INVESTIGATION OF TRINUCLEOTIDE REPEAT INSTABILITY IN THE *ESCHERICHIA COLI* CHROMOSOME

CAMELIA MIHAESCU

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2002

Declaration

I hereby declare that this thesis was composed by myself and that the work described is my own, unless stated otherwise.

Camelia Mihaescu

September 2002

ABSTRACT

The expansion of trinucleotide repeat tracts is the cause of nearly twenty genetic disorders. Almost all these diseases are characterised by anticipation, which means an earlier age of onset and an increased severity of the symptoms from one generation to the next. The mechanisms of trinucleotide repeat expansion are not understood.

In the course of this project, I have investigated the instability of a trinucleotide repeat array of 43 copies integrated at the *attB* site of chromosomes of various *Escherichia coli* mutants. The trinucleotide repeat tract $(CTG)_{43}$ was integrated into the *E. coli* chromosome in both possible orientations using an intermediate vector and exploiting site-specific recombination between the *attB* site of the chromosome and the *attP* site of the vector. Using this method I have constructed 60 mutant strains of *E. coli* which contain the trinucleotide repeat tract and are deficient in genes involved in replication, recombination, secondary structure repair or mismatch repair.

Techniques for the analysis of the instability of the trinucleotide repeat arrays were developed and used to quantify repeat instability. These included: digestion of chromosomal DNA with a rare-cutting restriction endonuclease and PAGE of the labelled fragments; PCR of the trinucleotide repeat tract, followed by restriction enzyme digestion and PAGE; fluorescent PCR and f-TRAMP (fluorescent trinucleotide amplification which uses just one primer in repeated cycles of linear

primer extension): products were separated by capillary electrophoresis and analysed using Gene Scan software.

Intensive analyses of different *E. coli* mutants showed that the trinucleotide repeat arrays integrated into the chromosome are stable. Except in one case, no instability was observed in any mutant deficient in replication, recombination, mismatch repair or secondary structure repair. The only strain, which showed instability, was a *mutD* mutant (impaired in the proof-reading activity of DNA Polymerase III). Possible explanations for this observation are discussed.

Acknowledgements

I would like to thank my supervisor Prof. Dr. David Leach for giving me the opportunity to work in his laboratory and for his help and support over the past few years. I would also like to thank all of the people in the Leach lab for creating such a great work environment and being good friends out of the lab as well.

An enormous thank goes to my family for everything you have done for me and enabling me to get this far.

Last, but by far most importantly, I would like to thank Ken and Noreen Murray and the Darwin Trust of Edinburgh for the generous financial support that made it possible for me to study in Edinburgh.

Table of Contents

Chapter 1

Introduction

1.1 Repeated DNA

The DNA sequences in the human cell nucleus which can be transcribed and translated account for 5% of the total. A small fraction represents regulatory elements required to control gene expression, but the great bulk serves no recognised function. Much of it is in the form of fairly short nucleotide sequences repeated over and over again, perhaps thousands of times. These tandem arrays seem to have arisen in the course of evolution by processes such as unequal crossing-over between chromosomes at meiosis and also slippage, generating not only multiple copies of the repeat units, but also considerable diversity within the species in size and chromosomal dispersion of the individual blocks of repetitive DNA.

Some of the repeated elements are highly characteristic of particular species so that, for example, DNA of mouse or human origin can be identified by looking for the individual "trade marks". In some species, the very highly repetitive component forms a separate peak when total DNA is run out on a density gradient. This has given rise to the name "satellite" DNA and, by analogy, less abundant repetitive elements are termed "minisatellites".

The diversity in size and number of individual blocks of repetitive DNA makes them useful for distinguishing one member of the species from another in the

technique known as "DNA fingerprinting". Each of us has a unique set of these blocks, half having been inherited from each parent, and we pass on a random mixture, totalling half our own set, to each of our children. Thus, by comparing the extent of similarity between the minisatellite block patterns of two individuals we can assess the closeness of the family relationship between them. Conversely, if we have a source of DNA from an assailant, large numbers of suspects can be screened with the secure knowledge that only the perpetrator of the crime will have a minisatellite block pattern corresponding exactly to that of the initial sample.

Microsatellites or simple sequence repeats (SSRs) are tandemly repeated tracts of DNA composed of 1-6 base pair (bp) long units. They are ubiquitous in prokaryotes and eukaryotes, both in protein-coding and non-coding regions (Toth *et al.*, 2000). Because of their high mutability, microsatellites are thought to play a significant role in genome evolution by creating and maintaining genetic variation.

A subset of SSRs, namely trinucleotide repeats (TR), are of great interest because of the role they play in many human neurodegenerative disorders (fragile-X syndrome, Huntington disease, myotonic dystrophy, Kennedy disease, spinal-bulbar muscular atrophy, spinocerebellar ataxia). The alteration responsible for these genetic diseases is the expansion of triplet repeats, where the rate of mutation depends on the number of tandem units within the repeat.

These sequences can either be perfect repeats, that is, with no variation in the base composition of the repeat motif, or imperfect repeats, where there are some copies that vary from the canonical repeat motif. Many repeats are composites of perfect and imperfect units.

Though there are 64 different codons, there are only 10 different types of trinucleotide repeat. From the 64 trinucleotides, one can subtract the four that contain only one type of base and the other 60 have to be divided by two because there is another trinucleotide on the complementary strand, and further divided by three because the frame of translation is not important as regards to the DNA sequence. The ten possible trinucleotide repeats are:

1. AAC•GTT = ACA•TGT = CAA•TTG
2. AAG•CTT = AGA•TCT = GAA•TTC
3. AAT•ATT = ATA•TAT = TAA•TTA
4. ACC•GGT = CCA•TGG = CAC•GTG
5. ACG•CGT = CGA•TCG = GAC•GTC
6. ACT•AGT = CTA•TAG = TAC•GTA
7. AGC•GCT = GCA•TGC = CAG•CTG
8. AGG•CCT = GGA•TCC = GAG•CTC
9. ATC•GAT = TCA·TGA = CAT•ATG
10. CCG•CGG = CGC•GCG = GCC•GGC

Of the 18 trinucleotide repeat disease loci that had been identified so far, all contained only three of these repeat sequences: CAG•CTG; CCG•CGG and GAA•TTC.

1.2 Dynamic Mutation

The rate of mutation in the human genome varies widely. Unique gene sequences, even of nonessential function, appear to change very slowly. Such mutations are static, in that the mutant has the same rate of mutation as its predecessor. Microsatellites have a unique form of mutation: variation in copy number. The rate of the mutation is related to the copy number, and therefore, the mutability of the product of a change in copy number is different from that of its predecessor. For this reason, Richards and Sutherland (Richards & Sutherland, 1992) have termed this mechanism dynamic mutation.

Dynamic mutation is a process of multiple steps rather than a single event. The first step in the process involves rare, small changes in repeat copy number and/ or composition, such that a relatively unstable number of perfect repeats is achieved. This repeat allele then becomes increasingly unstable with further increases in copy number. The general properties of dynamic mutations are:

-1) mutation rate is related to perfect-repeat copy number;

-2) rare initial events lead to alleles with an increase in the number of perfect repeats. These repeats have copy numbers near the top of the normal range and are a pool of unstable alleles with increased risk of expansion to disease-causing alleles.

-3) there is a relationship between repeat copy number and age-at-onset and severity of disease symptoms. Together, these properties account for the increasing severity of the disease in successive generations within a family - a phenomenon referred to as anticipation.

The normal (or wild type) alleles at loci that undergo dynamic mutation show polymorphism (variation in the number of repeat copies from chromosome to chromosome). As the total number of repeats in an allele increases, the instability increases into the premutation, or carrier, range until the expansion manifests as a chromosomal fragile site or phenotypic effect. Premutation, or carrier, alleles show instability upon transmission, but are of insufficient copy number to cause disease or fragile sites.

Dynamic mutation is not restricted to trinucleotides: massive expansion of a 33-bp AT-rich minisatellite repeat is the molecular basis of the fragile site FRA16B. The expansion of a 12-bp GC-rich repeat is the most common mutation at the EPM1 epilepsy locus (Lafreniere *et al.*, 1997).

Estimation studies of mutation rates at microsatellite loci showed the dinucleotide repeat loci appear to evolve at a rate 1.5-2 times greater than the tetranucleotide loci. The non-disease-related trinucleotide loci have mutation rates intermediate between the di- and tetranucleotides. In contrast, the disease-related trinucleotides have a mutation rate higher than the dinucleotides, even within the normal allele size range (Chakraborty *et al.*, 1997). These conclusions disagree with those of Weber and Wong (Weber & Wong, 1993) who found that the average mutation rate for tetranucleotides is nearly four times higher than that for dinucleotides.

1.3 Fragile Sites

Fragile sites on chromosomes are loci that show an increased frequency of gaps and breaks when the cells are exposed to specific conditions of tissue culture or chemical agents. They are divided into rare and common ones. All common fragile sites are present in all individuals, most likely as part of their normal chromosome structure. The rare ones range in frequency from 1 in 40 chromosomes for FRA16B and 1 in 80 chromosomes for FRA10B to just one single report for FRA1M at 1p21.3.

Most fragile sites are not expressed spontaneously, but require induction by manipulation of tissue culture conditions. Fragile sites are also divided according to the conditions of tissue culture required to induce their cytogenetic expression. The common fragile sites are divided into:

- aphidicolin inducible (FRA3B, FRA16D, FRAXD),
- 5-azacytidine inducible (FRA1J, FRA19A) and
- bromodeoxyuridine inducible (FRA6D, FRA10C).

The rare fragile sites are classified into:

- folate sensitive (FRA11B, FRA16A, FRAXA, FRAXE, FRAXF),
- distamycin A inducible (FRA16B, FRA17A, FRA8E) and
- bromodeoxyuridine requiring (FRA10B, FRA12C) (Sutherland & Richards, 1995).

The rare folate-sensitive fragile sites have all been shown to be dynamic mutations of naturally occurring CCG trinucleotide repeat sequences.

Fragile sites are discontinuities of chromatin in metaphase chromosomes. The chromosome can break at a fragile site, giving rise to a variety of unusual chromosomal configurations. Culturing somatic cells containing human chromosomes with fragile sites under conditions, which induce their cytogenetic expression, gives rise to rearrangements with breakpoints at the fragile sites. Whether such rearrangements occur *in vivo* has been the subject of controversy until it was shown that breakage near the rare fragile site at 11q23.3 (FRA11B) in humans gives rise to a chromosomal deletion ($11q^-$) resulting in Jacobsen syndrome (Jones *et al.*, 1994). FRAXA and FRAXE are two other fragile sites with clinical significance. FRAXA and FRAXE are the cytogenetic manifestations in fragile X syndrome and in mild mental retardation, respectively.

1.4 Trinucleotide Repeat Expansion Diseases

For some repeats, large (expanded) alleles above a certain threshold were shown to be pathogenic. There are nearly 20 known triplet repeat expansion diseases that are divided into three classes.

- A first class is characterised by large CAG expansions that code for a polyglutamine stretch. The polyglutamine disease class comprises Huntington disease (HD), five spinocerebellar ataxias (SCA1, 2, 3/ Machado-Joseph disease = MJD, SCA7 and SCA17), dentatorubral-pallidoluysian atrophy (DRPLA) and spinal

and bulbar muscular atrophy (SBMA) or Kennedy's disease. The pathogenic threshold (35-39 repeats) is similar in this class of neurodegenerative diseases and the clinical phenotypes include abnormal voluntary and involuntary movements, frequently accompanied by neurophsychiatric features. Pathologically, the neurodegeneration is probably due to a toxic gain of function of the large polyglutamine stretches. It is most likely that there is a selective neuronal vulnerability since the genes are ubiquitously expressed and only certain regions of the brain are damaged.

- Diseases that are caused by a short expansion of a translated triplet repeat are grouped into a second class. Spinocerebellar ataxia 6 (SCA6) is caused by an increase of a CAG repeat above 19 units in a neuronal P-type calcium channel (*CACNA1A*). Small increases in the length of the repeat might alter the channel function, thereby changing calcium fluxes resulting in damage to the cell. As the channel is greatly enriched in cerebellar Purkinje cells, this mechanism may hamper cerebellar function. Three diseases are caused by small increases in a triplet repeat that is translated as a stretch of alanine residues. Two are caused by short expansions of a GCN repeat, (cleidocranial dysplasia and synpolydactyly), and one by a short expansion in a GCG repeat, (oculopharyngeal muscular dystrophy). A last disease, multiple epiphyseal dysplasia, occurs when a $(CAG)_5$ contracts to four units or expands to six or seven units (Goodman *et al.*, 1997).

- The third class contains diseases caused by vast expansions in untranslated triplet repeats. Friedreich ataxia (FA) is caused by an expansion in an intronic GAA repeat. Two subtypes of Fragile X syndrome (FRAXA and FRAXE) are caused by enormous expansions at a CGG and a GCC repeat respectively, both

located in the 5' - untranslated region (UTR). These expansions result in hypermethylation on CpG islands in an adjacent promoter with consequent loss of gene transcription. Expansion of another 5'-UTR CGG repeat can result in the deletion of the distal portion of chromosome 11q leading to Jacobsen syndrome. Two spinocerebellar ataxias are associated with an expansion of untranslated triplet repeats, SCA12 with a 5'-UTR CAG repeat and SCA8 with a 3'-UTR CTG repeat. Also myotonic dystrophy (DM) is caused by expansion of a CTG repeat in a 3'UTR.

Anticipation

Ten of all these triplet repeat expansion disorders (TREDs) (the 7 large polyglutamine expansions, FRAXA, MD and FA) were reported to show anticipation. Anticipation is defined as a decrease in age at onset (AAO) and/or an increase severity in successive generations. The anticipation observed in the triplet repeat expansion disease has a biological explanation by the combination of two factors. The first factor is that, in contrast to repeats of normal length, the large, pathogenic repeats are unstably transmitted and have the tendency to expand even further when passed on to the next generation. The second factor is an inverse correlation between pathogenic repeat length and AAO. Neither factor is present in 100% of the cases: pathogenic repeats are occasionally observed to contract and the repeat length does not explain all the variance in AAO. Still, the sum of both trends results in the observed anticipation (Goossens *et al.*, 2001).

1.4.1 Huntington Disease (HD)

Huntington disease was first described by George Huntington in 1872. It is caused by the expansion of a CAG repeat sequence in the first exon of the huntingtin gene on chromosome 4p16.3. The mutant allele is dominant and causes severe neurodegenerative disease, which usually sets in in middle age and leads first to uncontrolled movements, then to mood disturbances, depression and death.

Huntington disease is one of eight currently known neurodegenerative diseases in which an unstable expansion of a CAG repeat within the coding region of the responsible gene is translated into a polyglutamine (polyQ) domain within the disease protein. The HD gene contains in excess of 36 CAG repeats in exon 1 resulting in an expanded polyQ tract near the N-terminus of the 350-KDa cytoplasmic huntingtin protein. Large CAG repeats are associated with an earlier age of onset and also with accelerated clinical progression.

The most crucial factor that is required for the engagement and maintenance of a progressive HD phenotype is the continued expression of mutant Huntingtin (Orr & Zoghbi, 2000). The most compelling evidence for this comes from a conditional transgenic mouse model expressing exon 1 of the human HD gene (with 94Q residues) under the control of a tetracyclin-dependent promoter which allows programmable "switching off" of the transgene expression. Switching off the transgene at a stage when mice are already symptomatic leads to an arrest of pathological changes, clearance of Htt aggregates from cerebral neurones and, most importantly, stabilization and even improvement of the established motor abnormalities (Yamamoto *et al.*, 2000).

Several lines of evidence from cell culture, transgenic animal models, and also the brains of HD patients, incriminate N-terminal products of the proteolytic cleavage of mutant Htt, and suggest the toxic effects are exerted predominantly in the nucleus. This basic mechanism seems to be shared by other polyQ disorders. Htt is a cytoplasmic protein and its cleavage probably occurs mostly in the cytoplasm before the translocation of cleaved product to the nucleus. The nuclear accumulation of cleaved mutant Htt seems to occur by passive diffusion rather than active transport (Hackam *et al.*, 1999) and increases the susceptibility of cultured cells to apoptotic death. In keeping with this, cell death *in vitro* can be exaggerated or prevented by manipulations that alter the nuclear accumulation of mutant Htt, such as attaching a nuclear localization signal or a nuclear export sequence to the construct used in cell transfection (Saudou *et al.*, 1998).

However, there is also evidence that toxic effects of N-terminally truncated mutant Htt are mediated within the cytoplasm; it is likely that both nuclear and cytoplasmic toxicity occurs *in vivo*. Currently, it is unclear how these basic processes, experimentally modelled in neuronal and non-neuronal cell lines, are "translated" into the selective neurodegeneration characteristic of HD. Candidate mechanisms could include the level of mutant Htt expression, the distribution of insoluble aggregates of mutant Htt, protein-protein interactions with specific intracellular proteins either in the cytoplasm or nucleus, or a cell type-specific activation of proapoptotic mechanisms involving cysteine kinases.

1.4.2 Spinocerebellar Ataxias (SCAs)

The spinocerebellar ataxias (SCAs) are a group of autosomal dominant progressive disorders that have overlapping and variable phenotypes. They cannot be distinguished based on clinical features alone, but 16 distinct SCAs (1-8 and 10-17) have been identified. Cerebellar dysfunction is a hallmark of all the SCAs, but many also include abnormalities in other regions in the central and/or peripheral nervous system. At the genetic level, most of the SCAs are caused by translation of an expanded CAG repeat into an abnormally long polyglutamine tract within the corresponding protein.

The mutations for SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA12 and SCA17 are identified and found to be caused by an expansion of a CAG or a CTG repeat sequence of these genes. Six additional loci for SCA4, SCA5, SCA10, SCA11, SCA13 and SCA14 are mapped. SCA4, mapped on chromosome 16, is characterised by a late-onset spinocerebellar ataxia with sensory axonal neuropathy. SCA5, mapped on chromosome 11, is a slowly progressive late-onset cerebellar syndrome. Linkage of cerebellar disease phenotypes in families to loci on chromosome 22 for SCA10 (Zu *et al.*, 1999), chromosome 15 for SCA11 (Worth *et al.*, 1999) and chromosome 19 for SCA13 and SCA14 (Herman-Bert *et al.*, 2000),(Yamashita *et al.*, 2000) were reported. SCA10 is the first SCA shown to be caused by expansion of a non-triplet repeat (the pentamer ATTCT).

1.4.2.1 SCA1

SCA1 is a relatively rare autosomal dominant neurodegenerative disease typically with mid-life onset characterised by motor symptoms in the absence of cognitive deficits. Death usually occurs between 10 and 15 years after the onset of symptoms. The clinical features of SCA1 vary depending on the stage of the disease, but typically in addition to ataxia, include dysarthria, swallowing and breathing problems. At the pathological level, the most frequent and severe alterations seen in SCA1 patients are the loss of Purkinje cells in the cerebellar cortex and degeneration of neurons in the inferior olivary nuclei, the cerebellar dentate nuclei and the red nuclei.

Characterisation of SCA1 revealed that it encoded a novel protein of about 800 amino acids; the absolute length depends on the number of CAG and CAT triplets. Wild-type alleles show between 6 and 44 CAG copies, while mutant alleles have 40-82 repeats. The SCA1 gene spans 450 kb and consists of nine exons. The SCA1 transcript was unique in that the 5'- untranslated region (5'-UTR) was encoded by exons 1-7 and a portion of exon 8. In addition, the 3'-UTR encoded by the last exon extended for 7277 bases. Expression analysis revealed that mRNA was expressed equally from both the wild-type and mutant alleles (Orr & Zoghbi, 2001).

SCA1 protein ataxin-1 is a nuclear protein and is likely to be cleaved in the nucleus (Orr & Zoghbi, 2000). The realisation that mutant ataxin-1 causes disease upon entering the nucleus has prompted efforts to understand the regulation of its nuclear localization and its function in the nucleus. Secondly, understanding that

expansion of the polyglutamine tract alters the folding of ataxin-1 highlighted the importance of protein folding and clearance in pathogenesis.

1.4.2.2 SCA2

Spinocerebellar ataxia type 2 is an autosomal dominantly inherited disorder that affects the cerebellum and inferior olivary nuclei. The disease is characterised by appendicular and gait ataxia, dementia, optic atrophy and slow saccadic eye movements. First symptoms usually appear between the age of 2 to 65 years, but in 40% of SCA2 cases onset occurs before the age of 25 (Imbert *et al.*, 1996). Linkage analysis assigned SCA2 to chromosome 12q23-q24 (Gispert *et al.*, 1993). Normal alleles have 14-32 CAG copies with one to three CAA interruptions. SCA2 patients have 33-77 CAG copies without interruptions. The SCA2 gene consists of 25 exons spanning 130 kb and codes for ataxin-2, a cytoplamatic protein.

1.4.2.3 SCA3 /Machado-Joseph Disease (MJD)

Machado-Joseph disease is a neurodegenerative disorder that shows autosomal dominant inheritance and is characterised by neuronal loss and gliosis in the substantia nigra and nuclei of vestibular and cranial nerves. Symptoms start around the age of 40 and include ataxia, dystonia, muscle atrophy, bulgy eyes, loss of leg reflexes, cerebellar tremors and Parkinsonian features. The causative mutation for

SCA3 has been identified as an unstable expansion of CAG trinucleotide repeats in the MJD1 gene at the 14q32.1 locus (Kawaguchi *et al.*, 1994). Wild type alleles of the ataxin-3 gene have 12-40 CAG copies; mutant ones show 55-86 CAG copies.

1.4.2.4 SCA6

Neuronal loss in SCA6 consists of the prominent loss of Purkinje cells from the cerebellar cortex, like in SCA1 and SCA2 (Gomez *et al.*, 1997). The protein affected by the polyglutamine expansion in SCA6 is the α1A voltage-dependent calcium channel (Zhuchenko *et al.*, 1997). Normal allles exhibit 4-18 CAG copies. The size range of the polyglutamine tract on mutant SCA6 alleles of 21-30 repeats is much shorter than those found for the other polyglutamine diseases. While aggregates of the α1A-subunit have been detected in brains of SCA6 patients (Ishikawa *et al.*, 1999), the polyglutamine expansion also alters the kinetic properties of the channel (Restituito *et al.*, 2000).

1.4.2.5 SCA7

SCA7 is a clinically and genetically distinct entity, in which progressive loss of photoreceptors and bipolar cells results in progressive macular dystrophy, decreased visual acuity and ultimately in blindness. The gene for SCA7 has been mapped to chromosome 3. SCA7 is caused by the expansion of a trinucleotide CAG

repeat in the coding region of Ataxin-7. Wild type alleles of the ataxin-7 gene have 7-17 trinucleotides. Mutant alleles have 38-200 CAG copies (David *et al.*, 1998).

1.4.2.6 SCA 8

SCA 8 is a progressive ataxia with cerebellar atrophy, decreased vibration sense and brisk reflexes, caused by an expanded CTG repeat in its 3'-terminal exon (Koob *et al.*, 1999a). Interestingly, SCA8 may have a CTG repeat range that is pathogenic (~110-250 repeats) with shorter and larger repeats not resulting in disease (Koob *et al.*, 1999b). Also uniquely amongst triplet repeat disorders, the SCA8 transcript does not code for a protein and may be an endogenous antisense RNA that regulates the expression of another gene.

1.4.2.7 SCA 12

SCA12 is a rare autosomal dominant cerebellar ataxia caused by a non-coding CAG trinucleotide repeat expansion in the 5'-UTR of the PPP2R2B gene (Holmes *et al.*, 1999). Patients present with action tremor of the upper extremities and progress to develop a wide range of signs and symptoms, including mild cerebellar dysfunction, hyperreflexia, subtle parkinsonian features and dementia. Age of onset ranges from 8 to 55, with most individuals presenting in the fourth decade.

Magnetic resonance images of affected individuals revealed generalised atrophy of the central nervous system dominantly affecting the cerebral cortex and cerebellum.

PPP2R2B encodes a brain-specific regulatory subunit of protein phosphatase 2A (PP2A). Although the CAG repeat is located near conserved promoter elements and transcriptional start sites, repeat-mediated transcriptional interference of PPP2R2B has not yet been shown.

1.4.2.8 SCA17

SCA17 is caused by an abnormal CAG expansion in the TATA-binding protein (TBP) gene, a general transcription initiation factor. This abnormal expansion of glutamine tracts in TBP bears 47-55 repeats, whereas the normal repeat number ranges from 29 to 42. Age at onset ranges from 19 to 49 years. Most individuals present in the third decade with gait ataxia and dementia (Nakamura *et al.*, 2001).

1.4.3 Dentatorubral-pallidoluysian Atrophy (DRPLA) /Haw-River Syndrome

Dentatorubral-pallidoluysian atrophy is a rare neurodegenerative disorder that shows autosomal dominant inheritance and is clinically characterised by ataxia. The causative mutation for DRPLA has been identified as an unstable expansion of

CAG trinucleotide repeats at 12pter-p12 (Koide *et al.*, 1994),(Nagafuchi *et al.*, 1994).

The normal DRPLA gene encodes the Antrophin protein of about 1184 amino acid residues, with a CAG array of 3-36 copies in normal individuals and expansion from 49 to 88 in affected patients. The repeat begins at residue 484, preceded by five histidines, plus serine repeats and segments rich in serines and prolines.

1.4.4 Spinobulbar Muscular Atrophy (SBMA) /Kennedy disease

SBMA, described for the first time in 1968 by William Kennedy, is a rare X-linked, adult-onset, motor neurone disorder resulting in spinal and bulbar muscular atrophy. It was the first disease found to be caused by a CAG repeat expansion situated in the first exon of the androgen receptor gene (AR) at the Xq13-21 locus. Normal copy number is 9-36, while affected individuals have 38- 62 copies. Kennedy's disease is a progressive neuromuscular disease that is characterised by proximal muscle weakness, atrophy and fasciculations.

Males who have point mutations in the androgen receptor gene exhibit feminisation, but they do not get motor neurone disease. This is convincing proof that Kennedy disease is due to gain rather than loss of function.

1.4.5 Cleidocranial Dysplasia

Cleidocranial dysplasia is caused by a polyalanine expansion in the gene encoding the α1 core-binding factor (Mundlos *et al.*, 1997).

1.4.6 Synpolydactyly (SPD)

Synpolydactyly is a rare dominantly inherited congenital limb malformation. Typical cases have 3/4 finger and 4/5 toe syndactyly, with a duplicated digit in the syndactylous web, but incomplete penetrance and variable expressivity are common. The condition has been thought to be caused by expansions of an imperfect trinucleotide repeat sequence encoding a 15-residue polyalanine tract in HOXD13 (Goodman *et al.*, 1997).

The mutations in these two rare diseases (cleidocranial dysplasia and synpolydactyly) are not triplet repeats, but duplications of "cryptic GCN repeats" composed of mixed synonymous codons (Warren, 1997).

1.4.7 Oculopharyngeal Muscular Dystrophy (OPMD)

OPMD is an autosomal dominant adult-onset disease. It usually presents in the sixth decade with progressive swallowing difficulties (dysphagia), eyelid

drooping (ptosis) and proximal limb weakness. Its pathological hallmark are the unique nuclear filament inclusions in scheletal muscle fibres. The normal $(GCG)_6$ repeat encoding a polyalanine tract located at the N terminus of the protein, expands to $(GCG)_{8-13}$ in affected individuals. The homozygosity for the $(GCG)_7$ allele leads to autosomal recessive OPMD. More severe phenotypes were described in compound heterozygotes for the $(GCG)_9$ mutation and a $(GCG)_7$ allele. Thus the $(GCG)_7$ allele is an example of polymorphism which can act either as a modifier of a dominant phenotype or as a recessive mutation (Brais *et al.*, 1998).

OPMD was the first description of short trinucleotide repeat expansion causing a human disease. The addition of only two GCG repeats is sufficient to cause dominant OPMD. Furthemore, there is a clear cut-off between the normal and abnormal alleles, with a single GCG expansion being a recessive mutation.

1.4.8 Multiple Epiphyseal Dysplasia

The human gene for cartilage oligomeric matrix protein contains five tandem repeats of the GAC trinucleotide. Its expansion by one repeat causes multiple epiphyseal dysplasia, while expansion by two repeats or, remarkably, deletion of one repeat causes pseudoachondroplasia (Vorlickova *et al.*, 2001).

1.4.9 Friedreich's Ataxia

Friedreich's ataxia (FA or FRDA) is an autosomal recessive disorder characterised by ataxia, dysarthria, diminished reflexes, cardiomyopathy and diabetes. FRDA is caused by a large intronic GAA repeat expansion in the X25 gene (or *frataxin*) which leads to reduced X25 expression. The expanded AT-rich sequence most probably causes transcriptional interference via a self-association of the GAA/TTC tract, which stabilizes the DNA in a triplex structure. Reduced X25 mRNA decreases frataxin levels, suggesting that FRDA results from a partial loss of frataxin function. Disruption of the yeast X25 homolog (YFH1) causes abnormal accumulation of mitochondrial iron, loss of mtDNA, respiratory dysfunction, multiple iron-sulfur-dependent enzyme deficiencies and increased sensitivity to oxidative stress (Wilson & Roof, 1997). These findings, together with highly conserved mitochondrial targeting signal sequences and mitochondrial membrane localization in yeast and human, suggest that frataxin is involved in iron homeostasis and respiratory function (Koutnikova *et al.*, 1997). In FRDA samples, studies have shown increased iron content, deficient activities of proteins involved in iron-sulfur homeostasis and hypersensitivity to iron and H_2O_2 (Wong *et al.*, 1999). Therefore, frataxin insufficiency may result in abnormal iron-sulfur homeostasis and, in turn, mitochondrial dysfunction, free radical production, oxidative stress and cellular degeneration.

1.4.10 Fragile-X Syndrome (FRAXA)

Fragile-X is an inherited, X-linked dominant retardation disorder with minor dysmorphic features, affecting about 1 male in 1500 and 1 female in 2500, which makes it one of the most common human genetic disease. It is so called because most affected individuals exhibit, cytogenetically, a folate-sensitive fragile site at the 5'-untranslated region (UTR) of the first exon of the fragile-X mental retardation gene on the X chromosome (Xq27.3) (Verkerk *et al.*, 1991). Fragile sites are identified as non-staining gaps on chromosomes that are inducible by certain conditions of tissue culture.

The instability is due to variation in copy number of a trinucleotide repeat $(CGG)_n$. The FMR1 (*fragile X mental retardation-1 gene*) polymorphic CGG repeat located in the 5' UTR of the gene can be categorised in four forms based on the size of the repeat: common (6-40 repeats), intermediate (41-60 repeats), premutation (61-200 repeats) and full mutation (>200-230 repeats). Among the general population, the common repeats are usually transmitted from parent to offspring in a stable manner. Intermediate alleles are larger repeats that may or may not be transmitted stably from parent to offspring. Thus, these alleles overlap the boundary between common and premutation alleles (Nolin *et al.*, 1996). While intermediate alleles may be unstable, very few expand to the disorder-causing mutation in the next generation (Crawford *et al.*, 2001).

Expansion of the CGG repeat beyond 230 trinucleotides results in its hypermethylation together with a CpG island within the FMR1 promoter region (Eichler *et al.*, 1994). This hypermethylation recruits transcriptional silencing

machinery to the FMR1 gene, followed by reduced FMR1 transcription and loss of gene product (FMRP) (Coffee *et al.*, 1999). The mutation in fragile X syndrome is therefore a loss of the normal function of FMRP.

1.4.11 Fragile XE MR (FRAXE)

FRAXE patients have mild mental retardation and variable behaviour abnormalities. FRAXE is caused by an expansion of a polymorphic $(GCC)_n$ repeat in the promoter region of the FMR2 gene (Knight *et al.*, 1993). The expanded repeats are hypermethylated, leading to transcriptional silencing of FMR2 and subsequent loss of gene product (FMR2). Two other proteins, AF4 and LAF-4, share motifs with FMR2, and all three proteins exhibit nuclear localisation, DNA-binding capacity and transcription transactivation potential (Miller *et al.*, 2000). The putative role of FMR2 as a transcriptional activator together with its high expression levels in the hippocampus and the amygdala suggest that the cognitive and behavioural deficits in FRAXE are due to alterations in neuronal gene regulation (Chakrabarti *et al.*, 1998).

1.4.12 Jacobsen Syndrome

Jones (Jones *et al.*, 1994) mapped FRA11B (11q23.3) to the human cellular homologue of the murine oncogene *v-cbl* and murine cellular proto-oncogene *Cbl-2*

and presented evidence for a possible association of FRA11B with the loss of part of the long arm of chromosome 11 which is characteristic of Jacobsen syndrome. This was the first time that a fragile site had been linked to chromosome breakage *in vivo*. A year later, the same laboratory mapped FRA11B to the CGG repeat within the *Cbl-2* gene and showed that the CGG repeat was expanded in some Jacobsen syndrome patients (Jones *et al.*, 1995).

1.4.13 Myotonic Dystrophy (Dystrophia Myotonica)

Myotonic Dystrophy (DM) is an autosomal, dominant, multisystemic disease with a variety of clinical features including myotonia, muscular dystrophy, cardiac conduction defects, posterior iridescent cataracts and endocrine disorders. With an incidence of 1 in 8,000 it is the most common form of muscular dystrophy in adults and can be caused by a mutation on either chromosome 19q13 (DM1) or 3q21 (DM2). DM1 is caused by a CTG expansion in the 3' untranslated region of the dystrophia myotonica-protein kinase gene (DMPK) on chromosome 19. The repeat, transcribed as CUG, is normally present in 5-27 copies, while carriers have >50 copies. There is an association between copy numbers and phenotype, and the copy number increases with subsequent generations through affected pedigrees, again accounting for the phenomenon of anticipation. It remains unclear how the CTG expansion in a non-coding region of a gene causes the complex DM phenotype. Suggested mechanisms include:

-1) haploinsufficiency of the dystrophia myotonica-protein kinase (DMPK) protein;

-2) altered expression of the neighbouring genes, including SIX5;

-3) pathogenic effects of the CUG expansion in RNA which accumulates as nuclear foci and disrupts cellular function.

Several mouse models have developed different aspects of DM1: a model expressing mRNA with CUG repeats manifests myotonia and the myopathic features of DM1; a DMPK knockout has cardiac abnormalities; and SIX5 knockouts have cataracts. Taken together, these data have been interpreted to suggest that each gene may contribute to DM1 pathogenesis.

DM2 or PROMM (proximal myotonic myopathy) is caused by a CCTG expansion (mean 5,000 repeats) located in intron 1 of the zinc finger protein 9 (ZNF9) gene. Although DM1 and DM2 phenotypes are strikingly similar, they are not identical. DM2 does not show a congenital form or the severe central nervous system involvement seen in DM1. Parallels between these mutations indicate that microsatellite expansions in RNA can be pathogenic and cause multisystemic features of DM1 and DM2 (Ranum & Day, 2002).

The discovery of each new triplet repeat disorder brings tremendous clinical benefits, offering better classification of the diseases and facilitating early diagnosis and genetic counselling.

1.5 Trinucleotide Repeat Arrays and Cancer

Since the initial description in 1993 of microsatellite instability (MIN) in a class of familial colon carcinomas, hereditary nonpolyposis colorectal carcinoma (HNPCC) (Peltomaki *et al.*, 1993), MIN has been identified in a wide variety of human cancers, both familial and sporadic. However, it is not clear that the conclusions drawn for different series are directly comparable, making it difficult to derive a consensus concerning the role of MIN in human carcinogenesis.

MIN is reflected in alterations in the patterns of polymorphic, short, tandem repeat segments (microsatellites) dispersed throughout the human genome. In the original descriptions of HNPCC, it was reported that MIN was associated with mutations in certain DNA repair genes (Fishel *et al.*, 1993).

Therefore the assumption is that MIN reflects an underlying genomic instability, resulting from inactivation of both alleles at a DNA repair gene locus (Parsons *et al.*, 1993). In familial cases, a single mutation in a DNA mismatch repair gene was inherited (Liu *et al.*, 1994a); MIN followed the loss of activity of the second allele at the same locus. In the non-familial colorectal carcinoma cases demonstrating MIN, inactivation of both alleles at a single DNA mismatch repair locus is presumed to occur as a somatic event in a target colon epithelial cell (Liu *et al.*, 1995b). Kunkel and colleagues (Koi *et al.*, 1994) showed that the introduction of a wild-type copy of a mismatch DNA repair gene can restore genomic stability of microsatellites. This finding clearly indicates that mutations in mismatch DNA repair genes constitute the underlying molecular defect responsible for MIN. Wooster

(Wooster *et al.*, 1994) found that 8% of the analysed samples had an allele of a different size in the tumour which was not present in the normal DNA.

It is known that androgens play a role in prostate cancer growth because dogs and males castrated before puberty do not get prostate cancer (Wilding, 1992). The main functions of the androgen receptor are the transcriptional regulation of target genes (Stanford *et al.*, 1997) and the transport of the androgen hormone (Ross *et al.*, 1998). This function is coded for by exon 1 which also contains the CAG repeat. A smaller CAG repeat size associated with increased activation of transcription might lead to an increased risk of prostate cancer. Schoenberg (Schoenberg *et al.*, 1994) reported the case of a prostate cancer patient who had a $(CAG)_{24}$ repeat in non-cancerous tissue and an additional contracted $(CAG)_{18}$ in the cancerous tissue. Irvine (Irvine *et al.*, 1995) identified the highest frequency of short CAG repeats in African-Americans who had the highest risk of prostate cancer in the analysed population, whereas short CAG repeats were less frequently found in Asians who also had the lowest cancer risk. Among the white population they observed that cancer patients had a higher frequency of short CAG repeats than unaffected individuals. They also detected a possible link between increased length of a nearby CGG repeat and an increased risk of prostate cancer. Often individuals with a short CAG repeat were found to have a long CGG repeat, whereas individuals with a long CAG repeat had a short CGG repeat.

Two studies (Correa-Cerro *et al.*, 1999; Edwards *et al.*, 1999) so far have not detected a statistically significant association between shorter CAG length and the occurrence of prostate cancer. Most studies reported to date suggest a positive association between shorter CAG repeat lengths and prostate cancer. It has been

observed that shorter androgen receptor (AR) CAG repeats impose a higher transactivation activity on the receptor and have an increased binding affinity for androgens (Feldman, 1997). This may make the prostate more vulnerable to chronic androgen overstimulation and increased proliferative activity, which in turn, could increase the risk of somatic mutations among tumor-suppresor genes (Nelson & Witte, 2002). In contrast, the expansion of the CAG repeat above 40 copies leads to a below normal AR level, reflected in lower frequency of prostate cancer.

1.6 Trinucleotide Repeat Tracts and Bipolar Disorders (BP)

Affective disorders are a group of psychiatric disorders characterised by disturbances in mood, ranging from mania to depression. Because triplet repeat expansions were shown to be correlated with genetic anticipation, it was hypothesised that the reported anticipation in BP disorders could also be caused by a triplet repeat expansion. Investigating this hypothesis with the repeat expansion detection (RED) method, Lindblad (Lindblad *et al.*, 1995) reported the first evidence for an association of large CAG/CTG repeats with BP disorder in Swedish and Belgian patients. Many studies using the RED method in a case-control set-up followed this publication.

The initial findings were confirmed by some of the published studies. O'Donovan (O'Donovan *et al.*, 1995), (O'Donovan *et al.*, 1996) showed for the first

time in a major psychiatric disorder a length of CAG repeats significantly higher in BP compared to normal controls. The same studies failed to show any correlation between maximum product size of CAG/CTG repeats and age at onset in BP disorders. It was only later that Mendlewicz (Mendlewicz *et al.*, 1997) showed for the first time a relation between trinucleotide expansions and phenotype severity in families with BP. A significant increase of CAG repeat length between generations G1 and G2 was observed when the phenotype increased in severity between generations to a more severe form of affective illness. On the other hand, studies that did not find association of large RED products with BP disorder were also reported (Craddock *et al.*, 1997).

The RED method (Schalling *et al.*, 1993) is a method to detect triplet repeat expansions in the genomic DNA of an individual, without prior knowledge of its chromosomal location. In the RED reaction, a thermostable ligase connects triplet repeat oligonucleotides that are adjacently annealed on single stranded template DNA. The reaction mixture containing genomic DNA, oligonucleotides and ligase undergoes 200-500 cycles of ligation (70-80°C) and denaturation (94°C). Then the products are separated on a denaturing polyacrilamide gel, transferred to a nylon membrane and hybridised with a radiolabelled probe complementary to the oligonucleotide in the RED reaction. Autoradiography results in a ladder of ligation products, the largest of which represents the size of the largest trinucleotide repeat present in the studied genome. Typical triplet repeat oligonucleotides used in the RED method are $(CTG)_{17}$ and $(CTG)_{10}$.

1.7 Mechanisms of Trinucleotide Repeat Instability

Polymorphic microsatellites have been identified in most organisms. They vary in copy number within and between species and are used in genotyping, prenatal diagnostics, forensic science and population studies. The previous section (1.4 Trinucleotide Repeat Expansion Diseases) has shown that premutated and expanded trinucleotide repeats in TREDs can undergo large-scale copy number changes upon a single parent-child transmission. The expansion mechanisms are not limited only to trinucleotide repeats. A 24-bp repeat in the coding region of the gene for prion protein (PRNP) has been associated with Creutzfeldt-Jackob disease (CJD). The fragile site FRA16B is an expanded 33-bp AT-rich minisatellite repeat.

It is agreed that small changes of one or two repeat units in either direction occur by strand-slippage, which means successive melting and reannealing of the newly synthesised strand in replication (Figure 1.1).

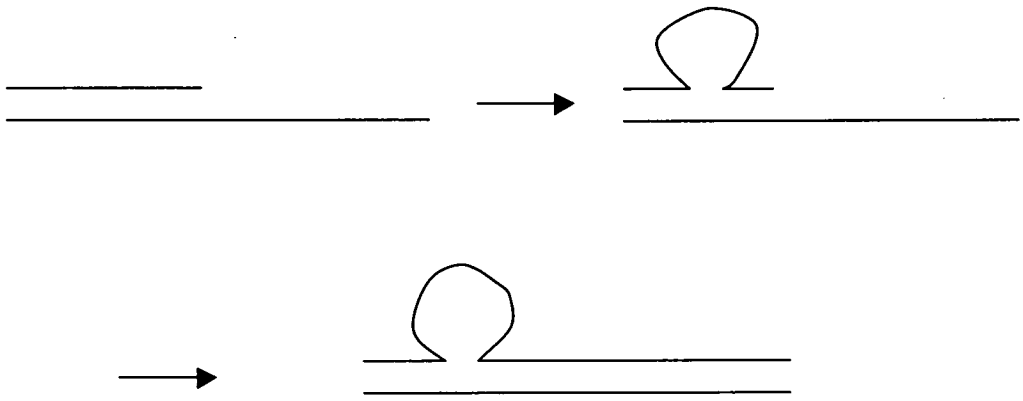


Figure 1.1. Slipped strand mispairing during DNA replication. The daughter strand dissociates from the template strand and reanneals at a distant position. Looped-out bases in the daughter strand lead to expansion, while looped-out bases in the template strand lead to deletions.

