# UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Centre for Biological Sciences

Volume 1 of 1

# Properties of "Sticky DNA" formed by GAA repeats

by

Mohammed Sulaiman Alsulami

Thesis for the degree of Doctor of Philosophy

August 2016

UNIVERSITY OF SOUTHAMPTON

# ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

CENTRE FOR BIOLOGICAL SCIENCES

Thesis for the degree of Doctor of Philosophy

**PROPERTIES OF "STICKY DNA" FORMED BY GAA REPEATS**

Mohammed Sulaiman Alsulami

Friedreich's ataxia (FRDA) is an autosomal recessive neurodegenerative disorder, caused by long repeats of the trinucleotide GAA.TTC in the first intron of the frataxin (FXN) gene. This expanded repeat results in lower than normal levels of mature FXN mRNA and thus reduced levels of frataxin protein, the FXN gene product. Normal individuals have 5 to 40 GAA repeat sequences, whereas affected individuals have from 70 to more than 1000 GAA triplets. The frataxin protein plays a role in assembly of iron-sulphur clusters which are needed for energy production. As GAA.TTC repeats contain only purine bases (Pu) on one strand and pyrimidine bases (Py) on the complementary DNA strand, they can form intramolecular triplexes under superhelical stress. "Sticky DNA" is a poorly characterised DNA structure that is generated by the association of two long stretches of the trinucleotide GAA.TTC sequence (>59 repeats), which are found in FRDA. The formation of this structure is dependent on the length and orientation of the repeats, negative supercoiling and the presence of divalent metal ions. Although sticky DNA was discovered over a decade ago, its exact structure remains unknown. Therefore, the principle aim of this thesis was to examine and characterize the formation of non-B-DNA structures in the FRDA-associated GAA.TTC repeats that were cloned in suitable vectors.

Various lengths of GAA.TTC repeats (up to 85 repeats) were cloned into two sites in both pUC18 and pGL3 plasmids, either singly or in combination. The clones in pUC18 were used to assess the formation of unusual DNA structures (H-DNA or sticky DNA), through the use of the chemical probes diethylpyrocarbonate (DEPC) and potassium permanganate ($KMnO_4$), and the single-strand specific endonuclease S1. These react with the regions of single stranded DNA that are generated within the non-B DNA structures. The use of DEPC and $KMnO_4$ together provided us a complete analysis of both the GAA- and TTC-containing DNA strands. DEPC and $KMnO_4$ revealed strong evidence for H-DNA formation in short $(GAA.TTC)_n$ sequences ($n \leq 29$) and some evidence in longer inserts. However, there was no evidence for 'sticky DNA' formation within our plasmids. High resolution probing with S1 nuclease confirmed the DEPC and $KMnO_4$ data for detecting H-DNA in short inserts, but failed to detect any unique structural perturbations within the longer repeats. However, when S1 nuclease was used to map the single strand regions that accompany H-DNA formation, S1 sensitive sites were apparent for both short

and long GAA.TTC repeats, indicating the presence of DNA secondary structure within this region, which was only present in the supercoiled species. The effects of GAA.TTC repeats on the structure of the plasmids were also examined by observing their electrophoretic mobility in agarose gels. The results showed that plasmids containing long GAA-repeats exhibited some unusual properties and were prone to form multimeric structures (dimers, trimers etc.). By using limited restriction digestions, we confirmed that the low mobility plasmid species is a dimer that consists of two monomer plasmids covalently ligated together. However, we failed to detect any 'sticky DNA' conformation in any of the cloned pUC18 plasmids, even those containing two tracts of long repeats. Finally, GAA-inserts in pGL3 were used to assess the effect of these sequences on transcription of the adjacent luciferase gene. These results reveal that a decrease in gene expression level from the vectors that carry short or long repeats at one or two positions in the luciferase reporter vector.

## DECLARATION OF AUTHORSHIP

I, Mohammed Sulaiman Alsulami declare that the thesis entitled **Properties of "sticky DNA" formed by GAA repeats** and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- None of this work has been published before submission.

Signed:

.....

Date:

.....

# ACKNOWLEDGENTS

I would firstly like to thank Keith Fox for his supervision throughout my research and Dr. David Rusling and Dr. Scott Kimber for their help, advice and tuition in the lab. Thanks also to Dr. Mark Coldwell and Dr. Helen Lewis for helping and guiding me in learning tissue culture and for their guidance. Thanks to the many people who have made this PhD an enjoyable experience, and finally I would like to thank my family for their love and support over these many years.

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| A | Adenine |
| bp | Base pair |
| C | Cytosine |
| dATP | deoxyadenosine triphosphate |
| DEPC | Diethylpyrocarbonate |
| DMS | Dimethyl sulphate |
| EDTA | Ethylenediaminetetraacetic acid |
| G | Guanine |
| Kb | Kilobase(s) |
| L | Linear (in respect to DNA) |
| LB | Luria-Bertani media |
| MCS | Multiple cloning site |
| NMR | Nuclear Magnetic Resonance |
| PAGE | Polyacrylamide gel electrophoresis |
| PCR | Polymerase chain reaction |
| SC | Supercoiled |
| SURE | Stop Unwanted Rearrangement Events |
| T | Thymine |
| TNR | Trinucleotide repeats |
| U | Uracil |
| UV | Ultra violet |

# 1. Introduction

## 1.1 Overview

Expansions of trinucleotide GAA.TTC tracts are associated with the human disease Friedreich's ataxia. The expansion is located within the first intron of the frataxin gene on chromosome 9 in both alleles and can reach lengths of more than 1000 repeats (1, 2). It is known that long GAA.TTC repeats are prone to form non-B-DNA structures such as triplexes and sticky DNA (3, 4). It is likely that these unusual DNA structures inhibit the transcriptional activity of the human frataxin gene producing low levels of mRNA and frataxin protein but do not alter the protein itself (5). Sticky DNA is a novel non-B-DNA conformation that is formed in supercoiled plasmids by the association of two long GAA.TTC repeats at lengths that are found in the frataxin gene in FRDA patients (6). The exact structure of sticky DNA still unknown.

## 1.2 Basic structures of nucleic acids

The discovery of DNA structure by James Watson and Francis Crick in 1953, has had a profound effect on molecular biology (7). Both DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are made up of nucleotides which consist of three units: a nitrogen-containing heterocyclic base, a pentose sugar and phosphate group. The major bases in nucleic acids are derivatives of monocyclic pyrimidines or bicyclic purines (Figure 1.1). Purine bases are adenine and guanine, and pyrimidine bases are cytosine and thymine in DNA and uracil in RNA. Nucleosides are subunits consisting of a pentose sugar and a nitrogenous base which are linked by a glycosidic bond. This bond is formed between N1 of pyrimidine bases or N9 of purine bases and C1' of the D-ribose sugar (8).

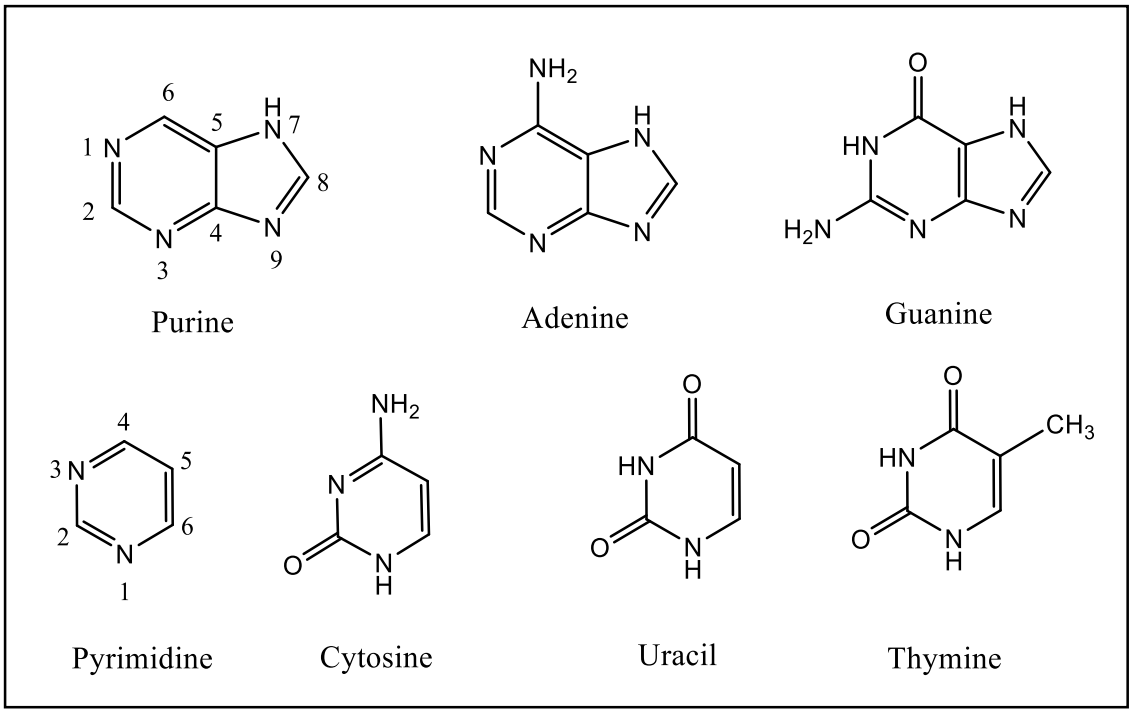



**Figure 1.1:** Structures of the five major purine and pyrimidine bases of nucleic acids in their dominant forms.

The most usual structure of DNA is a double-stranded right-handed helix (B-form) with a negatively charged backbone (deoxyribose phosphate) on the outside and stacked base pairs inside the double helix (Figure1.2). DNA consist of two long polymers of nucleotides joined by 3'-5'-phosphodiester linkages. The chains are arranged in an antiparallel orientation and follow a right handed helix. The two strands are held together by specific pairing of the bases through the formation of hydrogen bonds. Adenine pairs with thymine by two hydrogen bonds and guanine pairs with cytosine by three hydrogen bonds. Hoogsteen bonds, utilising acceptor and donor groups at positions 6 and 7 of purine rings can also be formed between a DNA third strand and a duplex in DNA triplexes. Hydrogen bonds and base-stacking interactions are extremely important forces for stabilising the DNA double helix (9). However, DNA can adopt a variety of unconventional non-B-form structures (Figure 1.3) under a variety of conditions, as well as several non-helical secondary structures as presented in Figure 1.4, including triplexes (H-DNA and sticky DNA), slipped (hairpin) structures and G-quadruplexes (10).

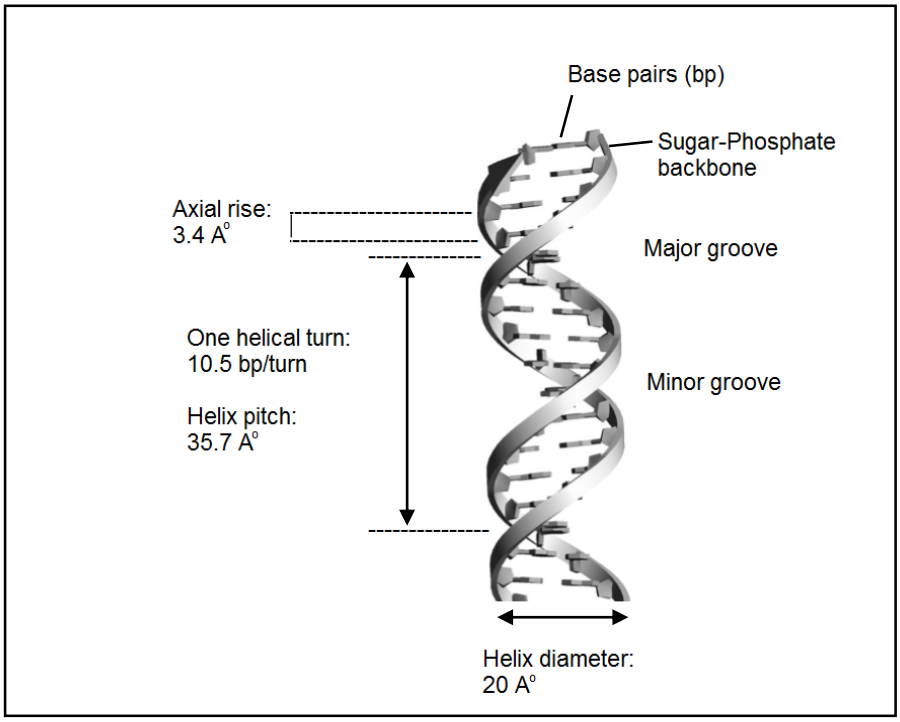



**Figure 1.2:** Structure and dimension of the DNA double helix (11, 12).

There are three main conformations of the DNA double helix, namely A-, B- and Z-DNA, summarized in Table 1.1. The B-form of DNA is the dominant right-handed helix in the biological state *in vivo* or in high humidity (92%). A-DNA also has a right handed double helix but appears as the relative humidity falls to about 75%, thus structural transitions between the A - and B - forms of DNA are possible. However, double-stranded DNA may also adopt a left-handed helical conformation (Z-DNA) under special conditions such as high salt and in sequences containing alternating purine-pyrimidine tracts of 12 to 15 base pairs (13). Compared to the B-DNA form, Z-DNA contains inverted purines in the *syn*-conformation while pyrimidines remain in the *anti*-conformation with the sugar-pucker altered from the C'2- to the C'3-endo position so as to maintain the Watson-Crick base-pairing (14). The left-handed Z-DNA and right-handed B-DNA are not mirror images but completely different structures. Z-DNA has only one groove, whereas B-DNA has a major and a minor groove (15). Recently, it has been shown that Z-DNA-forming sequences can induce genetic instability in a number of organisms but the underlying mechanisms remain unclear (16).

