zeocin HR should be reduced when the zeocin cassette is moved to an origin distal site with respect to the origin of replication and (ii.) TNR induced zeocin recombination should be reduced in a uvrD mutant. Indeed TNR induced zeocin recombination was lower at the origin distal site compared to the origin proximal site and any TNR induced stimulation of HR was lost in a uvrD background. Together these data support the hypothesis that TNR mediated zeocin HR is a result of RFR occurring at TNRs. Finally data presented in this chapter independently supports the hypothesis that RecG can antagonise RecA mediated strand exchange using the zeocin assay in a TNR-independent manner.
Chapter Five: Do expanded CTG and CAG TNRs stimulate the SOS response in *E. coli*?

CHAPTER FIVE

Do expanded CTG and CAG TNRs stimulate the SOS response in *E. coli*?
5.1 Introduction

Chapter five investigates whether expanded CAG or CTG TNRs generate sufficient single stranded DNA to stimulate the SOS response in *E. coli* in wild-type and cells deficient for the accessory helicase Rep. Previously 2-dimensional neutral / neutral gel electrophoresis analysis of multi-copy number plasmids harbouring TNRs and propagated in *E. coli* revealed replication fork stalling at TNRs (Samadashwily et al., 1997). I wished to investigate this further using TNRs present in the *E. coli* chromosome.

5.2 Expanded TNRs do not stimulate the SOS response.

Since expanded CTG TNRs have been shown genetically to induce recombination (Section 3.6) I wished to determine whether the level of level of single stranded DNA generated would be sufficient to stimulate the SOS response. The SOS response is a system set into action when a cell is subjected to DNA damage (Michel, 2005). Previously expanded CAG and CTG TNRs in plasmids were shown not to induce the SOS response (Hashem et al., 2004). However, I wished to re-test this observation but using TNRs harbouring expanded TNRs within the chromosome. I measured the level of SOS using plasmid pGB150, a gift of Dr. Garry Blakely, Edinburgh University (Fig. 5.1). pGB150 is a pACYC184 derivative plasmid containing the promoter of the *sfiA* gene fused in frame to the green fluorescent protein (GFP) open reading frame of pGFPmut3.1 (Clonetech). The *sfiA* gene is transcribed only during the SOS response and its function is to inhibit cell division (Bhattacharya and Beck, 2002). Therefore the amount of SOS inducing signal can be measured using pGB150.
Chapter Five: Do expanded CTG and CAG TNRs stimulate the SOS response in E. coli?

Initial experiments were carried out to determine levels of GFP intensity in cells grown in LB medium and LB medium supplemented with nalidixic acid as in section 2.4.1. Cells were grown at 37 °C from a single colony to an OD$_{600}$ = 0.6 in LB medium containing chloramphenicol. 5 mls of this culture were taken and grown for a further 1 ½ hours supplemented with nalidixic acid (final concentration 0.4 mM), which is known to induce the SOS response (Newmark et al., 2005). Untreated cells (pink) and cells treated with nalidixic acid (blue) were examined under the microscope and assayed for cell length and GFP intensity (Fig. 5.2). As expected cells grown in the absence of nalidixic acid grow to between 2 and 5 μM in length (mean 2.9 μM) and the mean level of corrected GFP intensity is 43 Gray units and does not beyond 414 Units (Fig. 5.2 pink scatter). In contrast the addition of nalidixic acid, which is known to induce the SOS response in a RecBCD dependent manner (Newmark et al., 2005), causes a dramatic shift in the scatter pattern (Fig. 5.2 blue scatter). Cells grown in the presence of nalidixic acid grow to between 3 and 44 μM.
in length (mean 5.8 μM); the mean level of corrected GFP intensity is 262 Gray units and does not exceed 886 Gray units. Considering that cells grown in nalidixic acid grew longer and fluoresced more than untreated cells, the plasmid pGB150 was deemed reliable for assaying levels of SOS induction in E. coli.