The unanswered question is what causes large changes to occur in a single generation. Most proposed mechanisms involve strand-slippage or recombination or a combination of the two. It might be that the trinucleotide repeat sequences associated with TREDs (CAG, CGG and GAA) possess intrinsic properties that promote instability. Yu (Yu *et al.*, 1995b) noticed that single strands of GC-rich palindromic sequences (CNG repeats) form hairpin structures *in vitro*. Hairpins formed by single strands of CAG and CGG repeats contain mismatched base pairs in addition to Watson-Crick base pairs and therefore they are less stable than hairpins formed by perfectly base-paired palindromic sequences. CAG homoduplexes disappear with increased length of the CAG repeat. Mariappan (Mariappan *et al.*, 1998) showed that $(CAG)_5$ exists both as hairpins and homoduplexes, while $(CAG)_{10}$ exists exclusively as hairpins. The pattern obtained from cleavage structures formed from single-stranded $(CAG)_{15}$ and $(GAC)_{15}$ repeats with single-strand specific P1 nuclease suggested to Yu (Yu *et al.*, 1995a) the formation of singly looped hairpins. P1 digestion of hairpins formed by single stranded $(CAG)_{10}$ and $(CAG)_{11}$ repeats favoured doubly looped hairpin structures with a single hydrogen bond in the A-A mispair. Petruska (Petruska *et al.*, 1996) showed that single strands consisting of ten CTG trinucleotides form slightly more stable hairpins than the complementary CAG single strand. This was attributed to the bulkier adenines in CAG repeats which may not fit into the helix as well as the smaller thymidines fit into the CTG hairpin stem. Moreover, T-T mispairs contain two hydrogen bonds and are more stable than A-A mispairs, which have only one hydrogen bond.

The discovery that all trinucleotide repeat sequences involved in TREDs can form partially Watson-Crick paired secondary structures with mispairs that contain

hydrogen bonds, together with the observation that complex secondary structures adopted by palindromic sequences promote specific replication-based deletion mutations (Trinh & Sinden, 1993), leads to the hypothesis that trinucleotide repeats may form relatively stable slipped structures during replication leading to large-scale repeat expansion (Darlow & Leach, 1998).

Repetitive sequences can slip and mispair when they are replicated. It was found that slippage of the two strands relative to each other occurred during *in vitro* DNA replication of short synthetic DNA fragments containing di-, tri- and tetranucleotide repeats (Wells *et al.*, 1965), (Wells *et al.*, 1967). The error rate of the amplification increased with increased length of the repeating unit. Looped-out bases in the daughter strand lead to expansion, while looped-out bases in the template strand lead to deletions.

Short hairpins reanneal to their complementary sequence at a rate that is indistinguishable from random DNA, while hairpins formed from longer repeats reduplex slowly. As a result, long repeats may exist long enough in a non-duplexed state *in vivo* to cause replication or repair errors. Gacy (Gacy & McMurray, 1998) proposed that hairpin formation is required but not sufficient for expansion.

Strand (Strand *et al.*, 1993) showed that $(GT)_n$ tracts in centromere-containing plasmids in yeast with mutations in three mismatch repair genes (*pms1, mlh1, msh2*) caused 100-700 fold increase in instability. When poly(GT) repeats were inserted into a chromosome, the mutation rate was only 77-fold elevated. They suggested that dramatic elevation of tract instability in repair-defective mutants reflects the failure to remove the mismatched bases present after a DNA polymerase slippage event. In wild-type cells, these mismatches would usually be corrected by

local excision of bases from the newly synthesised strand, followed by repair synthesis using the old strand as a template.

Petruska (Petruska *et al.*, 1998) proposed a model for slippage-mediated trinucleotide expansion (Figure 1.2). They proposed that during separation of the complementary strands prior to replication or transcription, a cruciform structure is formed by the trinucleotide repeat sequence. Then the hairpins move away from each other and a single strand nick is introduced opposite one of the loops. The hairpin can stretch and the gap is filled in, leading to expansion.

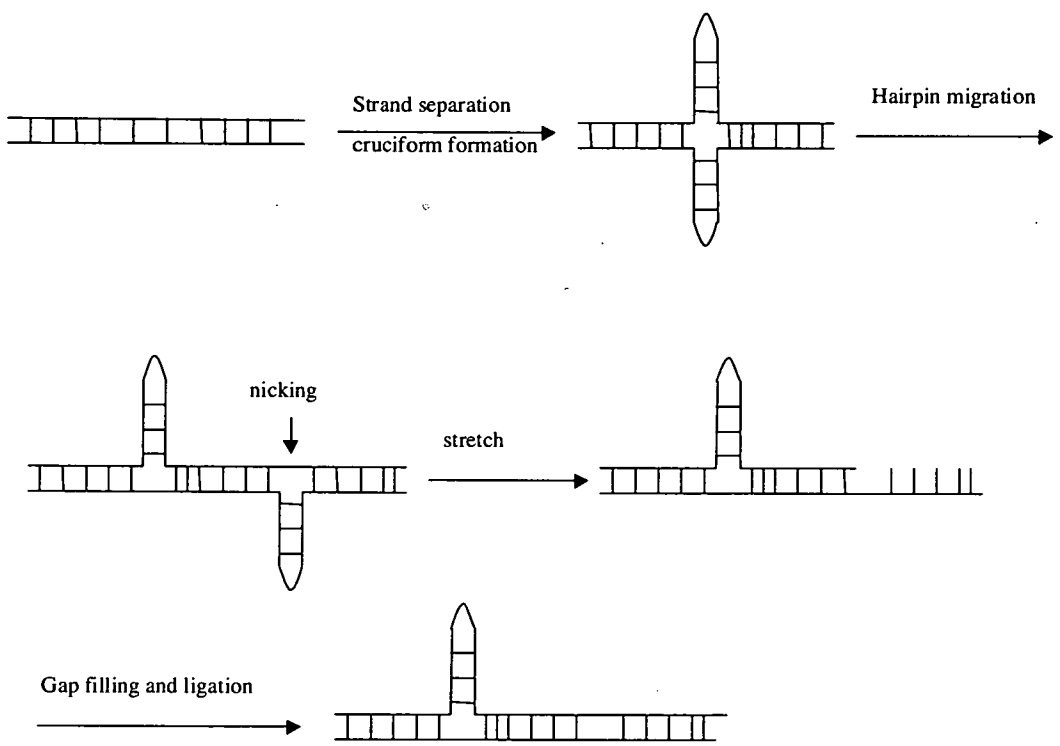


Figure 1.2. Model for *in vivo* trinucleotide repeat expansion by strand slippage.

The high flexibility of the CNG sequences suggested that it might cause a higher concentration of supercoiling within the CNG repeat sequence compared with the random DNA flanking the repeats. As a result, the high degree of supercoiling could overwhelm topoisomerase and slow down the DNA polymerase complex (Bacolla *et al.*, 1997). Kang (Kang *et al.*, 1995) found several pause sites in both strands of a $(CTG)_{130}$ repeat sequence, using two different DNA polymerases. The authors suggested that the daughter strand dissociates from its template, folds into a hairpin and reanneals to the template at a wrong position as a result of a replication block (Figure 1.3).

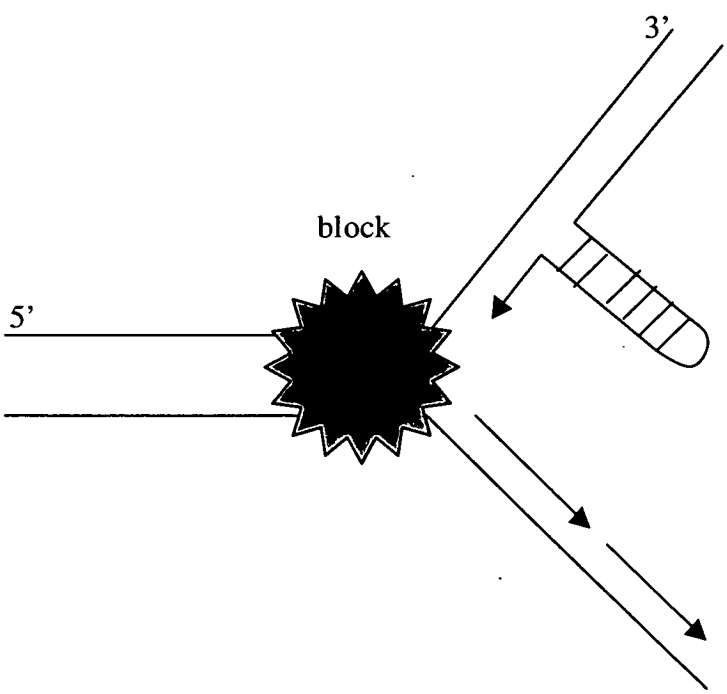


Figure 1.3. Model for trinucleotide repeat expansion. A replication block causes the DNA polymerase to pause.

Resnick (Gordenin *et al.*, 1997) suggested that a hairpin structure at the 5' end of the Okazaki fragment has a role in inhibiting FEN1 (Figure 1.4). This enzyme functions as an endonuclease metabolising the "flapped" end of the Okazaki fragment to allow ligation to the next fragment. Hairpin structures have the potential to interfere with this process. FEN1 is the human Rad27p homologue. Expanded CAG in yeast strains mutant for RAD27 supports a role for the Okazaki fragment in the repeat-expansion mechanism (Freudenreich *et al.*, 1998).

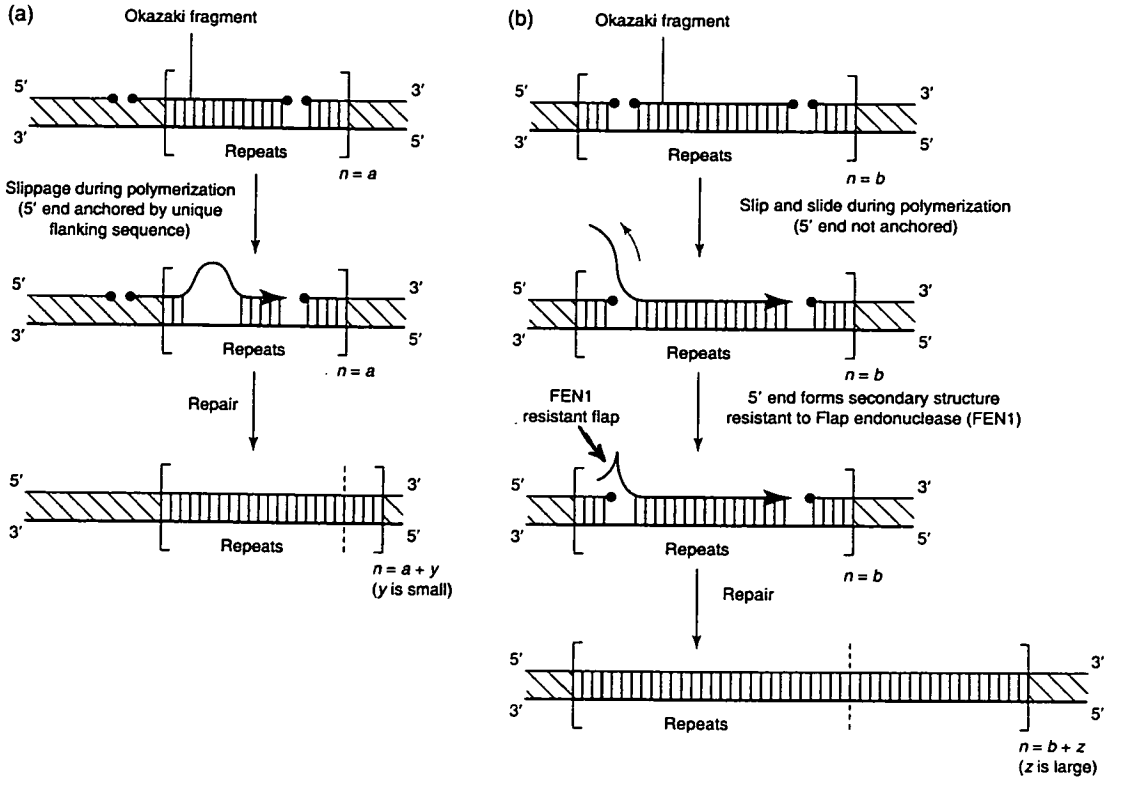


Figure 1.4. Models for Okazaki fragment-mediated increases in repeat copy numbers. A) For copy numbers less than about 80 only one gap is likely to occur within the repeat at any one time during replication. Slippage of the elongated strand during polymerisation can result in the addition of a few copies (y) of the repeat. B) For copy numbers greater than 80 , it is possible that two gaps are permitted within

the repeat during the replication. The strand between these two gaps is not anchored at either end by unique sequences and is therefore free to slide during polymerisation enabling the addition of more copies (z). The ability of the repeat sequences to form secondary structures might facilitate the formation of a FEN1 resistant hairpin at the 5' end of the Okazaki fragment.

Previous work on the instability of trinucleotide repeats in simple organisms suggested that it arises by replication slippage on either leading or lagging strand (Kang *et al.*, 1995), (Miret *et al.*, 1998), homologous recombination (Cemal *et al.*, 1999),(Jakupciak & Wells, 2000), gene conversion (Freudenreich *et al.*, 1998) and base excision repair (Lyons-Darden & Topal, 1999). These will be discussed in more detail in Chapter 6. It is not clear which of these mechanisms is used in mammalian cells *in vivo*. Leeflang (Leeflang *et al.*, 1999) suggested that expansion may occur by replication slippage during germ-cell proliferation or meiotic recombination. McMurray (Kovtun & McMurray, 2001) proposed a model of gap repair of strand breaks.

Repeated sequences tend to get deleted in prokaryotes, whereas in multicellular eukaryotes there is clearly a bias towards expansion (Levinson & Gutman, 1987). Bacteria might have evolved a bias towards deletions in order to minimise genome size because of selective pressure for rapid replication. Genome size might be less important in multicellular eukaryotes whose genetic apparatus might tend to favour insertions over deletions.

1.8 DNA Replication in *E. coli*

DNA replication is an essential event in the life of a cell. The replication process consists of three stages:

1) initiation, which ensures the separation of the two DNA strands of the template in the origin of replication region and the synthesis of the first primer;

2) elongation, which corresponds to the synthesis of the new DNA strands with the help of the replisome, a multiprotein complex;

3) termination, which occurs when the replisome meets the 5'-phosphate end of a DNA molecule or a specific termination complex.

The process of replication is precisely controlled, especially at the initiation step. At the end of replication, each of the two daughter cells will have an exact copy of the genetic material of the parent cell.

Replication is not a continuous process. It can be delayed or stopped by a variety of factors which prevent the replisome from advancing. Replication blocks can be caused by strand breaks, DNA binding proteins, RNA-DNA heteroduplexes or interactions between the replication and transcription machineries. It can lead to adaptative responses consisting of homologous recombination or rearrangements by illegitimate recombination.

The *E. coli* chromosome is a negatively supercoiled, circular, double-stranded DNA molecule of 4.6 Mbp. Its replication is bidirectional from an unique starting point named *oriC* situated at 84.3 min on the genetic map. For replication to

start, the DNA helix is required to open, so that the proteins responsible for strand synthesis can attach.

The origin of replication region has several binding sites for DnaA, an essential replication protein. Ten to twenty DnaA monomers bind to the *oriC* region forming a multiprotein complex named the initiation complex. The DNA is modified allowing separation of the two DNA strands at the AT-rich region, leading to the open complex. DnaB, an important helicase in *E. coli* replication, binds to the *oriC* region, forming the pre-initiation complex (Fang *et al.*, 1999).

DnaB is a hexamer which is loaded on the DNA by six molecules of DnaC, each of them being bound to an ATP molecule. The complex (DnaB- DnaC- ATP)$_6$ permits a direct physical interaction between DnaB and DnaA and as a result of this interaction, DnaB binds to DNA. The release of the DnaC associated with ATP hydrolysis, activates the helicase. Two DnaB helicases (one on each strand) separate the two DNA strands in a bidirectional manner. The binding of SSB (single-strand DNA binding protein) stabilizes the denatured regions of DNA.

The primase DnaG joins the DnaB-DNA complex and forms the complex of initiation (Figure 1.5). A primase is a polymerase which synthesizes DNA or RNA in the 5'->3' direction and is responsible for the synthesis of replication primers. These primers are small RNA molecules of 8-12 nucleotides complementary to the template strands which has a 3'-OH end recognized and elongated by DNA Pol lll in the opposite sense to DnaB progression.

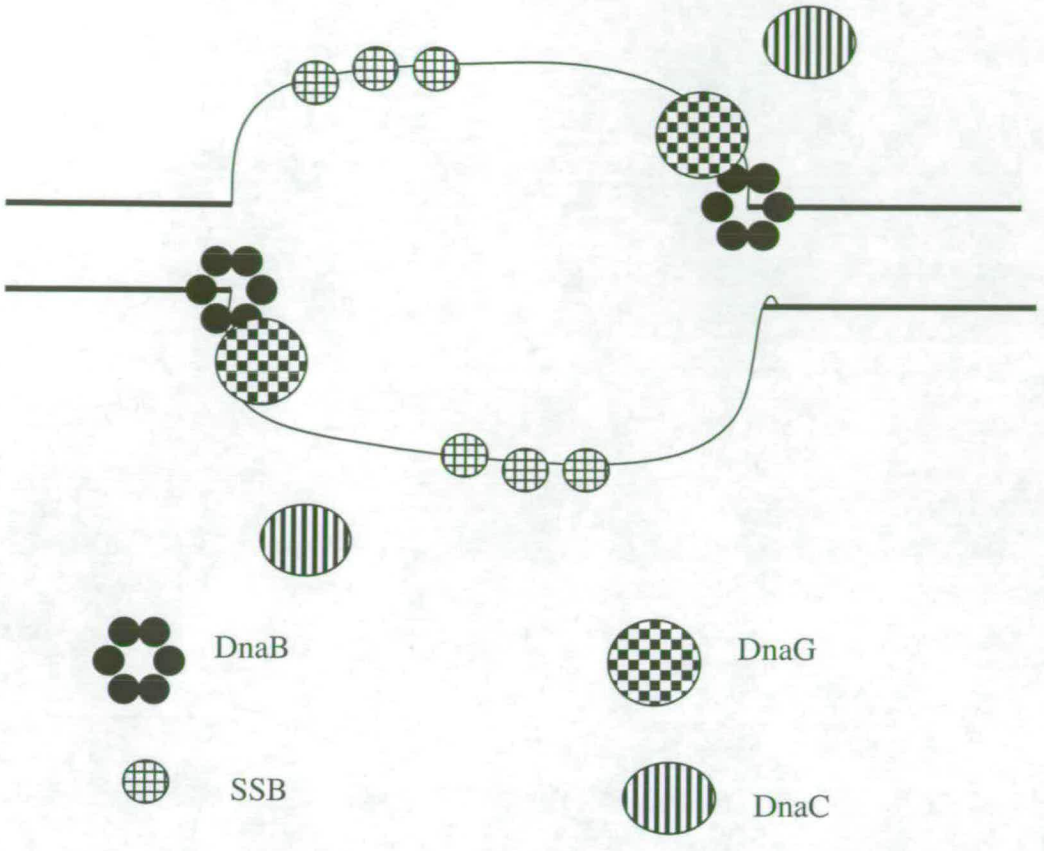


Figure 1.5. The initiation complex following the action of DnaA and DnaC which has led to the loading of DnaB.

For replication to be initiated, the DNA molecule is required to have negative supercoiling. The level of DNA supercoiling is controlled by topoisomerases which cut the DNA and ligate one strand (topoisomerases type I) or two DNA strands (type II or gyrase). Four topoisomerases have been identified in *E. coli* so far: topoisomerases I and III (type I); gyrase and topoisomerase IV (type II). Topoisomerase I, a monomeric protein, is encoded by *topA* and controls the general

level of supercoiling. Topoisomerase lll, encoded by *topB*, contributes to the resolution of the replicated chromosomes before segregation and cellular division. Gyrase has an opposite effect to topoisomerase I by introducing negative supercoils in relaxed DNA *in vitro* (Gellert *et al.*, 1976). Gyrase is ATP-dependent and is encoded by genes *gyrA* and *gyrB*. Topoisomerase lV, encoded by *parC* and *parE*, is involved in the partition of the chromosomes after replication.

The synthesis of new DNA requires the coordinated actions of 25-30 proteins which form the replisome. The two DNA strands are synthesized in a different manner because of the antiparallel orientation of the two DNA template strands and also because of the fact that DNA Pol lll elongates DNA only in the 5'->3' direction (Figure 1.6). Polymerisation is continuous for the leading strand and discontinuous for the lagging strand by formation of Okazaki fragments of 1-2kb in length.

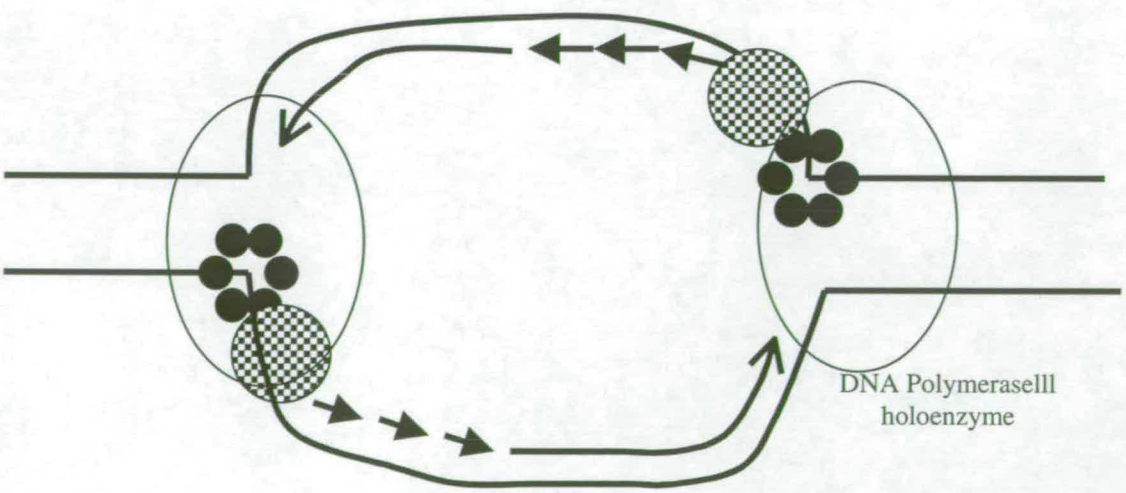


Figure 1.6 Bidirectional replication of the *E. coli* chromosome

ε (proofreading), and θ (unknown function, probably stimulates the proofreading activity of the ε) subunits. 2) a dimer of the τ subunit, which serves to bridge two core complexes; 3) a processivity clamp, which consists of a homodimer of the β subunit that tethers the core complex to the DNA; and 4) the five-subunit (γ, δ, δ', ψ, and χ) clamp loading complex termed the γ complex. The γ complex serves triple duty; in addition to actively recycling β clamps during replication of the lagging strand by loading a clamp at the start of each Okazaki fragment, it also interacts with SSB protein (Glover & McHenry, 1998). This interaction serves a critical role in displacing primase bound to the nascent RNA that serves as primer for each Okazaki fragment, thereby allowing the binding of Pol III for subsequent elongation (Yuzhakov *et al.*, 1999).

Table 2. DNA Polymerase III holoenzyme subunits

Subunit	Encoding gene	Function
α	*dnaE*	5'->3' DNA Polymerase
ε	*dnaQ (mutD)*	3'->5' exonuclease activity; proofreading
θ	*holE*	Stimulates ε
τ	*dnaX*	Dimerises the αεθ complexes; stimulates DnaB helicase
γ	*dnaX*	γ, δ and δ' subunits form a complex which loads the β ring in an ATP-dependent manner

δ	*holA*	
δ'	*holB*	
χ	*holC*	Accessory protein which attaches to SSB
ψ	*holD*	Bridge between subunits γ and χ
β	*dnaN*	Ensures the high processivity of the polymerase

Priming of the lagging strand in *E. coli* is performed by DnaG primase. This enzyme is targeted to the replication fork by a direct interaction with the moving replicative helicase, DnaB. It then synthesizes a short primer, to which it remains bound. The interaction between primase and the nascent primer is stabilized by its association with SSB. Interaction of the γ complex of Pol lll with SSB serves to displace primase from the primer-template junction, thereby permitting γ complex to load a β clamp. Subsequent translocation of the Pol lll assembly to the newly primed Okazaki fragment leads to elongation. When Pol lll reaches the previously synthesized Okazaki fragment, the process is repeated. Pol l with its associated 5'->3' activity, acts together with DNA ligase to carry out the task of Okazaki fragment maturation (Lee & Kornberg, 1992).

The termination step of replication involves a large number of events that result in the separation and segregation of the two chromosomes. Replication finishes when the two replication forks, which started at *oriC* and travelled along the chromosome in opposite directions, meet in the terminus region. When the replication forks reach the terminus region, they are stopped by *Ter* sites. Ten

different *Ter* sites (named *TerA to TerJ*) have been identified *in vivo*. They are divided into two groups:

1) sites which stop the fork advancing anti-clockwise: *Ter A, D, E, I, H*;

2) sites which stop the fork advancing clockwise: *Ter B, C, G, F, J.*

As mentioned earlier, the replication fork is blocked or stalled when it meets a DNA lesion which could have been caused by a variety of factors. If the lesion is situated on the template for the lagging strand, DnaG can start the synthesis of an Okazaki fragment and then the DNA Pol lll will stop at the lesion point, leaving behind a DNA region that was not replicated. Considering that the synthesis of the lagging strand is discontinuous, (meaning that it restarts for each new Okazaki fragment), the replication fork can advance, in spite of the existence of a shorter Okazaki fragment. If the lesion is situated on the template for the leading strand, DNA Pol lll stops at the lesion site. DnaB separates itself from the polymerase to continue the denaturation of the two strands. In conclusion, a lesion occurring on the template for the leading strand has more serious effects than when it occurs on the template for the lagging strand (Figure 1.7).

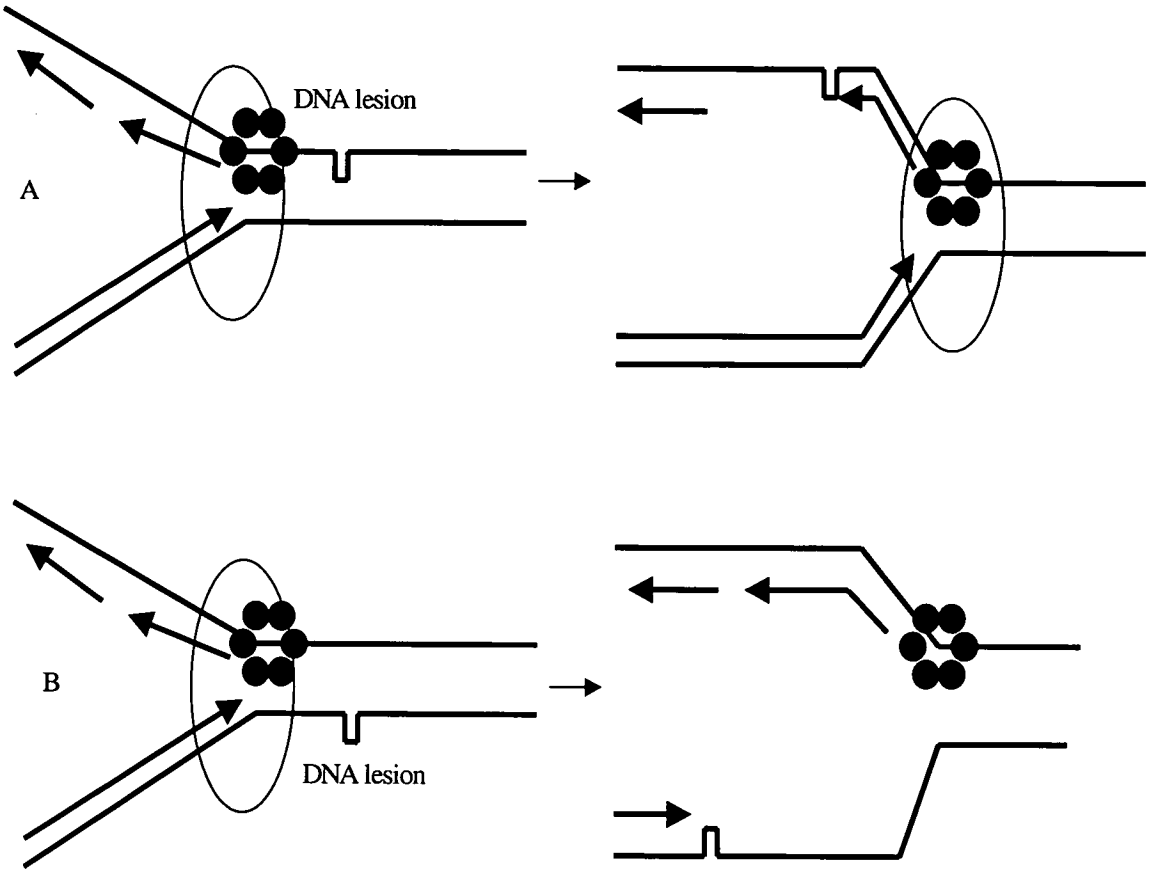


Figure 1.7. A) The DNA lesion occurs on the lagging strand template: the synthesis of the Okazaki fragment is stopped. This results in a discontinuity since a big DNA fragment will not be synthesized until the restart of a new Okazaki fragment. B) The DNA lesion is situated on the template for the leading strand. The replication is stopped; DNA Polymerase III (elipse) falls off; DnaB (hexamer) continues to denature the DNA double helix. DNA lesions occurring on the leading strand template have more dramatic consequences compared to the ones occurring on the template for the lagging strand during replication.

The rate of spontaneous substitution mutations is around 10^{-10} per replicated base. This high fidelity of bacterial DNA replication is the result of a multistage mechanism of error prevention (Loeb & Kunkel, 1982):

1) discrimination by the DNA polymerase against insertion of incorect nucleotides;
2) exonucleolytic removal of incorected inserted nucleotides by the 3'->5' exonuclease associated with the DNA polymerase (editing or proofreading);
3) DNA mismatch repair, which recognises and corrects mismatches shortly after replication.

Replicative DNA polymerases are extremely precise enzymes and can duplicate the sequence of the *E. coli* genome within approximately 45 min with an error rate of 10^{-5} per base (Kunkel, 1992). The high precision with which complementary nucleotides are added to the end of the primer strand is guaranteed by the formation of Watson-Crick base pairs in the active site of the enzyme. When a non-complementary nucleotide is incorporated at the end of the primer, extension from such a primer is highly inefficient (Benkovic & Cameron, 1995). The inability of DNA polymerases to go forward in such situations leads to a translocation of the mispaired primer terminus into the active site of the proofreading 3' ->5' exonuclease activity that is associated with all replicative polymerases and that adds a further two orders of magnitude to the fidelity of the replication process. The gene whose product is essential for proofreading is *mutD (dnaQ)*, which is located at about 5 min on the genetic map. *dnaQ* is the genetic locus for the ε subunit of the

DNA Polymerase III holoenzyme which carries the 3'->5' proofreading exonuclease of this enzyme (Echols *et al.*, 1983).

Occasionally, a mispair manages to elude the proofreading process. This happens mostly with the G/T wobble pair, which is stabilised by two hydrogen bonds and brings about only a slight distortion of the double helix (Hunter *et al.*, 1987). As it is the one mispair from which most polymerases are able to extend with the least difficulty, it also happens to escape the translocation into the proofreading site and exits the polymerase complex. A similar situation can arise when the primer and template strands slip with respect to one another (Kunkel, 1990). Such events occur relatively frequently in runs of repeated mono- or dinucleotides, where they give rise to loops containing extrahelical nucleotides, termed insertion/ deletion loops (IDL). When an IDL is formed by transient dissociation and reassociation of the primer and the template, or by simple slippage, the end of the primer strand will anneal with the template to produce a hydrogen-bonded terminus from which the polymerase can extend (Kunkel, 1993). Replication-associated transactions such as mispairs and IDLs that have escaped the proofreading exonuclease become substrates for mismatch repair (see section 1.9), whose task is to restore the information contained in the template strand.

In addition to their defect in proofreading, *mutD* strains are defective in mismatch repair. However, MutD is not thought to play a direct role in mismatch repair because transformation of *mutD* strains with multicopy plasmids expressing the *mutH* or *mutL* gene restores mismatch repair (Schaaper & Radman, 1989). These observations suggest that the mismatch repair deficiency of *mutD* strains results from a saturation of the MutHLS-mismatch repair system by an excess of primary DNA

replication errors due to the proofreading defect. The number of MutHLS complexes per cell is low (Lu *et al.*, 1984) and a 100-fold or more enhancement of DNA replication errors may overwhelm the system's capacity.

1.9 DNA Mismatch Repair in *E. coli*

Methyl-directed mismatch repair in *E. coli* improves the fidelity of DNA replication by a factor of 10^2 to 10^3 (Glickman & Radman, 1980). The removal and repair of replication errors, both mispaired and unpaired bases, results from excision of the newly synthesised strand including the misincorporated residue, followed by resynthesis, copying the parental strand. Full methylation of GATC sequences prevents mismatch repair. Transient undermethylation of GATC sequences makes the newly synthesised strands susceptible to MMR enzymes. The undermethylated GATC sequences are subsequently methylated by the *dam*-methylase at the N6 position of the adenine residue (Lyons & Schendel, 1984).

The proteins participating in MMR can be divided into two categories: those that are dedicated exclusively to mismatch correction (MutSLH polypeptides) and those that participate in other DNA metabolic pathways as well (helicase, exonuclease, polymerase, ligase).

The initial step of the correction cascade is mismatch recognition fulfilled by the homodimeric MutS protein (Su & Modrich, 1986a). MutS binds with high affinity to substrates containing base-base mispairs and IDLs up to four extrahelical nucleotides (Parker & Marinus, 1992). Malkov (Malkov *et al.*, 1997) demonstrated

that the MutS interaction with DNA is mediated via Phe39, which is situated within a short, but highly conserved motif in the N-terminal domain of the protein. Substitution of this residue for Ala decreased the affinity of the mutant protein for DNA by three orders of magnitude, but the mutation did not affect the ability of the protein to dimerize, nor did it alter its ATPase activity. This suggests that MutS and its homologues interact with DNA via their N-termini, while the C-terminal domains house the dimerization and ATP-binding domains.

Following mismatch recognition, the mismatch-bound MutS homodimer undergoes an ATP-driven translocation along the DNA (Allen *et al.*, 1997), such that a looped structure is formed, with the proteins sitting at its base (Figure 1). This structure is sometimes referred to as the α-loop or Ω-loop according to some authors (Jiricny, 1998). This process is accelerated by the addition of the MutL homodimer, which co-localizes with MutS.

Electron microscopic data (Allen *et al.*, 1997) imply that the MutS homodimer sitting at the base of the loop interacts with two double helices, i.e. that each subunit binds to the same homoduplex molecule. Gel-shift experiments (Drotschmann *et al.*, 1998) showed that the presence of MutL in the MutS binding assay increased the efficiency of MutS binding. MutL protein assists the loading of MutS onto the mismatched substrate, without being itself a part of the complex. MutL has the role of "molecular matchmaker" since it mediates the interaction between MutS and MutH (Modrich, 1991).

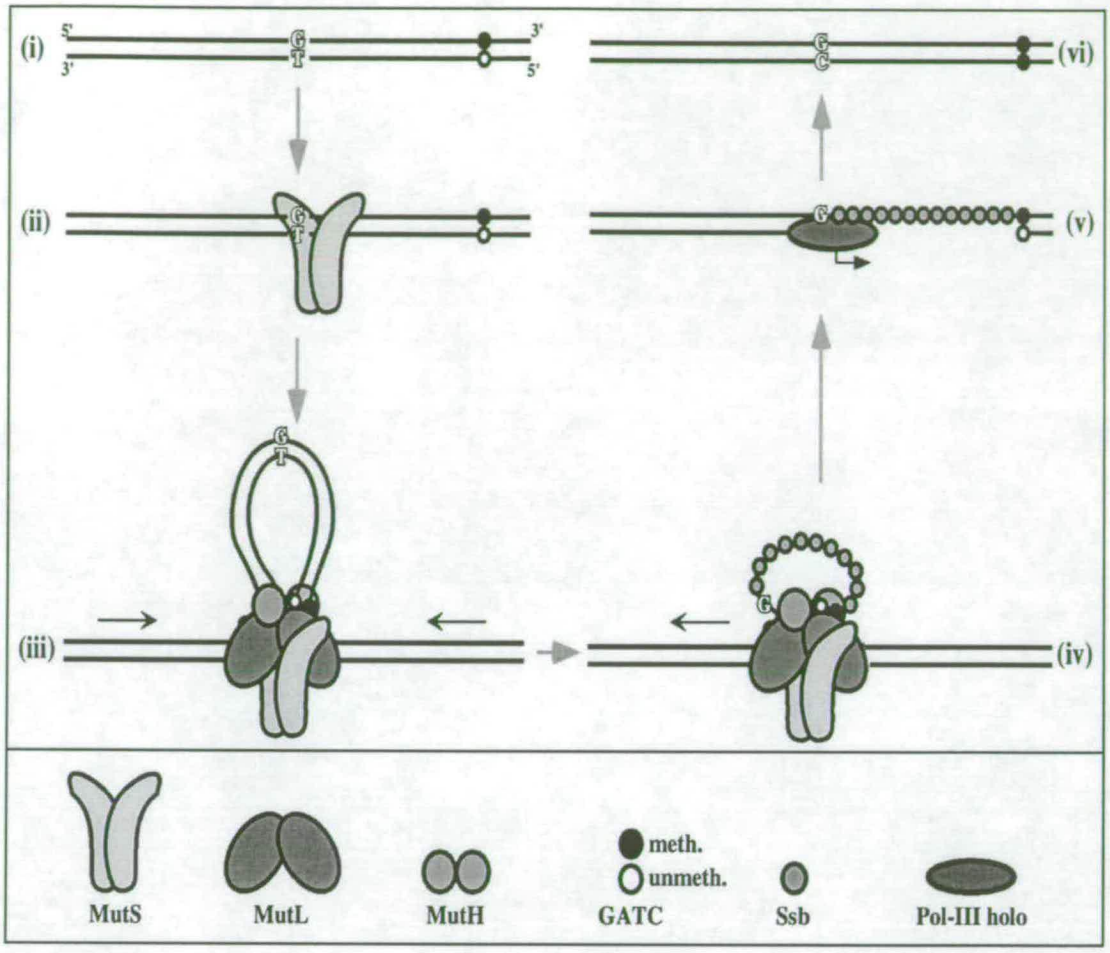


Figure 1.8. Mechanism for mismatch correction in *E. coli*. i) The G/T mispair that arose, as an error of DNA polymerisation is present in a DNA heteroduplex, which is transiently unmethylated at a GATC sequence in the newly replicated strand. The template GATC site is methylated. ii) Binding of the mismatch by the homodimeric MutS protein initiates a cascade of events that starts with an ATP-dependent conformational change of MutS, followed by a movement of the bound protein away from the mispair. iii) ATP hydrolysis drives bi-directional movement of the DNA through the bound MutS in the direction of the arrows, as well as the assembly of a multiprotein complex containing the MutS and MutL

homodimers and probably also two molecules of the strand discrimination factor MutH, all bound at the base of a looped structure. The assembly of the complex activates the endonucleolytic activity of MutH, which cleaves the newly synthesised DNA 5' from the unmethylated GATC sequence. iv) The cleaved strand is then degraded from the nick up to and slightly past the mismatch site by *Exo*VII, *Rec*J or *Exo*I. The single-stranded region thus generated is protected by SSB. v) Polymerase lll holoenzyme fills the gap and DNA ligase repairs the nick. vi) The process is completed by methylation of the GATC site by Dam methylase, at which point the substrate becomes refractory to further action by the MMR system (Jiricny, 1998).

Measurement of the loop size and of the mispair-to-terminus distances suggested that translocation results in a release of the mispair by the protein, such that it is now at the apex of the loop, as shown in Fig. 1.8. Binding experiments have demonstrated that the affinity of MutS for a homoduplex is several fold lower than for a mismatch-containing heteroduplex (Su & Modrich, 1986b).

MutS together with MutL, and in the presence of ATP, activate a third MMR protein, MutH endonuclease, that cleaves the transiently unmethylated daughter strand at hemimethylated GATC sequences. MutH confers strand specificity on MMR since the newly replicating strand containing the error is targeted for repair. The crystal structure of MutH reveals that it functions as a monomer and is evolutionarily related to type-II restriction endonucleases (Ban & Yang, 1998).

MutL also plays a role in loading the UvrD (MutU) helicase at the site of the nick (Dao & Modrich, 1998; Yamaguchi *et al.*, 1998). This is one of the key steps of the repair process, since it helps the helicase initiate the unwinding process.

MMR in *E. coli* targets seven of eight possible base-base mismatches, C·C mismatches being refractory to this system. MMR targets loops of 1-3 bp in length. In addition to mismatch recognition, MutS proteins possess an intrinsic ATPase activity and are members of the ABC (ATP binding cassette) transporter superfamily of ATPases (Gorbalenya & Koonin, 1990).

A hallmark of MMR is the bidirectionality of repair. In *E. coli*, the hemimethylated GATC site can reside on either side of the mismatch at distances up to 1kb (Modrich & Lahue, 1996). Because of this feature, single-stranded exonucleases with 5'-> 3' and 3'-> 5' polarities should be required. *In vitro* studies demonstrated that MMR requires ExoI (3'-> 5'), ExoVII (3'-> 5'), RecJ (5'-> 3') or ExoVII (5'->3') depending on the orientation of the nick relative to the mismatch. However, exonuclease-deficient strains show only modest mutator phenotypes (Viswanathan & Lovett, 1998).

