CAG.CTG TRINUCLEOTIDE REPEAT INSTABILITY IN THE *E. COLI* CHROMOSOME

RABAAB ZAHRA

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

November 2006

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude and regards to my supervisor David Leach for scientific guidance, help and understanding. Thanks David for always giving time, no matter how silly my discussion point could be! His encouragement and support throughout my PhD period is greatly appreciated.

I thank Jill Sales for her help in designing the instability assay and for on and off help with statistical analysis of my data.

My thanks are due to Commonwealth Scholarship Commission for awarding me scholarship to carry out PhD studies.

Thanks to the wonderful bunch of Leach Lab, who will be remembered with sweet thoughts. I am obliged to Elise Darmon for critically proofreading this thesis, for her help and friendship throughout…Thanks Elise for always being there!

My PhD research on TNRs will carry a sweet memory of TNR team and I thank John Blackwood for sharing the team spirit, for his help, interesting discussions and "Scottish" phrases…. Never will I forget those lessons!

I thank Federica Andreoni for a nice companionship, little Italian lessons and coffee sessions while listening to my non-sense. Teasing Fede has been my favorite fun activity in the lab and I will cherish our times together. It wouldn't have been any fun without you Fede!

Also wonderful thanks to Laura Wardrope, John Eykelenboom and Martin White for their help, appreciation and a share of good laughs.

I thank SBS sequencing department for help with GeneMapper® work and Anna Montazam for taking my plates to Ashworth and sending results with a smile.

I wish to thank my friends Angela, Soloman bhai and kids for keeping me special. Thanks to rest of my friends in Edinburgh for making my stay as joyful and memorable as I could ask for. I also admire my friends around the world for being with me regardless of the distances between us. I thank all those hands which are always up to pray for me, all those hearts which have kept me with love and all those voices which have given me strength. I thank you all!

Lastly, but above all, I am greatly indebted to my parents for allowing me to study abroad and for their love and understanding over the years. I adore you mom for your love and prayers, which always strengthened me in hard times giving me confidence in myself. I thank you dad for your confidence, appreciation and little poems for me. Lovely thanks go to my brothers and sisters for their love and trust from the time we got our spaces in this world. I treasure all those text messages and online chats that made me feel special, loved and tenderly cared.

I dedicate this thesis to the land I was born and will always belong.

ABSTRACT

Expanded CAG·CTG trinucleotide repeat tracts are associated with a number of hereditary neurodegenerative and neuromuscular diseases such as Huntington's disease, myotonic dystrophy and spinocerebellar ataxias. These diseases are characterized by the phenomenon of genetic anticipation, which is defined by a decrease in the age of onset and an increase in severity of the disease with successive generations. The mutational mechanism of repeat instability is still not fully understood.

In order to identify the molecular basis of genetic instability, a polymerization-independent strategy is developed to generate expanded repeat arrays. The repeat tracts are integrated in the 5'end of *lacZ* gene in the *Escherichia coli* chromosome. Using this model system, instability is studied in wild type *E. coli* and in strains deficient in cellular pathways such as DNA repair, replication and recombination. The work demonstrates that instability (expansion and contraction) in wild type cells is length and orientation dependent. Longer tracts are more unstable than shorter ones and the orientation where CAG repeats are on the leading strand template is more unstable than the opposite where CTG repeats are on the leading strand template. This orientation-dependence of CAG·CTG trinucleotide repeat instability is determined by the proofreading subunit of DNA polymerase III (DnaQ) in the presence of the hairpin nuclease SbcCD. The analysis of the sizes of deletions observed in wild type and mutant cells is consistent with the formation of secondary structures *in vivo*.

The mismatch repair pathway does not affect the instability of CTG repeats in the *E. coli* chromosome but influences the CAG orientation. It is suggested that MutS stabilizes CAG repeats by initiating a "repair" process and protecting hairpins from

SbcCD, which can cleave hairpins in the presence of MutL and MutH. No effect of recombination genes is observed on repeat instability in the *E. coli* chromosome.

No effect of transcription is observed in the wild type or mutant strains tested in this work. A mutation in *mfd* gene also does not affect instability. Furthermore, CAG·CTG repeats influence the yield of β-galactosidase in an orientation dependent manner.

Finally, the roles of two helicases, Rep and UvrD are analyzed. A mutation in *rep* helicase strongly destabilizes CTG repeats with no effect on the CAG orientation. UvrD mutants show instability in both orientations. The increase in instability in the *uvrD* mutant depends on RecF in the CTG orientation.

ABBREVIATIONS

%	percentage
°C	degree Celcius
AMP	Adenosine monophosphate
bp	base pairs
BSA	Bovine Serum Albumin
cAMP	cyclic AMP
cm	centimeter
DNA	Deoxyribonucleic Acid
dNTP	deoxyribonucleoside triphosphate
dsDNA	double stranded DNA
DSB	double strand break
DTT	Dithiothreitol
g	gram
kb	kilobase pair
L	litre
LB	Luria Broth
M	molar
µg	microgram
µl	microlitres
µM	micromolar
mg	milligram
ml	milliliter
mM	millimoles
mRNA	messenger ribonucleic acid
PCR	Polymerase Chain Reaction
pfu	plaque forming unit
pH	power of hydrogen
ssDNA	single-stranded DNA
SDS	Sodium dodecyl sulphate
Ta	annealing temperature

Tm	melting temperature
UTR	untranslated region
UV	ultra-violet light
V/v	volume per unit volume
W/v	weight per unit volume

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Tandem DNA Repeats

Tandem DNA repeats, sequences of consecutive repeats, are present in the DNA of all organisms whose genomes have been sequenced. Tandemly repeated sequences are commonly known as "satellite DNA" and have been classified into three categories; satellite, minisatellite and microsatellite DNA. Satellite sequences are highly repetitive with length in the range of one to thousand base pairs. Minisatellites are repeated with sequence length of 6-100 bp but typically are 15 bp long. Microsatellites are repeated with lengths from 1-6 bp and are present in prokaryotic and eukaryotic genomes. These sequences are highly polymorphic between individuals and have been used as major tools for genetic mapping (Dib et al., 1996; Weissenbach et al., 1992), studies of human and animal diversity (Bowcock et al., 1994; Bruford and Wayne, 1993) and for forensic investigations (Jeffreys et al., 1992). Where this polymorphism has been used clinically, it has shown high mutation rates, both in eukaryotes and prokaryotes. The study of tandem repeat sequence instability has been driven by its association with human diseases where it is linked with more than 40 neurological, neuromuscular and neurodegenerative diseases (Cleary and Pearson, 2005). It is important to note that these hypermutable tandem repeat sequences are not limited to eukaryotic genomes, but are also found in prokaryotic genomes where they play an important role in bacterial adaptation (van Belkum et al., 1998).

1.2 Simple Sequence Contingency Loci (SSCL)

Many pathogens have evolved the ability to alter surface-exposed molecules, most often in response to selective pressures associated with the host immune system (Brunham et al., 1993). Pathogenic bacteria can use this adaptive strategy under circumstances, and a range of molecular mechanisms has evolved in these bacteria for generating genetic variation at individual loci termed "contingency loci" (Moxon et al., 1994). SSCL are characterized by tandem repeats located within a coding sequence or a promoter region and have been identified in a variety of pathogenic bacteria (Bayliss et al., 2004; Bayliss et al., 2001). Particularly, an abundance of such loci has been found in the genomes of *Haemophilus influenzae* and *Neisseria meningitidis* (Hood et al., 1996; Saunders et al., 2000). Altering the number of repeat units in these loci can bring changes in the level of promoter activity or switching in the reading frame of translation. Two distinct features of these loci make them different from other types of phase variations. First, the hypermutation of these loci is independent of classic recombination genes. Second, the hypermutation results from expansion or contraction of repeats by one or more units (Bayliss et al., 2001).

1.3 Trinucleotide Repeats (TNRs)

Trinucleotide repeats (TNRs) are a class of microsatellites that have gained attention since the early nineties because of their association with human diseases. These sequences consist of three nucleotides, repeated in tandem. In the genome, there can be 64 possible trinucleotide repeats, since DNA is composed of four different nucleotides (see Table 1.1). Out of these, only three combinations of trinucleotide

repeats (CAG)·(CTG), (CCG)·(CGG) and (GAA)·(TTC) have been associated with human diseases.

These sequences stand out among their peers by exhibiting a unique behaviour termed "dynamic mutation" (Richards et al., 1992; Richards and Sutherland, 1992). This term was described to distinguish difference in the behaviour of repeated sequences from other types of mutations. Unlike classical mutations, where the mutations are retained in somatic tissue and mutant has the same rate of mutation as its predecessor, the repeat mutation process is "dynamic", where the rate of mutation is linked with the copy number of repeat sequences and the product continues to mutate within tissues and across generations, resulting in a product of a change with an altered mutation rate compared to that of its predecessor.

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

Table 1.1: Formation of all possible trinucleotide repeats
There can be 4^3=64 possible trinucleotide repeats, out of which four are mononucleotides (shown as bold italics). Among the rest of 60 combinations, there are 10 groups, which consist of 6 trinucleotides having the same double stranded repeat sequence. Of these ten groups, only three, GAA (in blue), CAG (in red) and CGG (in green) have been associated with disease process.

1.4 Trinucleotide Repeat Expansion Diseases (TREDs)

Expanded trinucleotide repeat sequences are associated with a number of neurodegenerative and neuromuscular diseases. TREDs are distinguished by the phenomenon of "genetic anticipation" which is characterized by an increase in disease severity and a decrease in age-of-onset as the mutation is transmitted from one generation to the next caused by the dynamic nature of the mutation event (Cleary and Pearson, 2005). These diseases can be categorized into two subclasses depending upon the repeat expansion in coding or non-coding region.

1.4.1 Non-Coding Trinucleotide Repeat Expansion Diseases

The non-coding TREDs typically are characterized by large and variable repeat expansions resulting in multiple tissue dysfunction or degeneration (Cummings and Zoghbi, 2000). The individual diseases with pathophysiology are described in the following sections and are shown in Table 1.2.

1.4.1.1 Fragile X Syndrome (FRAXA)

Fragile X syndrome is characterized by mental retardation, dysmorphic features and hyperactivity. The disease is caused by an expansion of a $(CGG)_n$ repeat in the 5'-UTR of the fragile X mental retardation gene (*FMR1*) (Fu et al., 1991; Verkerk et al., 1991). Expansion of CGG repeats beyond 230 results in hypermethylation of the gene, along with a CpG island within the *FMR1* promoter region, which recruits the transcription silencing machinery, and leads to reduced *FMR1* transcription and loss of the gene product (FMRP) (Eberhart and Warren, 1996).

1.4.1.2 Fragile X Mental Retardation (FRAXE)

This is caused by the expansion of $(GCC)_n$ repeats in the promoter region of the *FMR2* gene (Knight et al., 1993). As with FRAXA, the expanded repeats are hypermethylated, which results in transcriptional silencing and subsequent loss of gene product (FMR2).

1.4.1.3 Friedreich Ataxia (FRDA)

FRDA is an autosomal recessive condition and therefore the only triplet repeat disorder that does not exhibit genetic anticipation. This is caused by a large intronic expansion of GAA repeats in the *frataxin* (*X25*) gene, which results in reduced *X25* expression (Campuzano et al., 1996). Reduced *X25* mRNA decreases frataxin level, which suggests that FRDA results from the a partial loss of *frataxin* function (Campuzano et al., 1997).

1.4.1.4 Myotonic Dystrophy (DM)

DM is the most common form of muscular dystrophy in adults and is characterized by variable phenotypes and anticipation. Two types of DM have been identified. DM1 is caused by an expanded CTG repeat in the 3'-UTR of the protein kinase gene, *DMPK* (Brook et al., 1992). The CTG repeat may alter the DMPK level by interfering with the transcription of *DMPK* or RNA processing or translation. DM2 is caused by a CCTG expansion (mean ~5000 repeats) located in intron 1 of the zinc finger protein 9 (*ZNF9*) gene (Liquori et al., 2001).

1.4.1.5 Spinocerebellar ataxia type 8 (SCA8)

This is caused by an expanded CTG repeat in the 3'-terminal exon of *SCA8* (Koob et al., 1999). Uniquely among the triplet repeat disorders, the SCA8 transcripts do not code for a protein and may be an endogenous antisense RNA that regulates the expression of another gene(s).

1.4.1.6 Spinocerebellar ataxia type 12 (SCA12)

This is a rare disease caused by a non-coding expansion of CAG repeats in the 5'-UTR of the *PPP2R2B* gene (Holmes et al., 1999), which encodes a regulatory subunit of protein phosphatase 2A (PP2A).

Diseases	Gene	Protein	TNR	TNR size (Normal)	TNR size (Disease)
Fragile X Syndrome	*FMR1*	FMR-1 Protein	CGG	6-35	>230
Fragile XE Syndrome	*FMR2*	FMR-2 protein	GCC	6-35	>200
Friedreich ataxia	*X25*	Frataxin	GAA	7-34	>100
Myotonic Dystrophy	*DMPK*	Myotonic Dystrophy protein kinase (DMPK)	CTG	5-37	>50
Spinocerebellar ataxia type 8	*SCA8*	None	CTG	16-37	110-<250
Spinocerebellar ataxia type 12	*SCA12*	PP2A-PR55β	CAG	7-28	66-78

Table 1.2: Diseases caused by expansion of non-coding trinucleotide repeats (Cummings and Zoghbi, 2000)

1.4.2 Coding Trinucleotide Repeat Diseases

This class of TREDs comprises disorders with repeat expansion in a coding region and the expanded repeat is a CAG tract which encodes for amino acid, glutamine. In contrast to non-coding diseases, these polyglutamine diseases have small expansions of variable sizes. Studies of animal models and tissue culture systems have demonstrated that these diseases are caused by a "gain-of-function" mechanism, which results in gene products with new and abnormal function and the expanded polyglutamine tracts are at the core of pathogenesis (Zoghbi and Orr, 1999).

Disease	Gene	Protein	TNR	TNR Size (Normal)	TNR Size (Disease)
Spinobulbar muscular atrophy	*AR*	Androgen receptor (AR)	CAG	9-36	38-62
Huntington's disease	*HD*	Huntingtin	CAG	6-35	36-121
Dentatorubral-pallidoluysian atrophy	*DRPLA*	Atrophin-1	CAG	6-35	49-88
Spinocerebellar ataxia type 1	*SCA1*	Ataxin-1	CAG	6-44	39-82
Spinocerebellar ataxia type 2	*SCA2*	Ataxin-2	CAG	15-31	36-63
Spinocerebellar ataxia type 3	*SCA3*	Ataxin-3	CAG	12-40	55-84
Spinocerebellar ataxia type 6	*SCA6*	α1A-voltage-dependent calcium channel subunit	CAG	4-18	21-33
Spinocerebellar ataxia type 7	*SCA7*	Ataxin-7	CAG	4-35	37-306

Table 1.3: Diseases caused by expanded glutamine tracts (Cummings and Zoghbi, 2000)

1.5 *Cis*-elements and TNRs Instability

Cis acting factors associated with repeat instability can be both internal and external to the repeat tract. Internal factors can be repeat sequence, tract length and purity while the external factors may include CpG methylation, replication origins, flanking sequence elements and nucleosomes (Cleary and Pearson, 2003).

Repeat length has intrinsic connections with instability as evident from the human disease pattern. Only the tracts in the pre-mutation range or longer are prone to instability and longer lengths have more potential to expand than shorter tracts (Pearson et al., 2005; Richards, 2001).

Interruptions in repeat tracts can significantly alter the instability process. CAG tracts of 39 repeats with a single interruption of CAT repeats at the SCA1 locus are somatically stable, while the pure alleles of 40 repeats are unstable (Chong et al., 1995). Similarly, the interruptions in CCG·CGG tracts in the FRAXA locus suppress expansion potential and expansions are observed accompanied by the loss of interruptions (Kovtun et al., 2001).

CpG methylation may stabilize the CGG repeat tracts of FRAXA. In bacteria and primate cells, it was shown that CpG methylation stabilized the sequences from contraction (Nichol Edamura et al., 2005; Nichol and Pearson, 2002).

1.6 DNA Replication in *E. coli*

The DNA replication in *E. coli* is semi-discontinuous, with the leading strand synthesized as uninterrupted chain and lagging strand as a series of short Okazaki fragments. Numerous proteins act to advance the DNA replication fork, which moves about 1 kb per second under normal circumstances (Langston and O'Donnell, 2006).

Parental duplex is unwound by DnaB helicase as it translocates along the lagging strand template in 3'-5' direction, ahead of the leading strand template. The helicase activity of DnaB is stimulated by interaction with τ subunit of DNA plymerase III (Kim et al., 1996). Cellular DNA polymerases cannot initiate synthesis in the absence of a nucleic acid primer, so the first step is the formation of RNA primer by the RNA polymerases, primases. Leading strand synthesis requires only one priming event while continuous priming is required on the lagging strand. The distance of primers, approximately 1-2 kb apart on lagging strand is maintained by interaction of DnaB and DnaG primase (Tougu and Marians, 1996).

Polymerase III, majorly involved in the chromosomal replication is a complex enzyme with multi subunits encoded by different genes (Kelman and O'Donnell, 1995; Marians, 1992). The polymerase activity is provided by the α subunit, encoded by *dnaE*. The proofreading ε subunit, possessing exonuclease activity, is encoded by *dnaQ*. Both leading and lagging strands are co-ordinately synthesized by an asymmetric polymerase dimer. The τ subunit, encoded by *dnaX*, facilitates the interaction between the two core polymearse molecules (Studwell-Vaughan and O'Donnell, 1991).

PolI and II are involved in the repair pathways. PolymeraseI is required for some plasmid replication initiation and may also play a role in the joining of Okazaki fragments.

1.7 Proposed Molecular Mechanisms of Repeat Instability

Since the discovery of dynamic mutation process in 1992, the molecular mechanisms causing repeat instability have been extensively studied and debated. Conflicting

findings have come up in certain areas and main cellular processes replication, repair and recombination have been proposed as potential players.

1.7.1 Secondary Structures Formed by Trinucleotide Repeats

1.7.1.1 Hairpin Formation by CNG Repeats

Trinucleotide repeats, when single-stranded, have been proposed to fold into hairpin structures. Evidence for these structures *in vitro* has been provided by using different strategies such as chemical modification, enzyme probing, gel mobility, UV absorbance and nuclear magnetic resonance (NMR) studies (Gacy et al., 1995; Mitas et al., 1995a; Mitas et al., 1995b; Petruska et al., 1996; Smith et al., 1995; Yu et al., 1995a; Yu et al., 1995b; Zheng et al., 1996).

The hairpins formed by single strands of trinucleotide repeats are less stable than the hairpins formed by palindromes as they have mismatches as compared to the perfect palindromic hairpins (Figure 1.1). The hairpins formed by CAG·CTG and CCG·CGG repeats show differing degrees of stability *in vitro*. CTG hairpins are more stable than CAG hairpins and CGG hairpins are more stable than CGG hairpins (Gacy and McMurray, 1998; Mitas, 1997; Petruska et al., 1996). It was shown that a single strand of ten CTG repeats can form more stable hairpins than the complementary CAG strand (Petruska et al., 1996). This was attributed to the A·A mispairs being bulkier so they cannot fit into the helix while the small thymidines in T·T mispairs fit well into the stem of CTG hairpins. The T·T mispairs in the stem of CTG and GTC hairpins are resistant to the action of $KMnO_4$, indicating that they are stable and well stacked into the helix (Mitas et al., 1995b; Yu et al., 1995b). A·A mispairs are well stacked in GTC hairpins but less well in CTG hairpins as they are more prone to

modification by diethylpyrocarbonate (DEPC) (Zheng et al., 1996). Moreover, it was shown that T·T contains two hydrogen bonds (Mariappan et al., 1996) so is better stacked than an A·A mispair that has only one hydrogen bond.

1.7.1.2 Tetraplexes

Tetraplexes (four stranded DNA) assemble at G-rich sequences forming G-quartets. This tetrad formation has been reported for single stranded G-rich telomeric sequences (Neidle and Parkinson, 2003) and can also be formed by base pairing between cytosines of one duplex with the cytosine residues of another duplex resulting in tetraplexes known as *i*-motif (Fojtik and Vorlickova, 2001). $(CGG{\cdot}CCG)_n$ repeats have been suggested to form tetraplexes (Balakumaran et al., 2000; Fojtik and Vorlickova, 2001) along with hairpins. The hairpins formed have C·C mismatched base pairs that have no hydrogen bonding (Mariappan et al., 1996). Though the long CGG repeats may form hairpins, the association of two hairpins may lead to tetraplex. When the length of CGG exceeds 13 repeats, the hairpin can fold over and form an intramolecular tetraplex (Usdin and Woodford, 1995). These are held together by two G·G·G·G quartets and one C·C·C·C quartet per trinucleotide.

1.7.1.3 Triplexes

Unlike CAG and CTG repeats that form hairpins, GAA·TTC repeats have been observed to adopt both an inter- and intramolecular triplex structures (Potaman et al., 2004) consisting of a Watson-Crick paired GAA·TTC duplex and a third CTT strand, which is non-Watson-Crick paired with the duplex (Gacy et al., 1998). It was shown that GAA repeats can form intramolecular triplex structures by folding of a

pyrimidine rich strand back to pair with the duplex, under the influence of low pH (Hanvey et al., 1988). A bi-triplex structure, sticky DNA, formed by the association of the two py.pu.pu triplexes in negatively supercoiled DNA (Sakamoto et al., 1999) has been shown to inhibit transcription (Sakamoto et al., 2001).

1.7.1.4 Slipped Strand Structures

Slipped strand structures can form within repeated sequences by an out-of-register alignment of complementary duplex strands. The first observation of slipped strand structures was made by Pearson and Sinden (1996). It was demonstrated that alternative structures are formed when DNA containing long tracts of CAG·CTG and CGG·CCG repeats were denatured and renatured (Pearson and Sinden, 1996). The structures had low electrophoretic mobility on polyacrylamide gels and were stable at physiological salt concentrations. The authors named the slipped strand structures S-DNA. The fraction of molecules forming S-DNA increased with increasing repeat length. Further to these observations, S-DNA structures were shown by electron microscopy and denaturing and renaturing $(CTG)_{50}\cdot(CAG)_{50}$, it was demonstrated that S-DNA could form with slipped out regions on both strands (Pearson et al., 1998).

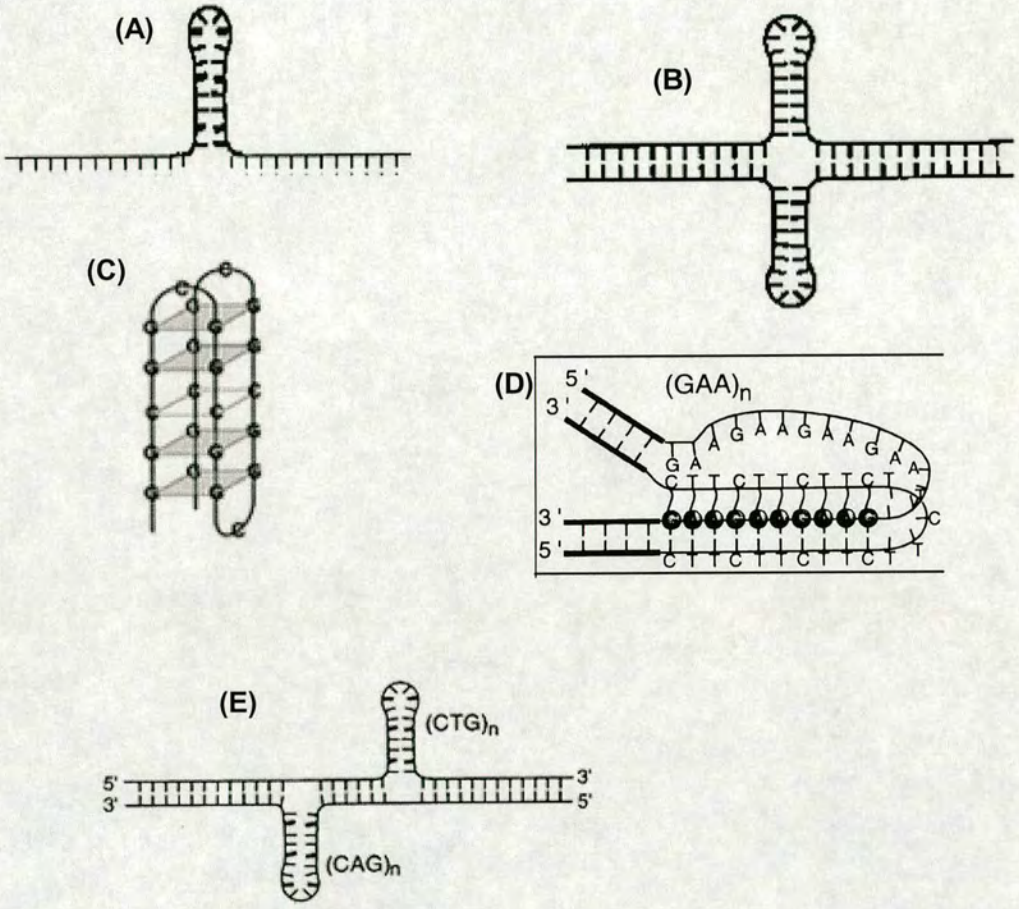


Figure 1.1: Secondary structures formed by triplet repeats (Sinden et al., 2002) (A) Hairpin structure formed by single strands of CAG, CTG, CCG, CGG, GTC and GAC (B) Cruciform structures are formed by perfect palindromes (C) Tetraplex structure as suggested to be formed by CGG repeats (D) Triplex structure as formed by GAA·TTC repeats associated with Friedreich's ataxia (E) Slipped strand structures are formed in CAG·CTG and CCG·CGG repeat sequences.

1.7.2 Replication associated Instability

Several studies have pointed out the importance of replication as a major source of instability of trinucleotide repeats. DNA synthesis can contribute to instability both in proliferative and non-proliferative cells, as DNA synthesis is also needed in repair processes. Using *E. coli* and yeast model systems, instability has been observed as a function of orientation of repeats relative to the replication origin (Freudenreich et al., 1997; Kang et al., 1995a; Maurer et al., 1996; Miret et al., 1998; Rosche et al., 1995), the growth phase of the host cell (Bowater et al., 1996) and the mode of replication of sequences in the plasmid (Schumacher et al., 1998). The observations of the involvement of direction of replication laid the foundation of lagging strand model to explain the expansion and deletion bias observed in *E. coli* and yeast (Figure 1.2). According to this model, deletions are prominent when the more stable secondary structure forming CTG repeats are on the lagging strand template, while expansions are prominent when the stable structure forming repeats are on the Okazaki fragment. Such bias has been explained by the greater stability of secondary structures formed by CTG repeats than the complementary CAG repeats (Petruska et al., 1996; Yu et al., 1995b) and the ease of formation of structures on the template for lagging strand that exhibit the single stranded region during replication (Trinh and Sinden, 1991).

However, this replication orientation bias for instability is not always observed (Schumacher et al., 1998) and using an *in vivo* phagemid system, it was shown that deletion and expansion of triplet repeats may occur on the leading strand (Iyer and Wells, 1999).

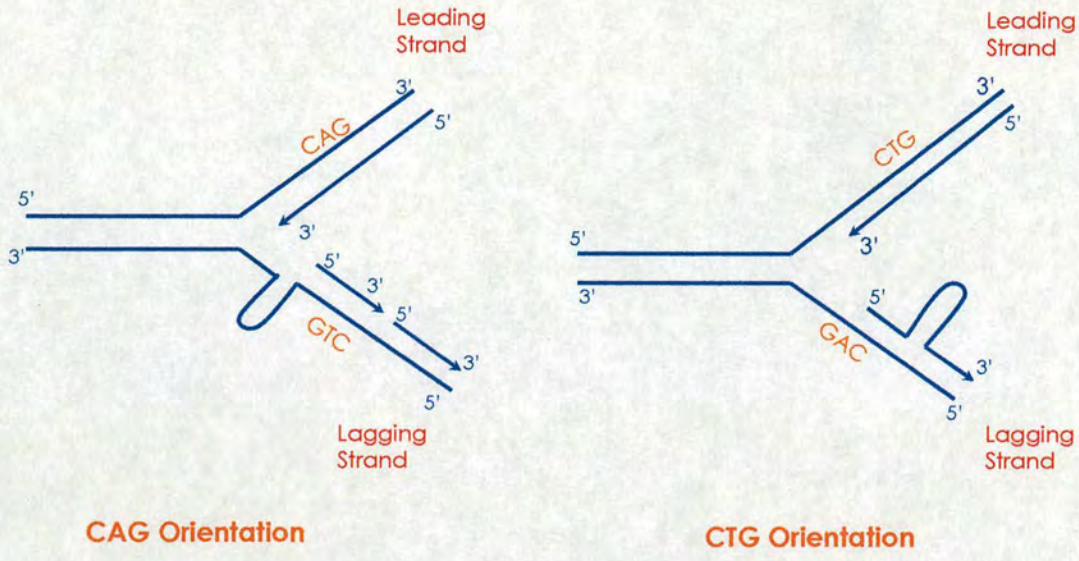


Figure 1.2: Traditional replication slippage model
The model proposes that deletions are favoured more frequently when the more stable secondary structure forming CTG repeats are present on the lagging strand template (CAG orientation) while expansions are more likely to occur when the stable structure forming CTG repeat tract is on the daughter lagging strand (CTG orientation).

1.7.3 Replication Origin or Replication Fork?

In a recent review, the dynamics of replication fork in repeat instability have been pondered upon (Cleary and Pearson, 2005). The authors have discussed three possible scenarios supported by studies in *E. coli*, yeast and primate cells (Figure 1.3).

According to the "origin-switch" model, the switch in the origin of replication changes the direction of replication. This changes the sequence for the lagging strand template and alters repeats instability, which depends on the repeat sequence on template and nascent strands (Cleary et al., 2002; Kang et al., 1995a; Maurer et al., 1996). This model gets further support from observations that replication fork stalling at repeat tracts depends partly on the orientation of the repeats relative to the

replication origin (Pelletier et al., 2003; Samadashwily et al., 1997) and the altered instability of repeat tracts in yeast FEN1/rad27 mutants (Freudenreich et al., 1998; Schweitzer and Livingston, 1998). This model suggests that in humans, the affected chromosome may have the origin of replication on the opposite side of the repeat tract as compared to the non-affected chromosome.

The "origin-shift" model gathers support from the observations made in primate cells where shifting the location of replication origin while maintaining the direction of replication, changed instability drastically (Cleary et al., 2002). A similar observation was reported recently for CGG·CCG repeats (Nichol Edamura et al., 2005).

In the "fork-shift" model, it is proposed that *cis*-elements within the repeats or flanking the repeat sequences affect the dynamics of the advancing replication fork to generate repeat instability (Figure 1.3). The position of the repeat tract within the advancing fork determines the location of Okazaki initiation, termination and processing and so determines the chances of any mutagenic process. So, repeat instability is dependent on the repeat sequence, length and portion of the repeat within the Okazaki initiation zone (OIZ), which is the ss region in the lagging strand template after which the lagging strand synthesis starts. This model attempts to explain the pattern of instability between different disease loci and between tissues of the same patient. Since it does not depend on the origin or direction of replication, it is responsive to epigenetic surroundings and cellular processes.

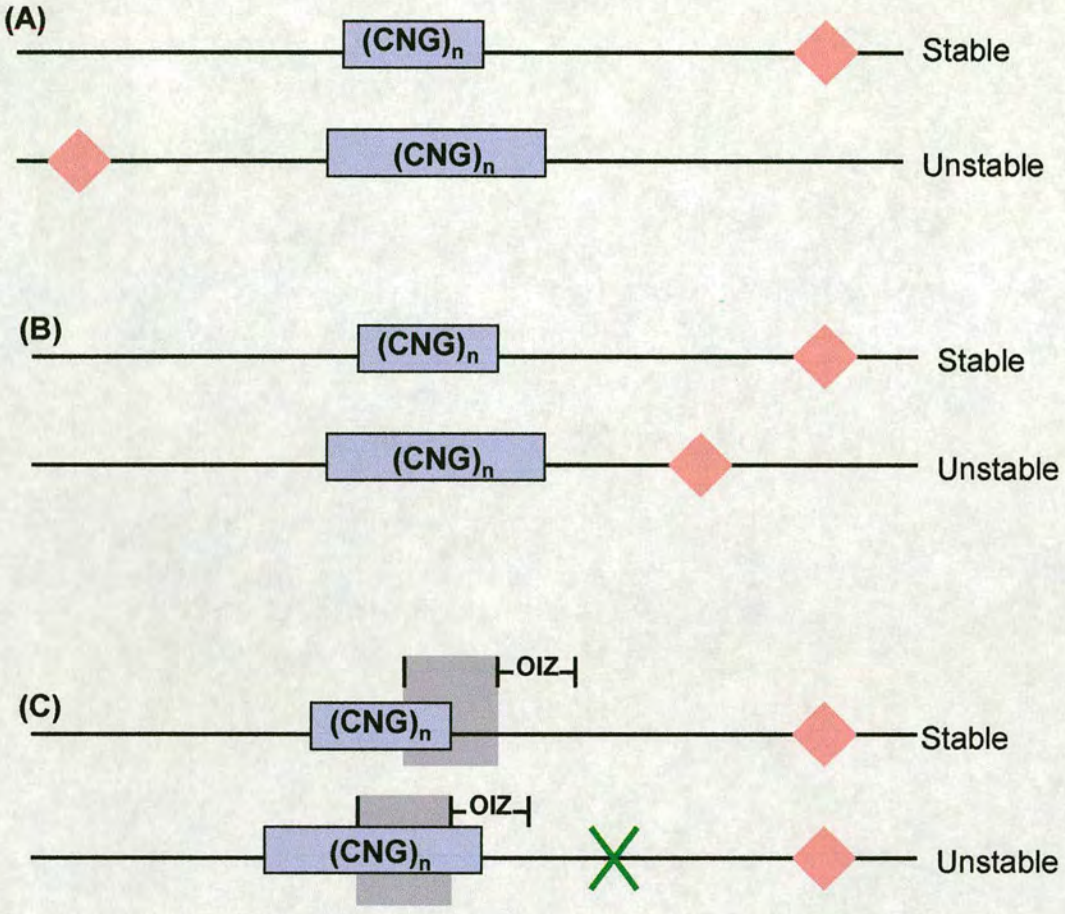


Figure 1.3: Dynamics of replication origin and replication fork in repeat instability (Cleary and Pearson, 2005)
(A) Origin-switch model: The direction of replication is determined by the origin of replication (pink diamond shape). A switch in the origin of replication will change the sequence for lagging strand template, which in turn may decide whether the instability will happen or not. (B) Origin-shift model: The location of origin of replication on one side affects the instability, because of the position of repeats within Okazaki initiation zone (OIZ). The unstable chromosome may have the origin of replication near to the repeats than on the stable chromosome, thus producing instability. (C) Fork-shift model: Any epigenetic event or cellular process (shown by X in green) that occurs in the path of replication fork can alter the dynamics of fork progression, which may shift the location of OIZ relative to the repeat tract. That may result in the formation of mutagenic DNA structure leading to instability.

1.7.4 Flap Endonuclease 1 (FEN 1) mediated instability

Replication associated proteins have also been investigated for their participation in repeat instability. Particularly, the inability of flap endonuclease 1 (FEN1) to process replication-induced errors has been proposed to generate expansions. FEN1 is a structure specific exo/endonuclease involved in removal of RNA primers on Okazaki fragments. It has a preference for substrates with a 5' unannealed flap (Harrington and Lieber, 1994). In *E. coli*, the homologue of FEN1 is the 5' to 3' exonucleolytic activity of DNA polymerase I and in yeast FEN1 shares homology with Rad27. The function of FEN1 is critical for maintaining genome stability. CAG and CCG sequences were shown to expand in *rad*27 mutants (Schweitzer and Livingston, 1998). The effect of mutations in human FEN1 on sequence expansion was assayed in yeast and FEN1 was found to compete with ligase I for suppressing expansions (Liu and Bambara, 2003). A model was proposed to explain the role of FEN1 in preventing expansions of repeat tracts (Figure 1.4). According to this model, strand displacement during lagging strand synthesis produces a flap of triplet repeats which can fold onto itself because of its repetitive nature. Such a folded flap is not a substrate for FEN 1. Through subsequent slippage, an equilibrium is generated between a hairpin structure and an unstable large bubble. DNA ligase cannot act on a large bubble with a large unannealed region. However, the large unstable bubble can equiliberate to a small stable bubble with a long annealed region, which can be a good substrate for DNA ligaseI, which seals the nick and generates an expanded product. Alternatively, if the unstable bubble forms a flap, FEN1 can remove it and DNA polymerase can synthesize through the gap leaving the nick to be sealed by DNA ligase. This results in an non-expanded product.

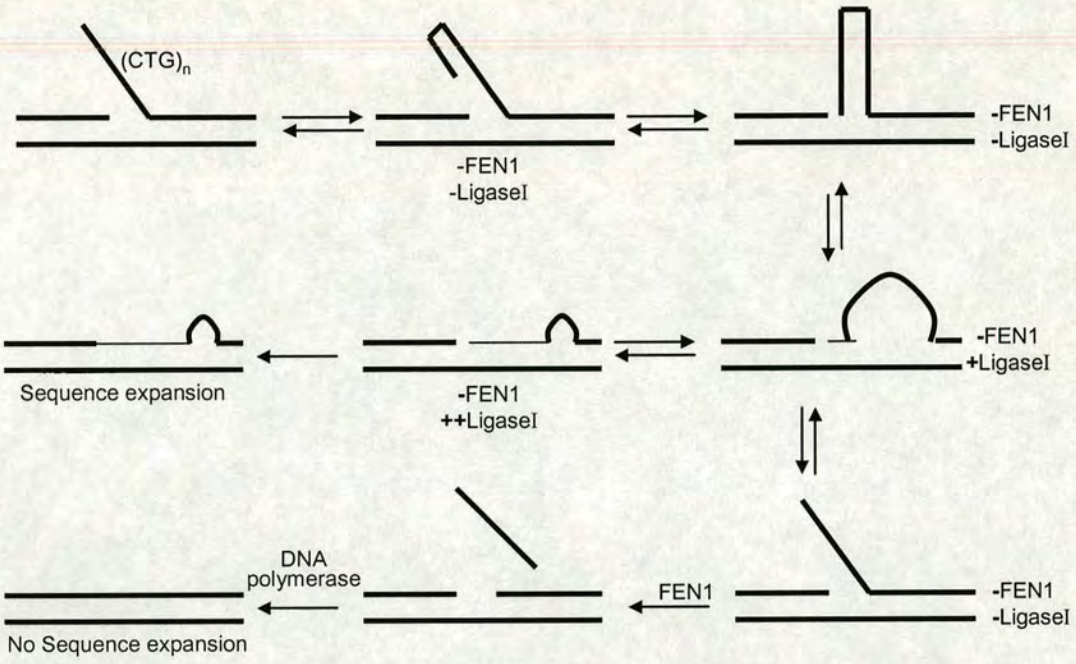


Figure 1.4: Proposed model of triplet repeat expansion involving FEN1 and DNA liagase I (Liu and Bambara, 2003)
During lagging strand synthesis, the CTG flap is created which can fold back to form hairpin structure. This cannot be utilized as substrate for FEN1 or DNA ligase I. Subsequent DNA slippage and misalignment allows the equilibrium between the hairpin and an unstable large bubble. DNA ligase I cannot act on a large bubble with a large unannealed region. However, the unstable large bubble can equilibrate to small stable bubble with a long annealed region, which can be a good substrate for DNA ligase I that seals the nick leaving an expansion product. Alternatively, the bubble forms a flap substrate that can be removed by FEN1 endonuclease. DNA polymerase fills the gap and the nick is sealed by DNA ligase I resulting in a non-expanded product.