**Figure 5.2 Nalidixic acid stimulates the SOS response**
The plasmid pGB150 (A gift of Dr. Garry Blakely) was introduced into strain DL1786 containing no TNRs rendering the cells resistant to the antibiotic chloramphenicol. Cells were grown from single colony to an OD₅₀₀ = 0.6 in LB medium containing chloramphenicol. 5 ml of this culture were taken and grown for a further 1 ½ hours supplemented with nalidixic acid, which is known to induce the SOS response (Newmark et al., 2005). Untreated (pink) (896 measurements) and cells treated with nalidixic acid (blue) (991 measurements) were examined under the microscope and assayed for cell length and GFP intensity. As LB medium can auto-fluoresce, GFP intensity was normalised against any background fluorescence. Units of fluorescence on the y-axis are Gray units (Gy).

pGB150 was introduced into strains with no TNRs (DL1786), CAG₈₄ (DL2639) and CTG₁₄₀ (DL2305) at lacZ and levels of SOS induction and cell length were determined as in Section 2.4.1 (Fig. 5.3). Cells were re-tested for TNR length to ensure that TNRs had not deleted during growth. Overall, the presence of TNRs in wild-type E. coli had no effect on cell length or level of induction of the SOS response. The no repeat control (Fig. 5.3 A) gave an average cell length of 3.08 μM and an average corrected GFP intensity level of 81.61 Gray units. Similarly strains harbouring CAG₈₄ (Fig. 5.3 B) and CTG₁₄₀ (Fig. 5.3 C) TNRs at lacZ have an average cell length of 3.54 and 4 μM respectively; corrected GFP intensity levels of 71.6 and
Chapter Five: Do expanded CTG and CAG TNRs stimulate the SOS response in E. coli?

54.4 μM respectively. From this we can conclude that expanded CAG and CTG repeat arrays do not induce the SOS response in a detectable proportion of cells in the population. This is consistent with the low level of TNR induced HR at the zeocin repeats.
Figure 5.3 Expanded TNRs do not stimulate the SOS response.
The plasmid pGB150 (A gift of Dr. Garry Blakely) was introduced into strains with (A) no TNRs (DL2861) (999 measurements), (B) (CAG)$_{84}$ (DL2865) (1004 measurements), and (C) (CTG)$_{140}$ (DL2864) (1063 measurements). Cells were grown in LB containing chloramphenicol to OD$_{600} = 0.6$ and examined under the microscope for GFP intensity and cell length. Units of fluorescence on the y-axis are Gray units (Gy).
5.2.1 TNR induced SOS induction in a rep background

Replication fork progression has been shown to be two fold slower in rep mutants compared to wild-type E. coli (Colasanti and Denhardt, 1987; Lane and Denhardt, 1975) and it is known, from plasmid studies in E. coli and yeast, that DNA replication fork progression is impeded upon encountering an expanded TNR array (Section 1.5). Furthermore rep mutants undergo elevated levels of RFR (Michel et al., 1997; Seigneur et al., 1998). Following on from Section 5.2, we hypothesised that the introduction of a rep mutation into E. coli cells harbouring expanded CAG and CTG TNRs might exacerbate the progression of the replisome through an expanded TNR array and increase levels of SOS induction.