The MutSLH proteins are involved not only in postreplicative MMR, but also in at least two other contexts. MMR of heteroduplex DNA occurs during homologous recombination between similar but not identical sequences leading to mismatch correction or abortion of the recombination process (Evans & Alani, 2000). MMR also occurs in the context of damaged or altered bases arising from modification by a variety of genotoxic agents including alkylating agents or bulky adducts (Harfe & Jinks-Robertson, 2000).

1.10 Recombination in *E. coli*

The boundaries that once separated the fields of DNA replication, recombination, and repair have become increasingly blurred in the last few years. Recent advances in each of these three fields have not only illuminated the molecular mechanisms of the individual processes, but have also provided significant insights into their interrelatedness and codependence. For example, recent studies indicate that the *Escherichia coli* RecA protein is not only required for homologous recombination, but is also required for efficient chromosomal DNA replication even under normal growth conditions, as well as for the regulation of cellular responses to DNA damage and the replication of damaged DNA (Sutton & Walker, 2001).

The genome of an organism can be modified by point mutations, such as base replacements or frameshifts, or by large DNA rearrangements, such as deletions, insertions, translocations or duplications. Genome rearrangements are important in evolution, medicine and biotechnology. Duplications provide supplementary copies of genes, which can accumulate mutations and thus evolve. Translocations and deletions alter the environment of a gene and can thus contribute to its integration within novel control circuits. Insertions of foreign material into a genome facilitate horizontal gene transfer and thereby bypass the need for similar functions to evolve repeatedly in different organisms.

Homologous recombination represents the exchange of homologous DNA sequences between two molecules or between two regions of the same molecule. Homologous recombination is one of the main pathways of repairing DNA breaks. It is also involved in the restart of blocked replication forks. The process of homologous recombination involves three stages: presynapsis, synapsis and

postsynapsis. The most important step in homologous recombination is the formation of a heteroduplex DNA region where exchange between two homologous sequences has taken place (Kowalczykowski, 2000). SSB protein binds to the single-stranded DNA region formed as a result of the action of helicases (RecQ, UvrD or helicase IV) or of the exonuclease/helicase complex RecBCD or of nucleases (RecJ, RecE), which facilitate the opening of the double helix. The presynaptic proteins RecBCD and RecFOR then facilitate the loading of RecA monomers on the single-stranded DNA, by facilitating the displacement of SSB (Kantake *et al.*, 2002).

The RecA nucleo-protein filament thus formed allows the association of two DNA molecules during synapsis at homologous regions, to form a Holliday junction.

The postsynaptic steps of homologous recombination consist in the separation and repair of the DNA strands. The Holliday junction formed at a region of exchange of genetic material, can be moved by branch-migration along the DNA molecule. This migration is catalysed rapidly by RuvAB or RecG proteins (McGlynn & Lloyd, 2001). RecA also can catalyse branch-migration, but not as efficiently as RuvAB or RecG (Mahdi *et al.*, 1996). RuvC resolves Holliday junctions by cleaving two of the four strands. RuvC is an endonuclease, which recognises Holliday junctions and the RuvAB complex (Muller & West, 1994). The breaks are repaired by DNA ligase and two heteroduplex DNA molecules are generated after the genetic exchange. The main genes involved in homologous recombination are listed in Table 3.

Table 3. The main genes involved in homologous recombination in *E. coli*

Gene	Biochemical activities	Functions
recA	Binds to ssDNA, forms RecA filaments (RecA-ssDNA), ATPase	Matches and exchanges homologous strands
recB *recC* *recD*	RecBCD complex: exonuclease (exonuclease V), endonuclease, helicase, ATPase, recognises the Chi sequence	Generates a 3'-OH extremity of the ssDNA; Assists loading of RecA on ssDNA coated with SSB
recF *recO* *recR*	Binding to ssDNA, binding ATP	Assists the loading of RecA on ssDNA coated with SSB
ruvA *ruvB* *ruvC*	Binding to Holliday junctions, 5'->3' helicase, endonuclease	Recognises and assists the migration of the Holliday junction DNA strands, resolves Holliday junctions
recG	Binds to DNA, 3'->5' helicase	Recognises Holliday junctions, insures the migration of DNA strands in Holliday junctions

Clark (Clark, 1973) proposed the concept of recombinational pathways, analogous to pathways of small molecule metabolism, such as those of amino acid biosynthesis. The RecBCD pathway is predominant in wild-type cells; the RecE pathway in *recBC sbcA* cells; the RecF pathway, in *recBC sbcD* cells. These pathways are not entirely independent, but they overlap, since all three require the *recA* protein (Smith, 1988). The RecBCD pathway is responsible for homologous recombination accompanying conjugation and transduction, and for repair of double-strand breaks through recombination between the damaged DNA and an intact homologous chromosome (Kogoma, 1996). These potentially fatal breaks can arise during DNA replication or as a direct result of high energy irradiation (Friedberg, 1995).

Genetic recombination is regulated by special sites. Recombination sites influence the frequency with which exchange occurs. Recombination-promoting proteins recognise these sites and act preferentially at or near them. Chi sites (*cross-over hotspot instigator*) and the recombination enzymes associated with their action provide an example of this type of regulation.

Chi sites stimulate the RecBC pathway, but have no detectable effect on the other recombination pathways tested (Smith, 1983). Chi sites occur in the *E. coli* chromosome at an average density of 1 per 5kb (about 1,000 sites per chromosome). Chi stimulates exchange maximally near itself and with decreasing frequency more than 10kb from itself. Stimulation is greater to one side of Chi (to the left on the conventional λ map) than to the other side (Stahl *et al.*, 1980). Chi is specified by the nucleotide sequence 5'-GCTGGTGG-3' or its complement or the duplex (Smith, 1981).

RecBCD enzyme, or exonuclease V, has multiple enzymatic activities. It has several activities *in vitro*:

1) helicase;
2) ATP-dependent endonuclease and exonuclease on double stranded DNA;
3) Endonuclease on single-stranded DNA;
4) DNA-dependent ATPase.

The RecBCD complex has three subunits: RecB, RecC and RecD (Kowalczykowski *et al.*, 1994) encoded by the genes *recB, recC and recD.* The RecBCD complex binds a double strand DNA end and forms an initiation complex (Anderson & Kowalczykowski, 1997) (Figure 1.9). This step is ATP-independent. Then RecBCD travels along the DNA by denaturing it thanks to its ATP-dependent helicase activity. RecBCD degrades DNA with a 3'->5' exonuclease polarity (Dixon & Kowalczykowski, 1993) until it reaches a Chi site. Upon meeting a Chi site, RecBCD changes its direction and degrades DNA with a 5'->3' polarity. This produces a single-stranded DNA region with a 3'-OH end. The ssDNA region is the substrate for loading the RecA protein followed by the formation of the RecA filament. Chi-activated RecBCD complex coordinates and facilitates the loading of the RecA protein on the ssDNA (Arnold & Kowalczykowski, 2000). This step is essential for RecBCD-mediated homologous recombination *in vivo*. It was shown that the mutant $RecB^{2109}CD$ enzyme is unable to coordinate the loading of RecA protein onto the ssDNA generated. This inability is the cause of the recombination-defective phenotype of the $RecB^{2109}$ allele.

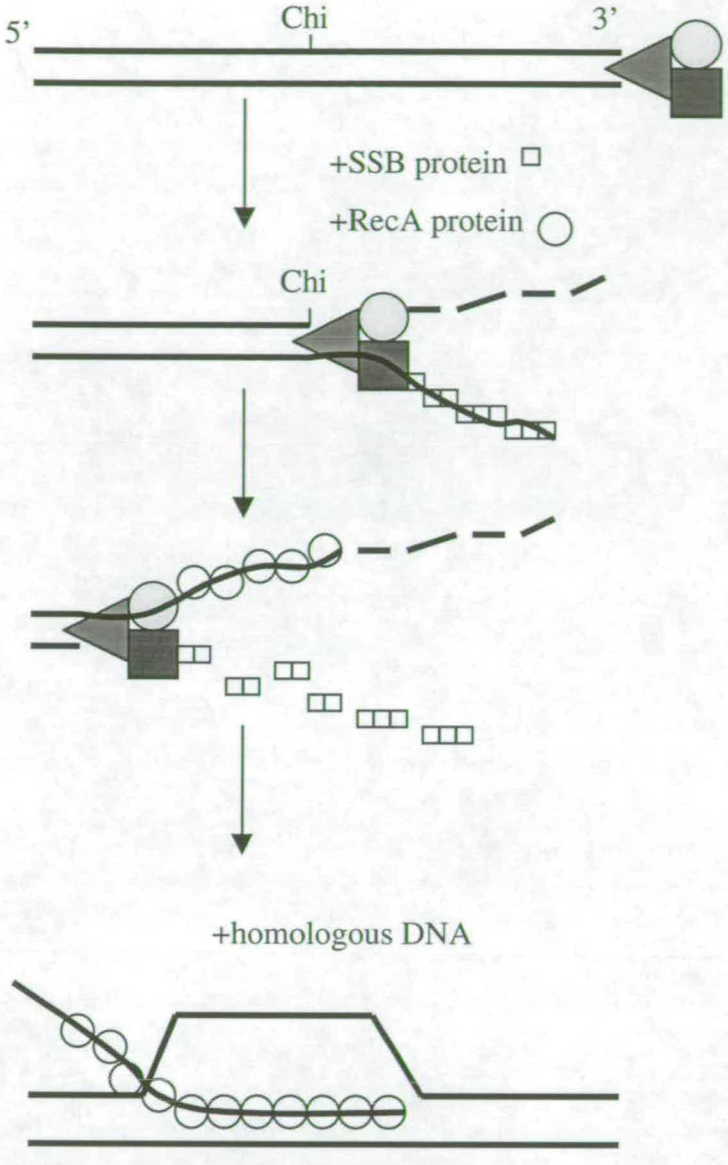


Figure 1.9. Model for the RecBCD-Chi recombination. RecBCD complex (red, yellow and blue subunits) binds to a dsDNA end; then travels through DNA digesting it with a 3'->5' polarity. When it reaches Chi the polarity is reversed to 5'->3' orientation. RecA protein (white circle) binds to the 3' single-stranded overhand coordinated by RecBCD. This results in the formation of a presynaptic filament which initiates strand invasion into a homologous duplex DNA molecule.

RecA can bind to ssDNA resulting in a filament, which represents the active RecA conformation. RecA can induce the renaturation of two single-stranded complementary DNA strands. RecA also catalyses the matching of homologous sequences and the exchange of complementary strands (Clark, 1991). SSB protein stimulates RecA presynaptic and postsynaptic activities by binding to single-stranded DNA and avoiding the formation of secondary structures (hairpins).

The higher residual levels of recombination following conjugation in *recBC* null mutants than in *recA* mutants suggested that *E. coli* has at least one other low-level pathway of recombination dependent upon RecA protein, but independent of RecBCD complex. Suppressor mutations designated *sbc* (for suppressor of *recBC*) were isolated. They increase the activity of the alternative pathways and thereby restore recombination proficiency to *recBC* mutants. It was found that in a *recBCD* strain, homologous recombination can be mediated by the RecF, RecO and RecR proteins via the RecF pathway (Ivancic-Bace *et al.*, 2003). RecO, RecR and probably RecF proteins are part of the RMP class (Recombination/Replication Mediator Proteins) (Beernink & Morrical, 1999).

Holliday junctions created by RecA are efficiently resolved by the RuvABC complex and the RecG protein. Bacterial strains deficient in *ruvA, ruvB or ruvC* are UV-sensitive and partially deficient in homologous recombination (Sharples *et al.*, 1990). RuvA and RuvB ensure the branch-migration of Holliday junctions. RuvC resolvase cleaves Holliday junctions. RuvA, B and C are interdependent: RuvC cannot resolve the recombination intermediates in the absence of RuvAB. RuvAB and RecG act together in the postsynaptic stage of homologous recombination (Kuzminov, 1996). *ruvAB recG* double mutants are much more sensitive to UV

radiation compared to single mutants *ruvAB* or *recG*. RecG is a DNA-dependent ATPase and helicase which binds to Holliday junctions and contributes to their resolution, perhaps by branch-migration to nicks in the DNA.

1.11 Processing of DNA Secondary Structures in *E. coli*

SbcC mutants were initially isolated as cosuppressors of recombination deficiency in *recBC* strains of *E. coli* (Lloyd & Buckman, 1985). SbcC (C suppressor of *recBC*) is co-transcribed with SbcD (D suppressor of *recBC*). They constitute the primary control for the replication of long palindromes in *E. coli*. SbcC and SbcD polypeptides have an ATP-dependent double-strand exonuclease activity and an ATP-independent DNA single-stranded endonuclease activity. SbcCD also acts as a hairpin endonuclease, cleaving hairpin loops near the 5' junction with the duplex stem of the secondary structure (Connelly *et al.*, 1998). Although both SbcC and SbcD polypeptides are required for the ATP-dependent double-strand exonuclease activity of SbcCD, only the SbcD polypeptide is required for the ATP-independent single-stranded endonuclease activity of SbcCD (Connelly & Leach, 1996).

The SbcC polypeptide belongs to the family of chromosome condensation and segregation proteins which includes SMC1 and SMC2 of *Saccharomyces cerevisiae* (Gasser, 1995), MukB protein of *E. coli* and RAD50 (recombination and repair protein of yeast). These polypeptides form a head-rod-tail structure with two globular domains linked to each other by a long filamentous coiled-coil region.

Secondary structures formed during DNA replication are cleaved by the SbcCD enzyme, resulting in a broken chromosome. The chromosome is then

repaired by homologous recombination. This was demonstrated by the requirement for the genes *recA*$^+$, *recB*$^+$ and *recC*$^+$ for the viability of an *sbcCD*$^+$ cell containing a 246 bp palindrome (Leach *et al.*, 1997) in experiments where *sbcCD*$^+$ strains were lysogenised with a palindrome-containing λ phage. The observation that mutations inactivating either RecA or RecBCD proteins have similar effects argues strongly that it is homologous recombination *per se* that is required for cell viability.

G. Cromie (Cromie *et al.*, 2000) showed that the genes of the RecF pathway (*recF*$^+$, *recO*$^+$, *recR*$^+$, *recJ*$^+$, *recQ*$^+$, *recN*$^+$) are also essential for viability in the presence of a chromosomal palindrome and *sbcCD*$^+$, but the gene *priA*$^+$ was not. The requirement for genes in both the RecBCD and the RecF pathways of recombination and the lack of requirement for PriA (needed for the loading of DnaB to form normal replication forks) have led to the model in Figure 1.10.

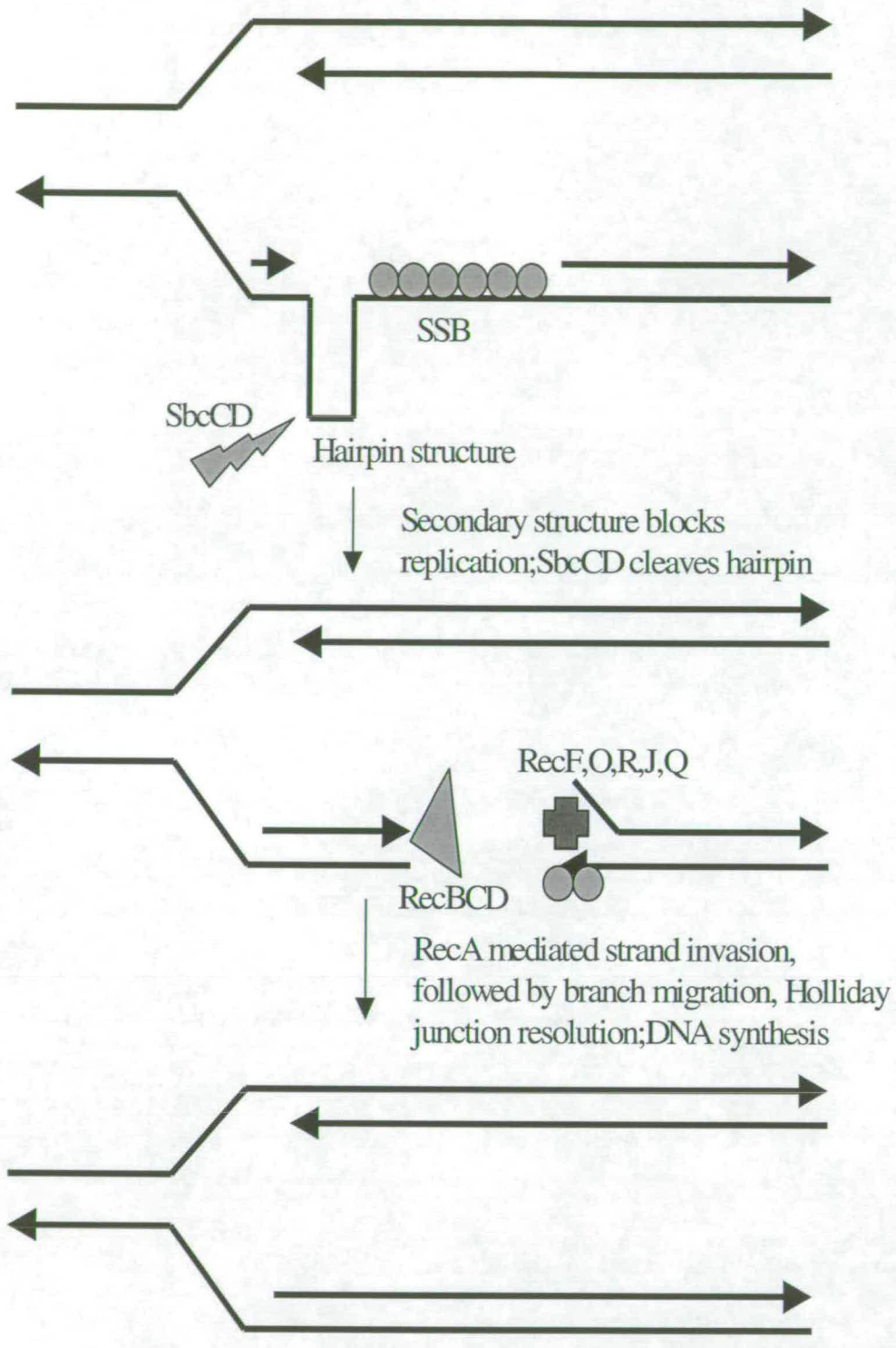


Figure 1.10. Recombinational repair of SbcCD-induced double-strand breaks.

1.12 Investigation of trinucleotide repeat instability in the *E. coli* chromosome

Trinucleotide repeat expansions are the cause for almost twenty neurodegenerative diseases but the mechanisms underlying the DNA instability are not known. The aim of my project was to study the behaviour of $(CAG)_{43}$ and $(CTG)_{43}$ in the *Escherichia coli* chromosome. The expectation was that this investigation would provide clues regarding the fundamental mechanisms that lead to expansion of trinucleotide repeat arrays. *E. coli* was considered a good organism in which to study the nature of the instability mechanisms because the pathways of recombination, replication and mismatch repair and the proteins involved in these pathways are well defined and characterized (Chapter 1.8-1.11).

The work in this thesis investigates trinucleotide repeat instability in bacterial strains deficient in genes with known roles in recombination, repair or replication. The expectation was that this system could distinguish between mechanisms based on DNA synthesis as opposed to mechanisms based on homologous recombination. Similarly, this system could distinguish between other mechanisms that have been proposed to play a part in TREDs, such as flap processing, mismatch repair and proofreading.

Investigating trinucleotide repeat instability in the *E. coli* chromosome avoids the difficulties associated with its study in plasmids, where selection for shorter repeat lengths complicates the analysis of repeat array length changes.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Microbiological Strains, Media and Solutions

2.1.1.1 Strains of Bacteria

Table 2-1. Strains of *E. coli* used in this work

Designation	Genetic Background	Notes	Reference/ Source
N2677 (DL513)	As AB1157, but *pro*$^+$	1	(Lloyd *et al.*, 1987)
N2679 (DL515)	As N2677, but *sbcC201*	1	(Naom *et al.*, 1989)
DL902	*mutS*::Tn*10*	2	D. Leach
DL936	*mutL*::Tn*10*	2	D. Leach
DL1179	*mutH471*::Tn*5*	2	(Schmidt *et al.*, 2000)
DL1116	As N2677, but *recG263*::*Kan*	1	D. Leach
BW1161 (DL1126)	As N2677, but *nfi1*::*Cm*	1	(Cunningham *et al.*, 1986)
DB1318 (DL654)	*recD1014, recA*::*Cm*		(Wertman *et al.*, 1986)
N2101 (DL974)	As N2677, but *recB268*::Tn*10*	1	(Lloyd *et al.*, 1987)
RH6972 (DL962)	*mutD*::miniTn*10*		Genevieve Maenhaar-Michel

Notes:

1. These strains are derivatives of AB1157 (Howards-Flanders and Theriot, 1966), F⁻ *hisG4 argE3 Δ(gpt- proA)62 thr-1 thi-1 leuB6 kdgK51 rfbD1 mgl-51 ara-14 lacY1 galK2 xyl-5 mtl-1 tsx-33 supE44 rpsL31* (Str^r).
2. These strains are derivatives of JM83 (Yanisch-Perron et al. 1985), F⁻ *ara Δ(lac-proAB) rpsL [ϕ80dlac Δ(lacZ)*M15] (Str^r).

2.1.1.2 Media

BBL Agar

10g Trypticase (Baltimore Biological Laboratories), 5g NaCl, 10g Bacto-agar (Difco) per litre, adjusted to pH 7.2 with NaOH.

BBL Top Agar

Same as BBL agar, but made with 6.5g Bact-agar (Difco) per litre.

LB Agar

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl, 15g Bacto-agar (Difco) per litre, adjusted to pH 7.2 with NaOH.

Lc Agar

10g Tryptone, 5g yeast extract, 5g NaCl and 10g Difco-agar per litre. The pH was adjusted to 7.2 using NaOH.

Lc Top Agar

As Lc agar, but containing 7g Difco-agar.

L Broth

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl per litre, adjusted to pH 7.2 with NaOH.

Phage Buffer

3g KH_2PO_4 ,7g HPO_4 , 5g NaCl, 1mM $MgSO_4$, 1mM $CaCl_2$,1% (w/v) gelatine.

2.1.1.3 Media Additives

500mM $CaCl_2$ stock

Made up in sterile, distilled water, autoclaved.

1M $MgSO_4$

Made up in sterile, distilled water, autoclaved.

20% (w/v) Glucose Stock

Made up in distilled water, filter sterilised.

Ampicillin (stock concentration 100mg ml^{-1})

Ampicillin (Beecham Pharmaceuticals) was used at 50µg ml^{-1} and stored at -20° C.

Chloramphenicol (stock concentration 50mg ml^{-1})

Chloramphenicol (Sigma Chemical Company) was made up in 100% ethanol and used at 25 µg ml^{-1}. It was stored at -20°C.

Kanamycin (stock concentration 50mg ml^{-1})

Kanamycin (Sigma Chemical Company) was made up in distilled sterile water and used at 50µg ml^{-1}. It was stored at -20°C.

Tetracycline (stock concentration 15mg ml^{-1})

Tetracycline (Calbiochem) was made up in 50% ethanol and used at 13 µg ml^{-1}. It was stored at -20°C.

2.1.1.4 Solutions for Microbiological Methods

Solutions for Transformation of *E. coli*

$CaCl_2$ (100mM)

Made up in distilled water, autoclaved.

MOPS-Glycerol

100mM MOPS [3-(4-morpholinyl) 1-propanesulfonic acid]- NaOH (Sigma Chemical Company, pH 6.5), 50mM $CaCl_2$, 20% (v/v) glycerol.

2.1.2 Materials for DNA Purification and Manipulation

2.1.2.1 General Solutions

500mM EDTA Stock

500mM EDTA (Sigma Chemical Company), adjusted to pH 8 using glacial acetic acid, autoclaved.

1M Tris-HCl Stock

1M Tris base, adjusted to pH 7.5 using concentrated HCl, autoclaved.

10mM Tris-HCl

10mM Tris base, adjusted to pH 8 using concentrated HCl, autoclaved.

10 x TBE (Tris-Borate-EDTA-Buffer) pH 8

54g Tris-base, 27.5 Boric acid (Fisher Scientific), 20 ml of 500mM EDTA per litre.

10 x Tris-EDTA buffer

100mM Tris-base (Sigma Chemical Company), 10mM EDTA, adjusted to pH 7.5 with concentrated NaOH, autoclaved.

3M Sodium Acetate (pH 5.3)

0.19 volumes of sterile 3M acetic acid were added to 0.81 volumes of sterile 3M sodium acetate, autoclaved.

BSA (Bovine Serum Albumin)

BSA (New England Biolabs) was supplied at a concentration of 20mg ml^{-1} and stored at -20°C.

2.1.2.2 Solutions for DNA Preparation

2.1.2.2.1 Solutions for Preparation of Plasmid DNA from *E. coli*

The following solutions supplied in the QIAGEN Plasmid Maxi and QIAGEN Plasmid Mini Spin Kit were used for the preparation of plasmid DNA from 100-ml and 5-ml overnight cultures of *E. coli*.

Resuspension Buffer P1

50mM Tris-HCl (pH 8.0), 10mM EDTA, 100µg ml^{-1} RnaseA, stored at 4°C.

Lysis Buffer P2

200mM NaOH, 1% (w/v) SDS (sodium dodecyl sulphate), stored at room temperature.

Neutralisation Buffer N3

3M potassium acetate, pH 5.5

Equilibration Buffer QBT

750 mM NaCl, 50mM MOPS (pH 7.0), 15% (v/v) ethanol.

Elution Buffer QF

1.25 M NaCl, 50mM Tris-HCl (pH 8.5), 15% (v/v) ethanol.

2.1.2.2.2 Solutions for the Preparation of Genomic DNA from *E. coli*

The Bacterial Genomic DNA Purification Kit (EdgeBio Systems) was used to extract high molecular weight chromosomal DNA from 5-ml overnight cultures of *E. coli*. Advamax™ beads (1.5% (w/v) latex) and the solutions listed below were supplied by the manufacturer. Information of the composition of solutions was limited.

Spheroplast Buffer

Contains lysozyme, RNAse, Sucrose, Tris and EDTA (pH 8)

Lysis Buffer 1

Contains SDS

Lysis Buffer 2

Contains NaCl

Extraction Buffer

Contains $MgCl_2$

2.1.2.3 Solutions for Gel Electrophoresis

2.1.2.3.1 Agarose Gel Electrophoresis

TBE Gel Electrophoresis Buffer (10 X)

0.9M Tris-borate, 20mM EDTA (pH 8.0)

Agarose Gel Loading Buffer (5X)

20% (v/v) sterile glycerol, 0.05% (w/v) bromophenol blue

Ethidium-Bromide Solution (50X)

1% (w/v) ethidium bromide in sterile, distilled water

2.1.2.3.2 Polyacrylamide Gel Electrophoresis

5% Long Ranger™ Gel Solution for Native Gels (1.2x TBE)

8.4ml of 10x TBE buffer, 7ml of 50% Long Ranger™ gel solution (Flowgen), 54ml of distilled water.

5%Long Ranger™ Gel Solution for Sequencing Gels (1.2x TBE)

8.4ml of 10x TBE buffer, 7ml of 50% Long Ranger ™ Gel Solution (Flowgen), 30g urea, volume adjusted to 70ml with distilled water.

0.6xTBE Gel Running Buffer

60ml of10x TBE buffer stock solution, 940ml of distilled water

10% (w/v) AMPS

10% (w/v) AMPS (ammonium persulphate; Sigma Chemical Company) was freshly prepared in distilled water.

TEMED

TEMED (N-N-N'-N'-tetra-methyl-1,2-diamino-ethane) (Sigma Chemical Company) was stored at 4°C.

Polyacrylamide Gel Loading Buffer (10x)

95% Formamide, 20mM EDTA, 0.05% (w/v) bromophenol blue.

Dextran blue loading buffer

This loading buffer was used for f-TRAMP and was supplied with the internal lane size standard GENESCAN-500™ ROX or TAMRA (Applied Biosystems).

1ml contains: 50mg of Ficoll 400-DL, 1.7mg dextran sulphate, 8.3mg blue dextran, buffered with 2 x TBE.

2.1.2.4 Enzymes and Buffers

2.1.2.4.1 Restriction Endonucleases and Incubation Buffers

All restriction endonucleases, listed in Table 2-2, were incubated with the buffers provided by the suppliers at the temperatures recommended in the manufacturers' instructions.

Table 2-2. Restriction endonucleases used in this work

Enzyme	Cleavage site (5'->3')	Supplier
EcoRI	G/AATTC C TTAA/G	New England Biolabs
Sau3A	/GATC CTAG/	Roche
XmaI	C/CCGG G G GGCC/C	Promega
NotI	GC/GGCC GC CG CCGG/CG	Promega
XhoI	C/TCGA G G AGCT/C	Roche
AvrII	C/CTAG G G GATC/C	New England Biolabs

2.1.2.4.2 Thermophilic and Other DNA Polymerases

Taq DNA Polymerase (5U μl^{-1})

Taq DNA Polymerase (Roche) was used with 1 x PCR reaction buffer as supplied by the manufacturer. Enzyme and buffer were stored at -20°C.

DNA Polymerase I Large (Klenow) Fragment

In this work, Klenow enzyme [supplied concentration: 2U μl^{-1} (Roche)] was used after incubation of plasmid DNA with *EcoRI* restriction endonuclease (20U μl^{-1}) or after incubation of chromosomal DNA with *AvrII* restriction enzyme. Klenow enzyme was added to unpurified reaction mixtures containing *EcoRI* and *1 x EcoRI* restriction endonuclease buffer or *AvrII* and *1 x AvrII* buffer.

2.1.2.4.3 Other enzymes

T4 DNA Ligase

T4 DNA Ligase (Roche) was incubated in the 1 x ligase buffer [30mMTris-HCl pH 7.8, 10mM $MgCl_2$, 10mM DTT and 1mM ATP] supplied by the manufacturer. The buffer was stored in small aliquots at -20°C to minimize degradation of the ATP and DTT because the performance of this buffer depends on the integrity of the ATP

Shrimp Alkaline Phosphatase

Shrimp Alkaline Phosphatase [4U μl^{-1}(United States Biochemical)] was used with *EcoRI* restriction endonuclease buffer (50mM NaCl, 100mM Tris-HCl, 10mM $MgCl_2$, 0.025% (v/v) Triton X-100, pH 7.5).

2.1.2.5 Plasmids and Synthetic Oligonucleotides

Table 2-3. Plasmids used in this work

Plasmid	Description	Notes	Reference/ Source
pDL915	pUC18, but $(CTG)_{43}$ in *EcoRI* site	Marker:Amp Orientation A: CTG repeat on the leading strand during replication	D. Leach
WM2269 pLDR8	Ori pSC101, λ *intI*, *rep Ts*, cI_{857}	Marker:Kan Temperature sensitive (Kan^s at 42°C; Kan^r at 30°C)	W. Messer (Diederich *et al.*, 1992)
WM2153 pLDR9	*OriColE1 rop* *λattP*	Marker:Amp, Kan	W. Messer (Diederich *et al.*, 1992)
WM2155 pLDR11	*OriColE1 rop* *λattP*	Marker:Amp, Kan	W. Messer (Diederich *et al.*, 1992)

Table 2-4. Synthetic oligonucleotides for PCR and primer extension reactions. All oligonucleotides designed in the course of this work were synthesised by OSWEL DNA Service (University of Southampton, UK) and MWG.

ID	Sequence (5'-> 3')	Modification	Reference/ Source
FAM-T2049	GCATCTTGGGAGCATCTTTG	5'-fam dye	K.Schmidt thesis
HEX-T2049	GCATCTTGGGAGCATCTTTG	5'-HEX dye	K.Schmidt thesis
M6833	CCCCTTTCTAGCCTTCTTCA	none	Abbott and Chambers (1994)
M6834	TTTGGTCCAAACGGGATGCT	none	Abbott and

			Chambers (1994)
AVRLEFT	CCGGAATTCCGCCTAGGTCCT CCCCTTTCTAGCCTTC	None	This work
AVRRIGHT	CCGGAATTCATCCTAGGCAGT GAGCCTGCTGCTGAT	None	This work
BAMECO	GATCAGGTACCCCTAGGGAAT TCCCTAGGT	None	This work
ECOBAM	AATTACCTAGGGAATTCCCTA GGGGTACCT	None	This work
PRIMER D	GATAAGCTTGGGCTGCAGGT	None	This work
GENOM1	ATCAGAAGGACGTTGATCGG	None	This work
GENOM2	GCAATGCCATCTGGTATCACT	None	This work
GENOM3	GAGTATTCAACATTTCCGGTG	None	This work
METAPH-LEFT	CAAGCATCTTGGGAGCATCT	None	This work
METAPH-LEFT-FAM	CAAGCATCTTGGGAGCATCT	5'- FAM	This work
METAPH-RIGHT	CAGTGAGCCTGCTGCTGAT	none	This work

2.1.2.6 Radionucleotides

[α-^{35}S] dATP

[α-^{35}S] dATP was supplied by the manufacturer at a concentration of > 1000Ci mmol^{-1} in 5 mM Tris-HCl (pH 7.4-7.5).

2.1.2.7 Materials and Solutions for DNA Amplification

dNTP-4 Stock Solution (10x)

dNTP-4 Stock Solution was a mixture of dATP, dTTP, dGTP and dCTP (Roche), each at a concentration of 2 mM, prepared with sterile distilled water and stored at -20°C.

dNTP-3 Stock Solution (10x)

dNTP-3 Stock Solution was a mixture of dTTP, dGTP and dCTP (Roche), each at a concentration of 2mM, prepared with sterile distilled water and stored at -20°C.

ddATP Stock Solution (10x)

ddATP was supplied by Roche at a concentration of 100mM. The 10x ddATP stock solution containing 1mM ddATP was prepared with sterile distilled water and stored at -20°C.

Mineral Oil

Sterile mineral oil (Sigma Chemical Company) was stored protected from light at room temperature.

Bio/Optimiser™ Kit for the Rapidcycler

Reaction buffers in this optimisation kit (Bio/Gene Ltd) vary in the concentration of Mg^{2+} (1-5mM). Some buffers contain Ficoll and dye to accommodate direct loading of the reaction mix onto an agarose gel for analysis.

10x BSA

BSA at a concentration of 5mg ml^{-1} is provided in the Bio/Optimiser kit (Bio/Gene Ltd) and is used at a final concentration of 500 $\mu g\ ml^{-1}$.

2.1.2.8 Materials and Solutions for DNA Sequencing

2.1.2.8.1 Manual DNA Sequencing

Manual DNA Sequencing was carried out using the Sequenase® v2.0 Sequencing kit (United States Biochemicals). Primer sequences are shown in Table2-4.

Sequenase Buffer (5x)

200mM Tris-HCl (pH 7.5), 100mM $MgCl_2$, 250mM NaCl

DTT

0.1 M DTT (1,4 dithio-threitol) prepared in distilled water.

Labelling mix (5x)

7.5 μM dGTP, 7.5 μM dCTP, 7.5 μM dTTP

ddG Termination Mix

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddGTP, 50 mM NaCl

ddA Termination Mix

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddATP, 50 mM NaCl

ddT Termination Mix

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddTTP, 50 mM NaCl

ddC Termination Mix

80 μM dGTP, 80 μM dATP, 80 μM dCTP, 80 μM dTTP, 8 μM ddCTP, 50 mM NaCl

Enzyme Dilution Buffer

10mM Tris-HCl (pH 7.5), 5 mM DTT, 0.5 mg ml^{-1} BSA

Stop Solution

95% formamide, 20mM EDTA, 0.05 % (w/v) bromophenol blue, 0.05 % (w/v) xylene cyanol.

2.1.2.8.2 Automated DNA Sequencing

Automated DNA Sequencing was carried out using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Sequencing reactions were analysed on a ABI PRISM 377 DNA Sequencer (Applied Biosystems).

2.1.2.9 Materials and Solutions for Colony Hybridisation

The DIG DNA Labelling and Detection Kit (Roche) was used in a non-radioactive protocol to screen bacterial colonies for the presence of CTG/CAG trinucleotide repeat arrays.

Hexanucleotide Mix (10x)

62.5 A_{260} ml^{-1} in reaction buffer.

dNTP Labelling Mixture (10x)

1mM dATP, dCTP, dGTP (each), 0.65 mM dTTP, 0.35 mM DIG-II-dUTP-alkalilabile.

0.5 N NaOH

Made up by diluting 25 ml of 10N NaOH in 500 ml distilled water.

Neutralisation buffer 1

0.1 M NaOH, 1.5 M NaCl

Neutralisation buffer 2

0.5 M Tris-HCl, 1.5 M NaCl (pH 7.5)

Solution 3

1 x SSC (sodium saline citrate), prepared by 10-fold diluting 20 x SSC.

20 x SSC: 175.3 g of NaCl and 88.2 g of sodium citrate are dissolved in 800 ml of sterile distilled water and adjust pH to 7.0 with 10N NaOH and adjust the volume to 1 l with sterile distilled water.

Easy-Hyb-Hybridisation Buffer

Easy-Hyb (Roche) is a "Ready-to-use" hybridisation solution stored at room temperature.

Wash Buffer 1

2 x SSC (see solution 3), 0.1 % (w/v) SDS (AMRESCO, Ohio)

Wash Buffer 2

1 x SSC, 0.1 % (w/v) SDS

Detection Solution 1 (10 x)

0.1 M maleic acid, 0.15 M NaCl, 0.3 % (v/v) Tween 20.

10 % (w/v) Blocking Reagent

Block solution was made up by mixing 10 g of blocking reagent (Roche) with 100 ml of detection buffer1 and then warming it up to 50-70°C to dissolve. The solution was autoclaved and stored at 4°C.

Detection Solution 2

0.1 M Maleic acid, 0.15 M NaCl, 1 % (w/v) block solution.

Detection Solution 3

1µl of Anti-DIG-AP-Conjugate was added to 30 ml of detection buffer 2. This solution was always made freshly.

Detection Solution 4

0.1 M Tris-base, 0.1 M NaCl, 0.05 M $MgCl_2$, adjusted to pH 9.5 before use.

CDP-Star™ (25mM, 100x)

CDP-Star™ [Disodium-4-chloro-3-(methoxyspiro {1,2- dioxetane- 3,2'-(5'- chloro) tricyclo [3.3.1.1.^{3,7}] decan}-4-yl) phenyl phosphate] was supplied with the DIG-DNA labelling kit (Roche).

Detection Solution 4 + CDP-Star™

Per filter, 0.6 µl of CDT-Star™ were mixed with 265 µl of detection buffer 3.

2.2 Methods

2.2.1 Bacterial Methods

2.2.1.1 Storage of Bacteria

Glycerol stocks *E. coli* strains were prepared in 1.5-ml Eppendorf tubes by adding 0.5 ml of sterile 100 % glycerol to 1 ml of a stationary phase bacterial culture. The tube was sealed with parafilm, labelled and stored at - 70°C.

Stabs of *E. coli* strains serve as a back-up to the glycerol stocks. They were made from a purified colony that was picked with a sterile toothpick from a freshly streaked LB agar plate. The colony was transferred to a small 2-ml plastic tube filled with LB agar, sealed with parafilm and incubated overnight at 37°C. This tube was then stored at room temperature.

2.2.1.2 Growth of Overnight Cultures

To make overnight cultures of *E. coli,* the desired strain was streaked out on LB agar plates from a frozen glycerol stock to obtain single colonies. A bacterial culture was grown overnight by inoculating 5 ml of L Broth with a single colony from the LB agar plate and shaking at 37°C.

2.2.1.3 Preparation and Storage of $CaCl_2$-Competent **E. coli** *cells*

Competent cells were made by diluting an overnight culture of the appropriate *E. coli* strain 10-fold in 20-ml of L broth. The culture was grown shaking at 37°C until OD_{650}=0.4 to 0.5 was reached. The cell culture was incubated on ice for 20 minutes and then centrifuged at 5 krpm for 5 minutes at 4°C (Centra-3, International Equipment Company, UK). The supernatant was discarded and the cell pellet was resuspended in 5 ml of ice-cold 100mM $CaCl_2$. This step was followed by centrifugation at 5 krpm for 5 minutes at 4°C and the supernatant was removed. The cell pellet was again resuspended in 5 ml of ice-cold 100mM $CaCl_2$ and incubated on ice for 20 minutes. After centrifugation at 5 krpm for 5 minutes at 4°C the supernatant was discarded and the cell pellet resuspended in 400 µl of MOPS-glycerol solution. Aliquots of 200 µl were dispensed into pre-cooled 1.5-ml Eppendorf tubes. Tubes were immediatelly transferred to -70°C.

2.2.1.4 Transformation of $CaCl_2$-Competent E. coli *Cells*

$CaCl_2$-competent cells of the appropriate *E. coli* strain were thawed on ice. 100ng (1 µl) of plasmid DNA were added to 100 µl of $CaCl_2$-competent cells and the mixture was vortexed (Whirli mixer, Fisons Scientific Apparatus Ltd.) for 1 second. After 20 minutes of incubation on ice the DNA/cell mixture was heat-shocked at 42°C for 90 seconds. It was then put on ice for no more than 2 minutes. The transformation mixture was plated on LB agar plates supplemented with the appropriate antibiotics to select for transformants. The plates were incubated at 37°C overnight.

2.2.1.5 Transfer of mutations between E. coli *strains by P1 transduction*

Upon infection of *E. coli* cells phage P1 packages random fragments of the bacterial chromosome (up to 2 minutes) into phage particles which can be injected into the recipient *E. coli* strain (for review see Margolin, 1987). Hence, the transfer of a mutation of interest from one *E. coli* strain to another can be achieved by P1 transduction. For that purpose, a P1 lysate was made on a *E. coli* strain that carries the mutation of interest and a cotransducible selective marker, such as a Tn10 or Tn5 insertion. The recipient *E. coli* strain is then transduced with the P1 lysate. Transductants are identified by selection for the cotransduced marker. The presence of the mutation of interest is then confirmed by testing for the mutant phenotype.

2.2.1.5.1 Preparation of P1 Plate Lysate

A fresh overnight culture of the appropriate *E. coli* strain was diluted 10-fold in fresh L broth supplemented with 2.5 mM $CaCl_2$. This culture was grown for 2 hours at 37°C with shaking. Of this culture 200 µl were added to 100 µl of P1 lysate (10^7 pfu ml^{-1}). After incubation at 37°C for 30 minutes 2.5 ml of LB top agar containing 5

mM $CaCl_2$ were added to the phage/ bacteria mixture and poured onto a fresh LB agar plate, also supplemented with 5 mM $CaCl_2$. After incubation at 37°C for 6 to 8 hours, 5 ml of phage buffer were applied onto the plate, the top agar was scraped off and filled into a 30-ml glass bottle containing 100 µl of chloroform. The mixture was vortexed and then incubated at room temperature for 30 minutes. After centrifugation at 5 krpm for 10 minutes the clear supernatant was transferred to a sterile 5-ml McCautney bottle containing 200 µl of chloroform. The P1 lysate was stored at 4°C.