1.7.5 TNRs cause DNA polymerase pausing

DNA containing TNRs is difficult for DNA polymerases to copy, which can be experienced in a standard PCR. Primer extension experiments in *E. coli* using DNA polymerase I, T7 polymerase have shown polymerase stalling in CTG and CCG repeats (Kang et al., 1995b). Using two dimensional gel electrophoresis, it has been shown that replication forks face hindrance while progressing through TNRs (Krasilnikova and Mirkin, 2004; Pelletier et al., 2003; Samadashwily et al., 1997).

The pausing was influenced by repeat length, purity and repeat orientation relative to replication origin.

The rescue from a paused replication fork may involve slippage events, double strand breaks, repair or recombination processes that may result in further instability. The instability associated with these processes is discussed in the corresponding Chapters along with discussion of this work.

1.8 Work in this thesis

This thesis describes work carried out to investigate CAG·CTG repeat instability in the *E. coli* chromosome. To study repeat instability in the *E. coli* chromosome was of interest for two main reasons. First, *E. coli* is a well-characterized organism with a fine understanding of its processes of DNA replication, repair, recombination and transcription, genetically and biochemically. Second, all studies of TNRs instability in *E. coli* have been carried out using plasmid substrates which have limitations of plasmid biology such as the copy number and the replicon size. In these systems, TNRs are localized on a plasmid, which may differ from chromosomal marker in terms of their repair and replication. Another disadvantage is that the plasmids bearing TNRs have to be introduced into the cells by a transformation step which has been shown to influence the frequency of instability of some inserts (Hashem et al., 2002). Furthermore, two plasmid based systems have been used to study the effect of recombination on repeat instability but the reproducibility of results obtained by such systems is already being questioned and will be discussed in detail in Chapter 5. All these studies have brought forward conflicting reports and have significantly complicated progress in the field. So, it was important to utilize efficiently a well-

studied model system, *E. coli*. Studying repeat instability in the *E. coli* chromosome is a novel investigation, which will aid in clarifying the ambiguities added by plasmid based studies.

First, a polymerisation independent strategy was developed to generate long CAG·CTG repeat arrays in plasmid p*Lac*D2. Then the repeat sequences were integrated in the 5' end of the *lacZ* gene of *E. coli*. Instability was studied in wild type *E. coli* and strains deficient in cellular processes of DNA replication, repair and recombination. The role of transcription was analysed in parallel in wild type and mutant strains.

In wild type cells, instability was investigated for various repeat lengths in both CAG and CTG repeat orientations (Chapter 3). Instability was found to be dependent on repeat length and orientation with respect to the direction of replication. Considering the observations of hairpin formation by repeat sequences *in vitro*, the role of hairpin nuclease SbcCD and the proofreading function of DNA polymerase III (DnaQ) was investigated. Orientation dependent instability observed in wild type cells was lost in a *dnaQ* mutant strain (Chapter 4). A model explaining the orientation dependent instability is proposed, based on the dynamics of secondary structure formation and processing.

Chapter 5 describes the investigation of the role of recombination in CAG·CTG repeat instability. The roles of homologous recombination genes *recA*, *recB*, *recF*, *ruvAC* and *recG* were investigated. Furthermore, dimer resolution was investigated with mutations in *xer* and *dif*. No effect of recombination was observed on instability.

The work was further extended to study the role of DNA repair in instability. The genes encoding proteins involved in mismatch (*mutH*, *mutL* and *mutH*) and

nucleotide excision repair (*uvrA*, *uvrB* and *uvrC*) were analysed. No effect of mismatch and nucleotide excision repair on CTG repeat instability was observed. MutS was found to have a stabilizing role in CAG orientation. Additionally, the role of mismatch repair genes was questioned in conjunction with SbcCD. It was observed that SbcCD could cleave hairpin structures in the presence of MutL and MutH in the CAG orientation. The results and a hypothesis predicted for the involvement of SbcCD and MutS are discussed in Chapter 6.

In Chapter 7, the role of transcription in repeat instability is discussed. Since, repeats were integrated in the *lacZ* gene, the phenotype of strains having *lacZ* with varying repeat length was characterized. Additionally, the size of β-galactosidase with various repeat lengths was determined by Western blotting.

Furthermore, the role of two DNA helicases; a repair helicase, UvrD and the replicative helicase, Rep was investigated (Chapter 8).

This work is a novel investigation of CAG·CTG repeat instability. Studying repeat instability in the *E. coli* chromosome has revealed interesting findings, which are a significant contribution in our endeavours to understand the dynamics of instability caused by these dynamic sequences.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 General Reagents

2.1.1.1 Chemicals

All chemicals were purchased from the following suppliers: Sigma, Fisher Scientific, Invitrogen, Malford Labs and Calbiochem.

2.1.1.2 Enzymes

All restriction enzymes were purchased from New England Biolabs (NEB) and Roche. *Taq* DNA polymerase was purchased from Roche and PFU polymerase from Promega.

2.1.1.3 Growth Reagents

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

2.1.2 Bacterial Growth Media

2.1.2.1 Bacterial Media

Table 2.1 Bacterial Media

Media	Composition
BBL Agar	1% Trypticase, 0.5% NaCl, 1% Bacto-agar, pH adjusted to 7.2 with NaOH
BBL Top Agar	1% Trypticase, 0.5% NaCl, 0.65% Bacto-agar, pH adjusted to 7.2 with NaOH
LB Agar	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% Bacto-agar, pH adjusted to 7.2 with NaOH
Lc Agar	1% Tryptone, 0.5% NaCl, 1% Difco-agar, pH adjusted to 7.2 with NaOH
Lc Top Agar	1% Tryptone, 0.5% NaCl, 0.7% Difco-agar, pH adjusted to 7.2 with NaOH
L Broth	1% Bacto-tryptone, 0.5% yeast extract, 1% NaCl, pH adjusted to 7.2 with NaOH

2.1.2.2 Antibiotics

Antibiotics were added to liquid media immediately before use. For the solid media, the antibiotics were added after autoclaving. Solid media was always provided autoclaved. It was melted using a microwave and stored at 55 °C if not to be used immediately. The antibiotics were added at media temperature ~ 55 °C (when easily touchable by naked hand). All antibiotics were stored at –20 °C. The stock and final concentrations of antibiotics are shown in Table 2.2.

Table 2.2 Antibiotics

			Stock Concentration	Final Concentration
Antibiotic	Abbreviation	Solvent	(mg/ml)	(μg/ml)
Chloramphenicol	Cm	100% Ethanol	50	50
Ampicillin	Amp	Water	100	100
Kanamycin	Km	Water	50	50
Tetracyclin	Tc	50% Ethanol	15	15
Rifampicin	Rif	Methanol	50	50

2.1.3 Buffers and Solutions

Table 2.3 Buffers

Buffer	Components
20X TAE	0.8M Tris, 0.4M Sodium Acetate, 0.02M EDTA, pH 8.2 (for Agarose gel)
Phage buffer	7g Na_2HPO_4, 3g KH_2PO_4, 5g NaCl, 10ml $MgSO_4$ (0.1M), 10ml $CaCl_2$ (0.01M), 1ml Gelatin (1%) for a volume of 1 Litre
10x Transfer buffer	30.25g Tris, 144g glycine for a volume of 1 Litre (filter the buffer) (for Western blotting)
20x MOPS	10g SDS, 104.6g MOPS, 60.6g Tris and 3g EDTA for a volume of 500ml (for Polyacrylamide gel)
Sample lysis buffer	Lysis buffer, inhibitor cocktail and pepstatin in a ratio of 100:10:0.1 v/v (for Western blotting samples)

2.1.4 Plasmids

Plasmids used in this work are listed below in Table 2.4. All plasmids were constructed in XL1-blue cells.

Table 2.4 Plasmids

Plasmid	Characteristics	Source	DL No.
pTOF24	*repAts sacB* Cm^r;used for *SalI-PstI* cloning	Millicent Masters	1605
p*Lac*D1	pTOF24 derivative; contains *Bbs*I, *Mfe*I & *Bsa*I sites in centre of two 400bp *lac* homology arms, L8 mutation	John Blackwood	1823
p*Lac*C1	p*Lac*D1 derivative; no L8 mutation	John Blackwood	1824
p*Lac*D2	p*Lac*D1 derivative; *Bbs*I site in *lacZ* homology arm removed	This work	2911
pLacC2	p*Lac*C1 derivative; *Bbs*I site in *lacZ* homology arm removed	This work	3215
p*Lac*D2 $(CAG)_5$	p*Lac*D2 derivative; $(CAG)_5$ in place of *Mfe*I site	This work	1816
p*Lac*D2 $(CAG)_8$	p*Lac*D2 derivative; $(CAG)_8$ in place of *Mfe*I site	This work	1892
p*Lac*D2 $(CTG)_8$	p*Lac*D2 derivative; $(CTG)_8$ in place of *Mfe*I site	This work	1893

p*Lac*D2 $(CAG)_{14}$	p*Lac*D2 derivative; $(CAG)_{14}$ in place of *Mfe*I site	This work	1899
p*Lac*D2 $(CAG)_{26}$	p*Lac*D2 derivative; $(CAG)_{26}$ in place of *Mfe*I site	This work	1894
p*Lac*D2 $(CTG)_{26}$	p*Lac*D2 derivative; $(CTG)_{26}$ in place of *Mfe*I site	This work	1895
p*Lac*D2 $(CAG)_{28}$	p*Lac*D2 derivative; $(CAG)_{28}$ in place of *Mfe*I site	This work	1900
p*Lac*D2 $(CTG)_{28}$	p*Lac*D2 derivative; $(CTG)_{28}$ in place of *Mfe*I site	This work	1901
p*Lac*D2 $(CAG)_{50}$	p*Lac*D2 derivative; $(CAG)_{50}$ in place of *Mfe*I site	This work	1911
p*Lac*D2 $(CTG)_{50}$	p*Lac*D2 derivative; $(CTG)_{50}$ in place of *Mfe*I site	This work	1912
p*Lac*D2 $(CAG)_{98}$	p*Lac*D2 derivative; $(CAG)_{98}$ in place of *Mfe*I site	This work	2912
p*Lac*D2 $(CTG)_{98}$	p*Lac*D2 derivative; $(CTG)_{98}$ in place of *Mfe*I site	This work	2913
p*Lac*D2 $(CAG)_{194}$	p*Lac*D2 derivative; $(CAG)_{194}$ in place of *Mfe*I site	This work	2914
p*Lac*C2 $(CGG)_5$	p*Lac*C2 derivative; $(CGG)_5$ in place of *Mfe*I site	This work	1993
p*Lac*C2 $(CCG)_5$	p*Lac*C2 derivative; $(CCG)_5$ in place of *Mfe*I site	This work	2007
p*Lac*C2 $(CGG)_8$	p*Lac*C2 derivative; $(CGG)_8$ in place of *Mfe*I site	This work	2368
p*Lac*C2 $(CCG)_8$	p*Lac*C2 derivative; $(CCG)_8$ in place of *Mfe*I site	This work	2254
pTOF24-*rep*	pTOF24 derivative; to integrate *rep* deletion	John Blackwood	2341
pTOF24-*uvrD*	pTOF24 derivative; to integrate *uvrD* deletion	John Blackwood	2391
pTOF24-*mfd*	pTOF24 derivative; to integrate *mfd* deletion	This work	2519
pTOF24-*uvr*A	pTOF24 derivative; to integrate *uvrA* deletion	Ewa Okely	2699
pTOF24-*uvrB*	pTOF24 derivative; to integrate *uvrB* deletion	Ewa Okely	2700
pTOF24-*uvrC*	pTOF24 derivative; to integrate *uvrC* deletion	Ewa Okely	2701
pTOF24-*recA*	pTOF24 derivative; to integrate *recA* deletion	Ewa Okely	2711
pTOF24-*mutH*	pTOF24 derivative; to integrate *mutH* deletion	Ewa Okely	2714
pTOF24-*mutL*	pTOF24 derivative; to integrate *mutL* deletion	Ewa Okely	2721

2.1.5 *E. coli* Strains

E. coli strains used in this work are listed below in Table 2.5.

Table 2.5 Strains

Strain No.	Genotype	Source
732	F⁻ *thr-1 leuB6 proA2 his4 thi1 argE3 lacY1 galK2 rpsL supE44 ara-14 xyl-15 mtl-1, txs-33) sbcCD*::Km	This Laboratory
815	*ara Δ(gpt-lac)5 rpsL mutS*::mini-Tn10	Cold Spring Harbour Lab (CSH115)
962	F⁻ *araD*139 Δ(*argF lacIPOZYA*)U169 *strA relA fla dnaQ*::mini-Tn10	Geneviève Maenhaut-Michel (Geuskens et al., 1991)

1675	MG1655	Millicent Masters
1786	MG1655 *lacZ*χ^- *lacI*^{q} Zeo$^R\chi^+$	John Eykelenboom
1991	DL1786 *lacZ*::$(CAG)_{28}$ L8	This work
1994	DL1786 *lacZ*::$(CTG)_{48}$ No L8	This work
1995	DL1786 *lacZ*::$(CAG)_{75}$ No L8	This work
2001	DL1675 *lacZ*::$(CAG)_{50}$	This work
2002	DL1675 *lacZ*::$(CAG)_{28}$	This work
2009	DL1786 *lacZ*::$(CTG)_{95}$ No L8	This work
2052	DL1786 *lacZ*::$(CTG)_{28}$ L8	This work
2079	DL2009 *recA*::Cm (DL654)	John Blackwood
2080	DL2009 *recB*::Km (DL1792)	John Blackwood
2081	DL2009 *recF*400::Km (DL1876)	John Blackwood
2099	DL2009 *recG*::Km (DL1077)	John Blackwood
2205	DL1786 *lacZ*::$(CAG)_{14}$ L8	This work
2104	DL2009 *sbcCD*::Km (DL732)	John Blackwood
2247	DL1786 *lacZ*::$(CAG)_8$ No L8	This work
2250	DL1786 *lacZ*::$(CAG)_{45}$ L8	This work
2266	DL1786 *lacZ*::p*Lac*D2 L8	This work
2278	DL2001 *mutS*::Tc *sbcCD*::Km	This work
2279	DL2300 *sbcCD*::Km (DL732)	This work
2280	DL1786 mutS::Tc *sbcCD*::Km	This work
2281	DL1994 *mutS*::Tc sbcCD::Km	This work
2300	DL2009 *mutS*::Tc (DL815)	This work
2301	DL1995 *dnaQ*::Tc (DL962)	This work
2302	DL1995 *mutS*::Tc (DL815)	This work
2303	DL1995 *sbcCD*::Km (DL732)	This work
2304	DL1995 *recA*::Cm (DL654)	This work
2305	DL2009 expansion to $(CTG)_{140}$	This work
2306	DL1786 *lacZ*::$(CCG)_8$	This work
2347	DL1786 *lacZ*::$(CTG)_{14}$ L8	This work
2384	DL2009 *rep*$^-$ (using pDL2341)	John Blackwood
2436	DL1995 *recF*400::Km (DL1876)	This work
2437	DL1995 *recB*::Km (DL1792)	This work
2438	DL2302 *sbcCD*::Km (DL732)	This work
2445	DL1786 *lacZ*::$(CTG)_{75}$ *dnaQ*::Tc (DL962)	This work
2550	DL1786 *lacZ*::$(CTG)_{87}$ *dif*::Km (DL1385)	John Blackwood
2555	DL2009 *ruvAC*::Tc (DL1076)	John Blackwood
2558	DL1995 *xerC*::minicat (DL1065)	This work

2559	DL2009 *xerC*::minicat (DL1065)	This work
2629	DL2009 *uvrD*$^-$ (using pDL2391)	John Blackwood
2639	DL1786 *lacZ*::$(CAG)_{84}$ No L8	This work
2678	DL2384 *sbcCD*::Km (DL732)	This work
2709	DL2305 expansion to $(CTG)_{163}$	John Blackwood
2728	DL2709 expansion to $(CTG)_{184}$	This work
2729	DL2009 *uvrB*$^-$ (using pDL2700)	This work
2730	DL2629 *recF*400::Km (DL1876)	This work
2737	DL2550 *recA*$^-$ (using pDL2711)	This work
2744	DL2009 *uvrC*$^-$ (using pDL2701)	This work
2751	DL2009 *uvrA*$^-$ (using pDL2699)	This work
2781	DL2009 *mutH*$^-$ (using pDL2714)	This work
2782	DL2009 *mutL*$^-$ (using pDL2721)	This work
2783	DL1995 *mutH*$^-$ (using pDL2714)	This work
2784	DL1995 *mutL*$^-$ (using pDL2721)	This work
2785	DL2104 *mutH*$^-$ (using pDL2714)	This work
2786	DL2104 *mutL*$^-$ (using pDL2721)	This work
2787	DL2303 *mutH*$^-$ (using pDL2714)	This work
2788	DL2639 *rep*::Km (DL2633)	This work
2789	DL2639 *uvrD*$^-$ (using pDL2391)	This work
2831	DL1995*mfd*$^-$ (using pDL2519)	This work
2915	DL2009*mfd*$^-$ (using pDL2519)	This work
2976	DL2303 *dnaQ*::Tc (DL962)	This work
3046	DL2104 *dnaQ*::Tc (DL962)	This work
3052	DL2303 *mutL* (using pDL2721)	This work
3138	DL1786 *lacZ*::$(CAG)_{80}$ *recF*400::Km (DL1876)	This work

2.1.6 Bacteriophage λ strains

Table 2.6 Bacteriophage λ Strains

Strain	Genotype	Source
λDRL152	*pal571* Δ*spi6 cI857* χ^+	This laboratory
λDRL154	*pal571* Δ*spi6 cI857*χ^+ 571 bp palindromic sequence	This laboratory

2.1.7 Oligonucleotides

All oligonucleotides were purchased from MWG apart from the fluorescent oligonucleotides, which were purchased from Metabion. A list of oligonucleotides used in this work is shown in Table 2.7 where restriction enzyme sites and any changes (in red) are underlined.

Table 2.7 Oligonucleotides

Name	Sequence (5' to 3')
ExCAG-01	CTATGACCATGGTCTCGCAGCAGCAGCAGCAGGTCTTCGTCGTTTTAC
ExCTG-01	GTAAAACGACGAAGACCTGCTGCTGCTGCTGCGAGACCATGGTCATAG
ExCAG-02	CTATGACCATGGTCTCGCTGCTGCTGCTGCTGGTCTTCGTCGTTTTAC
ExCTG-02	GTAAAACGACGAAGACCAGCAGCAGCAGCAGCGAGACCATGGTCATAG
Ex-test-F	TTATGCTTCCGGCTCGTATG
Ex-test-R	GGCGATTAAGTTGGGTAACG
FAM-Ex-test-F	Fam- TTATGCTTCCGGCTCGTATG
FAM-Ex-test-R	Fam- GGCGATTAAGTTGGGTAACG
MFD-1	AAAAACTGCAGCACAGCGTGCATTCAG
MFD-2	TGCGGCTGTTTTGGAAGAGTTTCCCAGTCC
MFD-3	AAAAAGGTACCGCCACGTTGTGTCTCAAAATC
MFD-4	AAAAAGTCGACAATTTGTTTTCGCCAACCAG
SDM_BbsI_F	GGGATACGACGATACCGAGGACAGCTCATG
SDM_BbsI_R	CATCAGCTGTCCTCGGTATCGTCGTATCCC
ExCCG-01	CTATGACCATGGTCTCGCCGCCGCCGCCGCCGGTCTTCGTCGTTTTAC
ExCGG-01	GTAAAACGACGAAGACCGGCGGCGGCGGCGGCGAGACCATGGTCATAG
ExCCG-02	CTATGACCATGGTCTCGCGGCGGCGGCGGCGGGTCTTCGTCGTTTTAC
ExCGG-02	GTAAAACGACGAAGACCCGCCGCCGCCGCCGCGAGACCATGGTCATAG

2.2 Microbiological Methods

2.2.1 Growth of Bacteria

E. coli strains were streaked out from the –70 °C glycerol stock onto LB agar plates or LB containing the appropriate antibiotic to obtain single colonies. A single colony was used to inoculate 5 ml of L Broth in a 10ml 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 at 4 °C on solid medium plates 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. 700 µl of cell culture was mixed with 700 µl of 80% sterile glycerol and mixed gently. The tubes were sealed with Parafilm™ before storing at –70 °C.

2.2.3 $CaCl_2$ Transformation

E. coli strain to be transformed was grown overnight at 37 °C. It was then diluted 1:50 in LB and grown for 2 hours at 37 °C. 1ml of this culture was taken into 2ml Eppendorf tubes and spun down for 30 seconds. Supernatant was removed and pellet was resuspended into 500 µl of 0.1 M Cold $CaCl_2$ (freshly prepared). This was kept on ice for 30 minutes and then centrifuged for 30 seconds in cold room. The presence of a hole in cell pellet indicated that the cells were efficiently competent but its

absence does not mean that cells are not competent. Supernatant was discarded and pellet was resuspended in 100 μl of 0.1M $CaCl_2$. Plasmid DNA or ligation mix was added to the cells and the mix was kept on ice for 30 minutes. The cells were subsequently heat-shocked for 5 minutes at 37 °C. Following, 400 μl of LB was added to the cells that were kept for ~1 hour recovery at appropriate temperature. Finally 100 μl of cells were plated on selective media and incubated at appropriate temperature.

2.2.4 Integration of pKO24 derivative Plasmid

The plasmid was transformed into recipient cells using chloramphenicol selection at 30 °C. Single colony was streaked onto fresh chloramphenicol LB plate and grown overnight at 42 °C. This resulted in a mixture of very small colonies and larger healthier colonies. Cells in larger colonies would have integrated the plasmid into their genome. These integrants were purified onto new chloramphenicol LB plates and grown overnight at 42 °C. A large colony was picked and grown overnight in 5 ml LB at 30 °C. This culture was diluted 10^{-5} in fresh LB and plated onto LB agar containing 5 % sucrose. This provided a selection for the excision of the plasmid from the chromosome. Sucrose resistant colonies were 'patch-tested' simultaneously onto LB containing 5 % sucrose and LB containing 5 % sucrose and chloramphenicol. Chloramphenicol sensitive colonies were tested for the required alteration via polymerase chain reaction (PCR) and sequencing.

2.2.5 P1 Transduction

Upon infecting *E. coli*, phage P1 packages random fragments of bacterial chromosome (up to 2 minutes) into phage particles that can be injected into the recipient *E. coli* strains. So the transfer of a mutation of interest from one strain to another can be achieved by P1 transduction. For that purpose, a P1 lysate was grown on *E. coli* that carries the desired mutation and a cotransducible marker, generally an antibiotic resistance, which helps in the selection of transductants. The recipient *E. coli* is then transduced with P1 lysate. Transductants are identified by selection for linked antibiotic resistance and the presence of the mutation is confirmed by testing for the mutant phenotype or by PCR.

2.2.5.1 P1 Plate Lysate Preparation

An overnight culture was diluted 1/10 in 5ml of LB containing 2.5mM $CaCl_2$ and grown for 2 hours at 37 °C while shaking. 200 µl of the this culture was mixed with 100 µl of diluted P1 lysate and incubated at 37 °C for 30 minutes to allow for phage adsorption. 2.5 ml of Lc top agar containing 5mM $CaCl_2$ was added to the cells and poured onto fresh Lc plates containing 5mM $CaCl_2$. The plates were incubated upright at 37 °C. The recommended time for incubation was 6-8 hours but I always incubated overnight, doing it as the last thing on the day. Cells only control was included in each P1 lysate preparation.

Next day, the P1 dilution giving confluent lysis was selected. 5 ml of phage buffer was poured onto the plate and the top agar was scraped off and pipetted into a bijoux bottle. Few drops of chloroform were added (done in fume hood). The mixture

was mixed gently and then left to stand for 30 minutes at 4 °C. 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.

Notes:

The P1 lysate used to carry the new mutation of interest should not already harbour the same antibiotic resistance gene.

Sterile bottles used for storing P1 lysates should not be washed in detergent. This increases the 'life' of the lysate.

2.2.5.2 P1 Transduction

The strain to be genetically altered was grown overnight in L-broth containing 2.5mM $CaCl_2$. 1 ml of overnight culture was taken in 1.5ml Eppendorf tube and spun at full speed for 1 minute. The cells pellet was resuspended in 100 μl of L-broth containing 2.5mM $CaCl_2$ and 100 μl P1 lysate was added. This mix was incubated at 37 °C under agitation for 20 minutes. 800 μl of L-broth containing 2mM Sodium Citrate was added to prevent further infection. This was again incubated at 37 °C shaking for 30-60 minutes (expression time). 100 μl of the aliquot was plated on selective plates and incubated overnight at appropriate temperature.

Cells only and P1 only controls were used for each P1 transduction experiment.

2.2.6 Mutations Phenotypic Tests

2.2.6.1 UV Irradiation

Ultra-violet (UV) radiation was used to phenotypically test mutations in genes involved in DNA recombination such as *recA*, *recBCD* and *recFOR*.

The strain of interest was diluted 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 seconds to 1 min, using UV lamp on 8th Floor. Both plates were incubated overnight at 37 °C. The cultures showing no/poor growth after UV exposure were taken for further work.

2.2.6.2 *sbcCD* Phenotypic Test

An overnight culture of the strain to be tested for *sbcCD* mutation was diluted 1:250 in 5ml L-broth containing 5mM $MgSO_4$ and 0.2% maltose and grown for one and half hour at 37 °C. 250μl of culture was added to 2ml of top BBL and the mix was poured on BBL plates containing $CaCl_2$ (freshly prepared). Once settled, dilutions of phages DRL152 (-pal) and DRL154 (+pal) were spotted. Plates were incubated at 37 °C overnight. The cultures giving more plaques for DRL154 (+pal) indicated successful *sbcCD* mutation and were chosen for further analysis.

2.2.6.3 *dnaQ* and *mutS* Phenotypic Test

dnaQ and *mut*S mutations are hypermutator phenotypes, which render cells resistant to antibiotic rifampicin. The strain to be tested was grown overnight in liquid culture and diluted to 10^{-2}. Then 10μl of undiluted culture and the dilutions (10^{-1} and 10^{-2}) were spotted on rifampicin plates. The culture showing higher resistance to rifampicin was selected for further work.

2.3 DNA Methods

2.3.1 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 5-7ml of *E. coli* culture grown to stationary phase with appropriate antibiotic and eluted in 25 – 50 µl of TE and stored at –20 °C.

2.3.2 *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.

2.3.3 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 and BSA, where applicable. Digestions were incubated at the appropriate temperature between 2 to 5 hours. Digested DNA was analysed on 1% agarose gel electrophoresis. Where necessary, the DNA of required size was extracted from gel.

2.3.4 Agarose gel electrophoresis

DNA was analysed for size, quantity or quality on 1- 2% (w/v) agarose gels. Gels were prepared by melting agarose in 1X TAE buffer using a microwave. The gel solution was kept at 55-60 °C for a period of 1 to 3 weeks without loss of band separation resolution. Usually 2 µl of 60% glycerol was added to 5 µl of DNA to be analysed and loaded directly into the gel wells, immersed in 1X running TAE buffer. A potential difference was supplied across the gel so that the DNA fragments separated relative to their size (100 to 150 volts). The DNA was stained by immersing the gel in a solution of ethidium bromide (2μg/ml) for 10 minutes 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 and intensity of DNA bands to be determined.

2.3.5 Purification of DNA from Agarose Gels

DNA fragments were purified from a population of differently sized DNA molecules by separating on agarose gel electrophoresis. The desired DNA band size 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.

2.3.6 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). For ligation of two fragments of p*Lac*D2, equal volume of fragments was taken. Reactions were allowed to proceed at 16 °C overnight.

2.3.7 *E. coli* boiled Cells

Often boiling cells produced DNA of adequate quality for PCR and sequencing. A colony was picked up with a sterile yellow tip, suspended in 0.5 ml tube containing 30µl MilliQ™ and boiled in a PCR machine at 99.9 °C for 10 minutes. For a pool of 5 colonies, 50µl MilliQ™ was used. After centrifugation for 5-10 minutes, 2 µl of supernatant was typically used for PCR.

Note:

The use of wooden toothpicks for suspending cells is not recommended from the experience of other lab fellows.

2.3.8 Polymerase Chain Reaction

Regions of DNA were specifically amplified using the polymerase chain reaction (PCR). A typical PCR reaction (50μl volume) using plasmid or genomic DNA as template was as follows:

Table 2.8 A typical PCR reaction (50μl volume)

Ingredients	Stock Concentration	Final Concentration (per reaction)	Volume per reaction (μl)
Forward Primer	5pmol/μl	0.1pmol/μl	1
Reverse Primer	5pmol/μl	0.1pmol/μl	1
Polymerase buffer	10x	1x	5
dNTPs	2mM	0.2mM	5
Water (MQ)	-	-	35.8
DNA polymerase	5 units/μl	1 unit/μl	0.2

DNA polymerases used were either *Taq* DNA polymerase (Roche) or *Pfu* DNA polymerase (Promega). PCR reactions were carried out in either a Hybaid PCR Express or an Applied Biosystems GeneAmp PCR System 2400. The reaction conditions used were as follows:

95 °C for 5 minutes	
95°C for 30seconds X °C for 15seconds 72°C for 1.15minutes	30 cycles
72 °C for 10 minutes	
4 °C hold.	

X= Ta (i.e. Tm-5 °C)

2.3.9 Cross-over PCR

Cross-over PCR is a useful technique for constructing homology arms for pKO integrative vector pTOF24. Often mutations can be introduced into DNA during construction preventing a round of site-directed mutagenesis after cloning.

Briefly, two pairs of primers are designed with annealing sites within and the region flanking the gene of interest (see Figure 2.1). The two sets of primers have the following features: the first forward (F1) and second reverse primer (R2) have restriction site added to 5' end followed by five A nucleotides allowing cloning into the plasmid. The first reverse (R1) and the second forward primer (F2) have 12 bp of homology to one another. The precaution is taken to keep the sequence of gene in frame after the deletion event.

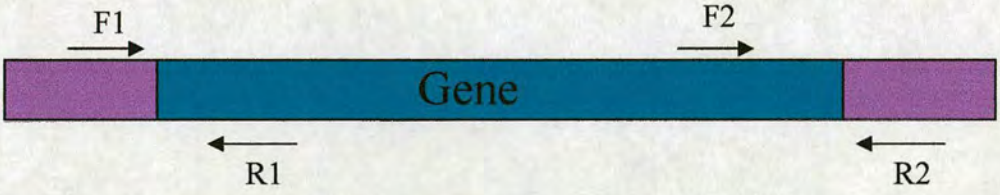


Figure 2.1: The figure shows the positions of primers for cross-over PCR to create gene knock-outs in pTOF24 plasmid

After the first round of PCR, the two PCR products are purified to remove any unused primers and mixed together in a new PCR reaction together with the first forward (F1) and the second reverse (R2) primers. After initial denaturation, the two PCR products anneal at the 12 bp of homology. This fused piece of DNA is then amplified with the two flanking primers.

2.3.10 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.

2.3.11 DNA sequencing

Sequencing reactions were performed using BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems). Plasmid DNA or PCR products to be sequenced were run on an agarose gel and bands intensity were compared with that of DNA marker for quantification. A typical DNA sequencing reaction mix was as follows:

Template DNA	Variable
Terminator Mix	2 µl
Primer	0.5 pmol/µl
MilliQ water	up to 10 µl

Reaction DNA concentration was dependent on size and type of DNA being sequenced – manufacturer's guidelines were followed. The reaction conditions were as following:

96 °C for 1 minute	
96°C for 10seconds 50°C for 5 seconds 60°C for 4 minutes	25 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 Sequence Navigator computer program.

2.3.12 Site directed mutagenesis

The site directed mutagenesis was done by following the instruction of Stratagene's QuickChange® Site-Directed Mutagenesis Kit.

2.4 Repeat Instability Assay

2.4.1 Assay

Every strain to be tested was streaked from glycerol stock onto LB plate to get single colonies. 12 colonies were selected in each assay. Each colony was suspended in 100μl of LB and 50μl of suspension was added to 5ml of LB and LB with 2mM IPTG (to induce transcription). The cultures were grown overnight at 37 °C with shaking. Then they were diluted till 10^{-6} in LB and 100μl was plated onto LB plates. The plates were incubated at 37 °C overnight. For each parental colony, 10 sibling colonies were selected for analysis from IPTG and no IPTG cultures. PCR was carried out to check the length of the repeat tract. The schematic representation of the assay is in the Figure 2.2.

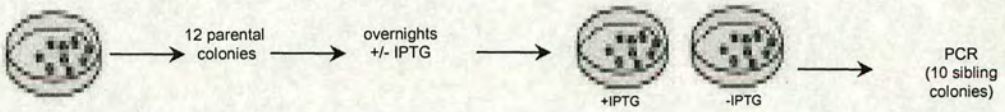


Figure 2.2: Schematic representation of Repeat Instability Assay
12 parental colonies are grown overnight at 37 °C in the presence and absence of IPTG (2mM). The cultures are diluted to 10^{-6} and plated on L-agar plates. PCR is carried out on 10 sibling colonies each from + and – IPTG culture plates.

2.4.2 GeneMapper® analysis of repeats

An ABI 3730 genetic analyser (Applied Biosystems) was used to determine the length of repeat tracts. This automatically detects and determines the sizes of PCR products based on electrophoresis separation (in a capillary). A size standard is run in parallel allowing the sizing of PCR fragments, which is performed by a laser registering when fragments pass a certain point of the capillary.

2.4.2.1 Samples Preparation

For the PCR reaction, the set of primers consisted of one unmodified (Ex-test-R) and one fluorescent primer labelled with 5'FAM (FAM-Ex-test-F). The PCR products were diluted 1:100 in MilliQ water and 1μl of this dilution was mixed with 9μl of Hi-Di™ Formamide containing the GeneScan™-500 LIZ™- Size Standard. 1μl of size standard was added to 1ml of Hi-Di™ Formamide. The bar coded plates were always run either half (48 reactions) or full (96 reactions) as advised by the sequencing service.

2.4.2.2 Data Analysis

The data collected from ABI 3730 Genetic Analyser were analysed using the GeneMapper® software version 4.0. The fragments are visualized as peaks on a graph. The abscissa is a time axis representing when fragments have passed the laser. However, after having run the size standard the time scale is converted to a size scale permitting the sizing of PCR fragments. Characteristic result outputs are shown in Figures 2.3 and 2.4. Here it can be seen that in addition to the main peaks characteristic of the repeat array lengths, several "stutter" peaks are observed. These represent deletions and expansions that have arisen during the PCR reaction and not *in vivo*.

2.5 Western Blotting

Western blotting is the technique used to detect one protein in a mixture of proteins. The basic principle is to transfer the electrophoretically separated proteins on to a solid support and probe with antibody directed against a desired protein.

2.5.1 Preparation of samples

The cultures were grown overnight at 37 °C while agitating. 4 ml of culture was centrifuged and the pellet was dissolved in 150μl of lysis buffer. The samples were kept on ice for 30 minutes and then centrifuged for 20 minutes in cold room. The supernatant was separated into new 2ml tubes and equal volume of SDS (2x)+DTT was added. The samples were boiled for 10 minutes and immediately put on ice.

2.5.2 Gel electrophoresis and transfer

The samples were run on a pre-cast 4-12% NuPAGE Bis-Tris gel (Invitrogen) in 1x MOPS buffer at 150 volts. The gel was transferred onto nitrocellulose membrane using wet-transfer method. Every step was carried out in the transfer buffer. The transfer apparatus was set up in the following manner:

The black side of the cassette (the cathode side) was laid flat on the bench. A thick paper sheet (provided with the apparatus) pre-soaked in transfer buffer was placed on the black side. Then a layer of wet Whatmann paper was made. Then the gel, already immersed in transfer buffer was placed on it and was topped with the nitrocellulose membrane. Then another layer of Whattmann paper was placed and any air bubbles were squeezed out by gently rolling a glass pipette on it. Another thick paper sheet was placed on it. The cassette was closed and transfer was carried out in cold room at 80 volts for 2 hours.

2.5.3 Blocking

The membrane was blocked in 3% milk (made in PBS-Tween) and left overnight in the antibody, anti β-galactosidase rabbit IgG antibody (2mg/ml) in cold room while agitating. Then it was washed in PBS-tween for one hour and second antibody was added with 1% milk and left for one hour. The membrane was again washed for one hour and treated with ECL western blotting substrate (PIERCE). It was then developed on X-ray film.