**Table 1.1:** Comparison of A-, B-, and Z-DNA structures (11, 17).

| **Helix form** | **A-DNA** | **B-DNA** | **Z-DNA** |
|---|---|---|---|
| Shape | Widest | Intermediate | Narrowest |
| Rise per base pair | 2.3 A° | 3.4 A° | 3.8 A° |
| Helix diameter | 25.5 A° | 23.7 A° | 18.4 A° |
| Screw sense | Right-handed sense | Right-handed sense | Left-handed sense |
| Glycosidic bond | Anti- | Anti- | Anti- and Syn- |
| Base pairs per helical turn | 11 | 10.5 | 12 |
| Pitch per helical turn | 25.3 A° | 35.7 A° | 45.6 A° |
| Major groove | Narrow and very deep | Wide and quite deep | Flat |
| Minor groove | Very wide and shallow | Narrow and quite deep | Very narrow and deep |

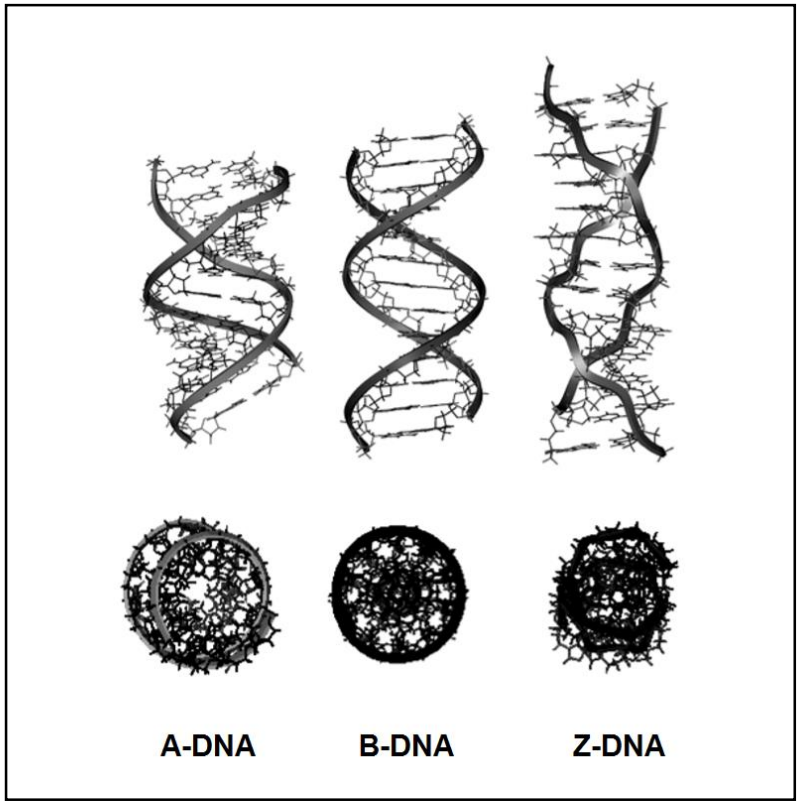



**Figure 1.3:** The major nucleic acid duplex conformations in DNA (A-form, B-form, and Z-form) and RNA (A-form). The sugar-phosphate backbone is represented by the ribbon. Bottom views: orthogonal representations (11).

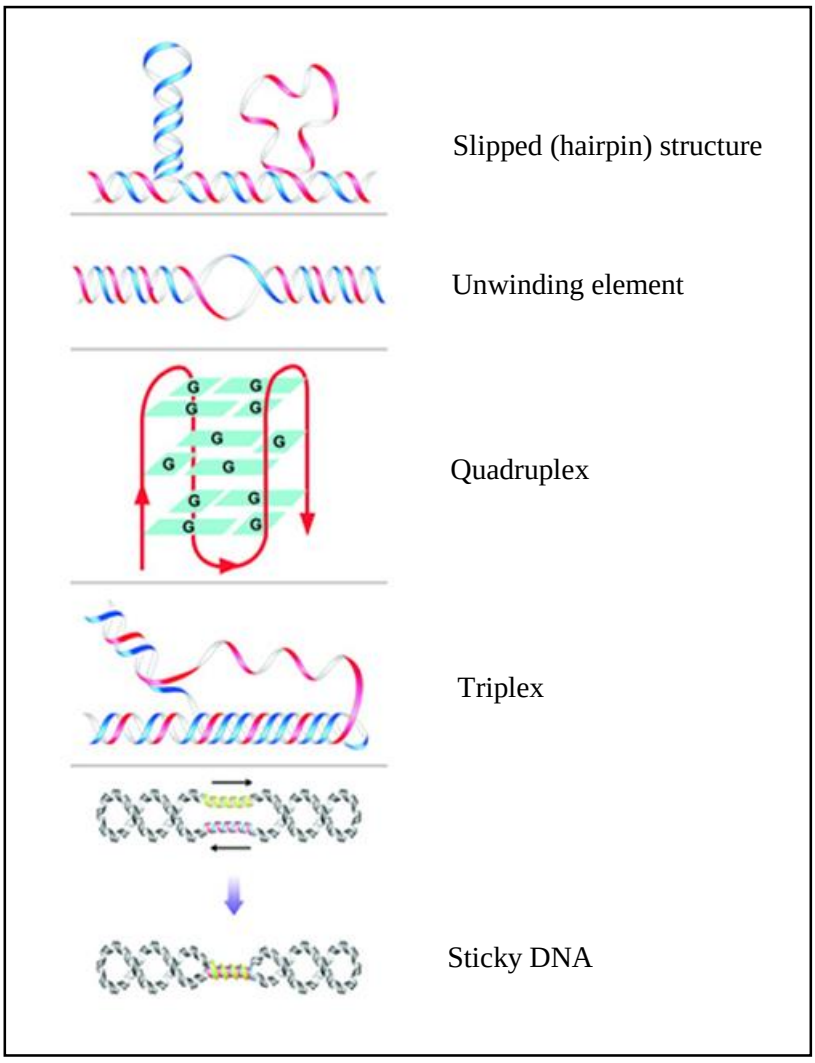



**Figure 1.4:** Non-helical DNA conformations formed by repeating sequences (18).

These non-B-DNA structures are often formed in stretches of repeat sequences and may affect a wide range of genetic transactions such as DNA replication, transcription, recombination and repair which can then lead to genetic instability and human disease (19). One such unusual DNA structure is associated with long repeats of the trinucleotide sequence GAA, which causes the genetic disease Friedreich's ataxia, and which is the subject of this thesis. The (GAA.TTC) repeating sequence is an example of a poly (purine.pyrimidine)-rich sequence that has only purines (Pu) in one strand and pyrimidines (Py) on the other strand. *In vitro* and *in vivo* studies have shown that long GAA.TTC repeats can form an intramolecular triplex conformation. Recently, GAA-containing repeat sequences have been shown to adopt higher-order conformations, called sticky DNA. This is generated by two long stretches of the GAA trinucleotide repeats and has been shown to inhibit transcription *in vitro* and in mammalian cells (3, 20). Various

studies, described in detail below, have shown that the formation of 'sticky DNA' in plasmids requires a pair of GAA-repeat sequences (each of which must contain at least 59 repeats), negative supercoiling and the presence of divalent metal ions. The first evidence for the existence of sticky DNA came from the presence of a retarded band in agarose gel electrophoresis of restriction fragments derived from plasmids containing pairs of long GAA-repeats. Its structure has been poorly defined, but it is often depicted as a dumbbell-shaped structure (Figure 1.4, sticky DNA) (6), arising from association of two intramolecular triplexes (3). However, there is no clear evidence of this structure being present inside living cells.

## 1.3 Triplex DNA

The first formation of triple helices in nucleic acids was observed in 1957 by Felsenfeld *et al*. (21) using X-ray fibre diffraction and UV melting on mixing polyribonucleotides poly(U) and poly(A) in the ratio 2:1 in the presence of magnesium chloride. Further experiments showed that poly(C) and poly(G) can create a similar structure at low pH (22). Duplexes capable of forming triple helices contain long stretches of polypyrimidine.polypurine sequences and the third strand (Pu or Py) only associates with the purine strand of the Watson-Crick duplex (22, 23). However, it has been shown that various structures of DNA and RNA triple helices can exist, including inter- and intra-molecular triplexes in parallel and antiparallel conformations (9, 23-26). In these structures the third strand lies in the major groove of the target DNA duplex (Figure 1.5), making alternative hydrogen bonding with substituents on the exposed faces of the duplex DNA base pairs. The base pairs formed between the third strand and the duplex DNA are different from normal Watson-Crick base pairing which are known as Hoogsteen base pairs (27-29). In this thesis, triplets are described by placing the pyrimidine base of the duplex as the first letter, the purine base as the second letter and the third letter as the third strand base. *i.e.* YR.Z indicates interaction of the third strand base Z with a YR base pair.

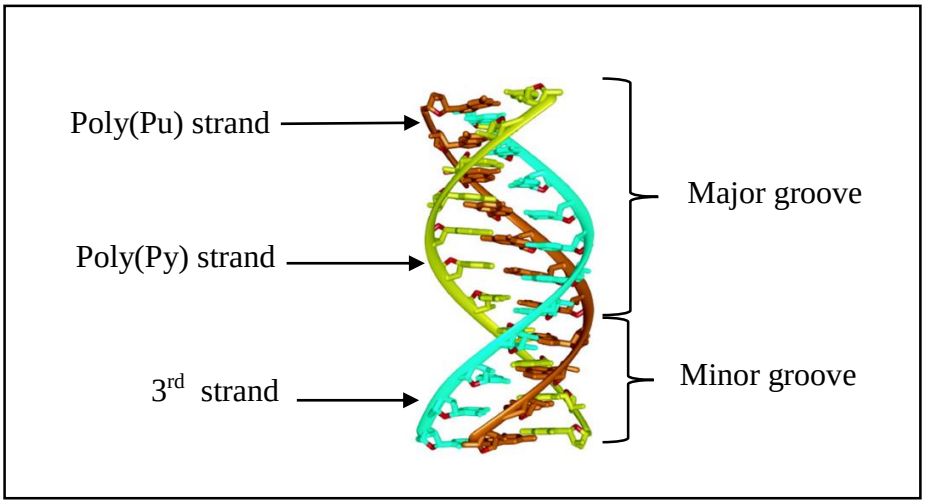



**Figure 1.5:** Molecular model of a DNA triple helix. The brown and yellow strands represent the polypurine and polypyrimidine sequences of the duplex DNA, respectively, whereas the green strand represents the third strand in the major groove of DNA, adapted from reference (30).

### 1.3.1 Triplex conformations

Triplexes can be divided into two main dissimilar types; intermolecular and intramolecular. An intermolecular DNA triple helix forms when a free third strand, triplex-forming oligonucleotide (TFO), binds to one strand of a region of duplex DNA (shown in Figure 1.6A). Intramolecular triplexes, generally named H-DNA, form when the third strand comes from the same molecule (for example a region of duplex DNA partially unwinds and one strand folds back on itself to form a triple helix leaving an orphaned strand) as shown in Figure 1.6B. Intermolecular triplexes have been the subject of much attention because of their therapeutic potential and biotechnological applications (31, 32), while intramolecular triplexes have significant roles in DNA function (33-35).

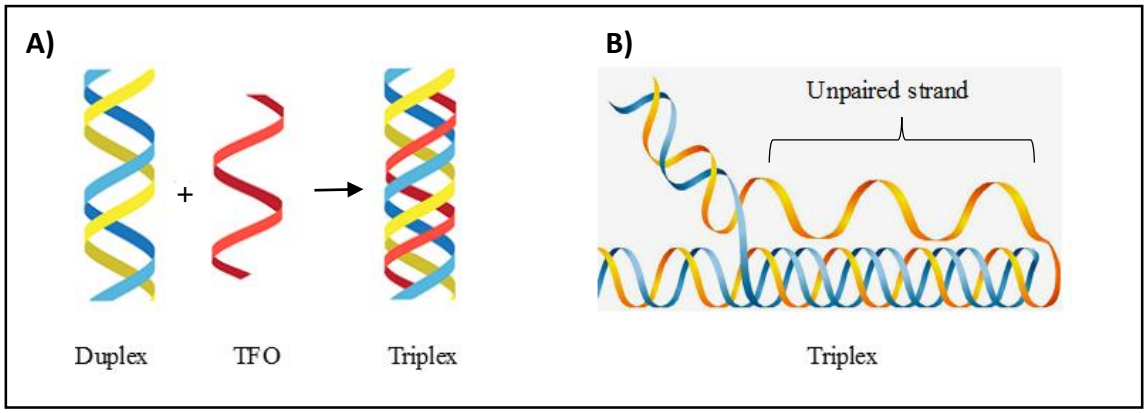



**Figure 1.6:** Illustrates schematic representation of intermolecular DNA triplex formation **(A)** and model of intramolecular triplex DNA **(B)**. (A) and (B) taken from reference (32).

#### 1.3.1.1 Intermolecular triplex DNA

Intermolecular triplexes have been the subject of extensive research in recent years due to their potential therapeutic application in inhibiting the expression of genes involved in cancer and other human diseases. They can also be used for targeting gene disorders for inactivation, stimulating DNA repair and/or homologous recombination pathways (32). However, triplex technology has some disadvantages in cellular systems, including low stability due to electrostatic repulsion between strands (32).

Two types of triplexes can be divided depending on the orientation of the third strand (Figure1.7). Those in which the third strand runs parallel to the duplex purine strand are characterised by (PyPu.Py) and stabilised by the formation of Hoogsteen base pairs (27, 28). In contrast antiparallel triplexes are characterised by (PyPu.Pu) and reverse-Hoogsteen base pairs enable reasonable stacking interactions for the triplex to be stabilised (22, 36). Parallel triplex formation generally requires low pH conditions, whereas antiparallel triplexes are pH-independent (37).

The chemical structures of the various triplets are shown in Figure 1.8. It can be seen that the third strand base only makes hydrogen bond contacts with one of the two duplex strands, the purine strand. The purine duplex strand therefore becomes the central strand of the triplex, so one side of the base binds the third strand (forming Hoogsteen base pairs), and the other side binds the other duplex strand (forming Watson-Crick base pairs). The third strand binds to the target purine strand in either a parallel or antiparallel orientation. CT-containing (pyrimidine) third strands bind parallel to the duplex purine strand. The most stable parallel triplets are TA.T (or TA.U) and CG.$C^+$; T (or U in RNA) (27, 38, 39). Protonation of the third strand cytosine is essential to form a second hydrogen bond with guanine; this will be covered in more detail later. AT.G and GC.T triplets have also been described, but these are much less stable than TA.T or CG. $C^+$ and create a different bonding pattern and triplet geometry (40-42). Purine-rich strands (GA- or GT-containing) bind to the central purine strand in an antiparallel orientation forming TA.A, CG.G and TA.T triplets. The TA.T triplet can also be formed in this orientation as it binds to adenine in a reverse Hoogsteen orientation (43-45).

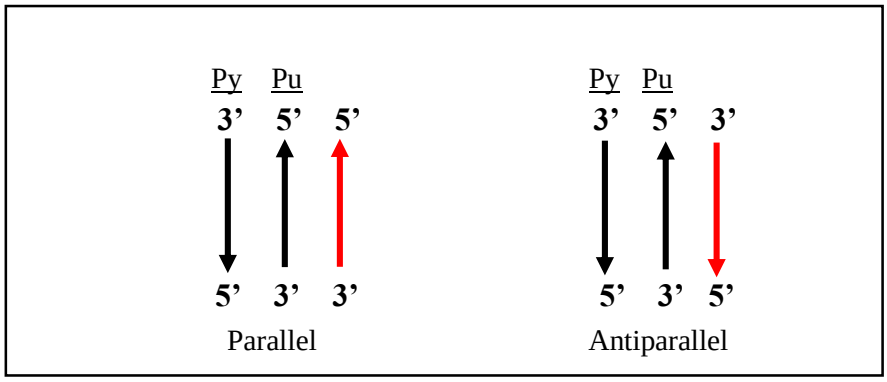



**Figure 1.7:** Illustrates the orientation of the third strand of the triplex motif .The third strand in red and binds with purine rich strand. Py and Pu represents pyrimidine- and purine-rich strands, respectively.