A pKO vector (pDL2341) (Fig. 5.4) for deleting a central region of the rep gene was constructed. After integration and excision of pDL2341, rep mutants had an in-frame rep deletion, 195 bp in length. pDL2341 contains two 400 bp regions of homology to the beginning (Rep1F, Rep1R) and end (Rep2F, Rep2R) of the rep gene amplified from DL1675 using the primers indicated in parenthesis. The two homology arms were fused using primers Rep1F, Rep2R during cross-over PCR and cloned into pTOF24 using the restriction sites PstI and SalI.
The *rep* gene was deleted from DL1786 and DL2639 using a P1 lysate harvested from DL2633 (A gift of B. Michel) and from DL2009 using pDL2341, creating the strains DL2647, DL2788 and DL2384 respectively. The plasmid pGB150 was introduced into the *rep* mutants creating DL2905 (CNG0), DL2907 (CAG86) and DL2906 (CTG94). Experiments with isogenic *rep* mutants are to be carried out. All strains analysed exhibited increased cell length as previously reported for *E. coli* *rep* mutants (Lane and Denhardt, 1975) The no repeat control (Fig. 5.5 A) gave an average cell length of 5.06 μM and an average corrected GFP intensity level of 43.93 Gray units. Similarly a *rep* mutant harbouring CAG84 (Fig. 5.5 C) gave an average cell length of 5.63 μM and an average GFP intensity level of 43.36 Gray Units. Interestingly a *rep* mutant harbouring CTG94 exhibited an elevated SOS response compared to the *rep*, no repeat control and the *rep*, CAG84 strain (Fig 5.5 B). The *rep*, CTG94 strain gave an average cell length of 4.31 μM and an average GFP intensity level of 63.20 Gray units. These data correlate with data presented in chapter three, reporting that the CTG94 but not the CAG84 or the no repeat control stimulate HR at zeocin repeats.
Chapter Five: Do expanded CTG and CAG TNRs stimulate the SOS response in E. coli?

Figure 5.5 Cell length and GFP intensity in rep mutant strains
The plasmid pGB150 (A gift of Dr. Garry Blakely) was introduced into rep mutants strains with (A) no TNRs (DL2905) (2980 measurements), (B) (CTG)$_{94}$ (DL2906) (2986 measurements) and (C) (CAG)$_{35}$ (DL2907) (1977 measurements). Data presented represent pooled data from three individual experiments for DL2905 and DL2905 and two individual experiments for DL2907. Cells were grown in LB containing chloramphenicol to OD$_{600}$ = 0.6 - 0.64 and examined under the microscope for GFP intensity and cell length. As LB medium can autofluoresce, GFP intensity was normalised against any background GFP fluorescence. Units of fluorescence on the y-axis are Gray units (Gy).
5.3 Two dimensional neutral / neutral gel electrophoresis

Section 5.1.2 describes that removing the rep helicase in a CTG₉₅ strain increases the amount of SOS signal and suggests replisome progression through CTG TNRs is inhibited in rep mutants. Previously, TNRs present in high copy number plasmids have been shown to stall or impede replication fork progression (Samadashwily et al., 1997). However it is not clear whether expanded TNRs would have the same effect in a chromosome. I therefore wished to use 2-D gel analysis to investigate the effects of chromosomal TNRs on replication fork progression. Although a technically difficult method to setup and interpret, 2-D gel analysis would have the potential to show not only replisome pausing (indicated as a DNA ‘blob’ on the y-arc) and events such as replication fork reversal but also TNR induced DNA breaks. These effects may be especially pronounced in a rep mutant.

The ideal size of DNA fragment for use in 2-D gel electrophoresis is around 5 Kbp with the TNRs positioned 1/3 towards the end of the DNA fragment with respect to the direction of replication. It is important to use a restriction enzyme that will efficiently cleave the chromosome and is not too expensive so that many units of enzyme can be added. A naturally occurring EcoRI site existed at the 5’ end of the lacZ gene. As the restriction enzyme EcoRI is known to cleave well, is inexpensive and supplied in large quantity it was decided to construct a pKO vector (pDL2908) to integrate another EcoRI restriction site in the gene mhpR (Fig. 5.6 A). pDL2908 contains two 400 bp regions of homology to 5’ end of the mhpR gene (2D_EcoRI_F1, 2D_EcoRI_R1 and 2D_EcoRI_F2, 2D_EcoRI_R2) amplified from DL1675 using the primers indicated in parenthesis. The two homology arms were fused together using
primers 2D_EcoRI_F1 and 2D_EcoRI_R2 during cross-over PCR and cloned into pTOF24 using the restriction sites PstI and Sall.