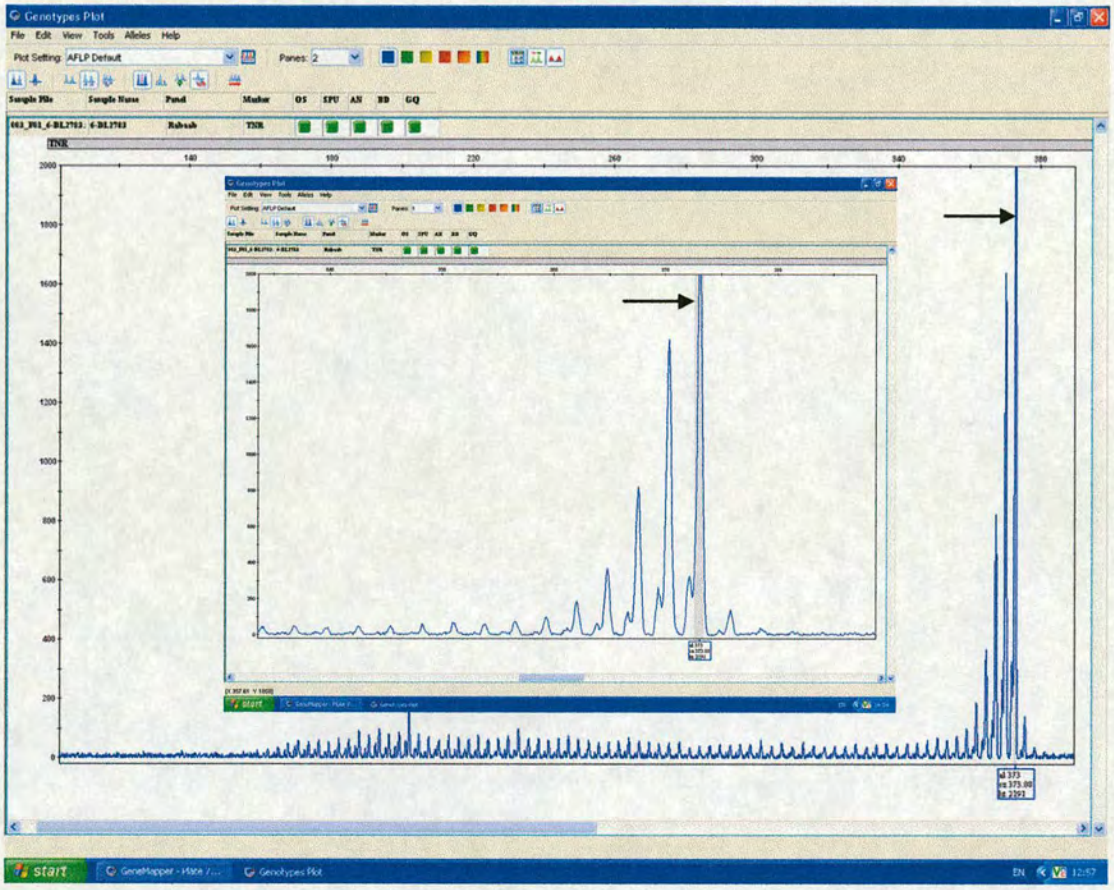
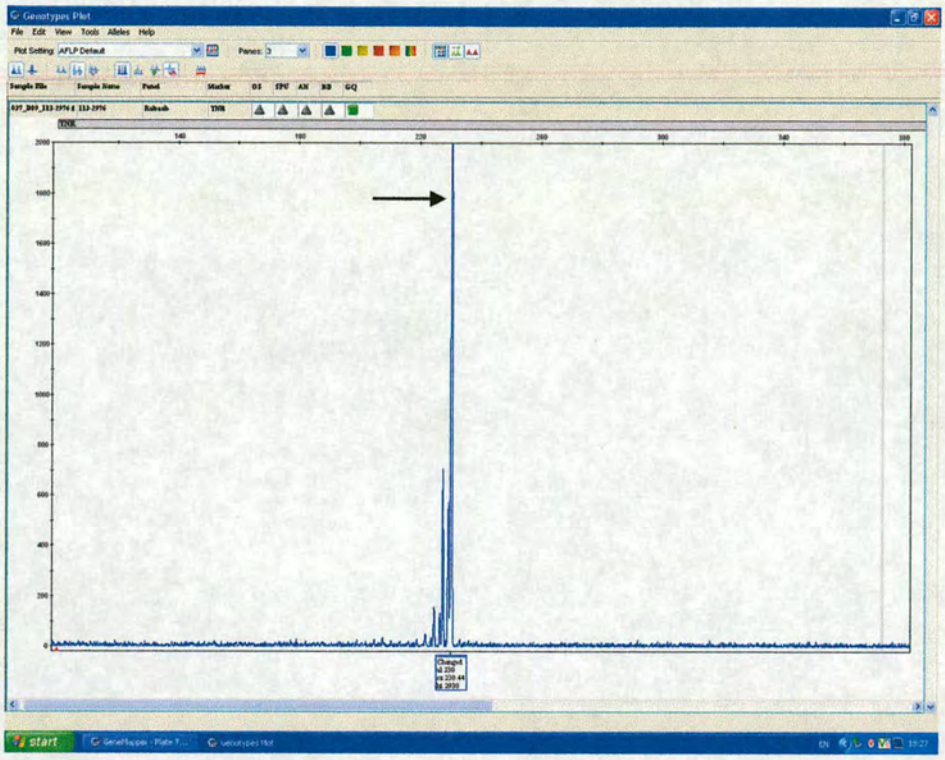


Figure 2.3: Example showing the data output of GeneMapper®
The arrow points to a 373bp peak, which corresponds to the repeat length of 75 (373 bp=3x75 bp+148 bp) as the PCR product size without repeats is 148 bp. The insert (a small window) shows a magnification of the region around the 373 bp peak to illustrate the "stutter" peaks more clearly. Stutter peaks are the by-products of a PCR reaction and generally appear immediately before a real peak.

(A)



(B)

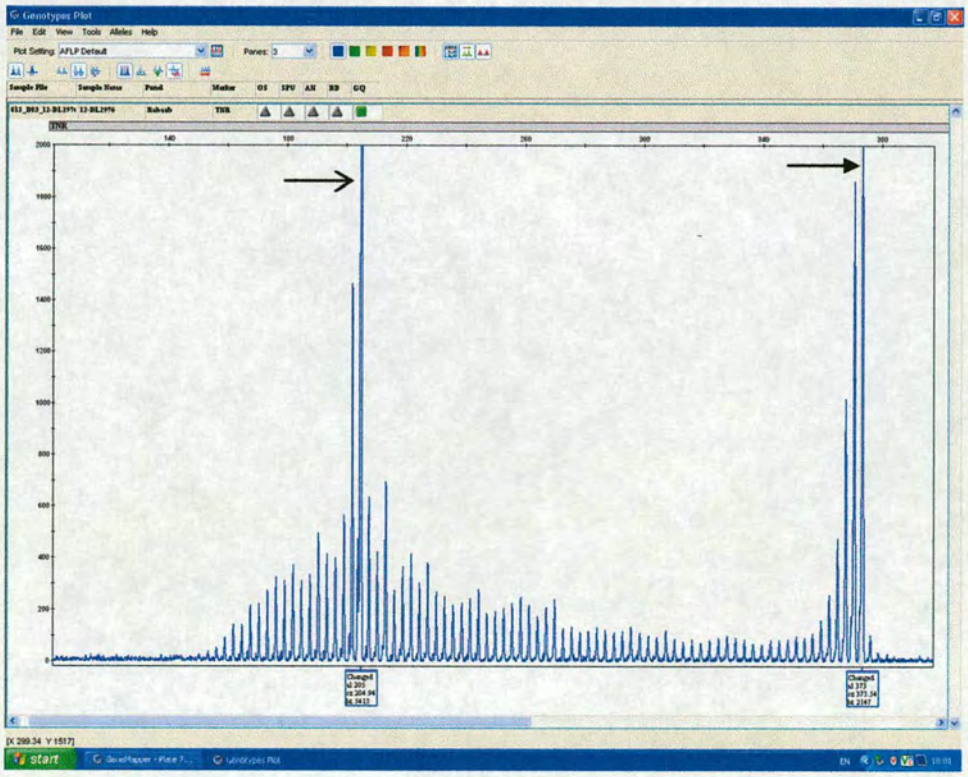


Figure 2.4: Examples showing the data output of GeneMapper®
(A) The arrow points to a peak of 230 bp, which is a deletion of $(CAG)_{75}$ to $(CAG)_{27}$. (B) An example of a mixed colony showing a parental length of $(CAG)_{75}$ with a 373 bp peak (thick arrow) along with a deletion peak of 205 bp (thin arrow) which corresponds to $(CAG)_{19}$.

2.6 Statistical Analysis

Logistic regression models were fitted to the CAG and CTG arrays separately, using Genstat 8th edition, to compare the instability proportions in the different arrays. Approximate 95% confidence intervals were calculated for each estimated instability proportion as the mean $\pm$ 2 $\times$ standard error.

The instability proportion represents the frequency of sibling colonies that had a repeat length changed (expanded or contracted) from the parental length. To avoid counting deletions or expansions that had arisen on the plates, mixed colonies (containing cells with parental and new lengths) were classified as parental in the instability proportion. However, all new lengths were included in the analysis of deletion length distributions. Rare sibling colonies, derived from one parental colony, containing the same length of deletion were only counted once on the assumption that they were sister clones.

CHAPTER 3

CAG·CTG REPEAT INSTABILITY IN TERMS OF LENGTH AND ORIENTATION

3.1 Introduction

This chapter describes the instability (expansions and deletions) of CAG·CTG repeats in the *E. coli* chromosome in terms of repeat length and orientation. In order to study repeats of various lengths, a strategy was developed to generate a library of uninterrupted repeat tracts of different lengths in both CAG-leading and CTG-leading orientations. CAG-leading is the orientation where the CAG repeat tract is on the leading strand template while CTG-leading is the orientation where the CTG repeat tract is on the leading strand template (see Figure 3.1 and table 3.1). For simplicity, CAG-leading and CTG-leading will be described as 'CAG' and 'CTG', respectively, from now onwards.

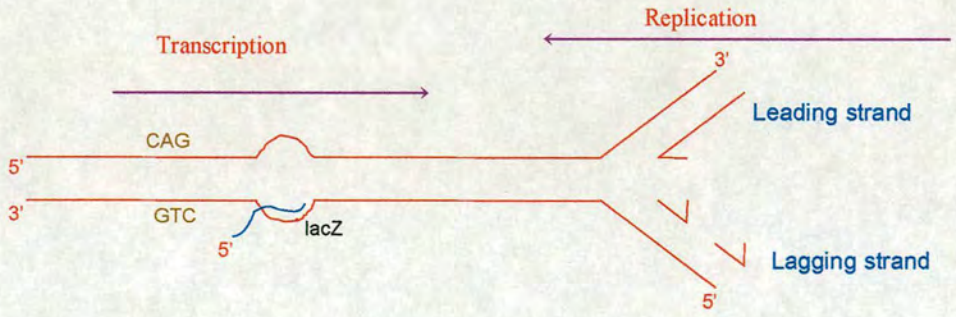


Figure 3.1: Location of repeats integrated in the *E. coli* chromosome.
Repeats were integrated in the 5' end of the *lacZ* gene. The construct shows the CAG-leading (CAG) orientation where CAG repeats are on the leading strand template and CTG repeats on the lagging strand template.

Table 3.1

The CAG and CTG Repeat Orientations

CAG-leading (CAG)	CTG-leading (CTG)
CAG on leading strand template	CTG on leading strand template
CTG on lagging strand template	CAG on lagging strand template
CTG template for transcription	CAG template for transcription

The instability assay described in Chapter 2 was done on CAG repeats of lengths 45, 75 and 84 and CTG repeat lengths of 48, 95 and 140 in wild type *E. coli* cells (MG1655). Results of the instability assay show a dependence on repeat length and orientation with respect to the direction of replication.

3.2 Generation of Long Repeat Arrays

As discussed in Chapter 1, trinucleotide repeat instability associated with human diseases depends on length and purity of repeat tract. Repeats become unstable in humans when they cross a certain size threshold level and repeats with interruptions are more stable than pure tracts. Therefore, in order to study the molecular mechanism of instability, it was important to have uninterrupted repeat tracts of various lengths. Previously, CTG repeat tracts have been produced *in vitro* using thermal cycling protocols (Ordway and Detloff, 1996; Takahashi et al., 1999) and ligation strategies (Grabczyk and Usdin, 1999; Krasilnikova and Mirkin, 2004). Another method developed recently, uses a cloning vector, pACT, to generate repeat fragments containing a G/C overhang that can be ligated efficiently to produce repeats as long as 800 bp (Kim et al., 2005).

3.2.1 Strategy developed in this work

A simple polymerisation independent strategy was developed to generate a set of uninterrupted repeat tracts of various lengths in both CAG and CTG orientations in a plasmid, p*Lac*D2, which was further used to integrate the repeats into the *E. coli* chromosome.

3.2.1.1 p*Lac*D2 Construction

The plasmid p*Lac*D2 was constructed using p*Lac*D1 (constructed by John Blackwood), a derivative of pTOF24 (Merlin et al., 2002) which has a temperature sensitive replication protein encoding gene (*repAts*), a chloramphenicol resistance gene (*cat*) and a *sacB* gene, encoding sensitivity to sucrose. Two *lacZ* homology arms of 400bp, *lac*1 and *lac*2 were introduced into pTOF24 to enable the integration of the plasmid into the *E. coli* chromosome. The two arms were separated by a region of unique restriction sites for *Mfe*I, *Bsa*I and *Bbs*I, resulting in plasmid p*Lac*D1 (see Figure 3.2).

p*Lac*D1 had an extra restriction site for *Bbs*I in one of the *lacZ* homology arms and that was knocked out by changing A to G using primers SDM_BbsI_F and SDM_BbsI_R by site directed mutagensis (SDM) technique, resulting in plasmid p*Lac*D2. Both plasmid maps are shown in Figure 3.2 with the positions of restriction sites.

3.2.1.2 Construction of repeat tracts

$(CAG)_5$ and $(CTG)_5$ repeats were introduced between *Bsa*I and *Bbs*I restriction sites of plasmid p*Lac*D2 by SDM using primer pairs ExCAG-01, ExCTG-01 and ExCAG-02, ExCTG-02 respectively, removing the *Mfe*I site. The partial sequence of the *lacZ* homology arms before and after the introduction of repeats is shown in Figure 3.3.

Another unique restriction site, *Hind*III, in p*Lac*D2 was used to perform the cloning of longer repeats. The p*Lac*D2 plasmid containing the repeats was digested by *Bsa*I and *Hind*III and *Bbs*I and *Hind*III. Each digestion reaction gave two fragments of sizes 2.8 and 3.7 kb. In the case of the *Bsa*I and *Hind*III digest, the 2.8 kb fragment contained the repeats and in the case of *Bbs*I and *Hind*III, repeats were with the 3.7 kb fragment (Figure 3.4). The fragments containing the repeats were extracted from a 1% agarose gel and ligated together to increase the repeat number.

The recognition sites for *Bsa*I and *Bbs*I direct cleavage inside the repeat sequence (Figure 3.5) so in every round of restriction and ligation, there was a doubling of the repeat array length coupled to the loss of two repeats. This method follows the following formula:

$$n_x = 2n_{x-1} - 2$$

n = number of repeat units in the repeat tract

x = round of restriction and ligation

Using this method, repeat arrays of length 5, 8, 14, 26, 50 and 98 were generated for both CAG and CTG sequences. For CAG repeats, another length of 194

was also constructed. For building repeats, the vector was propagated in XL1-blue cells but was also easily propagated in other backgrounds like JM83 and MG1655.

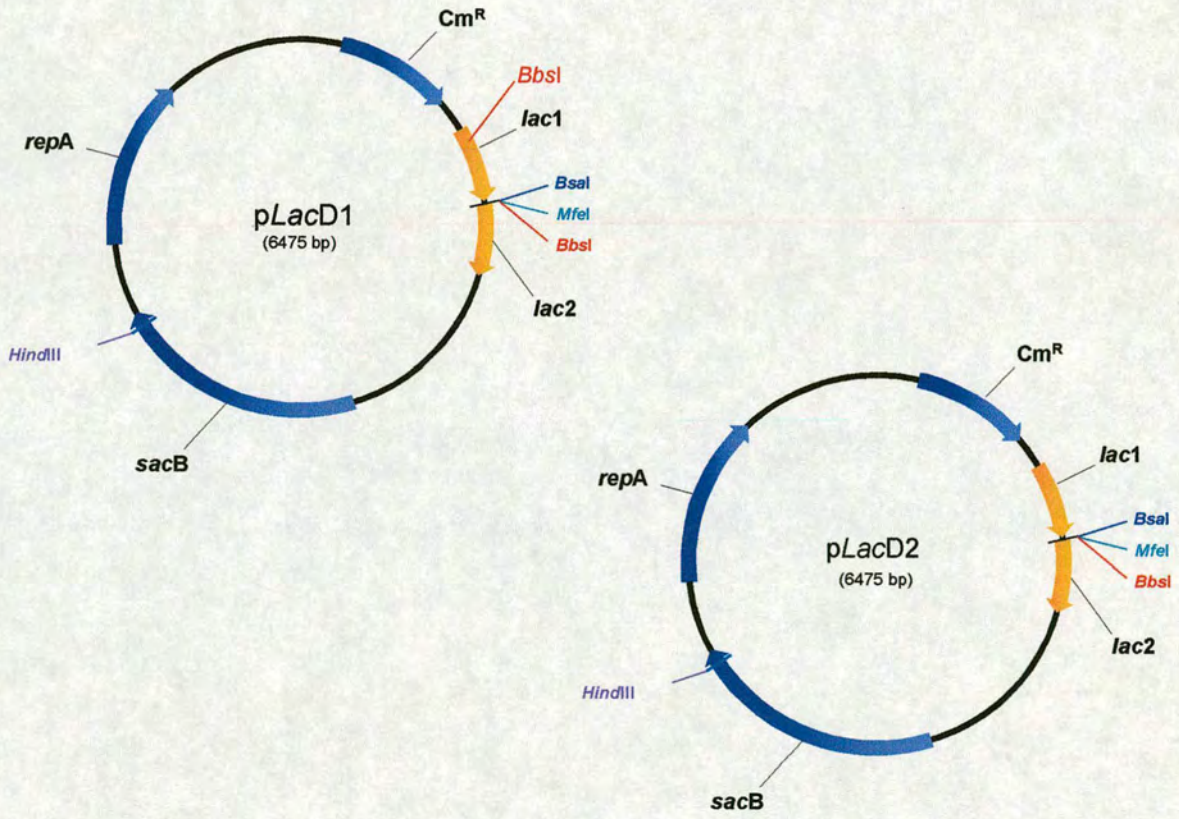


Figure 3.2: Plasmid maps of p*Lac*D1 and p*Lac*D2
The three restriction sites *Bsa*I, *Mfe*I and *Bbs*I are shown along with *Hind*III. Cm^R represents chloramphenicol resistance gene (*cat*), *lac*1 and *lac*2 correspond to two *lacZ* homology arms, *repA* encodes for the temperature sensitive replication protein that is functional at 30 °C but not at 42 °C, *sacB* gene renders cells sensitive to sucrose. The *Bbs*I site in the *lac*1 homology arm of p*Lac*D1 was removed by SDM resulting in p*Lac*D2.

(A) ***lac*1** *BsaI MfeI BbsI* ***lac*2**

CT ATG ACC ATG GTC TCC AAT TGC GTC TTC GTC GTT TTA CAA CGT CGT

(B) ***lac*1** *BsaI* $(CAG)_5$ *BbsI* ***lac*2**

CT ATG ACC ATG GTC TCG CAG CAG CAG CAG CAG GTC TTC GTC GTT TTA C

(C) ***lac*1** *BsaI* $(CTG)_5$ *BbsI* ***lac*2**

CT ATG ACC ATG GTC TCG CTG CTG CTG CTG CTG GTC TTC GTC GTT TTA C

Figure 3.3: (A) Plasmid sequence of *lacZ* homology arms (in green) showing restriction sites of *Bsa*I (in blue), *Mfe*I (in red) and *Bbs*I (in blue). (B) $(CAG)_5$ introduced between *Bsa*I and *Bbs*I by SDM (C) $(CTG)_5$ introduced between *Bsa*I and *Bbs*I by SDM knocking out the *Mfe*I site (in red).

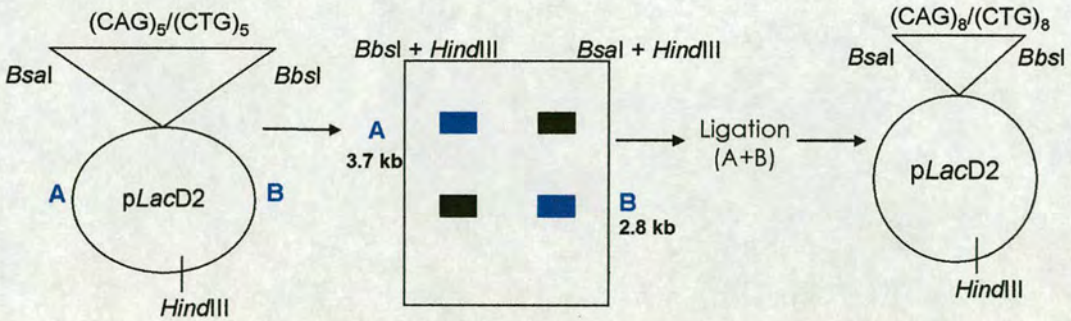


Figure 3.4: Schematic representation of the strategy for building long repeat arrays.
$(CAG)_5$ and $(CTG)_5$ were introduced between *Bsa*I and *Bbs*I restriction sites by site directed mutagenesis. Double digests of plasmid DNA were carried out using *Bbs*I+*Hind*III or *Bsa*I+*Hind*III. Bands were separated on a 1% agarose gel. The fragments containing repeats (A and B) were extracted from gel and ligated to obtain a longer repeat tract length than in the original plasmid

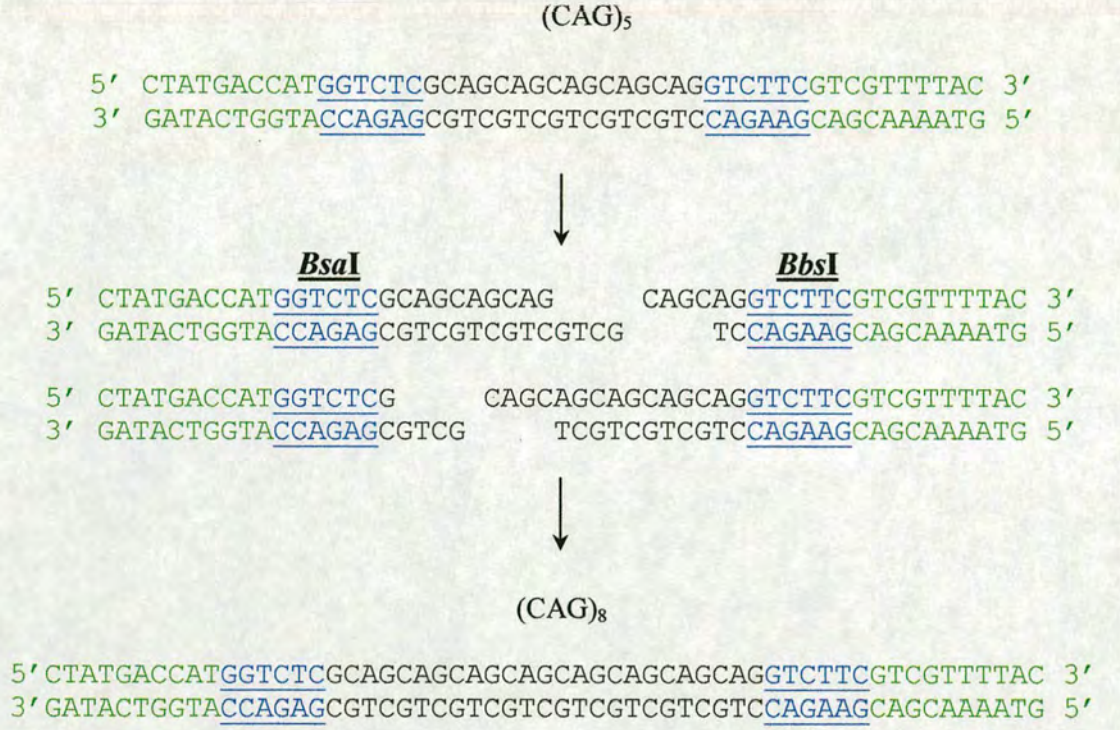


Figure 3.5: Restriction sites for *Bsa*I and *Bbs*I direct cleavage inside the repeat sequence so every new repeat tract length will be twice the original tract length minus 2 repeats. As shown, $(CAG)_5$, after cleavage by *Bsa*I and *Bbs*I and ligation, gives rise to $(CAG)_8$.

3.2.1.3 Integration into the *E. coli* chromosome

Repeats constructed in plasmid p*Lac*D2 were integrated into the *E. coli* chromosome in the 5' end of the *lacZ* gene using the method described in Chapter 2 (Section 2.2.4). Though the repeat lengths generated in plasmid p*Lac*D2 were 50 and 98 in both orientations, the lengths were reduced to 45 and 75 for CAG repeats and to 48 and 95 for CTG repeats during integration into the chromosome. Another attempt of integrating $(CAG)_{98}$ into the chromosome resulted in a length of $(CAG)_{84}$. $(CTG)_{140}$ is the length expanded from $(CTG)_{95}$ in the chromosome.

3.3 Results

3.3.1 Instability increases with increasing repeat length and depends on repeat orientation with respect to direction of replication

Three different lengths of the two orientations of CAG·CTG repeats were studied in wild type cells (MG1655). In both orientations, instability was found to be dependent on the repeat length as the proportion of instability increased with the length of the repeat tract. The instability proportion represents the frequency of sibling colonies that had a repeat length changed (expanded or contracted) from the parental length. To avoid counting deletions or expansions that had arisen on the plates, mixed colonies (containing cells with parental and new lengths) were classified as parental in the instability proportion. However, these new lengths were included in the analysis of deletion length distributions. Rare sibling colonies, derived from one parental colony, containing the same length of deletion were only counted once on the assumption that they were sister clones.

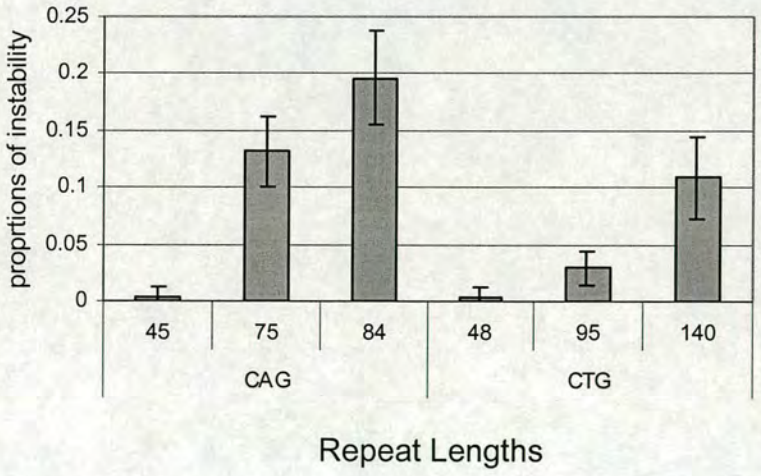


Figure 3.6: Instability proportion of different repeat lengths for CAG and CTG repeat orientations.
The instability proportion is the frequency of sibling colonies that had a repeat length changed from the parental length. Each bar represents the pooled data of two independent assays and corresponds to the individual analysis of 480 clones. The error bars show 95% confidence intervals.

Instability proportions of all repeat lengths studied in wild type cells are shown in Figure 3.6. $(CAG)_{84}$ is 31-fold more unstable than $(CAG)_{45}$. Similarly, $(CTG)_{140}$ is 28-fold more unstable than $(CTG)_{48}$. Notably, the highest repeat length studied in the CAG orientation, $(CAG)_{84}$ is 1.7-fold more unstable than the highest length of the CTG orientation, $(CTG)_{140}$. With respect to DNA replication, the orientation where CAG repeats lie on the leading strand template is more unstable than the opposite orientation, where CTG repeats lie on the leading-strand template.

3.3.2 Distribution of Deletion Sizes

In order to see the sizes of deletions obtained in CAG·CTG repeats in wild type cells, all observed deletions were plotted as a function of percentage of deletion size against the number of times the deletions were observed.

3.3.2.1 CTG Repeat Deletions

Figure 3.7 (panels A and B) shows the distributions of repeat deletion sizes and number of deletion events in the CTG orientation. Deletion distributions have a median (midpoint of the distribution) at 58% for $(CTG)_{95}$ and 61% for $(CTG)_{140}$. $(CTG)_{95}$ has not enough events to document its distribution pattern but $(CTG)_{140}$ shows clearly a negatively skewed distribution (long tail towards left), suggesting a preference for large deletions in this orientation.

3.3.2.2 CAG Repeat Deletions

Figure 3.7 (panels C and D) shows the distributions of sizes of deletions in the CAG orientation. Both lengths 75 and 84 display negatively skewed distributions as seen

from the long tails towards left with predominant larger deletions. The median (midpoint) for $(CAG)_{75}$ comes at 65% and for $(CAG)_{84}$ at 70%.

3.3.3 Distribution of Expansion Sizes

For both CAG and CTG repeat orientations, very few expansions were observed (Figure 3.8). Small expansions of up to 14% of the repeat array were observed for the CAG orientation with the exception of one expansion of 28%. $(CTG)_{95}$ was found to have a rare potential to expand. $(CTG)_{140}$ is an expanded version of $(CTG)_{95}$. Both repeat lengths showed few expansions of up to 10% of the repeat length. $(CTG)_{140}$ gave a higher expansion to $(CTG)_{163}$, which also further expanded to $(CTG)_{184}$. However, these higher lengths could not be analysed because of the limitations of the assay.

3.4 Discussion

The results demonstrate that CAG·CTG repeat instability in the *E. coli* chromosome increases with increasing repeat length and depends on the orientation of repeat tract with respect to the direction of replication. The orientation where CAG repeats lie on the leading strand template is more unstable than the opposite orientation where CTG repeats are on the leading strand template.

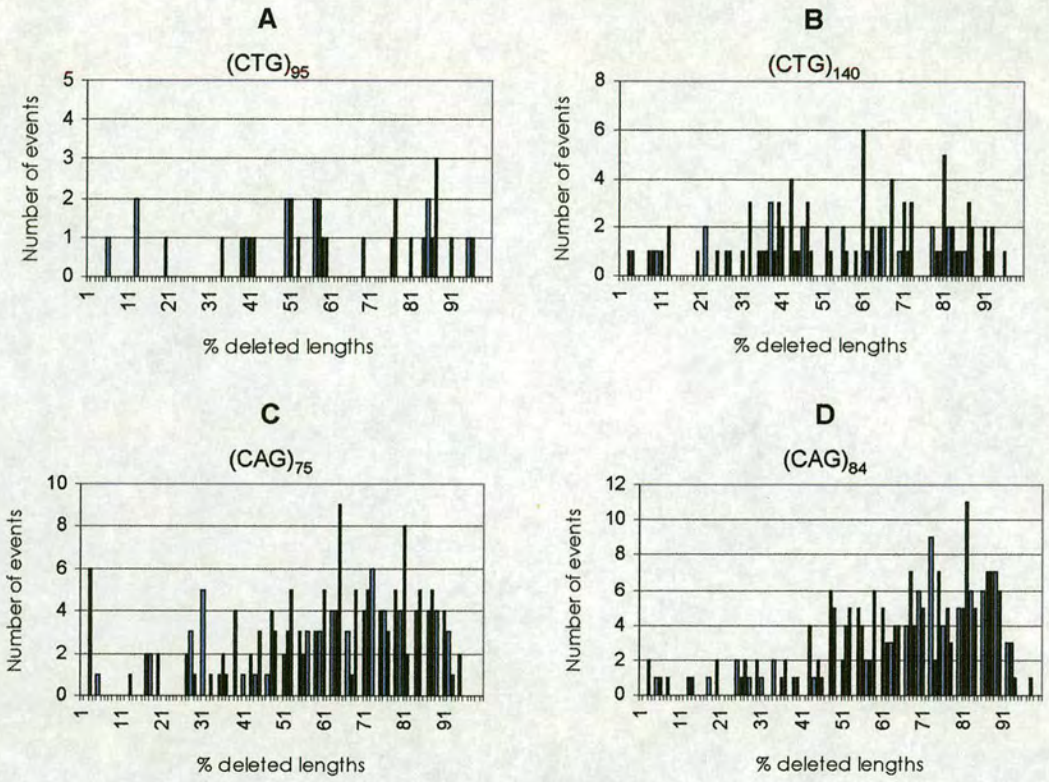


Figure 3.7: Distribution of deletion sizes of CAG·CTG repeats
The deletions observed are plotted as percentage of the tract deleted against the number of times the particular deletions were observed in wild type cells.

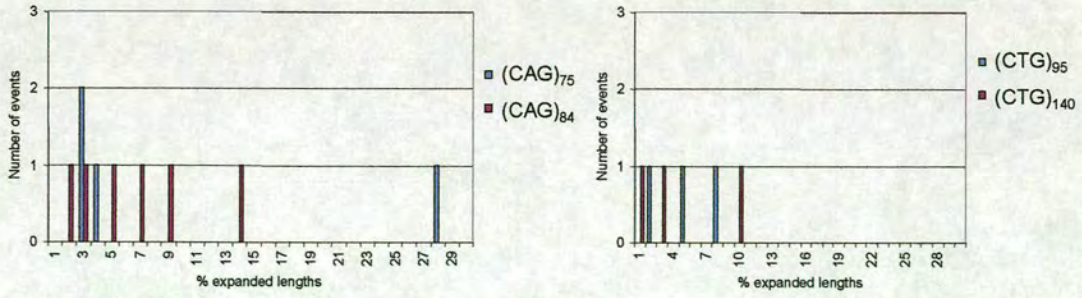


Figure 3.8: Distribution of expansion sizes of CAG·CTG repeats
The expansions observed are plotted as percentage of the tract expanded against the number of times the particular expansions were observed in wild type cells.

3.4.1 Length Dependent Instability

As discussed in Chapter 1, length dependence of repeat array mutation contributes to the phenomenon of dynamic mutation where expanded repeat arrays have an increased probability of further expansion leading to anticipation in the inheritance of disease phenotypes in humans. Previous studies in *E. coli* using plasmid substrates, and in yeast have demonstrated that longer lengths of trinucleotide repeats are more unstable than smaller lengths (Freudenreich et al., 1998; Kang et al., 1995a; Richard et al., 2000; Rolfsmeier et al., 2001; Sarkar et al., 1998; Shimizu et al., 1996). The results in this chromosomal study also demonstrate the same observations that CAG·CTG repeat instability is a length dependent phenomenon. Instability increases with increasing repeat length in both orientations. *In vitro*, long repeat tracts form hairpins with long lifetimes than shorter tracts (Gacy and McMurray, 1998). It has been shown that DNA polymerase pauses at specific loci in CTG and CGG sequences *in vitro*, depending upon the length of the repeat tract and temperature of the reaction (Kang et al., 1995b). The strength of the pausing increased as the CTG repeat length increased and pausings were stronger at lower temperature. As the longer repeats are more prone to form stable hairpins, the instability has also been suggested to be caused by reiterative synthesis on the secondary structures (Lenzmeier and Freudenreich, 2003).

Although the lengths studied in this work are in the human disease causing range, they do not show the bias towards expansions seen in humans. Deletions account for most of the instability observed. The deletion distributions suggest a preference for large deletions indicating the formation of secondary structures encompassing large areas of repeat tract.

3.4.2 Orientation Dependent Instability

Studies in *E. coli* and yeast have suggested the importance of the direction of replication through CAG·CTG tracts (Freudenreich et al., 1997; Kang et al., 1995a; Maurer et al., 1996; Miret et al., 1998) and have proposed a replication slippage model (discussed in Chapter 1) to explain this orientation dependent instability. The model uses lagging strand replication dynamics along with the differential stabilities of hairpin structures (Figure 1.2). According to this model, deletions are prominent when the more stable secondary structure forming CTG repeats are on the lagging strand template, while expansions are prominent when the stable structure forming repeats are on the Okazaki fragment. However, in both model systems, contractions have been found more frequently than expansions.

In this study of CAG·CTG repeat instability in the *E. coli* chromosome, orientation dependence was observed with mostly deletions in both orientations. Furthermore, CAG orientation was found to give bigger expansions than CTG, so the traditional model is not the best explanation of this orientation dependent instability. This study suggests another explanation of orientation dependence of TNRs instability, which will be revealed in the next Chapter.

CHAPTER 4

CAG·CTG REPEAT INSTABILITY: A CONSEQUENCE OF PROCESSING OF INTERMEDIATES DURING DNA REPLICATION

4.1 Introduction

This chapter describes the study carried out to address the link between CAG·CTG repeat instability and processing of secondary structures formed by repeats during cellular processes, particularly during DNA replication. As discussed in Chapter 1, when single stranded, repeats can fold into hairpin structures of different stability depending upon the sequence and length. These structures can be a target of nucleases such as the SbcCD complex and other repair pathways proteins. Additionally, the proofreading activity of the DNA polymerase III, performed by the DnaQ protein, can also influence the processing of these structures or intermediates.

The instability assay described in Chapter 2 was performed on CAG and CTG repeats in an *sbcCD* and *dnaQ* mutants. CAG repeat instability was decreased in an *sbcCD* mutant strain and both repeat orientations were destabilized in a *dnaQ* mutant. Based on these findings, a new model is proposed to explain the molecular basis of orientation dependent repeat instability.

4.1.1 SbcCD Complex

The *sbcC* and *sbcD* genes of *E. coli* encode a nuclease complex called SbcCD. Based on its molecular organization, the SbcCD complex has been predicted to belong to the SMC (structural maintenance of chromosome) family (Sharples and Leach, 1995). Furthermore, it has been shown by electron microscopy that this complex forms a head-rod-tail structure, characteristic of an SMC protein (Connelly et al., 1998). SbcC belongs to a subclass of SMC proteins that also includes human and *Saccharomyces cerevisiae* Rad50 proteins (Sharples and Leach, 1995). Both human Rad50 (hRad50) and yeast Rad50 (scRad50) are involved in the double strand break repair pathway and interact with their respective Mre11 proteins (Johzuka and Ogawa, 1995). Mre11 shares sequence similarity with SbcD and both proteins belong to a family of phosphoesterases (Sharples and Leach, 1995).

The 1.2 MDa SbcCD complex functions as an ATP-dependent dsDNA exonuclease and ATP-independent ssDNA endonuclease and acts on a variety of substrates including DNA hairpins (Connelly et al., 1999; Connelly et al., 1997; Connelly and Leach, 1996).

4.1.2 SbcCD and Repeat Instability

Trinucleotide repeats, when single-stranded, can fold into hairpins structures *in vitro* (Gacy et al., 1995), which can be attacked by the SbcCD complex (Connelly, de Leau et al. 1999). The observations that CAG and CTG repeats form hairpins (Gacy et al., 1995; Mitas et al., 1995b; Petruska et al., 1996; Smith et al., 1995; Yu et al., 1995a) and slipped mis-pairing structures *in vitro* (Pearson et al., 1998; Sinden et al., 2002) led to the hypothesis that misfolded structures play a role in TNR instability. In *E.*

coli and yeast, several observations have implicated the role of SbcCD and Rad50-Mre11 complexes in repeat instability. The first evidence of the involvement of SbcCD in repeat instability in *E. coli* came from the work of Sarkar et al. (1998) where using a plasmid based system, the bimodal pattern of CTG repeat amplification was reported. It describes incremental expansions for repeat length less than the size of an Okazaki fragment ($(CTG)_{100\text{-}200}$) and saltatory expansions for repeat length of more than 300 repeats when the experiment was carried out in the multiply mutant strain (SURE) at 25 °C in the absence of an active *sbcC* gene product (Sarkar et al., 1998). In yeast, a reduction in CAG repeat expansions associated with double-strand break repair was suppressed by over-expressing the Rad50-Mre11 complex, suggesting that it may cleave hairpin structures (Richard et al., 2000). Furthermore, CTG repeat-induced spontaneous double-strand breaks were reduced in a *rad50* mutant (Freudenreich et al., 1998).

4.1.3 DnaQ and Repeat Instability

In *E. coli*, the fidelity of DNA replication is 10^{-10} errors per replicated base (Drake, 1991; Schaaper, 1993) and the contribution of the proofreading process to the overall fidelity is 10^{-2} (Schaaper, 1993). The proofreading function during replication is performed by the 3' to 5' exonucleolytic ε-subunit of the DNA polymerase III, which is encoded by the *dnaQ* gene. DnaQ protein functions to prevent slipped-strand pairing events during replication that can lead to instability of repeated sequences.

Mutations in the proofreading function of DNA Polymerase III (*dnaQ49*^{ts} and the *mutD5*) have been shown to be involved in the instability of TNRs and other tandemly repeated sequences in *E. coli* (Bzymek et al., 1999; Iyer et al., 2000; Saveson and Lovett, 1997).