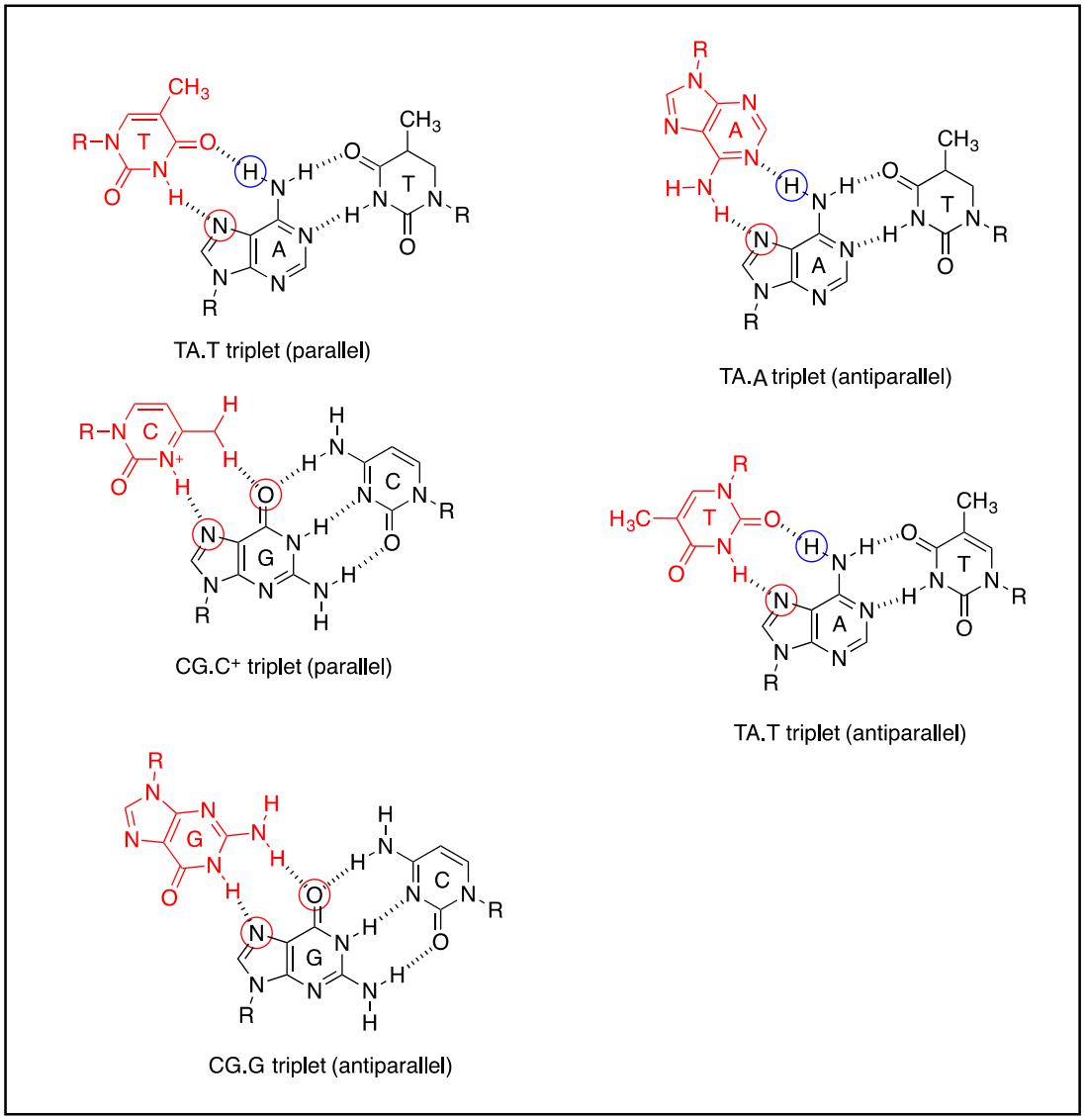



**Figure 1.8:** The structures of parallel and antiparallel triplets. Blue circles indicate hydrogen bond donors and red circles indicate hydrogen bond acceptors on the purine bases. The third strand is shown in red, while the Watson-Crick base pairs are shown in black.

#### 1.3.1.2 Structure of triplexes

To obtain more details about the exact structure of triplexes, various NMR and X-ray crystallographic studies have been carried out (46-51). The type of these triplexes were intermolecular oligomer complexes and intramolecular triplexes which formed in oligonucleotides folded twice upon themselves. In 1974, Arnott *et al.* reported the first study of triplex structure on the sequence poly(dT).poly(dA).poly(dT) using X-ray diffraction analysis (46). His studies reported that the triple helix has a similar radius to the corresponding duplex poly(dA).poly(dT). This led him to suggest that TFOs altered the basic DNA duplex to an A-DNA like structure, which has a particularly deep major groove that could accommodate the third strand without disrupting the double helix. Molecular dynamic investigations of the sequence poly(dT).poly(dA).poly(dT) demonstrated that the Watson-Crick duplex portion of the triplex has a completely separate conformation and differs from both A- and B-DNA (52). The sugar pucker, the major groove and the base tilt are similar to B-DNA, while the axial base displacement and helical twist resemble A-DNA. However, it was later shown that the actual structure of triplex DNA depends on the sequence and type of the third strand (*i.e.* DNA or RNA), but generally contains mixtures of characteristics for both A and B conformations (Figure 1.9) (53-55). NMR investigations have shown a general trend of increasing A-form with increasing RNA content. B-like conformations are more likely to be adopted with a DNA purine strand, whereas an A-like conformation could be formed with an RNA purine strand (56).

The first crystallographic study of a DNA triplex and its junction with a duplex was performed in 1999 by Rhee S. *et al.* (50). It was shown that general features of the triplex are similar to the conclusions from NMR studies unlike some differences in the helical parameters. It was deduced that formation of CG.$C^+$ triplets causes large changes in the phosphate backbone torsion angles of the central purine strand, probably due to electrostatic interaction between $C^+$ and the phosphate groups. These changes narrow the minor groove size and may provide the triplex structure with unique features to recognize DNA binding proteins. However, there is no data available on GAA-containing triplexes using X-ray crystallography.

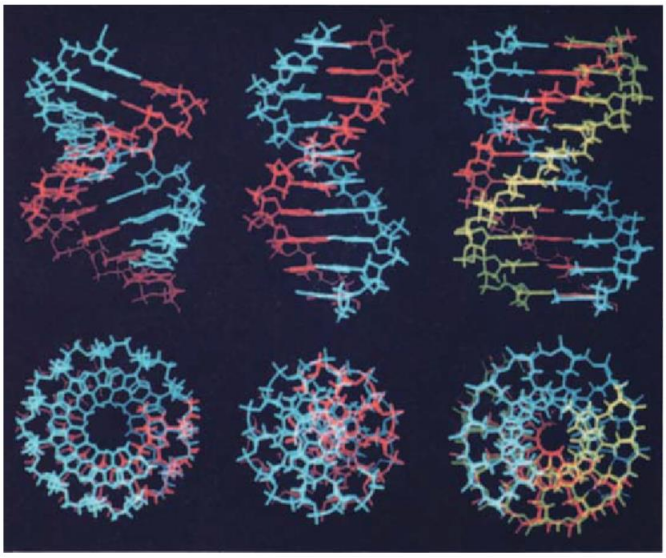


**Figure 1.9:** Model of A-DNA (left), B-DNA (middle) and triple helical DNA (right), third strand shown in yellow. Top: Side on view, Bottom: top down view. These structures obtained by X-ray diffraction (57).

#### 1.3.1.3 Conditions that required for intermolecular triplex formation

Triplex formation is affected by several factors including divalent cations, pH and temperature as well as base composition and length of the third strand itself. Triplexes in general are much weaker than their duplex counterparts. This is partly due to the bringing together of three polyanionic DNA strands in close proximity and subsequently greater energetic repulsion between the three negatively charged phosphodiester backbones. Triplex formation also involves a conformational change in the target duplex; to accommodate the third strand and this is may lead to slowing the process of triplex formation (58, 59).

I) Metal ions

Neutralization of phosphate groups reduces the repulsion of the negatively charged sugar-phosphate backbones and facilitates the three strands approaching each other. Monovalent cations; especially $Na^+$ (21), divalent cations; especially $Mg^{2+}$ and $Zn^{2+}$ (60), and polyamines such as spermine and spermidine (61), are known to have triplex-stabilising effects. It has been shown that the PyPu.Py triplex becomes more stable when the $Na^+$ and $K^+$ concentration increases beyond 0.1 M, but triplexes containing protonated cytosines are stabilised with up to 0.4 M $Na^+$ (60, 62). However, at higher $Na^+$ concentrations (>0.4 M $Na^+$), the $Na^+$ has a destabilising effect on the CG.$C^+$ triplet (63).

Additionally, $K^+$ can destabilise G-rich triplexes at physiological pH (64, 65), since G-rich sequences can adopt competing four stranded DNA structures. $K^+$ stabilises quadruple helices because it fits perfectly into the central cavity of the quadruplex. Thus, $K^+$ converts the triplex to a duplex and G-quadruplex DNA (66).

Divalent metal ions such as $Mg^{2+}$, $Mn^{2+}$or $Zn^{2+}$ , with a concentration of only 10 mM, are more effective than monovalent cations in stabilising triplexes (21). It was shown that these divalent cations are necessary to stabilise the (PyPu.Pu) triplex, presumably because the ions bind to the negatively charged phosphate backbone of the DNA (67-69), though there may also be specific interaction with the bases themselves. Some multivalent cations including polyamines provide a stabilising effect on both (PyPu.Py) and (PyPu.Pu) triplexes (61, 70-72). Polyamines such as spermine, spermidine and the diamine putrescine are present in mammalian cells in low concentrations. Their amino groups are protonated at physiological pH and bind strongly with the DNA bases; therefore polyamines can neutralize the repulsive charges between the phosphate groups within DNA and so stabilise intramolecular triplexes (73-75).

II) pH

In some cases, specific pH values are required for stabilising triplex structure. For example, at low pH (<6.0), N3 of the third strand cytosine of (PyPu.Py) tract should be protonated in order to form second hydrogen bond with the duplex guanosine. The pK of free cytosine is about 4.5, but this is increased within a triplex and may be as high as 9.0 for an isolated internal cytosine (76). This is due to the polyanionic environment and the binding energy of triplex formation. In addition, terminal cytosine bases are less stable, with lower pKs, because the protonated cytosines are exposed to solution leading to interaction with only one closest neighbour at the 3' or 5' end, instead of interaction with two neighbour bases (76).

It has been reported that $CG.C^+$ is more stable than the TA.T triplet (63, 76, 77). This is not only due to the presence of an additional hydrogen bond in the target duplex DNA but this can also be attributed to favourable electrostatic interactions between the positive charge on a protonated cytosine and the negatively charged sugar-phosphate backbone producing stronger hydrogen binding of $CG.C^+$ than TA.T (45, 63, 76, 77). Generally, parallel triplex (PyPu.Py) formation requires conditions of low pH (< 6.0) whereas antiparallel (PyPu.Pu) triplex occurs at neutral pH (76, 78-80).

III) Composition of the polypyrimidine.polypurine sequences

To bind a third strand of DNA (Pu or Py), the target duplex must contain oligopurine tracts, indicating an importance of hydrogen-bonding capabilities of purine bases in the formation of triplex (78, 81, 82). Attempts to generate triplexes at mixed purine-pyrimidine sequences, with oligonucleotides that contain natural nucleotides, have either failed or produced complexes with much lower affinity (83). This failure arises from the fact that within the major groove of duplex DNA, the pyrimidine bases have only one vacant hydrogen bonding site whereas the purine bases have two vacant Hoogsteen hydrogen bonding (Figure 1.10), therefore purine bases can form stable triplexes (84).

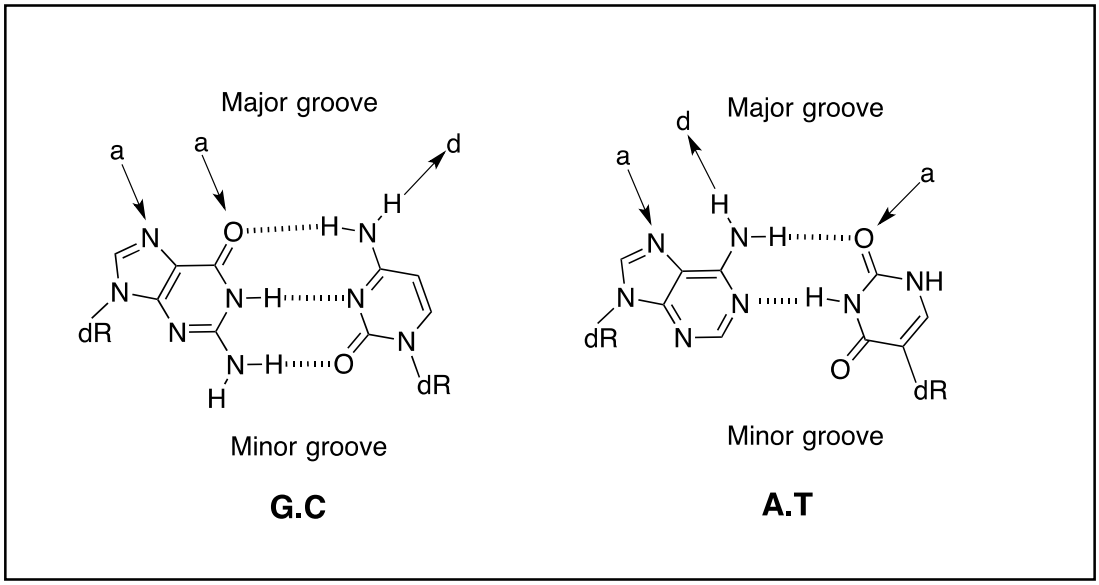



**Figure 1.10:** Standard duplex GC and AT base pairing. (a) and (b) indicate acceptor and donor, respectively, within the major groove of DNA. Purines have distinct binding profiles with bidentate vacant Hoogsteen sites whereas pyrimidines have only one.

### 1.3.1.4 Intramolecular triplex DNA or H-DNA

Most intramolecular triplexes are formed within a single poly(Py).poly(Pu) duplex DNA region having mirror symmetry in supercoiled DNA (85, 86). This is of interest because many sequences in the human genome have the potential to adopt intramolecular triplexes. Furthermore, these sequences are commonly distributed in the regulatory regions of genes and are generally located near promoter regions and recombination hotspots (86, 87). The most common structure is the PyPu.Py configuration in which half of the pyrimidine strand pairs as the third strand and the complementary strand of this

region remain unpaired. Two isomers of the PyPu.Py structure can form, depending on which half of the pyrimidine sequence, either the 5' (Hy5) or the 3' (Hy3), pairs as the third strand (88). Similar structures (Hu5 and Hu3) can be formed when the purine strand is folded back and donated as the third strand. All four possible isomers are presented in Figure 1.11.