Figure 5.6 Plasmid pDL2908 and an example of a 2D-gel
(A) Plasmid pDL2908 used for integrating the EcoRI restriction site at mhpR (B) An example of a 2-D gel showing a Y arc and a bubble arc (Section 2.3.19) taken from Joel Humberman's Website (http://asajj.roswellpark.org/huberman/jh.html).

pDL2908 was integrated into DL1768 (CNG)₀, DL2305 (CTG)₁₄₀ and DL2639 (CAG)₈₈. Unfortunately DL2639 deleted from CAG₈₆ to CAG₄₃ during construction. Integration of pDL2908 resulted in the integration of the EcoRI restriction site in the mhpR gene. After integration, a 4995 bp is liberated when cut with restriction enzyme, EcoRI (Fig. 5.7)

Figure 5.7 ~ 5 Kbp region used for two dimension gel analysis
A 4947bp region is excised from the E. coli genome after integration of pFL2908 when cleaved with EcoRI. pDL2908 was integrated into DL1768 (CNG)₀, DL2305 (CTG)₁₄₀ and DL2639 (CAG)₄₃, creating DL2932, DL2937 and DL2936 respectively.
The integration of pDL2908 was confirmed using Southern analysis (Fig. 5.8). Genomic DNA was extracted from DL2932 (lanes 2 - 4), DL2936 (lanes 5 - 9) and DL2937 (lanes 10 - 14) and run in a 1 % agarose gel for 16 hours at 60 V. The gel was then blotted onto nylon membrane, cross-linked and probed with 1 Kbp lacI and lacZ probes using random priming and $^{32}$P-dCTP.

For an extended period of time I could not obtain a replication γ-arc showing replication intermediates (Fig. 5.9). Initial two dimensional analysis of DL2932 provided a strong monomer signal (Fig. 5.9, arrow) but no replication arc. As DNA present in replication intermediates is often single stranded and especially fragile, it was assumed that excessive DNA degradation was taking place, most probably from UV exposure when visualising the DNA on a UV transilluminator. Acting on advice from Dr. Maria Vogelauer (Wellcome Trust for Cell Biology, University of Edinburgh), I minimised UV irradiation exposure and by doing so finally achieved γ-arc (Fig. 5.10) (confirmed by M. Vogelauer).

**Figure 5.8 Southern analysis to confirm EcoRI restriction site integration**
After integration and excision of pDL2908, the integration of the EcoRI site at mhpR was confirmed by digesting chromosomal DNA with EcoRI, Southern blotting to nylon membrane and probing with $^{32}$P labelled 1 Kbp probes to lacZ and lacI generated by random priming. Lanes: (1) 4.5 Kbp PCR fragment containing the lacZ and lacI DNA region, (2 - 4) DL2932 no TNR control (5 - 9) (DL2936) CAG$_{41}$, (10 - 14) (DL2937) CTG$_{140}$. Lanes 5 - 14, which contain TNRs, migrate slower than lanes 2 - 4.
Figure 5.9 Initial two dimensional gel analysis
Initial two dimension gel analysis of DL2932 (in triplicate) yielded a strong monomer signal (arrow) but did not show a replication arc.

Although the y-arc observed was weak (Fig. 5.10, arrow), I am confident that it could be improved through further minimising UV exposure, baking the membrane to fix DNA to Hybond™ membrane instead of cross-linking with a Stratalinker™, enriching for replication intermediates using BND cellulose, not depurinating DNA with HCl during Southern transfer, and by using more specific DNA probes.

Figure 5.10 Two dimensional gel analysis – the y-arc
Two dimensional gel analysis of DL2932 (in triplicate). A faint y-arc can be seen (arrow) in all three samples.

If continued, it would be interesting to investigate replication arc progression through TNRs at lacZ in wild-type and in rep mutant cells.