4.2 Results

4.2.1 CAG repeat instability is reduced in an *sbcCD* mutant

The effect of *sbcCD* mutation was investigated for both CAG and CTG repeat orientations. Figure 4.1 shows the proportions of instability of CAG·CTG repeats as a function of the absence of SbcCD and DnaQ. The $(CAG)_{75}$ repeat array was significantly stabilised 1.8-fold in an *sbcCD* mutant. The CTG orientation was too stable at the studied length, $(CTG)_{95}$, to obtain sufficient data permitting a statistical distinction between the wild type and the *sbcCD* mutant. The results suggest that SbcCD is involved in the instability pathway in the CAG orientation but do not provide evidence for destabilisation of CTG orientation.

4.2.2 CAG·CTG repeat arrays are destabilised in a *dnaQ* mutant and orientation dependence is lost

In a PolIII proofreading (*dnaQ*) mutant, the instability proportion increased for both repeat orientations, with a more pronounced effect in the CTG orientation (Figure 4.1). The instability proportion for $(CTG)_{75}$ was increased 8-fold in a *dnaQ* mutant compared with that of $(CTG)_{95}$ in wild type. It should be noted here that the CTG repeat length in *dnaQ* mutant was smaller than the wild type because the repeats were deleted down to 75 during the construction of the mutant strain. In the other orientation, $(CAG)_{75}$ had an instability proportion only 1.6-fold higher than $(CAG)_{75}$ in wild type ($p=0.008$). It is notable that the level of instability is the same in both orientations in a *dnaQ* mutant (Figure 4.1), which indicates that orientation

dependence could be a direct consequence of different efficiencies of proofreading of CAG and CTG templates, with the CTG orientation being proofread more efficiently.

4.2.3 A mutation in *sbcCD* in the *dnaQ* mutant reduces CTG repeat instability

As mentioned in section 4.2.1, the CTG orientation was too stable at the length studied to obtain sufficient data permitting a statistical distinction between wild type and *sbcCD* mutant (Figure 4.1). However, an effect of the *sbcCD* mutation could be measured in an *sbcCD dnaQ* double mutant, where a significant decrease in the instability proportion was observed compared to that in *dnaQ*, even though the repeat length was shorter in the *dnaQ* mutant (75 instead of 95) ($p=0.004$). In contrast to the *dnaQ* single mutant, orientation dependence of CAG·CTG repeats was retained in *sbcCD dnaQ* double mutant ($p<0.001$) (Figure 4.1). This suggests that intermediates in the CTG orientation pathway can escape deletion in the absence of SbcCD and proofread more easily than intermediates in the CAG orientation pathway.

4.2.4 Large deletions predominate over small deletions in *sbcCD* and *sbcCD dnaQ* mutants in the CAG orientation and in a *dnaQ* mutant in both repeat orientations

In order to investigate the sizes of deletions obtained in CAG·CTG repeats in *sbcCD*, *dnaQ* and *sbcCD dnaQ* mutant cells, all observed deletions were plotted as a function of percentage of deletion size against the number of events (Figure 4.2). All CAG mutants in this study, show repeat deletion distributions like wild type, negatively skewed as seen from the long tails towards left. This shows that the distributions of deletion lengths in CAG repeat tracts are not influenced substantially

by mutations in *dnaQ* or *sbcCD* suggesting that these genes do not affect the nature of the primary intermediate in the pathway but the frequency of its processing to a product with a new repeat length. In the CTG repeat orientation, the *dnaQ* single mutant gives a negatively skewed distribution while no conclusion could be made about the distribution in the *sbcCD* mutant due to insufficient number of deletion events observed. But a comparison between the *dnaQ* and *dnaQ sbcCD* mutants in this orientation brings out an interesting conclusion. The negatively skewed distribution observed in the *dnaQ* single mutant (median 63%) disappears in the *dnaQ sbcCD* double mutant giving a flat distributon with a median of 51% (Figure 4.2). This suggests that, contrary to the CAG orientation, the presence of SbcCD (in the absence of proofreading) favours the formation of large deletions in the CTG orientation.

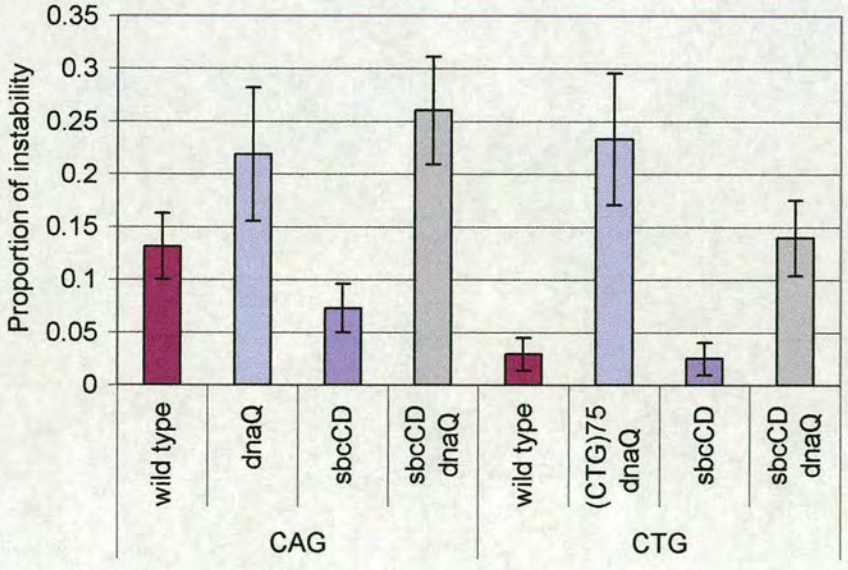


Figure 4.1: Instability proportions for *sbcCD*, *dnaQ* and *sbcCD dnaQ* mutants of CAG·CTG repeats

CAG repeat length studied in wild type and all mutants was 75 and CTG repeat length was 95 except in the *dnaQ* mutant where it was 75. Each bar represents the proportion of instability (pooled data of two independent assays – 480 clones). The error bars show 95% confidence intervals.

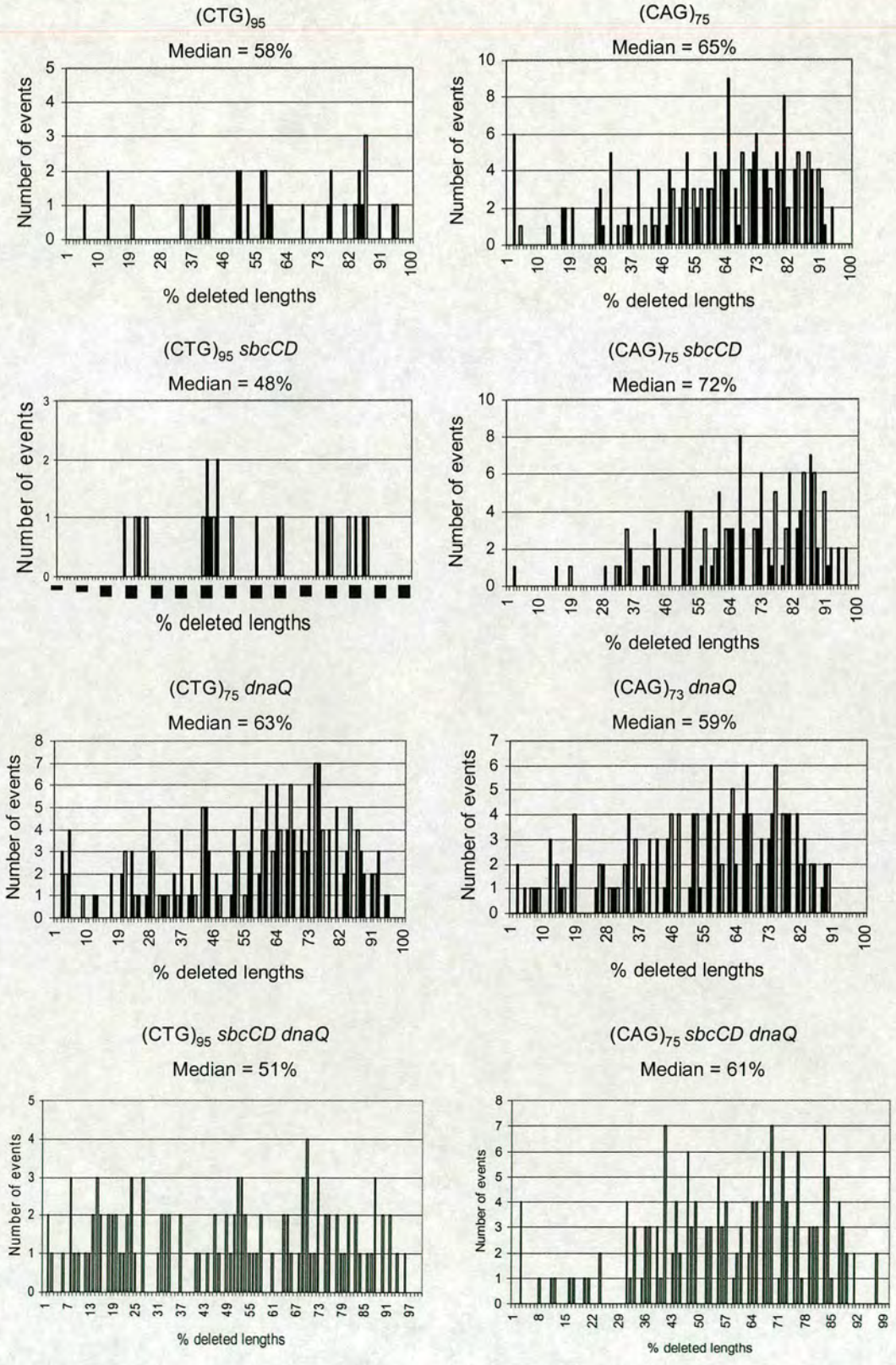


Figure 4.2: Distribution of deletion sizes in wild type cells and in *sbcCD*, *dnaQ* and *sbcCD dnaQ* mutants containing CAG CTG repeats

The observed deletions are plotted as percentage of the tract deleted against the number of times the particular deletions were observed.

4.3 Discussion

4.3.1 The proofreading subunit of DNA polymerase III (DnaQ) determines orientation dependence of replicative instability in cells with active SbcCD nuclease

This work demonstrates that a mutation in *dnaQ* destabilizes CAG·CTG repeat arrays and that orientation dependence of instability in wild type cells is lost in this mutant. This is a specific effect of *dnaQ* mutation since a mutation in *mutS* does not have a corresponding effect (see Chapter 6) suggesting that the instability observed in the *dnaQ* mutant is likely to be attributable to its proofreading defect and not to a non-specific effect of elevated mutagenesis. Previous studies in the literature suggested that the orientation dependence of CAG·CTG repeat instability is caused by the dynamics of lagging-strand DNA synthesis accompanied by the greater thermodynamic stability of CTG repeat hairpins compared to CAG repeat hairpins (Freudenreich et al., 1997; Maurer et al., 1996; Miret et al., 1998). However, it has not previously been shown how this process is mediated. The data in this study suggest that proofreading is inefficient on the CTG repeat template of the lagging strand (CAG orientation) leading to the orientation dependent instability. It is the more foldable CTG repeat strand that is more refractory to proofreading.

In *E. coli*, two mutations in DnaQ (*dnaQ49*^{ts} and *mutD5)* have been shown to destabilize and stabilize CAG·CTG repeats in plasmids (Iyer et al., 2000) but in yeast, it was demonstrated that proofreading mutants of DNA polymerases delta and epsilon do not destabilize these repeats (Schweitzer and Livingston, 1999). The yeast results

are interesting given that homo and dinucleotide repeats are destabilized in these mutants (Strand et al., 1993; Tran et al., 1997), suggesting some particular resistance to proofreading of CAG·CTG triplet repeats by polymerases delta and epsilon. Interestingly, DnaQ has been shown to share sequence homology with human DNA editing enzyme DNase III (Hoss et al., 1999). This enzyme is present in equal amounts in non-dividing and proliferating cells, which suggests that it is involved in repair processes as well as replication. It is therefore plausible that efficient proofreading during replication and repair in human cells may contribute to repeat stability.

4.3.2 The SbcCD nuclease increases CAG repeat instability when proofreading is active and CTG repeat instability when proofreading is inactive

This work demonstrates that CAG·CTG repeat instability is reduced in an *sbcCD* mutant when the CTG containing strand is the template for the lagging-strand of the replication fork (CAG orientation). This stabilizing effect of an *sbcCD* mutation is lost in a *dnaQ sbcCD* double mutant. This result argues for an antagonistic action of SbcCD and the proofreading subunit of DNA polymerase III. It is consistent with the existence of an SbcCD-dependent pathway of deletion formation for the CAG orientation that is only significant in the presence of proofreading. This may be because the SbcCD-independent pathway of deletion is fully functional in the *dnaQ* mutant and swamps the SbcCD-dependent pathway.

In the CTG orientation (CAG on the lagging-strand template), SbcCD plays an active role in stimulating deletions in the absence of proofreading. This is evidenced

by the small but significant decrease in instability in an *sbcCD dnaQ* strain compared to *dnaQ* mutant and the shift from a skewed distribution of deletion sizes in *dnaQ* to a flat distribution in *sbcCD dnaQ* for the CTG orientation. Furthermore, the observation of orientation dependence in *sbcCD dnaQ* but not in *dnaQ* implies a role of SbcCD in removing the orientation dependence in the absence of proofreading. These observations are consistent with SbcCD having access to an intermediate in the CTG deletion pathway (CAG on the lagging-strand template) in absence of DnaQ and stimulating its conversion to a deletion product rather than its return to a parental template.

It has recently been shown that fluorescent tagged fusions of SbcC protein of *Bacillus subtilis* are localized with a similar pattern to that of the replication factory, consistent with action of SbcCD at the site of DNA replication (Meile et al., 2006). A similar co-localisation of fluorescent tagged SbcC with a replication factory protein has also been observed in *E. coli* (Darmon and Leach, personal communication)

4.3.3 A model for orientation dependent replicative instability of CAG·CTG repeats

Combining the results of *sbcCD* and *dnaQ* mutations, we propose a model to explain the orientation dependence of CAG·CTG repeat instability during replication (Figure 4.3). Orientation dependence is determined by proofreading of slippage intermediates formed during replication of the lagging-strand. We suggest that the CTG repeat template for the lagging-strand (CAG orientation) is partially refractory to proofreading because of the formation of stable structures, which can cause more slippage (Figure 4.3 (3 and 4)) leading to elevated frequencies of deletions in wild type cells. Intermediates in the slippage reaction in this orientation of the repeat array are more accessible to the SbcCD nuclease, which can increase instability by

digesting the strands that signal the presence of a substrate for proofreading. In the absence of proofreading, it would appear that SbcCD can no longer affect instability possibly because its main effect is to divert intermediates from effective proofreading or because cleavage is not possible in a proofreading mutant. The latter hypothesis is made less likely with the observation of an effect of SbcCD on instability in a *dnaQ* mutant when CAG repeats are the template for the lagging-strand (CTG orientation). In this orientation, we hypothesise that more unstable intermediates are formed and the effect of SbcCD is to divert them from a proofreading-independent pathway of return to parental length (Figure 4.3 (f)).

4.3.4 Hairpin formation *in vivo* and SbcCD

From these data, it is not possible to be certain whether hairpin structures have an effect on instability. However, several observations suggest them to be playing a role *in vivo*. First, in most situations, the distribution of sizes of deletion products is negatively skewed consistent with a preference for large deletions comprising enough repeats to form stable hairpins. Second, an effect of SbcCD is observed on instability in the orientation predicted to form the more stable CTG repeat hairpins (CAG orientation) consistent with the role of SbcCD being a hairpin nuclease that has been shown to cleave CTG repeat hairpins (Connelly et al., 1999). However, the effect of sbcCD is also observed in the opposite orientation in the absence of proofreading. Further, the role of SbcCD cannot be restricted to cleavage of hairpin structures. It has been shown that the *sbcCD* genotype can affect deletions between 101-bp direct repeats in a situation where no obvious hairpin is present which suggests that the SbcCD complex has a wider specificity of action (Bzymek and Lovett, 2001).

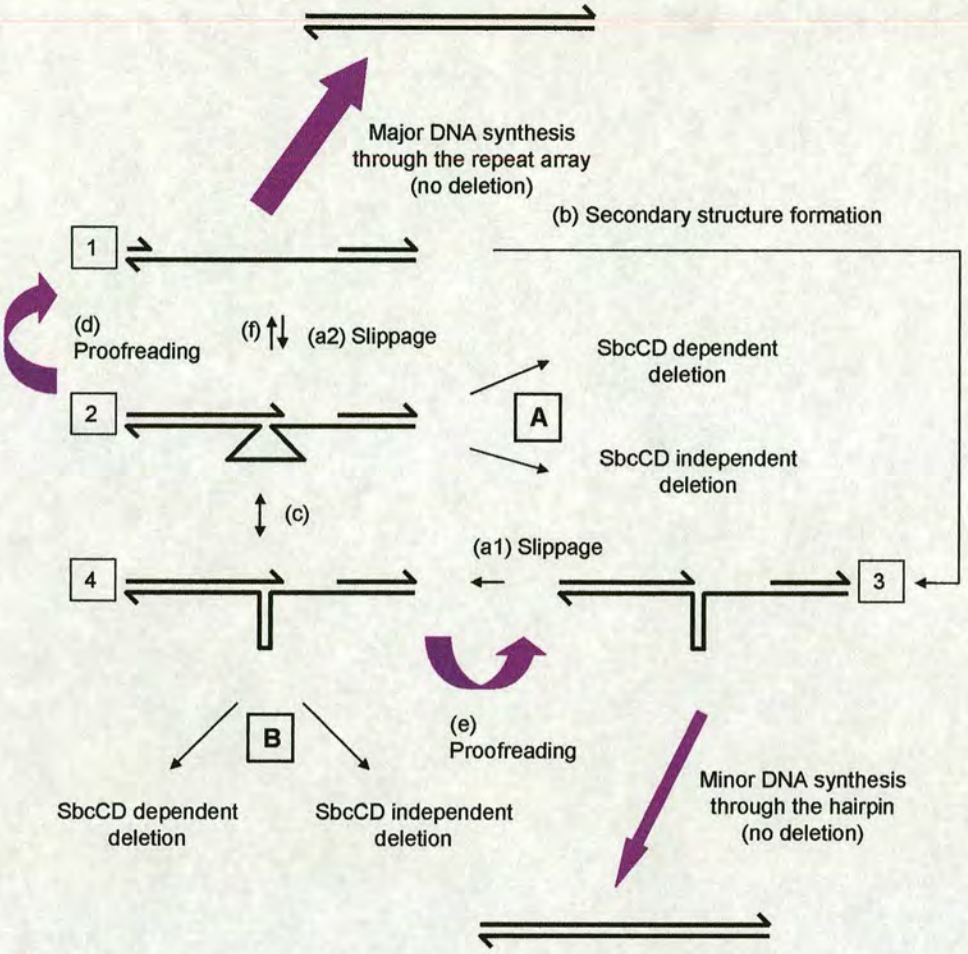


Figure 4.3: Model for orientation dependent replicative instability
DNA synthesis arrests within a CAG or CTG repeat array (1) and strand-slippage occurs either accompanied (a1) or not accompanied (a2) with stable secondary structure formation (b). A lagging strand template containing CAG repeats will form less stable secondary structures (2) and so primarily will adopt pathway A whereas a template with CTG repeats will form more stable secondary structures (3 and 4) and so primarily will adopt pathway B. Some interconversion between well-folded and less well-folded structures may also be possible (interchange c). The observation of the loss of orientation dependent repeat instability in a *dnaQ* mutant implies that the initiation of slippage is independent of whether the template contains CAG or CTG repeats. This has the further implication that initiation of slippage occurs independently of the potential of the strand to form stable secondary structures. The requirement for the presence of SbcCD for loss of orientation dependence in a *dnaQ* mutant implies that SbcCD ensures the efficient processing of slippage intermediates initiated on both CAG and CTG repeats. Orientation dependence is generated by poor proofreading of the CTG template strand. In this model, 3' to 5' exonucleolytic proofreading does not remove a secondary structure if it is present (compare e with d) so new synthesis has the potential to slip again on such a template (3 to 4). In the absence of proofreading, intermediates with stable structures are committed to deletion (whether or not SbcCD is present) while intermediates with less stable secondary structures can escape through an inefficient disassembly pathway (f) that is only significant in the absence of SbcCD and proofreading.

CHAPTER 5

CAG·CTG REPEAT INSTABILITY AND RECOMBINATION

5.1 Introduction

This study was carried out to investigate the role of DNA recombination in CAG·CTG repeat instability. Repeats can form secondary structures that can be cleaved by SbcCD, which might result in double strand breaks (DSBs) implicating the repair by RecA-mediated recombination. Further, replication paused at trinucleotide repeats might require recombination for its resumption, which might also lead to trinucleotide repeat instability. Previously, no chromosomal study has been carried out in *E. coli* to investigate the role of recombination in repeat instability. Since the system developed in this work allowed the analysis of repeat instability in the *E.coli* chromosome so it was of interest to test recombination mutants.

The instability of both repeat orientations, CAG and CTG on the leading strand templates, was analysed in *recA, recB*, *recF* and *xerC* mutant strains of *E. coli*. $(CTG)_{95}$ repeat instability was also assessed in *recG* and *ruvAC* mutants along with *dif* and *dif recA* mutants.

5.1.1 Homologous Recombination (HR) in *E. coli*

Homologous recombination is the process by which DNA recombines with a homologous partner. This mechanism not only protects the cell from double and single strand breaks but also participates in the rescue of blocked replication forks (Michel, 2000).

There are several proteins that can be involved in a recombination reaction depending upon the nature of the DNA substrates involved in this reaction. In general terms, HR involves three steps: initiation (pre-synapsis), homologous pairing and strand exchange (synapsis) and branch migration and resolution (post-synapsis).

Repair of DSBs depends on RecBCD (exonuclease V) whose function is to provide a 3' overhang for RecA-mediated D-loop formation (Wang and Smith, 1989). Upon contacting a DSB, RecBCD simultaneously unwinds and degrades the 3'-strand until it reaches a *chi* sequence (5'-GCTGGTGG-3') (Anderson and Kowalczykowski, 1997) which modifies its activity. The nuclease activity of the enzyme is altered so that degradation of the 3'-strand is down-regulated and the nuclease activity of 5'-strand is up-regulated (Dixon and Kowalczykowski, 1993). This produces a 3'-ssDNA where single strand binding (SSB) protein is removed and the strand exchange protein RecA is loaded (Anderson and Kowalczykowski, 1997), which forms a nucleoprotein filament with single stranded DNA that extends 5' to 3' (Register and Griffith, 1985). On finding a homologous sequence, the RecA-ssDNA nucleofilament invades the duplex DNA and extends into it.

HR can also repair single stranded gaps which is mediated via the RecFOR pathway (Kowalczykowski, 2000). The RecFOR protein complex removes SSB and loads on RecA, promoting strand exchange. This results in setting up of a Holliday

junction, which is driven by RecG and RuvAB to the point where RuvC resolves it. Depending on the strands cleaved by RuvC, either crossover or non-crossover products are generated.

5.1.2 RecG – a Branch-Specific Helicase

RecG is a branch-specific 3'→ 5' helicase, which functions in branch migration and in the resolution of Holliday junctions. It was demonstrated that RecG could resolve Holliday junctions without cleavage by forcing branch migration in the reverse direction to Rec-A driven strand exchange (Whitby et al., 1993). It has also been proposed to promote RFR at replication forks blocked by UV lesions (McGlynn and Lloyd, 2000) but this reaction has been regarded as controversial in later studies (Donaldson et al., 2004; Wang, 2005).

5.1.3 Xer site-specific Recombination

Xer site-specific recombination functions in the *Escherichia coli* chromosome segregation and cell division apparently by resolving chromosome dimers, which arise through homologous recombination, to monomers. It was initially identified through its role in the resolution of ColE1 plasmid multimers (Summers and Sherratt, 1984). Two site-specific recombinases, XerC and XerD bind cooperatively to the recombination site *dif* (deletion induced filamentation) in the chromosome replication terminus region and catalyse separate pairs of strand exchanges (Blakely et al., 1991; Kuempel et al., 1991).

5.2 Results

5.2.1 CAG·CTG repeat instability is unaffected in *recA*, *recB* and *recF* mutants

Mutations in recombination genes, *recA*, *recB* and *recF* did not affect significantly the proportion of instabilities of $(CAG)_{75}$ and $(CTG)_{95}$ repeat tracts (Figure 5.1). This suggests that recombination does not contribute to instability of CAG·CTG repeats in the *E. coli* chromosome at the lengths studied.

5.2.2 Mutations in *recG* and *ruvAC* do not affect CTG repeat instability

Figure 5.2 shows the proportions of instability observed for *recG* and *ruvAC* mutant strains containing $(CTG)_{95}$ repeats on the leading strand template. The mutation in *recG* did not affect CTG repeat instability ($p=0.108$) and the same was observed for *ruvAC* mutation ($p=0.239$) when instability was compared with wild type levels. This suggests that these genes are not involved in the pathway(s) of instability of CTG repeats in the *E. coli* chromosome.

5.2.3 CAG·CTG repeat instability is unaffected in *xerC* mutants and mutations in *dif* and *dif recA* do not affect CTG repeat instability

As shown in Figure 5.3, a mutation in *xerC* did not affect the proportion of instability in both repeat orientations, implicating no influence of chromosome dimer resolution in repeat instability.

Further the effect of *dif* mutation was investigated on CTG repeat instability. Figure 5.4 shows the proportions of instability observed for *dif* and *dif recA* mutants

containing CTG repeats which were not significantly different from that in the wild type strain. The repeat length in the *dif* mutant was deleted down to 87 during construction of the mutant strain and the *recA* mutation was introduced in the same strain. Therefore, both *dif* and *dif recA* mutants had a repeat length smaller than the wild type. The results imply that the genes for chromosome dimer resolution do not influence CTG repeat instability in the *E. coli* chromosome.

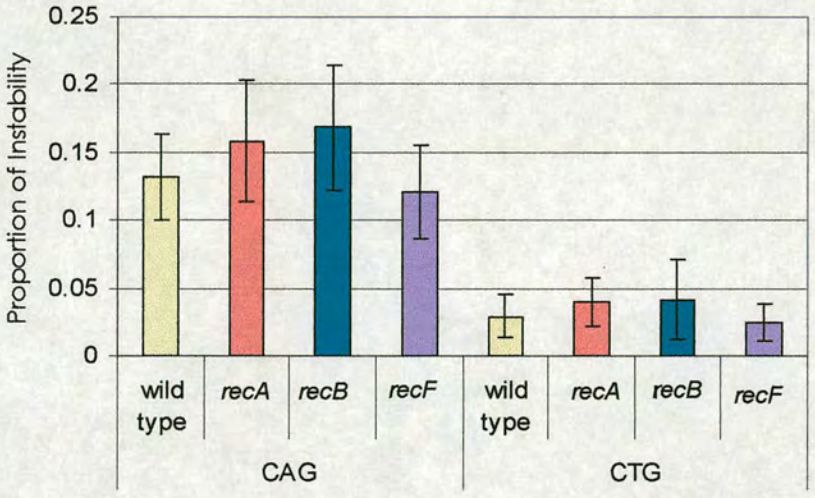


Figure 5.1: Instability proportions of *recA*, *recB* and *recF* mutants containing CAG·CTG repeats

The length studied for CTG repeats is 95 in wild type and mutants. For CAG repeats, it is 75 for all mutants except in *recF*, where it is 80. Each bar represents the instability proportion calculated from the data of two independent experiments (480 clones). Error bars represent 95% confidence intervals.

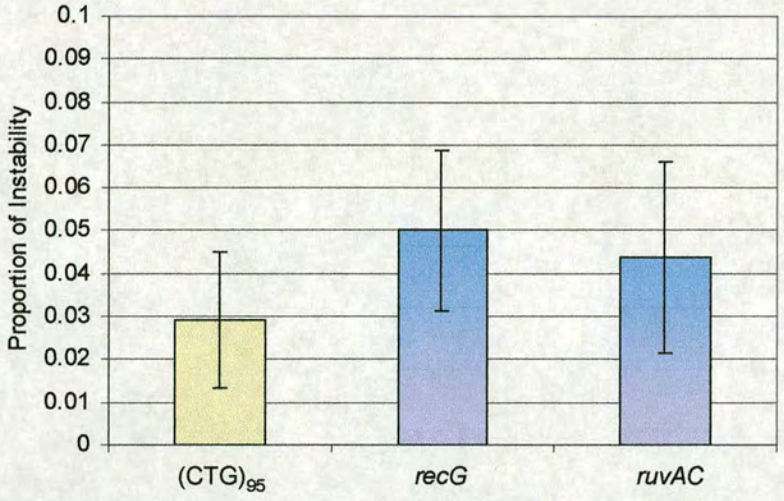


Figure 5.2: Instability proportions for *recG* and *ruvAC* mutants containing $(CTG)_{95}$ repeats on the leading strand template
The bars represent the pooled data of two independent experiments (480 clones). The error bars represent 95% confidence intervals.

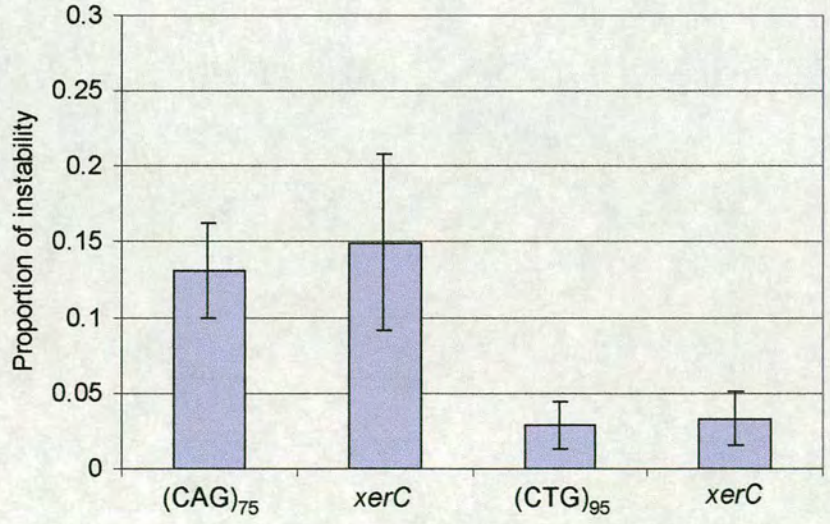


Figure 5.3: Instability proportions for *xerC* mutants containing CAG.CTG repeats
The bars represent the pooled data of two independent experiments. The error bars represent 95% confidence intervals.

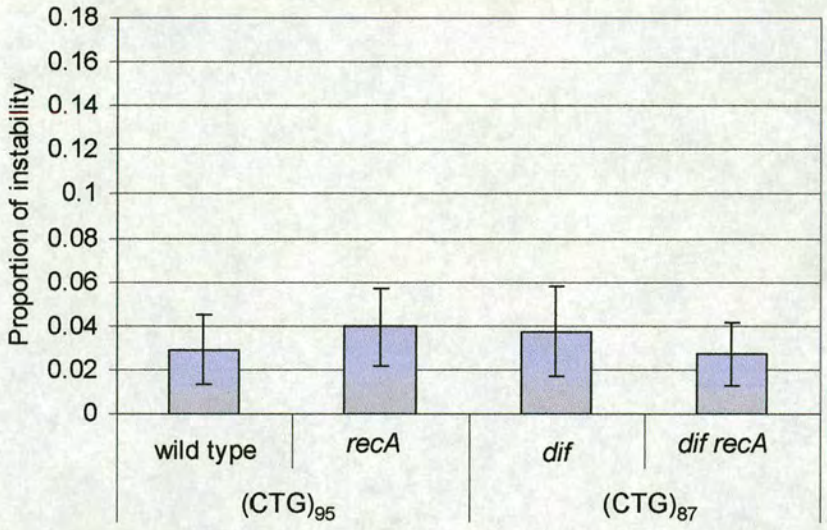


Figure 5.4: Instability proportions for *dif* and *dif recA* mutants containing CTG repeats
The wild type and *recA* mutants had CTG repeats of length 95 while the length was deleted down to 87 in *dif* mutant, in which *recA* mutation was introduced. The bars represent the data from two independent experiments (480 clones). The error bars represent 95% confidence intervals.

5.2.4 Distribution of Deletion Sizes in recombination mutants

In order to analyse the sizes of deletions obtained in CAG·CTG repeats in recombination deficient mutant cells, all observed deletions were plotted as a function of percentage of deletion size against the number of times the deletions were observed.

5.2.4.1 CAG repeat deletions

CAG repeat deletions distributions in *recA*, *recB* and *recF* mutants are shown in Figure 5.5. All mutants show the negatively skewed distributions similar to that seen in wild type cells. This shows that large deletions are preferred in these mutants and the nature of intermediates involved in instability is the same as in wild type cells.

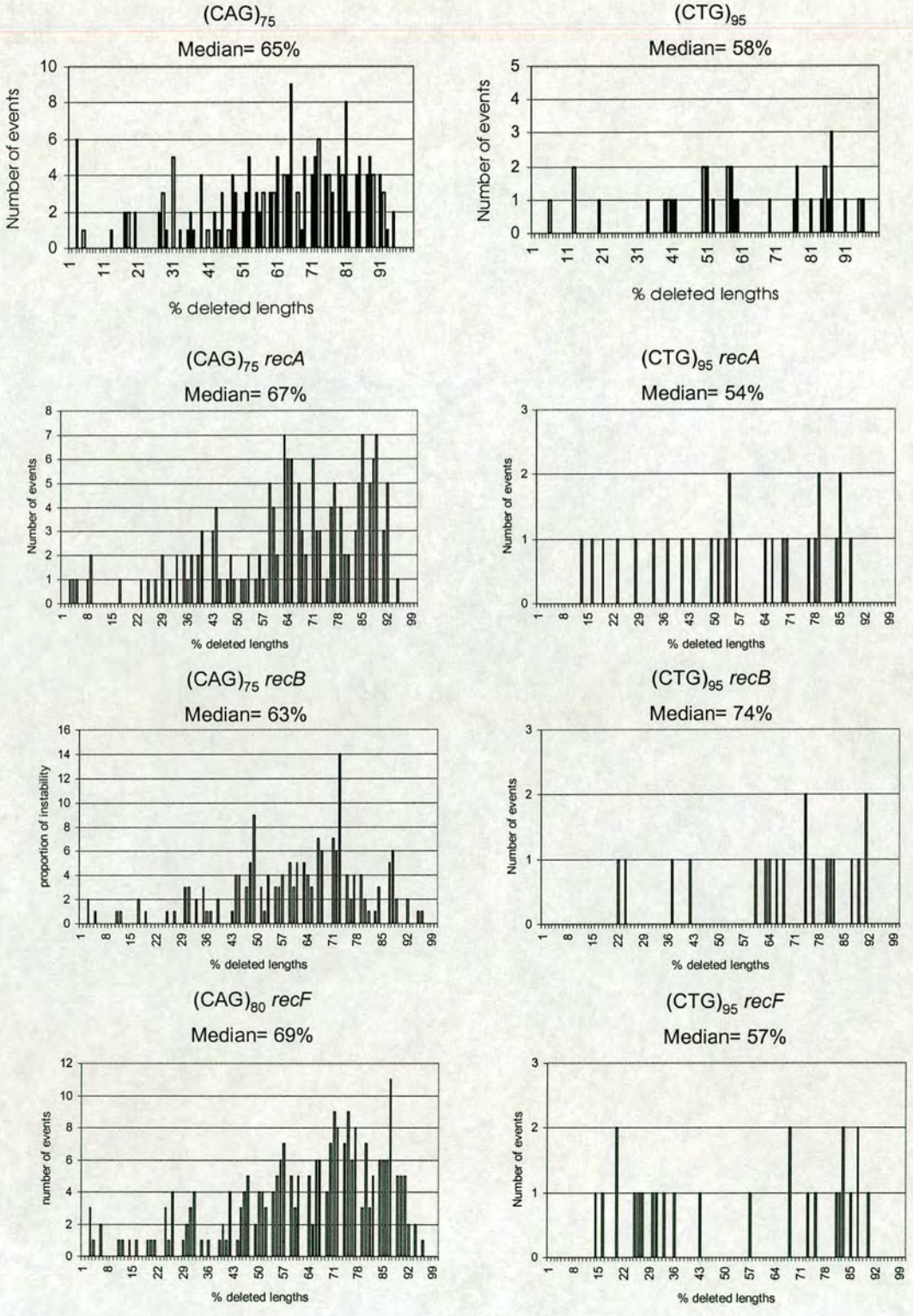


Figure 5.5: Distributions of deletion sizes in *recA*, *recB* and *recF* mutants containing CAG·CTG repeats

The deletions observed are plotted as percentage of the tract deleted against the number of times the particular deletions were observed.

5.2.4.2 CTG repeat deletions

Figure 5.5 shows the sizes of deletions observed for the recombination mutants, *recA*, *recB* and *recF* containing CTG repeats. In all three mutants, the observed numbers of deletions are not sufficient to make any significant conclusion about the distribution pattern but the patterns display negatively skewed distributions like wild type cells.

The pattern of deletion distributions in a *recG* mutant is positively skewed with a median of 41% (Figure 5.6) as compared to negatively skewed distribution observed in wild type and most of the mutants studied in this work. A flat distribution is observed in a *ruvAC* mutant with a median of 50%, though the numbers of observed deletions are not sufficient to reach a significant conclusion. These observations suggest the possibility that the mutations in *recG* and *ruvAC* influence the nature of primary intermediates in the instability pathway(s).

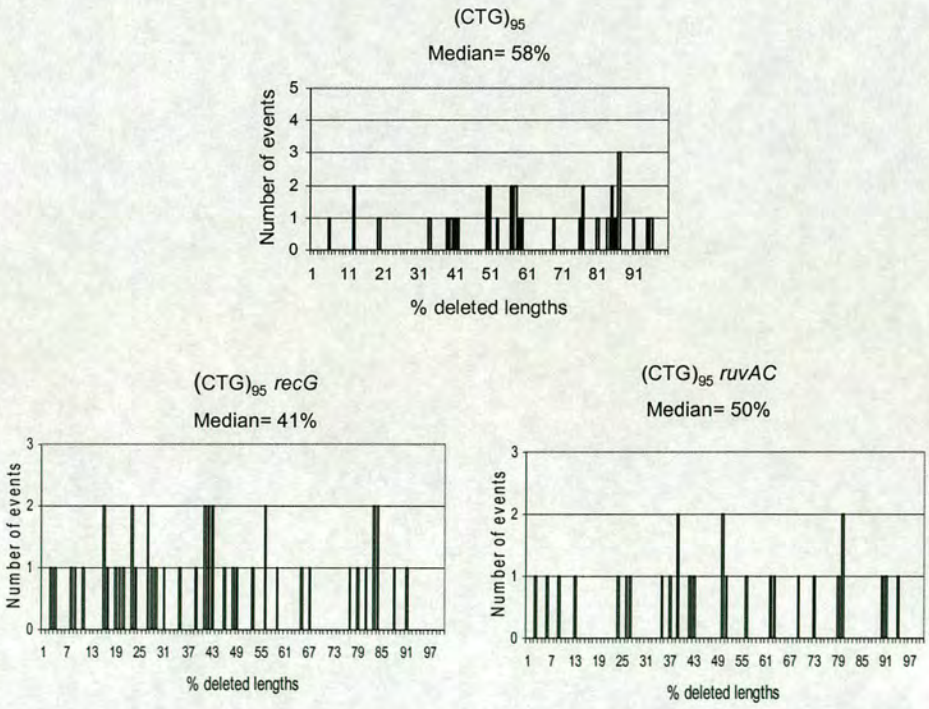


Figure 5.6: Distributions of deletion sizes in *recG* and *ruvAC* mutants containing CTG repeats
The deletions observed are plotted as percentage of the tract deleted against the number of times the particular deletions were observed.