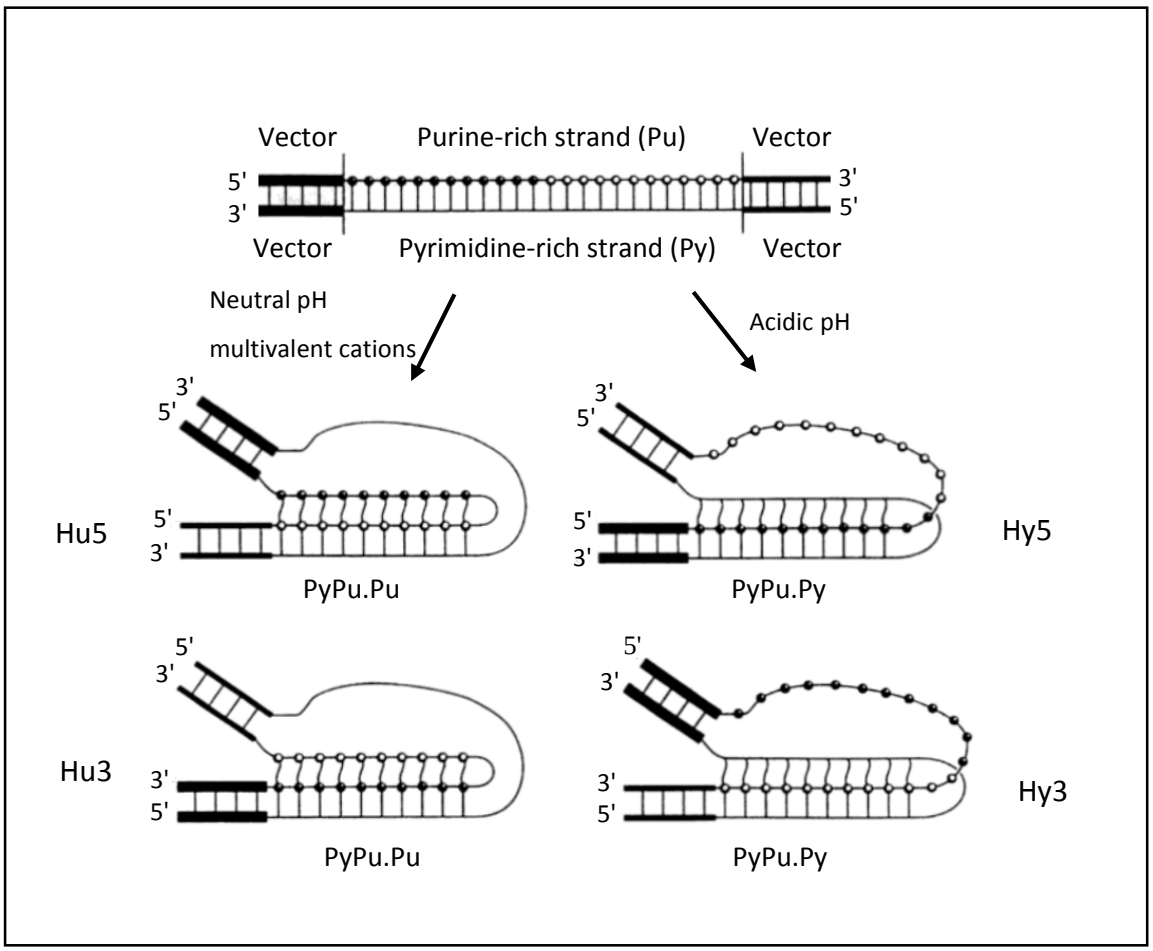



**Figure 1.11:** Illustrates four possible isomers of intramolecular triplex DNA under two main categories depending on pH and multivalent cations. The four triplex isomers are called Hu5, Hu3, Hy5 and Hy3. H denotes H-DNA; y and u refer to the pyrimidine-rich and purine-rich strands, respectively, as the third strand in the triplex. The numbers 5 and 3 indicate the 5' and 3' end of the third strand that binds to the duplex. Adapted from reference (89).

Generally, formation of a PyPu.Py or PyPu.Pu triplex depends on the pH and presence of multivalent cations as shown in Figure 1.11. The (PyPu.Py) class is characterised by the formation of TA.T and CG.$C^+$ triplets. As mentioned previously, N3 atoms of the third strand cytosine bases should be protonated which is possible in mildly acidic conditions

(pH ~ 4-5) (90). This type of triplex was discovered first and was denoted the "H form" to indicate that the medium required high concentrations of $H^+$ ions to form this triplex (91). The PyPu.Pu triplex is characterised by CG.G and TA.A triplets and does not require acidic pH for its formation, instead requires the presence of multivalent cations for stabilization. It was termed the "H* form" to distinguish it from the H form (92). For both types of H-DNA (H-DNA and H*-DNA, both will be referred to as H-DNA in the following sections).

#### 1.3.1.5 Requirement of supercoiling for intramolecular triplex formation

Although intramolecular triplex can form under similar conditions to intermolecular formation, DNA supercoiling is essential to form and stabilise H-DNA. Supercoiling is torsional stress on double stranded DNA that is maintained either by protein interaction or by the effect of the DNA being a closed circular form such as a plasmid. Early studies have shown that the torsional stress generated from negative supercoiling can facilitate the formation of alternative DNA secondary structures such as H-DNA (4, 39, 93), cruciforms (94, 95), and Z-DNA (96, 97). The free energy generated by negative supercoiling is necessary to drive the melting of the DNA at the centre of the Py.Pu region to allow H-DNA formation, which means that structural transitions within linear DNAs are not feasible (98). In addition, relaxation of torsional stress in supercoiled plasmid may result from locally unwound conformation. Therefore, increased levels of DNA supercoiling can provide the formation and stabilisation of H-DNA (4, 39, 93).

In living cells, most DNA molecules (and all plasmids) are present as circular loops and are negatively supercoiled; *i.e.* underwound (99). Positively supercoiled (overwound) DNA is transiently generated during DNA replication and transcription (100, 101). DNA isolated from a bacteriophage-like plasmid molecule from a *Sulfolobus* species, an archebacterium living at high temperature and low pH is positively supercoiled (102). DNA overwinding makes it more difficult to open the two strands of the double helix by heat and acid. Packaging DNA in a positively supercoiled form may protect the genetic material from denaturation (89).

The superhelical density of cellular DNA varies between -0.02 and -0.09, meaning there are 2-9 supercoils per 100 helical turns of DNA (103). There is interdependence between the low pH and plasmid negative superhelical density (104). For example, at pH 4, no negative supercoiling was required for PyPu.Py triplex formation and extreme negative

superhelical density (-σ = 1) was required for PyPu.Py triplex formation at neutral pH. In contrast, certain long Py.Pu sequences can form intramolecular triplexes in linear DNA at low pH (105).

#### 1.3.1.6 Evidence for the existence of triplex structures *in vivo*

The Py.Pu tracts are abundant in the eukaryotic genome, which constitute up to 1% of total DNA (106, 107) and less frequent in prokaryotic DNA (108). The most abundant tract is $(GA.TC)_n$ which accounts for 0.4% and 0.75% respectively of the primate and rodent genomes (109). An interesting thing of the Py.Pu sequences is their non-random distribution throughout the eukaryotic genome implying they may participate in various biological process (110, 111). Direct detection and role of H-DNA in eukaryotic cells is very difficult due to the high complexity of genomic DNA. Therefore, several research groups have employed *E.coli* cells bearing recombinant plasmids with triplex-forming inserts as a convenient model system. However, substantial progress in this field proposed three possible roles of triplexes formed by Py.Pu sequences in the eukaryotic chromosomes (Figure 1.12). First, intramolecular triplexes or H-DNA may present in the loops of DNA and might play a role in the control of gene expression. Second, intramolecular triplexes between the Pu.Pu tract and the third strand donated by another distant Py.Pu tract. This type is also called transmolecular triplex and could form at the base of the loop and be involved in assembling the loops of DNA on to the nuclear scaffold. Finally, transmolecular triplexes may form between loops and organise them into an ordered array (112). In support of this model was the finding that *in vitro* linear plasmids containing two separated Py.Pu tracts were able to form looped DNA structures as a result of triplex formation between the tracts (113).

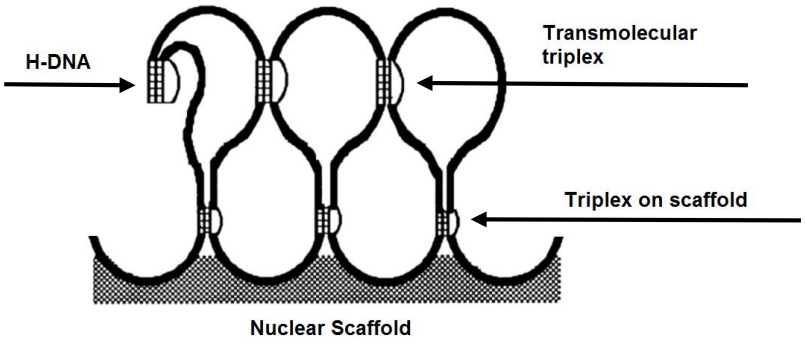



**Figure 1.12:** Model for triplex mediated genome organization. The formation of H-DNA and transmolecular triplexes between adjacent chromatin loops mediates genome organisation at nuclear scaffold or nuclear matrix. Taken from reference (112).

More evidence for the existence of triplex structures *in vivo* has come from immunofluorescent staining of fixed metaphase chromosomes with the triplex-specific monoclonal antibodies (114-117). These specific antibodies were prepared initially against duplex DNA/DNA third strand and their interaction examined for binding to cell nuclei and chromosomes by immunofluorescence. They showed significant binding to the nuclei which could be suppressed by adding competing triplex DNA but not by adding *E.coli* DNA. The results revealed a granular staining pattern throughout the nucleus with abundant triplex signals suggesting the presence of H-DNA and transmolecular triplexes between adjacent chromatin loops (112, 118, 119). Further work employing the same antibodies against duplex DNA/RNA third strand supported this observation. It was shown that large triplex signals were detected which may reflect DNA/RNA triple helices (120, 121).

In addition, chemical probing using osmium tetroxide ($OsO_4$) was applied to detect triplexes in a plasmid within *E.coli* cells. Osmium tetroxide reacts strongly with thymine and less with cytosine in single-stranded DNA, but only when these are in single stranded DNA or exposed in distorted regions of duplex DNA. It does not react with bases in duplex DNA (122, 123). Karlovsky *et al.*, in 1990 (124) utilised osmium tetroxide and bipyridine as a probe for intramolecular triplex formation at slightly acidic pH. *E.coli* cells containing a plasmid with a $(GA)_n$ triplex region were incubated in a buffer at pH 4.5-5 and treated with $OsO_4$. The plasmid DNA was isolated from the treated cells and showed that $OsO_4$ reacted with thymines at the centre loop of intramolecular triplex and at the triplex-duplex junctions. This demonstrated that intramolecular triplexes can form inside cells.

## 1.4 Sticky DNA

### 1.4.1 The discovery and properties of sticky DNA

Sticky DNA is a novel unusual DNA structure that is described as a self-associated complex formed by two stretches of FRDA-related long $(GAA.TTC)_n$ repeats ($n \geq 59$), under conditions that are more strict than those necessary for intramolecular triplex formation (shown in section 1.4.2). The two long repeat sequences must be within one supercoiled plasmid and in a direct repeat orientation (Figure 1.13) (125, 126). Sticky DNA was first observed by Sakamoto *et al*., 1999 (3). He discovered new non-B DNA conformation by abnormal changes in the mobility of DNA in gel electrophoresis

experiments. When linearised plasmids harbouring $(GAA.TTC)_n$ (n = 59 – 150) repeating tracts were run on agarose gels, a retarded band, migrating seven times slower (equivalent to 42 Kb) than the linear DNA fragment (7.1 Kb) was observed. However, the amount of the "17 kbp" retarded band increases as the repeat length increases, whilst the linearised plasmids containing 270 GAA repeats did not exhibit any retarded band. His investigation revealed that this retarded band is sticky DNA conformation and consists of the association of two H-DNAs from two separated plasmids, each plasmid containing a single GAA repeats and forming individual H-DNA. Thus, sticky DNA is considered a very stable bi-triplex conformation. To explain this extremely stable structure, a strand exchanged model was proposed in which the pyrimidine strands of two PyPu.Pu triplexes were exchanged with each other as shown in Figure 1.14. Since this dimeric DNA association has single stranded region as a part of the pyrimidine strand, this model agrees with the results of P1 nuclease and osmium tetroxide experiments. It was found that the TTC strand was much more sensitive than the GAA strand to P1 nuclease and osmium tetroxide as expected for triplex (PyPu.Pu) (43, 127). Furthermore, unpaired nucleotides of the TTC strand can associate with the GAA strand of the other molecule producing different lengths of the TTC strand. Again, this interpretation supports the P1 nuclease sensitivity of the entire TTC sequence. Prior to this occurring, the loops of the GAA strands of the triplexes would be increased due to the dissociation of the TTC strand from the triplex (PyPu.Pu) causing the GAA loops to destabilise. Therefore, it was believed that length of the GAA.TTC repeat should be at least 59 repeats to form DNA.DNA associated complex.

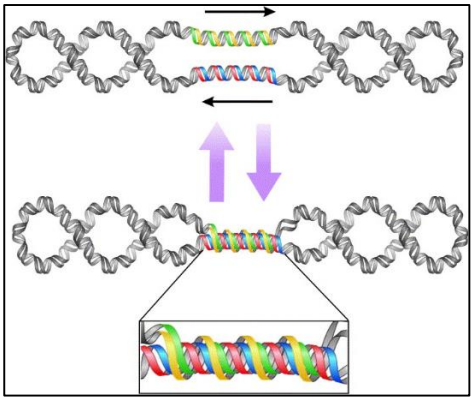


**Figure 1.13:** Model for sticky DNA structure in a supercoiled plasmid. Sticky DNA is formed by the association of two long repeats of GAA.TTC in one closed circular DNA. The green and yellow strands represent one GAA.TTC duplex and the red and blue represent the other duplex (6).