2.2.1.5.2 P1 Transduction

An overnight culture of the recipient *E. coli* strain was grown in 5 ml of L broth supplemented with 2.5 mM $CaCl_2$. 1-ml aliquots of this culture were transferred to 1.5-ml Eppendorf tubes and cell pellets were obtained by centrifugation at 9 krpm for 5 minutes. The supernatant was removed from each tube and each cell pellet was resuspended in 100 µl of L broth containing 2.5 mM $CaCl_2$. To the first tube 100 µl of undiluted P1 lysate were added while 100 µl of a 10-fold dilution of the P1 lysate were added to the second tube. 100 µl of phage buffer were added to the third tube, which served as a negative control. A fourth tube contained 100 µl of undiluted P1 lysate, but no recipient. All four tubes were incubated at 37°C for 20 minutes. Then 800 µl of L broth supplemented with 2 mM sodium citrate (to stop P1 infection) were added to each tube and incubation was continued at 37°C for 60 minutes. Of all four transduction mixtures, 100 µl of a 10^{-2} dilution and 100 µl of the undiluted transduction mixture were plated on LB agar plates containing the appropriate antibiotic drug for selection of transductants.

Plating of recipient cells (tube 3) and of P1 lysate (tube 4) alone should not yield any colonies that are resistant to the appropriate antibiotic drug. Only if this was the case, single antibiotic-resistant colonies that had been obtained by plating transduction mixtures from tubes 1 and 2 were purified and tested for the mutant phenotype. Glycerol stocks of purifies transductants were stored at -70°C.

2.2.2 Non-Radioactive In Situ Plaque Hybridisation

In the course of this work a DNA sequence including a trinucleotide repeat array was cloned into three different vectors in two possible orientations for each vector. The DNA fragment whose cloning had been attempted was non-radioactively labelled and used as a probe in a screen for recombinant plasmids. DNA fragments were labelled using the DIG-DNA Labelling Kit (Roche) by following the manufacturer's instructions. The procedure relies on the detection of hybridised DNA probes, which have been labelled by incorporation of DIG-11-dUTP using a random priming method and Klenow enzyme.

2.2.2.1 Preparation of a DIG-Labelled Hybridisation Probe

15 µl to 30 µl of gel-extracted DNA were transferred onto a 0.5-ml Eppendorf tube, boiled for 10 minutes in a water-bath and then immediately placed on ice for 3 minutes. 3.3 µl of 10 x Hexanucleotide mixture, 3.7 µl of 10 x dNTP labelling mix and 2U Klenow enzyme (all Roche) were added to the denatured DNA fragment, mixed thoroughly and incubated at 37°C overnight. This labelling reaction was inactivated by adding 1.5 µl of 0.5 M EDTA. The DNA probe was ethanol-precipitated. The pellet was dissolved in 50 µl of 1 x TE complemented with 1 µl of 5% (w/v) SDS, vortexed, centrifuged and stored at -20°C. DIG-labelled DNA fragments were used as hybridisation probes for up to 6 months.

2.2.2.2 Colony Transfer onto Nylon Membranes

Round Hybond™-N Nylon membranes (Amersham Life Science) were labelled so that later the exact position of a membrane on the corresponding agar plate could be

identified. The labelled membrane was carefully placed on a BBL agar plate. After 2 minutes the membrane was transferred onto two layers of blotting paper [Fords Gold Medal Blotting, 140g $(m^2)^{-1}$] which had been moistened with some 0.5 N NaOH and were lying on a piece of Saranwrap (The Dow Chemical Company). To ensure complete denaturation it had to be made sure that there were no air bubbles or creases in the blotting paper and the paper was reasonably wet. After 5 minutes on the denaturation paper the membranes were washed in neutralisation buffers 1 and 2 for 20 and 40 seconds, respectively. Membranes were than washed in 2 x SSC for another 20 seconds. The membranes were dried briefly between two pieces of blotting paper soaked with 2 x SSC. The DNA was cross-linked by exposing the membranes to 3 x 120 mJ using a UV-Stratalinker (Stratagene).

2.2.2.3 DNA Hybridisation

The membranes were moistened with some sterile distilled water and transferred to a plastic bag. 24 ml of Easy-Hyb® buffer were added to the bag and the bag was sealed. The membranes were pre-hybridised at 37°C for 1 hour. In the meantime, a mixture of 4 µl of DIG-labelled probe and 100 µl of Easy-Hyb® buffer was boiled for 10 minutes, and then placed on ice/water. After the 1 hour incubation period was complete, the 24 ml of Easy-Hyb® buffer were removed from the plastic bag. The entire volume of boiled probe was added to 5.5 ml of fresh Easy-Hyb® buffer and added to membranes in the plastic bag. The membranes were incubated with the probe at 37°C overnight.

2.2.2.4 Washing the Membranes

After overnight incubation the membranes were removed from the plastic bag and placed in a small plastic tray. The membranes were washed in 100 ml of wash buffer 1 at room temperature for 5 minutes with shaking. This first wash step was repeated using a further 100 ml of wash buffer 1. Then the membranes were washed twice for 15 minutes in wash buffer 2 with continuing agitation.

2.2.2.5 Signal Detection Using CDP-Star™

The membranes were washed for 2 minutes in detection buffer 1 before they were incubated with 100 ml of detection buffer 1 containing 1 % (w/v) blocking reagent. After shaking for 30 minutes, all liquid was removed from the tray and 30 ml of detection solution 3 containing the Anti-DIG-AP-Conjugate were added to the membranes and incubated for a further 30 minutes. Then the membranes were washed twice for 15 minutes in detection solution 1 followed by a single wash in detection solution 4 for 5 minutes. CDP-Star™ was used as a substrate for Alkaline Phosphatase linked to the DIG-antibody. CDP-Star™ was diluted in detection solution 4 to a final concentration of 100 µm of which 250 µl were distributed over each membrane. The membranes were incubated with this substrate solution for 5 minutes at room temperature. After the membranes have been dried briefly between two sheets of blotting paper, they were covered with Saranwrap and exposed to an X-ray film (DuPont). The first film was developed after an exposure time of 20 minutes so that a second film could be put down for the appropriate length of time necessary to obtain readable signals.

The signals on the films were aligned with the colonies in the corresponding agar plate and positive colonies were marked. Several colonies that gave positive

hybridisation signals were picked. They were screened for the insert by PCR and by incubation with appropriate restriction endonucleases.

2.2.3 Methods of DNA Purification and Manipulation

2.2.3.1 DNA Precipitation

Unless otherwise stated, DNA was precipitated in 2 volumes of a mixture of 96% ethanol and sodium acetate (20:1). For instance, 50 µl of a DNA solution were added to a mixture of 100 µl of 96 % ethanol and 4 µl of 3 M sodium acetate. The mixture was briefly vortexed. After overnight incubation at -20°C, the DNA was pelleted by centrifugation at 15 krpm at 4°C for 30 minutes. The precipitate was washed twice with 70 % (v/v) ethanol, air-dried, resuspended in the appropriate volume (20-50 µl) of 10 mM Tris-HCl buffer and stored at -20°C.

2.2.3.2 Measurement of the Concentration of DNA

The calculation of the amount of DNA in a solution is based on the maximal UV-light absorption of DNA at a wavelength of 260nm. An absorption value of 1 measured at this wavelength equals 50 µg of double-stranded DNA per ml (Maniatis et al.1989). The absorption spectrum between 200nm and 300nm was recorded as an indication of the purity of the DNA preparation. The quotient of the absorption values measured at 280nm and 260nm should be greater than 1.5 in a pure DNA solution.

2.2.3.3 DNA Restriction Digests

Unless stated otherwise DNA digestion was carried out in volumes of 20 µl using 2 U of restriction endonuclease per µg plasmid DNA with the incubation buffer provided by the manufacturer. The reaction mixture was incubated for 2 hours at the temperature recommended by the manufacturer. If the DNA had to be modified subsequently (e.g. end-labelling, ligation, sequencing) the restriction endonuclease was heat inactivated at 65°C for 20 minutes. If heat inactivation was not applicable the reaction was stopped using the QIAquick Nucleotide Removal kit (QIAGEN) which removes enzymes and buffers, but retains DNA fragments ≥ 17bp (80-95 % of fragments consisting of 100bp- 10kb).

2.2.3.4 DNA Dephosphorylation

Plasmid DNA was dephosphorylated after digestion with EcoRI restriction endonuclease. 1 x Shrimp Alkaline Phosphatase buffer and 1 U of Shrimp Alkaline Phosphatase were added to the reaction mixture which contained 2 to 3 µg of DNA. The mixture was incubated at 37°C for 1 hour. The reaction was stopped by heating to 65°C for 10 minutes.

2.2.3.5 DNA Ligation

DNA ligation was carried out in 25 to 50 µl of T4 DNA ligase buffer. Dephosphorylated vector DNA was ligated to *EcoR*I fragments in a molar ratio of vector to insert of approximately 1:1 to 2:1.

The ligation reaction was incubated in a 16°C waterbath. To inactivate DNA ligase the reaction mixture was transferred to a dry heating block that had been preheated to

70°C. After 15 minutes the heating block was switched off and allowed to cool down to 30°C. The mobile metal insert was then removed from the heating block and put on ice for 2 hours. The DNA was ethanol precipitated following standard procedure. The air-dried DNA pellet was resuspended in 5 µl of 1 x TE buffer. Resuspension of the DNA pellet was completed by incubation at room temperature for 1 hour. This was followed by incubation on ice for 2 hours. DNA was stored at - 20°C.

2.2.3.6 Extraction of Plasmid DNA from E. coli

Principle of the Procedure

All plasmid DNA was prepared using the solutions provided by the range of QIAprep Kits (Qiagen Inc.) and by following the manufacturer's instructions. The protocol is an application of a modified alkaline lysis method by Birnboim & Doly (1979) followed by binding of the DNA to a silica-gel membrane in the presence of high salt which allows its purification from RNA, cellular proteins and metabolites (Vogelstein and Gillespie, 1979).

2.2.3.6.1 Small Scale Preparation of Plasmid DNA

Small amounts of plasmid DNA (10-15 µg) were prepared from 5-ml overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen). For this purpose, cells from 4.5 ml (3 x 1.5 ml) of a fresh overnight culture were pelleted in a single 1.5 ml Eppendorf tube by centrifugation at 15 krpm for 1 minute in a bench top centrifuge (Sorvall Microspin 24). This and all subsequent steps were carried out at room temperature. The cell pellet was resuspended in 250 µl of buffer P1 using a vortex. Bacteria were lysed by adding 250 µl of buffer P2. The tube was inverted gently to achieve sufficient mixing of the solutions without shearing the bacterial chromosomal DNA. Addition of 350 µl of buffer P3 neutralises and adjusts to high-

salt binding conditions in one step. The solution was centrifuged for 10 minutes at 15 krpm and the clear supernatant was loaded onto a QIAprep spin column. After another round of centrifugation (1 minute, 15 krpm) the flow-through was discarded and 500 µl of wash buffer PB were added to the column to remove trace nuclease activity and carbohydrate. The column was washed by adding 750 µl of buffer PE and centrifuging for 1 minute at 15 krpm. This centrifugation step was repeated after the flow-through had been discarded from the collection tube. The column was placed in a sterile 1.5-ml Eppendorf tube and the DNA was eluted from the membrane by adding 55 µl of 10 mM Tris-buffer and centrifuging at 15 krpm for 1 minute. The DNA was stored at -20°C.

2.2.3.6.2 Large Scale Preparation of Plasmid DNA

A single colony picked from a freshly streaked LB agar plate (supplemented with the appropriate antibiotic drug) was inoculated into 100 ml L broth containing the antibiotic drug as before. The culture was grown with vigorous shaking overnight at 37°C. The cells were collected by centrifugation at 12 krpm for 6 minutes at 4°C (Sorvall centrifuge, GSA or SS34 rotor). All subsequent centrifugations were carried out at 16 krpm and 4°C (Sorvall centrifuge, GSA or SS34 rotor).

The cell pellet was resuspended in 4 ml of buffer P1. The cells were lysed in the presence of RNase (provided in buffer P1) by adding 4 ml of buffer P2. The viscous solution was mixed gently by inverting the tube a few times. Bacterial lysis was allowed at room temperature for 5 minutes. The solution was neutralised and lysis stopped with 4 ml of chilled buffer P3. The samples were mixed and incubated on ice for 15 minutes. The mixture was centrifuged for 30 minutes to separate cell debris and chromosomal DNA from the clear solution containing the plasmid DNA. The supernatant was carefully decanted and applied to a QIAGEN-tip 500, which had been equilibrated with 4 ml of QBT buffer. The column was allowed to empty by gravity flow. Then the column was washed twice by permitting 2 x 10 ml of buffer QC to passthrough the column by gravity flow. The DNA was eluted from the column with 5 ml of buffer QF. The DNA was precipitated with 0.7 volumes of

isopropanol and centrifugation for 30 minutes. Since DNA pellets resulting from an isopropanol precipitation are transparent, the outside of the glass tube was marked at the position where a precipitation would be expected. The supernatant was removed and 2 ml of ice-cold 70 % (v/v) ethanol were added to wash the DNA pellet. A final centrifugation was carried out for 10 minutes, all liquid was removed from the tube and the pellet was air-dried. DNA was resuspended in 500 µl of 10 mM Tris-HCl buffer and stored at -20°C.

2.2.3.7 Extraction of Genomic DNA from E. coli

Genomic DNA was extracted from *E. coli* overnight cultures using the AGTC® Bacterial Genomic DNA Purification Kit (Advanced Genetic Technologies Corp.) which contains all necessary buffers and ADVAMAX™ beads for DNA binding. DNA was prepared following the protocol provided by the manufacturer.
Bacteria were collected from a 5-ml overnight culture by centrifugation at 15 krpm for 1 minute. The supernatant was discarded and the cell pellet was resuspended in 400 µl of Spheroblast buffer. After incubation at 37°C for 10 minutes 100 µl lysis buffer 1 and 100 µl of lysis buffer 2 were added. The suspension was mixed gently and incubated at 65°C for 5 minutes after which 100 µl of ADVAMAX™ beads and 100 µl of extraction buffer were added. The mixture was vortexed for 10 seconds and centrifuged at 15 krpm for 3 minutes. The supernatant was transferred to a sterile 1.5-ml Eppendorf tube and an equal volume of isopropanol was added. The suspension was inverted several times until the white DNA precipitate became clearly visible. Using a blue (1 ml) pipette tip the DNA was transferred to a sterile 1.5-ml Eppendorf tube which contained 500 µl of 70 % (v/v) ethanol. To wash the DNA the tube was inverted 5-10 times. The precipitate was pelleted by centrifugation at 15 krpm for 5 minutes. The ethanol was removed and the DNA pellet was air-dried. The DNA was dissolved in 50 µl of 1 x TE buffer and stored at -20°C.

2.2.3.8 DNA Sequencing

2.2.3.8.1 Manual Radioactive Sequencing

Sequencing of double stranded template DNA carrying direct repeats was carried out using the Sequenase® v2.0 Sequencing Kit (United States Biochemical) which uses a genetic variant of T7 DNA Polymerase to incorporate α^{35}S-dATP into the newly synthesised DNA strand during primer extension *in vitro*. All reaction components mentioned below except template DNA and [α^{35}S]-dATP were provided in the Sequenase ® v2.0 Sequencing Kit.

Double stranded plasmid DNA was purified using the QIAquick Nucleotide Removal Kit (QIAGEN). In a 0.5-ml Eppendorf tube, 3 µg to 5 µg (9 µl) of purified plasmid DNA were mixed with 11 pmole (1 µl) of sequencing primer D. This mixture was boiled for 4 minutes and immediately put on dry ice to keep the DNA single stranded. A sequencing mix was prepared by mixing 4 µl of 100 mM DTT solution and 4 µl of 10 x Reaction buffer with 1.4 µl of 5 x Labelling mix. Sequenase® DNA Polymerase (13 U μl^{-1}) was diluted 1:5 in Sequenase® dilution buffer to yield a final volume of 5 µl. Just prior to starting the sequencing reaction the sequencing-mix was completed by adding 5 µl of the Sequenase® DNA Polymerase dilution and 1 µl (10 µCi) of [α-^{35}S] dATP.

Four Eppendorf tubes were labelled G, A, T, C and 2 µl of the appropriate Termination-Mix were added to each of them. One DNA/primer mix was taken off the ice, defrosted and centrifuged briefly before 7.7 µl of sequencing mix were added to the tube. The reaction components were mixed well and the tube was incubated at 20°C for 5 minutes. The four tubes containing the termination mixtures were transferred to a 40°C-waterbath approximately 1 minute before they had to be used. Then 4 µl of the sequencing mixture were added to each of the tubes with the termination mixtures. The sequencing reactions were terminated after incubation for 3 minutes at 40°C by adding 4 µl of stop solution to each tube.

Prior to loading, samples were boiled for 3 minutes and immediately put on ice to denature the extension products. 5 µl of each sample were loaded on a preheated

(55°C) denaturing 5 % Long Ranger™ gel (see below) in the order G, A, T, C. If the sequencing products were not analysed immediately they were stored at -20°C for up to one week.

2.2.3.8.2 Non-Radioactive Automated Cycle Sequencing

As an alternative, recombinant DNA was sequenced using a non-radioactive cycle-sequencing using the ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). This method is based on the incorporation of fluorescent-dye labelled chain terminators by the heat-stable AmpliTaq® DNA Polymerase (FS) in a cycling single-tube reaction.
The detection and analysis of the sequencing products was carried out on a ABI PRISM 377 DNA Sequencer (Applied Biosystems). For one sample, the following reagents were mixed in a 0.5-ml Eppendorf tube:
8 µl Terminator Ready Reaction Mix
1 µl DNA (~ 500 ng of ds plasmid DNA, ~ 200 ng of ds PCR product)
1 µl primer (3.2 pmole)
10 µl sterile distilled water
20 µl final reaction volume

Cycle sequencing was performed in a PCR Express Hybaid Thermal Cycler, preheated to 94°C:

Denaturation	Primer annealing	Elongation
94°C - 30 seconds	50°C - 15 seconds	60°C - 4 minutes
	30 cycles	

Prior to the sample analysis on the ABI PRISM™ 377 DNA Sequencer the sequencing products were ethanol-precipitated in order to remove unincorporated dye-labelled terminators and to concentrate the labelled extension products. The air-

dried DNA pellet was handed over to the ABI PRISM 377 DNA Sequencing facility at the Institute of Cell and Molecular Biology (University of Edinburgh). Collected data were analysed using the GeneJockey™ Sequence Processor (Biosoft, Cambridge).

2.2.3.9 The Polymerase Chain Reaction (PCR)

Usually, PCR was carried out using purified plasmid DNA or bacterial genomic DNA. PCR was carried out in a volume of 25 µl using 0.2-ml PCR tubes in a PCR Express Hybaid Thermal Cycler. The reaction mixture consisted of:

2.5 µl 10 x Taq-Polymerase buffer (including 25 mM $MgCl_2$)

2.5 µl 2 mM dNTP-4 mix

5 µl DNA (plasmid DNA: 10-100 ng; genomic DNA: 200-400 ng)

0.5 µl Taq-Polymerase (2.5 Units)

1 µl primer 1 (20 pmole)

1 µl primer 2 (20 pmole)

13 µl sterile, distilled water

A master-mix containing all reaction components except template DNA was prepared on ice with Taq-DNA Polymerase being added last. Aliquots of 20 µl were dispensed into 0.2-ml PCR tubes, which already contained 5 µl of DNA solution. In a control reaction the DNA solution was replaced by 5 µl of sterile distilled water. The solution was mixed well and centrifuged briefly at 9 krpm. The PCR block was pre-heated to 94°C before the samples were inserted. PCR programmes that were developed in the course of this project are described at appropriate places in Chapter 5. The PCR programme for the amplification of CTG repeats using primers M6833 and M6834 was designed by Abbott and Chambers (1994). This is shown in Table 2-5.

Table 2-5. PCR programme used on the Hybaid PCR Express Thermal Cycler for the amplification of trinucleotide repeats in pDL915 and pDL915R, using primers M6833 and M6834 (Abbott and Chambers, 1994). Time is shown in minutes.

Initial Denat.	Denaturation	Annealing	Elongation	Final Elong.
Time/ Temp				
3 / 94	1/94	1/55	1/72	5/72
1 cycle	30 cycles			1 cycle

2.2.3.10 Purification of PCR Products

DNA fragments ranging from 100 bp to 10 kb were purified from primers, dNTPs, salts and enzymes using the QIAquick PCR Purification Kit (Qiagen). This method of purification was used to ensure optimal conditions during an enzymatic reaction when, for example, incubation with a restriction endonuclease followed PCR. Like the Gel Extraction Kit the PCR Purification Kit uses silica-gel membranes. DNA binds to silica at pH $\leq$ 7.5 in the presence of a high concentration of chaotropic salts while other components of the PCR reaction mixture are found in the flow-through. The purification was carried out using the solutions provided by the QIAquick kit according to the protocol supplied by the manufacturer. 5 volumes of buffer PB were added to 1 volume of PCR reaction. After mixing thoroughly the solution was applied to a QIAquick column which was standing in a 2-ml collection tube. The column was centrifuged at 15 krpm for 1 minute at room temperature. The flow-through containing salts, dNTPs, primers and enzymes was discarded. The column was washed with 750 µl of buffer PE to remove residual salts and other PCR reaction components. Flow-through was discarded and the centrifugation step was repeated to remove all of buffer PE. The DNA was eluted from the column with 30 µl to 50 µl of 10 mM Tris-HCl and stored at -20°C.

2.2.3.11 Radioactive Labelling of Double-Stranded DNA Fragments

DNA fragments were labelled after incubation of plasmid DNA (e.g. pDL915, pDL1400, 1401, etc) with restriction endonucleases (e.g. *EcoRI, AvrII*), that produce 3'- recessed ends which can be filled-in by Klenow enzyme. Because of the sequence of the overhang produced by *EcoRI* and *AvrII*, [α-^{35}S] dATP was used in the labelling reaction. During this project, radioactive end-labelling was used to determine the number of CTG trinucleotides in *attP*-derived plasmids which carry CTG repeat arrays of varying length and also the number of CTG trinucleotides integrated at the *attB* site in chromosomes (Chapter 5). For this purpose, 17 µl of plasmid DNA prepared by the small scale method were mixed with 2 µl of 10 x *EcoRI* restriction buffer (New England Biolabs) and 1 µl of *EcoRI* (20 U µl^{-1}). After 2 hours of incubation at 37°C EcoRI was heat-inactivated at 65°C for 20 minutes. The samples were spun in a bench-top centrifuge at 9 krpm fpr 30 seconds. Without any further modification the samples were used in the following radioactive end-labelling reaction. Per sample, 1 µl (10 µCi) of [α-^{35}S] dATP and 0.5 µl (2 U µl^{-1}) Klenow enzyme were combined in an Eppendorf tube. Of this labelling mix 1.5 µl were added to 20 µl of EcoRI-digested plasmid DNA. After incubation at room temperature for 5 minutes 2 µl of the 2 mM dNTP-4 mix were added to the labelling reaction and incubation was continued for 10 minutes at room temperature. 3 µl of polyacrylamide gel loading buffer were added to the reaction. Immediately, 20 to 25 µl of this reaction mixture were loaded onto a native 5 % Long Ranger gel to separate the labelled DNA fragments.

2.2.3.12 Non-Radioactive Linear Amplification of CTG Repeats

For the rapid analysis of the length of triplet repeats on the *E. coli* chromosome as well as in plasmids, a non-radioactive method was improved as an alternative to error-prone PCR that is frequently used to measure triplet repeat length. The original

idea proposed by Yamamoto et al. (1992) was modified by K. Schmidt to include fluorescent dye labelled primers and to eliminate error prone PCR from the process. The method's principle, optimisation and application to plasmid and genomic DNA of *E. coli* are described in Chapter 5.

2.2.3.13 Gel Electrophoresis

2.2.3.13.1 Agarose Gel Electrophoresis

The length of double-stranded DNA fragments was analysed by electrophoretic migration in horizontal agarose gels in 1 x TBE-buffer at approximately 5-8 V cm^{-1}. Depending on the size of the DNA fragments, the concentration of routine electrophoresis grade agarose (Flowgen) was 1 % to 2 % (w/v) while higher percentage gels (3 % to 4 %) were made from Methaphor agarose (Flowgen). To pour an agarose gel, the appropriate amount of agarose was added to 1 x TBE-buffer and mixed in a conical flask. A magnetic stirrer bar was placed into the flask and the agarose was melted in a microwave oven at setting "High" for 1 - 3 minutes (depending on the gel volume). When all agarose particles were melted, the flask was placed on a magnetic stirrer and left there to cool down. Just before the gel was poured, 1 x ethidium bromide stock solution was mixed into the agarose solution (2 µl of 50 x ethidium bromide stock solution for a 100-ml gel or 6 µl for a 300-ml gel). The gel was allowed to set for 1 hour at room temperature. (Et Br migrated towards the anode in an electrical field). Prior to loading, 1 x agarose gel loading buffer was added to each sample and mixed well. DNA was visualised using a C-62 BlackRay transilluminator (Ultraviolet Products Incorporated). Pictures of gels were taken using GRAB-IT™ software (Ultraviolet Products Incorporated). Metaphor agarose gel electrophoreses are described in Chapter 5.

2.2.3.13.2 Native Polyacrylamide Gel Electrophoresis

Native polyacrylamide gels were used to analyse length changes in the highly unstable CTG trinucleotide repeats. The gels were run in the Sequi-Gen® Nucleic Acid Sequencing Cell (Biorad) according to the manufacturer's instructions. The gel was prepared aproximately 3 hours before it was needed. For a gel measuring 40 cm x 21 cm x 0.4 mm, 70 ml of gel solution were required (20 ml to seal the bottom of the glass plate sandwich and 50 ml to pour the gel). In order to make a 5 % native gel containing 1.2 x TBE, 7 ml of 50 % Long Ranger™ stock solution (Flowgen), 8.4 ml of 10 x TBE and 54 ml of sterile distilled water were mixed thoroughly but carefully in a glass beaker by slowly pipetting up and down. To seal the bottom of the glass plate sandwich 70 µl of TEMED and 170 µl of freshly prepared 10 % (w/v) AMPS were added to 20 ml of the gel solution. When the gel had set, 50 µl of TEMED and 150 µl of freshly prepared 10 % (w/v) AMPS were added to the remaining 50 ml of gel solution. This gel mixture was immediately poured between the two glass plates with the help of a 50-ml plastic syringe. A square-toothed comb was introduced at the top of the gel. The gel remained at room temperature for at least 3 hours to set. After 3 hours the gel was removed from the casting tray and placed into the sequencing apparatus (Biorad). Upper and lower buffer tanks were filled with 700 ml and 300 ml of 0.6 x TBE buffer, respectively. Just before the samples were loaded the comb was removed and the slots were thoroughly washed with 0.6 x TBE using a syringe with a needle. Into each slot 20 µl to 25 µl of sample containing polyacrylamide gel loading buffer was loaded. The gel was run at a constant power of 35W at a temperature between 45-50°C until the bromophenol blue band of the loading buffer had reached the bottom of the gel. After the electrophoresis was completed, the gel was removed from the glass plate sandwich, transferred onto wet blotting paper (Ford Goldmedal), covered with Saranwrap and dried in a Biorad Gel Dryer (Model 583) at 80°C for 45 minutes.

2.2.3.13.3 Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gels were used to separate products of manual sequencing reactions (see 2.2.4.3.1 Manual Radioactive Sequencing). In addition to the ingredients mentioned above, the gel solution contained 7 M urea. To make a 5 % denaturing Long Ranger™ gel, the following components were mixed: 7 ml of 50 % Long Ranger ™ stock solution, 8.4 ml of 10 x TBE, 50 ml of sterile, distilled water and 24 g urea (Sigma Chemical Company). This mixture was warmed in a 50°C waterbath to help disolve the urea. Sterile, distilled water was added to reach a final volume of 70 ml. This gel solution was used to pour the gel as it was described above. A sharktooth comb was inserted at the top of the gel with the flat side facing the gel. The gel was allowed to set for at least 2 hours at room temperature and was then inserted into the sequencing cell (Biorad). The upper and lower buffer tanks were filled with 0.6 x TBE before the comb was removed. The gel surface was flushed with 0.6 x TBE using a syringe with a needle to remove urea and unpolymerised polyacrylamide. The gel was pre-run at 46W until the gel had reached a temperature of 55°C (approximately 1 hour). Before loading, the samples were boiled for 3 minutes and put on ice immediately. The sharktooth comb was so re-inserted that the teeth formed wells in which the samples could be loaded. The slots were rinsed again and 5 µl of sample were loaded. The gel was run at constant power of 40 W for 1-2 hours. After completion of electrophoresis the gel was removed from the glass plate sandwich, transferred onto wet blotting paper (Ford Goldmedal), covered with Saranwrap and dried in a Biorad Gel Dryer (Model 583) at 80°C for 45 minutes.

2.2.3.13.4 Metaphor Agarose Gel Electrophoresis

The MetaPhor agarose gels were prepared according to the manufacturer's instructions. Precautions were taken when dissolving agarose against scalding solutions.

1) 75 ml of chilled 1 X TBE buffer were poured into a 500-ml beaker.

2) The agarose powder was slowly sprinkled while the solution was rapidly stirred with a Teflon® coated stir bar.

3) The agarose was soaked in the buffer for 15 minutes before heating. This reduced the tendency of the agarose solution to foam during heating.

4) The beaker and solution were weighted before heating.

5) The beaker was covered with plastic wrap, with a small hole pierced for ventilation.

6) The beaker was heated in the microwave oven on medium power for 2 minutes.

7) The beaker was removed from the microwave oven, gently swirling it to resuspend any settled powder and gel pieces.

8) The beaker was reheated on high power until the solution boiled for 1 minute or until all the particles were dissolved.

9) The beaker was removed from the microwave oven, gently swirling it to thoroughly mix the agarose solution.

10) Sufficient hot distilled water was added and mixed thoroughly to obtain the initial weight.

11) The solution was cooled to 50 - 60°C prior to casting.

12) Once the gel was cast, the molten agarose was cooled at room temperature. The gel was then placed at 4°C for 20 minutes to obtain optimal resolution and gel handling characteristics.

2.2.3.14 Extraction of DNA Fragments from Agarose Gels

PCR products and DNA restriction fragments were extracted from agarose gels when it was necessary to select particular restriction fragments or PCR products for labelling of DNA probes, for ligation or DNA sequencing. DNA fragments were separated on horizontal agarose gels made from a Genetic Technology Grade (GTG) agarose (SeaKem Incorporated) which is special agarose for preparative electrophoresis of DNA. After gel electrophoresis was completed, the gel was placed on an UV transilluminator and the selected DNA band was excised from the agarose gel using a sterile scalpel. The gel slice was put into a sterile 1.5-ml Eppendorf tube and was weighted. The QIAquick Gel Extraction kit (Qiagen) was used to extract the DNA from the agarose slice. First, 3 gel-volumes (w/v) of buffer QG were added to the gel slice and the gel/ buffer mix was incubated at 50°C until the gel had completely dissolved (approximately 5 minutes). The colour of buffer QG, which is yellow indicating a pH ≤ 7.5, should not change during this procedure. If it changed to orange or violet the pH was too high and had to be re-adjusted to pH 7.5 to ensure optimal binding of the DNA to the silica-gel membrane. If small DNA fragments (< 500 bp) were extracted 1 gel-volume isopropanol was added to the solution to increase DNA recovery. Solutions containing longer DNA fragments were loaded onto the column without any modification. The columns were spun at 15 krpm for 1 minute. The flow-through was discarded and 500 µl of buffer QG were added to the column to remove traces of agarose. The columns were centrifuged at 15 krpm for 1 minute. The flow-through was again discarded and 750 µl of ethanol-containing buffer PE were added to eliminate salts. The column was spun twice at 15 krpm for 1 minute and the collection tube was emptied between centrifugations to ensure complete removal of buffer PE from the column. The column was placed in a sterile 1.5-ml Eppendorf tube and 30 µl of 10 mM Tris-HCl buffer were added to the centre of the membrane to elute the DNA. After 1 minute the column was centrifuged for 1 minute at 15 krpm. Since the lid had to be cut off the tube for centrifugation, the eluate was transferred to a new, sterile Eppendorf tube. The sample was stored at -20°C.

2.2.3.15 Autoradiography

Dried Long Ranger™ gels were exposed to Cronex 4 X-ray films (DuPont). Exposure was carried out in Cronex cassettes (DuPont), at room temperature and overnight. Depending on the strength of the signal a second film was put down for a more suitable length of time. Films were developed in a X-OGRAPH Compact X2 automatic film processor.

Chapter 3

Strategies for Constructing New Plasmids Containing the CAG/CTG Repeat Tract

3.1 Introduction

More than a dozen TREDs are known to be caused by expanded CAG/CTG repeats, including Huntington disease, spinocerebellar ataxia types 1, 2, 3, 6, 7, 8, 12 and 17, myotonic dystrophy, DRPLA and SBMA (Chapter 1). Trinucleotide repeat expansion is a well established, but poorly understood mutational mechanism of human genetic disease. It is known that the CTG repeats are located in transcribed and, with the exception of myotonic dystrophy, SCA8 and 12, translated regions of genes. It had been shown that mutated chromosomes in carriers and affected individuals with Huntington Disease generally harboured more than 39 CTG trinucleotides while the length of these CTG repeats in unaffected individuals ranged from 5 to 39 trinucleotides.

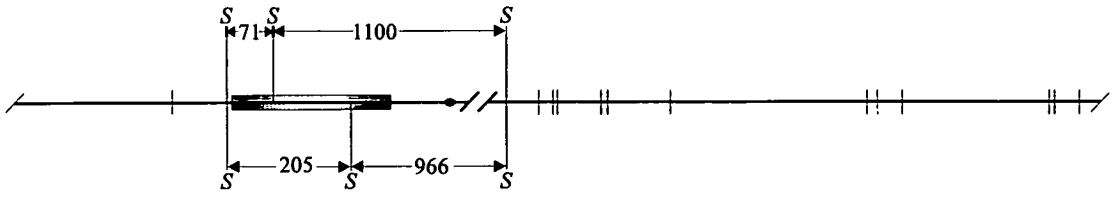
Previous studies of long lengths of trinucleotide repeat tracts ($CTG_{130\text{-}180}$) (Kang *et al.*, 1995) showed both expansion and contraction depend on the orientation of the $(CTG)_n$ repeat with respect to the unidirectional replication origin of pUC-based plasmids (ColE1) as well as on the copy number of plasmids, the host

genotype, the number of cell generations, the length of the repeat and its location in the plasmid.

To gain insight into the mutational mechanisms by which shorter trinucleotide repeats integrated in plasmids, are destabilised, a CTG repeat consisting of just 43 trinucleotides integrated in pUC18 was studied in the Leach laboratory in wild-type *E. coli* and in mutator mutants of *E. coli* that are deficient in post-replicative mismatch-repair (K. Schmidt's PhD thesis, University of Edinburgh).

CTG repeat instability was studied using two pUC-derived plasmids, pDL915 and pDL915R. They carry a $(CTG)_{43}$ trinucleotide repeat in opposite orientations as insertions in the single *EcoRI* site of pUC18. Plasmid pDL915 was constructed in the laboratory of Dr. C. Abbott (Western General, Edinburgh) using the following procedure. A DNA fragment containing a $(CTG)_{25}$ repeat and 88 bp of flanking sequence was amplified by PCR from the mouse metallothionein-III (Mt-III) gene which maps on chromosome 8 and contains a highly polymorphic CTG repeat in its 5' untranslated region (see Appendix 1). The MT-III gene codes for Metallothionein-III, a small, central nervous system-specific metal-binding protein (Blaauwgeers *et al.*, 1996). The primer pair M3/M4 (Table 2.4) was designed to amplify the $(CTG)_{25}$ repeat and 88 bp of flanking sequence from the inbred strain of laboratory mouse AKR/J. This PCR product was ligated to *EcoRI* linkers (5' GAATTCCT 3') and inserted into the single *EcoRI* site of the high copy number plasmid pUC18. This plasmid was labelled pDL913. After passage of pDL913 through a *mutS*-deficient strain, an expanded plasmid was isolated and labelled pDL915. Sequencing revealed that this plasmid contained an *EcoRI* fragment with an expanded repeat array containing approximately 41 CTG trinucleotides. Plasmids

with the *EcoRI* fragments in the opposite orientations, labelled pDL913R and pDL915R, were constructed in the course of an earlier project in this laboratory by cleaving both plasmids with *EcoRI* and religating the fragments. Figure 3.1 explains how the presence of the single asymmetric *Sau*3A site in the *EcoRI* fragment permitted the identification of the orientation of the *EcoRI* fragments with respect to the unidirectional origin of replication in pUC18 (ColE1).



CTG repeat on leading strand: orientation A (pDL915)
CTG repeat on lagging strand: orientation B (pDL915R)

Figure 3.1 Partial *Sau*3A restriction map of pDL915 (upper line) and pDL915R (down). The single *Sau*3A site in the PCR fragment amplified from the mouse MT-III gene and the two flanking *Sau*3A sites in pUC18 could be used to determine the orientation of the *EcoRI* fragments in pDL915 and pDL915R with respect to the origin of replication. The remaining 13 *Sau*3A restriction sites in pUC18 yield 14 *Sau*3A restriction fragments whose length is independent of the orientation of the *EcoRI* fragment.

For example, if the $(CTG)_{43}$ repeat array is oriented so that the CTG sequence is on the template for the leading strand during replication, a *Sau*3A restriction digest would produce 198-bp and 966-bp fragments in addition to the 14 *Sau*3A restriction fragments whose length is not affected by the presence of the *EcoRI* fragment. If the *EcoRI* fragment was in the opposite orientation, the same digest would produce 70-bp and 1094-bp fragments. Henceforth the orientation in

which the *EcoRI* fragment is inserted so that CTG is the template for the leading strand during replication will be referred to as orientation A. If the *EcoRI* fragment is found to be the other way around, *i.e.* if CTG is on the template for the lagging strand during replication, then this will be referred to as orientation B.

In the beginning of my project, the trinucleotide repeat tract $(CTG \bullet CAG)_{43}$ flanked by two *EcoRI* sites from the plasmid described above, was available in the Leach lab to be mobilised into the *Escherichia coli* chromosome. This procedure was attempted by employing the help of two helper vectors, kindly provided by Prof. W. Messer (Berlin, Germany) : pLDR9 and pLDR11, in the presence of the *Int Protein Expression* vector, pLDR8 (Diederich *et al.*, 1992).

3.2 Inserting the Trinucleotide Repeat Tract CTG/CAG into Vectors pLDR9 and pLDR11

Prof. Messer's team developed an improved plasmid vector-based system for integration of DNA fragments into the λ attachment site *attB*. The method is applicable to any DNA fragment and any strain as long as the IHF protein is produced by the cells. Furthermore, the method allows the direct integration of the DNA fragment under consideration into the strain of choice, without the need of subsequent transduction.

The system consists of two parts: a helper plasmid that provides the Int protein from phage λ, which is necessary for integration into the *attB* site of the chromosome (Weisberg *et al.*, 1983) and another cloning vector, containing the *attP* site and an easily removable origin cassette (Figure 3.2).

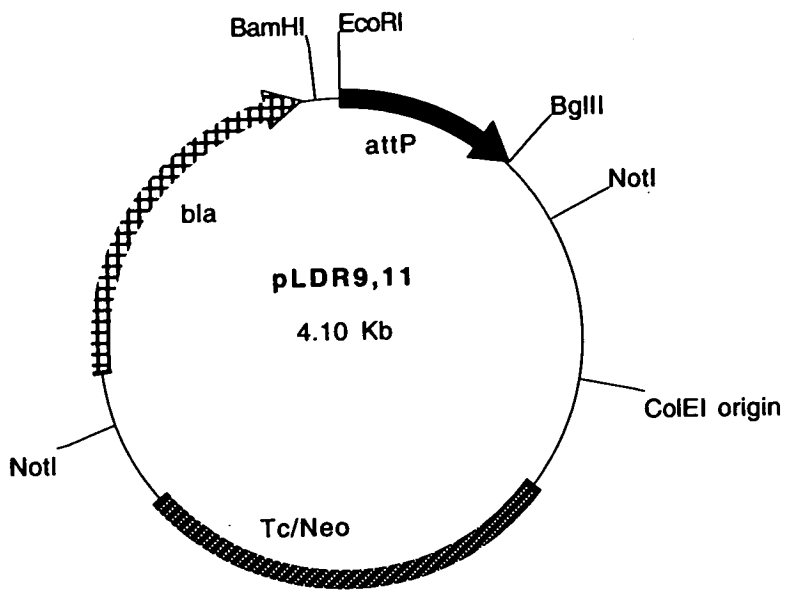


Figure 3.2 Schematic representation of the plasmid vectors pLDR9 and pLRD11.

The existence of various markers (resistance to antibiotics) within this cassette makes selection for it possible. The method proceeds as follows: first the DNA fragment in question is cloned in one of the *attP* vectors (Figures 3.3 and 3.4). Second, the origin cassette is cut out of the plasmid obtained and the fragment is ligated, leading to a closed circular DNA molecule lacking a replication origin. Third, the *E. coli* strain of choice carrying the Int protein producing helper plasmid pLDR8 (Chapter 4) is transformed with the ligation products.