5.3 Discussion

In the previous Chapters, it was demonstrated that in wild type cells, the orientation where CAG repeats lie on the leading strand template, is more unstable than when CTG repeats are on the leading strand template (Chapter 3). From further investigations, it was observed that an *sbcCD* mutation stabilizes repeats in the CAG orientation (Chapter 4), which suggested the formation of secondary structures that are cleaved by SbcCD. So, this study was carried out to test the hypothesis that if the cleavage of secondary structures by SbcCD leads to double strand breaks, their repair by *recA*-mediated homologous recombination might influence instability but no significant effect of recombination genes was observed.

5.3.1 CAG·CTG Repeat Instability is unaffected in *recA*, *recB* and *recF* mutants of *E. coli*

The lack of effect of *recA*, *recB* and *recF* mutations on instability of both repeat orientations suggests that instability of CAG·CTG repeats of the lengths studied does not involve homologous recombination. Furthermore, these mutations do not affect the distributions of deletion sizes.

Previously, numerous studies have reported roles of recombination in repeat instability in *E. coli* using plasmid substrates (Jakupciak and Wells, 2000a; Jakupciak and Wells, 2000b; Napierala et al., 2002; Pluciennik et al., 2002). However, this has faced some ambiguities since both stabilizing and destabilizing effects have been reported. In favour of its stabilizing role, CAG repeats were shown to have increased contractions in a *recA* strain (Sopher et al., 2000) and similarly, RecA and RecB

dependent DSB repair, induced in a CAG sequence, was shown to protect the sequence from deletions (Hebert et al., 2004). In other reports, recombination proficiency was correlated with a high rate of instability. Using a plasmid system, a dramatic decrease in instability was observed in *recA* and *recB* deficient cells (Hashem et al., 2004). Moreover, other studies showed that recombination destabilizes CAG·CTG repeats by promoting both expansions and deletions (Jakupciak and Wells, 1999; Jakupciak and Wells, 2000b; Napierala et al., 2002; Pluciennik et al., 2002).

The work carried out in this study differs from previous works done by using plasmid systems. The plasmid-based systems have utilized two approaches to study recombination. The first is the use of a single multicopy plasmid carrying repeat sequences propagated in recombination deficient strains (Hashem et al., 2004; Sopher et al., 2000). The second strategy is to study recombination between two different compatible plasmids (Jakupciak and Wells, 2000a; Jakupciak and Wells, 2000b; Napierala et al., 2002; Pluciennik et al., 2002). A problem associated with these systems is the accurate measuring of the recombination frequencies, which are determined by the appearance of cells demonstrating a recombinant phenotype. To express the recombinant phenotype, a cell requires formation of a recombinant plasmid and its manifestation among non-recombinant parental plasmids. Such competition between the parental and recombinant plasmids has been documented to significantly bias the measured frequency of recombination (Bierne et al., 1995).

Furthermore, in yeast and mouse, no change in repeat instability was observed when the recombination genes *rad52*, *rad54* and non-homologous end joining DNA-*pkcs* were knocked out (Miret et al., 1998; Savouret et al., 2003)

The lack of effect of the *recF* mutation on instability in both orientations suggests that the gap repair is not involved in the instability mechanism. However, RecF was found to have effect on CTG repeat deletions in a *uvrD* mutant (see Chapter 8), which suggests that RecF plays a role in instability but in the absence of UvrD helicase. The implications of this observation are discussed in Chapter 8.

5.3.2 CTG repeat instability is not affected by mutations in *recG* and *ruvAC*

The proportion of CTG repeat instability was unaffected by mutations in *recG* and *ruvAC* genes which suggests that these genes do not play a major role in CTG repeat instability. It was recently shown that mutations in *recG* and *ruvAC* decreased the deletion rates of CAG·CTG repeats in plasmids (Kim et al., 2006b) implicating the roles of these genes in a replication restart pathway influencing instability. The same group showed that a *recG* mutation affects CAG·CTG repeat instability in the *E. coli* chromosome in a length dependent manner (Kim et al., 2006a). They demonstrated a decrease in deletion rate in *recG* mutants for $(CTG)_{43}$ and no change at $(CTG)_{60}$ while in CAG orientation, *recG* mutation did not affect $(CAG)_{43}$ but increased deletion rate of $(CAG)_{52}$. The authors suggest the involvement of RecG in replication restart of the fork blocked by structure on the leading strand. This is confusing because if this were the case, we would expect them to have detected higher instability in *recG* mutants for longer CTG repeats considering the differential stabilities of hairpins in a length dependent manner formed by CAG and CTG repeats (discussed in Chapter 1).

The lack of effect of these mutations on instability proportions of CTG repeats in this work suggests that the events involved in chromosome are different from those in plasmids. Though the mutations do not affect the proportions of instability, *recG*

does appear to affect the distribution of deletions, which suggests that it may play a role in determining the primary intermediate in the instability pathway(s). This is interesting since a similar deletion pattern was observed in a *rep* helicase mutant (Chapter 8), which also possesses 3'-5' helicase activity like RecG. This may define a new class of instability events being generated on the leading strand. RecG might be acting on replication fork like Rep helicase, unwinding the nascent lagging strand. The dynamics of fork in the absence of such helicase activity are discussed in detail in Chapter 8.

5.3.3 A mutation in *xerC* does not affect CAG·CTG repeat instability and CTG repeat instability is unaffected in *dif* and *dif recA* mutants

The gene involved in chromosome dimer resolution, *xerC* was found to have no effect on CAG·CTG repeat instability. This suggests that the mechanism of instability does not generate crossover products. Furthermore, *dif* and *dif recA* mutants showed no difference in instability of CTG repeats, which is consistent with no role of recombination process in instability. These results are interesting in the context of the observation of crossing-over associated with tandem repeat instability. Using a 787 bp tandem repeat, it was observed that *recA*-independent deletions were often accompanied by plasmid dimerization, suggesting a crossing-over mechanism (Lovett et al., 1993). The model proposed by authors suggests that *recA*-independent crossovers between replicating sister strands can be associated with deletion or amplification of repeated sequence. An absence of effect of dimer resolution genes in this study suggests that deletion pathway(s) of CAG·CTG repeat instability do not involve generation of crossover products and these events might be happening with long tandem repeat sequences.

CHAPTER 6

CAG·CTG REPEAT INSTABILITY AND DNA REPAIR

6.1 Introduction

DNA in all organisms is repeatedly submitted to chemical and physical assaults. The cell can face extracellular threats from the environment, such as radiation and mutagens, or intracellular threats, such as hydrolysis and oxidation. Unfaithful DNA replication can also be a threat, if mistakes created during replication are left unattended and passed to the next generation. Therefore, cells have developed a variety of DNA repair mechanisms to maintain genome integrity. In *E. coli*, the fidelity of DNA replication is 10^{-10} errors per replicated base (Drake, 1991; Schaaper, 1993). This high fidelity is achieved in three steps, faithful synthesis in terms of base selection by DNA polymerase, 3'-5' exonuclease proofreading activity and post-replicative mismatch repair (MMR).

Small misalignments of 3 bp occasionally generated during replication of trinucleotide repeats can be recognized and corrected by the MMR system, preventing a single triplet insertion or deletion. Studies in *E. coli* regarding the role of MMR in repeat instability have come up with conflicting findings, suggesting both a stabilizing and destabilizing effect of MMR on plasmids-born trinucleotide repeats (Jaworski et al., 1995; Schmidt et al., 2000; Schumacher et al., 1998). This study was carried out to analyse the role of this faithful repair system on CAG·CTG repeat instability in the *E. coli* chromosome. Three main genes of MMR, *mutS*, *mutH* and *mutL* were tested in both orientations of the repeat array.

TNRs can form hairpin structures, which contain a mismatch, every two base pairs. In mice, MMR has been documented to be a contributing factor in expansions. It has been suggested that the Msh2/Msh3 complex (MutS homologue) binds and stabilizes hairpin structures (Kovtun and McMurray, 2001; Sinden, 2001). As hairpins might be a substrate for the structure dependent nuclease SbcCD (Connelly et al., 1999; Connelly and Leach, 1996), this work was further expanded to test the effect of SbcCD in MMR mutants to investigate whether a competition exists between SbcCD and MMR proteins for action on the hairpins.

Whereas MMR recognizes mismatches as long as 3 bp, the nucleotide excision repair system (NER) recognizes bulky DNA adducts at damaged sites and can be involved in large loop processing. The deletion sizes observed in a wild type background suggested the formation of large loops as intermediates in CAG·CTG instability (Chapter 3). Therefore, the role of NER proteins UvrA, UvrB and UvrC in CTG repeat instability was analysed.

6.1.1 Mismatch Repair (MMR) in *E. coli*

In *E. coli*, MMR is the system that repairs mismatches in DNA after replication. The proteins involved in MMR are MutS, MutH, MutL, UvrD, SSB, ExoI, RecJ, ExoVII, DNA polymerase III and DNA ligase.

MutS, encoded by the *mutS* gene, recognizes and binds up to four mismatched base pairs. Then MutS goes through a conformational change and translocates the DNA flanking the mismatch, forming an alpha shaped heteroduplex loop (Allen et al., 1997). MutL binds to the MutS-heteroduplex complex (Grilley et al., 1989). Then MutH, a methyl-directed and sequence specific endonuclease, is activated upon encountering the MutS-MutL complex at a hemimethylated 5'-GATC-3' site and

incises the unmethylated DNA 5' to the G of the GATC site either 3' or 5' of the mismatch (Modrich, 1991). Next, DNA helicase II, encoded by the *uvrD* gene, unwinds the strands so that the error-containing strand can be degraded exonucleolytically.

RecJ, ExoI and ExoVII, three single-stranded exonucleases in *E. coli*, have been shown to act during MMR, depending on the polarity of the exonucleolytic degradation. Exo VII or RecJ carries out the degradation if the nick is 5' of mismatch, and Exo I if the nick was made 3' of the mismatch. Finally, polymerase III holoenzyme fills in the gap and ligase repairs the nick.

6.1.2 MMR and Repeat Instability

The literature, including studies in model systems ranging from bacteria to human cell culture, reveals conflicting findings regarding the effect of MMR on repeat instability.

In *E. coli*, the frequency of long deletions of $(CTG)_{180}$ repeats was reduced in *mutS*, *mutL* and *mutH* strains, when CTG repeats were on the leading strand template (Jaworski et al., 1995). This suggested that MMR functions to increase large deletions of long repeated sequences. Another report showed that a mutation in *mutS* led to instability of $(CTG)_{64}$, independent of the orientation of the repeats (Schumacher et al., 1998). A study carried out in this lab using $(CTG)_{43}$ confirmed these opposing findings arguing that MMR can have both stabilizing and destabilizing effect on repeat instability (Schmidt et al., 2000). The model proposed by Schmidt et al., 2000 (shown in Figure 6.1) suggested that 3 bp misalignments occur on the lagging strand template and MMR can avoid single triplet expansion or contraction by repairing these 3 bp misalignments. In the presence of functional MMR, the single strand

generated during repair may fold up into a hairpin and lead to contraction during repair DNA synthesis. In another study, it was also suggested that MMR promoted large deletions when a CTG sequence of more than 100 repeats was located on the lagging strand template and cells were propagated for a large number of generations (Parniewski et al., 2000).

Studies in yeast have demonstrated that MMR does not affect the frequency of large contractions (Miret et al., 1997; Schweitzer and Livingston, 1997) or large expansions (Miret et al., 1998; Rolfsmeier et al., 2000; Schweitzer and Livingston, 1997). It was also shown to block the expansions of interrupted repeat tracts (Rolfsmeier et al., 2000).

In the mouse, MMR has been shown to contribute to TNR expansions. Using transgenic *msh2*$^{-/-}$ knockout mice, it was shown that both somatic and germline repeat expansions depend on Msh2 protein (Kovtun and McMurray, 2001; Manley et al., 1999). The model for Msh2 dependent instability argues that the protein binds to and stabilizes hairpins, which allows the hairpin to persist and expand during subsequent DNA synthesis (Figure 6.2). The binding effect of Msh2 leading to expansions appears to be independent of the MMR pathway, as Msh2 seems to bind and stabilize hairpin during gap repair of single and double strand breaks (Kovtun and McMurray, 2001). Different approaches suggested in literature for the involvement of MMR in repeat instability are shown in Figure 6.2.

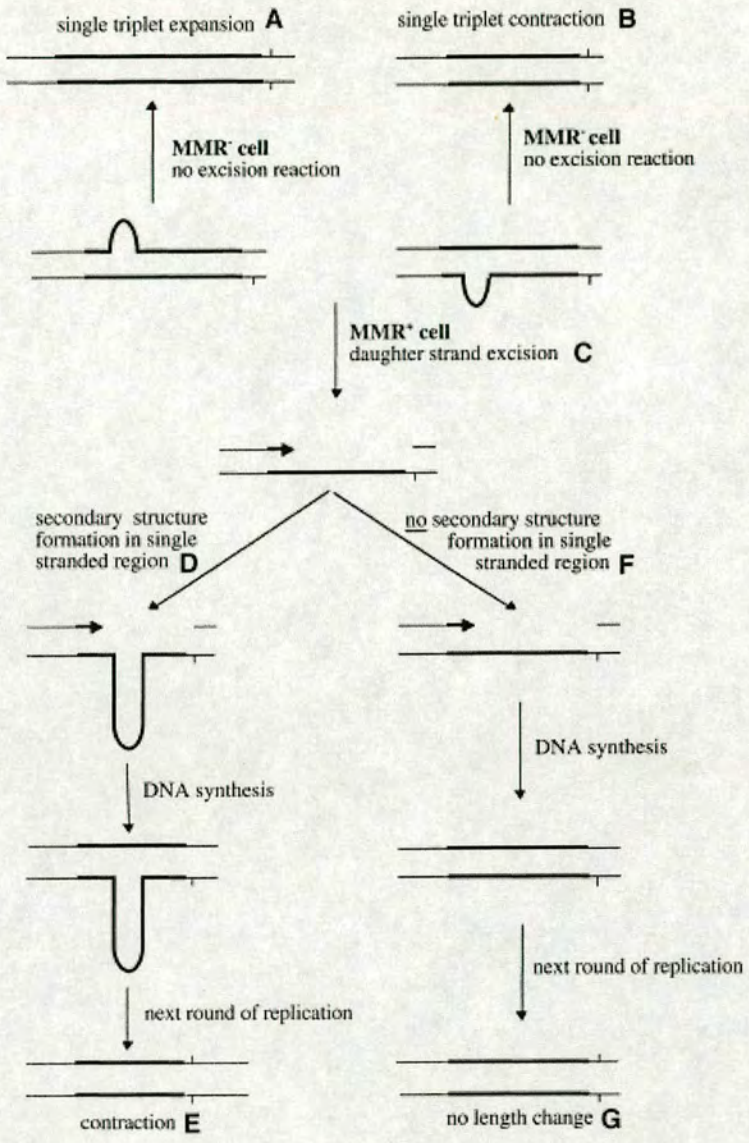


Figure 6.1: Model for MMR-promoted contractions of short plasmid born repeats in *E. coli* (Schmidt et al., 2000)
Mispairs of 3 bp preferentially happen during lagging strand synthesis and in the absence of MMR can lead to expansion (A) or contraction (B). In MMR+ cells, ss region is generated (C) and if the sequence folds into a hairpin structure (D), an MMR dependent contraction happens during repair (E). In the absence of secondary structure (F), functional MMR does not generate contraction.

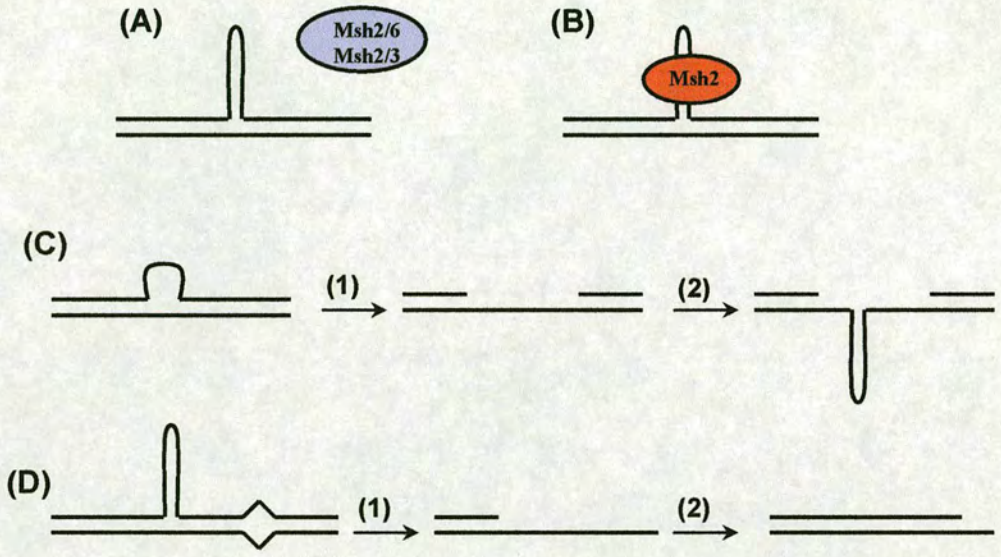


Figure 6.2: Models proposed for the involvement of MMR in TNRs instability, based on studies in *E. coli*, yeast and mice, adapted from (Lahue and Slater, 2003)
(A) A hairpin is formed during DNA synthesis and the expansion/contraction occurs unaffected by MMR. (B) Msh2/Msh3 complex binds mismatches within hairpin and stabilizes them leading to expansions. (C) MMR recognizes the small loops formed by polymerase slippage (1). Following excision, hairpin can form within the single strand region (2) resulting in MMR induced contractions. (D) MMR recognizes mismatched bases near a hairpin. (1) MMR proteins start co-excision of mismatch and hairpin (2) so reducing the instability by removing instability intermediates.

6.1.3 Nucleotide Excision Repair (NER) in *E. coli*

NER in *E. coli* removes bulky adduct DNA, acts on unnatural base modifications and excises damaged bases by an ATP dependent nuclease. The excision nuclease complex, Excinuclease (ABC) comprises UvrA, UvrB and UvrC proteins (Sancar and Rupp, 1983). UvrA is the specific DNA-damage binding protein, which makes a complex A_2B_1 with UvrB. Upon delivering UvrB to DNA, UvrA dissociates from the UvrB-DNA complex (Orren and Sancar, 1989). UvrC recognizes this complex, causing a conformational change in UvrB, which hydrolyses the fifth phospohodiester bond 3' of the lesion. The eighth phospohodiester bond 5' of the lesion is hydrolysed by UvrC. Then UvrD helicase displaces UvrC and the excised DNA. DNA polymerase I fills in the gap, displaces UvrB and the final nick is sealed by ligase (Lin et al., 1992; Orren et al., 1992).

6.2 Results

6.2.1 CAG repeat instability is unaffected in *mutL* and *mutH* mutants but increases in a *mutS* mutant

Instability for CAG repeat orientation was analysed in *mutS*, *mutL* and *mutH* mutants using the instability assay described in Chapter 2. Figure 6.3 shows that as compared to wild type, no significant difference in the proportions of instability was observed in *mutL* and *mutH* mutants. An increase in proportion of instability was observed in a *mutS* mutant ($p=0.050$) suggesting a stabilizing role of MutS protein. Though the marginal p-value of 0.050 renders the result to be either statistically

significant or non-significant, it was considered significant, taking into account the observations of other mutants, which are discussed later in this Chapter.

6.2.2 CTG repeat instability is unaffected in *mutS*, *mutL*, *mutH*, *mutS sbcCD*, *mutL sbcCD* and *mutH sbcCD* mutants

As shown in Figure 6.3, all three mutants of MMR containing CTG repeats had no significant difference in the proportion of instability from that in wild type background cells. Similarly, no effect was observed when *sbcCD* mutation was introduced in these mutants implicating no influence of these proteins in CTG repeat instability.

Very few deletion and expansion events were observed in the CTG orientation, which were not sufficient to make a comparison of the distributions observed in wild type. The actual percentages of deletions and expansion sizes are shown in Table 6.1. On the whole, deletions were distributed randomly in single and double mutants. A few expansions were observed too, in a range of 1-13% of the repeat sequence.

6.2.3 The stabilizing effect of *sbcCD* mutation is lost in *mutL* and *mutH* mutants in the CAG orientation

The increase in proportion of instability of CAG repeats in a *mutS* mutant was reduced less than two-fold in a *mutS sbcCD* mutant suggesting that SbcCD has a destabilizing effect in a *mutS* mutant similar to its effect in wild type background (Figure 6.3).

The levels of instability observed in *mutL* and *mutH* mutants were retained in *mutL sbcCD* and *mutH sbcCD* double mutants (Figure 6.3), which are higher than the level observed in a single *sbcCD* mutant. This suggests that these mutations have a destabilizing role in an *sbcCD* mutant.

6.2.4 Few small and big expansions but large deletions are observed in *mutS, mutL, mutH, mutS sbcCD, mutL sbcCD* and *mutH sbcCD* mutants in the CAG orientation

Figure 6.4 shows the expansion sizes of CAG repeats observed in *mutS, mutL, mutH, mutS sbcCD, mutL sbcCD* and *mutH sbcCD* mutants. The expansions in a *mutS* mutant fall in the range of 11-13% of the repeat tract with the exception of two expansions of 16 and 19%. In the *mutH* and *mutL* mutants, the expansions range from 3-8% of the repeat area. The expansions in double mutants, *mutS sbcCD, mutL sbcCD* and *mutH sbcCD* are concentrated in the range of 3-6% apart from a rare expansion of 46% of the repeat tract in the *mutL sbcCD* mutant.

The distributions of the deletion sizes of CAG repeats observed in *mutS, mutL, mutH, mutS sbcCD, mutL sbcCD* and *mutH sbcCD* mutants are shown in Figure 6.5. All observed deletions were plotted as a function of percentage of deletion size against the number of events. The distributions obtained in all mutants share a negatively skewed pattern with wild type distributions (see Chapter 3), as seen from the long tail towards the left. The medians of distributions of both single and double mutants fall in a range of 58-71% (Figure 6.5), which are similar to the medians observed in wild type strains. This shows that the studied MMR proteins do not influence the nature of the primary intermediate(s) involved in instability pathway(s).

6.2.5 CTG repeat instability is unaffected in *uvrA*, *uvrB* and *uvrC* mutants

Figure 6.6 shows that proportions of instability are unaffected by mutations in *uvrA*, *uvrB* and *uvrC* in the CTG orientation suggesting that these genes do not play a role in CTG repeat instability.

Table 6. 1

Deletions and expansions observed in MMR and sbcCD mutant cells containing CTG repeats

Strains characteristics	Deletions (%)	Expansions (%)
$(CTG)_{95}$ *mutS*	5, 7, 13, 21, 46, 47, 56, 58, 83, 84, 88	3, 10
$(CTG)_{95}$ *mutL*	1, 4, 59	13
$(CTG)_{95}$ *mutH*	1, 36, 62, 64	7, 13
$(CTG)_{95}$ *mutS sbcCD*	2, 17, 54, 60, 65, 67, 79	2, 2, 4, 9
$(CTG)_{95}$ *mutL sbcCD*	30, 31, 54, 54, 56	1
$(CTG)_{95}$ *mutH sbcCD*	2, 3, 19, 25, 39, 51, 57, 57, 63, 66, 79	0

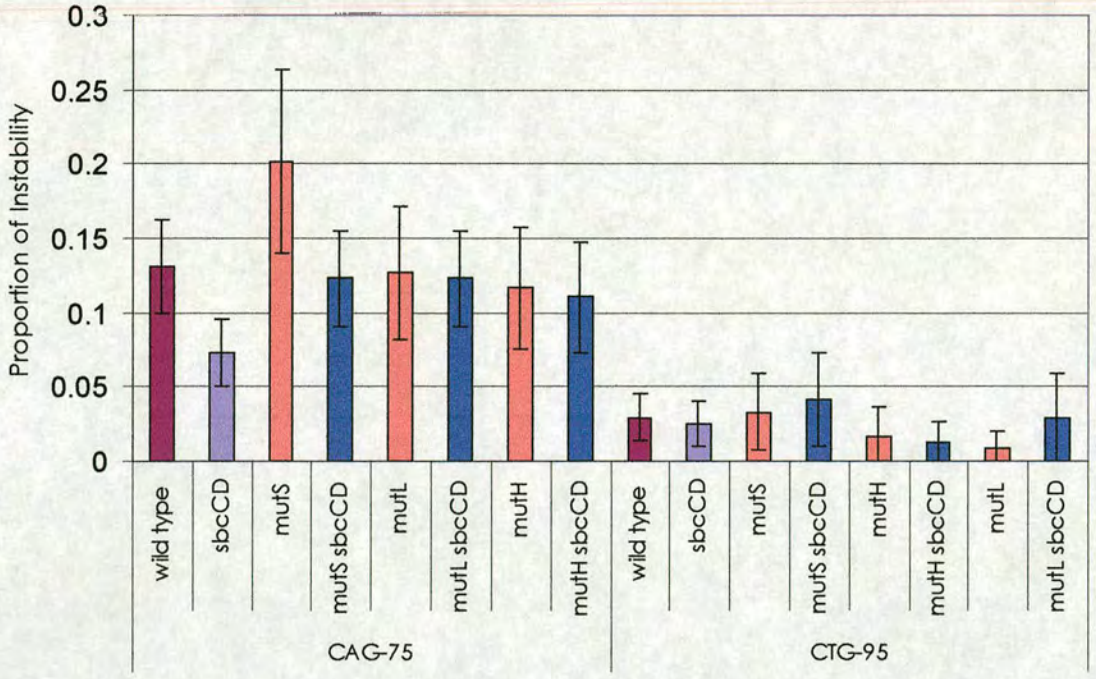
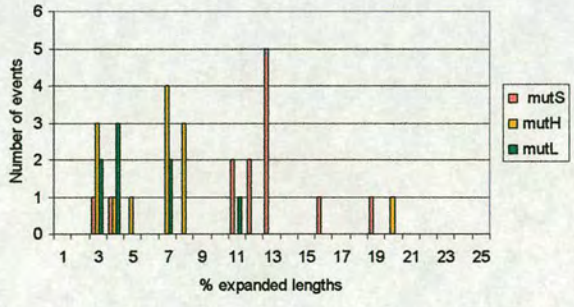


Figure 6.3: Instability Proportions for *mutS*, *mutL*, *mutH*, *mutS sbcCD*, *mutL sbcCD* and *mutH sbcCD* mutants containing CAG·CTG repeats
Each bar represents the pooled data of two independent experiments (480 clones). The error bars show 95% confidence intervals.



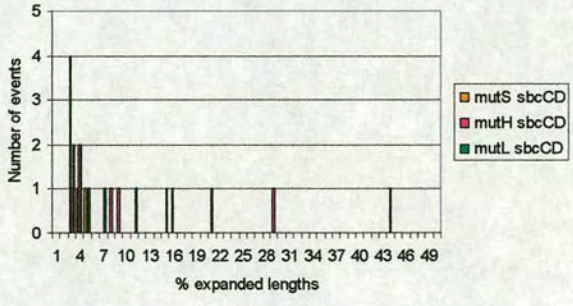


Figure 6.4: Distribution of expansion sizes in *mutS*, *mutH*, *mutL*, *mutS sbcCD*, *mutH sbcCD* and *mutL sbcCD* mutant cells containing CAG repeats
The expansions observed are plotted as percentage of tract expanded against the number of times the particular expansions were observed.

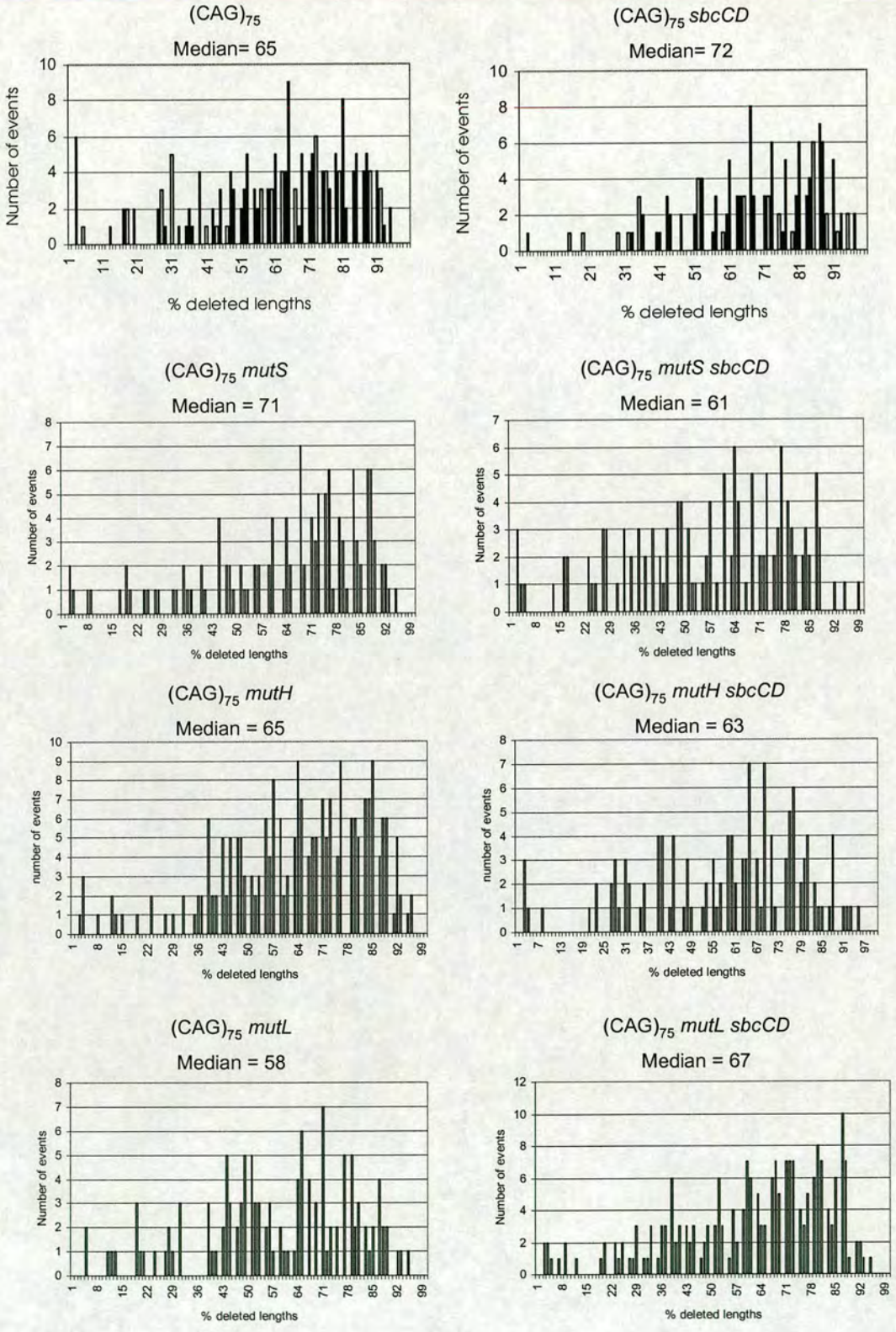


Figure 6.5: Distributions of deletion sizes in *mutS*, *mutH*, *mutL*, *mutS sbcCD*, *mutH sbcCD* and *mutL sbcCD* cells containing CAG repeats
The deletions observed are plotted as percentage of tract deleted against the number of times the particular deletions were observed.

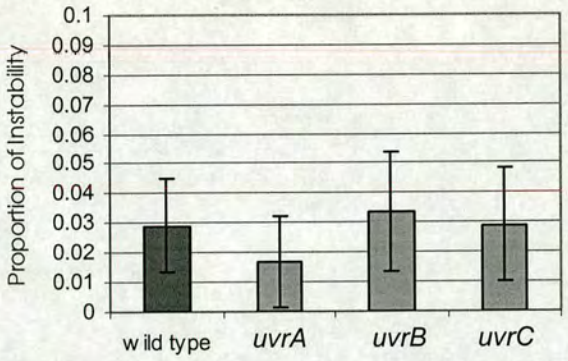


Figure 6.6: Instability proportions for *uvrA*, *uvrB* and *uvrC* mutants containing $(CTG)_{95}$ repeats
The error bars show 95% confidence intervals.

6.3 Discussion

The role of MMR pathway was studied in CAG·CTG repeat instability. The results demonstrate that MutS has a stabilizing role in CAG orientation and SbcCD can cleave hairpins in the presence of MutL and MutH. No effect of MMR proteins was found on CTG repeat instability. Furthermore, the NER pathway was also found to have no effect on the CTG repeat instability.

6.3.1 Mutations in genes encoding *mutL* and *mutH* do not affect CAG·CTG repeat instability except *mutS* in the CAG orientation

In this study, the effect of mutations in *mutS*, *mutL* and *mutH* on CAG·CTG repeat instability in the *E. coli* chromosome was investigated. In contrast to previous studies carried out in plasmids, no significant effect of MMR genes was observed on CTG repeat instability. Similarly, mutations in *mutL* and *mutH* did not affect CAG repeat instability in wild type cells while a small increase in instability was observed in a *mutS* mutant, suggesting that MutS has a stabilizing role in the CAG orientation.

E. coli MMR can correct heteroduplex loops of up to 4 bp (Parker and Marinus, 1992). So, it can only recognize small misalignments of 3 bp generated during replication. The distributions of deletion sizes observed in all three mutants of MMR in CAG orientation are negatively skewed as seen in wild type cells, showing that slippage during replication of TNRs generates big loops which will not be recognized by MMR and deleted by other pathway(s), as discussed in Chapter 4. Interestingly, accumulation of small slippages of plasmid born repeats has been observed in MMR mutants (Schmidt et al., 2000). The absence of these small changes in repeats integrated in chromosome suggests that either these slippage events do not happen in the chromosome or they are corrected by another route such as proofreading, in the absence of MMR.

6.3.2 Hairpins are inaccessible to SbcCD in the absence of MutL and MutH but not in the absence of MutS

In this work, the *sbcCD* mutation was found to have a stabilizing effect when CAG repeats were on the leading strand template (Chapter 4). This effect, along with the distributions of the repeat deletion sizes, both in wild type and in the *sbcCD* mutant, suggested the formation of secondary structures such as hairpins *in vivo*. These hairpins are targets of the SbcCD nuclease (Connelly et al., 1999) and have been suggested as binding sites for MutS homologues, Msh2 and Msh3 (Kovtun and McMurray, 2001; Manley et al., 1999; Sinden, 2001). Therefore, the *sbcCD* mutation was introduced into MMR deficient strains. The increase in the proportion of instability in a *mutS* single mutant was reduced to wild type level in a *mutS sbcCD* double mutant. This result suggests that MutS has a stabilizing role both in wild type cells and *sbcCD* mutant. Furthermore, the stabilizing effect of *sbcCD* mutation was

lost in *mutL* and *mutH* mutants, suggesting that these proteins have a stabilizing role in an *sbcCD* mutant and SbcCD cannot access hairpin structures in the absence of these proteins.

These results shed more light on the processes involved in generating substrates available for the action of SbcCD and add into the observations in the model for orientation dependent replicative instability described in Chapter 4. The model proposed (Figure 6.7) suggests that CAG repeats will form more stable structures (CTG hairpins) on the lagging strand template. MutS can bind hairpins and initiate a "repair" process, which contributes towards repeat stability. It is proposed that binding of MutS will trigger a "MMR-like" pathway which would essentially be a mismatch repair process since it requires all genes of MMR but differs from classical MMR in respects that it directs the removal of the new strand even when MutS encircles a mismatch exclusively on the template strand (Figure 6.7). It is speculated that this "repair" might happen during replication as well as after replication just like the well-defined post-replication MMR.

MutS binding will also protect the structures from being cleaved by SbcCD, which can have access to hairpins only if MutL and MutH induce MutS to cycle off the DNA (Figure 6.7). In the absence of MutS, the "repair" pathway would be blocked and SbcCD would have more access to its substrate hairpins (instability intermediates) resulting in an increase in instability.

These observations give further insight into orientation dependent instability. CAG repeats are more unstable as they form stable structures on the lagging strand template, which can face multiple pathways of deletion. The deletion distributions in wild type and mutants have also suggested the importance of these hairpin structures as intermediates of instability. CTG repeats are relatively stable because the structures

formed on the lagging strand template are unstable and the strand is proofread efficiently (Chapter 4).

6.3.3 NER is not involved in CTG repeat instability

This study demonstrates that the nucleotide excision repair pathway does not affect repeat instability when the CTG tract is on the leading strand template. Studies in literature have suggested both stabilizing and destabilizing role of NER in repeat instability. Increased instability was observed in a *uvrA* mutant but not in a *uvrB* mutant when $(CTG)_{175}$ repeats were studied but no effect of *uvrA* mutation was observed in a $(CTG)_{50}$ repeats containing strain (Parniewski et al., 1999). This stabilizing role of UvrA was attributed to its binding to hairpins and a consequent block to mutagenic bypass of the DNA hairpins. In another study, a decreased rate of deletions was observed in *uvrA* mutants containing $(CAG)\cdot(CTG)_{25\text{-}79}$ repeats (Oussatcheva et al., 2001). This destabilizing effect was suggested to be caused by the ability of UvrABC complex to recognize and excise loops formed during DNA slippage. The absence of an effect of NER genes on CTG repeat instability in this chromosomal study suggests that either large loops are not formed in this orientation or that they are not processed by this repair pathway. Though this orientation is stable enough to not give sufficient observations of deletions/expansions to have a distribution trend in mutants, the wild type distributions indicate the formation of big loops in this orientation too (shown in Chapter 3). Therefore, this work shows that big loops are formed in the CTG orientation but are not processed by the NER pathway. Differences observed between this chromosomal study and plasmid-based studies

could be attributed to the drawbacks associated with plasmid-based systems (as discussed in the introduction of this work in Chapter 1).