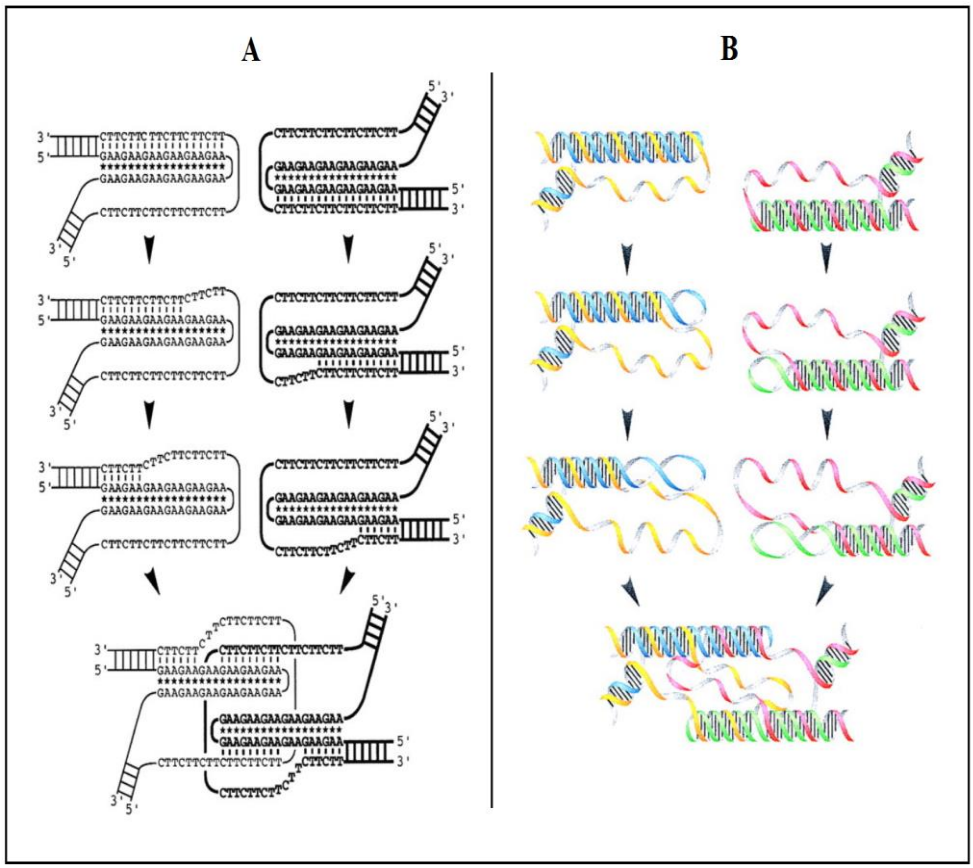
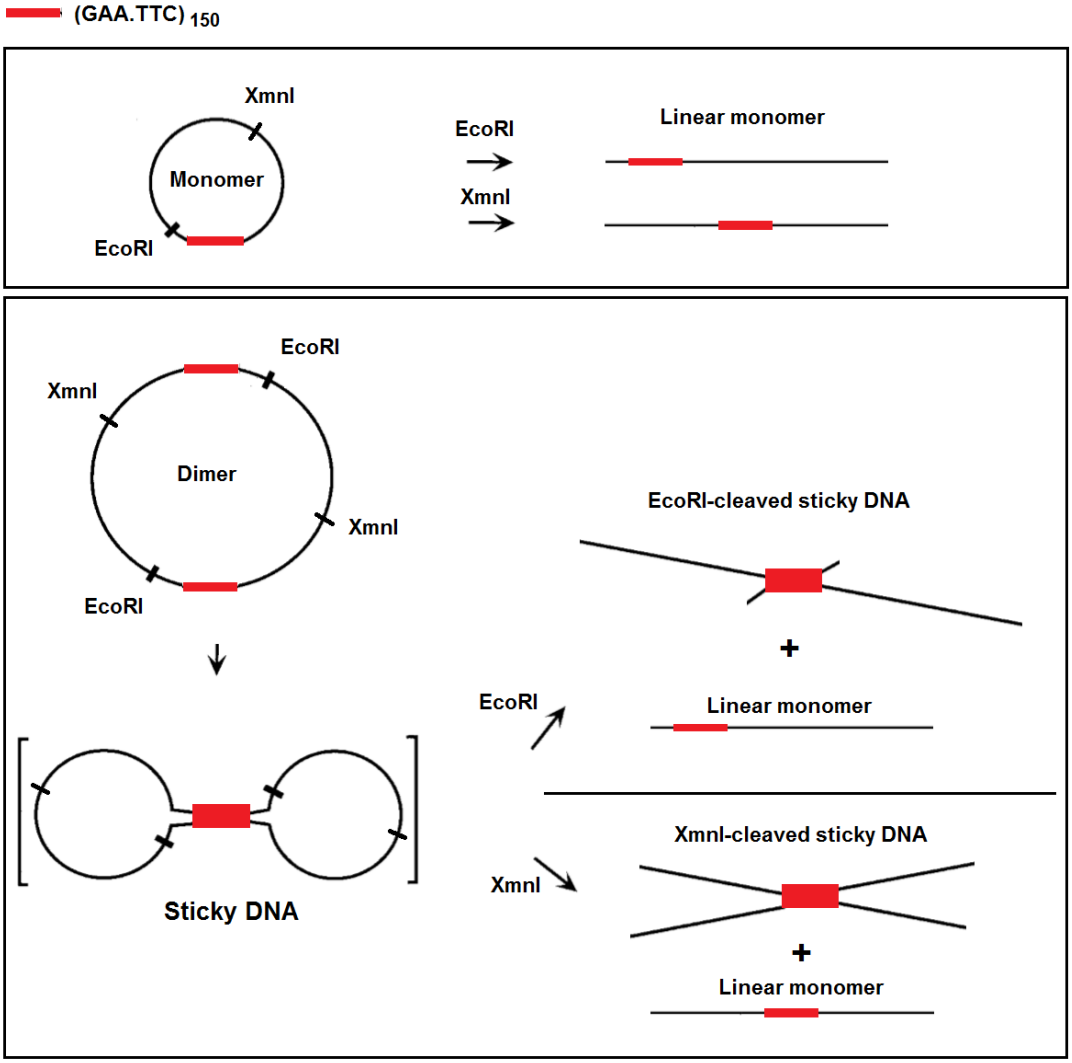



**Figure 1.14:** Model for the association of two triplexes formed by long GAA.TTC tracts. **(A)** Schematic picture of the strand exchange model. The two triplexes are represented as thin and thick lines. The short vertical lines between the bases represent Watson-Crick pairs, and the stars represent the reversed Hoogsteen base pairs. **(B)** Three-dimensional picture of the strand exchange model. Different colours represent different strands. In the left molecule, blue shows the purine strand, while yellow shows the pyrimidine strand. (A) and (B) taken from reference (3).

The hypothesis of sticky DNA formation has been modified by Vetcher *et al.*, 2002 (125), since it has been shown that sticky DNA is only formed intramolecularly from two stretches of GAA repeats in a direct orientation within one supercoiled plasmid. Plasmids harbouring one tract of $(GAA.TTC)_{150}$ repeats were analysed for their capacity to form sticky DNA. Previous studies show that plasmids containing GAA repeats exhibit monomers, dimers and higher oligomeric DNA forms on agarose gel. Monomers and dimers were isolated and cleaved by EcoRI or XmnI to allow retarded band formation (Figure 1.15A). The restriction products were analysed on agarose gel as shown in Figure 1.15B. The retarded band was formed only from the dimeric, not the monomeric form, of the construct plasmid suggesting that two long GAA.TTC repeats must be present within one closed plasmid to observe sticky DNA.

**(A)**


(GAA.TTC) 150
XmnI
Monomer
EcoRI
EcoRI
XmnI
Linear monomer
EcoRI
XmnI
Dimer
XmnI
EcoRI
Sticky DNA
EcoRI-cleaved sticky DNA
+
EcoRI
Linear monomer
XmnI-cleaved sticky DNA
XmnI
+
Linear monomer

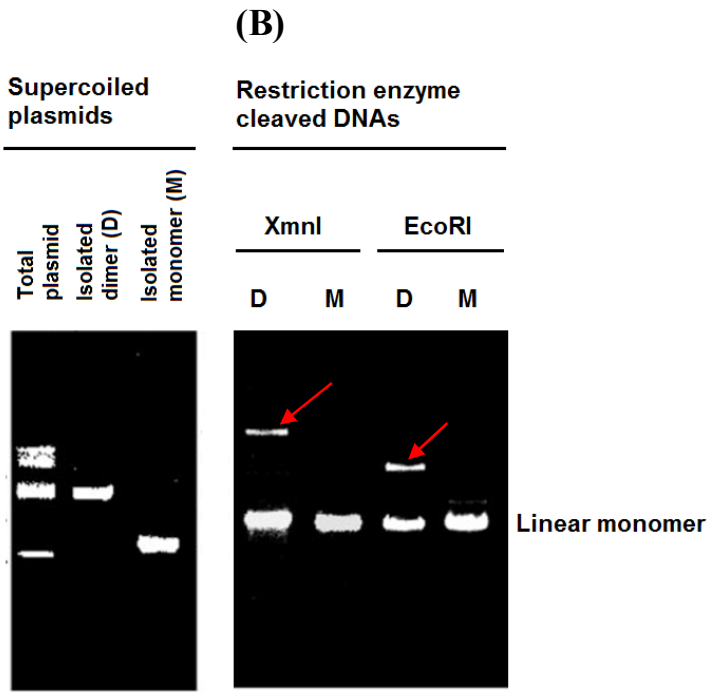



**Figure 1.15:** Illustrates contribution of the dimer or monomer forms of plasmid containing $(GAA.TTC)_{150}$ to the yield of retarded band (sticky DNA). **(A)** Schematic representation of the enzymatic cleavage, with EcoRI or XmnI, of the construct plasmid in the dimeric or monomeric forms. Retarded band or sticky DNA formation was observed only when more than one GAA.TTC tract (dimer) was present in the supercoiled DNA molecule. **(B)** Agarose gel (0.7%) electrophoresis of native the construct plasmid, dimers (D) and monomers (M) of the plasmid (left panel). The dimeric and monomeric forms were cleaved with EcoRI or XmnI and analysed on a 0.7% agarose gel (right panel). Red arrows represent retarded band (sticky DNA) from XmnI or EcoRI cleavage. The extent of the retardations was related to the distance between the digestion positions and the insert. (A) and (B) adapted from reference (125).

This model of sticky DNA was confirmed by Son *et al*., 2006 (6), on single digestion of plasmid containing two tracts of $(GAA.TTC)_{176}$ repeats in the direct orientation (Figure 1.16). The cleavage products consisted of a retarded band in addition to the presence of the reminder linear fragment. This retarded band consisted of one side of the enzyme-cleaved plasmid in the form of two linear duplex free arms and the other side was supercoiled domain (structure 2). In addition, the nature of the retarded DNA band species was studied by incubation of the digested plasmid at 80 °C for 10 minutes with EDTA, chelating any cations that might stabilise unusual DNA structure formed by the repeat tracts. The results revealed that the retarded band disappeared with complete restoration of the linear monomer DNA product (structure 3). The dissociation of the sticky DNA in a supercoiled molecule was observed after treatment with EDTA at 80 °C for 10 minutes (structure 4). The resultant DNA was then digested with unique restriction enzyme

generating one linear DNA fragment (structure 5). To demonstrate the ability of the GAA.TTC tracts to re-associate under superhelical stress, the supercoiled non-sticky DNA (structure 4) was incubated with 5 mM $MgCl_2$ for 2 hours at 37 °C. The sticky DNA was formed (structure 6) and was observed as a cleaved sticky retarded band after digestion of the supercoiled sticky DNA with single restriction enzyme (structure 7), followed by treatment with EDTA at 80 °C for 10 minutes producing only a linear DNA (structure 8). This result of dissociation and re-association of the supercoiled sticky DNA indicates that divalent cations play an important role in formation and stability of the structure, since chelating by EDTA dissociated the sticky DNA and addition of $MgCl_2$ to supercoiled non-sticky DNA aided in re-association of the structure. In addition, it was found that absence of $MgCl_2$ resulted in the lack of the DNA.DNA association under superhelical stress, even after an incubation time of 24 hours.

The role of supercoiling on association of a pair of GAA.TTC tracts was also examined. This was performed by relaxation of the supercoiled non-sticky DNA with topoisomerase I to obtain a relaxed closed plasmid (structure 9). The relaxed DNA was then incubated with $MgCl_2$ concentrations of 5 mM -50 mM at 37 °C for up to 24 hours and subsequently digested with the same unique restriction enzyme. No cleaved sticky retarded band for this relaxed DNA molecule was observed (structure 10), concluding that supercoiling is required for sticky DNA formation, and also $MgCl_2$ alone was not sufficient to promote the reconstitution of the GAA.TTC associated structure. All data in Figure 1.16 was detected by agarose gel electrophoresis.

Parallel experiments were performed on plasmid containing the same length of pair of GAA.TTC tract but in the inverted orientation (6). No sticky DNA was observed even with high concentrations of $MgCl_2$ for up to 24 hours incubation time. Hence, sticky DNA formation also depends on the orientation of the cloned inserts.

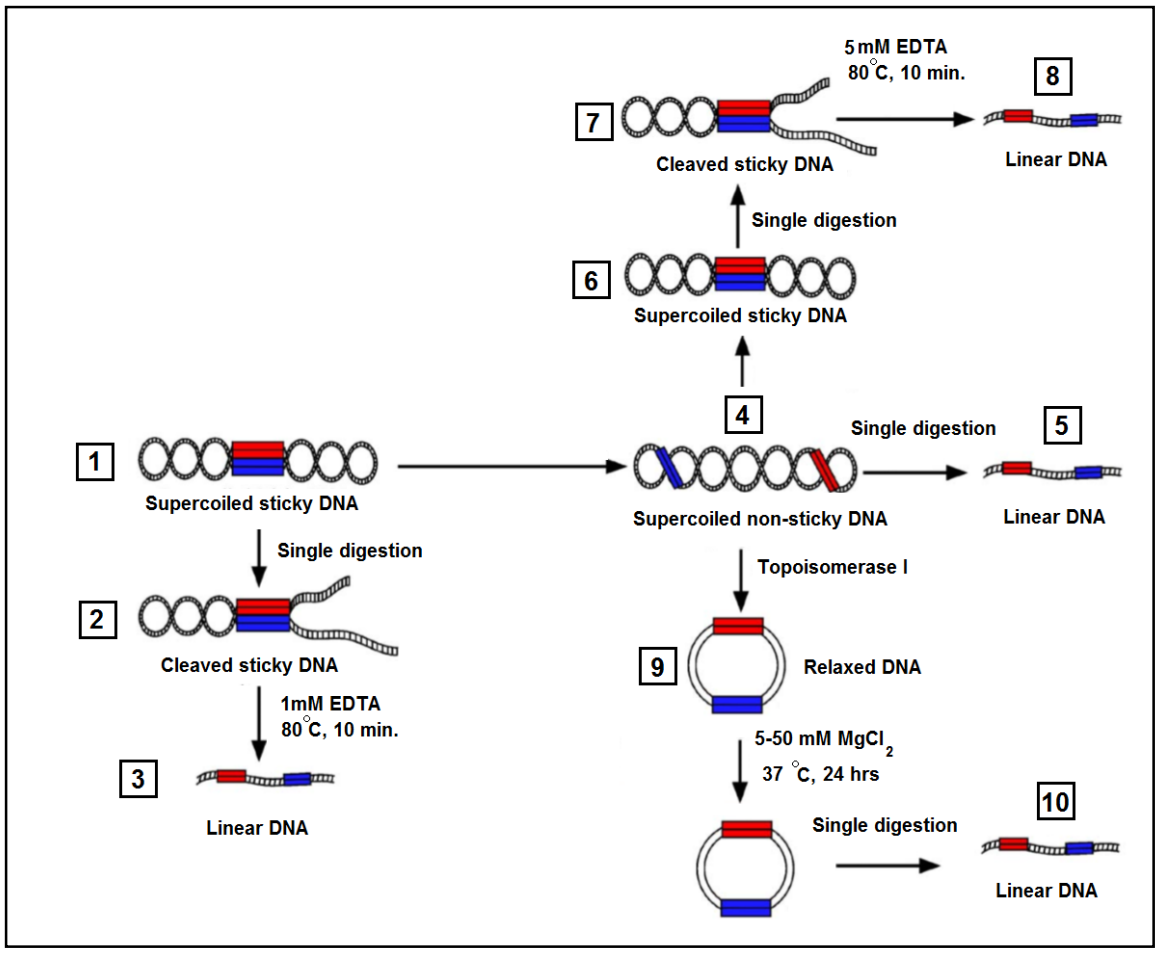



**Figure 1.16:** *In vitro* dissociation and reconstitution of sticky DNA with supercoiling and $MgCl_2$. Adapted from reference (6).