5.4 Conclusions
This final chapter investigates whether CAG or CTG TNRs stimulate a SOS response. In wild-type cells, there was no effect of expanded CTG or CAG TNRs on the SOS response. However, upon introduction of the rep mutation, the CTG orientation displayed a clear increase in SOS response. The observation that the CTG TNRs but not CAG TNRs stimulate the SOS response is in agreement with results presented in chapter three. I also wished to use two dimensional neutral / neutral gel electrophoresis to observe replisome progression through TNRs. However, despite best efforts, I only began to observe a replication y-arc in no-TNR control cells towards the end of my PhD and could not undertake further analysis. Regardless, I feel that it would not take much more optimisation (using a BND cellulose enrichment stage for replication intermediates and also higher specificity DNA probes for Southern analysis) to be able to use this technique to address the two questions: (i.) does replisome stalling occur in wild-type cells harbouring expanded CTG and CAG TNRs at the lacZ gene and (ii.) same experiment as (i.) but in a rep mutant background.
6.1 Thesis overview

For the first time, different lengths of TNRs have been integrated into the *E. coli* genome (Section 3.1 to 3.4) allowing novel investigations on TNR instability (Rabaab Zahra) and on TNR induced HR (this thesis). This thesis intended to explore the mechanism by which expanded TNRs stimulate homologous recombination. I have established that expanded CTG TNRs on the leading strand template stimulate HR (Section 3.6), and that the opposite CAG orientation, stimulates recombination to a lesser extent (Section 3.9). This observation was especially interesting considering that studies from Rabaab Zahra clearly indicate that the CAG orientation is much more unstable than the CTG orientation and is the first time a direct comparison has been made between CAG and CTG orientations. This observation raises questions about the link between TNR instability and stimulation of HR by expanded TNRs. Considering the strong link between HR and replication, one new hypothesis could be that whatever role TNRs have on perturbing replisome progression is rescued by HR, through which the length of the TNR array is maintained. In the absence of HR, as in the case of the CAG orientation, there is no rescue and through some aberrant DNA process, the repeat length is altered without any change to flanking sequence.

A significant body of TNR research has focused on the possibility that TNRs form secondary structures and that the formation of these secondary structures, such as hairpins, could be linked to instability (Section 3.8). In *E. coli*, the protein complex SbcCD has been shown to cleave hairpin structures formed by palindromes during DNA synthesis, resulting in a DNA break that is repaired by HR. A prediction made in this thesis is that if TNRs do form stable hairpin secondary structures capable of being processed like palindromes then the rate of HR would be altered in *sbcCD* mutants. Data presented in Section 3.8 does not support this hypothesis and argues
against the formation of recombinogenic hairpin structures by expanded CTG TNR arrays.

Another interesting question raised by this thesis is whether there is a link between TNR instability and TNR stimulated HR (Section 3.10). As data presented in chapter three clearly demonstrate a link between CTG TNR length and stimulation of HR, I wished to determine whether CTG instability was higher in zeocin resistant cells compared to zeocin sensitive cells. Genemapper analysis undertaken here supports the conclusion that the level of instability observed in both zeocin resistant and sensitive CTG140 cells was very low. No stimulation of CTG140 instability in cells that had undergone HR was observed.

This thesis continues by attempting to physically demonstrate the creation of double strand breaks by expanded TNRs using pulse field gel electrophoresis (Section 4.2). After initial attempts to detect breakage using the rare cutting enzyme NotI, I developed a system using S. cerevisiae IScel restriction. This allowed me to cleave the chromosome only twice at engineered IScel sites, liberating a large DNA fragment that could be probed using Southern blotting for DNA breaks with minimal background signal. Unfortunately I was still unable to physically detect TNR induced DNA breaks above background signal. I was able to clearly detect DNA breaks using an IScel cleavage control and using a strain harbouring a 246 bp imperfect palindrome in the presence of SbcCD (a gift of John K. Eykelenboom). It is not surprising that detection of a TNR induced DNA break has been a difficult task. Using the same zeocin assay, a 246 bp imperfect palindrome at lacZ in a sbcCD+ strain shows a 50-fold higher rate of HR compared to CTG140. However, the demonstration that CTG TNRs do stimulate the rate of zeocin HR does argue that CTG TNRs cause DNA breaks at a low frequency. Although HR does seem to be stimulated by expanded
TNRs in *E. coli* and yeast and that DNA double strand breaks have been detected in yeast, it remains to be determined whether breaks are made in the chromosome of human patients with Huntington’s disease or Myotonic dystrophy.