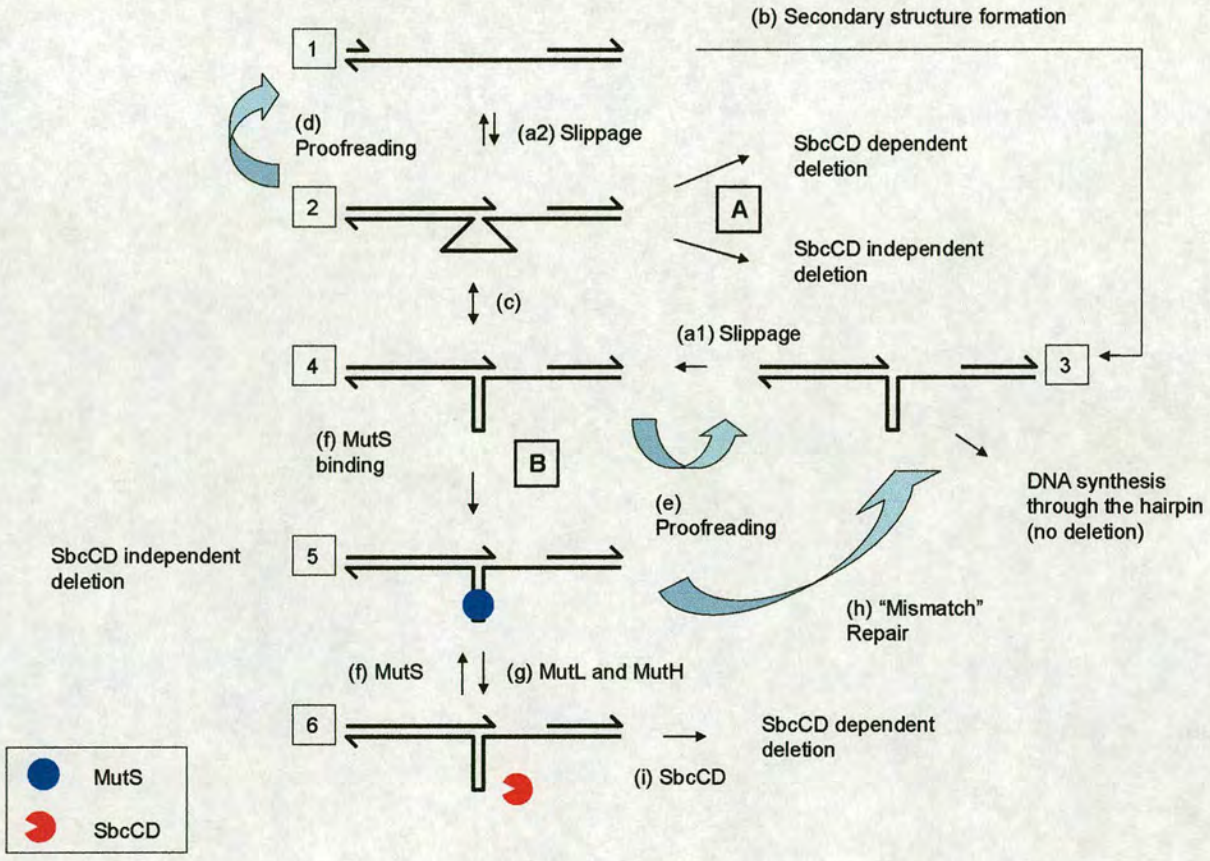


Figure 6.7: Model for processing of intermediates of instability in an orientation dependent manner

A lagging strand template containing CAG repeats will form less stable secondary structures (2) and so primarily will adopt pathway A whereas a template with CTG repeats (CAG orientation) will form more stable secondary structures (3 and 4) and so primarily will adopt pathway B. MutS can bind these stable structures (f) and direct to "MMR-like" repair (h). It is proposed in this model that MutS binding also protects the hairpins from being cleaved by SbcCD, which can have access to the structures only in the presence of MutL and MutH. These proteins can cycle MutS off the hairpin (g), giving SbcCD the freedom to cleave its substrates (i).

CHAPTER 7

CAG·CTG REPEAT INSTABILITY AND TRANSCRIPTION

7.1 Introduction

TNRs might form hairpin structures during cellular processes that involve opening of the DNA double helix, creating a single stranded region. These hairpin structures are hypothesized to be the intermediates of repeat instability as discussed in Chapters 4 and 6. Melting of duplex DNA during initiation of transcription and progression of RNA polymerase may favour the formation of these structures and could influence instability. Apart from the formation of secondary structures, transcription could also affect TNRs instability by destabilizing DNA polymerase during replication when a replication fork comes across a stalled transcription complex.

Transcription coupled repair (TCR) was first discovered in eukaryotes in the mid-1980s and was thought to be a mechanism specific to higher organisms. However, a study in 1989 showed that prokaryotes also exhibit preferential repair of transcribed strands compared to the rest of the genome (Mellon and Hanawalt, 1989). This repair process is needed when RNA polymerase encounters a DNA lesion that prevents further transcription. The *E. coli*, Mfd protein was identified as a transcription-repair coupling factor as it ensures the repair of DNA damage in transcribed strands of active genes (Selby and Sancar, 1993). Mfd is able to bind

DNA, RNA polymerase and the UvrA protein. It removes RNA polymerase from the DNA and recruits the excision repair apparatus to the damaged site. It is also required in the removal of stalled transcription complexes (Park et al., 2002).

In this study, CAG·CTG repeats of different lengths were integrated in frame at the 5'end of the *lacZ* gene adding CAG or CTG codons to the gene. Transcription of *lacZ* was induced by growing cells in the presence of IPTG and the effect of this process on CAG·CTG repeat instability was investigated. Further, repeat instability was studied in *mfd* mutants in both orientations.

7.1.1 *lacZ* gene

*lac*Z encodes the β-galactosidase enzyme, which cleaves lactose into glucose and galactose. Lactose is an inducer of the *lac* operon that binds the *lacI* repressor, preventing it from repressing *lacZ* expression. Studies on this operon have been facilitated by the use of an analogue of lactose, isopropyl beta-D-thiogalactoside (IPTG), and a non-inducing substrate, 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-gal). β-galactosidase hydrolyses X-gal giving a blue product. Under uninduced conditions (in the absence of the inducer IPTG), transcription from the *lacZ* promoter is repressed by the *lacI* repressor, while cells grown in the presence of IPTG remove *lacI* from the promoter and induce transcription.

The L8 mutation in the *lac* promoter is a nucleotide exchange of GC for AT. This mutation is in the cAMP receptor protein (CRP) binding site of the *lac* promoter and reduces promoter activity by inhibiting CRP-cAMP binding, or by altering the conformation of CRP-DNA (Fried and Crothers, 1983). With this mutation, the rate of

synthesis of β-galactosidase enzyme is only 2-6% of the fully induced wild type level (Yudkin, 1970).

7.1.2 Transcription and Repeat Instability

Trinucleotide repeats involved in diseases, such as myotonic dystrophy and spinal/bulbar muscular atrophy, show reduced transcription of the adjacent gene (Choong et al., 1996; Klesert et al., 1997). Different studies have shown that transcription through the repeated sequences can modulate their instability. In *E. coli*, transcription has been shown to reduce the frequency of expansions (Schumacher et al., 2001) and increase the deletion events of long CAG·CTG repeats (Bowater et al., 1997; Mochmann and Wells, 2004; Schumacher et al., 2001). In human cells, transcription promotes CAG repeat contractions via a pathway that requires repair processes, MMR and NER but does not depend on DNA replication (Lin et al., 2006). It appears that structures formed by trinucleotide repeats are a key source of instability since no effect of transcription was found on the instability of dinucleotide repeats (Bichara et al., 2000).

Many suggestions have been put forward by different studies regarding the link between transcription and repeat instability. It has been hypothesized that the DNA replication machinery can stall briefly on collision with RNA polymerase when both processes happen to meet (Cleary and Pearson, 2005) or the replication fork can collide with a repressor bound to an operator sequence (Schumacher et al., 2001). In *E. coli*, the rate of replication is faster than the rate of transcription, so the collisions are predestined. Head-on-collision of replication fork and RNA polymerase can lead to an increase in illegitimate recombination and deletions within plasmid vectors (Vilette et al., 1995; Vilette et al., 1996). Transcription can also influence the

formation of secondary structures such as S-DNA while opening of the DNA duplex by the transcription apparatus.

7.2 Results

7.2.1 Transcription does not influence CAG·CTG repeat instability

To analyse the influence of transcription on the instability of CAG·CTG repeats, different repeat lengths were integrated in the 5'end of the *lacZ* gene in the strain DL1786 that also bears a $lacI^q$ mutation in the promoter of the repressor gene. Transcription was induced by growing cells in the presence of 2mM IPTG. In CTG orientation, the transcribed strand is the CTG strand, which is also the leading strand template, while in the CAG orientation; it is the CAG repeat containing strand that acts as leading strand template and transcribed strand. The repeat lengths studied in both orientations in wild type cells showed no significant difference in the proportion of instability in the presence or absence of IPTG (Figure 7.1). Similar results were observed in all mutants analysed in this work. Since, no difference was observed, the +/- IPTG data were pooled in subsequent analyses (so individual +/- IPTG data for mutants are not shown).

7.2.2 A mutation in *mfd* does not affect CAG·CTG repeat instability

As shown in Figure 7.2, *mfd* mutants did not show a difference in the proportion of instability compared to the wild type levels in either repeat orientation. This suggests that Mfd protein is not involved in CAG·CTG repeat instability.

7.2.3 CAG·CTG repeats affect the LacZ phenotype in an orientation dependent manner

With repeats integrated in the *lacZ* gene, their effect on β-galactosidase activity was studied by plating cells on L-agar containing 2mM of IPTG and 40μg/ml of X-gal. As shown in Figure 7.3, the wild type strain (DL1786) displayed a blue phenotype as expected. The strain containing the vector p*Lac*D2 integrated in the chromosome, $(CNG)_0$, demonstrated a faint blue phenotype. This strain also contains the L8 mutation, which explains the faint phenotype. As described in the strain characteristics in Chapter 2, most of the strains constructed carried the L8 mutation after the integration of repeats into the chromosome, except $(CAG)_{8,\ 75,\ 84}$ and $(CTG)_{48,\ 95}$ lengths. Cells containing CAG repeats with no L8 mutation demonstrated a blue phenotype indicating a functional β-galactosidase. On the other hand, cells containing CTG repeats showed a white phenotype irrespective of the presence or absence of the L8 mutation. Therefore, the CAG orientation of the repeats in *lacZ* gene did not affect the gene expression but the presence of CTG repeats produced a non-functional or non-existent β-galactosidase. Repeat lengths were checked by PCR in white and blue colonies containing CTG or CAG repeats, respectively. Both orientations had the original repeat length, suggesting that the blue phenotype in CAG repeats was not due to a deleted repeat tract in *lacZ* and the white phenotype in CTG repeats was not due to a deleted *lacZ* gene. Colonies of cells containing CTG repeats could show some sectors of blue colour, so colonies were grown from a spot of culture of 1μl for five days at 37 °C. Blue sectors were developed in the colonies by the cells that deleted repeat tracts (Figure 7.4). The repeat length of cells from these blue sectors was checked by PCR and it was deleted from the original parental length.

Taking into account the phenotypic difference of the cells containing the two repeat orientations, the size and quantity of the β-galactosidase in strains containing repeats was determined by Western blotting. As shown in Figure 7.5, the protein expression observed in both repeat orientations was consistent with the phenotype of the strains. Strains containing CAG repeats but no L8 mutation showed β-galactosidase while no protein could be detected in strains with the L8 mutation. No protein was detected in strains containing CTG repeats irrespective of whether they contained the L8 mutation or not. Therefore, the quantity of β-galactosidase was different in function of the orientation of the repeats in the gene.

Additionally, the presence of CAG repeats in *lacZ* reduced the expression of the protein compared to the wild type level and according to the intensity of bands on the Western blot, this reduction seems to be dependent on the repeat length (Figure 7.5). The quantity of protein in the strain containing $(CAG)_8$ repeats was significantly less than in the wild type and it was further reduced with $(CAG)_{75}$ and $(CAG)_{84}$. The intensity of LacZ protein band appears the same in all three CAG repeat lengths ($(CAG)_8$, $(CAG)_{75}$ and $(CAG)_{84}$) but the amount of protein loaded in $(CAG)_8$ was considerably less than in the wild type and higher repeat lengths of 75 and 84 as seen from more intense bands throughout in the lanes of wild type, $(CAG)_{75}$ and $(CAG)_{84}$ as compared to $(CAG)_8$. It was also confirmed by calculating the band intensity in these lanes using ImageJ software version 1.36b. Therefore, it is speculated that there would have been more β-galactosidase in $(CAG)_8$ than in higher repeat lengths, should there have been the same amount of protein loaded.

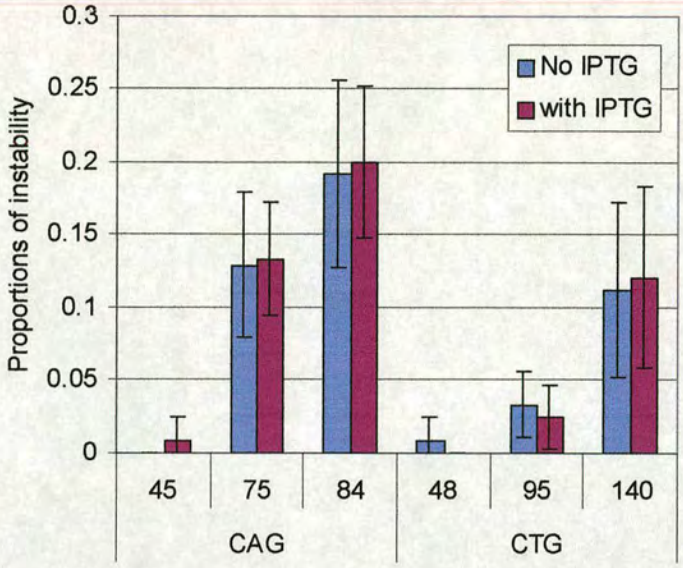


Figure 7.1: Proportions of instability for CAG·CTG repeats in the presence or absence of IPTG

The cells were grown overnight in the presence and absence of IPTG (2mM). X-axis shows the repeat lengths for CAG and CTG repeat orientations. The bars represent the data from two independent assays and correspond to the individual analysis of 240 clones. The error bars represent 95% confidence limit.

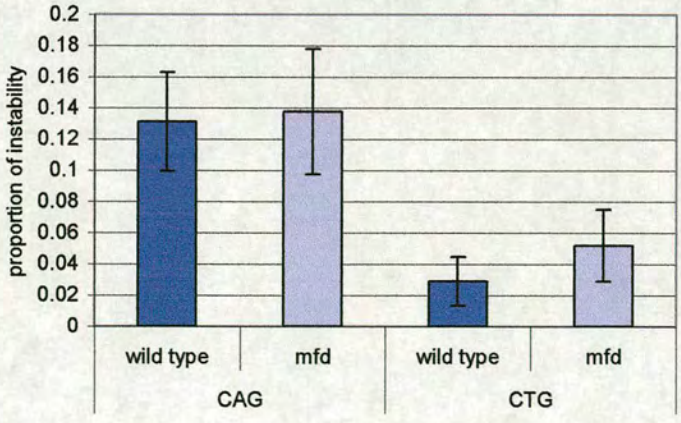


Figure 7.2: Instability proportions of $(CAG)_{75}$ and $(CTG)_{95}$ repeats in *mfd* mutants

Each bar represents the pooled data of two independent experiments (480 clones). The error bars show 95% confidence intervals.

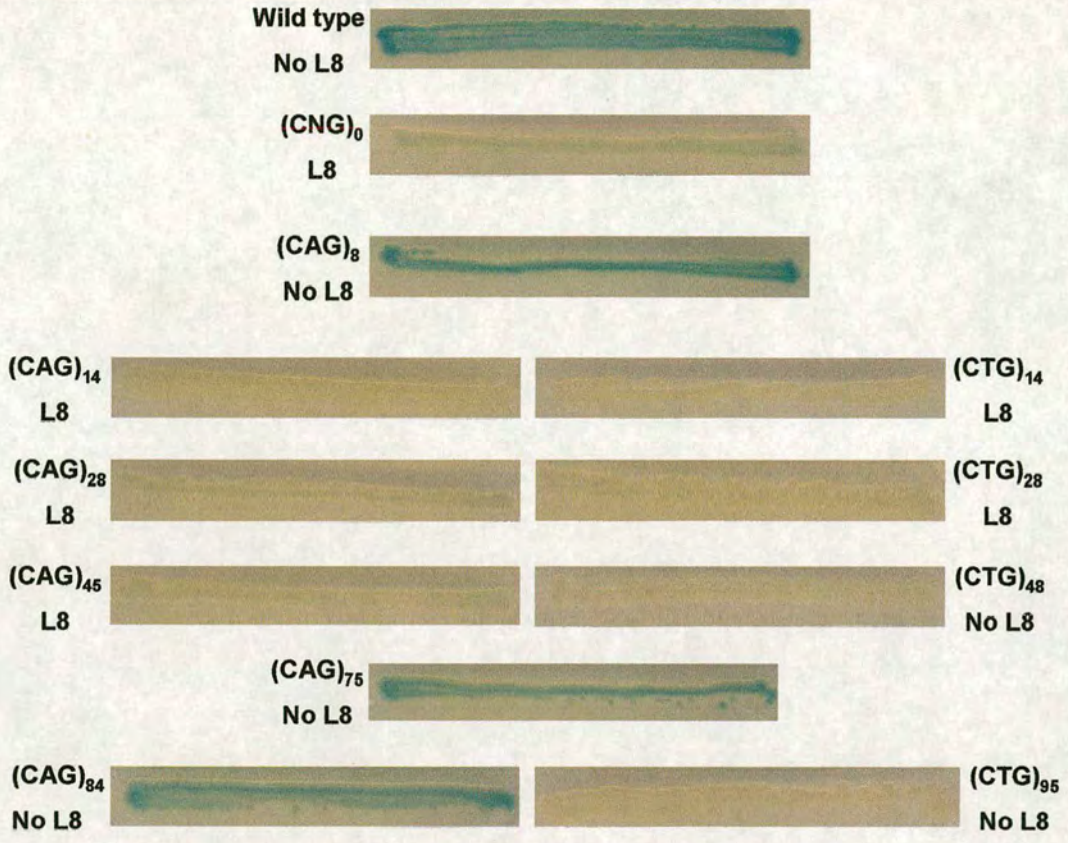


Figure 7.3: β-galactosidase phenotype of strains containing CAG·CTG repeats in the *lacZ* gene

The strains were streaked on L-agar containing 2mM IPTG and 40µg/ml X-gal. Wild type cells displayed a blue colour, indicative of a functional β-galactosidase. The introduction of the L8 mutation strongly reduces *lacZ* expression, as seen from the very faint blue colour of $(CNG)_0$. In CAG repeats, the only repeat lengths that gave blue colour did not have the L8 mutation (i.e. $(CAG)_8$, $(CAG)_{75}$ and $(CAG)_{84}$) while all CTG repeat lengths gave a white phenotype, irrespective of the L8 mutation. $(CTG)_{48}$ and $(CTG)_{95}$ had not the L8 but still displayed a white phenotype indicative of a non-functional or non-existent β-galactosidase in the presence of CTG codons.

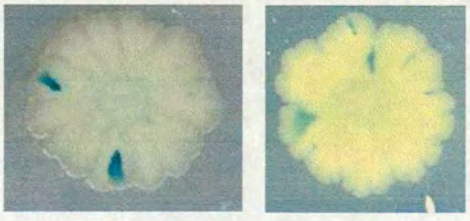


Figure 7.4: Blue sectors developed by deleted CTG tracts in *lacZ*

A 1 µl spot of culture was grown on L-agar containing 2mM IPTG and 40µg/ml X-gal. The spot grows developing blue sectors containing cells that deleted repeat tracts.

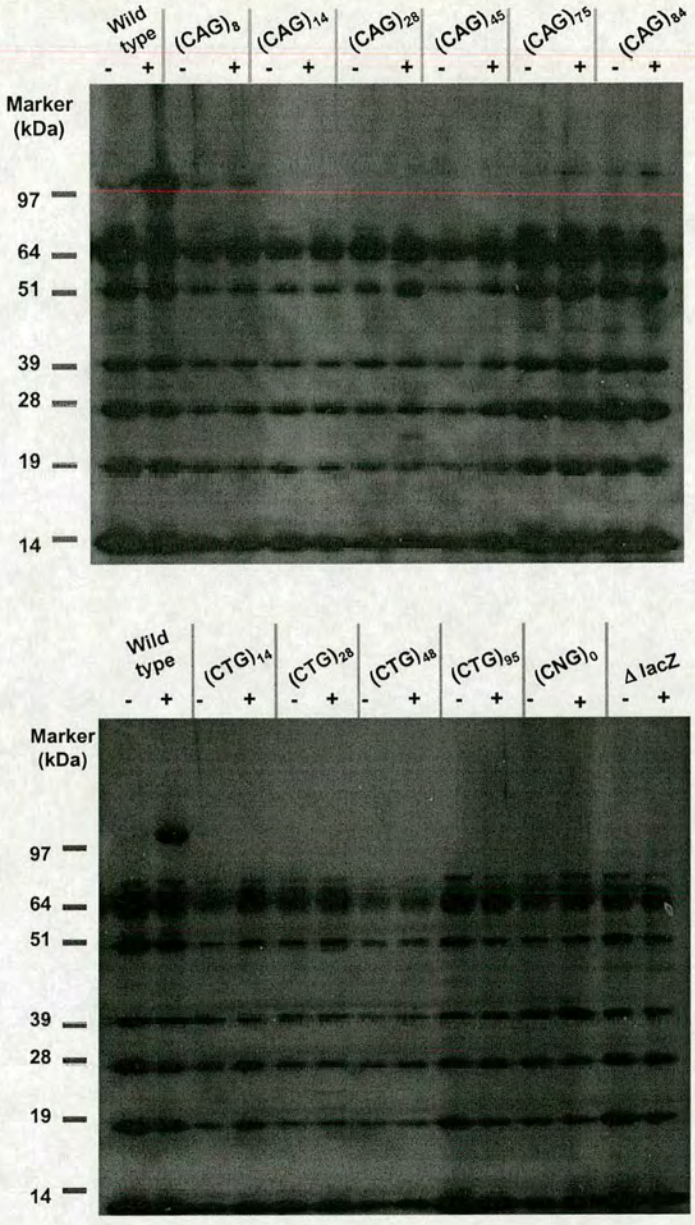


Figure 7.5: Western blots for β-galactosidase in strains containing CAG·CTG repeats

Top: $(CAG)_8$ displays β-galactosidase of the similar size of wild type, 114kDa. $(CAG)_{75}$ and $(CAG)_{84}$ display β-galactosidase of larger size than the wild type. The yield of protein in all three repeat lengths is significantly less than in the wild type.

Bottom: No β-galactosidase was detected in any of the strains containing CTG repeats 14, 28, 45, and 95. The strain containing p*Lac*D2 integrated with $(CNG)_0$ also showed no protein product since it contains L8 mutation. A strain with *lacZ* deletion was used as negative control.

+ and – correspond to the presence and absence of 2mM IPTG, respectively.

7.3 Discussion

In this study, no significant effect of transcription was observed on CAG·CTG repeat instability in the *E. coli* chromosome. It was expected that instability would be affected by transcription since the repeats are located in the *lacZ* gene as shown in Figure 3.1 (Chapter 3) where replication and transcription work in opposite directions. In many experimental systems it has been suggested that problems arise when a replication fork meets a transcription complex on plasmid DNA (French, 1992; Vilette et al., 1995). Previous studies of repeat instability have suggested such collisions of replication fork and transcription complex to give rise to repeat instability in plasmids (Bowater et al., 1997; Schumacher et al., 2001). A lack of effect of transcription in wild type and the mutants studied in this work suggests that this process does not influence repeat instability in the *E. coli* chromosome.

A mutation in the *mfd* gene did not change the proportion of instability in CAG and CTG repeat orientations, suggesting no role of the transcription repair coupling factor Mfd in repeat instability. This result indicates that transcription through CAG·CTG repeats does not seem to cause any block to the RNA polymerase, hence does not require any repair coupled with transcription.

Repeat sequences were built in plasmid p*Lac*D2, which contains the L8 mutation in the *lacZ* promoter region. While integrating repeats in the chromosome, the L8 mutation was transferred in many of the strains but the main lengths studied in all mutants $(CAG)_{75}$ and $(CTG)_{95}$ had no L8 mutation. This mutation decreases the synthesis of β-galactosidase enzyme to only 2-6% of the fully induced wild type level (Yudkin, 1970) explaining the absence of protein in Western blots in all strains carrying the L8 mutation (Figure 7.5). However, there is a considerable difference between the two repeat orientations both in phenotype and β-galactosidase level. *lacZ*

containing CAG repeats encodes a functional protein, as it can metabolise X-gal into a blue product in the presence of IPTG (Figure 7.3) but the *lacZ* containing CTG repeats encodes a non-functional or non-existent protein as it cannot metabolise X-gal and displays a white phenotype. This suggests that either the β-galactosidase encoded by the gene containing CTG repeats is not being formed or it is produced but is non-functional because of a poly-leucine stretch or it is being degraded. Since no protein with CTG repeats was detected on Western blot, the β-galactosidase with the poly-leucine stretch was either not formed or quickly degraded.

CHAPTER 8

Rep AND UvrD HELICASES: IMPLICATIONS IN REPEAT INSTABILITY

8.1 Introduction

Replication fork progression can be impeded upon inactivation of essential components of the replisome, collision with the transcription apparatus, encountering DNA lesions and DNA bound proteins (Liu and Alberts, 1995; Vilette et al., 1992). Among choices for the cell to resume replication, one is to reverse the fork and transform it into a recombination substrate (Seigneur et al., 1998). Recent studies of replication restart using this fork reversal strategy have manifested the involvement of two helicases, Rep and UvrD, along with other players. Pausing of DNA synthesis has been found *in vitro* at specific loci in double stranded TNRs (Kang et al., 1995b). My study was carried out to address the questions: how do the cells maintain repeats when a fork is forced to progress slowly by introducing a *rep* mutation and how does the unwinding function of replicative and repair helicases, Rep and UvrD, respectively affect CAG·CTG repeat instability? The work described in Chapters 5 and 7 shows that the genes involved in DNA recombination and transcription have no significant effect on repeat instability in the *E. coli* chromosome. These findings reinforce the major role of replication in repeat instability, as discussed in Chapter 4.

The instability assay described in Chapter 2 was used to study the effects of *rep* and *uvrD* genes on $(CTG)_{95}$ and $(CAG)_{84}$ repeats. These were then compared to the instabilities studied in wild type *E. coli* cells. Finally, the role of *recF* in a *urvD*

mutant was studied since RecF was shown to be involved in instability of tandem repeats in *uvrD* mutants (Bierne et al., 1997). The data reflect upon the role of Rep and UvrD helicases in repeat instability.

8.1.1 DNA Helicases

Helicases are the enzymes that act as molecular motors to unwind the duplex DNA using the energy released by the hydrolysis of nucleoside phosphates (van Brabant et al., 2000). *E. coli* has eleven distinct DNA helicases including replicative helicases, DnaB, PriA, Rep; recombination and repair helicases, UvrAB complex, helicase II, RecQ, RecG, RecBCD and RuvAB (Matson et al., 1994). Two helicases, Rep and UvrD, used in this study will be discussed in detail here.

8.1.1.1 Rep Helicase

The replicative helicase, Rep, belongs to helicase superfamily I. The exact role of Rep in cellular processes is not yet known but several observations suggest its function in DNA replication. The replication fork progresses two times slower in a *rep* mutant than in a *rep*+ strain (Lane and Denhardt, 1974; Lane and Denhardt, 1975). Rep is required for the replication of a number of ssDNA phages like ΦX174 and M13 (Takahashi et al., 1979). Cells lacking Rep are viable, suggesting that a second helicase may compensate for loss of Rep. This could be UvrD since *rep uvrD* double mutants are non-viable (Washburn and Kushner, 1991). Rep unwinds the nascent lagging strand in the presence of SSB on the leading strand template (Heller and Marians, 2005) and has been proposed to limit the loading of RecA on damaged forks

by facilitating the closure of any gaps on the leading strand (Mahdi et al., 2006). RecBC is essential for the viability of a *rep* mutant (Uzest et al., 1995; Washburn and Kushner, 1991).

8.1.1.2 UvrD Helicase

UvrD, also known as helicase II, is a dimeric protein that belongs to helicase superfamily I. It acts in a 3' to 5' direction. At low concentration, it can unwind only a partially single stranded substrate but at high concentration, it can start activity at a nick or a blunt end (Runyon et al., 1990). It has been documented to be involved in the last steps of DNA repair processes. During nucleotide excision repair, it removes the 12 nucleotide long DNA segment containing the lesion, after its incision by UvrA, UvrB and UvrC (Orren et al., 1992). During mismatch repair, it removes the DNA segment containing the mismatched DNA after its incision by MutS, MutL and MutH (Modrich, 1994). The enzyme is conserved in prokaryotes and shares a eukaryotic orthologue Srs2, known to be involved in replication and recombination in yeast. In addition to its helicase activity, it plays a role as protein remover *in vivo* since it can dislodge RecA filaments from single stranded DNA, suggested to be the reason of lethality of the *rep uvrD* mutant (Veaute et al., 2005).

UvrD can play a role in DNA replication and this notion is supported by several observations. Several alleles of *uvrD* constitutively induce the SOS response (Ossanna and Mount, 1989). *uvrD rep* double mutants are non-viable (Washburn and Kushner, 1991). *uvrD* mutations have been found to increase the frequency of homologous recombination, either because of their replication defects or the anti-recombinase activity of the helicase.

8.1.2 Replication Fork Reversal (RFR) Reaction

Stalled replication forks can restart by reassembly of the replication machinery. Among several routes that a fork can take to restart replication, one is replication fork reversal where the nascent leading and lagging strands can anneal resulting in the formation of a four armed DNA structure (chicken foot) with a DNA double-strand end (Figure 8.1). The recombinase exonuclease RecBCD processes the DNA double strand end, unwinding and degrading the DNA until it reaches an octameric sequence, c*hi*, where the activity of enzyme is modified and it loads RecA onto DNA. RecA catalyses strand-exchange resulting in the formation of a Holliday junction, which is resolved by RuvABC resolvase.

The fork reversal model has been supported by genetic data and direct analysis of fork breakage. This was observed in several *E. coli* mutants including three Polymerase III mutants, *dnaEts*, *dnaNts* and $holD^{Q10am}$ (Flores et al., 2001; Grompone et al., 2002) *rep* and *dnaB*ts helicase mutants (Seigneur et al., 2000) and the *priA* mutant (Grompone et al., 2004). The DNA repair helicase UvrD was shown to be essential for RFR in *dnaEts* and *dnaNts* Polymerase III mutants (Flores et al., 2004) which was attributed to its role in removing RecA filaments from the fork (Flores et al., 2005).

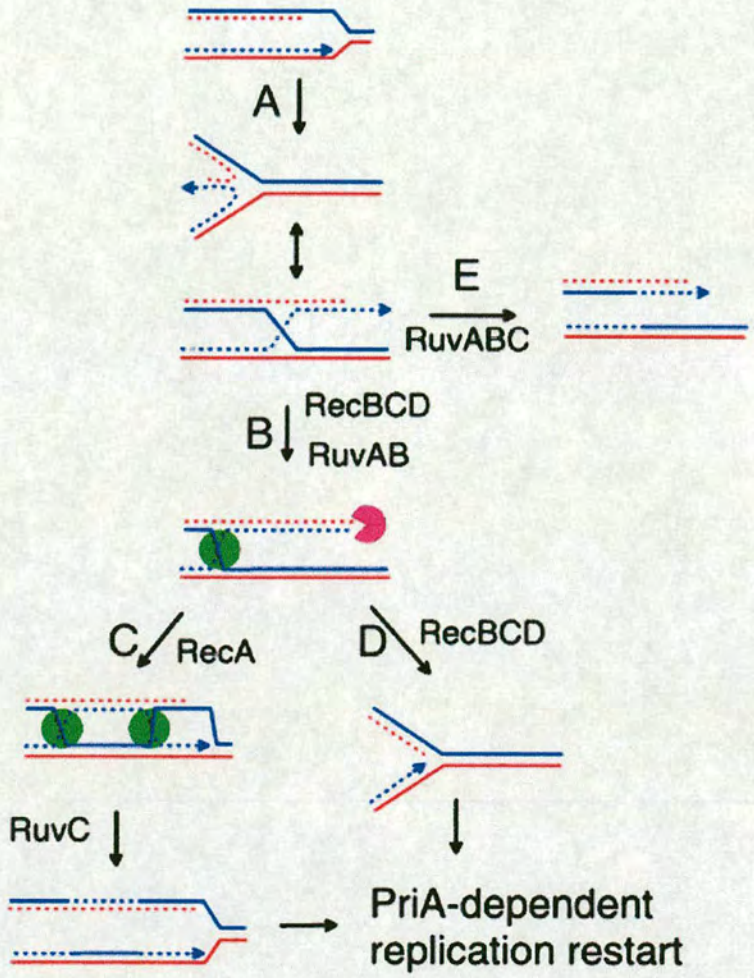


Figure 8.1: Replication fork reversal model (Michel et al., 2004)
In the first step (A), the replication fork is arrested, causing fork reversal, which forms a Holliday junction (HJ). In Rec+ cells (B and C), RecBCD (in pink) resolves this fork by initiating RecA-dependent homologous recombination (HR) at a *chi* site on the DNA double strand end. RuvC (green circle) resolves the two HJs, one formed by fork reversal and the other by HR. Alternatively, if RecBCD encounters the HJ before *chi* or in the absence of RecA, the double strand end is degraded up to the HJ and replication fork is restored. In the absence of RecBCD (E), RuvABC resolves the HJ resulting in chromosome linearization.

8.2 Results

8.2.1 CTG repeat instability increases while CAG repeat instability is unaffected in a *rep* mutant strain

Figure 8.2 shows the instability proportions for *rep* mutant strains containing $(CTG)_{95}$ and $(CAG)_{84}$ repeats along with their corresponding wild type proportions. The bars represent the average of instability proportions analysed in two independent experiments. The proportion of instability was increased 4.8-fold for $(CTG)_{95}$ in a *rep* mutant compared to wild type while a non-significant 1.2-fold increase was observed for $(CAG)_{84}$.

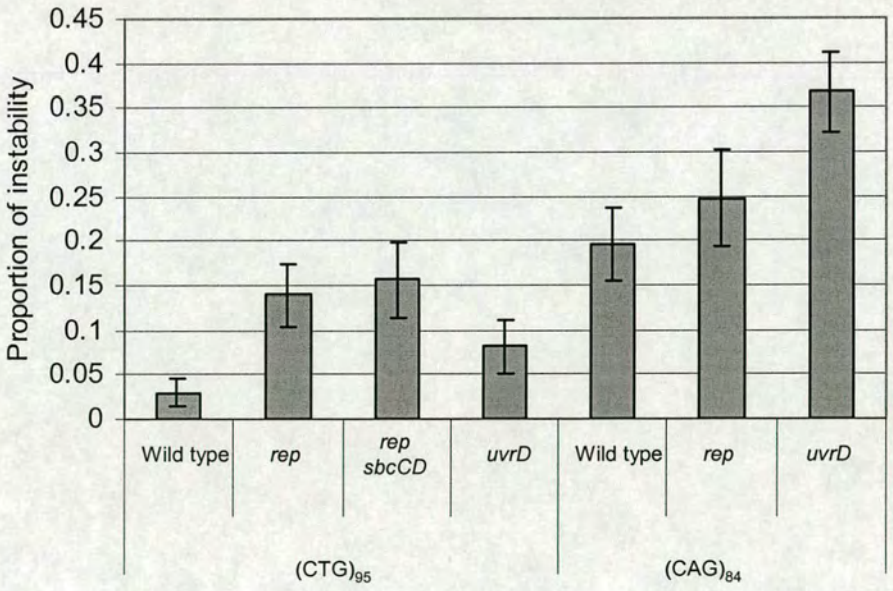


Figure 8.2: Instability proportions for *rep*, *rep sbcCD* and *uvrD* mutants of CAG·CTG repeats compared with wild type cells
Repeat length for CTG is 95 and for CAG is 84. Each bar represents the proportion of instability (pooled data of two independent assays – 480 clones). The error bars show 95% confidence intervals.

8.2.2 Large deletions are predominant in the CAG orientation while the CTG orientation shows both large and small deletions in a *rep* mutant

In order to see the sizes of deletions obtained in the *rep* mutant, all observed deletions were plotted as a function of percentage of deletion sizes against the number of events (Figure 8.3). CAG repeats show a negatively skewed distribution (long tail towards left), with a median of 62% giving the same pattern seen in wild type cells. This suggests that the intermediates formed in a CAG *rep* mutant are of same nature as in wild type since it also does not affect the frequency of deletions in this orientation.

The CTG deletion distribution shows few large deletions but predominantly an extreme positive skewness is observed with a median of 30% as compared to 58% in wild type (Figure 8.3). Since, large deletions are also observed in wild type cells, the intermediates generating these deletions may be the same as in wild type. The pattern of small deletions suggests the generation of a new class of intermediates with small loops during frequent slippage in a *rep* mutant. So Rep helicase participates in determining the nature of the primary intermediate in instability pathway(s) of CTG repeats.

8.2.3 Small expansions of CAG·CTG repeats are observed in a *rep* mutant

Both CAG and CTG repeats showed predominantly deletions as a result of instability in wild type with few expansions. But in a *rep* mutant, frequent small and few large expansions were observed in both orientations (Figure 8.3). Most expansions were up to 10-15% of the repeat tract in both orientations with few big expansions. Interestingly, the CAG orientation displays bigger expansions than the CTG

orientation. The biggest expansion observed in the CTG orientation was of 27% of the repeat sequence while in the CAG orientation, it was 37%.

8.2.4 CTG repeat expansions in the *rep* mutant are independent of SbcCD nuclease

The instability in the *rep* mutant was also found to be unaffected by an *sbcCD* mutation as seen from the same proportion of instability in both *rep* and *rep sbcCD* mutants (Figure 8.2).

CTG repeat expansions observed in the *rep* mutant were found to be independent of the presence of SbcCD nuclease since the pattern of expansions observed in a *rep sbcCD* double mutant was the same as in a *rep* mutant (Figure 8.4).

8.2.5 *uvrD* mutation enhances the CAG·CTG repeat instability

As shown in Figure 8.2, instability was increased for both repeat orientations in a *uvrD* mutant. The instability proportion for $(CTG)_{95}$ was increased 2.8-fold in the *uvrD* mutant compared with that of wild type while $(CAG)_{84}$ had an instability proportion only 1.9-fold higher than that in the wild type.

8.2.6 RecF is required for CTG repeat deletions in a *uvrD* mutant

Figure 8.5 shows that the increase in proportion of instability in a *uvrD* mutant was suppressed by introducing a *recF* mutation. The 2.8-fold increase of CTG repeat instability in the *uvrD* mutant went down to wild type level in a double *recF uvrD*

mutant, which implied that RecF was required for the deletion events of CTG repeats in the *uvrD* mutant.