Sticky DNA was also detected by electron microscopy (EM) as an X-shape (Figure 1.17) after digestion of the dimer plasmid, which harboured two long (GAA.TTC) tracts, with one restriction enzyme (Figure 1.18) (125). This technique provided strong evidence that the DNA.DNA associated region consists of a single long triplex and is formed intramolecularly. This was attributed to the length of the sticky DNA region being longer than half the known length of the GAA.TTC tract, since the strand-exchanged bi-triplex model (Figure 1.14) consists of two folded back intramolecular triplexes. In addition, EM determinations revealed that sticky DNA would be very rigid compared with the duplex DNA. This is because of the absence of bulges, bubbles and flexible regions in the region of association, although the strand-exchanged bi-triplex model could be more flexible due to the extent of unpaired bases between the two triplexes.

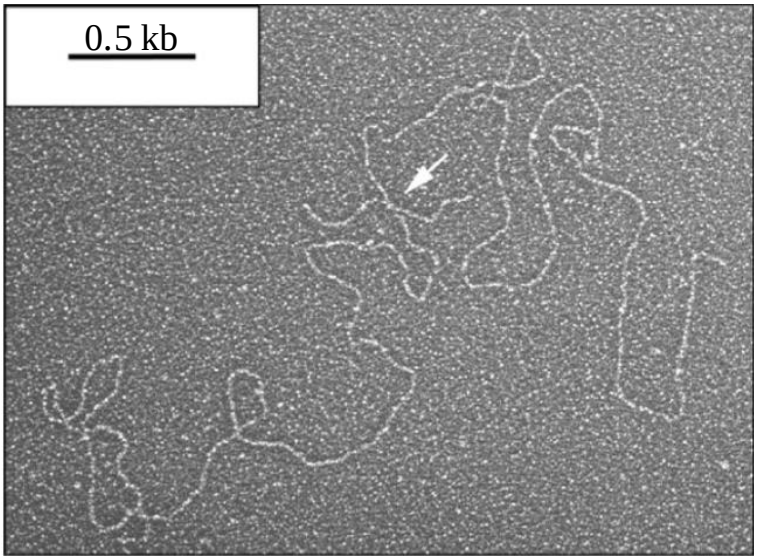



**Figure 1.17:** EM visualization of plasmid dimer cleaved to place the region of sticky DNA near one end. The arrow reveals the region of sticky DNA. Adapted from reference (125).

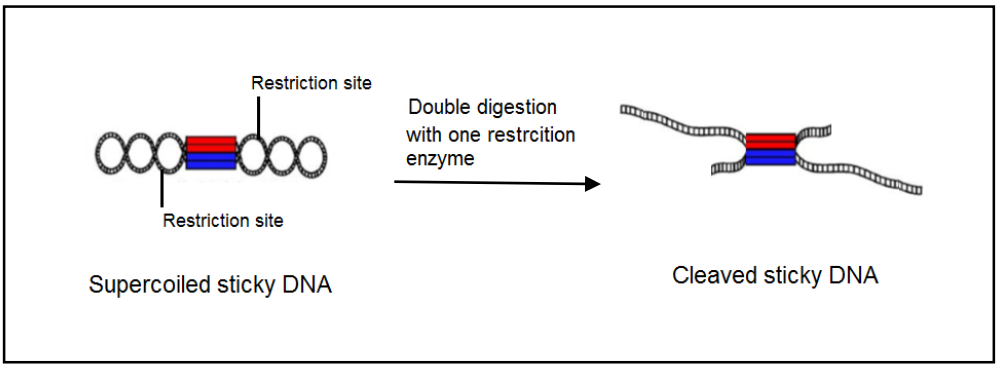



**Figure 1.18**: Example of a plasmid dimer digested with one restriction enzyme which has two recognition sites that are separated by the insert.

### 1.4.2 Requirements for sticky DNA formation

(I) Metal ions

Divalent metal ions ($Mg^{2+}$ and $Mn^{2+}$) promote and stabilise sticky DNA formation (6). Son *et al*., 2006 (6) showed that different concentrations of $MgCl_2$ or $MnCl_2$, together with negative supercoiling, facilitate the formation of sticky DNA between the GAA.TTC tracts. The association process was dependent on time and the temperature of incubation of the DNA with the metal ions. It was found that 1 mM $MgCl_2$ with a 4 hours incubation time is sufficient for sticky DNA formation and that increasing the concentration of $MgCl_2$ decreased the time required to associate the GAA.TTC tracts. In addition, effects of other monovalent (KCl, LiCl) and divalent ($ZnCl_2$, $CaCl_2$, $CoCl_2$) cations on association of the GAA.TTC tracts were examined but all of them failed to stabilise sticky DNA structure.

(II) pH

The effect of pH on sticky DNA formation has been investigated and the amount formed at pH 8.0 was about 3.5 times greater than at pH 5.0. It is known that neutral pH in the presence of divalent metal ions favours the H*-form of intramolecular triplexes, while the H-form is favoured below pH 6.0. This suggests that sticky DNA is generated by the association between two (PyPu.Pu) triplexes (78, 82, 128, 129).

(III) Supercoiling

Negative supercoiling is also required for the formation of sticky DNA and for the two long GAA.TTC tracts to interact with each other. By using topoisomerase I treatment in the presence of different concentrations of ethidium bromide Son *et al*., 2006 (6) generated plasmids (the length of the vector is 4361 bp excluding the inserts), containing a pair of $(GAA.TTC)_{176}$ tracts (direct and inverted orientation) with different superhelical densities. A cleaved sticky DNA retarded band was observed only at supercoil density values ($-\sigma$) between 0.051-0.075; higher or lower levels of negative supercoiling ($<0.05$ and $>0.075$) did not generate the slow migrating sticky DNA species as shown in Figure 1.19. This could be due to two reasons: low levels of supercoiling may not be sufficient to allow slithering of the different regions of the closed plasmid (130-133), while higher superhelical density may cause the DNA structure to become inflexible, thereby hindering association of the two distal GAA.TTC tracts. The two distal GAA.TTC repeats may also

adopt two separate intramolecular triplexes at high superhelical densities, which then prevent sticky DNA formation (134).

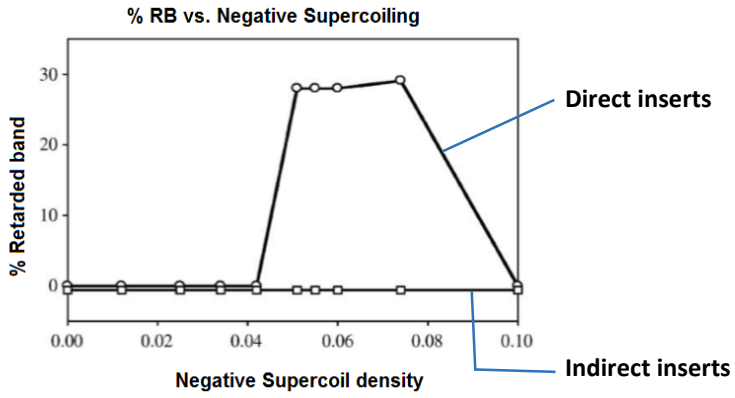



**Figure 1.19:** Illustrates the supercoil densities required for sticky DNA formation. Plasmids containing a pair of $(GAA.TTC)_{176}$ , either in direct orientation (circles) or inverted orientation (squares), were relaxed by topoisomerase I and treated with different concentrations of ethidium bromide ranging from 0-5 µg/ml to generate topoisomer families with different negative supercoil densities (-σ = 0-1.0). The topoisomers were cleaved with unique restriction enzyme to determine the amount of cleaved sticky DNA retarded band at certain negative supercoil density. The retarded band was observed with plasmid containing direct inserts at (-σ) between 0.051-0.075, (on average, 29% of the DNA was cleaved sticky DNA and the remainder was in the linear monomer form. For plasmid containing inverted inserts, no retarded band was found at any supercoil density. Taken from reference (6).

(IV) Length and orientation of the GAA.TTC repeats

As mentioned previously, the minimum length of the triplet repeat required for DNA.DNA interaction must be at least 59 repeats and lengths shorter than 59 repeats failed to form this novel DNA structure (125, 126). Correlation was found between the lengths of GAA·TTC sequence and the formation of sticky DNA: FRDA patients have 66 or more repeats, sticky DNA was found only for repeats longer than 59 units (3). Therefore, it has been proposed from *in vitro* transcription studies of GAA.TTC repeats that transcriptional inhibition by sticky DNA invlove sequestration of the RNA polymerases by direct binding to the complex DNA structure (2). The repeating hexanucleotide sequence $(GAAGGA.TCCTTC)_{65}$ was also found in the first intron of the frataxin gene, in which no inhibition of transcription was observed or associated with FRDA (135). Thus, it has been suggested that interruption in the GAA·TTC sequence may destabilise sticky DNA structure and facilitate transcription. It was found that the insertion of >11% of the GGA.TCC sequence into a pure $(GAA.TTC)_{150}$ repeating

sequence inhibited sticky DNA formation and relieved the inhibition of transcription. This effect depended on the length of the inserted sequence $(GGA.TCC)_n$ (136).

Influence of orientation of the inserts on retarded band formation was investigated by several studies (6, 125, 137); both GAA.TTC inserts should be in the head-to-tail orientation. For example, each of the GAA repeat tracts on the same DNA strand should be in the same 5'→3' orientation. In a study by Vetcher *et al.*, 2002 (125), four plasmids were digested, each containing a pair of $(GAA.TTC)_{60}$, in all possible orientations of the inserts (Figure 1.20). The ability of these plasmids to form sticky DNA was examined by cleavage with restriction enzyme AccI, which had two recognition sites that are separated by the GAA.TTC inserts. Only plasmids which harboured two tracts of $(GAA.TTC)_{60}$ in the same orientation produced cleaved sticky DNA, but this was not found when the tracts were in the inverted orientation. In a further study the effect of orientation of the inserts on the existence of sticky DNA was investigated by using different negative supercoil densities and high concentration of magnesium (6). The results revealed that no retarded bands were observed with inverted inserts but only with direct repeats, confirming that inserts orientation play a significant role in the DNA.DNA interaction under certain superhelical stress.

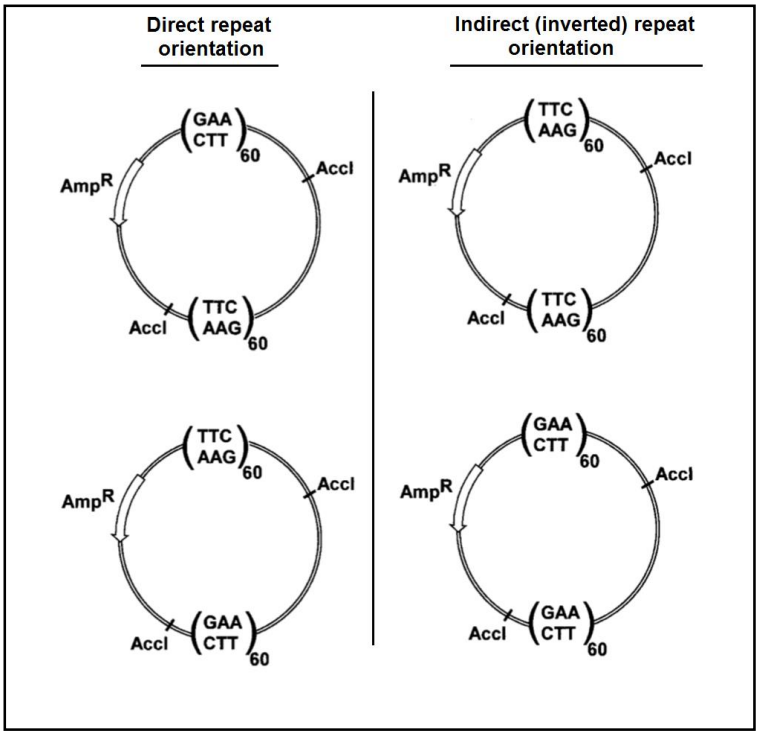



**Figure 1.20:** Schematic diagram of plasmids containing two GAA.TTC tracts used to determine the influence of the insert orientation on retarded band formation. Taken from reference (125).

## 1.5 Friedreich's ataxia

Friedreich ataxia (FRDA) is an autosomal-recessive disease primarily characterized by progressive damage to the nervous system (138), affecting about 2 people in 50,000 (20). In 1863, Nikolaus Friedreich (139), described FRDA as a neurological disorder caused by mutation in the frataxin (FXN) gene. This mutation was discovered by Campuzano *et al.* in 1996 (140). It was defined as a homozygous expanded GAA trinucleotide repeat in the first intron of the frataxin gene (140). Onset can start in very young children or in adulthood, but usually before the age of 25 (141). Clinical features include progressive gait and limb ataxia, muscular weakness of the legs and positive extensor planter response. At present, there is no effective treatment for FRDA and the affected people usually die at a young age (142).

### 1.5.1 FXN gene: location, structure and expression

The first mapping of the FXN gene was accomplished by Chamberlain *et al.* in 1988 (143). The human FXN gene, which is responsible for FRDA, was mapped to chromosome 9 and is localised in the proximal long arm at position 9q13-q21.1, as shown in Figure 1.21. The FXN gene consists of seven exons: exons 1-4, alternate exons 5a/5b and a sixth non-coding exon (present in exon 5b-containing transcripts) (140). Transcription goes in the centromere to telomere direction (Figure 1.22). The major transcript, exons 1 to 5a, is a 1.3kb mRNA that encodes a 210 amino acid protein called frataxin. By alternative splicing, exon 5b can be transcribed producing different protein isoform (140). The FXN gene is expressed in all cells with variable levels in different tissues and during development (140, 144). In adult human, frataxin mRNA is mainly expressed in the heart and central nervous system, followed by liver, skeletal muscle and pancreas (145, 146).

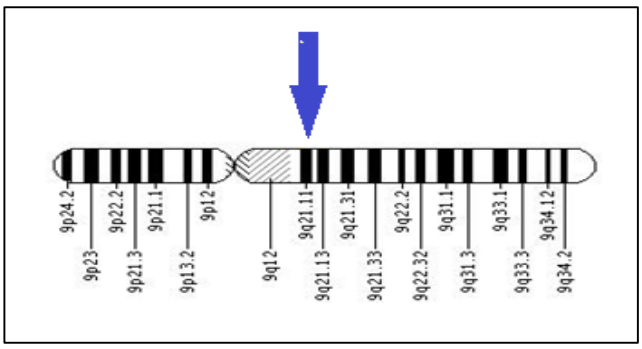



**Figure 1.21:** Schematic representation of human chromosome 9. The blue arrow indicates the location of the FXN gene on the long arm (q).