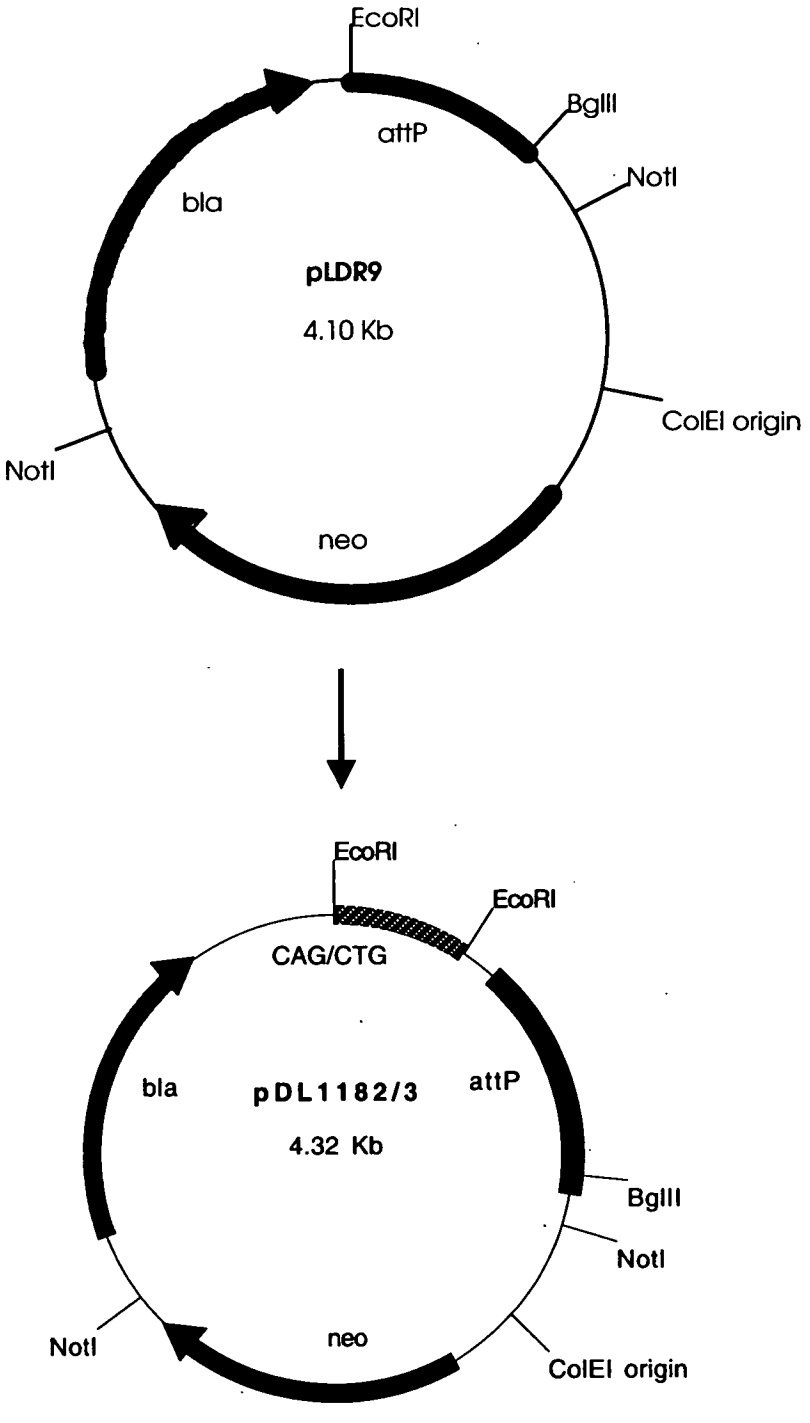


Figure 3.3: Schematic representation of the vector pLDR9 and the new constructed plasmids, with the trinucleotide repeats [$(CTG)_{43}$ or $(CAG)_{43}$ on the leading strand] integrated in the *EcoRI* site (pDL1182 and pDL1183).

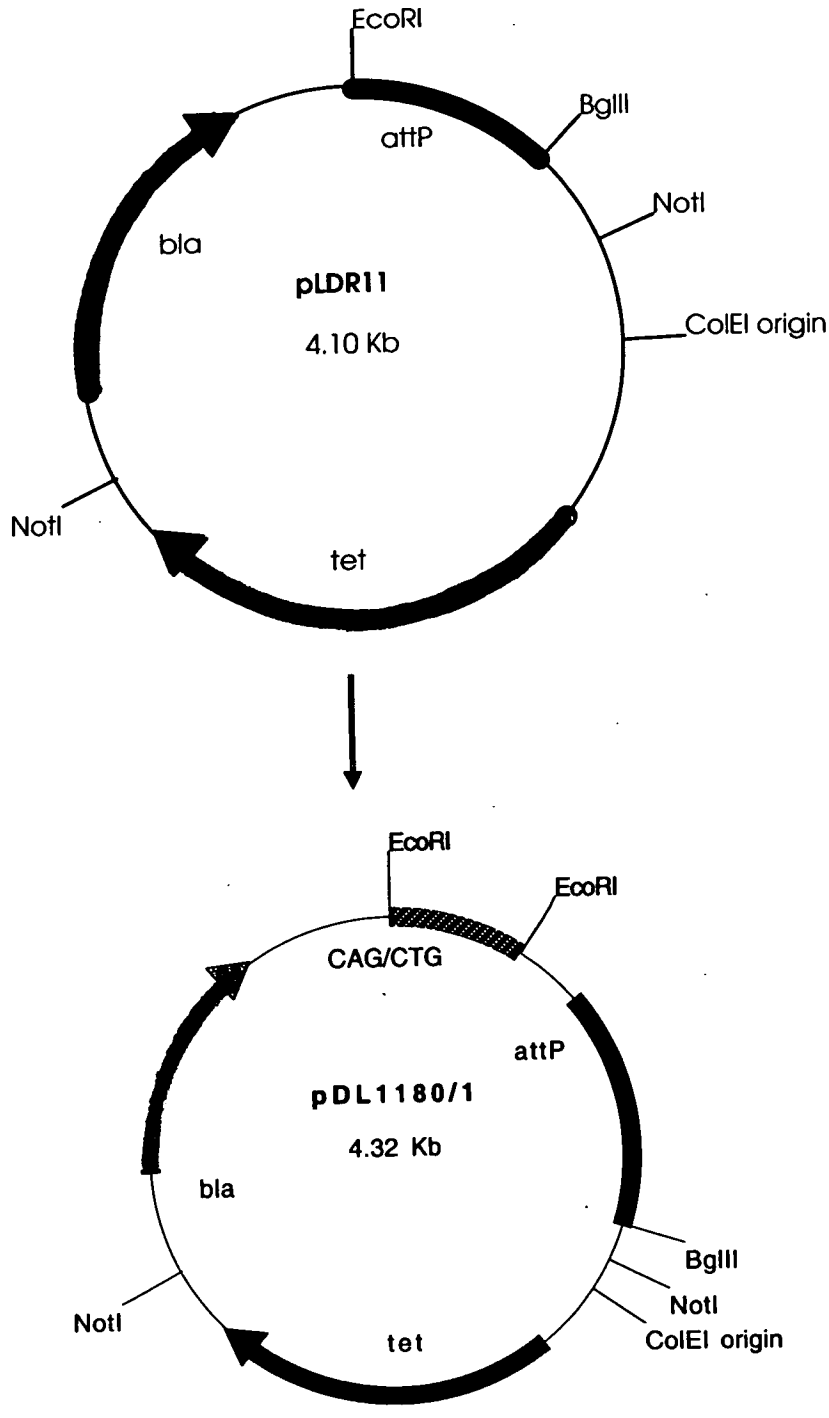


Figure 3.4: Schematic representation of the vector pLDR11 and the construction of the new plasmids with the trinucleotide repeat tract [$(CTG)_{39}$ or $(CAG)_{39}$ on the leading strand] integrated into the *EcoRI* site (pDL1180 and pDL1181).

The first step of this project consisted in integrating the trinucleotide repeats in the vectors pLDR9 and pLDR11 which allow the integration of any DNA fragment into the bacteriophage λ attachment site *attB* of the *Escherichia coli* chromosome. The repeat tract $(CAG)_{43}$ and 88 bp flanking sequence of the murine metallothineine III gene (Mt3) was separated from pDL915 by digestion with *EcoRI* and ligated to the *EcoRI* sites of the vectors pLDR9 and pLDR11. A nonradioactive *in situ* hybridization (the DIG-system hybridisation) was used to screen for correctly constructed colonies, that had the trinucleotide repeat tract integrated into the *EcoRI* site of the vectors pLDR9 and pLDR11.

The vectors pLDR9 and pLDR11 have important features, making them very useful for integrating DNA fragments into the chromosome of *Escherichia coli*:

- they contain the λ attachment site *attP*;
- they have double resistance to antibiotics (pLDR9: Amp^R, Km^R; pLDR11: Amp^R, Tc^R);
- they have an origin cassette easily removable, by digestion with *NotI*;
- the screening for the correctly constructed colonies is easy, since the gene coding the resistance to the second antibiotic (Kan, Tc) is removed at the same time as the origin cassette.

The newly constructed plasmids were checked by digestion with *NotI* or *EcoRI* (Figure 3.5) and by PCR to determine the orientation of the repeats (Figure 3.7). Three primers were designed to perform the polymerase chain reactions: primer D with the property to anneal to the vector in the neighbouring region to the *EcoRI* site; primers M3 and M4 which anneal to the ends of the trinucleotide repeat tract (Figure 3.6).

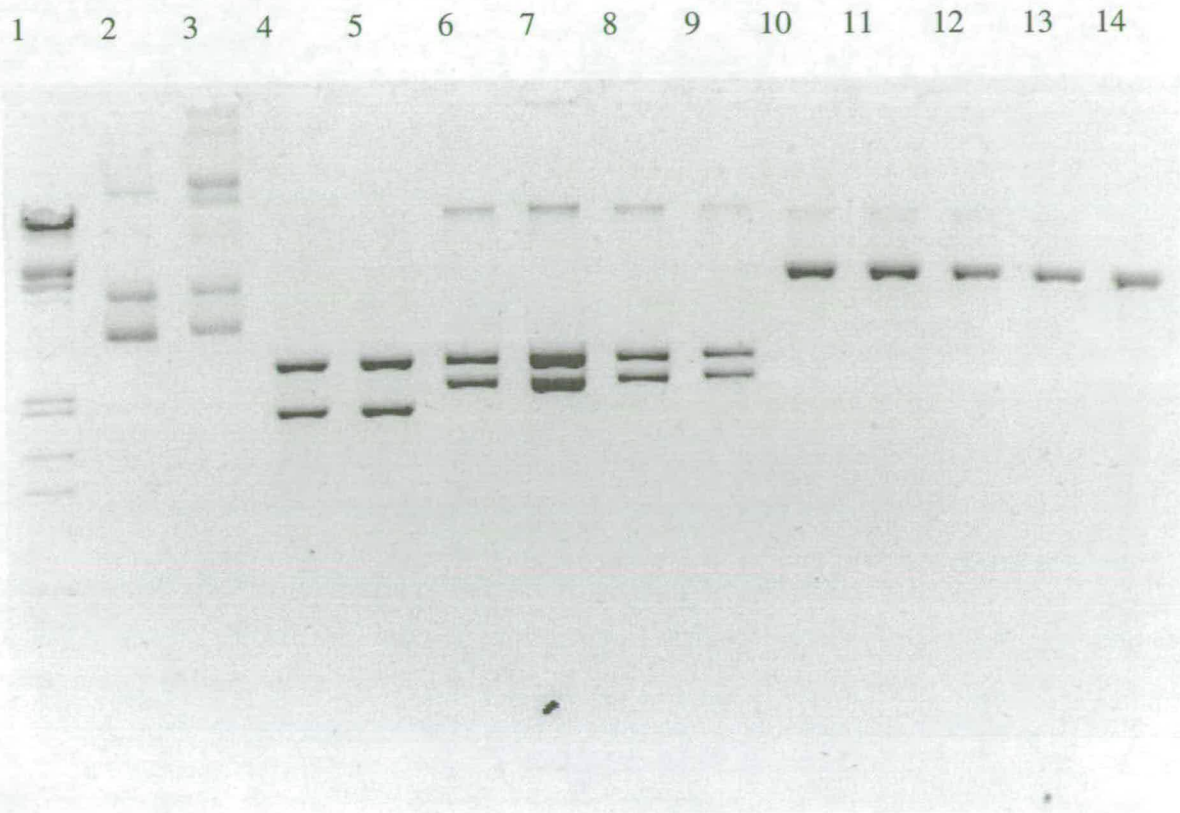


Figure 3.5: 1% agarose gel loaded with the products of the digestion of the vectors and the new constructed plasmids with *NotI* or *EcoRI*. Lane 1, marker III; 2, vector pLDR9; 3, vector pLDR11; 4, vector pLDR9 cut with *NotI*; 5, vector pLDR11 cut with *NotI*; 6-9, new constructed plasmids cut with *NotI*: 6, DL1180; 7, DL1181; 8, DL1182; 9, DL1183; 10-13, new constructed plasmids cut with *EcoRI*: 10, DL1180; 11, DL1181; 12, DL1182; 13, DL1183; 14, vector pDLR9 cut with *EcoRI*.

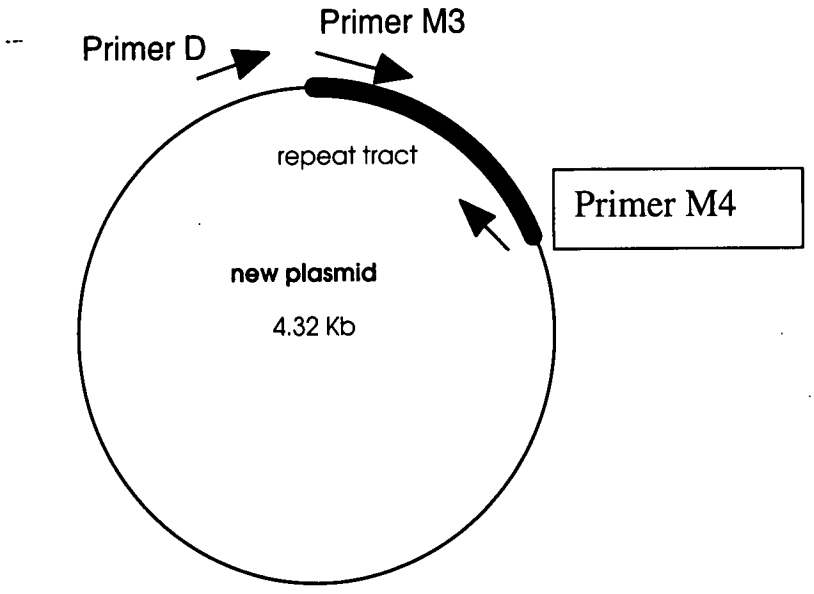


Figure 3.6: Schematic representation of the experiment establishing the orientation of the repeat tract (black area represents the trinucleotide repeat tract). Primer D was designed to anneal outside the trinucleotide repeat tract; primers M3 and M4 anneal to the Mt-III flanking sequence. PCR using primers D and M3 or D and M4 was performed in order to establish the orientation of the repeats in the new plasmid.

PCR was performed to find both orientations of the triplet repeat tract in the vectors (figure 3.7), according to the following protocol: 94°C for 2 min, 25 cycles x (94°C for 40 sec, 62°C for 40 sec, 72°C for 40sec), 72°C for 10 min.

This first objective of the project was successful; four new plasmids were produced:

pDL 1180 (pLDR11 + $(CTG)_{39}$ on the leading strand), Tc^R, Amp^R

pDL 1181 (pLDR11 + $(CAG)_{39}$ on the leading strand), Tc^R, Amp^R

pDL 1182 (pLDR9 + $(CTG)_{43}$ on the leading strand), Km^R, Amp^R

pDL1183 (pLDR9+$(CAG)_{43}$ on the leading strand), Km^R, Amp^R .

3.3 Modifying the restriction sites of the attP site containing vector pLDR9

A novel strategy was attempted in the course of my project. It was based on the advantage offered by the *AvrII* restriction endonuclease to cleave the *Escherichia coli* chromosomal DNA at a low frequency. The average fragment size resulting by such a digestion is 150kb, with the largest fragment observed being of 1Mb (New England Biolabs catalogue). The recognition site for the *AvrII* enzyme is:

5'.....C/CTAG G.....3'

3'.....G GATC/C.....5'

The strategy behind this new experiment was to engineer two flanking *AvrII* sites on both sides of the trinucleotide repeat tract. Following integration into the chromosome of this new fragment, and restriction with *AvrII*, the CAG/CTG array of 220bp could easily be separated on a polyacrylamide gel from the rest of the fragments of 150kb.

The vector pLDR9 was chosen to have the restriction sites modified because it was known that it had only one site for the *BamHI* enzyme (Figure 3.8). The modification consisted in inactivating the existing sites for *EcoRI* and *BamHI*, by performing a double-digest with both enzymes and then, ligating a new fragment, called "the AvrII fragment", to the open vector.

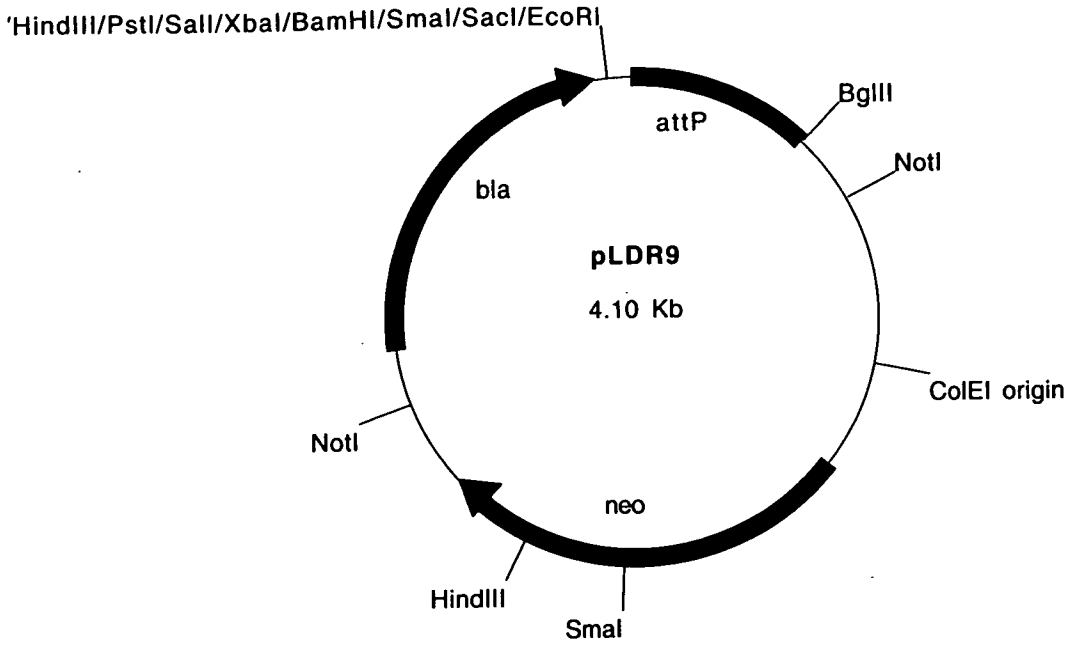


Figure 3.8 Vector pLDR9 was chosen to be modified because it has only one restriction site for the enzyme BamHI. It is resistant to Km (*neo*).

The *AvrII* fragment was constructed by the annealing of two asymmetric oligonucleotides, BAMECO and ECOBAM. The two oligonucleotides were boiled together in equimolar amounts in TE buffer, 100mM NaCl and dH_2O and left to anneal overnight by cooling down at room temperature. The sequence of the two oligonucleotides was:

BAMECO

5'....GATCAGGTACCCCTAGGGAATTCCCTAGGT.....3'

ECOBAM

5'.....AATTACCTAGGGAATTCCCTAGGGGTACCT.....3'

The trinucleotide repeat tracts CAG/CTG were introduced into the newly modified vector, pLDR9M, as it was described in subchapter 3.1 for the initial unmodified vectors pLDR9 and pLDR11. The two new plasmids pDL1400 and 1401, derived from pLDR9M, have the CAG repeats on the template for the lagging and the template for the leading strand, respectively (Figure 3.10).

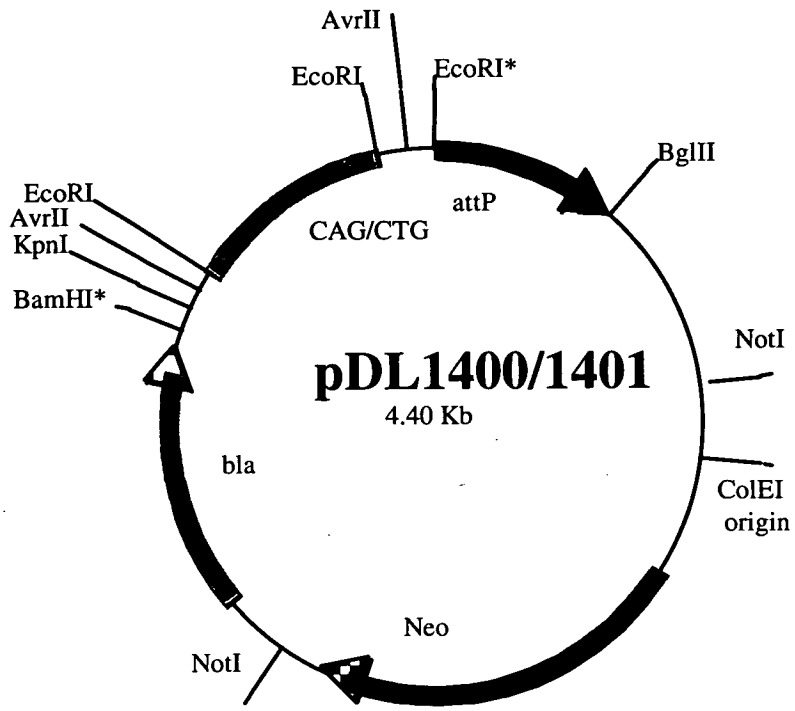


Figure 3.10 Schematic representation of the new plasmids containing a trinucleotide repeat array flanked by *EcoRI* and *AvrII* sites.

The new plasmids pDL1400 and pDL1401 were checked as described in subchapter 3.1, and also by cleavage with *EcoRI* and *AvrII*, followed by Klenow enzyme filling-in with radioactive nucleotides and separation by 6% PAGE (Fig 3.11). The difference in size between *EcoRI* digestion products and *AvrII* is of 12 bp, according to the DNA sequence of the BAMECO and ECOBAM oligonucleotides.

Later in my project, a trinucleotide repeat array of 84 copies became available from pUC18, thanks to G. Cromie. I attempted to transfer the $(CAG/CTG)_{84}$ from pUC18 into pDL1400/1401. This repeat tract proved to be highly unstable in the vectors pDL1400/1401. In spite of intensive efforts to integrate this repetitive fragment into the *attB* vectors, it turned out to be fruitless. The longest sequence, which was successfully introduced, was of 56 copies (Figure 3.12).

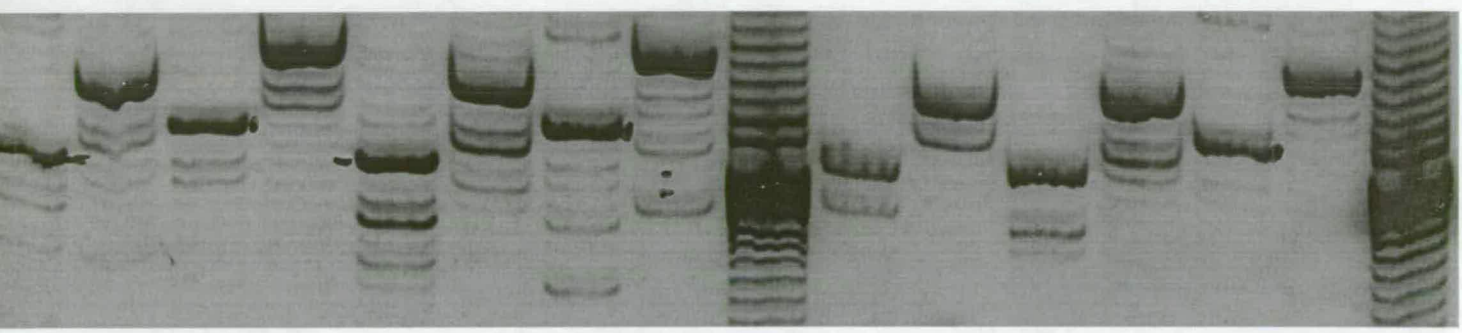


Figure 3.11. Checking colonies pDL1400 and pDL1401 for the number of trinucleotide repeat copies integrated into pLDR9M. *EcoRI* and *AvrII* restriction products were radioactively labelled with α-35S dATP and separated by 6%PAGE.

There are two putative reasons for the failure to construct new plasmids with $(CTG)_{84}$ integrated into pLDR9M: 1) longer repeat tracts form more stable hairpins compared to shorter ones, so the efficiency of $(CTG)_{84}$ ligating to the vector is smaller in competition with shorter arrays present in the reaction (as it can be noticed in Figure 3.12, apart from the main population of $(CTG)_{84}$, all other smaller lengths are also present); 2) the process of transformation into competent cells might have a mutagenic effect for certain DNA sequences. This observation is in agreement with Hashem's paper (Hashem *et al.*, 2002) which showed that instabilities of $(CAG)(CTG)_{76\text{-}79}$ are dramatically elevated upon transformation. The authors found that the mutation frequency for deletion of the CAG tract upon transformation was 75 to 145-fold higher than the mutation frequency for the same plasmids maintained in cells. Several aspects of the transformation process are to be taken into consideration to explain the highly mutagenic effect for the repeated sequences: 1) the process of making the cells competent may alter their natural metabolic state and the rate of spontaneous mutation such that the results do not reflect those occurring in a normal cellular environment; 2) the process of transport past the bacterial membrane may also introduce damage to the DNA; 3) the plasmid DNA presented to cells lacks its normal complement of DNA binding and packaging proteins and the initial interactions with the replication and transcription machinery may be abnormal.

The successfully constructed plasmids as described in this chapter were later used for integrating the trinucleotide repeat tract into the *E. coli* chromosome (Chapter 4).

Chapter 4

Construction of New Bacterial Strains with Trinucleotide Repeat Tracts Integrated into the *E. coli* chromosome

4.1 Site-Specific Recombination Occurring at the *attB* Site of the *E. coli* Chromosome

Following the successful construction of new plasmids with $(CAG)_{43}$ and $(CTG)_{43}$ integrated into pLDR9M and pLDR11 vectors (Chapter 3), the next step of this project was to integrate the trinucleotide repeat tracts into the λ attachment site *attB* of the *Escherichia coli* chromosome. This step was based on the same principle as the integration of the λ phage into the specific attachment site *attB* on the *E. coli* chromosome (Figure 4.1).

The plasmids pDL1180, 1181, 1182, 1183, 1400 and 1401, described in Chapter 3, contain the λ *attP* site on a fragment which can be separated from the one containing the origin of replication by digestion with the restriction enzyme *NotI*. Furthermore, detection of both fragments is possible due to different resistance markers located on the fragments. Accordingly, to carry out the integration process, separation of the two fragments is not necessary.

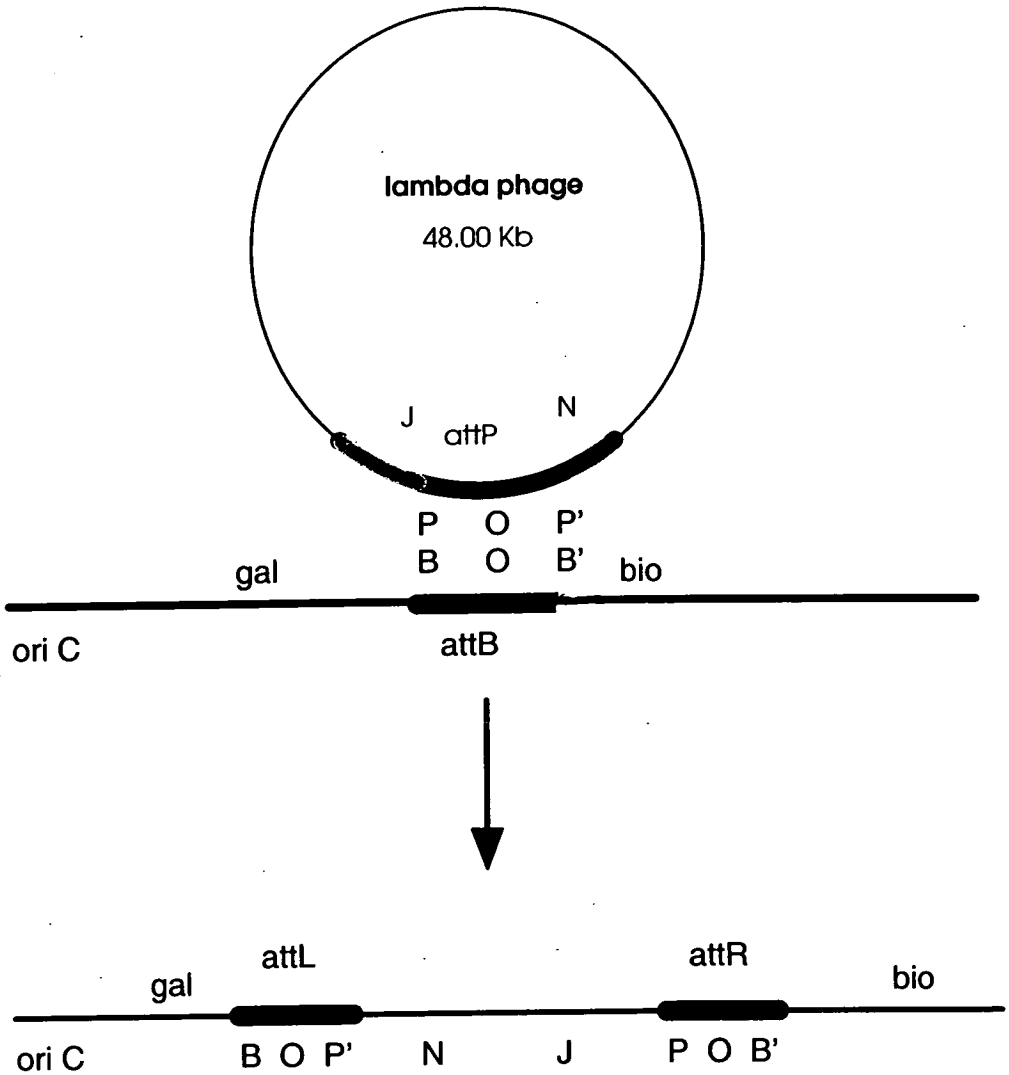


Figure 4.1: Integration of lambda phage into the *attB* site on the *E. coli* chromosome. The attachment site in the bacterial chromosome contains a core sequence, O, flanked by sequences B and B'. The core sequence is identical to the core sequence on the lambda phage, which is flanked by sequences P and P'. Phage encoded integrase together with integration host factor (IHF) catalyses the site-specific recombination event.

The origin cassette was cut out with *NotI* from the newly constructed plasmids (Chapter 3). The fragment containing the triplet repeats, the *attP* site and the gene coding the resistance for ampicillin was ligated, leading to a closed circular DNA molecule lacking a replication origin (Figure 4.2).

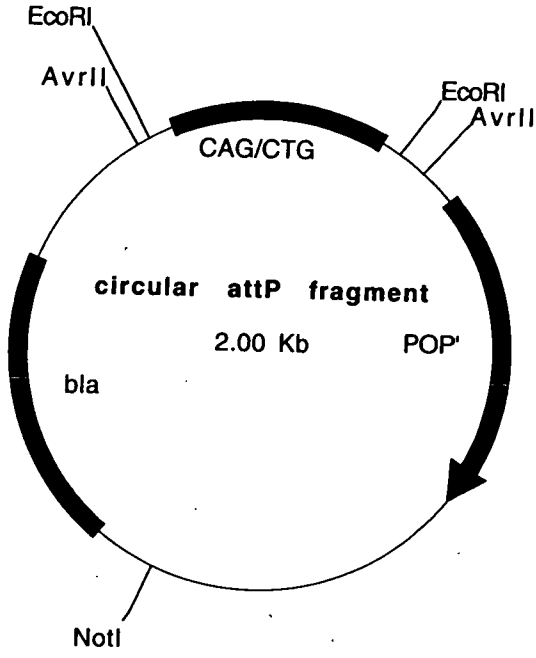


Figure 4.2. Schematic representation of the circular DNA fragment containing the trinucleotide repeat tract, the *attP* site and the gene coding resistance to *Ampicillin.*

Escherichia coli DL513 (wild-type) and DL515 (*sbcC*) carrying the Int protein producing helper plasmid pDLR8 were transformed with the ligation product. pDLR8 (Km^R) is a helper plasmid, bearing the λ *int* gene under the control of the temperature-sensitive repressor cI_{857} (Figure 4.3).

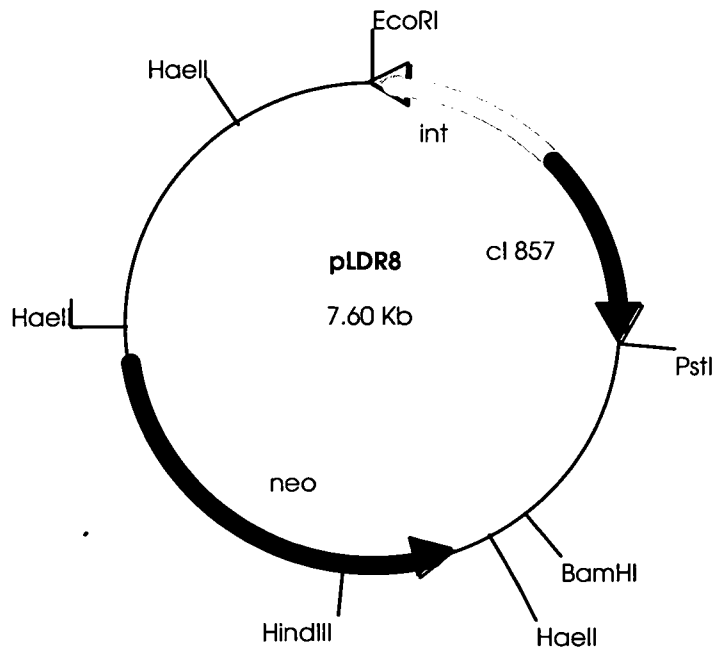


Figure 4.3: The helper plasmid pLDR8 contains the *int* gene, encoding Integrase under the control of a temperature-sensitive replicon and carries a Km resistance gene.

The *Escherichia coli* strains N2677 (*wild-type*) and N2679 (*sbcC201*) carrying the helper plasmid pLDR8, were grown at 30°C overnight in LB supplemented with Km. These cultures were diluted 1:20 and grown for 90 minutes at 37°C. At this stage, the cells were made competent and transformed with the ligation mixture. The ligation mixture had previously been digested with *Xho*I restriction enzyme in order to make sure that the only circular fragments present in the mixture are the ones containing the triplet repeats, the *attP* site and the gene encoding the resistance to ampicillin. The cells were plated on LB agar supplemented with Amp and incubated at 42°C overnight.

Plasmid pLDR8 carries a temperature-sensitive replicon and a Km-resistance gene. Moreover, the plasmid contains the *int* gene encoding the integrase transcribed from the λ P_R promoter under the control of the λ cI_{857} repressor.

When cells containing the helper plasmid pLDR8 are grown at 30°C, the replication of the plasmid is normal and the expression of the *int* gene is repressed. When the temperature is increased to 42°C, the cI_{857} repressor is inactivated and transcription from the λ P_R promoter leads to expression of the *int* gene. Simultaneously, the replication of pLDR8 is blocked due to inactivation of the Rep protein, resulting in the loss of the plasmid after a few cell generations. Therefore, the *int* gene product is only temporarily present in the cell.

Four new bacterial strains were initially constructed:

DL 1184 (DL513 Amp^R with $(CAG)_{39}$)

DL 1185 (DL513 Amp^R with $(CAG)_{43}$)

DL 1186 (DL515 Amp^R with $(CTG)_{39}$)

DL 1187 (DL515 Amp^R with $(CTG)_{43}$).

PCR was performed to check the number of repeat tracts in the chromosomes of the newly constructed bacterial strains, using primers Genom1 and Genom2, designed to anneal outside the *attB* site (Figure 4.4). The PCR protocol was performed according to the one in Table 4.1. The PCR products electrophoresed on 1% agarose gel proved that only one copy of the repeat tracts was integrated into each of the newly constructed strains chromosomes, since all the PCR products have the expected size of 2.2kb (Figure 4.5).

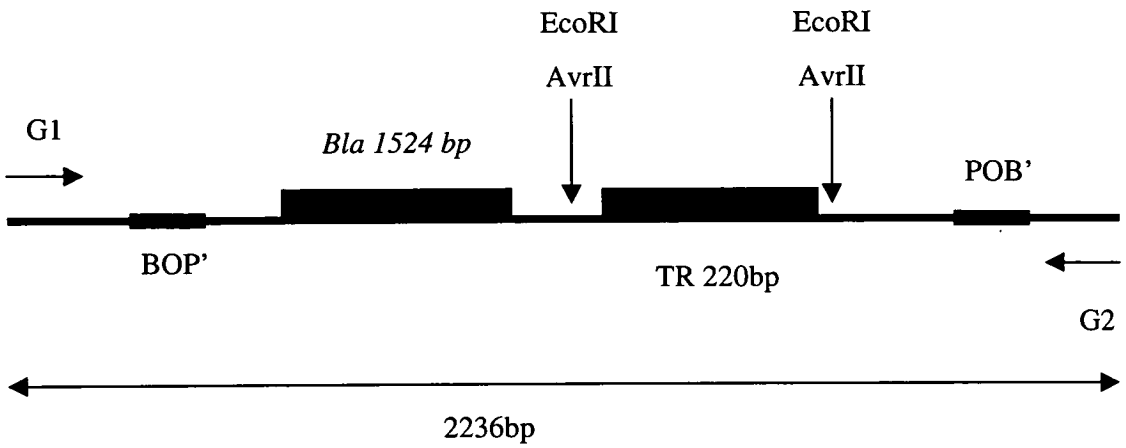


Figure 4.4: Schematic representation of the PCR product amplified by primers Genom1 and Genom2.

Table 4.1. Genom1+Genom2 PCR programme

Denaturation		Annealing		Elongation		Cycles
Time	Temp	Time	Temp	Time	Temp	
3 min	94°C					1
45 sec	94°C	45 sec	68°C	2min 30 sec	72°C	30

4.2 Constructing Various *E. coli* Mutant Strains with the Trinucleotide Repeat Tract Integrated at the *attB* Site

Following the successful construction of the initial strains containing the trinucleotide repeat tract $(CTG)_{43}$ or $(CAG)_{43}$ at the *attB* site, 60 new strains (Table 4.2) were constructed and checked for the existence of the newly introduced mutations. Several mutated genes, coding for proteins thought to be involved in trinucleotide instability, were introduced into the chromosomes of the previously constructed strains (DL1184- DL1187) (Chapter 4.1) by P1 transduction, as described in Chapter 2 (Materials and Methods). All newly constructed strains from Table 4.1 were checked for the correct propagation of the trinucleotide repeat tract by PCR followed by 1% agarose gel electrophoresis as described previously in Chapter 4.1 (Figure 4.4).

Table 4.2. Newly constructed bacterial strains with TR and the original *E. coli* strains they were derived from.

E.coli strain of origin	CAG = leading strand template	CTG = leading strand template
N2677/ DL513(wt)	DL1442 / 1302 / 1354	DL1404 / 1184 / 1185
N2679/ DL515 (*sbcC*)	DL1413 / 1186 / 1187	DL1406 / 1355
513, *recQ::km*	DL1453	DL1454
513, *recQ::km,sbcC*	DL1455	DL1456

DL1347 (*polA polB)*	DL1443	DL1428
DL1347 (*polA polB) recQ::km*	DL1457	DL1458
DL1348 (*polA*)	DL1444	DL1430
DL1348 *(polA) recQ::km*	DL1459	DL1460
DL1349 *(wt)*	DL1445	DL1432
DL1349 *recQ::km*	DL1461	DL1462
DL1350 *(polB)*	DL1446	DL1434
DL1350 *(polB) recQ::km*	DL1463	DL1464
DL1487 (DL513, *recA::cm)*	DL1488	DL1489 / 1272 / 1273
DL1490 (DL515, *recA::cm)*	DL1491 / 1264 / 1274	DL1492
DL1493 (DL513, *recB::Tc*)	DL1494	DL1495
DL962(*mutD::Tn10)*	DL1568	DL1567
DL1496 (DL515, *recB::Tc)*	DL1497	DL1498
DL1477 (DL513, *mutS::Tc)*	DL1478	DL1479 / 1420 / 1294 / 1295
DL1480 (DL515, *mutS::Tc)*	DL1481 / 1422 / 1296 / 1297	DL1482

DL1471 (DL513, *nfi::cm)*	DL1472	DL1473
DL1474 (DL515, *nfi::cm)*	DL1475 / 1243	DL1476
DL1499/1270 (DL513, *recG::km)*	DL1500	DL1501 / 1416 / 1236 / 1237
DL1502/1271 (DL515, *recG::km)*	DL1503 / 1238 / 1239	DL1504
DL513, *recF::cm*	DL1516	DL1517
DL1507 (DL515, *recF::cm)*	DL1508	DL1509
DL1226 (*radC)*	DL1249 / 1251	DL1252 / 1256
DL1224 *(radA)*	DL1250	DL1254
DL1225 (*wt*)		DL1255

A large number of control strains were constructed by the integration at the *attB* site of the circular DNA fragment originating from the vectors pLDR9, pLDR11 or pLDR9M which did not have the trinucleotide array.

All newly constructed strains of *E. coli* containing the trinucleotide repeat array were checked for the double resistance to antibiotics: *Ampicillin* (linked to the CAG/CTG tract) and also for the second antibiotic introduced by P1 transduction, together with the new mutation. The successful introduction of the new mutation was also checked by PCR, as exemplified in Figure 4.6 for the *recG* mutation. The

primers used for this step were "*E. coli* ORFmer™ PCR Primer Pairs" provided by Genosys Biotechnologies Limited. The *E. coli* ORFmer PCR primer pairs have been designed to amplify all putative open reading frames (ORFs) in *Escherichia coli K-12*. Each ORFmer PCR primer pair consists of an "A-primer" (amino-terminal primer) directed at the start codon and a "C-primer" (carboxyl-terminal primer) directed at the stop codon of a specific ORF. The ORF-specific sequences of the primers have an average Tm of 66.7°C (±4.7°C).

Several colonies of the newly constructed strain were picked and analysed for the successful transduction of the required mutation. PCR was performed using the primer pair of the respective gene and also a control pair for an additional gene, which was not modified.