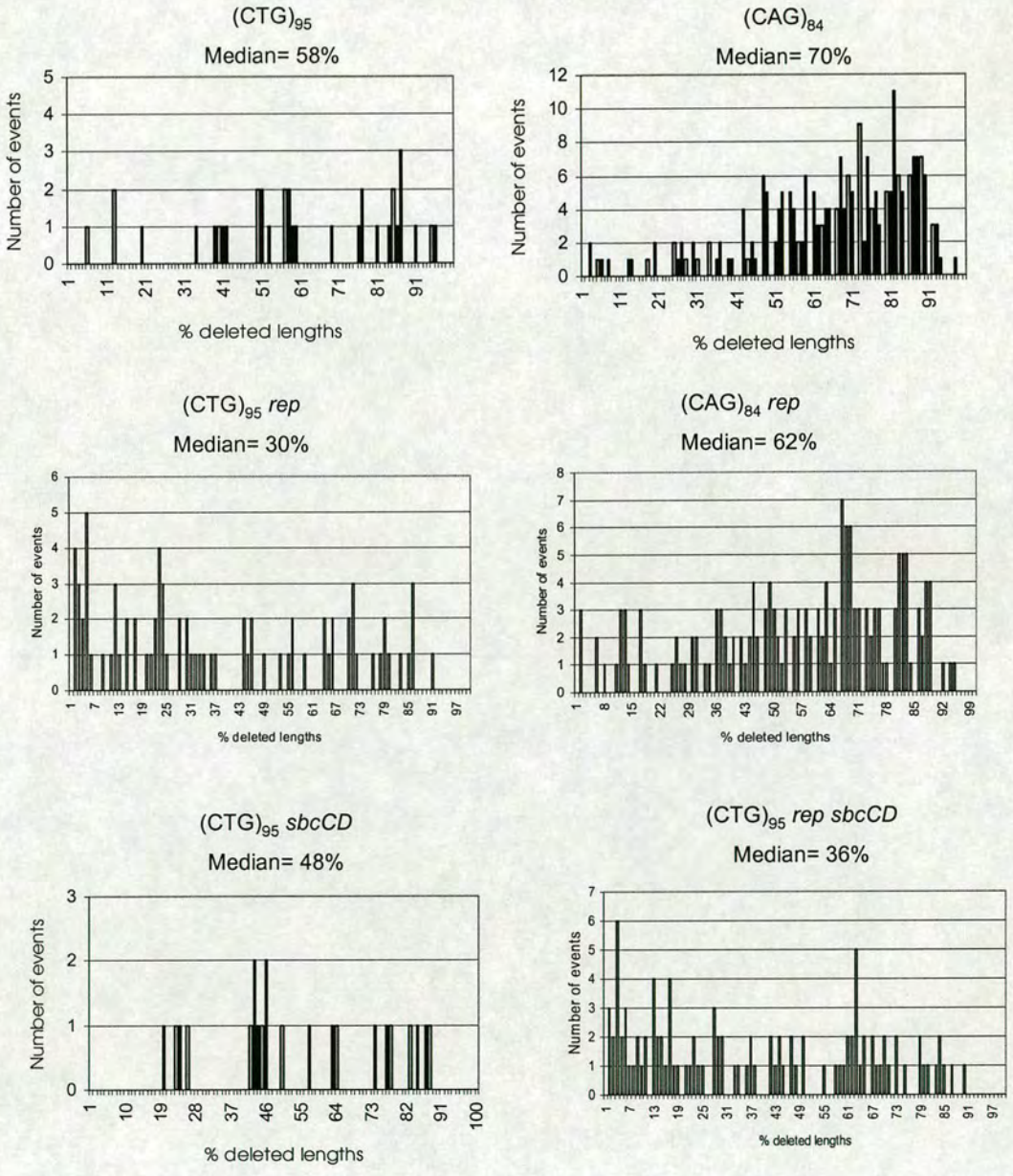


Figure 8.3: Distributions of deletion sizes in *rep* and *rep sbcCD* mutant cells containing CAG·CTG repeats.
The deletions observed are plotted as percentage of the tract deleted against the number of times the particular deletions were observed.

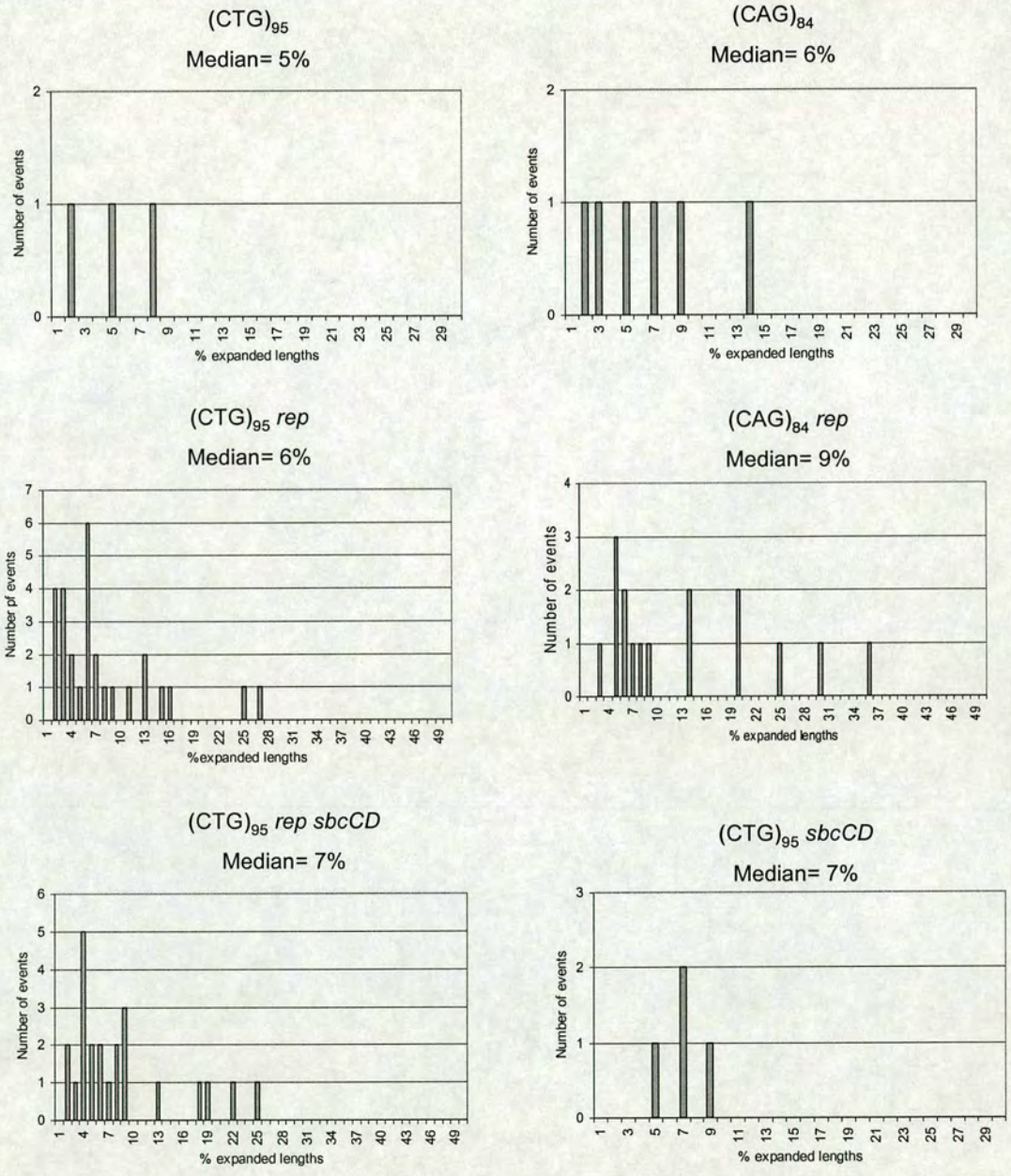


Figure 8.4: Distributions of expansion sizes in a *rep* mutant containing CAG·CTG repeats.

The expansions observed are plotted as percentage of the tract expanded against the number of times the particular expansions were observed.

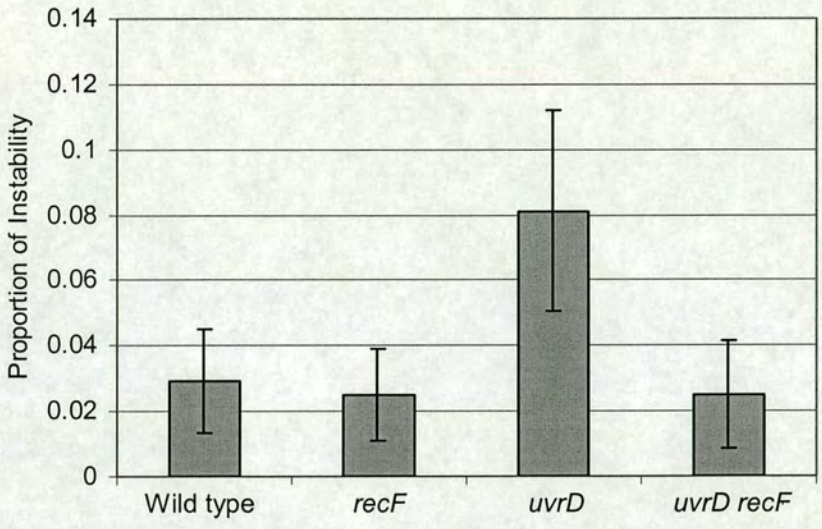


Figure 8.5: Instability Proportions for *recF*, *uvrD* and *recF uvrD* mutants containing $(CTG)_{95}$ repeats
Each bar represents the proportion of instability (pooled data of two independent assays – 480 clones). The error bars show 95% confidence intervals.

8.2.7 Large deletions predominate over small deletions in a *uvrD* mutant in CAG orientation

Figure 8.6 shows the distributions of repeat sizes deletions observed in both orientations of the repeat tracts. CAG orientation displays a pattern of large deletions with skewness in negative direction with a median of 60%. The CTG distribution shows a slight positive skewness with a median of 46% in contrast to negatively skewed distribution in wild type. This suggests that the absence of UvrD affects the nature of the precursor intermediates of instability in CTG orientation along with their processing to new products with altered repeat length.

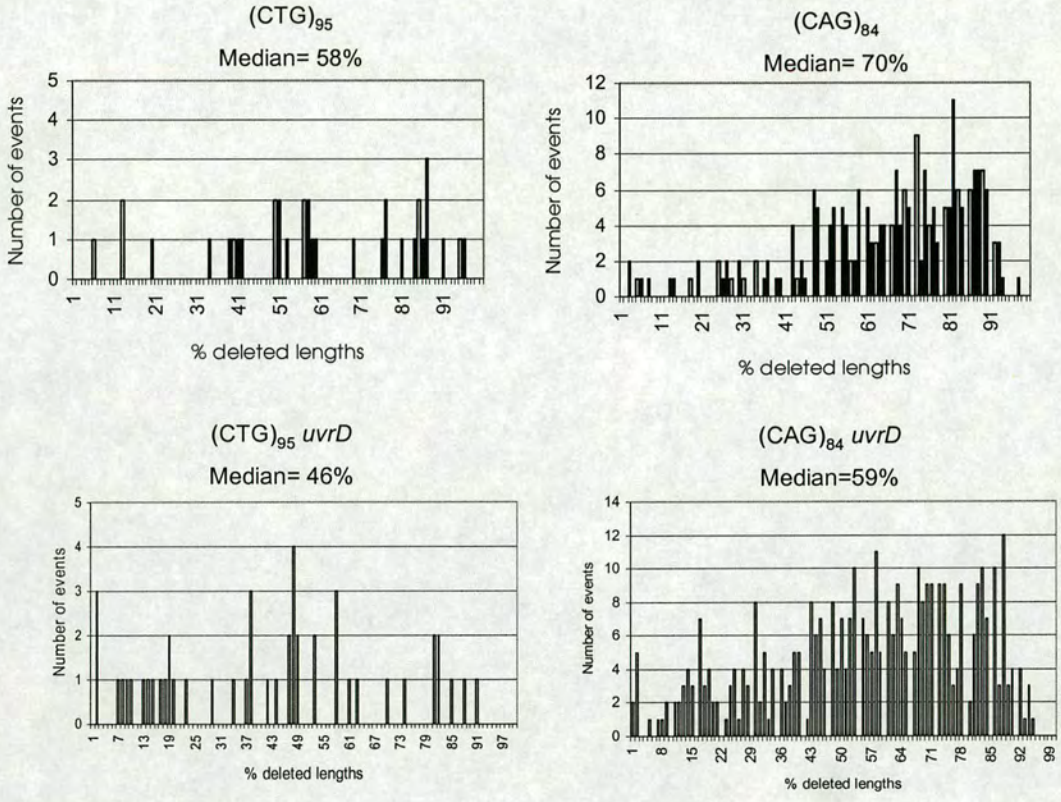


Figure 8.6: Distributions of deletion sizes in *uvrD* mutant cells containing CAG·CTG repeats.
The deletions observed are plotted as percentage of the tract deleted against the number of times the particular deletions were observed

8.3 Discussion

The roles of Rep and UvrD helicases in CAG·CTG repeat instability were studied. CTG repeats were destabilized in a *rep* mutant while a mutation in *uvrD* affects instability in both repeat orientations.

8.3.1 *rep* mutation destabilises CAG·CTG repeats when the CTG repeat tract is on the leading strand template

The work demonstrates that a mutation in Rep helicase affects CAG·CTG repeat instability depending upon the repeat orientation with respect to the direction of replication. The orientation, where CTG repeats were on the leading strand template, was significantly destabilised as compared to wild type while no significant difference in instability was observed when CAG repeats were on the leading strand template.

In a *rep* mutant, the replication fork progresses two times slower than in wild type (Lane and Denhardt, 1975). It has been shown that $(CTG)_n$ hairpins are more stable than $(CAG)_n$ hairpins (Petruska et al., 1996). The data in this study suggest that in a *rep* mutant, pausing is more frequent when stable structure forming CTG repeats are on the leading strand template. Such frequent pausing gives rise to frequent slippage and more small deletions as evident from the CTG repeat deletion pattern.

The *rep* mutation does not affect either the frequency of CAG repeat instability or the repeat deletion distribution, suggesting no contribution of Rep helicase to the deletion events involved in CAG repeat instability. Repeats are difficult sequences for DNA polymerase to copy as experienced from difficulties in amplification of longer sequences by PCR and *in vitro* studies have shown that CAG·CTG repeats cause length dependent pausing by bacterial and phage polymerases (Kang et al., 1995b; Ohshima and Wells, 1997). The model proposed in this study (Figure 8.7) suggests that in the absence of Rep helicase replication pausing causes the slow progression of the fork. The leading and lagging strand polymerases can become uncoupled while allowing the synthesis of lagging strand. This will generate gaps on the leading strand template. It has been shown that Rep helicase

unwinds the nascent lagging strand, to provide a "landing pad" for PriC to load DnaB to restart replication (Heller and Marians, 2005). Absence of Rep will permit the longer persistence of gaps on the leading strand template, which can fold and be deleted when CTG repeats are on the leading strand template (Figure 8.7).

The absence of effect of the *rep* mutation on CAG repeats suggest that either frequent pausing does not happen with CAG repeats on the leading strand template or the gaps created do not fold into a stable structure since CAG repeats form less stable hairpins and replication continues as it would in a rep mutant irrespective of repeated sequence or not (Figure 8.6).

8.3.2 *rep* mutation generates expansions in both orientations of CAG·CTG repeats

Mutation of the *rep* helicase gives rise to an increase in the frequency of expansions in both orientations of repeat tract, which are supposed to arise from a structure formed on the nascent lagging strand. According to the traditional replication slippage model (discussed in Chapter 1), expansions are preferred when CTG repeats are on the nascent leading strand. Since, in this study, both orientations show expansions with CAG repeats giving larger expansions, the traditional model based on hairpin stabilities cannot be applied to explain this scenario. So the question was what would drive the formation of a hairpin on one of the nascent strands in both orientations?

Previously, work done in this laboratory (Darlow and Leach, unpublished) suggested the phenomenon of micro-folding in dsDNA. A plaque area assay was carried out on $(CAG{\cdot}CTG)_{1\text{-}3}$ cloned in the centre of a 462 bp palindrome in both orientations and no difference in the two orientations was observed (Darlow and

Leach, 1998). This finding of orientation-independent micro-folding led them to suggest that dsDNA can unfold and refold and micro-folding reactions are possible *in vivo* within double-stranded repeat sequences. Based on this work, they proposed a model for the growth of a quasi-hairpin arising from the small bulges arriving from one side of the double-strand (Figure 8.8). It has been shown that di- and trinucleotide repeated DNA could be synthesized starting from short complementary primers, based on polymerase slippage (Schlotterer and Tautz, 1992). It was concluded that after melting of 3'end from its template and reannealing, the little bulge behind could move along the nascent strand and come off at the other end. So, a hairpin might also grow in length by accumulating the ripples coming from one end (Harvey, 1997).

The hairpin growth model suggested by Darlow and Leach (unpublished) provides an explanation for the expansions seen in CAG·CTG repeats in a *rep* mutant. Based on this model, Figure 8.9 shows that in a *rep* mutant, a small bulge arising from the end of a nascent lagging strand can move towards the end of the repeat tract, creating a hairpin. Such creation of a hairpin is independent of the sequence of the repeat tract, as the free end in both orientations would have equal chances to create a bulge, eventually ending up in a hairpin that will generate expansion of the tract. The observations of bigger expansions in the CAG orientation suggests that bulges generated in this orientation might be able to travel faster than the bulges in the CTG orientation. Another suggestion is that the CTG hairpin once created might hinder the further movement of bulges. But it seems unlikely considering the model (Figure 8.8), where a bulge becomes part of an existing hairpin. So, it is proposed that CAG repeats give bigger expansions because the bulges generated can move faster or they are created faster in this orientation.

Since, SbcCD can cleave the hairpin structures (Connelly et al., 1999), a double mutant *rep sbcCD* was tested in the CTG orientation with the hypothesis that expansions would increase, if SbcCD happens to cleave the hairpins started by small bulges. No increase in expansions was observed in the double mutant suggesting that SbcCD does not cleave these hairpins either because of their size or conformation. Furthermore, SbcCD was also shown to have no role in DSB formation in *rep* mutant (Michel et al., 1997).

8.3.3 *uvrD* mutation destabilizes both CAG and CTG repeats

Both CAG and CTG repeats were destabilised in a *uvrD* mutant, which suggested a role of UvrD helicase in repeat stability. UvrD helicase acts at gaps created by the action of mismatch and nucleotide excision repair processes and has also been shown to be essential for fork reversal in Pol III replication mutants (Flores et al., 2004). So, UvrD could stabilize repeats by its activity during repair processes or by reversing the fork. The work carried out in this laboratory by John Blackwood suggests that RFR occurs with CTG repeats on the leading strand template but not in the opposite orientation i.e. CAG on the leading strand template. In Chapter 6, it was demonstrated that an "MMR-like" process might have a stabilizing role in an *sbcCD* mutant in the CAG orientation. The destabilization of CAG repeats in *uvrD* mutant might also be the same effect as of *mutS* mutant. It is proposed that *uvrD* mutant being deficient in "MMR-like" repair process gives license to MutL and MutH for the removal of MutS and cleavage of hairpins by SbcCD, hence leading to instability in the CAG orientation.

In this study, the increase in CTG repeat instability by *uvrD* mutation was found to be dependent on RecF (Figure 8.4). Michel and colleagues have shown that *uvrD* mutations enhance tandem repeat deletion via SOS induction of a RecF recombination pathway (Bierne et al., 1997) and RecF, RuvAB and RecQ are required for more than 90% of the deletion events. The authors propose that in *uvrD* mutants, the final gap fillings during mismatch and nucleotide excision repair pathways require an alternative helicase. They suggest that RecQ can functionally replace UvrD as the alternative helicase. SOS is induced in *uvrD* mutants, which would allow the production of an increased quantity of RecQ, which will initiate RecF mediated recombination. Whether the same events are involved in the CTG repeat instability was not tested in this study but this effect of *recF* in a *uvrD* mutant favours a role of UvrD in RFR for CTG repeat stability.

The observations of *uvrD recF* mutant are consistent with the proposal that CTG repeat instability is increased in a *uvrD* mutant because replication fork reversal cannot be carried out in this mutant. UvrD was shown to be essential for replication fork reversal in Polymerase III *dnaE*ts and *dnaN*ts mutants (Flores et al., 2004). It is speculated that in these mutants, when UvrD is not present, RecFOR can prevent the replication fork reversal by forming RecA filaments on the lagging strand template. UvrD can remove RecA from single-stranded DNA (Veaute et al., 2005) like its yeast orthologue Srs2 helicase (Krejci et al., 2003; Veaute et al., 2003). By removing the toxic RecA filaments form the fork, UvrD facilitates the fork reversal reaction (Flores et al., 2005). So, it is suggested here that in wild type cells, fork reversal may contribute to repeat stability. A mutation in *recF* allows the fork to be reversed, as there will be no blocking *recA* filaments loaded onto the lagging strand so fork could

be reversed without needing UvrD, giving a decrease in instability in the *uvrD recF* mutant.

Though the previous work done in this laboratory seems to suggest that RFR does not contribute in CAG repeat stability, it cannot be fully ruled out. It is speculated that some fork reversal might be happening in this orientation too. That would be clear by some further work, which is discussed in Chapter 9.

8.3.4 Interplay of Rep and UvrD helicases in CAG·CTG repeat instability

Both Rep and UvrD helicases share 40% amino acid homology but differ in their functions. Where much is known about UvrD, the role of Rep is still not well understood. The *in vivo* role of UvrD is to prevent homologous recombination as opposed to Rep helicase, which participates in replication (Veaute et al., 2005). Though our understanding about the role of these helicases is still lacking, their effects on repeat instability provide new understanding of mechanism of repeat instability. Where Rep utilizes its helicase function during replication, UvrD participates in both repair and replication to maintain the repeats stable.

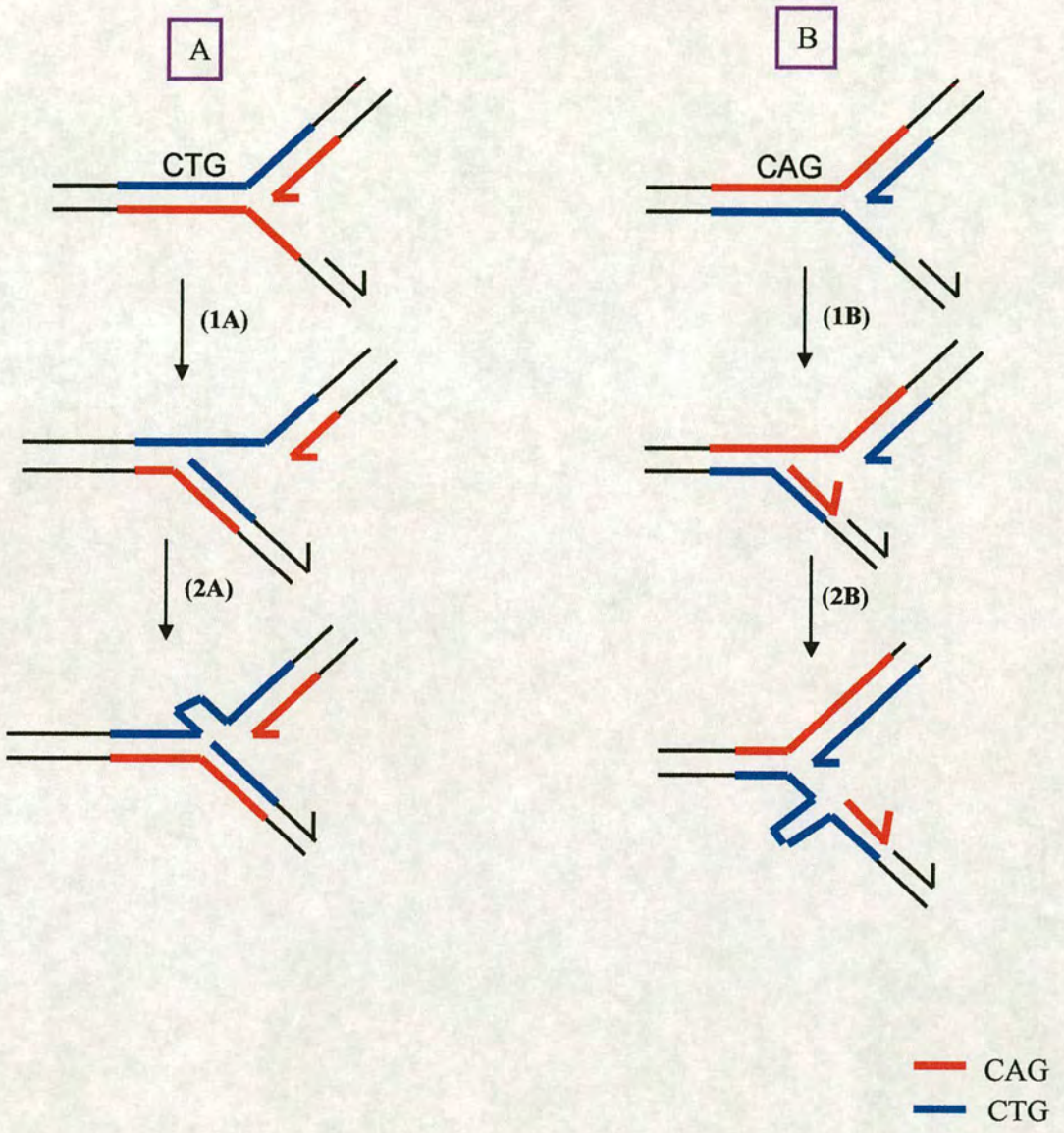


Figure 8.7: Dynamics of replication fork pausing at CAG·CTG repeats in a *rep* mutant

(A) Replication pauses at CTG repeats while lagging strand synthesis continues (1A). Absence of Rep helicase permits a leading strand gap to persist for longer, which can fold into a hairpin (2A) resulting in deletion event. **(B)** Replication may or may not pause at CAG repeats. If it pauses, the lack of Rep will create a leading strand gap (1B). The gap does not fold and replication progresses slowly through (2B). So the deletions happening in a CAG *rep* mutant are associated with structures being formed on lagging strand as in wild type.

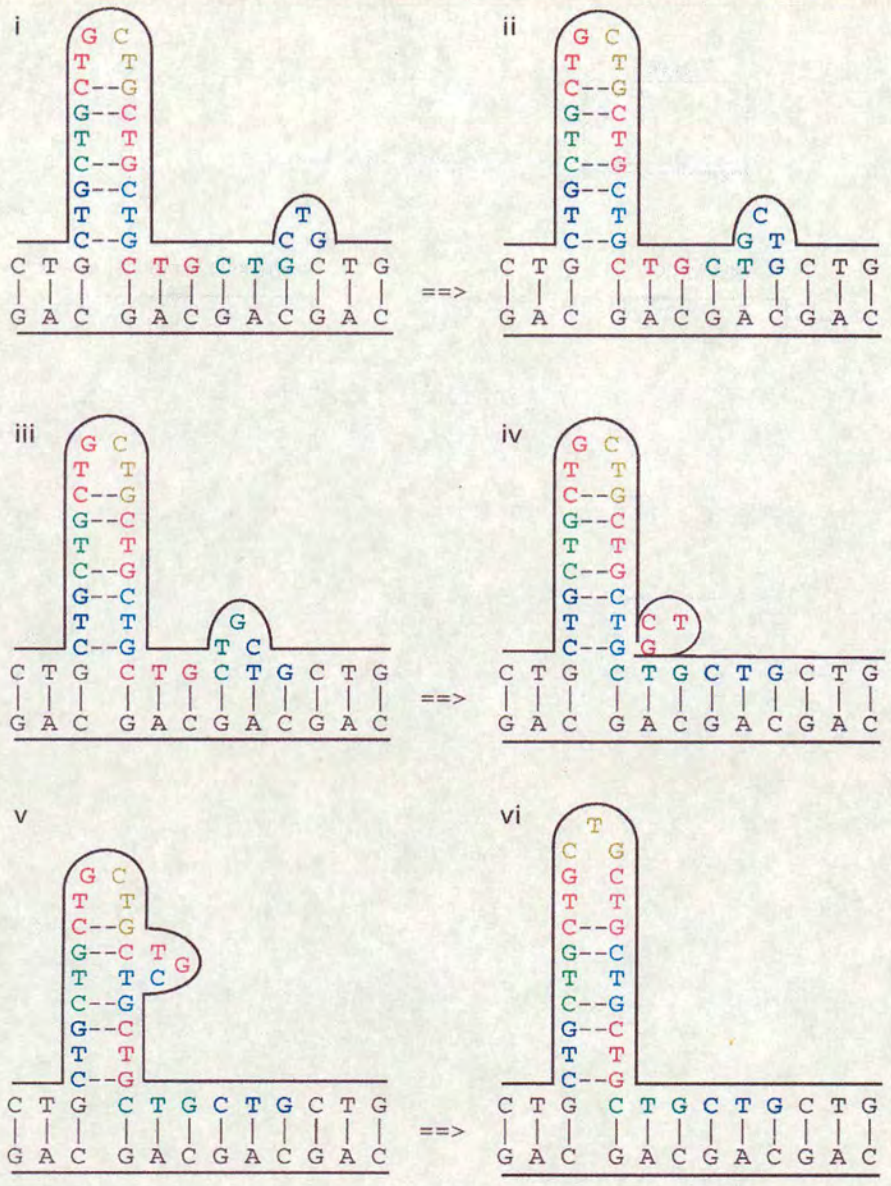


Figure 8.8: Possible mechanism for the formation of a quasi-hairpin of TNRs by accumulation of bulges (Darlow and Leach, Unpublished)
Local melting of a very short sequence within TNRs generates a small bulge (i), which travels on by breaking of one hydrogen bond behind and formation of next (ii-v). The bulge becomes part of the hairpin (if already present as shown in (i)) when it reaches its tip (vi) or stays till the end of repeat tract and a new bulge can start from the free end again. This movement of bulges would be iso-energetic, since the same number of bases remains unpaired throughout migration.

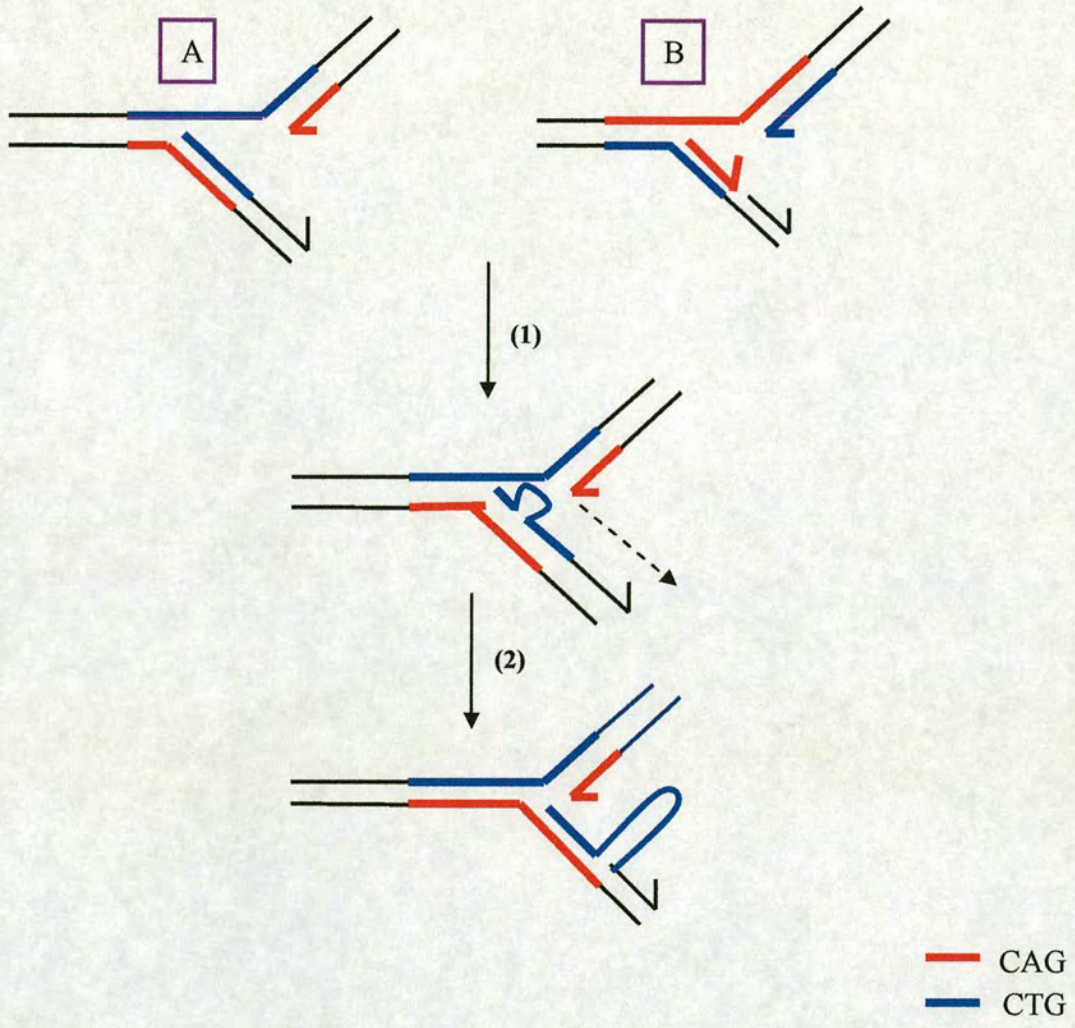


Figure 8.9: Model for expansions of CAG·CTG repeats in a *rep* mutant
A small bulge can arise at the free end the nascent lagging strand in both orientations – (A) CTG on the leading strand template, (B) CAG on the leading strand template (1). This bulge can travel along the tract until it reaches the end of the repeat tract, where it can end in a small hairpin structure (2). The continuous creation of bulges at the free end and travelling towards the end of repeat tract would give rise to a hairpin. In the figure, only CTG tract has been shown forming the hairpin. The dynamics would be the same for CAG repeat tract but it is suggested that the movement of bulges would be faster with CAG repeats.

CHAPTER 9

CONCLUDING REMARKS

9.1 Summary of work presented in this thesis

This work presents a novel investigation of CAG·CTG repeat instability in the *E. coli* chromosome. *E. coli* enjoys the reputation of a simple standard model because of the detailed understanding of its genetic and biochemical pathways. This study has revealed new paradigms of instability, which have previously been complicated by limitations of plasmid-based studies.

First, a library of uninterrupted CAG·CTG repeats was generated using a polymerisation independent strategy developed in this work. The repeat sequences of various lengths were integrated in the 5' end of the *lacZ* gene in the *E. coli* chromosome and instability was studied by measuring the lengths of repeat tracts using polyacrylamide electrophoresis based separation following PCR amplification.

CAG·CTG repeat instability increases with increasing repeat length and is orientation dependent with CAG repeats being more unstable when on the leading strand template. Further investigations suggest that the instability is mainly associated with replication and is provoked by the processing of slippage intermediates during replication.

SbcCD nuclease processes hairpin structures on the lagging strand template in the CAG orientation and orientation dependent instability observed in wild type cells is lost in DNA polymerase III proofreading (*dnaQ*) mutants. This suggests that

efficient proofreading of the lagging strand template determines orientation dependence of CAG·CTG repeat instability in the presence of SbcCD nuclease.

No significant role of recombination genes *recA*, *recB*, *recF*, *ruvAC*, *recG* and dimer resolution genes *xer* and *dif* is observed. The mismatch repair proteins play a role in the processing of intermediates during replication in the CAG orientation. It is suggested that MutS stabilizes CAG repeats by initiating a "repair" process and protecting the hairpins from cleavage by SbcCD. The presence of MutL and MutH cause MutS to cycle off the DNA, which allows SbcCD to access the hairpins. Furthermore, nucleotide excision repair process is not involved in CTG repeat instability as no effect of mutations in *uvrA*, *uvrB* and *uvrC* is observed.

Transcription does not influence repeat instability in wild type cells or any mutant studied in this work. A mutation in the *mfd* gene also does not affect the instability. CAG·CTG repeats influence the yield of β-galactosidase in an orientation dependent manner. β-galactosidase is formed but at reduced level when containing a CAG tract (poly-glutamine stretch). The presence of CTG repeats (poly-leucine stretch) does not produce any detectable β-galactosidase as observed by Western blotting.

A mutation in the gene encoding replicative helicase, Rep, destabilizes CTG repeats and *rep* mutants generate small expansions in both repeat orientations. The repair helicase, UvrD, affects instability in both orientations and RecF is essential for CTG repeat instability in the *uvrD* mutant. The results suggest the involvement of replication fork reversal reaction in maintaining repeat stability.

9.2 Advances made in this work

This study describes a novel system to study CAG·CTG repeat instability in the *E. coli* chromosome, which is a clear system with no complications associated with plasmid systems. This has revealed new and profound findings giving new directions to our classical ways of thinking about instability events.

The work demonstrates that the instability of these repeats is mainly associated with DNA synthesis, primarily caused by the processing of slippage intermediates. The absence of effects of transcription and recombination further favours the instability to be of replicative nature.

Both in bacteria and yeast, the instability has been documented to be an orientation dependent phenomenon (Freudenreich et al., 1997; Kang et al., 1995a; Maurer et al., 1996; Miret et al., 1998) but the molecular mechanism of this dependence was never investigated. So far, the bulk of the literature had relied on the traditional replication slippage model (discussed in Chapter 1), which actually does not explain the instability bias observed in these model systems. This study, for the first time, has provided an understanding of the molecular mechanism of this orientation dependence, which is a significant contribution in our endeavours to understand the mechanism(s) of instability. The work demonstrates that this orientation dependent instability is a consequence of the processing of slippage intermediates during replication or repair associated DNA synthesis and is interplay of proofreading (DnaQ), hairpin nuclease (SbcCD), mismatch repair proteins and helicases. The CTG orientation is stable because it is efficiently proofread and the fork reversal reaction seems to operate in this orientation to maintain repeats stable. In contrast, proofreading is inefficient in the CAG orientation, which also faces another

route of deletion mediated by MutL and MutH proteins, involving SbcCD nuclease. Furthermore, replication fork reversal does not seem to happen in this orientation.

This study also demonstrates that bigger expansions occur when CAG repeats are on the leading strand template that again brings the replication slippage model in question. According to the traditional model, expansions are likely to happen when CTG repeats are on the nascent lagging strand. The observations made in this work that expansions happen in both orientations and are bigger when CAG repeats are on the nascent lagging strand requires an alternative explanation. This study suggests a model for expansions, which is sequence independent and attempts to explain the events that might be generating expansions in both orientations.

The deletion length distributions observed imply that secondary structures are intermediates in instability pathway(s) and their processing is influenced by their differential stabilities, which eventually determine the orientation dependence. The CAG orientation would have stable structures on the lagging strand template, which would persist for long as compared to unstable structures in the CTG orientation.

This work also raises the possibility of a "mismatch repair" process, which differs from the classical one. This seems to be occurring during DNA synthesis and is being promoted by the binding of MutS to the hairpin structure.

Furthermore, this work highlights the role of helicases influencing repeat instability. It suggests the existence of a new class of instability events that might be generated because of the action of helicases on the fork. This further brings replication into light and strengthens the stance that the CAG·CTG repeat instability in the *E. coli* chromosome revolves around replication.

9.2 Future directions

This study of CAG·CTG repeat instability in the *E. coli* chromosome has addressed many areas but still many aspects remain uncovered. The study leaves several questions that can be looked at in future.

In this work a polymerisation independent strategy was developed to build long repeat sequences in a plasmid that can eventually be used to integrate sequences in the *E. coli* chromosome. The method was used mainly to build CAG·CTG repeats, which were the focus of this study but the construction of other tri, tetra and pentanucleotide sequences was also initiated. A TNR sequence CCG·CGG was built to a length of 8, a pentanucleotide, ATTCT·AGAAT involved in SCA10 was successfully introduced in the plasmid while an attempt was also made for the tetranucleotide, CCTG·CAGG involved in myotonic dystrophy type 2. So, this method can essentially be applied to build any repeat sequence. Various lengths of sequence of choice can be built and their behaviour can be studied in this chromosomal system.