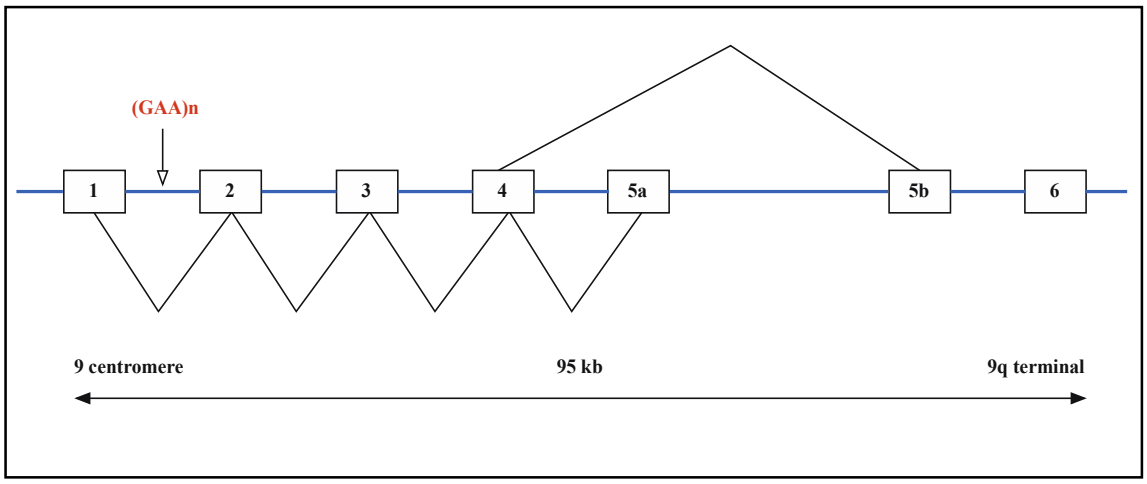



**Figure 1.22**: Transcription map of the frataxin gene. The gene extends from centromere to telomere of the 9q chromosome. The GAA repeats are situated in intron1. Adapted from reference (147).

### 1.5.2 Frataxin protein

Frataxin is a highly conserved mitochondrial protein that is found in all cells from bacteria to humans (140, 148). It plays a significant role in the assembly of iron-sulphur clusters and in maintaining the iron homeostasis in the mitochondria which is needed for energy production (149, 150). Immunofluorescence and immunoelectron microscopy have shown that frataxin is located in the mitochondria near its inner membrane via an N-terminal localisation signal (145). The crystal structure of mature frataxin protein (Figure 1.23) shows a compact αβ sandwich, including 7 β-sheets (β1-β5, β6 and β7) interacting with 2 α-helices (α1 and α2). The α1- and α2-helices are almost parallel to each other and the large β-sheets (β1-β5) are antiparallel sheets. The β6 and -β7 intersect the planes to give an overall compact αβ sandwich structure (151). Frataxin protein has large, negatively-charged side chains on its surface and that helps the protein to bind iron contributing to cellular iron homeostasis (151, 152).

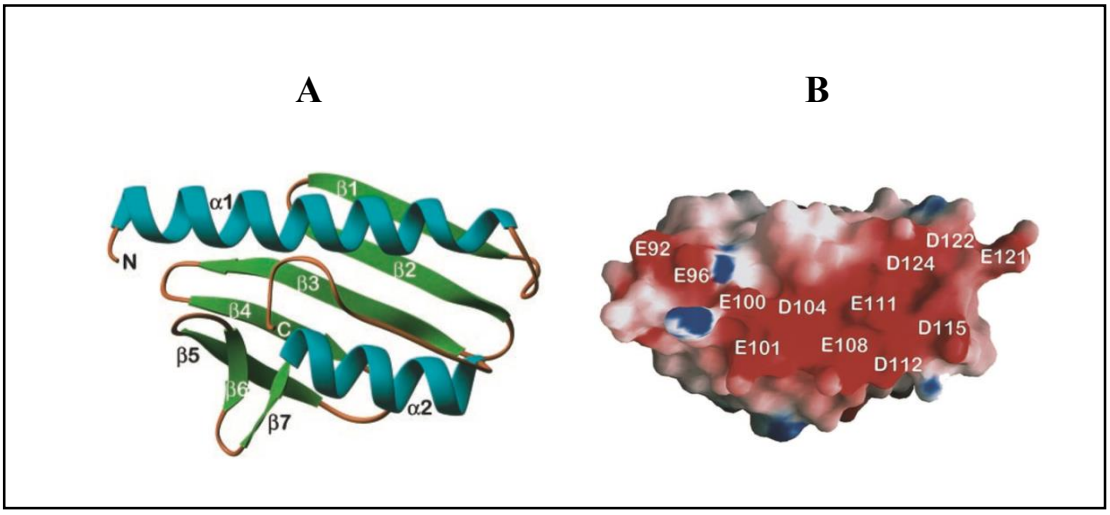



**Figure 1.23:** Structure of frataxin. **A)** Crystal structure of human frataxin shows the fold of frataxin, a compact αβ sandwich, with α helices and β sheets. Strands β1 – β5 forming a flat antiparallel construct that interacts with the two helices, α1 and α2. **B)** Molecular surface representation of frataxin. Several anionic residues in α1 and β1 are shown in red. The anionic surface might be critical for the function of frataxin. Acidic residues (glutamic acid and aspartic acid) are represented by the letters E and D, respectively. Taken from reference (151).

### 1.5.3 Effect of GAA repeat expansion on frataxin gene expression

Several genes are known to contain transcriptional enhancers within their first introns, suggesting that the GAA triplet repeat sequence can potentially function as a factor in regulating gene expression (153, 154). The expanded GAA repeat within the first intron of the FXN gene may impede the transcription machinery resulting in immature mRNA and then lower generation of frataxin protein (Figure 1.24), causing the clinical phenotype of FRDA (20, 155, 156). In contrast to other TNR disorders such as Huntington disease, long stretches of GAA.TTC in the FXN gene do not cause changes in the frataxin amino acid sequence, but result in decreased expression of the normal frataxin protein (142).

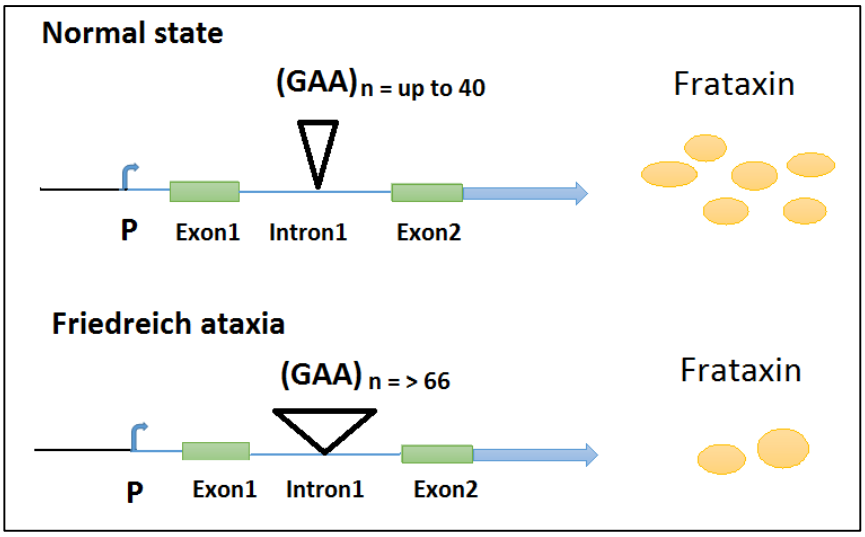



**Figure 1.24:** Schematic representation of frataxin expression. Unaffected individuals, who have up to 40 GAA repeats have more frataxin protein, whereas affected state (FRDA) produce less frataxin. The size of intron 1 is 12 kb and the triplet repeat is located in 1.4 kb after exon 1 (157). P is the promoter.

All molecular biological studies in the FRDA field believe that inhibition of FXN transcription is the main cause of this neurological disease. However, exactly how the GAA.TTC repeats in intron 1 cause the FXN mRNA deficit is still unknown. *In vitro* and *in vivo* investigations suggested three different hypotheses; abnormal non-B DNA conformations and/or heterochromatin-mediated gene silencing or splicing process (20, 155, 158, 159). Firstly, it has been suggested that the GAA.TTC repeat expansion may adopt unusual DNA structures such as triplexes or sticky DNA, or DNA.RNA hybrid structures (R loops), which impede the process of RNA polymerase and thus inhibit frataxin gene transcription. A simple model (Figure 1.25) was proposed to explain the inhibitory effects of triplex formation on transcription elongation. During transcription of a long GAA.TTC repeat a transient intramolecular triplex (PyPu.Pu), pausing RNA polymerase. The polymerase moves along the template and causes the formation of supercoils in the DNA, with negative supercoiling upstream of the transcription bubble and positive supercoiling downstream. The negative superhelical energy that formed behind the polymerase promoted the formation of a short–lived intramolecular triplex. The junction between the triplex and the duplex DNA at the distal end of the repeat tract causes polymerase arrest, resulting in a transcript truncated at the 3' end of the structure (155, 160, 161).

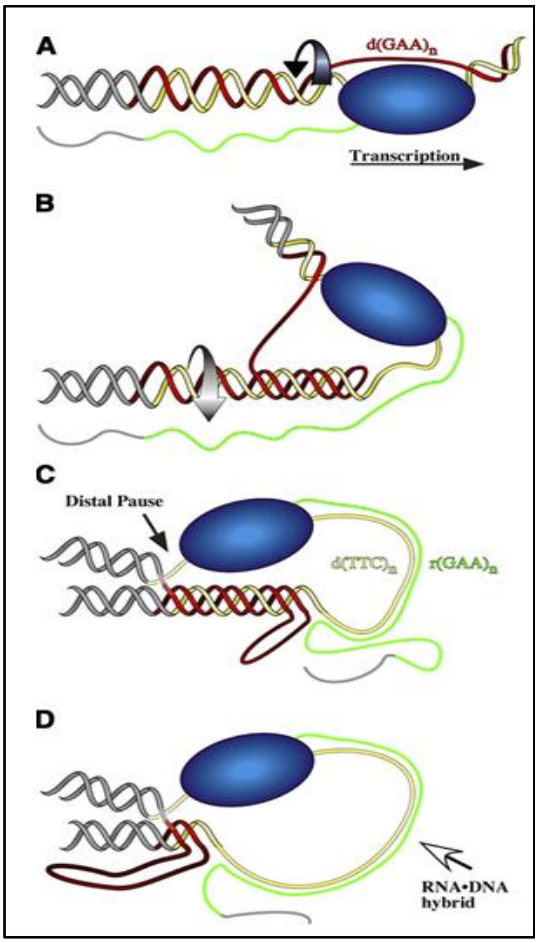



**Figure 1.25:** Model for transient transcription-dependent triplex formation leading to pausing of RNA polymerase and RNA·DNA hybrid formation. The purine (GAA or Pu) strand of the repeat is shown in red, while the pyrimidine (TTC or Py) strand is in yellow, and the flanking DNA is grey. **A**) A standing wave of negative supercoiling follows RNA polymerase. At the transcription bubble, the non-template (GAA) strand is available to fold back into an RY.R interaction; the template strand is covered by RNA polymerase. **B**) Rotation of the helix (curved arrow), as it winds in the third strand, relaxes the negative supercoils caused by transcription and leaves a length of the template single-stranded. **C**) RNAP is impeded at the distal template-triplex junction and the nascent transcript (green) can anneal to the single-stranded stretch of template. **D**) The RNA·DNA hybrid displaces the less stable triplex structure.

Secondly, the FRDA GAA.TTC repeats have shown to decrease mRNA splicing efficiency producing a deficit of mature RNAs. This results from aberrant splicing in which intron 1 is retained. This effect of the triplet repeat was attributed to the ability of the repeats to bind splicing factors (Figure 1.26) which can interfere with normal turnover of intronic RNA and thus lead to its degradation and yield lower amounts of mature RNA (162). In addition, the defect in pre-mRNA processing was observed to be position and context dependent; the insertion of the repeat at various distances from the reporter exons and not in the first intron of the gene showed a variable effect on splicing efficiency (162).

However, no abnormal splicing was observed in FRDA cells with transcripts produced from the intact FXN gene (154, 158).

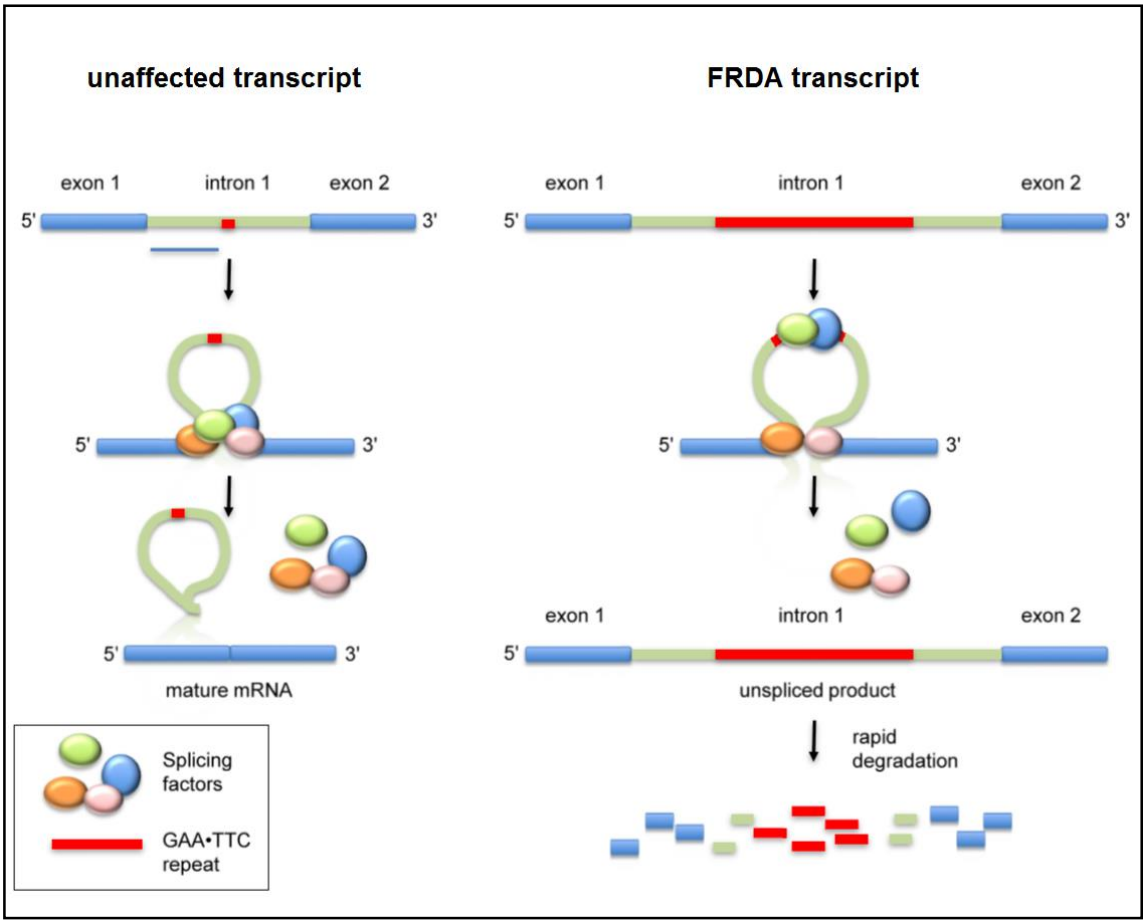



**Figure 1.26:** Altered splicing model for FRDA (162). On unaffected frataxin (FXN) alleles the GAA.TTC repeat is not sufficient to affect splicing. However, once the repeat number exceeds 90, splicing factors that are normally involved in the proper splicing of the FXN gene become mislocalised such that normal splicing is prevented. This could be the result of binding of splicing factors to the repeat, preventing their normal assembly at the splice junctions. This could lead to the failure to remove intron 1. Taken from reference (159).