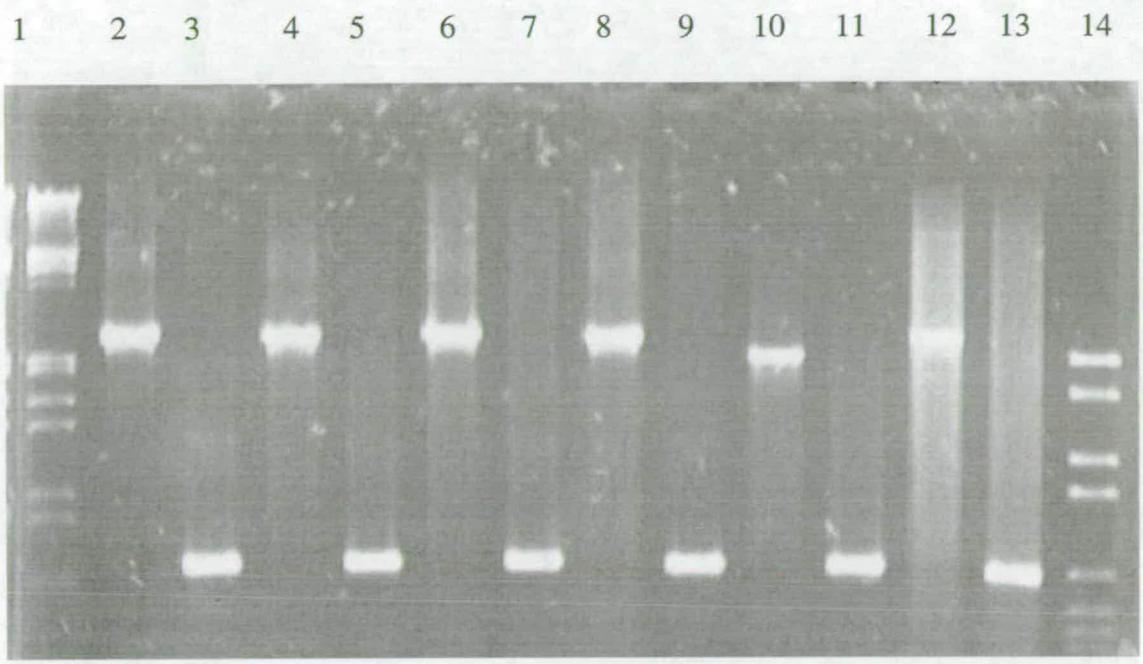


Figure 4.6. 1% agarose gel: lane 1) Marker 3; lane 14) Marker 6; even lanes: PCR products resulted from the amplification with the recG primer pair; odd lanes: PCR products from the amplification using the *recR* primer pair.

All newly constructed strains with a trinucleotide repeat array integrated at the *attB* site of the *E. coli* chromosome, were recorded in the Leach laboratory strain collection and stored as described in Chapter 2.2 (Methods). These strains constitute the material of analysis for investigating trinucleotide repeat instability (Chapter 6).

Chapter 5

Methods Employed in the Analysis of Trinucleotide Repeats Integrated into the *E. coli* Chromosome

5.1 Introduction

The aim of my project was to study the instability of trinucleotide repeat arrays in the *E. coli* strains, which I constructed (as described in Chapter 4). New methods had to be developed, investigated and improved in the course of my project in order to test the subtle changes of triplet arrays of only one, two or three trinucleotides, as it will be shown in detail in this chapter.

The trinucleotide repeat expansion is the cause of a large number of human diseases (Chapter 1). A trinucleotide repeat array of 43 GAC/CTG integrated into the pUC18 vector was previously studied in the Leach laboratory (K. Schmidt's PhD thesis, University of Edinburgh). Studying the unstable trinucleotide repeats in plasmids has some disadvantages: the maintenance of the unstable insert is difficult (Shimizu *et al.*, 1996), since they are deleted readily to sequences of heterogeneous lengths. The number of plasmids is likely to differ between strains, indirectly influencing plasmid stability within the population. The characterisation of the triplet

repeats integrated in plasmids is influenced by the copy number of the plasmid and is also orientation-dependent on the origin of replication of the plasmid, which is unidirectional. Kang (Kang *et al.*, 1996) reported that the region distal from the unidirectional replication origin expands much more than the proximal region. CTG repeats greater than 80 copies in plasmids seem more susceptible to deletions than expansions in *E. coli*, whereas they expand and remain long in human diseases. The reason for this apparent difference could be due to the size of the plasmids and to the mechanisms of replication. These disadvantages are overcome by studying the behaviour of trinucleotide repeats integrated in the chromosome of *E. coli*, which has a bi-directional replication origin, is 1,000 times larger than an average plasmid and may be a better mimic of the events occurring during replication in human chromosomes.

The next step, when my project started, was to assess the instability of these sequences in the *E. coli* chromosomes. Following the successful construction of new bacterial strains with the trinucleotide repeat arrays integrated at the *attB* site (as described in Chapters 3 and 4), my objective was to find an appropriate method to observe and analyse the subtle changes in the repeat lengths.

The analysis of trinucleotide repeat length is facilitated when high copy-number plasmids, such as pUC-derived vectors, pLDR9 and pLDR11, are used for the propagation of trinucleotide repeats. In this case, amplification of the DNA fragment of interest prior to analysis is dispensable as sufficient amounts of plasmid can be extracted from hosts. The fragments containing trinucleotide repeats are excised from plasmid DNA by cleavage with a restriction endonuclease that cleaves in the immediate flanking sequence of the repeat and the exact number of

trinucleotides is determined by separation of these restriction fragments in polyacrylamide gels. Since there is no *in vitro* amplification involved in this method, results reflect the *in vivo* situation. The disadvantage is that the method is no longer available once the CTG repeats had been inserted into the chromosome of *E. coli* (Chapter 4).

The existent methods used for the diagnostic of the trinucleotide repeat diseases are: 1) PCR using flanking primers to the TR array, not reliable above 100 trinucleotides; 2) Southern blot probing with a locus specific fragment flanking or containing the repeats; 3) fluorescent TP-PCR (Triplet Primed-PCR) (Warner *et al.*, 1996) that can identify, but not size the largest alleles.

Five new methods of trinucleotide repeat detection were investigated during this project in order to find the best one which fulfilled the requirement of distinguishing between trinucleotide repeat tracts of only 3 bp apart: short PCR followed by the separation of the products on MetaPhor agarose, long PCR products digested with *EcoRI* or *AvrII* and separated by 6% PAGE, *AvrII* digestion of chromosomal DNA, short fluorescent PCR products analysed by Gene Scan and *f*-TRAMP.

5.2 MetaPhor® Agarose Analysis of Short PCR Products Containing Trinucleotide Repeat Tracts

The first method investigated was the separation of short PCR products containing the trinucleotide repeat arrays on 3% MetaPhor agarose gels. Metaphor agarose comes highly recommended by its manufacturer (FMC®), as "the high resolution agarose that challenges polyacrylamide". It has an intermediate melting temperature (75°C) and it has twice the resolution capabilities of the finest-sieving agarose products.

A primer pair (primers Metaphor-Left and Metaphor-Right, Table 2.4) was designed to amplify the shortest possible fragment containing the trinucleotide repeat array (Figure 5.1). The PCR programme was successfully improved for picking colonies with toothpicks by using the gradient setting for the annealing temperature. Table 5.1 shows the final version of the Metaphor PCR programmes.

Table 5.1. MetaPhor PCR programme

Denaturation		Annealing		Elongation		Cycles
Time	Temp	Time	Temp	Time	Temp	
3 min	94°C					1
30 sec	94°C	30 sec	65°C	30 sec	72°C	25

The expected size for a PCR product containing 43 trinucleotide repeats was of 176 bp. Differences of 3 or 4 bp between the products were supposed to be clearly seen on the MetaPhor agarose gels since it can resolve small DNA fragments that differ in size by 2%.

The MetaPhor agarose gels were prepared according to the manufacturer's instructions. Precautions were taken when dissolving agarose against scalding solutions.

1) 75 ml of chilled 1 X TBE buffer were poured into a 500-ml beaker.
2) The agarose powder was slowly sprinkled while the solution was rapidly stirred with a Teflon® coated stir bar.
3) The agarose was soaked in the buffer for 15 minutes before heating. This reduced the tendency of the agarose solution to foam during heating.
4) The beaker and solution were weighted before heating.
5) The beaker was covered with plastic wrap, with a small hole pierced for ventilation.
6) The beaker was heated in the microwave oven on medium power for 2 minutes.
7) The beaker was removed from the microwave oven, gently swirling it to resuspend any settled powder and gel pieces.
8) The beaker was reheated on high power until the solution boiled for 1 minute or until all the particles were dissolved.
9) The beaker was removed from the microwave oven, gently swirling it to thoroughly mix the agarose solution.
10) Sufficient hot distilled water was added and mixed thoroughly to obtain the initial weight.

11) The solution was cooled to 50 - 60°C prior to casting.

12) Once the gel was cast, the molten agarose was cooled at room temperature. The gel was then placed at 4°C for 20 minutes to obtain optimal resolution and gel handling characteristics.

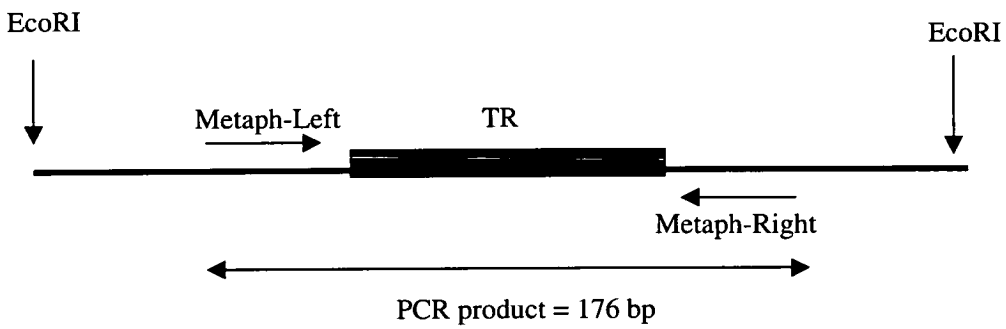


Figure 5.1. Schematic representation of the DNA fragment amplified by primers Metaph-Left and Metaph-Right.

Trinucleotide repeat arrays of different lengths (CAG_{39} and CTG_{43} from plasmids pDL1181 and pDL 1183) were amplified as controls. Their electrophoresis did not show a conclusive difference of 12 bp as expected (Figure 5.2). PCR products obtained from chromosomal amplification with primers Metaph-Left and Metaph-Right showed only one DNA product (Figure 5.2).

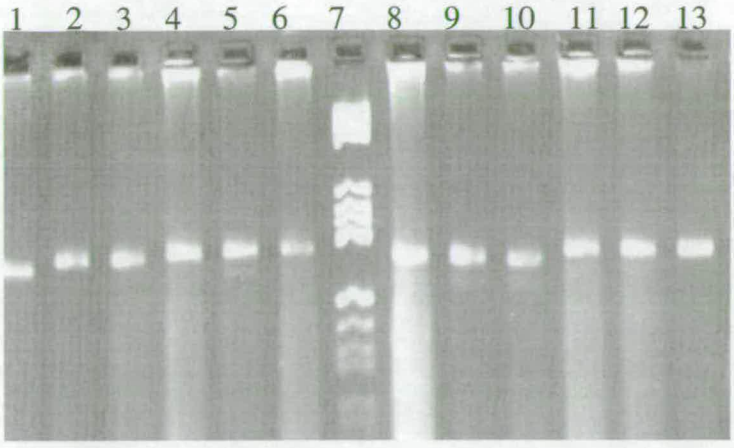


Figure 5.2. 3% MetaPhor agarose gel. Lanes 1 and 10: $(CAG)_{39}$; lanes 2 and 11: $(CAG)_{43}$; lanes 3, 4, 5, 6, 8, 9, 12, 13: PCR products from chromosomal amplification of DL1294 (*mutS*); lane 7: marker V.

The conclusion of the MetaPhor analyses was that differences of only one trinucleotide could not be easily noticed. This method did not show the expected efficiency, so other new methods had to be investigated, as described in the next sub-chapters.

5.3 Analysis of the Trinucleotide Repeat Tracts by 6% PAGE of *EcoRI* Digested Long PCR Products

A new strategy was employed in order to analyse subtle changes in trinucleotide repeat lengths, following the poor efficiency of the MetaPhor agarose method. The new strategy was based on the existing method for analysing TR in plasmids (Chapter 2.2.3. Radioactive labelling of double-stranded DNA fragments), but modified for TR in chromosomes, which meant the introduction of an additional PCR step. The primers for this method were already available, primers Genom1 and Genom2, as described in Chapter 4 (Figure 4.3). They amplify 2.2 kb of the chromosomal region containing the trinucleotide repeat tract integrated at the *attB* site.

Cells were cultured on LB agar in the presence of the appropriate antibiotics. Single colonies were transferred into 30 µl sterile water and heated to 99°C for 5 min to lyse the cells. Lysates were spun at high speed in a microcentrifuge for 2 min to pellet cell debris. The supernatant was used as template DNA for PCR reactions using primers Genom1 and Genom2. Table 5.2 shows the Genom(1+2) PCR programme. Products were checked by agarose gel electrophoresis (Figure 5.3). The PCR products were purified from excess deoxynucleoside triphosphate (dNTP) by ethanol precipitation: 45 µl PCR product were added to 180 µl cold ethanol and incubated at -70°C for 30 min. Samples were spun at high speed in a microcentrifuge for 15 min and the ethanol removed by blotting. Another 180 µl

cold ethanol were added and then centrifuged for 5 min and the ethanol removed. The DNA was left to dry completely prior to addition of 40 µl deionised water.

Table 5.2. Genom(1+2) PCR programme

Denaturation		Annealing		Elongation		Cycles
Time	Temp	Time	Temp	Time	Temp	
4 min	94°C					1
45 sec	94°C	45 sec	68°C	2min 30sec	72°C	30
				10 min	72°C	1

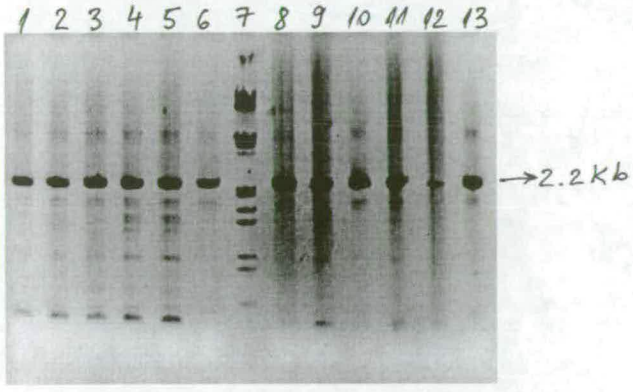


Figure 5.3. Agarose gel electrophoresis of PCR products. Main products have 2.2 kb. Lanes 1-6)wild-type strains; lane 7) Marker III; lanes 9-13) *mutS* strains.

The 2.2 kb fragment was digested with *EcoRI* to give three fragments, one of them of 220 bp containing the trinucleotide repeat tract (Figure 5.4).

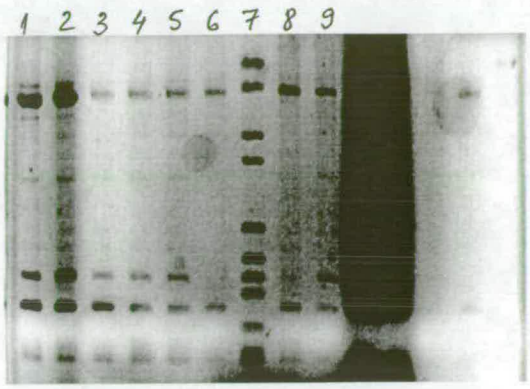


Figure 5.4. 1% agarose gel electrophoresis of *EcoRI* digestion of PCR products from Figure 5.3.

The 3' recessed ends generated by *EcoRI* restriction were labelled with α-^{35}S dATP using the Klenow fragment of DNA Polymerase I. 15 µl of each digested sample were incubated with 0.5 µl (1U) Klenow enzyme and 0.5 µl (5 µCi) α-^{35}S dATP at room temperature for 20 min. dNTP-4 mix solution was then added and incubated at room temperature for a further 10 min. 3.5 µl of loading buffer were added and the samples loaded on a 6% non-denaturing polyacrylamide gel (Sequi-Gen® apparatus from BioRad). Electrophoresis was carried out at 40W until the dye front reached the bottom of the gel (approx. 1.5 hours). The separated DNA fragments were visualised by autoradiography (Chapter 2.2.3).

Figure 5.5 shows DNA from strains DL1184(wt) and DL1294(*mutS*). Trinucleotide repeat instability can be observed in both strains in the form of ladders of bands differing in size by multiples of three bp.

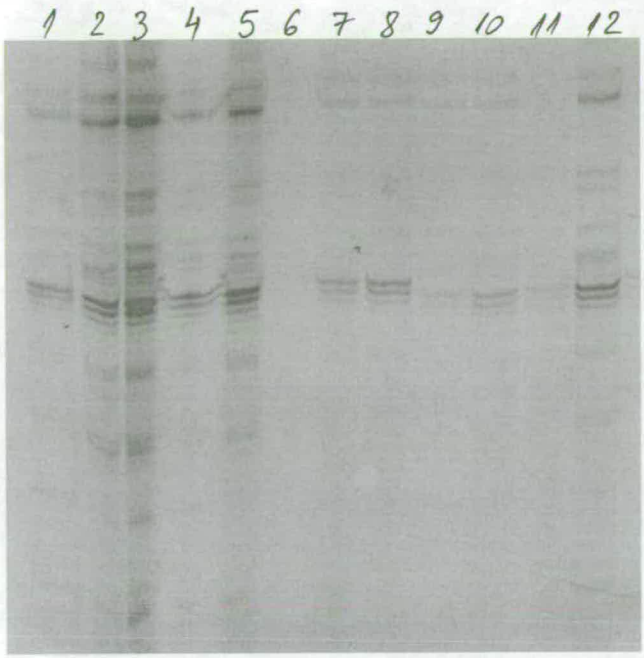


Figure 5.5. Non-denaturing polyacrylamide gel electrophoresis of Klenow-labelled DNA fragments. Lanes 1-4) DL1184 (wt), lanes 5-12) DL1294 (*mutS*). The lack of difference between the wild-type strains (DL1184) and *mutS* deficient strains (DL1294) may have a technical explanation.

The single trinucleotide expansions and deletions seen in Figure 5.5 may be due to slippage during the extension phase of the PCR. This highlights a potential problem with this method of analysis as it means that one cannot be certain whether differences in apparent stability are real or PCR artefacts.

It was found that successful Klenow labeling of the *EcoRI*-digested PCR product was dependent on the careful execution of the ethanol precipitation step. This was probably due to the presence of dNTP molecules from the PCR reaction in

non-purified samples. These molecules could potentially interfere with the labeling reaction in two ways. Firstly, it is thought that the Klenow enzyme must first use its 3'→ 5' exonuclease activity to further degrade the 3' recesses of the DNA fragments before filling them in. It is possible that the presence of excess dNTP molecules at this stage could shift the equilibrium of the enzyme towards filling in, rather than degrading the ends of the DNA fragments. Secondly, an excessively high concentration of non-labeled dATP in the reaction mixture could lead to a lower frequency of incorporation of ^{35}S-dATP.

It can be seen, therefore, that this method of trinucleotide repeat stability analysis is not without its problems. These problems were not observed in experiments where the trinucleotide repeats were integrated in plasmids as there is no PCR step in such protocols. This was the underlying idea of the next investigated strategy, described in sub-chapter 5.4.

5.4 *AvrII* digestion of Chromosomal DNA

The consecrated methods for studying the changes in TR tracts in plasmids were not applicable to the TRs integrated in chromosomes. For example, digestion of the entire *E. coli* chromosome with *EcoRI* (method previously developed in the Leach laboratory for plasmids described in Chapter 2.2.3.11 *Radioactive labelling of double-stranded DNA fragments*) would generate a large number of fragments. The

separation of the *EcoRI* fragment containing the TR would be almost impossible among such a large number of fragments with identical sticky ends and various lengths. My novel approach was to find an alternative restriction endonuclease, which would cleave the DNA sequence of the *E. coli* chromosome in a limited number of fragments of large size. Three restriction endonucleases: *NotI*, *AvrII* and *SfiI* fit this description, according to the New England Biolabs catalogue. *NotI* would be the best choice since the fragments generated by digesting the *E. coli* chromosome have an average size of 200,000 bp. Unfortunately, using the *NotI* enzyme was not possible at this stage of the project, taking into consideration that *NotI* sites were present in plasmids pLDR9 and pLDR11 (Figure 3.2). This meant that there was a *NotI* site in the close proximity of the TR fragment, as shown in Figure 4.3. Another technical difficulty was created by the *NotI* recognition sequence itself (GC/GGCCGC), which would not allow Klenow-filling in with [α-^{35}S] *dATP* (cheaper and easily available) at a later stage. This disadvantage was overcome by using the *AvrII* enzyme, which generates average products of 150,000 bp and has the recognition sequence C/CTAGG. Radioactive α-^{35}S *dCTP* was used in the labelling reaction since the first nucleotide to be added in order to fill in the sticky *AvrII* ends was *dCTP*. Two main reasons led to discontinue the use of α-^{35}S-*dCTP*: high cost and poor labelling efficiency. The old protocol developed in the Leach laboratory for Klenow-filling in of sticky ends was designed for *EcoRI* ends (Chapter 2.2.3 Radioactive labelling of double-stranded DNA fragments). It had to be modified because the *AvrII* sticky ends require both *dCTP* and *dTTP* in order to be radioactively labelled with α-^{35}S-*dATP*. The new adapted protocol used in the course of this project is described below:

Chromosomal DNA was incubated with restriction endonuclease *AvrII* at 37°C for 2h. *AvrII* produces 3'-recessed ends which can be filled-in by Klenow enzyme. There was not any inactivation step since *AvrII* enzyme can not be heat-inactivated. The samples were spun in a bench-top centrifuge at 9 krpm for 30 seconds. Because of the sequence of the overhang produced by *AvrII*, non-radioactive *dCTP* and *dTTP* were used together with [α-^{35}S] *dATP* in the end-labelling reaction. After incubation at room temperature for 10 minutes, 2 µl of the 2 mM dNTP-4 mix were added to the labelling reaction and incubation continued for 10 minutes at room temperature. Immediately, 20 to 25 µl of this reaction mixture were loaded onto a 6% polyacrylamide gel to separate the labelled DNA fragments.

The *AvrII* protocol produced faint gel images, which showed only one fragment of DNA containing the trinucleotide repeat tract (Figure 5.6). This result is backed up by other methods (fluorescent PCR and *f*-TRAMP) described in sub-chapters 5.4 and 5.5.

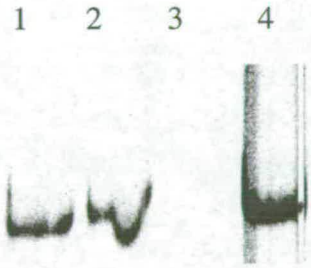


Figure 5.6. PAGE of DNA fragments containing chromosome-borne $(CTG)_{43}$ does not show obvious deletions or expansions. Lanes: 1, 2, 4) chromosomal DNA containing $(CTG)_{43}$; lane 3) control: chromosomal DNA without $(CTG)_{43}$.

5.5 Fluorescent PCR

Currently used methods for the measurement of the length of trinucleotide repeats that are not on high-copy number plasmids often require amplification of the trinucleotide repeat array by PCR prior to sizing by electrophoresis.

During the course of my project, a 310 Genetic Analyser (a very modern apparatus) became available for the study of the trinucleotide repeat instability. Fluorescently labelled DNA fragments are introduced into a polymer-filled capillary for electrophoresis and then separated according to their size. The fragments travel through capillary into the read window, where laser excites fluorescent labels. Fluorescent emissions are collected and analysed with a software programme, GeneScan. The 310 capillary electrophoresis takes place in a very narrow capillary (50 μm) which is filled with a separation medium (Performance Optimiser Polymer – POP4). Sample loading into the capillary is by electrokinetic injection (capillary is dipped into the sample and voltage is applied for a specified time). GeneScan software allows the use of up to three experimental colours in association with a size standard. It collects the fluorescent emissions of the dyes through one of four filter sets. Filter set D was set-up for all the runs on the 310 Genetic Analyser. It allows the use of FAM and HEX labels, together with a ROX labelled size standard and it is the best filter since the wavelengths of the fluorescent dyes are well separated apart. For most of the experiments, FAM-labelled PCR products were run together with the corresponding HEX-labelled *f*-TRAMP products (Figure 5.9). The 310 Gentic Analyser has a high output (47 samples/day) and it allows accurate separation of DNA fragments, which are only 1 base pair different in size.

The same two primers mentioned in Chapter 5.1 were used, the only difference this time being the FAM dye attached to the Mataph-Left primer (Figure 5.7). As a result of successfully improving the Metaphor PCR protocol for single colonies, the PCR products were fluorescently labelled, which made possible their detection and analysis on the 310 Genetic Analyser using the GeneScan® software.

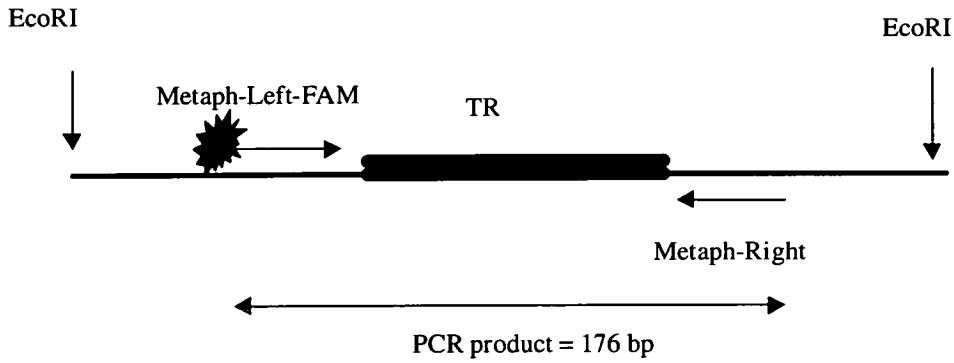


Figure 5.7. Schematic representation of the Fluorescent Metaphor PCR.

Due to the repetitive nature of the template, such PCR-based methods can be plagued with *in vitro* artefacts, mostly due to slipped strand mispairing, that are exponentially amplified during the PCR reaction leading to stutter bands, and thus to uncertainty as to the exact number of repeat units (Figure 5.8). In addition, the requirement for a matching primer pair leaves usually little choice of PCR product length and primer sequence. The smaller peaks from Figure 5.8 represent inherent artefacts of trinucleotide repeat amplification. They are caused by the limited processivity of Taq DNA polymerase enzymes (approxymately 60 nucleotides). The polymerase dissociation allows template strands to breathe apart, which leads to

slipped strand mispairing. At the same time, the stutter products represent a good diagnostic for identifying the genuine PCR products, as opposed to other artefact PCR products (*e.g.*:primer-dimer).

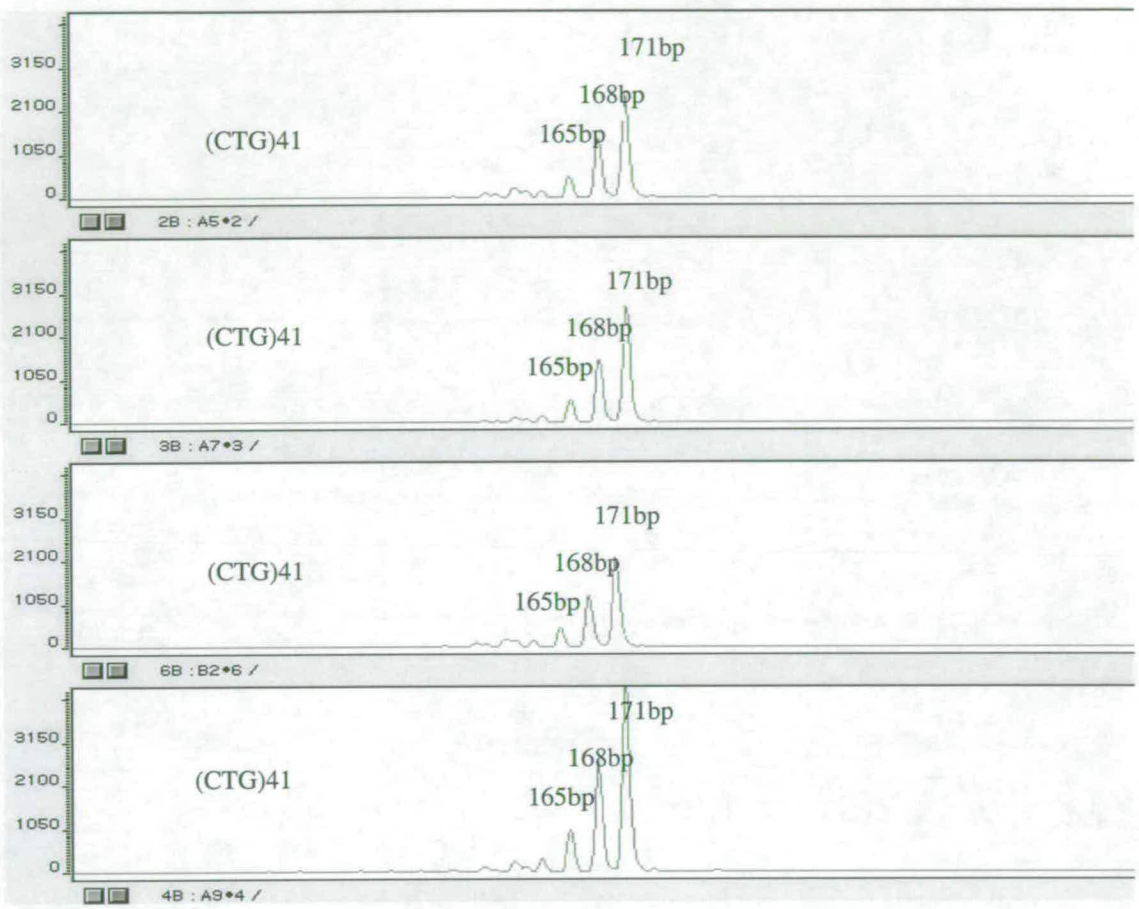


Figure 5.8. Gene Scan analysis of fluorescently labelled PCR products containing the trinucleotide repeat array. The four pannels show four different single colonies of DL1404 (wild-type with CTG on the template for the leading strand).

Such PCR artefacts hamper analysis when the trinucleotide repeat is unstable *in vivo* [*e.g.* $(CTG)_{84}$] because the expected variation in repeat array length due to *in vivo* instability is indistinguishable from additional repeat length variation

that may have been generated *in vitro*. This uncertainty always exists with PCR unless it is carried out on a single cell that carries a single copy of the repeat (small-pool PCR). Taking into consideration these disadvantages created by the PCR, the fifth method was investigated, as detailed in the next section.

5.6 *f*-TRAMP: *f*luorescent Trinucleotide Amplification of DNA Fragments

In order to avoid the disadvantages of PCR-based methods, I employed an assay already existent in the Leach laboratory: *f*-TRAMP (K. Schmidt's PhD thesis, University of Edinburgh). *f*-TRAMP allows amplification and length measurement of trinucleotide repeats, and any other DNA sequence that has at least one nucleotide missing, in plasmids and in the chromosome of *E. coli* without the need for restriction sites, radioactive labels and amplification by PCR. This assay is based on an idea of Yamamoto and co-workers (Yamamoto *et al.*, 1992) who used only three of the four dNTPs in a single round of primer extension to radioactively label CAG repeats which had been amplified by PCR from the human androgen receptor gene. The advantage of the *f*-TRAMP over PCR is that it produces only one peak for a trinucleotide repeat tract, unlike PCR, which gives additional stutter bands (Figure 5.8).

The Principle of *f*-TRAMP: a thermostable DNA polymerase is used to extend a single fluorescently labelled oligonucleotide in repeated cycles of DNA melting, primer annealing and primer extension. Primer extension in each cycle is automatically terminated by the absence of the nucleotide, which is not present in the amplified sequence. Copying of the CAG repeat requires only dCTP, dGTP and dTTP so that primer extension is terminated at the first dTTP when dATP is required but not present in the reaction mixture. Optionally, ddATP can be added to the reaction mixture as a fourth nucleotide in order to guarantee chain termination.

Because of the single primer in the reaction mixture, products of one cycle are not used as templates in the next round of amplification. Hence, every cycle of amplification demands the original template and errors introduced into the sequence by polymerase slippage during *in vitro* amplification do not give rise to artificial repeat populations that are commonly reported after amplification of trinucleotide repeats by PCR (Rubinsztein *et al.*, 1996) and are also clearly visible in Figure 5.8.

In contrast to the method used by Yamamoto, error-prone PCR has been eliminated from the process thereby increasing the confidence in the sizing of trinucleotide repeats in the population. In addition, the utilisation of fluorescent dye labels allows the rapid analysis on an automated DNA sequencer and subsequent automated quantification and sizing of amplified products using GeneScan™ analysis software.

f-TRAMP was applied to measure CTG repeats both in plasmids and in the *E. coli* chromosome (Figure 5.9). Single colonies of the *E. coli* strains of interest were used to inoculate 5 ml of L broth. After overnight incubation at 37 °C and under vigorous shaking, genomic DNA was prepared using the Genomic DNA Kit from

AGTG®. DNA was precipitated at room temperature with isopropanol, washed twice with 70 % ethanol, air-dried and dissolved in 50 µl to 100 µl of sterile 10 mM Tris-buffer. A Rapidcycler (Idaho Technology) was available for carrying out the primer extension reactions. This cycler uses micro-capillary tubes as reaction containers and high velocity air as a heating and cooling medium. Some changes in the composition of the reaction mixture and the amplification programme were required.

In general all PCR-buffers used with the Rapidcycler are free of potassium ions and PCR reactions contain 500 µg ml^{-1} crystalline BSA to prevent denaturation of the polymerase on the surface of the glass capillary. The Bio/Optimizer™ Kit for Rapidcycler provides appropriate buffers with Mg^{2+} concentrations varying between 1 – 5 mM. For amplification from genomic DNA approximately 100 ng to 200 ng of DNA were used in 10-µl reactions. Typical cycling conditions are listed in Table 5-3.

Table 5-3. Cycling conditions for the amplification of $(CTG)_{43}$ repeats using the Rapidcycler. Times are shown in seconds and temperatures are in degrees Celsius (* an initial denaturation step of 1 minute at 94 °C and a final elongation step of 1 minute at 72 °C were added to the programme).

Type of template	Denaturation		Annealing		Elongation	
	Time	Temp	Time	Temp	Time	Temp
Plasmid DNA	0	94	0	55	15-20	72
Genomic *E. coli**	2	94	0	55	20	72

f-TRAMP was carried out in 10-µl reaction volumes of the following composition: 3 µl of genomic DNA (500 ng to 1 µg), 1 x Buffer 1769 (Bio/Gene

Ltd.), 200 µM of each dNTP except dATP, 20 pmole primer T2049 (HEX or FAM labelled) and 2.5 U Taq-Polymerase (Boehringer Mannheim). The cycling conditions for primer extension were:

1 cycle of	denaturation	at 94 °C	for 2 minutes
50 cycles of	denaturation	at 94 °C	for 5 seconds
	primer annealing	at 50 °C	for 0 seconds
	primer extension	at 72 °C	for 20 seconds

In order to preserve all potential primer extension products, samples were not purified after the programme had been completed. Instead, 2 µl were taken directly from each sample and mixed with 0.5 µl (4 nM) of internal lane standard (GeneScan-500™ ROX / Applied Biosystems) and 1 µl of Hi-di formamide which was supplied with the internal lane standard. Samples were boiled for 2 minutes and analysed on an ABI 310 Genetic Analyser.

f-TRAMP gives clear reproducible results for chromosomal DNA when the template was previously digested with a rare-cutting restriction endonuclease (*AvrII*) (Figure 5.9). This suggests that the fragment containing the trinucleotide repeat array is much more easily accessible to the single primer when the chromosomal DNA is fragmented. Further improvement is required for the *f*-TRAMP method in order to produce clear results without the additional step of restriction digestion of the chromosomal template.

f-TRAMP and *f*-PCR are valuable methods which give clear answers concerning the trinucleotide repeat tracts lengths, especially when used together. The disadvantage of "stutter" bands of the *f*-PCR is overcome by *f*-TRAMP, which gives only one signal peak for one length of the trinucleotide repeat tract. The potential problem for the *f*-TRAMP method would be that products could be shorter than the real length because of interruptions in the trinucleotide repeat array which include the fourth nucleotide. Performing *f*-PCR on the same template DNA could rapidly signal such interruptions. The size difference between *f*-PCR and *f*-TRAMP products from the same template should be of 27 bp, since the expected PCR product with 43 copies of repeats has 176 bp and the *f*-TRAMP product only 149 bp. The difference of approximatively 27 bp is consistent in all four pannels in Figure 5.9, indicating that no interruptions occurred in the trinucleotide repeat tract. It is acceptable for the difference not to be exactly 27 bp since different fluorescent dyes affect the mobility of the DNA fragments in different amounts. The different repeat array sizes in the four pannels in Figure 5.9 are attributed to the fact that different lengths were integrated into the chromosome in the process described in Chapter 4.2 (Constructing various *E. coli* mutant strains with the trinucleotide repeat tract integrated at the *attB* site).

These two methods proved to be the most powerful ones in investigating trinucleotide repeat instability and they were the preferred methods used for the analysis described in Chapter 6.

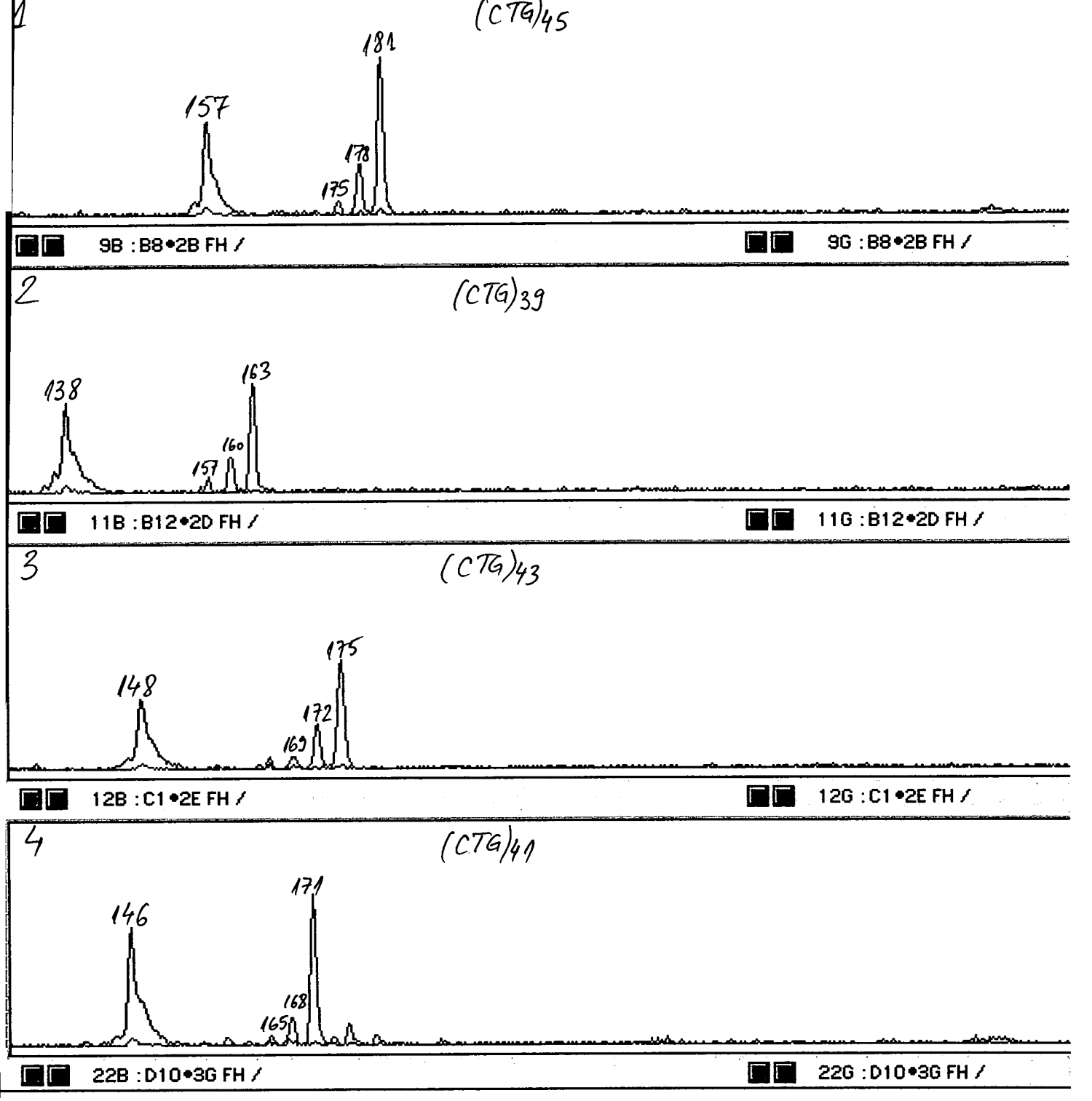


Figure 5.9. GeneScan analysis of *f*-TRAMP products (green) and *f*-PCR (blue). Four pannels with four different strains with CTG on the template for the leading strand: 1)DL1432 (wt); 2)DL1428 (*polA polB)*; 3) DL1430 (*polA)*; 4) plasmid pUC18 with $(CTG)_{41}$.

Chapter 6

Analysis of the $(CAG)(CTG)_{43}$ Repeat Instability in Replication, Recombination, Post-Replicative Mismatch Repair Deficient and Wild-Type *Escherichia coli*

6.1 Introduction

The expansion of triplet repeat sequences is the causative mutation in a large number of human hereditary neurodegenerative diseases (Chapter 1). The mechanisms by which trinucleotide arrays expand are not well understood. The aim of my project was to investigate the instability of a trinucleotide repeat array at the *attB* site of the *E. coli* chromosome in various genetic backgrounds.

Unstable transmission can occur either as a large increase (class I) (generally > 10 repeats) in which the rate of expansion far exceeds the rate of contraction, or as smaller changes (class II) (generally < 4 repeats) that are associated with a relatively equal rate of insertion and deletion mutation (McMurray, 1995). Class I changes are associated with HD, Fragile X, SBMA, DM, SCA1, DRPLA, MJD (see Chapter 1). Class II instability are linked to hereditary colon cancer.