In this work, a CTG repeat containing strain was found to have the potential to expand and expanded lengths of up to 184 were collected, which could not be studied because of the detection limit of the size standard used in GeneMapper® analysis. A GeneScan™-Rox™-2500 Rox standard, advised by ABI customer support, was tried but was found to be incompatible with the settings of ABI genetic analyser. Designing a size standard with bigger fragments labelled with different dyes might be helpful in the analysis of these longer lengths. A study of these longer repeat lengths might give more expanded lengths and reveal alternative pathways, which may involve recombination responsible for instability at longer lengths.

In the past, small slippages of plasmid born TNRs have been observed in MMR mutants (Schmidt et al., 2000). In this work, most of the deletions observed are of large sizes. This raises the question whether small slippages happen in the chromosome and if they happen, is there any other route to correct them? One candidate is the proofreading function of the DNA polymerase. Investigating a double mutant of MMR and proofreading may provide an answer to this.

In this study, the role of the NER pathway was analysed only in the CTG orientation so an analysis of this pathway in the CAG orientation would be helpful in understanding the role of this repair process in instability. It would be of interest since the CAG orientation shows a predominance of larger deletions, consistent with the formation of large loops and NER might be involved in processing of these large loops. Furthermore, a complete analysis of both orientations in a chromosomal system, free from the artefacts of plasmid systems, would clarify the conflicts in the literature.

Two helicases, Rep and UvrD, significantly affected instability and RecF was found to be essential for instability in the *uvrD* mutant as shown previously for tandem repeats (Bierne et al., 1997). They attributed this effect of RecF to the overproduction of RecQ in *uvrD* mutants. Investigating this effect of RecF in a *uvrD* mutant containing CAG repeats would be interesting to compare both orientations and would also allow us to have a better understanding of whether RFR happens in this orientation or not. Furthermore, studying the role of RecQ helicase in instability would explore whether this third helicase is involved in CAG·CTG repeat instability and whether the instability pathway(s) involve the same events observed by Bierne et al. (1997).

In this work, the role of *recG* was investigated only in the CTG orientation so investigating this helicase in the CAG orientation would be interesting, considering the involvement of helicases in repeat instability observed in this study. Similarly, studying the effect of *ruvAC* mutation in the CAG orientation would give us a clear understanding of the events involved in both repeat orientations.

Furthermore, this study demonstrated that *lacZ* containing CTG repeats encodes a non-functional or non-existent β-galactosidase as observed from the absence of protein on Western blot and a white phenotype on media containing X-gal and IPTG. So it would be interesting to investigate why is the protein not made or quickly degraded.

9.3 Outlook on TNRs research

Research in the field of TNRs over the past fifteen years has provided significant understanding of the molecular mechanisms involved in genetic instability but unravelling the mysteries of repeat instability has been complicated. Where many findings of the involvement of repair, replication, transcription and repair-mediated events have been reported, the inter-relationship of these processes is yet to be established.

A wide range of model systems, with their strengths and limitations has been unable to recapitulate the instability seen in humans, which leave us with the question of whether there is a suitable model system for this repeat expansion phenomenon, predicting that this is a human specific process. Though *E. coli* is a well-defined model, its limitations are that its cellular processes have not evolved to manage long trinucleotide repeat sequences, which tend to delete in this organism. Despite its

shortcomings, it has been employed in studying trinucleotide repeat instability. But, previously this simple model was not used efficiently because of the lack of a controlled chromosomal system. The system developed in this work is better controlled and it is hoped that findings will be more easily correlated with systems in eukaryotes.

This study raises a possibility that proofreading during DNA synthesis might contribute to repeat stability in humans. *E. coli* DnaQ shares sequence homology with the human DNA editing enzyme DNase III/TREX1 (Hoss et al., 1999). Mutations in the gene encoding TREX1 cause human Aicardi-Goutieres Syndrome, which is a genetically determined encephalopathy at the *AGS1* locus (Crow et al., 2006). The enzyme is present in equal amounts in proliferating and non-proliferating cells, suggesting that its editing function is involved in both replication and repair associated DNA synthesis. Therefore, it is plausible that efficient proofreading during replication and repair in human cells may contribute to repeat stability. Though Aicardi-Goutieres Syndrome is a recessive disorder and does not involve a repeat expansion process, it manifests brain atrophy and loss of white matter and it is plausible that repeat expansion disease process may involve the same defects of proofreading leading to neurodegeneration.

Since the discovery of this dynamic mutation process, all investigators have agreed on the formation of non B-DNA structures *in vitro* and their involvement in instability has been seriously debated. This study advances the evidence for the formation of hairpins *in vivo* as evident from the deletion size distributions observed. This further opens ways into how these mutagenic intermediates might be formed (discussed in next paragraph). However, we still need to know about their formation *in vivo* in eukaryotic cells.

The complex pattern of repeat instability observed between different disease loci and between tissues of the same patient suggests that instability might not be determined by the location or direction of replication origin as per our conventional thoughts. A fork-shift model (discussed in Chapter 1) proposed by Cleary and Pearson, (2005) suggests that the *cis*-elements within or flanking the repeat tracts may alter the dynamics of the advancing replication fork to produce instability. Although this proposal is a good attempt to explain the numerous differences among different TREDs and different tissues, it takes into account only *cis*-elements, the sequence of the repeat tract or the portion of the sequence in the Okazaki initiation zone ignoring the possibility of *trans*-acting factors that might influence this fork shift model. This work possibly explains such dynamics of the fork bringing in *cis*- and *trans*-factors. Since the instability is replication oriented, it is suggested that most of the instability events would be happening at the fork. It takes into account the action of helicases, which might determine both the formation of a mutagenic intermediate and the portion of repeats going into the Okazaki initiation zone. The observations in this study favour these possibilities. The absence of Rep and UvrD helicase elevates CTG repeat instability suggesting that helicase activity is involved in maintaining stable repeats. RecG also changes the sizes of deletions suggesting its involvement in determining the formation of instability intermediates. The proposal is that the helicase action avoids the formation of mutagenic intermediates in this orientation. In yeast, Srs 2 helicase (orthologue of UvrD) has been shown to block expansions *in vivo* (Bhattacharyya and Lahue, 2004) and it also selectively unwinds triplet repeat DNA (Bhattacharyya and Lahue, 2005). So, integrating the role of helicases along with the Rad50/Mre11 complex might provide an answer to one of the outstanding

questions in TNRs research. How are the mutagenic intermediates formed and how are they subsequently handled in the light of fork dynamics during replication?

A long-term goal in the field of repeat instability is to come up with clinical benefits but there are many steps to cross before reaching clinical targets. Aspects of mismatch repair appears to be required for expansions in mouse model systems (Manley et al., 1999; Savouret et al., 2003) so determining their role in proliferative and non-proliferative tissues would be an interesting advance. Locus specific factors that influence instability need to be determined. The tissue specific and germline instability with the pattern and timing of instability remains to be known. Since, each repeat sequence and disease locus exhibits different patterns of instability in humans and in model system, it will take a significant amount of time and research to reveal the mechanism(s) responsible for instability.

Despite our failings and limitations, the field is gaining interest due to its clinical implications and we should be optimistic for fruitful advances.

BIBLIOGRAPHY

Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P. and Griffith, J. D. (1997). MutS mediates heteroduplex loop formation by a translocation mechanism. *Embo J* **16**, 4467-76.

Anderson, D. G. and Kowalczykowski, S. C. (1997). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77-86.

Balakumaran, B. S., Freudenreich, C. H. and Zakian, V. A. (2000). CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in Saccharomyces cerevisiae. *Hum Mol Genet* **9**, 93-100.

Bayliss, C. D., Dixon, K. M. and Moxon, E. R. (2004). Simple sequence repeats (microsatellites): mutational mechanisms and contributions to bacterial pathogenesis. A meeting review. *FEMS Immunol Med Microbiol* **40**, 11-9.

Bayliss, C. D., Field, D. and Moxon, E. R. (2001). The simple sequence contingency loci of Haemophilus influenzae and Neisseria meningitidis. *J Clin Invest* **107**, 657-62.

Bhattacharyya, S. and Lahue, R. S. (2004). Saccharomyces cerevisiae Srs2 DNA helicase selectively blocks expansions of trinucleotide repeats. *Mol Cell Biol* **24**, 7324-30.

Bhattacharyya, S. and Lahue, R. S. (2005). Srs2 helicase of Saccharomyces cerevisiae selectively unwinds triplet repeat DNA. *J Biol Chem* **280**, 33311-7.

Bichara, M., Pinet, I., Schumacher, S. and Fuchs, R. P. (2000). Mechanisms of dinucleotide repeat instability in Escherichia coli. *Genetics* **154**, 533-42.

Bierne, H., Ehrlich, S. D. and Michel, B. (1995). Competition between parental and recombinant plasmids affects the measure of recombination frequencies. *Plasmid* **33**, 101-12.

Bierne, H., Seigneur, M., Ehrlich, S. D. and Michel, B. (1997). uvrD mutations enhance tandem repeat deletion in the Escherichia coli chromosome via SOS induction of the RecF recombination pathway. *Mol Microbiol* **26**, 557-67.

Blakely, G., Colloms, S., May, G., Burke, M. and Sherratt, D. (1991). Escherichia coli XerC recombinase is required for chromosomal segregation at cell division. *New Biol* **3**, 789-98.

Bowater, R. P., Jaworski, A., Larson, J. E., Parniewski, P. and Wells, R. D. (1997). Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in Escherichia coli. *Nucleic Acids Res* **25**, 2861-8.

Bowater, R. P., Rosche, W. A., Jaworski, A., Sinden, R. R. and Wells, R. D. (1996). Relationship between Escherichia coli growth and deletions of CTG.CAG triplet repeats in plasmids. *J Mol Biol* **264**, 82-96.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-7.

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T. et al. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**, 799-808.

Bruford, M. W. and Wayne, R. K. (1993). Microsatellites and their application to population genetic studies. *Curr Opin Genet Dev* **3**, 939-43.

Brunham, R. C., Plummer, F. A. and Stephens, R. S. (1993). Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun* **61**, 2273-6.

Bzymek, M. and Lovett, S. T. (2001). Evidence for two mechanisms of palindrome-stimulated deletion in Escherichia coli: single-strand annealing and replication slipped mispairing. *Genetics* **158**, 527-40.

Bzymek, M., Saveson, C. J., Feschenko, V. V. and Lovett, S. T. (1999). Slipped misalignment mechanisms of deletion formation: in vivo susceptibility to nucleases. *J Bacteriol* **181**, 477-82.

Campuzano, V., Montermini, L., Lutz, Y., Cova, L., Hindelang, C., Jiralerspong, S., Trottier, Y., Kish, S. J., Faucheux, B., Trouillas, P. et al. (1997). Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum Mol Genet* **6**, 1771-80.

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A. et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**, 1423-7.

Chong, S. S., McCall, A. E., Cota, J., Subramony, S. H., Orr, H. T., Hughes, M. R. and Zoghbi, H. Y. (1995). Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* **10**, 344-50.

Choong, C. S., Kemppainen, J. A., Zhou, Z. X. and Wilson, E. M. (1996). Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Mol Endocrinol* **10**, 1527-35.

Cleary, J. D., Nichol, K., Wang, Y. H. and Pearson, C. E. (2002). Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat Genet* **31**, 37-46.

Cleary, J. D. and Pearson, C. E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res* **100**, 25-55.

Cleary, J. D. and Pearson, C. E. (2005). Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet* **21**, 272-80.

Connelly, J. C., de Leau, E. S. and Leach, D. R. (1999). DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. *Nucleic Acids Res* **27**, 1039-46.

Connelly, J. C., de Leau, E. S., Okely, E. A. and Leach, D. R. (1997). Overexpression, purification, and characterization of the SbcCD protein from Escherichia coli. *J Biol Chem* **272**, 19819-26.

Connelly, J. C., Kirkham, L. A. and Leach, D. R. (1998). The SbcCD nuclease of Escherichia coli is a structural maintenance of chromosomes (SMC) family protein that cleaves hairpin DNA. *Proc Natl Acad Sci U S A* **95**, 7969-74.

Connelly, J. C. and Leach, D. R. (1996). The sbcC and sbcD genes of Escherichia coli encode a nuclease involved in palindrome inviability and genetic recombination. *Genes Cells* **1**, 285-91.

Crow, Y. J., Hayward, B. E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D. N., van Bokhoven, H., Brunner, H. G., Hamel, B. C. et al. (2006). Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutieres syndrome at the AGS1 locus. *Nat Genet* **38**, 917-20.

Cummings, C. J. and Zoghbi, H. Y. (2000). Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* **9**, 909-16.

Darlow, J. M. and Leach, D. R. (1998). Evidence for two preferred hairpin folding patterns in d(CGG).d(CCG) repeat tracts in vivo. *J Mol Biol* **275**, 17-23.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E. et al. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-4.

Dixon, D. A. and Kowalczykowski, S. C. (1993). The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell* **73**, 87-96.

Donaldson, J. R., Courcelle, C. T. and Courcelle, J. (2004). RuvAB and RecG are not essential for the recovery of DNA synthesis following UV-induced DNA damage in Escherichia coli. *Genetics* **166**, 1631-40.

Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A* **88**, 7160-4.

Eberhart, D. E. and Warren, S. T. (1996). Nuclease sensitivity of permeabilized cells confirms altered chromatin formation at the fragile X locus. *Somat Cell Mol Genet* **22**, 435-41.

Flores, M. J., Bidnenko, V. and Michel, B. (2004). The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. *EMBO Rep* **5**, 983-8.

Flores, M. J., Bierne, H., Ehrlich, S. D. and Michel, B. (2001). Impairment of lagging strand synthesis triggers the formation of a RuvABC substrate at replication forks. *Embo J* **20**, 619-29.

Flores, M. J., Sanchez, N. and Michel, B. (2005). A fork-clearing role for UvrD. *Mol Microbiol* **57**, 1664-75.

Fojtik, P. and Vorlickova, M. (2001). The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res* **29**, 4684-90.

French, S. (1992). Consequences of replication fork movement through transcription units in vivo. *Science* **258**, 1362-5.

Freudenreich, C. H., Kantrow, S. M. and Zakian, V. A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853-6.

Freudenreich, C. H., Stavenhagen, J. B. and Zakian, V. A. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol* **17**, 2090-8.

Fried, M. G. and Crothers, D. M. (1983). CAP and RNA polymerase interactions with the lac promoter: binding stoichiometry and long range effects. *Nucleic Acids Res* **11**, 141-58.

Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., Jr., Warren, S. T. et al. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047-58.

Gacy, A. M., Goellner, G., Juranic, N., Macura, S. and McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533-40.

Gacy, A. M., Goellner, G. M., Spiro, C., Chen, X., Gupta, G., Bradbury, E. M., Dyer, R. B., Mikesell, M. J., Yao, J. Z., Johnson, A. J. et al. (1998). GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell* **1**, 583-93.

Gacy, A. M. and McMurray, C. T. (1998). Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry* **37**, 9426-34.

Geuskens, V., Vogel, J. L., Grimaud, R., Desmet, L., Higgins, N. P. and Toussaint, A. (1991). Frameshift mutations in the bacteriophage Mu repressor gene can confer a trans-dominant virulent phenotype to the phage. *J Bacteriol* **173**, 6578-85.

Grabczyk, E. and Usdin, K. (1999). Generation of microgram quantities of trinucleotide repeat tracts of defined length, interspersion pattern, and orientation. *Anal Biochem* **267**, 241-3.

Grilley, M., Welsh, K. M., Su, S. S. and Modrich, P. (1989). Isolation and characterization of the Escherichia coli mutL gene product. *J Biol Chem* **264**, 1000-4.

Grompone, G., Ehrlich, D. and Michel, B. (2004). Cells defective for replication restart undergo replication fork reversal. *EMBO Rep* **5**, 607-12.

Grompone, G., Seigneur, M., Ehrlich, S. D. and Michel, B. (2002). Replication fork reversal in DNA polymerase III mutants of Escherichia coli: a role for the beta clamp. *Mol Microbiol* **44**, 1331-9.

Hanvey, J. C., Shimizu, M. and Wells, R. D. (1988). Intramolecular DNA triplexes in supercoiled plasmids. *Proc Natl Acad Sci U S A* **85**, 6292-6.

Harrington, J. J. and Lieber, M. R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. *Embo J* **13**, 1235-46.

Harvey, S. C. (1997). Slipped structures in DNA triplet repeat sequences: entropic contributions to genetic instabilities. *Biochemistry* **36**, 3047-9.

Hashem, V. I., Klysik, E. A., Rosche, W. A. and Sinden, R. R. (2002). Instability of repeated DNAs during transformation in Escherichia coli. *Mutat Res* **502**, 39-46.

Hashem, V. I., Rosche, W. A. and Sinden, R. R. (2004). Genetic recombination destabilizes (CTG)n.(CAG)n repeats in E. coli. *Mutat Res* **554**, 95-109.

Hebert, M. L., Spitz, L. A. and Wells, R. D. (2004). DNA double-strand breaks induce deletion of CTG.CAG repeats in an orientation-dependent manner in Escherichia coli. *J Mol Biol* **336**, 655-72.

Heller, R. C. and Marians, K. J. (2005). Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks. *J Biol Chem* **280**, 34143-51.

Holmes, S. E., O'Hearn, E. E., McInnis, M. G., Gorelick-Feldman, D. A., Kleiderlein, J. J., Callahan, C., Kwak, N. G., Ingersoll-Ashworth, R. G., Sherr, M., Sumner, A. J. et al. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat Genet* **23**, 391-2.

Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C. and Moxon, E. R. (1996). DNA repeats identify novel virulence genes in Haemophilus influenzae. *Proc Natl Acad Sci U S A* **93**, 11121-5.

Hoss, M., Robins, P., Naven, T. J., Pappin, D. J., Sgouros, J. and Lindahl, T. (1999). A human DNA editing enzyme homologous to the Escherichia coli DnaQ/MutD protein. *Embo J* **18**, 3868-75.

Iyer, R. R., Pluciennik, A., Rosche, W. A., Sinden, R. R. and Wells, R. D. (2000). DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in Escherichia coli. *J Biol Chem* **275**, 2174-84.

Iyer, R. R. and Wells, R. D. (1999). Expansion and deletion of triplet repeat sequences in Escherichia coli occur on the leading strand of DNA replication. *J Biol Chem* **274**, 3865-77.

Jakupciak, J. P. and Wells, R. D. (1999). Genetic instabilities in (CTG.CAG) repeats occur by recombination. *J Biol Chem* **274**, 23468-79.

Jakupciak, J. P. and Wells, R. D. (2000a). Gene conversion (recombination) mediates expansions of CTG[middle dot]CAG repeats. *J Biol Chem* **275**, 40003-13.

Jakupciak, J. P. and Wells, R. D. (2000b). Genetic instabilities of triplet repeat sequences by recombination. *IUBMB Life* **50**, 355-9.

Jaworski, A., Rosche, W. A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R. P., Sinden, R. R. and Wells, R. D. (1995). Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. *Proc Natl Acad Sci U S A* **92**, 11019-23.

Jeffreys, A. J., Allen, M. J., Hagelberg, E. and Sonnberg, A. (1992). Identification of the skeletal remains of Josef Mengele by DNA analysis. *Forensic Sci Int* **56**, 65-76.

Johzuka, K. and Ogawa, H. (1995). Interaction of Mre11 and Rad50: two proteins required for DNA repair and meiosis-specific double-strand break formation in Saccharomyces cerevisiae. *Genetics* **139**, 1521-32.

Kang, S., Jaworski, A., Ohshima, K. and Wells, R. D. (1995a). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet* **10**, 213-8.

Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S. and Wells, R. D. (1995b). Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J Biol Chem* **270**, 27014-21.

Kelman, Z. and O'Donnell, M. (1995). DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem* **64**, 171-200.

Kim, S., Dallmann, H. G., McHenry, C. S. and Marians, K. J. (1996). Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. *Cell* **84**, 643-50.

Kim, S. H., Cai, L., Pytlos, M. J., Edwards, S. F. and Sinden, R. R. (2005). Generation of long tracts of disease-associated DNA repeats. *Biotechniques* **38**, 247-53.

Kim, S. H., Pytlos, M. J., Rosche, W. A. and Sinden, R. R. (2006a). (CAG)*(CTG) repeats associated with neurodegenerative diseases are stable in the Escherichia coli chromosome. *J Biol Chem* **281**, 27950-5.

Kim, S. H., Pytlos, M. J. and Sinden, R. R. (2006b). Replication restart: A pathway for (CTG).(CAG) repeat deletion in Escherichia coli. *Mutat Res*.

Klesert, T. R., Otten, A. D., Bird, T. D. and Tapscott, S. J. (1997). Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat Genet* **16**, 402-6.

Knight, S. J., Flannery, A. V., Hirst, M. C., Campbell, L., Christodoulou, Z., Phelps, S. R., Pointon, J., Middleton-Price, H. R., Barnicoat, A., Pembrey, M. E. et al. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127-34.

Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W. and Ranum, L. P. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet* **21**, 379-84.

Kovtun, I. V., Goellner, G. and McMurray, C. T. (2001). Structural features of trinucleotide repeats associated with DNA expansion. *Biochem Cell Biol* **79**, 325-36.

Kovtun, I. V. and McMurray, C. T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet* **27**, 407-11.

Kowalczykowski, S. C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* **25**, 156-65.

Krasilnikova, M. M. and Mirkin, S. M. (2004). Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Mol Cell Biol* **24**, 2286-95.

Krejci, L., Van Komen, S., Li, Y., Villemain, J., Reddy, M. S., Klein, H., Ellenberger, T. and Sung, P. (2003). DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* **423**, 305-9.

Kuempel, P. L., Henson, J. M., Dircks, L., Tecklenburg, M. and Lim, D. F. (1991). dif, a recA-independent recombination site in the terminus region of the chromosome of Escherichia coli. *New Biol* **3**, 799-811.

Lahue, R. S. and Slater, D. L. (2003). DNA repair and trinucleotide repeat instability. *Front Biosci* **8**, s653-65.

Lane, H. E. and Denhardt, D. T. (1974). The rep mutation. III. Altered structure of the replicating Escherichia coli chromosome. *J Bacteriol* **120**, 805-14.

Lane, H. E. and Denhardt, D. T. (1975). The rep mutation. IV. Slower movement of replication forks in Escherichia coli rep strains. *J Mol Biol* **97**, 99-112.

Langston, L. D. and O'Donnell, M. (2006). DNA replication: keep moving and don't mind the gap. *Mol Cell* **23**, 155-60.

Lenzmeier, B. A. and Freudenreich, C. H. (2003). Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res* **100**, 7-24.

Lin, J. J., Phillips, A. M., Hearst, J. E. and Sancar, A. (1992). Active site of (A)BC excinuclease. II. Binding, bending, and catalysis mutants of UvrB reveal a direct role in 3' and an indirect role in 5' incision. *J Biol Chem* **267**, 17693-700.

Lin, Y., Dion, V. and Wilson, J. H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol* **13**, 179-80.

Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W. and Ranum, L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293**, 864-7.

Liu, B. and Alberts, B. M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* **267**, 1131-7.

Liu, Y. and Bambara, R. A. (2003). Analysis of human flap endonuclease 1 mutants reveals a mechanism to prevent triplet repeat expansion. *J Biol Chem* **278**, 13728-39.

Lovett, S. T., Drapkin, P. T., Sutera, V. A., Jr. and Gluckman-Peskind, T. J. (1993). A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in Escherichia coli. *Genetics* **135**, 631-42.

Mahdi, A. A., Buckman, C., Harris, L. and Lloyd, R. G. (2006). Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. *Genes Dev* **20**, 2135-47.

Manley, K., Shirley, T. L., Flaherty, L. and Messer, A. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* **23**, 471-3.

Marians, K. J. (1992). Prokaryotic DNA replication. *Annu Rev Biochem* **61**, 673-719.

Mariappan, S. V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R. K., Bradbury, E. M. and Gupta, G. (1996). Solution structures of the individual single strands of the fragile X DNA triplets (GCC)n.(GGC)n. *Nucleic Acids Res* **24**, 784-92.

Matson, S. W., Bean, D. W. and George, J. W. (1994). DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* **16**, 13-22.

Maurer, D. J., O'Callaghan, B. L. and Livingston, D. M. (1996). Orientation dependence of trinucleotide CAG repeat instability in Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 6617-22.

McGlynn, P. and Lloyd, R. G. (2000). Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. *Cell* **101**, 35-45.

Meile, J. C., Wu, L. J., Ehrlich, S. D., Errington, J. and Noirot, P. (2006). Systematic localisation of proteins fused to the green fluorescent protein in Bacillus subtilis: identification of new proteins at the DNA replication factory. *Proteomics* **6**, 2135-46.

Mellon, I. and Hanawalt, P. C. (1989). Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* **342**, 95-8.

Merlin, C., McAteer, S. and Masters, M. (2002). Tools for characterization of Escherichia coli genes of unknown function. *J Bacteriol* **184**, 4573-81.

Michel, B. (2000). Replication fork arrest and DNA recombination. *Trends Biochem Sci* **25**, 173-8.

Michel, B., Ehrlich, S. D. and Uzest, M. (1997). DNA double-strand breaks caused by replication arrest. *Embo J* **16**, 430-8.

Michel, B., Grompone, G., Flores, M. J. and Bidnenko, V. (2004). Multiple pathways process stalled replication forks. *Proc Natl Acad Sci U S A* **101**, 12783-8.

Miret, J. J., Pessoa-Brandao, L. and Lahue, R. S. (1997). Instability of CAG and CTG trinucleotide repeats in Saccharomyces cerevisiae. *Mol Cell Biol* **17**, 3382-7.

Miret, J. J., Pessoa-Brandao, L. and Lahue, R. S. (1998). Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **95**, 12438-43.

Mitas, M. (1997). Trinucleotide repeats associated with human disease. *Nucleic Acids Res* **25**, 2245-54.

Mitas, M., Yu, A., Dill, J. and Haworth, I. S. (1995a). The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn.Ganti base pairs. *Biochemistry* **34**, 12803-11.

Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J. and Haworth, I. S. (1995b). Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res* **23**, 1050-9.

Mochmann, L. H. and Wells, R. D. (2004). Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC*GTC repeats. *Nucleic Acids Res* **32**, 4469-79.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* **25**, 229-53.

Modrich, P. (1994). Mismatch repair, genetic stability, and cancer. *Science* **266**, 1959-60.

Moxon, E. R., Rainey, P. B., Nowak, M. A. and Lenski, R. E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**, 24-33.

Napierala, M., Parniewski, P., Pluciennik, A. and Wells, R. D. (2002). Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination. *J Biol Chem* **277**, 34087-100.

Neidle, S. and Parkinson, G. N. (2003). The structure of telomeric DNA. *Curr Opin Struct Biol* **13**, 275-83.

Nichol Edamura, K., Leonard, M. R. and Pearson, C. E. (2005). Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am J Hum Genet* **76**, 302-11.

Nichol, K. and Pearson, C. E. (2002). CpG methylation modifies the genetic stability of cloned repeat sequences. *Genome Res* **12**, 1246-56.

Ohshima, K. and Wells, R. D. (1997). Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. *J Biol Chem* **272**, 16798-806.

Ordway, J. M. and Detloff, P. J. (1996). In vitro synthesis and cloning of long CAG repeats. *Biotechniques* **21**, 609-10, 612.

Orren, D. K. and Sancar, A. (1989). The (A)BC excinuclease of Escherichia coli has only the UvrB and UvrC subunits in the incision complex. *Proc Natl Acad Sci U S A* **86**, 5237-41.

Orren, D. K., Selby, C. P., Hearst, J. E. and Sancar, A. (1992). Post-incision steps of nucleotide excision repair in Escherichia coli. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. *J Biol Chem* **267**, 780-8.

Ossanna, N. and Mount, D. W. (1989). Mutations in uvrD induce the SOS response in Escherichia coli. *J Bacteriol* **171**, 303-7.

Oussatcheva, E. A., Hashem, V. I., Zou, Y., Sinden, R. R. and Potaman, V. N. (2001). Involvement of the nucleotide excision repair protein UvrA in instability of CAG*CTG repeat sequences in Escherichia coli. *J Biol Chem* **276**, 30878-84.

Park, J. S., Marr, M. T. and Roberts, J. W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* **109**, 757-67.

Parker, B. O. and Marinus, M. G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in Escherichia coli. *Proc Natl Acad Sci U S A* **89**, 1730-4.

Parniewski, P., Bacolla, A., Jaworski, A. and Wells, R. D. (1999). Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli. *Nucleic Acids Res* **27**, 616-23.

Parniewski, P., Jaworski, A., Wells, R. D. and Bowater, R. P. (2000). Length of CTG.CAG repeats determines the influence of mismatch repair on genetic instability. *J Mol Biol* **299**, 865-74.

Pearson, C. E., Nichol Edamura, K. and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet* **6**, 729-42.

Pearson, C. E. and Sinden, R. R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35**, 5041-53.

Pearson, C. E., Wang, Y. H., Griffith, J. D. and Sinden, R. R. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n. (CAG)n repeats from the myotonic dystrophy locus. *Nucleic Acids Res* **26**, 816-23.

Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R. and Mirkin, S. M. (2003). Replication and expansion of trinucleotide repeats in yeast. *Mol Cell Biol* **23**, 1349-57.

Petruska, J., Arnheim, N. and Goodman, M. F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res* **24**, 1992-8.

Pluciennik, A., Iyer, R. R., Napierala, M., Larson, J. E., Filutowicz, M. and Wells, R. D. (2002). Long CTG.CAG repeats from myotonic dystrophy are preferred sites for intermolecular recombination. *J Biol Chem* **277**, 34074-86.

Potaman, V. N., Oussatcheva, E. A., Lyubchenko, Y. L., Shlyakhtenko, L. S., Bidichandani, S. I., Ashizawa, T. and Sinden, R. R. (2004). Length-dependent structure formation in Friedreich ataxia (GAA)n*(TTC)n repeats at neutral pH. *Nucleic Acids Res* **32**, 1224-31.

Register, J. C., 3rd and Griffith, J. (1985). The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. *J Biol Chem* **260**, 12308-12.

Richard, G. F., Goellner, G. M., McMurray, C. T. and Haber, J. E. (2000). Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. *Embo J* **19**, 2381-90.

Richards, R. I. (2001). Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* **10**, 2187-94.

Richards, R. I., Holman, K., Friend, K., Kremer, E., Hillen, D., Staples, A., Brown, W. T., Goonewardena, P., Tarleton, J., Schwartz, C. et al. (1992). Evidence of founder chromosomes in fragile X syndrome. *Nat Genet* **1**, 257-60.

Richards, R. I. and Sutherland, G. R. (1992). Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**, 709-12.

Rolfsmeier, M. L., Dixon, M. J. and Lahue, R. S. (2000). Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Mol Cell* **6**, 1501-7.

Rolfsmeier, M. L., Dixon, M. J., Pessoa-Brandao, L., Pelletier, R., Miret, J. J. and Lahue, R. S. (2001). Cis-elements governing trinucleotide repeat instability in Saccharomyces cerevisiae. *Genetics* **157**, 1569-79.

Rosche, W. A., Trinh, T. Q. and Sinden, R. R. (1995). Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J Bacteriol* **177**, 4385-91.

Runyon, G. T., Bear, D. G. and Lohman, T. M. (1990). Escherichia coli helicase II (UvrD) protein initiates DNA unwinding at nicks and blunt ends. *Proc Natl Acad Sci U S A* **87**, 6383-7.

Sakamoto, N., Chastain, P. D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J. D. and Wells, R. D. (1999). Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol Cell* **3**, 465-75.

Sakamoto, N., Ohshima, K., Montermini, L., Pandolfo, M. and Wells, R. D. (2001). Sticky DNA, a self-associated complex formed at long GAA*TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J Biol Chem* **276**, 27171-7.

Samadashwily, G. M., Raca, G. and Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication in vivo. *Nat Genet* **17**, 298-304.

Sancar, A. and Rupp, W. D. (1983). A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell* **33**, 249-60.

Sarkar, P. S., Chang, H. C., Boudi, F. B. and Reddy, S. (1998). CTG repeats show bimodal amplification in E. coli. *Cell* **95**, 531-40.

Saunders, N. J., Jeffries, A. C., Peden, J. F., Hood, D. W., Tettelin, H., Rappuoli, R. and Moxon, E. R. (2000). Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58. *Mol Microbiol* **37**, 207-15.

Saveson, C. J. and Lovett, S. T. (1997). Enhanced deletion formation by aberrant DNA replication in Escherichia coli. *Genetics* **146**, 457-70.

Savouret, C., Brisson, E., Essers, J., Kanaar, R., Pastink, A., te Riele, H., Junien, C. and Gourdon, G. (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. *Embo J* **22**, 2264-73.

Schaaper, R. M. (1993). Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *J Biol Chem* **268**, 23762-5.

Schlotterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20**, 211-5.

Schmidt, K. H., Abbott, C. M. and Leach, D. R. (2000). Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli. *Mol Microbiol* **35**, 463-71.

Schumacher, S., Fuchs, R. P. and Bichara, M. (1998). Expansion of CTG repeats from human disease genes is dependent upon replication mechanisms in Escherichia coli: the effect of long patch mismatch repair revisited. *J Mol Biol* **279**, 1101-10.

Schumacher, S., Pinet, I. and Bichara, M. (2001). Modulation of transcription reveals a new mechanism of triplet repeat instability in Escherichia coli. *J Mol Biol* **307**, 39-49.

Schweitzer, J. K. and Livingston, D. M. (1997). Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Hum Mol Genet* **6**, 349-55.

Schweitzer, J. K. and Livingston, D. M. (1998). Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum Mol Genet* **7**, 69-74.

Schweitzer, J. K. and Livingston, D. M. (1999). The effect of DNA replication mutations on CAG tract stability in yeast. *Genetics* **152**, 953-63.

Seigneur, M., Bidnenko, V., Ehrlich, S. D. and Michel, B. (1998). RuvAB acts at arrested replication forks. *Cell* **95**, 419-30.

Seigneur, M., Ehrlich, S. D. and Michel, B. (2000). RuvABC-dependent double-strand breaks in dnaBts mutants require recA. *Mol Microbiol* **38**, 565-74.

Selby, C. P. and Sancar, A. (1993). Molecular mechanism of transcription-repair coupling. *Science* **260**, 53-8.

Sharples, G. J. and Leach, D. R. (1995). Structural and functional similarities between the SbcCD proteins of Escherichia coli and the RAD50 and MRE11 (RAD32) recombination and repair proteins of yeast. *Mol Microbiol* **17**, 1215-7.

Shimizu, M., Gellibolian, R., Oostra, B. A. and Wells, R. D. (1996). Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J Mol Biol* **258**, 614-26.

Sinden, R. R. (2001). Neurodegenerative diseases. Origins of instability. *Nature* **411**, 757-8.

Sinden, R. R., Potaman, V. N., Oussatcheva, E. A., Pearson, C. E., Lyubchenko, Y. L. and Shlyakhtenko, L. S. (2002). Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* **27**, 53-65.

Smith, G. K., Jie, J., Fox, G. E. and Gao, X. (1995). DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic Acids Res* **23**, 4303-11.

Sopher, B. L., Myrick, S. B., Hong, J. Y., Smith, A. C. and La Spada, A. R. (2000). In vivo expansion of trinucleotide repeats yields plasmid and YAC constructs for targeting and transgenesis. *Gene* **261**, 383-90.

Strand, M., Prolla, T. A., Liskay, R. M. and Petes, T. D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-6.

Studwell-Vaughan, P. S. and O'Donnell, M. (1991). Constitution of the twin polymerase of DNA polymerase III holoenzyme. *J Biol Chem* **266**, 19833-41.

Summers, D. K. and Sherratt, D. J. (1984). Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**, 1097-103.

Takahashi, N., Sasagawa, N., Suzuki, K. and Ishiura, S. (1999). Synthesis of long trinucleotide repeats in vitro. *Neurosci Lett* **262**, 45-8.

Takahashi, S., Hours, C., Chu, A. and Denhardt, D. T. (1979). The rep mutation. VI. Purification and properties of the Escherichia coli rep protein, DNA helicase III. *Can J Biochem* **57**, 855-66.

Tougu, K. and Marians, K. J. (1996). The interaction between helicase and primase sets the replication fork clock. *J Biol Chem* **271**, 21398-405.

Tran, H. T., Keen, J. D., Kricker, M., Resnick, M. A. and Gordenin, D. A. (1997). Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol Cell Biol* **17**, 2859-65.

Trinh, T. Q. and Sinden, R. R. (1991). Preferential DNA secondary structure mutagenesis in the lagging strand of replication in E. coli. *Nature* **352**, 544-7.

Usdin, K. and Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res* **23**, 4202-9.

Uzest, M., Ehrlich, S. D. and Michel, B. (1995). Lethality of rep recB and rep recC double mutants of Escherichia coli. *Mol Microbiol* **17**, 1177-88.

van Belkum, A., Scherer, S., van Alphen, L. and Verbrugh, H. (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev* **62**, 275-93.

van Brabant, A. J., Stan, R. and Ellis, N. A. (2000). DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet* **1**, 409-59.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F. and Petit, M. A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. *Embo J* **24**, 180-9.

Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S. C., Le Cam, E. and Fabre, F. (2003). The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309-12.

Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P. et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905-14.

Vilette, D., Ehrlich, S. D. and Michel, B. (1995). Transcription-induced deletions in Escherichia coli plasmids. *Mol Microbiol* **17**, 493-504.

Vilette, D., Ehrlich, S. D. and Michel, B. (1996). Transcription-induced deletions in plasmid vectors: M13 DNA replication as a source of instability. *Mol Gen Genet* **252**, 398-403.

Vilette, D., Uzest, M., Ehrlich, S. D. and Michel, B. (1992). DNA transcription and repressor binding affect deletion formation in Escherichia coli plasmids. *Embo J* **11**, 3629-34.

Wang, T. C. (2005). Discontinuous or semi-discontinuous DNA replication in Escherichia coli? *Bioessays* **27**, 633-6.

Wang, T. C. and Smith, K. C. (1989). The roles of RecBCD, Ssb and RecA proteins in the formation of heteroduplexes from linear-duplex DNA in vitro. *Mol Gen Genet* **216**, 315-20.

Washburn, B. K. and Kushner, S. R. (1991). Construction and analysis of deletions in the structural gene (uvrD) for DNA helicase II of Escherichia coli. *J Bacteriol* **173**, 2569-75.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature* **359**, 794-801.

Whitby, M. C., Ryder, L. and Lloyd, R. G. (1993). Reverse branch migration of Holliday junctions by RecG protein: a new mechanism for resolution of intermediates in recombination and DNA repair. *Cell* **75**, 341-50.

Yu, A., Dill, J. and Mitas, M. (1995a). The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. *Nucleic Acids Res* **23**, 4055-7.

Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Haworth, I. S. and Mitas, M. (1995b). The trinucleotide repeat sequence d(GTC)15 adopts a hairpin conformation. *Nucleic Acids Res* **23**, 2706-14.

Yudkin, M. D. (1970). Catabolite repression of the lac operon. Effect of mutations in the lac promoter. *Biochem J* **118**, 741-6.

Zheng, M., Huang, X., Smith, G. K., Yang, X. and Gao, X. (1996). Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. *J Mol Biol* **264**, 323-36.

Zoghbi, H. Y. and Orr, H. T. (1999). Polyglutamine diseases: protein cleavage and aggregation. *Curr Opin Neurobiol* **9**, 566-70.