Considering that TNRs stimulate HR and may perturb DNA replication, I hypothesised that TNRs might stimulate replication fork reversal (RFR) (Section 4.3). I tested the prediction that RFR-stimulated HR would be lost in *uvrD* mutants, a prediction made on the basis that the UvrD helicase is known to be essential for RFR in certain replication mutants. I also tested the prediction that RFR-stimulated HR would be restricted to the origin-proximal side of the TNR array. I observed that stimulation of HR by CTG₉₅ and CTG₁₄₀ was lost in a *uvrD* background (Section 4.5) and stimulation of HR by CTG₉₅ was restricted to the origin proximal side of the TNR array (Section 4.4). Together these data support the hypothesis that TNRs do cause RFR at a low frequency resulting in HR.

I was able to confirm, using the zeocin assay, the hypothesis that the helicase RecG antagonises and possibly aborts RecA mediated HR (Section 4.6). Independent of dO TNRs, *recG* mutants demonstrate a stimulation of recombination over wild-type cells. Although the mechanism by which RecG acts against RecA is unknown one possibility is that RecG drives branch migration in the reverse direction to that naturally catalysed by RecA. These results are inconsistent with RecG-stimulated RFR (Section 1.11).

The final chapter of this thesis (chapter five) explores the possibility that the level of single-stranded DNA generated by TNRs might be sufficient to stimulate the SOS response in *E. coli*. Analysis was undertaken in wild-type cells harbouring TNRs (Section 5.2) and in *rep* helicase mutants (Section 5.2.1) using a P₉₆,:GFP reporter system kindly provided by Dr. Garry W. Blakely. In wild-type cells, TNRs did not
stimulate the SOS response. The rep mutation was then introduced into the same strains, based on the knowledge that replication fork progression was two fold slower and might exacerbate any events occurring when the replisome encounters an expanded TNR array. Interestingly, in a rep, CTG₉₄ strain a clear SOS response was observed compared to the no repeat, rep control and a CAG₈₄, rep strain.

Two dimensional neutral / neutral gel electrophoresis is a technique that would allow me to monitor the progression of the replisome through TNRs and would indicate replication fork stalling. Since the only evidence for TNR induced replisome stalling is in high copy number plasmids, data obtained from TNRs present in the chromosome would be a useful contribution to the TNR research field. However, this technique is notoriously difficult to set up. Towards the end of my PhD I was beginning to observe a y-arc in wild-type cells with no TNRs (Section 5.3). However the strength of the signal was very weak and further optimisation was required. However, due to completing my PhD and writing my thesis, I was unable to continue analysis.