The discovery that mutant DNA mismatch repair genes are a cause of instability in hereditary forms of colon cancer suggested that mistakes made during replication are at the source of the expansion mechanism.

Trinucleotide repeat instability was intensively studied in plasmids in different *E. coli* strains. Jaworski (Jaworski *et al.*, 1995) reported that the long triplet repeat $(CAG)_{175}(CTG)_{175}$ had a reduced frequency of large deletions in *mutH, mutL* or *mutS* backgrounds when the CTG tract comprised the leading template strand. Little effect of methyl-directed mismatch-repair was observed when the CTG tract was the lagging template strand. In an apparently contradictory report, Schumacher (Schumacher *et al.*, 1998) reported that mutations in methyl directed mismatch repair did lead to instability of $(CTG)_{64}$ in both orientations in *E. coli*. K. Schmidt (Schmidt *et al.*, 2000) reconciled these findings, demonstrating that both conclusions are valid and that the apparent inconsistencies result from differences in experimental systems, including the homogeneity and length of the repeat, plasmid and bacterial strain. Schmidt examined changes in the distribution of mutant lengths from a starting tract of $(CAG)_{43}$ $(CTG)_{43}$ at intervals up to about 140 generations. Over time the proportion of plasmids with the starting length diminishes and the proportions with other lengths increase. In wild-type cells a range of 5 – 69 repeats was reached, but the rate of dispersion from the original length was greater when the lagging strand template carried CTG than when it carried CAG. In mismatch-repair deficient cells, the repeat tract was more unstable and this was independent of orientation. The increase in instability was mainly due to changes of +1 and –1 repeat from 43 which were barely seen in wild type as they are most efficiently detected and corrected by the repair system. Large changes in either direction were attained by gradual accumulation of small changes. These results therefore indicate that large deletions are orientation- and mismatch-repair-dependent and are eliminated by mismatch repair deficiency.

The general conclusion is that expansions have an increased frequency when the CTG sequence comprises the template for the leading strand, while deletions are more prominent when the CTGs are on the template for the lagging strand (Kang *et al.*, 1996). This can be explained by a DNA slippage model and also taking into consideration the findings that CTG hairpins are more stable than CAG hairpins. During replication, directly repeated sequences can provide multiple sites for pairing of a complimentary strand (Figure 6.1).

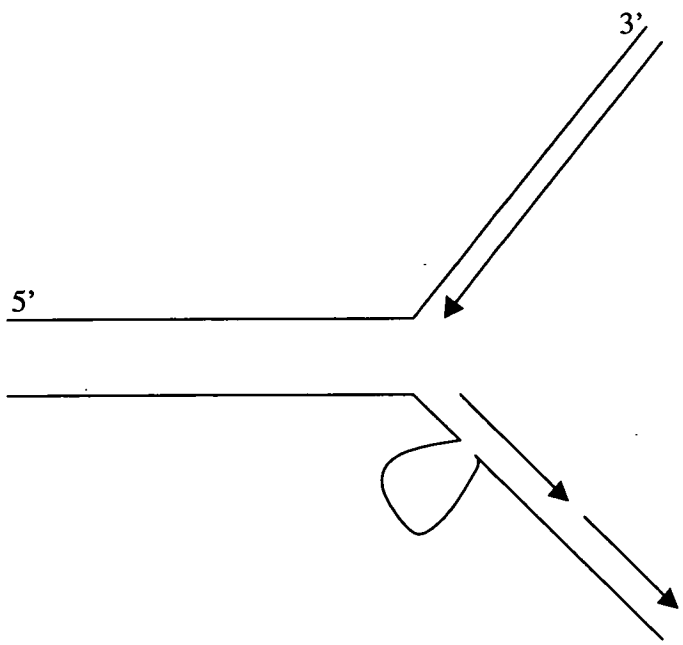


Figure 6.1. DNA slippage during replication when CTG repeats are on the template for the lagging strand.

Another model for instability (McMurray, 1995) suggests that DNA replication is first blocked by a hairpin that formed on the lagging strand and then wrongly re-initiated at the single-stranded hairpin-loop. During replication, the leading strand is copied continuously and has a very short lifetime as a single strand.

In contrast, lagging strand exists as a single strand for a significantly longer period, until the replication fork moves ahead (Nossal, 1992). Since the DNA is repetitive, hairpin formation may occur immediately following the generation of the single strand on the lagging strand. This may initially block polymerase at the fork from replicating DNA at the base of the hairpin. SSB binds to the single stranded loop of a stable hairpin where lagging strand replication may be re-initiated. While copying along one arm of the hairpin, hydrogen bonds are undone, the hairpin is destabilised and the initial replication block is resolved and normal DNA replication is resumed. Assuming that the entire trinucleotide repeat forms one large hairpin, McMurray suggests that an extra half-copy of the repeat is subsequently ligated to the preceding Okazaki fragment. It is not clear from this model why the entire repeat should fold into a hairpin, considering that the instability of the repeats does not always show an increase of an extra half-copy of the parental repeat length.

The expansion of triplet repeats represents a novel type of mutagenic event that has not been identified to occur spontaneously, like it happens in humans, in genetically tractable organisms such as bacteriophage, *E. coli*, yeast or *Drosophila* (Sinden, 1999). In order to shed some light on the mechanisms causing human diseases, the DNA mechanisms of instability are investigated in this Chapter in the model organism *E. coli*.

6.2 CTG Repeat Instability in *E. coli* Strains Deficient in Proofreading and Post-Replicative Mismatch Repair

Radiation and other mutagens from the environment constantly threaten the integrity of the genome. Mistakes in the DNA replication can alter the integrity of the genome. The fidelity of DNA replication in *E. coli* is approximately of 10^{-10} errors/ replicated base (Drake, 1991). This high fidelity of base incorporation is achieved in three steps: base selection by the DNA polymerase, the 3' to 5' exonucleolytic proofreading of the inserted bases and postreplicative MMR. Ten factors are known to be involved in MMR of *E. coli*: MutS (mismatch binding protein), MutL, MutH (a latent methylation-directed and sequence-specific endonuclease), UvrD (DNA helicase II), SSB (single-strand binding protein), ExoI (3' to 5' exonuclease), RecJ, ExoVII (5' to 3' exonucleases), DNA polymerase III and ligase.

MutS, encoded for by the *mutS* gene, is the mispair recognition protein of *E. coli*. MutS recruits MutL and MutH to the mismatched site initiating repair. Then MutS undergoes an ATP-dependent conformational change and loses affinity for the mispair. MutS translocates the DNA through itself and an α-loop forms containing the mismatch (Allen *et al.*, 1997). Then MutH is activated upon encounter with the MutS-MutL complex. DNA helicase II separates the strands so that the error-containing strand is available for exonucleolytic degradation (Friedberg, 1995). The error-containing strand is degraded by the exonucleolytic activity of ExoVII or RecJ or ExoI. Most likely, Polymerase III holoenzyme fills-in the gap and ligase repairs the nick.

In the course of my project I investigated the instability of the trinucleotide repeat arrays in *mutS, mutL* and *mutH* backgrounds. 47 single colonies from each background were initially analysed. Following the finding that in these mutant strains both orientations of trinucleotide repeats were completely stable, I suspected that instability occurs with a much lower frequency, so a much larger number of colonies should be investigated. In order to investigate a larger number of colonies, a different strategy was employed. Ten distinct single colonies were pooled together and analysed. This allowed 470 colonies to be screened at once. The clear result of these analyses was that there are no changes in the repeat tract length in strains deficient in *mutS, mutH* or *mutL.*

Strains deficient in MutS activity were analysed in a separate experiment after they were cultivated at higher and lower temperatures than normal (42°C and 30°C) since previous work showed that the formation of hairpin and triplexes is temperature-dependent (Lyons-Darden & Topal, 1999). They were also analysed after they were kept in stationary phase for a long time (2 weeks), using as control cells which were cycled in and out of the stationary phase for the same period of time. Bowater (Bowater *et al.*, 1996) observed that frequencies of deletions were dramatically increased if the cells passed through stationary phase before subculturing. Typical Gene Scan traces of *mutS* strains with CAG or CTG are shown in Figure 6.2. First panel represents *mutS* grown at 30°C; second panel, at 42°C; third panel, the strain kept in stationary phase; fourth panel, the cells subcultured after they reached stationary phase. All 200 colonies analysed showed the same length of the trinucleotide repeat array.

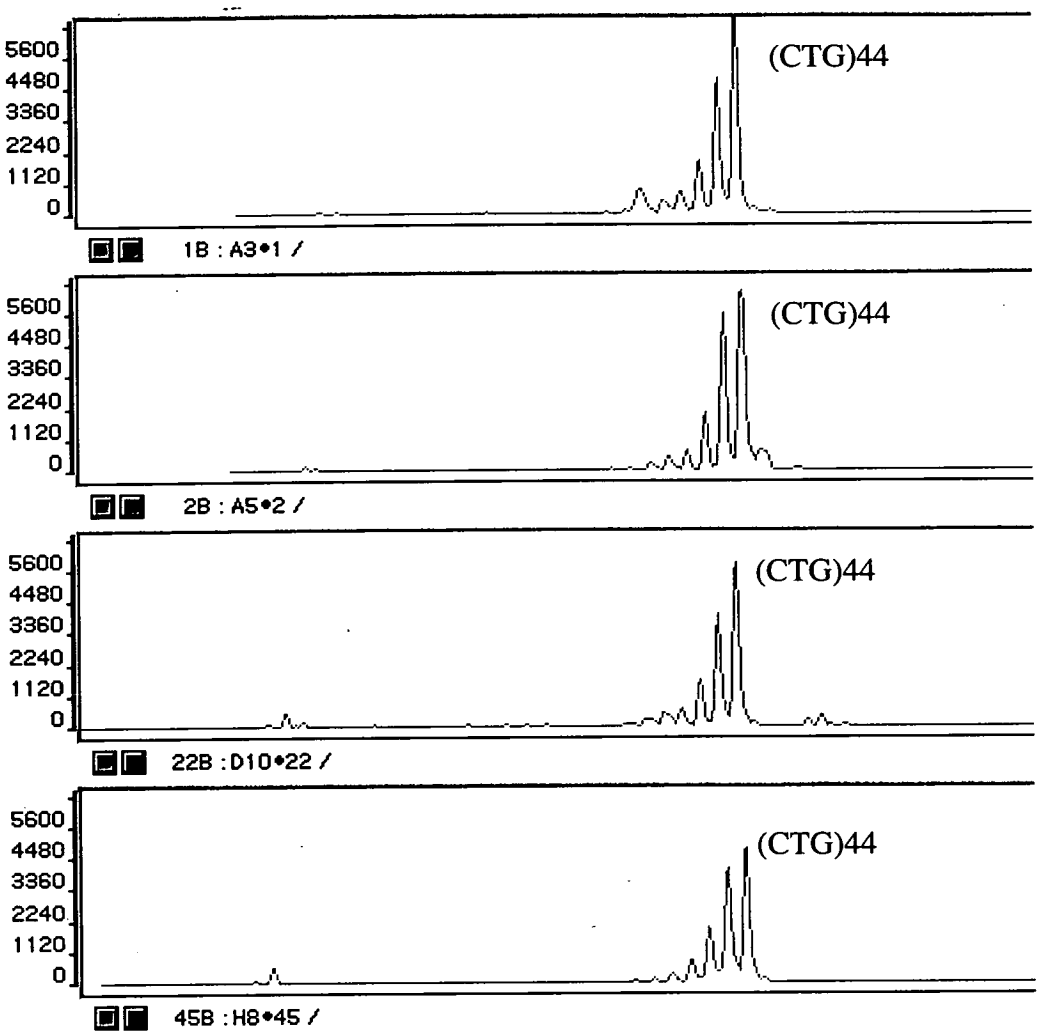


Figure 6.2. No changes were noticed in the length of the trinucleotide repeats for cells grown at 30°C, 42°C, kept in stationary phase or subcultured after reaching stationary phase.

The proofreading function in *E. coli* is performed by the 3' to 5' exonucleolytic ε-subunit of DNA polymerase III, which is encoded by the *dnaQ* gene. On average 1 in 10^5 replicated bases is misincorporated by DNA polymerase at the 3'-end of the daughter strand (Schaaper & Radman, 1989) and the proofreading

activity of the DNA polymerase holoenzyme corrects these mistakes decreasing the error-rate to 10^{-7}. The contribution of proofreading to the overall fidelity is 10^{2}.

mutD is the most potent known mutator in *E. coli* (Schaaper & Radman, 1989). The high mutagenic capacity of the *mutD* deficient strains led my project into investigating the instability of trinucleotide repeat tracts in a *mutD* background. Some colonies of *mutD* with CAG repeats on the template for the leading strand showed deletions (Figure 6.3).

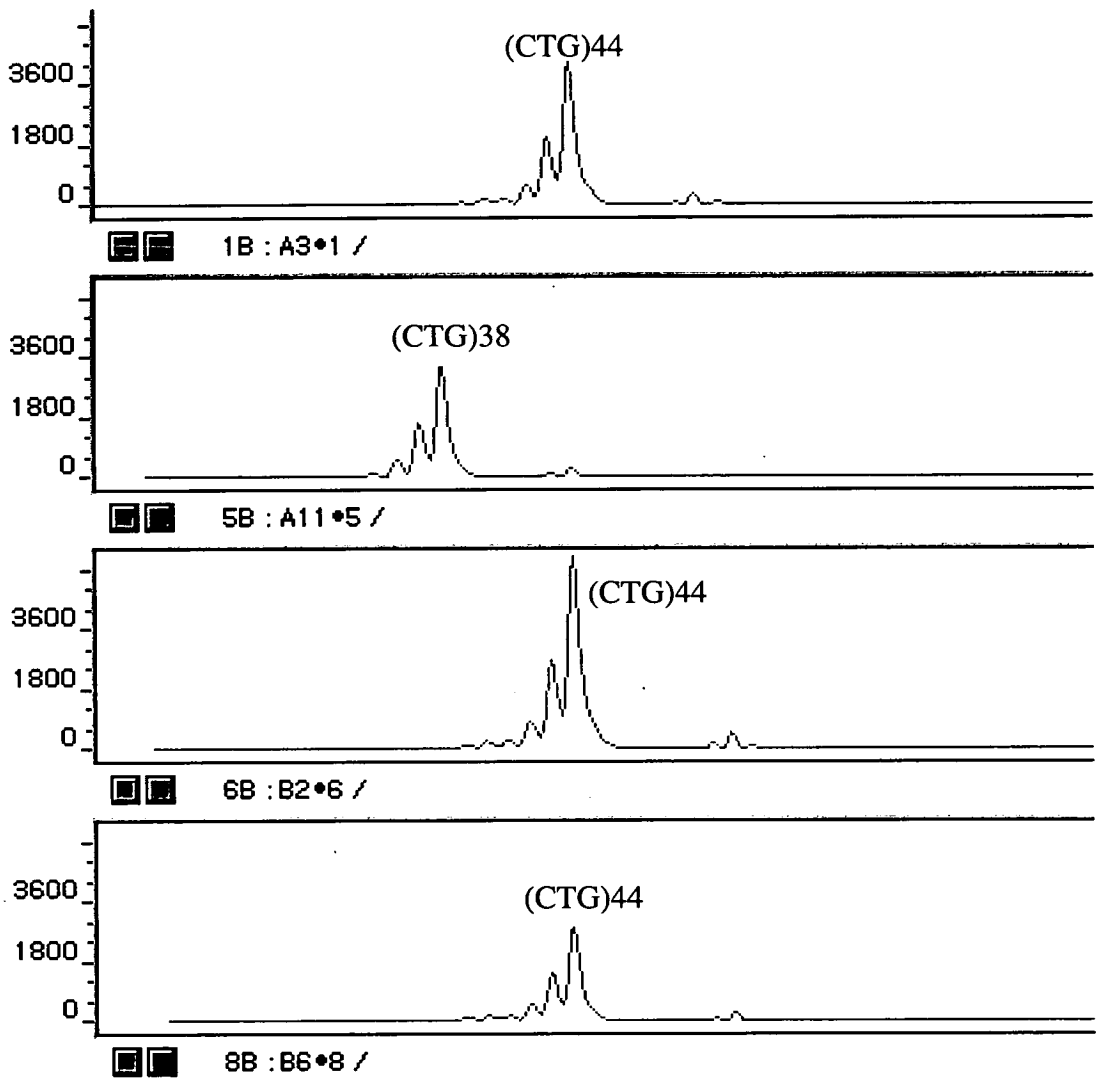


Figure 6.3 A deletion of $(CTG)_6$ was noticed in a *mutD* mutant.

In an endeavour to establish the frequency of deletions in the *mutD* mutant, ten colonies were pooled together. A total number of 470 colonies were analysed. The electrophoregrams (Figure 6.4) did not show any instability, in spite of the expected result.

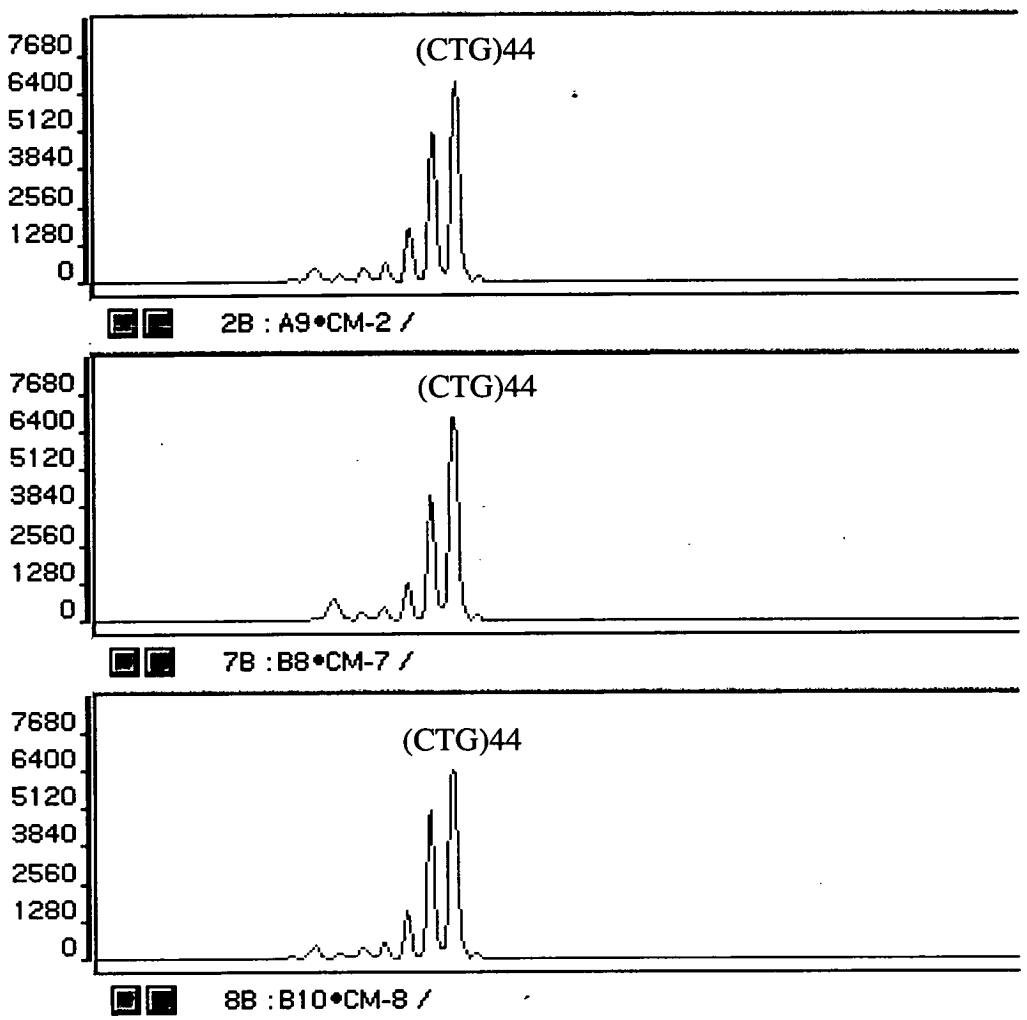


Figure 6.4. Ten colonies of *mutD* with CAG on the template for the leading strand pooled together do not show instability.

I suspected it might be some problem with the detection method since single colonies show deletions, but ten colonies pooled together do not. It might be that if only one colony in ten has a deletion, then in the process of amplification, that repeat length loses out, in competition with 9 times more template for a different length.

In order to determine the accuracy of the results generated by pooling ten colonies together, I investigated which would be the ideal number of colonies to be pooled together. The DNA from a colony which was already known to have a significant deletion in the trinucleotide repeat tract was mixed with the DNA from a colony known to have the full length, without deletion, in order to mimic the pooling of two colonies together. Similarly, the deleted DNA was mixed with the DNA from two colonies with full lengths, pooling three colonies. And so on, up to mixing the colony with shorter repeat tract with nine full-length colonies, pooling ten colonies in total. The results of this experiment for pooling one to five colonies are reproduced in Figure 6.5, while pooling six to ten colonies are shown in Figure 6.6. The conclusion was to pool six colonies together since there is not any risk of missing noticing any existent changes.

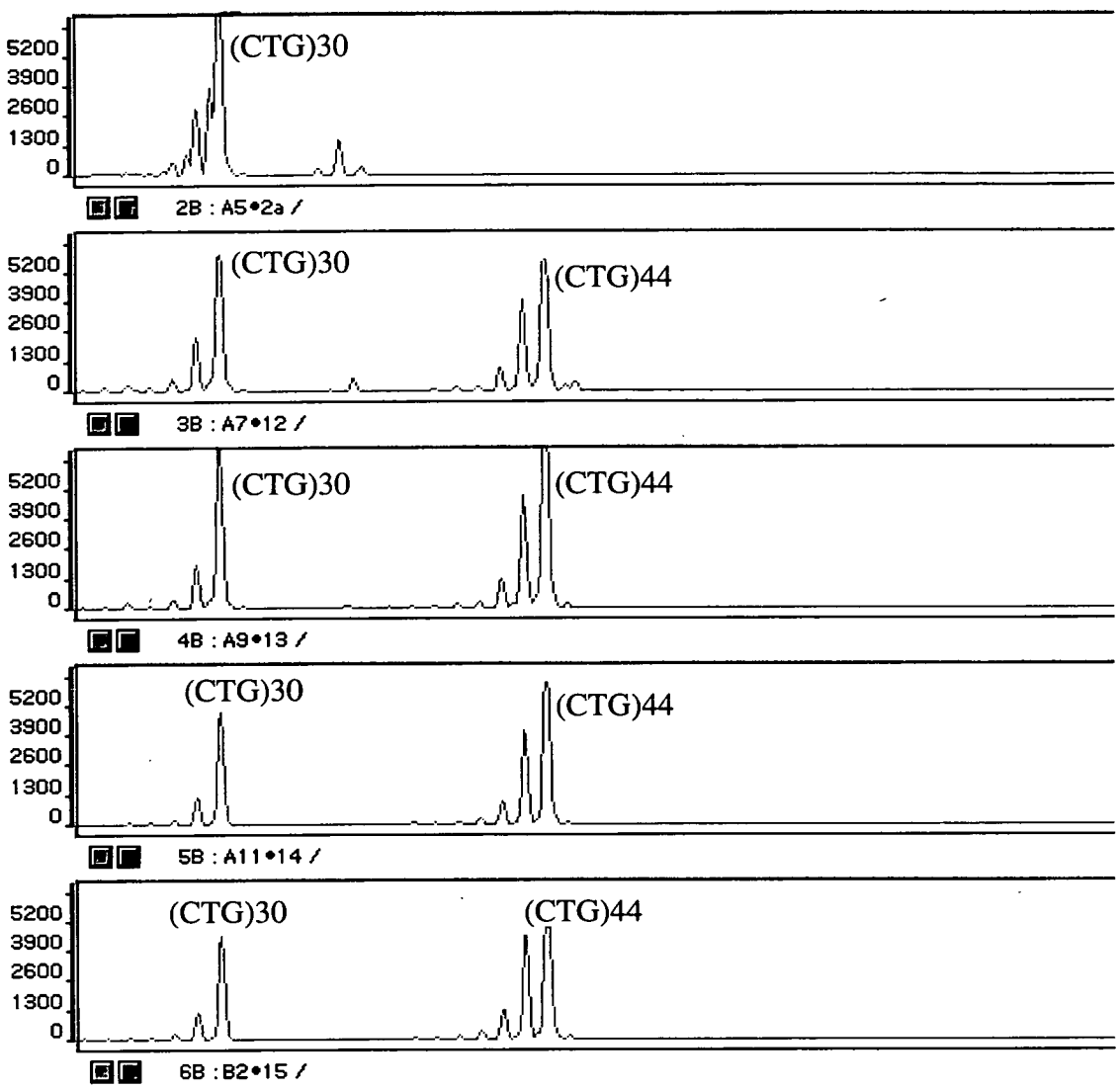


Figure 6.5. Gene Scan analysis of pooling different number of colonies together. First panel) one colony with a deletion in the repeat tract; 2) 2 colonies pooled; 3) 3 colonies; 4) 4 colonies; 5) 5 colonies.

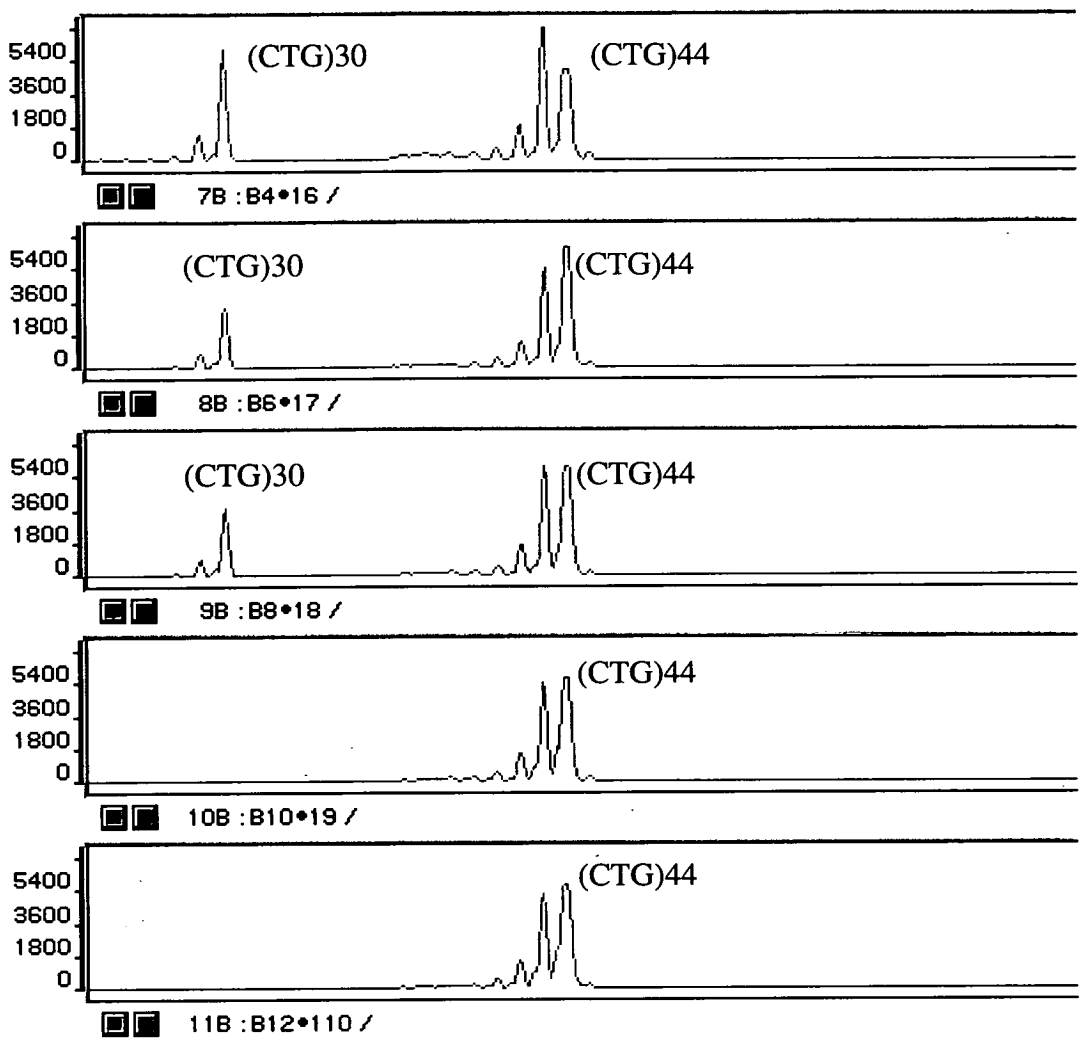


Figure 6.6. Gene Scan analysis of pooling together six to ten colonies. The electrophoregrams show that the shorter length losses out in the competition with nine times more template of a different length.

The overall frequency of deletions occurring in *mutD* CAG on the template for the leading strand, after pooling six colonies together was found to be one in twelve colonies. Examples of different lengths of deletions are presented in Figure 6.7.

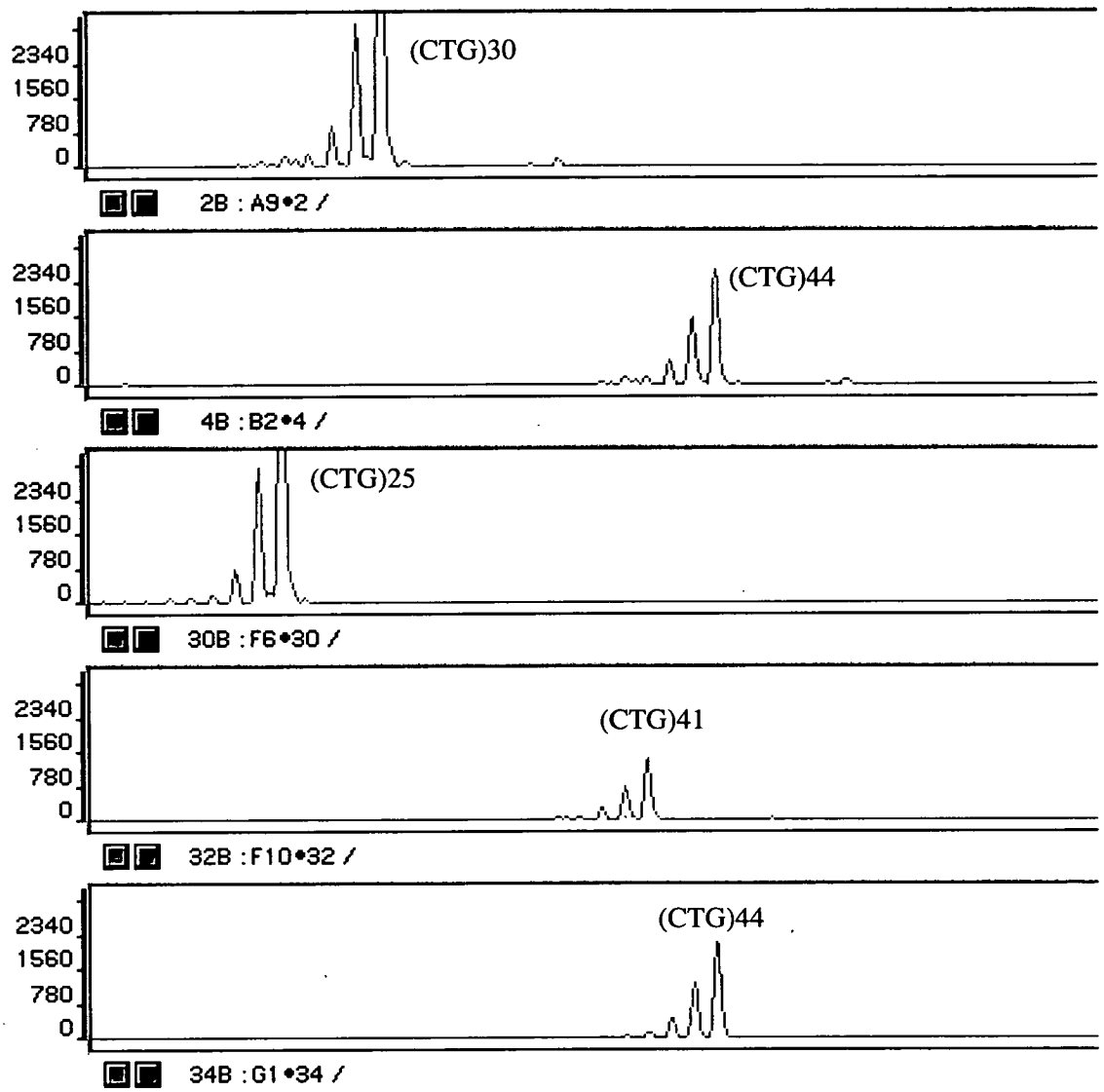


Figure 6.7 Different deletion lengths in *mutD* CAG template for the leading strand.

It is puzzling that deletions in *mutD* strains are detected only when CTG repeats are the template for the lagging strand and not also in the opposite orientation. It might be that instability when CTG comprises the lagging strand occurs also in the wild-type background, but with a very low frequency, which I could not detect. This instability would be a consequence of the activity of an intact mismatch repair system, in agreement with the previous observations from the Leach laboratory (K. Schmidt). In the absence of the proofreading activity of the MutD protein, 10^2-10^3 more errors are generated (Schaaper & Radman, 1989). The mismatch repair proteins have much more errors to correct (around 10^2- 10^3 fold). It is likely that the system does not cope well with such a large amount of work and as a result, more deletions are generated. This brings the instability frequency to a detectable level (1 in 12 colonies).

6.3 CTG Repeat Instability in *E. coli* wild-type and recombination deficient strains

Several studies showed that (CAG/CTG) repeats behave like palindromes forming hairpins *in vitro* (Connelly *et al.*, 1999). It was also shown that SbcCD cleaves pseudo-hairpins (Leach *et al.*, 1997). It was expected that *recA* or *recB* mutants not to replicate successfully the trinucleotide repeat tract integrated into the chromosome. This deficiency was expected to be overcome in *sbcC* mutants. The

analysis of trinucleotide repeat tracts in the chromosomes of *recA*, *recB* and other mutants with altered proteins involved in replication showed the same pattern of stability like the wild-type (Figure 6.8).

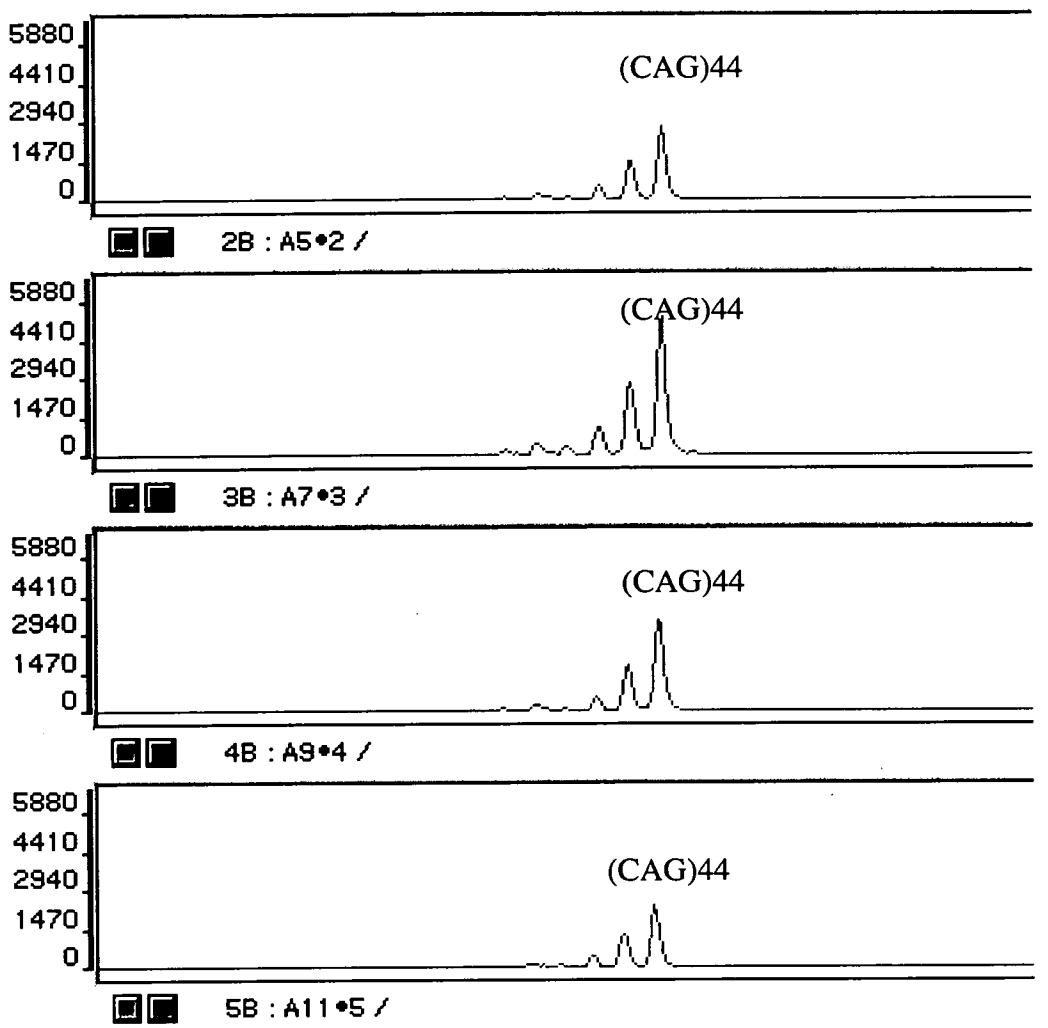


Figure 6.8. Analysis of the trinucleotide repeat tracts in wild-type cells.

It might be that the repeat tract forms secondary structures cleaved by SbcC in the chromosomes at a low frequency, which could not be detected. The evidence so far showed that SbcC cleaves hairpins *in vitro,* but we do not know for sure if it also cleaves hairpins *in vivo.* Another scenario would be that the hairpin formed by

the trinucleotide repeat tract is protected from SbcC cleavage *in vivo* by something else, maybe MutS.

All mutant strains with trinucleotide repeat tracts constructed (Chapter 4) were analysed. It is clear that each strain could integrate a different length of the repeat tract. It is intriguing that the *polA* mutant integrated a larger array than all the other mutant strains (Figure 6.9).

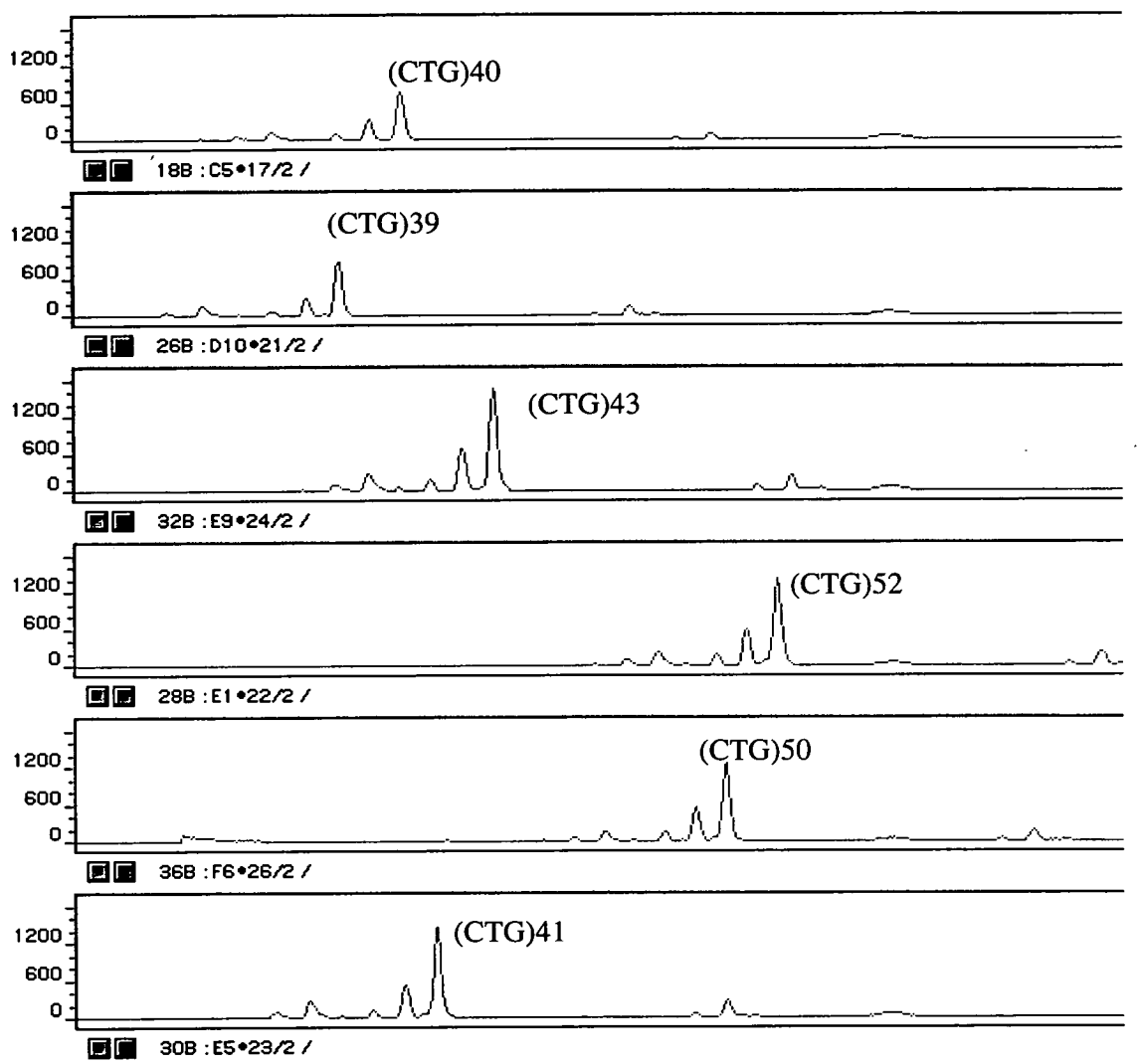


Figure 6.9. Different mutant strains integrated different lengths of the repeat array.

It is interesting to notice that the strains with CAG on the template for the leading strand integrated 40 copies for *polA polB*, 43 copies for *polA* and 50 copies for the *polB* mutant. The mutant strains with CTG on the template for the leading strand integrated different lengths of the repeat array: 39 copies for *polA polB*, 52 copies for *polA* and 41 copies for *polB* mutant. It is puzzling if these differences took place at the moment of the integration of the repeat array into the chromosome or soon after, but before the moment of the first analyses. In spite of the differences observed between different mutants, further instability was no longer detected, or in other words: each strain showed in all analysed colonies the same constant length. Sinden (Hashem *et al.*, 2002) showed that instability of the repeat arrays in plasmids occurs upon transformation. In order to construct different mutant strains with a repeat tract integrated into the chromosome, I did not transform the competent cells with plasmids containing the repeats, but circular DNA fragments which did not have the capacity to replicate. It is unlikely that the changes in repeat length occurred during the transformation, as shown by Sinden (Hashem *et al.*, 2002). It sounds more plausible that the different lengths noticed in the *polA, polB* mutants occurred at the moment of integration into the chromosome or in the very first rounds of replication. This is an interesting issue, which needs to be followed in future experiments.

Other single mutant strains (*sbcC; recF; recG; nfi1; recQ)* and double mutants (*polA recQ; polB recQ; polA polB recQ; recQ sbcC; recA sbcC; recB sbcC; mutS sbcC; nfi1 sbcC; recG sbcC; recF sbcC)* did not show any instability in the analysed colonies (Figure 6.10).

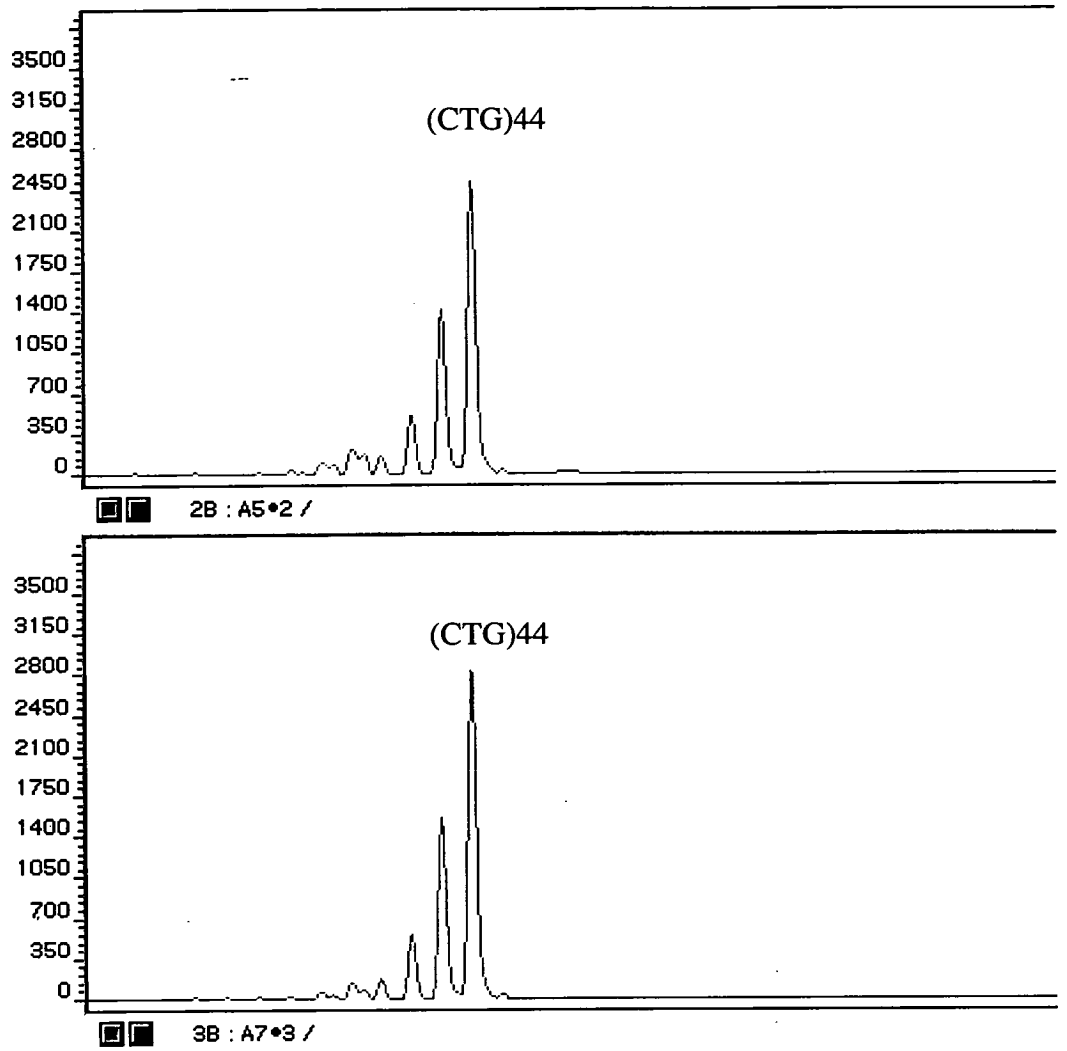


Figure 6.10. Different colonies of *recG* with CAG repeats as template for the leading strand show the same trinucleotide array length.

6.4 Concluding Remarks

The stability of the trinucleotide repeat tract is most likely influenced by the genetic environment. It is difficult to know whether the lower rate of tract instability for the chromosomal insertions relative to the plasmid ones represents a chromosome-plasmid difference or an effect of the flanking DNA sequences.

Instability of the trinucleotide repeats might be a function of the position where they are integrated, since it has already been shown that identical DNA palindromes are deleted at very different frequencies from different positions in plasmids (Leach, 1996). Two identical palindromes located only one base pair apart are deleted at frequencies that differ by a factor of more than 10^3. Compiling these observations with the finding that $(CTG)_n$ behaves like a palindrome *in vitro* (Connelly *et al.*, 1999), one can hypothesise that the particular position where the trinucleotide repeats are integrated into the chromosome could be the cause for reduced instability. It would be interesting to follow this hypothesis by analysing the repeat instability integrated at different locations in the chromosomes of *E. coli*. Instability of the TR tracts is probably greater at some chromosomal locations and lower at others. The differences in instability between repeat arrays integrated in plasmids and the ones in chromosomes might involve precise effects of long range flanking DNA sequences or it could be a result of the different numbers of rounds of replication between plasmids and chromosome. The fact that the plasmids undergo more cycles of replication than the chromosomes might result in a higher frequency of replication errors and a higher instability in plasmids. It has already been mentioned that the repeat instability in plasmids is strongly influenced by the process of transformation into competent cells. Trinucleotide repeats probably have a lower

instability in chromosomes compared to the ones in plasmids, since there is not any transformation step for the trinucleotide arrays already integrated into the chromosome.

The probability of expansion occurs only when there is a critical repeat length, at the upper range of the normal length. Expansion observed in all genetic diseases requires the presence of an allele with a threshold copy number. The significance of the length dependence is that little to no change in copy number occurs below the critical length. It might be that 43 copies integrated at the *attB* site in *E. coli* are under the critical size required for expansion. The dependence of expansion on the repeat length is most likely correlated with the genetic environment (organism, chromosomal location, flanking regions). In Fragile X, expansion is improbable near 50 copies, but almost certain at 80. For the *attB* site in *E. coli*, a length of 43 CAG or CTG seems not to be enough to undergo expansion.

Further experiments should focus on the *mutD* strain and trying to understand the mechanism of these deletions. New double mutant strains for *mutD* and other genes involved in replication, mismatch repair or recombination should be constructed and analysed to produce a clearer picture of the trinucleotide repeat instability in the *E. coli* chromosomes. Understanding factors that promote deletion may be important in developing therapeutic strategies to prevent expansion or mediate a reduction in repeat tract length.

For a number of reasons including the fact that long repeated sequences tend to be reduced in length rapidly when cloned in prokaryotes, work with *E. coli* cannot be expected to provide all the answers to the repeat expansion disorder puzzle, but the organism is easy to use and can still tell us a lot.

REFERENCES

Allen, D.J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P. & Griffith, J.D. (1997a). MutS mediates heteroduplex loop formation by a translocation mechanism. *EMBO J.* **16**, 4467-4476.

Anderson, D.G. & 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.

Arnold, D.A. & Kowalczykowski, S.C. (2000). Facilitated loading of RecA protein is essential to recombination by RecBCD enzyme. *J.Biol.Chem.* **275**, 12261-12265.

Bacolla, A., Gellibolian, R., Shimizu, M., Amirhaeri, S., Kang, S., Ohshima, K., Larson, J.E., Harvey, S.C., Stollar, B.D. & Wells, R.D. (1997). Flexible DNA: genetically unstable CTG.CAG and CGG.CCG from human hereditary neuromuscular disease genes. *J.Biol.Chem.* **272**, 16783-16792.

Ban, C. & Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell* **95**, 541-552.

Beernink, H.T. & Morrical, S.W. (1999). RMPs: recombination/replication mediator proteins. *Trends.Biochem.Sci.* **24**, 385-389.

Benkovic, S.J. & Cameron, C.E. (1995). Kinetic analysis of nucleotide incorporation and misincorporation by Klenow fragment of Escherichia coli DNA polymerase I. *Methods Enzymol.* **262:257-69.**, 257-269.

Blaauwgeers, H.G., Anwar, C.M., van den Berg, F.M., Vianney, d.J.J. & Troost, D. (1996). Expression of different metallothionein messenger ribonucleic acids in motor cortex, spinal cord and liver from patients with amyotrophic lateral sclerosis. *J.Neurol.Sci.* **142**, 39-44.

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

Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chretien, N., Tome, F.M., Lafreniere, R.G., Rommens, J.M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A.D., Heutink, P., Mathieu, J., Duranceau, A., Codere, F., Fardeau, M., Rouleau, G.A. & Korcyn, A.D. (1998). Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat.Genet.* **18**, 164-167.

Cemal, C.K., Huxley, C. & Chamberlain, S. (1999). Insertion of expanded CAG trinucleotide repeat motifs into a yeast artificial chromosome containing the human Machado-Joseph disease gene. *Gene* **236**, 53-61.

Chakrabarti, L., Bristulf, J., Foss, G.S. & Davies, K.E. (1998). Expression of the murine homologue of FMR2 in mouse brain and during development. *Hum.Mol.Genet.* **7**, 441-448.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J. & Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc.Natl.Acad.Sci.U.S.A.* **94**, 1041-1046.

Clark, A.J. (1973). Recombination deficient mutants of E. coli and other bacteria. *Annu.Rev.Genet.* **7:67-86.**, 67-86.

Clark, A.J. (1991). rec genes and homologous recombination proteins in Escherichia coli. *Biochimie* **73**, 523-532.

Coffee, B., Zhang, F., Warren, S.T. & Reines, D. (1999). Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat.Genet.* **22**, 98-101.

Connelly, J.C., de Leau, E.S. & Leach, D.R. (1999). DNA cleavage and degradation by the SbcCD protein complex from Escherichia coli. *Nucleic.Acids.Res.* **27**, 1039-1046.

Connelly, J.C., Kirkham, L.A. & 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**, 7969-7974.

Connelly, J.C. & 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**, 285-291.

Correa-Cerro, L., Wohr, G., Haussler, J., Berthon, P., Drelon, E., Mangin, P., Fournier, G., Cussenot, O., Kraus, P., Just, W., Paiss, T., Cantu, J.M. & Vogel, W. (1999). (CAG)nCAA and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. *Eur.J.Hum.Genet.* **7**, 357-362.

Craddock, N., McKeon, P., Moorhead, S., Guy, C., Harrison, D., Mynett-Johnson, L., Claffey, E., Feldman, E., McGuffin, P., Owen, M.J. & O'Donovan, M.C. (1997). Expanded CAG/CTG repeats in bipolar disorder: no correlation with phenotypic measures of illness severity. *Biol.Psychiatry* **42**, 876-881.

Crawford, D.C., Acuna, J.M. & Sherman, S.L. (2001). FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet.Med.* **3**, 359-371.

Cromie, G.A., Millar, C.B., Schmidt, K.H. & Leach, D.R. (2000). Palindromes as substrates for multiple pathways of recombination in Escherichia coli. *Genetics* **154**, 513-522.

Cunningham, R.P., Saporito, S.M., Spitzer, S.G. & Weiss, B. (1986). Endonuclease IV (nfo) mutant of Escherichia coli. *J.Bacteriol.* **168**, 1120-1127.

Dao, V. & Modrich, P. (1998). Mismatch-, MutS-, MutL-, and helicase II-dependent unwinding from the single-strand break of an incised heteroduplex. *J.Biol.Chem.* **273**, 9202-9207.

Darlow, J.M. & 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.

David, G., Durr, A., Stevanin, G., Cancel, G., Abbas, N., Benomar, A., Belal, S., Lebre, A.S., Abada-Bendib, M., Grid, D., Holmberg, M., Yahyaoui, M., Hentati, F., Chkili, T., Agid, Y. & Brice, A. (1998). Molecular and clinical correlations in autosomal dominant cerebellar ataxia with progressive macular dystrophy (SCA7). *Hum.Mol.Genet.* **7**, 165-170.

Diederich, L., Rasmussen, L.J. & Messer, W. (1992). New cloning vectors for integration in the lambda attachment site attB of the Escherichia coli chromosome. *Plasmid* **28**, 14-24.

Dixon, D.A. & 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.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc.Natl.Acad.Sci.U.S.A.* **88**, 7160-7164.

Drotschmann, K., Aronshtam, A., Fritz, H.J. & Marinus, M.G. (1998). The Escherichia coli MutL protein stimulates binding of Vsr and MutS to heteroduplex DNA. *Nucleic.Acids.Res.* **26**, 948-953.

Echols, H., Lu, C. & Burgers, P.M. (1983). Mutator strains of Escherichia coli, mutD and dnaQ, with defective exonucleolytic editing by DNA polymerase III holoenzyme. *Proc.Natl.Acad.Sci.U.S.A.* **80**, 2189-2192.

Edwards, S.M., Badzioch, M.D., Minter, R., Hamoudi, R., Collins, N., Ardern-Jones, A., Dowe, A., Osborne, S., Kelly, J., Shearer, R., Easton, D.F., Saunders, G.F., Dearnaley, D.P. & Eeles, R.A. (1999). Androgen receptor polymorphisms: association with prostate cancer risk, relapse and overall survival. *Int.J.Cancer* **84**, 458-465.

Eichler, E.E., Holden, J.J., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A. & Nelson, D.L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat.Genet.* **8**, 88-94.

Evans, E. & Alani, E. (2000). Roles for mismatch repair factors in regulating genetic recombination. *Mol.Cell Biol.* **20**, 7839-7844.

Fang, L., Davey, M.J. & O'Donnell, M. (1999). Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. *Mol.Cell* **4**, 541-553.

Feldman, D. (1997). Androgen and vitamin D receptor gene polymorphisms: the long and short of prostate cancer risk. *J.Natl.Cancer Inst.* **89**, 109-111.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M. & Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027-1038.

Freudenreich, C.H., Kantrow, S.M. & Zakian, V.A. (1998). Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853-856.

Friedberg, E.C. (1995a). Out of the shadows and into the light: the emergence of DNA repair. *Trends.Biochem.Sci.* **20**, 381

Gacy, A.M. & McMurray, C.T. (1998). Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry* **37**, 9426-9434.

Gasser, S.M. (1995). Chromosome structure. Coiling up chromosomes. *Curr.Biol.* **5**, 357-360.

Gellert, M., Mizuuchi, K., O'Dea, M.H. & Nash, H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc.Natl.Acad.Sci.U.S.A.* **73**, 3872-3876.

Gispert, S., Twells, R., Orozco, G., Brice, A., Weber, J., Heredero, L., Scheufler, K., Riley, B., Allotey, R. & Nothers, C. (1993). Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23-24.1. *Nat.Genet.* **4**, 295-299.

Glickman, B.W. & Radman, M. (1980). Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction. *Proc.Natl.Acad.Sci.U.S.A.* **77**, 1063-1067.

Glover, B.P. & McHenry, C.S. (1998). The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. *J.Biol.Chem.* **273**, 23476-23484.

Gomez, C.M., Thompson, R.M., Gammack, J.T., Perlman, S.L., Dobyns, W.B., Truwit, C.L., Zee, D.S., Clark, H.B. & Anderson, J.H. (1997). Spinocerebellar ataxia type 6: gaze-evoked and vertical nystagmus, Purkinje cell degeneration, and variable age of onset. *Ann.Neurol.* **42**, 933-950.

Goodman, F.R., Mundlos, S., Muragaki, Y., Donnai, D., Giovannucci-Uzielli, M.L., Lapi, E., Majewski, F., McGaughran, J., McKeown, C., Reardon, W., Upton, J., Winter, R.M., Olsen, B.R. & Scambler, P.J. (1997). Synpolydactyly phenotypes correlate with size of expansions in HOXD13 polyalanine tract. *Proc.Natl.Acad.Sci.U.S.A.* **94**, 7458-7463.

Goossens, D., Del-Favero, J. & Van Broeckhoven, C. (2001). Trinucleotide repeat expansions: do they contribute to bipolar disorder? *Brain Res.Bull.* **56**, 243-257.

Gorbalenya, A.E. & Koonin, E.V. (1990). Superfamily of UvrA-related NTP-binding proteins. Implications for rational classification of recombination/repair systems. *J.Mol.Biol.* **213**, 583-591.

Gordenin, D.A., Kunkel, T.A. & Resnick, M.A. (1997). Repeat expansion--all in a flap? *Nat.Genet.* **16**, 116-118.

Hackam, A.S., Singaraja, R., Zhang, T., Gan, L. & Hayden, M.R. (1999). In vitro evidence for both the nucleus and cytoplasm as subcellular sites of pathogenesis in Huntington's disease. *Hum.Mol.Genet.* **8**, 25-33.

Harfe, B.D. & Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. *Annu.Rev.Genet.* **34:359-399.**, 359-399.

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

Herman-Bert, A., Stevanin, G., Netter, J.C., Rascol, O., Brassat, D., Calvas, P., Camuzat, A., Yuan, Q., Schalling, M., Durr, A. & Brice, A. (2000). Mapping of spinocerebellar ataxia 13 to chromosome 19q13.3-q13.4 in a family with

autosomal dominant cerebellar ataxia and mental retardation. *Am.J.Hum.Genet.* **67**, 229-235.

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., Sharp, A.H., Ananth, U., Seltzer, W.K., Boss, M.A., Vieria-Saecker, A.M., Epplen, J.T., Riess, O., Ross, C.A. & Margolis, R.L. (1999). Expansion of a novel CAG trinucleotide repeat in the 5' region of PPP2R2B is associated with SCA12. *Nat.Genet.* **23**, 391-392.

Hunter, W.N., Brown, T. & Kennard, O. (1987). Structural features and hydration of a dodecamer duplex containing two C.A mispairs. *Nucleic.Acids.Res.* **15**, 6589-6606.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J.M., Weber, C., Mandel, J.L., Cancel, G., Abbas, N., Durr, A., Didierjean, O., Stevanin, G., Agid, Y. & Brice, A. (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat.Genet.* **14**, 285-291.

Irvine, R.A., Yu, M.C., Ross, R.K. & Coetzee, G.A. (1995). The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res.* **55**, 1937-1940.

Ishikawa, K., Fujigasaki, H., Saegusa, H., Ohwada, K., Fujita, T., Iwamoto, H., Komatsuzaki, Y., Toru, S., Toriyama, H., Watanabe, M., Ohkoshi, N., Shoji, S., Kanazawa, I., Tanabe, T. & Mizusawa, H. (1999). Abundant expression and cytoplasmic aggregations of [alpha]1A voltage-dependent calcium channel protein associated with neurodegeneration in spinocerebellar ataxia type 6. *Hum.Mol.Genet.* **8**, 1185-1193.

Ivancic-Bace, I., Peharec, P., Moslavac, S., Skrobot, N., Salaj-Smic, E. & Brcic-Kostic, K. (2003). RecFOR Function Is Required for DNA Repair and Recombination in a RecA Loading-Deficient recB Mutant of Escherichia coli. *Genetics* **163**, 485-494.

Jakupciak, J.P. & Wells, R.D. (2000). Genetic instabilities of triplet repeat sequences by recombination. *IUBMB.Life* **50**, 355-359.

Jaworski, A., Rosche, W.A., Gellibolian, R., Kang, S., Shimizu, M., Bowater, R.P., Sinden, R.R. & 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**, 11019-11023.

Jiricny, J. (1998). Replication errors: cha(lle)nging the genome. *EMBO J.* **17**, 6427-6436.

Jones, C., Penny, L., Mattina, T., Yu, S., Baker, E., Voullaire, L., Langdon, W.Y., Sutherland, G.R., Richards, R.I. & Tunnacliffe, A. (1995). Association of a

chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* **376**, 145-149.

Jones, C., Slijepcevic, P., Marsh, S., Baker, E., Langdon, W.Y., Richards, R.I. & Tunnacliffe, A. (1994). Physical linkage of the fragile site FRA11B and a Jacobsen syndrome chromosome deletion breakpoint in 11q23.3. *Hum.Mol.Genet.* **3**, 2123-2130.

Kang, S., Jaworski, A., Ohshima, K. & Wells, R.D. (1995). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat.Genet.* **10**, 213-218.

Kang, S., Ohshima, K., Jaworski, A. & Wells, R.D. (1996a). CTG triplet repeats from the myotonic dystrophy gene are expanded in Escherichia coli distal to the replication origin as a single large event. *J.Mol.Biol.* **258**, 543-547.

Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S. & Wells, R.D. (1995). Pausing of DNA synthesis in vitro at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J.Biol.Chem.* **270**, 27014-27021.

Kantake, N., Madiraju, M.V., Sugiyama, T. & Kowalczykowski, S.C. (2002). Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. *Proc.Natl.Acad.Sci.U.S.A.* **99**, 15327-15332.

Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, S., Kawakami, H., Nakamura, S., Nishimura, M. & Akiguchi, I. (1994). CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat.Genet.* **8**, 221-228.

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

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. (1993). Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* **74**, 127-134.

Kogoma, T. (1996). Recombination by replication. *Cell* **85**, 625-627.

Koi, M., Umar, A., Chauhan, D.P., Cherian, S.P., Carethers, J.M., Kunkel, T.A. & Boland, C.R. (1994). Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Res.* **54**, 4308-4312.

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., Takahashi, H., Kondo, R., Ishikawa, A. & Hayashi, T. (1994). Unstable expansion of

CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat.Genet.* **6**, 9-13.

Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W. & Ranum, L.P. (1999a). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat.Genet.* **21**, 379-384.

Koutnikova, H., Campuzano, V., Foury, F., Dolle, P., Cazzalini, O. & Koenig, M. (1997). Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Nat.Genet.* **16**, 345-351.

Kovtun, I.V. & McMurray, C.T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nat.Genet.* **27**, 407-411.

Kowalczykowski, S.C. (2000). Initiation of genetic recombination and recombination-dependent replication. *Trends.Biochem.Sci.* **25**, 156-165.

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. (1994). Biochemistry of homologous recombination in Escherichia coli. *Microbiol.Rev.* **58**, 401-465.

Kunkel, T.A. (1990). Misalignment-mediated DNA synthesis errors. *Biochemistry* **29**, 8003-8011.

Kunkel, T.A. (1992). DNA replication fidelity. *J.Biol.Chem.* **267**, 18251-18254.

Kunkel, T.A. (1993). Nucleotide repeats. Slippery DNA and diseases. *Nature* **365**, 207-208.

Kuzminov, A. (1996). Unraveling the late stages of recombinational repair: metabolism of DNA junctions in Escherichia coli. *Bioessays* **18**, 757-765.

Lafreniere, R.G., Rochefort, D.L., Chretien, N., Rommens, J.M., Cochius, J.I., Kalviainen, R., Nousiainen, U., Patry, G., Farrell, K., Soderfeldt, B., Federico, A., Hale, B.R., Cossio, O.H., Sorensen, T., Pouliot, M.A., Kmiec, T., Uldall, P., Janszky, J., Pranzatelli, M.R., Andermann, F., Andermann, E. & Rouleau, G.A. (1997). Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat.Genet.* **15**, 298-302.

Leach, D.R. (1996). Cloning and characterization of DNAs with palindromic sequences. *Genet.Eng.(N.Y.)* **18:1-11.**, 1-11.

Leach, D.R., Okely, E.A. & Pinder, D.J. (1997a). Repair by recombination of DNA containing a palindromic sequence. *Mol.Microbiol.* **26**, 597-606.

Lee, E.H. & Kornberg, A. (1992). Features of replication fork blockage by the Escherichia coli terminus-binding protein. *J.Biol.Chem.* **267**, 8778-8784.

Leeflang, E.P., Tavare, S., Marjoram, P., Neal, C.O., Srinidhi, J., MacFarlane, H., MacDonald, M.E., Gusella, J.F., de Young, M., Wexler, N.S. & Arnheim, N.

(1999). Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum.Mol.Genet.* **8**, 173-183.

Levinson, G. & Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol.Biol.Evol.* **4**, 203-221.

Lindblad, K., Nylander, P.O., De bruyn, A., Sourey, D., Zander, C., Engstrom, C., Holmgren, G., Hudson, T., Chotai, J. & Mendlewicz, J. (1995). Detection of expanded CAG repeats in bipolar affective disorder using the repeat expansion detection (RED) method. *Neurobiol.Dis.* **2**, 55-62.

Liu, B., Nicolaides, N.C., Markowitz, S., Willson, J.K., Parsons, R.E., Jen, J., Papadopolous, N., Peltomaki, P., de la Chapelle, A. & Hamilton, S.R. (1995b). Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat.Genet.* **9**, 48-55.

Liu, B., Parsons, R.E., Hamilton, S.R., Petersen, G.M., Lynch, H.T., Watson, P., Markowitz, S., Willson, J.K., Green, J. & de la Chapelle, A. (1994a). hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. *Cancer Res.* **54**, 4590-4594.

Lloyd, R.G. & Buckman, C. (1985). Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J.Bacteriol.* **164**, 836-844.

Lloyd, R.G., Buckman, C. & Benson, F.E. (1987). Genetic analysis of conjugational recombination in Escherichia coli K12 strains deficient in RecBCD enzyme. *J.Gen.Microbiol.* **133**, 2531-2538.

Loeb, L.A. & Kunkel, T.A. (1982). Fidelity of DNA synthesis. *Annu.Rev.Biochem.* **51:429-57.**, 429-457.

Lu, A.L., Welsh, K., Clark, S., Su, S.S. & Modrich, P. (1984). Repair of DNA base-pair mismatches in extracts of Escherichia coli. *Cold Spring Harb.Symp.Quant.Biol.* **49:589-96.**, 589-596.

Lyons-Darden, T. & Topal, M.D. (1999). Abasic sites induce triplet-repeat expansion during DNA replication in vitro. *J.Biol.Chem.* **274**, 25975-25978.

Lyons-Darden, T. & Topal, M.D. (1999). Effects of temperature, Mg2+ concentration and mismatches on triplet-repeat expansion during DNA replication in vitro. *Nucleic.Acids.Res.* **27**, 2235-2240.

Lyons, S.M. & Schendel, P.F. (1984). Kinetics of methylation in Escherichia coli K-12. *J.Bacteriol.* **159**, 421-423.

Mahdi, A.A., Sharples, G.J., Mandal, T.N. & Lloyd, R.G. (1996). Holliday junction resolvases encoded by homologous rusA genes in Escherichia coli K-12 and phage 82. *J.Mol.Biol.* **257**, 561-573.

Malkov, V.A., Biswas, I., Camerini-Otero, R.D. & Hsieh, P. (1997). Photocross-linking of the NH2-terminal region of Taq MutS protein to the major groove of a heteroduplex DNA. *J.Biol.Chem.* **272**, 23811-23817.

Mariappan, S.V., Silks, L.A., Chen, X., Springer, P.A., Wu, R., Moyzis, R.K., Bradbury, E.M., Garcia, A.E. & Gupta, G. (1998). Solution structures of the Huntington's disease DNA triplets, (CAG)n. *J.Biomol.Struct.Dyn.* **15**, 723-744.

McGlynn, P. & Lloyd, R.G. (2001). Action of RuvAB at replication fork structures. *J.Biol.Chem.* **276**, 41938-41944.

McMurray, C.T. (1995). Mechanisms of DNA expansion. *Chromosoma* **104**, 2-13.

Mendlewicz, J., Lindbald, K., Souery, D., Mahieu, B., Nylander, P.O., De bruyn, A., Zander, C., Engstrom, C., Adolfsson, R., Van Broeckhoven, C., Schalling, M. & Lipp, O. (1997). Expanded trinucleotide CAG repeats in families with bipolar affective disorder. *Biol.Psychiatry* **42**, 1115-1122.

Miller, W.J., Skinner, J.A., Foss, G.S. & Davies, K.E. (2000). Localization of the fragile X mental retardation 2 (FMR2) protein in mammalian brain. *Eur.J.Neurosci.* **12**, 381-384.

Miret, J.J., Pessoa-Brandao, L. & 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**, 12438-12443.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu.Rev.Genet.* **25:229-53.**, 229-253.

Modrich, P. & Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu.Rev.Biochem.* **65:101-33.**, 101-133.

Muller, B. & West, S.C. (1994). Processing of Holliday junctions by the Escherichia coli RuvA, RuvB, RuvC and RecG proteins. *Experientia* **50**, 216-222.

Mundlos, S., Otto, F., Mundlos, C., Mulliken, J.B., Aylsworth, A.S., Albright, S., Lindhout, D., Cole, W.G., Henn, W., Knoll, J.H., Owen, M.J., Mertelsmann, R., Zabel, B.U. & Olsen, B.R. (1997). Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* **89**, 773-779.

Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I. & Murayama, N. (1994). Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat.Genet.* **6**, 14-18.

Nakamura, K., Jeong, S.Y., Uchihara, T., Anno, M., Nagashima, K., Nagashima, T., Ikeda, S., Tsuji, S. & Kanazawa, I. (2001). SCA17, a novel autosomal

dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum.Mol.Genet.* **10**, 1441-1448.

Naom, I.S., Morton, S.J., Leach, D.R. & Lloyd, R.G. (1989). Molecular organization of sbcC, a gene that affects genetic recombination and the viability of DNA palindromes in Escherichia coli K-12. *Nucleic.Acids.Res.* **17**, 8033-8045.

Nelson, K.A. & Witte, J.S. (2002). Androgen receptor CAG repeats and prostate cancer. *Am.J.Epidemiol.* **155**, 883-890.

Nolin, S.L., Lewis, F.A., Ye, L.L., Houck, G.E.J., Glicksman, A.E., Limprasert, P., Li, S.Y., Zhong, N., Ashley, A.E., Feingold, E., Sherman, S.L. & Brown, W.T. (1996). Familial transmission of the FMR1 CGG repeat. *Am.J.Hum.Genet.* **59**, 1252-1261.

Nossal, N.G. (1992). Protein-protein interactions at a DNA replication fork: bacteriophage T4 as a model. *FASEB J.* **6**, 871-878.

O'Donovan, M.C., Guy, C., Craddock, N., Bowen, T., McKeon, P., Macedo, A., Maier, W., Wildenauer, D., Aschauer, H.N., Sorbi, S., Feldman, E., Mynett-Johnson, L., Claffey, E., Nacmias, B., Valente, J., Dourado, A., Grassi, E., Lenzinger, E., Heiden, A.M., Moorhead, S., Harrison, D., Williams, J., McGuffin, P. & Owen, M.J. (1996). Confirmation of association between expanded CAG/CTG repeats and both schizophrenia and bipolar disorder. *Psychol.Med.* **26**, 1145-1153.

O'Donovan, M.C., Guy, C., Craddock, N., Murphy, K.C., Cardno, A.G., Jones, L.A., Owen, M.J. & McGuffin, P. (1995). Expanded CAG repeats in schizophrenia and bipolar disorder. *Nat.Genet.* **10**, 380-381.

Orr, H.T. & Zoghbi, H.Y. (2000). Reversing neurodegeneration: a promise unfolds. *Cell* **101**, 1-4.

Orr, H.T. & Zoghbi, H.Y. (2001). SCA1 molecular genetics: a history of a 13 year collaboration against glutamines. *Hum.Mol.Genet.* **10**, 2307-2311.

Parker, B.O. & Marinus, M.G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in Escherichia coli. *Proc.Natl.Acad.Sci.U.S.A.* **89**, 1730-1734.

Parsons, R., Li, G.M., Longley, M.J., Fang, W.H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K.W., Vogelstein, B. & Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227-1236.

Peltomaki, P., Lothe, R.A., Aaltonen, L.A., Pylkkanen, L., Nystrom-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D. & Haugen, A. (1993). Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res.* **53**, 5853-5855.

Petruska, J., Arnheim, N. & 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**, 1992-1998.

Petruska, J., Hartenstine, M.J. & Goodman, M.F. (1998). Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease. *J.Biol.Chem.* **273**, 5204-5210.

Ranum, L.P. & Day, J.W. (2002). Myotonic dystrophy: clinical and molecular parallels between myotonic dystrophy type 1 and type 2. *Curr.Neurol.Neurosci.Rep.* **2**, 465-470.

Restituito, S., Thompson, R.M., Eliet, J., Raike, R.S., Riedl, M., Charnet, P. & Gomez, C.M. (2000). The polyglutamine expansion in spinocerebellar ataxia type 6 causes a beta subunit-specific enhanced activation of P/Q-type calcium channels in Xenopus oocytes. *J.Neurosci.* **20**, 6394-6403.

Richards, R.I. & Sutherland, G.R. (1992). Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**, 709-712.

Ross, R.K., Pike, M.C., Coetzee, G.A., Reichardt, J.K., Yu, M.C., Feigelson, H., Stanczyk, F.Z., Kolonel, L.N. & Henderson, B.E. (1998). Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. *Cancer Res.* **58**, 4497-4504.

Rubinsztein, D.C., Leggo, J., Coles, R., Almqvist, E., Biancalana, V., Cassiman, J.J., Chotai, K., Connarty, M., Crauford, D., Curtis, A., Curtis, D., Davidson, M.J., Differ, A.M., Dode, C., Dodge, A., Frontali, M., Ranen, N.G., Stine, O.C., Sherr, M., Abbott, M.H., Franz, M.L., Graham, C.A., Harper, P.S., Hedreen, J.C. & Hayden, M.R. (1996). Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats. *Am.J.Hum.Genet.* **59**, 16-22.

Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M.E. (1998). Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55-66.

Schaaper, R.M. & Radman, M. (1989). The extreme mutator effect of Escherichia coli mutD5 results from saturation of mismatch repair by excessive DNA replication errors. *EMBO J.* **8**, 3511-3516.

Schalling, M., Hudson, T.J., Buetow, K.H. & Housman, D.E. (1993). Direct detection of novel expanded trinucleotide repeats in the human genome. *Nat.Genet.* **4**, 135-139.

Schmidt, K.H., Abbott, C.M. & Leach, D.R. (2000a). Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli. *Mol.Microbiol.* **35**, 463-471.

Schoenberg, M.P., Hakimi, J.M., Wang, S., Bova, G.S., Epstein, J.I., Fischbeck, K.H., Isaacs, W.B., Walsh, P.C. & Barrack, E.R. (1994). Microsatellite mutation (CAG24-->18) in the androgen receptor gene in human prostate cancer. *Biochem.Biophys.Res.Commun.* **198**, 74-80.

Schumacher, S., Fuchs, R.P. & 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**, 1101-1110.

Sharples, G.J., Benson, F.E., Illing, G.T. & Lloyd, R.G. (1990). Molecular and functional analysis of the ruv region of Escherichia coli K-12 reveals three genes involved in DNA repair and recombination. *Mol.Gen.Genet.* **221**, 219-226.

Shimizu, M., Gellibolian, R., Oostra, B.A. & Wells, R.D. (1996). Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J.Mol.Biol.* **258**, 614-626.

Sinden, R.R. (1999). Biological implications of the DNA structures associated with disease-causing triplet repeats. *Am.J.Hum.Genet.* **64**, 346-353.

Smith, G.R. (1981). DNA supercoiling: another level for regulating gene expression. *Cell* **24**, 599-600.

Smith, G.R. (1983). Chi hotspots of generalized recombination. *Cell* **34**, 709-710.

Smith, G.R. (1988). Homologous recombination in procaryotes. *Microbiol.Rev.* **52**, 1-28.

Stahl, F.W., Stahl, M.M., Malone, R.E. & Crasemann, J.M. (1980). Directionality and nonreciprocality of Chi-stimulated recombination in phage lambda. *Genetics* **94**, 235-248.

Stanford, J.L., Just, J.J., Gibbs, M., Wicklund, K.G., Neal, C.L., Blumenstein, B.A. & Ostrander, E.A. (1997). Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res.* **57**, 1194-1198.

Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274-276.

Su, S.S. & Modrich, P. (1986a). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. *Proc.Natl.Acad.Sci.U.S.A.* **83**, 5057-5061.

Sutherland, G.R. & Richards, R.I. (1995). The molecular basis of fragile sites in human chromosomes. *Curr.Opin.Genet.Dev.* **5**, 323-327.

Sutton, M.D. & Walker, G.C. (2001). Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc.Natl.Acad.Sci.U.S.A.* **98**, 8342-8349.

Toth, G., Gaspari, Z. & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967-981.

Tougu, K. & Marians, K.J. (1996). The interaction between helicase and primase sets the replication fork clock. *J.Biol.Chem.* **271**, 21398-21405.

Trinh, T.Q. & Sinden, R.R. (1993). The influence of primary and secondary DNA structure in deletion and duplication between direct repeats in Escherichia coli. *Genetics* **134**, 409-422.

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. (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**, 905-914.

Viswanathan, M. & Lovett, S.T. (1998). Single-strand DNA-specific exonucleases in Escherichia coli. Roles in repair and mutation avoidance. *Genetics* **149**, 7-16.

Vorlickova, M., Kejnovska, I., Tumova, M. & Kypr, J. (2001). Conformational properties of DNA fragments containing GAC trinucleotide repeats associated with skeletal displasias. *Eur.Biophys.J.* **30**, 179-185.

Warner, J.P., Barron, L.H., Goudie, D., Kelly, K., Dow, D., Fitzpatrick, D.R. & Brock, D.J. (1996). A general method for the detection of large CAG repeat expansions by fluorescent PCR. *J.Med.Genet.* **33**, 1022-1026.

Warren, S.T. (1997). Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* **275**, 408-409.

Weber, J.L. & Wong, C. (1993). Mutation of human short tandem repeats. *Hum.Mol.Genet.* **2**, 1123-1128.

Weisberg, R.A., Enquist, L.W., Foeller, C. & Landy, A. (1983). Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J.Mol.Biol.* **170**, 319-342.

Wells, R.D., Jacob, T.M., Narang, S.A. & Khorana, H.G. (1967). Studies on polynucleotides. LXIX. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of Escherichia coli: DNA-like polymers containing repeating trinucleotide sequences. *J.Mol.Biol.* **27**, 237-263.

Wells, R.D., Ohtsuka, E. & Khorana, H.G. (1965). Studies on polynucleotides. L. Synthetic deoxyribopolynucleotides as templates for the DNA polymerase of Escherichia coli: a new double-stranded DNA-like polymer containing repeating dinucleotide sequences. *J.Mol.Biol.* **14**, 221-237.

Wertman, K.F., Wyman, A.R. & Botstein, D. (1986). Host/vector interactions which affect the viability of recombinant phage lambda clones. *Gene* **49**, 253-262.

Wilding, G. (1992). The importance of steroid hormones in prostate cancer. *Cancer Surv.* **14:113-30.**, 113-130.

Wilson, R.B. & Roof, D.M. (1997). Respiratory deficiency due to loss of mitochondrial DNA in yeast lacking the frataxin homologue. *Nat.Genet.* **16**, 352-357.

Wong, A., Yang, J., Cavadini, P., Gellera, C., Lonnerdal, B., Taroni, F. & Cortopassi, G. (1999). The Friedreich's ataxia mutation confers cellular sensitivity to oxidant stress which is rescued by chelators of iron and calcium and inhibitors of apoptosis. *Hum.Mol.Genet.* **8**, 425-430.

Wooster, R., Cleton-Jansen, A.M., Collins, N., Mangion, J., Cornelis, R.S., Cooper, C.S., Gusterson, B.A., Ponder, B.A., von Deimling, A. & Wiestler, O.D. (1994). Instability of short tandem repeats (microsatellites) in human cancers. *Nat.Genet.* **6**, 152-156.

Worth, P.F., Giunti, P., Gardner-Thorpe, C., Dixon, P.H., Davis, M.B. & Wood, N.W. (1999). Autosomal dominant cerebellar ataxia type III: linkage in a large British family to a 7.6-cM region on chromosome 15q14-21.3. *Am.J.Hum.Genet.* **65**, 420-426.

Yamaguchi, M., Dao, V. & Modrich, P. (1998). MutS and MutL activate DNA helicase II in a mismatch-dependent manner. *J.Biol.Chem.* **273**, 9197-9201.

Yamamoto, A., Lucas, J.J. & Hen, R. (2000). Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. *Cell* **101**, 57-66.

Yamamoto, Y., Kawai, H., Nakahara, K., Osame, M., Nakatsuji, Y., Kishimoto, T. & Sakoda, S. (1992). A novel primer extension method to detect the number of CAG repeats in the androgen receptor gene in families with X-linked spinal and bulbar muscular atrophy. *Biochem.Biophys.Res.Commun.* **182**, 507-513.

Yamashita, I., Sasaki, H., Yabe, I., Fukazawa, T., Nogoshi, S., Komeichi, K., Takada, A., Shiraishi, K., Takiyama, Y., Nishizawa, M., Kaneko, J., Tanaka, H., Tsuji, S. & Tashiro, K. (2000). A novel locus for dominant cerebellar ataxia (SCA14) maps to a 10.2-cM interval flanked by D19S206 and D19S605 on chromosome 19q13.4-qter. *Ann.Neurol.* **48**, 156-163.

Yu, A., Dill, J. & Mitas, M. (1995b). The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. *Nucleic.Acids.Res.* **23**, 4055-4057.

Yu, A., Dill, J., Wirth, S.S., Huang, G., Lee, V.H., Haworth, I.S. & Mitas, M. (1995a). The trinucleotide repeat sequence d(GTC)15 adopts a hairpin conformation. *Nucleic.Acids.Res.* **23**, 2706-2714.

Yuzhakov, A., Kelman, Z. & O'Donnell, M. (1999). Trading places on DNA--a three-point switch underlies primer handoff from primase to the replicative DNA polymerase. *Cell* **96**, 153-163.

Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y. & Lee, C.C. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat.Genet.* **15**, 62-69.

Zu, L., Figueroa, K.P., Grewal, R. & Pulst, S.M. (1999). Mapping of a new autosomal dominant spinocerebellar ataxia to chromosome 22. *Am.J.Hum.Genet.* **64**,.594-599.