6.2 The molecular mechanism behind TNR induced HR

This thesis for the first time attempts to investigate that expanded CTG TNRs cause replication fork reversal and has built up a body of evidence to suggest that RFR may be the cause of CTG TNR zeocin HR. Studies of instability have shown that CTG on the lagging DNA strand during replication folds into a secondary structure and is more prone to instability compared to CTG on the leading strand. In contrast to these instability data I have observed that CTG on the leading strand, the more stable orientation, gives a stimulation of zeocin HR but not CAG on the leading strand.
Furthermore, the events that are leading to zeocin resistance do so at a very low frequency. This orientation effect and low level of HR suggests that replication may be perturbed, possibly leading to RFR. The role of RFR has been investigated genetically, hypothesising that (i.) moving the zeocin repeat to an origin distal position would reduce CTG zeocin HR and (ii.) CTG dependent zeocin HR would be reduced in a mutant known to be involved in the early stages of RFR, namely \textit{uvrD}. CTG zeocin HR is significantly reduced the zeocin repeats are positioned in an origin distal position and when the \textit{uvrD} helicase is mutated supporting the hypothesis that expanded CTG TNRs lead to RFR. Furthermore \textit{rep} mutants are known to undergo elevated levels of RFR. SOS induction was elevated in a \textit{rep} mutant only in cells harbouring expanded CTG TNRs. Taken together these data support the hypothesis that expanded CTG TNRs perturb DNA replication in an orientation dependent manner leading to RFR.

Freudenreich et al., 1998, using \textit{Saccharomyces cerevisiae}, noted a much higher stimulation of recombination compared to the data produced here. This may suggest that some eukaryotic specific factors are present that modulate TNR dependent HR. Perhaps one of the most notable differences between a prokaryotic system and an eukaryotic system is the huge difference in Okazaki fragment length generated during DNA synthesis. The length of bacterial Okazaki fragments range between 1 – 2 kilobases compared to 150 bp in yeast (Gerbi and Bielinsky, 1999; Liu et al., 2004). Since 150 bp equates to around 50 TNRs it is plausible that whole Okazaki fragments consisting entirely of TNRs could be generated. Displacement DNA synthesis and flap processing of a TNR Okazaki fragment could account for the increased instability (Section 1.6) and TNR fragility (Liu et al., 2004) of TNRs observed in eukaryotic systems. However models for TNR instability during Okazaki
fragment processing depend on the formation of secondary structures during strand
displacement. It is unclear, also technically difficult to demonstrate, that these
structures are formed or indeed persist during or after strand displacement. As no
research has been undertaken into RFR in eukaryotic systems it is difficult to
determine whether the mechanism of TNR dependent HR described here could
account for the stimulation of HR seen in eukaryotic systems.

Whether TNR cause double strand breaks and stimulate HR in human cells is
currently unclear. Recently dramatic expansion in TNR length was observed in human
striatal cells early in the disease course (Kennedy et al., 2003) and in a murine
Huntington’s disease model (Kennedy and Shelbourne, 2000). It would be interesting
to investigate the formation of double strand breaks using physical analysis of striatal
tissues from deceased TRED patients and in murine models.

Studies of TNR instability have ruled out the possibility that the low rate of
CTG dependent HR was because CTG TNRs were deleting during culture.
Fluctuation analysis was also used to address TNR instability when assaying for HR
as any TNR deletion would have resulted in a lower HR rate. Furthermore, when
assaying CTG95, the rate of HR between different colonies was homogeneous arguing
against TNR deletion during culturing.

6.3 The future of TNR research

I predict that the molecular mechanism underlying TNR expansion will remain
enigmatic for some time to come. Furthermore, even if a mechanism of expansion in
human cells is discovered I feel that ways of preventing TNR expansion in humans
will prove difficult if not impossible. Any events that lead to the conversion of type I
TNR instability into type II TNR instability are likely to be rare founder events in a
small number of cells, probably germ cells. In terms of making an effective contribution to the sufferers of TREDs, the best direction of research maybe to understand and develop treatments for toxic aggregated proteins (type I TREDs) or to reverse expression patterns in cells that are no longer expressing essential genes that have been silenced during TNR expansion (type II TREDs). One example of this is the discovery that the simple sugar trehalose alleviates disease characteristics seen in murine models of Huntington’s disease (Tanaka et al., 2004).


References


the UvrBC-DNA complex by helicase II and DNA polymerase I. J Biol Chem 267, 780-80.


repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. J Med Genet 28, 818-23.


mutation at the FMR1 locus through multiple exchanges between maternal X chromosomes. *Hum Mol Genet* 3, 1823-7.