Thirdly, the FXN GAA.TTC repeat expansions can also cause epigenetic modifications (Figure 1.27). These epigenetic changes include increased DNA methylation, reduction of histone H3 and H4 acetylation levels, chromatin remodelling and noncoding RNAs, which affect expression of the FXN gene in a number of different ways, but do not cause changes in DNA sequence (163-165). This effect may occur near to the start of transcription mediated by chromatin changes on the promoter. Recently, it has been shown that the region flanking the repeats in the FXN gene is enriched for epigenetic marks characteristic of transcriptionally repressed regions of the genome. For example,

CTCF binding to the promoter region is necessary in FXN expression. It has been shown that CTCF is involved in a variety of transcriptional regulatory functions, including transcriptional activation, transcriptional repression and genomic imprinting (166). Furthermore, CTCF has a significant role in inducing RNA polymerase II mediated alternative splicing (167). Therefore, loss of this factor in patient cells could lead to reduced rates of transcription (168). Since DNA methylation in mammalian cells can inhibit transcription elongation, an effect on transcription through the intron is also possible (169). However, the exact mechanism by which the GAA repeat in intron 1 leads to the eventual loss of frataxin protein is still not fully understood. Identifying the mechanism responsible for the FXN mRNA deficit in FRDA is important for the development of treatments for this disorder (159).

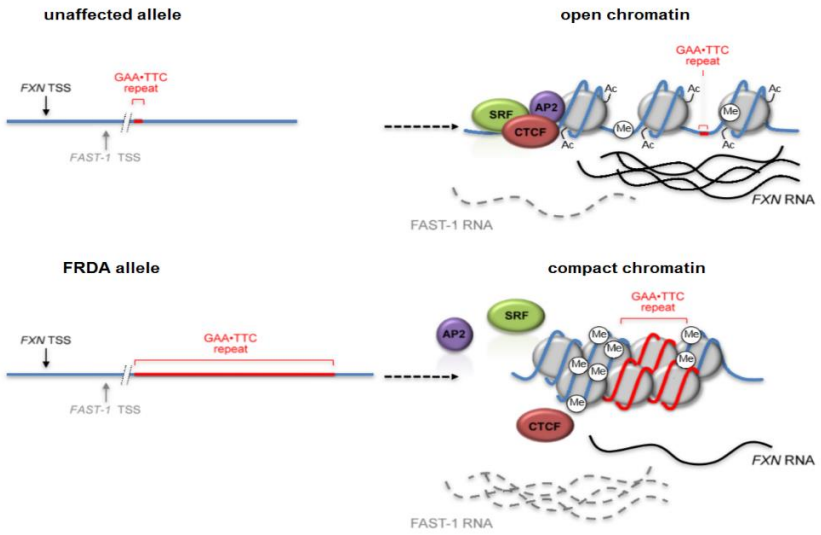



**Figure 1.27:** The FXN chromatin organisation in normal individuals and FRDA patients. Normal alleles contain open chromatin with acetylated histones (Ac) and methylated histones (Me) in the region flanking the repeat. Transcription factors including serum response factor (SRF), activator protein 2 (AP2) and CCCTC-binding factor (CTCF) associate with the 5' end of the gene. Non B-DNA structures, as a consequence of expanded GAA repeats, may induce histone deacetylase activity that along with histone methyltransferases generating more extensive methylated histones. These modifications lead to formation of a compact chromatin configuration and subsequently reduce binding of transcription factors and both FXN transcription initiation and elongation are reduced. Loss of CTCF binding is correlated with an increase in the amount of FXN antisense transcript-1 (FAST-1) RNA that is transcribed antisense to FXN, but how this relates to silencing is unclear. TSS: transcription start site. Taken from reference (159).

### 1.5.4 Therapeutic strategies for FRDA by targeting the GAA.TTC repeats

Now it is evident that expanded GAA.TTC repeats are involved in Friedreich's ataxia disease etiology and these repeats adopt triple helical structures that impede transcription and finally reduce levels of frataxin protein in patients (4). Inhibition of H-DNA formation by using molecules that bind the single-stranded GAA sequence during transcription in the FXN gene alleviates the transcription deficiency of the FXN gene (161). This could be one of the possible approaches to prevent or treat FRDA. It was suggested that short oligonucleotides like $(GAA)_7$ can block particular types of triplex formation and increase transcription of $(GAA.TTC)_{88}$ *in vitro* when used at high concentration (1μM). However, these oligonucleotides penetrate cells *in vivo* with poor efficiency, thus TFOs cannot provide an effective therapeutic approach (161).

Burnett *et al.* in 2006 (170) designed sequence-specific polyamides to prevent the intramolecular sticky DNA structure formed by long GAA.TTC repeats. It is known that pyrrole-imidazole polyamides are small molecules and penetrate living cells efficiently, therefore a class of these molecules, called "linear β-alanine-linked polyamides" was designed to target purine tracts (adenines/guanines) in DNA such as GAGAA.TTCTC repeats (171, 172). Linear β-alanine-linked polyamides have been shown to alleviate the transcription inhibition caused by long GAA.TTC repeats by binding to the duplex DNA in the frataxin gene thereby preventing the formation of sticky DNA. The capacity of these molecules to disrupt sticky DNA formation was also examined. As shown in Figure 1.28, after digestion of the supercoiled sticky DNA (structure 2) with a unique restriction enzyme, the sticky DNA was seen by agarose gel electrophoresis as a cleaved sticky retarded band (structure 1). Linear DNA (structure 3) is indicative of disruption of the sticky-DNA structure by a polyamide. Briefly, plasmid harbouring two tracts of $(GAA.TTC)_{176}$ sequence in direct orientation was incubated with and without polyamide at 37 °C for one hour followed by digestion with a single restriction enzyme. The products were then electrophoresed on 1% agarose gel to determine the presence of the cleaved sticky DNA band. The results revealed that sticky DNA conformation was significantly affected by polyamide molecules which dissociate the associated DNA.DNA region (170). In addition, polyamides may relieve heterochromatin-mediated repression by opening the chromatin domain containing the frataxin gene (172).

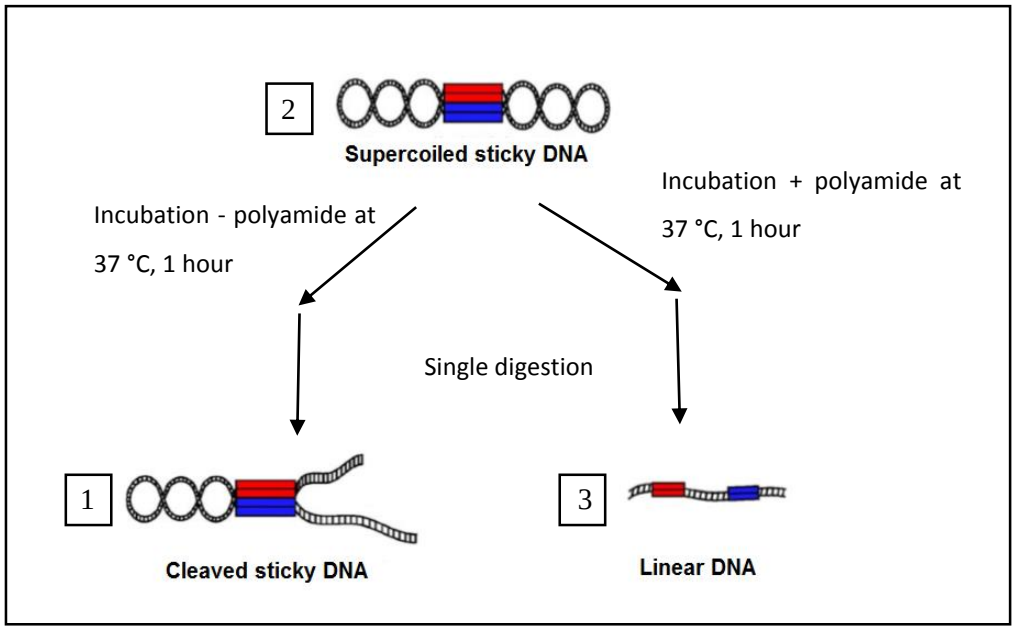



**Figure 1.28:** Effect of polyamide binding to plasmid DNA on sticky DNA stability. Adapted from reference (170).

Another different therapeutic strategy focused on regulation of the FXN gene expression at the epigenetic level. It is recognised that acetylation of histones bound to DNA is essential for transcription and hypoacetylation of histones H3 and H4 may cause FXN silencing in FRDA patients carrying expanded GAA trinucleotide repeats. Thus, histone deacetylase (HDAC) inhibitors have been developed as a new therapy for FRDA (165, 173). Several investigations have shown the potential therapeutic effects of commercial HDACi drugs in reversing heterochromatin and reactivating the silenced FXN gene in FRDA (20, 165, 174). These studies confirmed that the acetylation status of the histones has a significant effect on transcription at the FXN locus. Furthermore, these compounds have the ability to increase FXN gene expression in both pathogenic and non-pathogenic alleles, therefore HDAC inhibitors are considered as a general positive inducer on the expression of the FXN gene.

## 1.6 Methods to detect H-DNA and sticky DNA

Non-B DNA structures such as H-DNA and sticky DNA can be detected by similar techniques. These abnormal DNA structures contain single-stranded regions that can serve as substrates for single-strand specific chemicals and enzymes providing fine structural analysis of these regions. This section will briefly describe the principles of these techniques:

### 1.6.1 Chemical methods

(I) Bromoacetaldehyde (BAA) and Chloroacetaldehyde (CAA)

BAA and its less potent analogue CAA, react at the N-3 and N4 positions of unpaired cytosines and at the N-1 and N-6 positions of unpaired adenines (175-177). In addition, the N-1 and N-2 positions of guanine residues can react with these chemicals but are less sensitive (178). DNA can be cleaved at the modified residues with piperidine treatment after incubation of the DNA sample with hydrazine or formic acid. Cleavage can also be achieved by treating the modified DNA with S1 nuclease and the modification sites can then be identified in a sequence gel (43, 179).

(II) Osmium Tetroxide ($OsO_4$)

$OsO_4$ can be used as a site specific probe for single-stranded pyrimidines, where it reacts strongly with thymines and, to a lesser extent, with cytosines at the C5,C6 double bond forming C5 and C6 osmate ester bonds with $OsO_4$. In the presence of a ligand (pyridine), $OsO_4$ forms stable complexes with pyrimidine bases with greatly increased formation of esters. However, in the absence of a ligand the reaction products are different than those with a ligand (122, 123). Modifications require a second step such as piperidine cleavage of the modified base to convert the modification site into a break and sequencing according to Maxam and Gilbert (122, 180). Furthermore, $OsO_4$ and pyridine can penetrate the cells, thus it is possible to use them in probing the triplex structure in *E.coli* cells (124, 181). $OsO_4$ probes of DNA structure have a number of advantages as compared to other chemical probes; among these is that it is more compatible with other functional groups than $KMnO_4$. For example, $OsO_4$ can catalyses the oxidation of alkynes and alcohols but these reactions have not found wide synthetic applications because of the availability of other methods. In addition, antibodies are available against osmium-modified DNAs (182).

(III) Dimethylsulfate (DMS)

In duplex DNA, DMS methylates predominantly the N7 of guanine and is less reactive to N3 of adenine which are not involved in Watson-Crick hydrogen bonding. In the triplex conformation, DMS is unable to attack the N7 of guanine since it is involved in Hoogsteen base pairing. Therefore, the DMS technique can be employed to determine those guanines that are participating in the Hoogsteen-hydrogen bonding of the triplex structure. Phosphodiester bond at the site of methylation can also be broken by heating with piperidine to determine DMS-modified bases. These broken strands can then be identified by gel electrophoresis (183-186).

(IV) Diethylpyrocarbonate (DEPC)

DEPC reacts efficiently with N7 position of unpaired purines (especially adenines) or when the purine base exists in the *syn* configuration. DEPC has relative hyporeactivity toward duplex DNA and hyperreactivity to purines in single-stranded DNA. This specificity is the reason why DEPC is widely used in structural studies of triplexes (93, 184, 187-189). The modified bases are also identified by chemical cleavage with piperidine.

(V) Potassium permanganate ($KMnO_4$)

$KMnO_4$ is one of the most effective and versatile reagents used in detection regions of B-DNA that have undergone a Watson-Crick base paired disruption. $KMnO_4$ attacks the C5-C6 double bond of pyrimidines, especially thymines in single-stranded DNA (190-192). In H-DNAs, $KMnO_4$ is used to map single-stranded regions in a sequencing gel after hot piperidine treatment (193-195).

DEPC and $KMnO_4$ reagents that were employed in this thesis to probe H-DNA are fully described in Chapter 4.

### 1.6.2 Enzymatic methods

Although more than 30 single-strand specific nucleases from various sources have been isolated, only a few enzymes such as S1 nuclease from *Aspergillus oryzae* and P1 nuclease from *Penicillium citrinum* have been characterised to a significant extent (196). S1 and P1 nucleases have the ability to act selectively on single-stranded regions in

double-stranded nucleic acids and can be used to identify non-B DNA structures in one step, but this type of reaction usually requires more stringent conditions. For example, the activity of S1 nuclease is very specific to single-stranded DNA and has been widely used in the discovery of unusual DNA structures including H-DNA (184, 197, 198), cruciform and hairpins (199, 200), quadruplex (201) and Z-DNA (202, 203), but it works under acidic conditions (pH 4-5) and requires the presence of $Zn^{2+}$ for activity. Working with S1 nuclease in acidic media is disadvantageous since lower pH values lead to considerable depurination of DNA. P1 nuclease is less sensitive than S1 nuclease, but it can work under neutral conditions (204). It can also probe for single-stranded regions of H-DNA (205-207). The S1 nuclease that was used in this thesis is described in detail in Chapter 4.

## 1.7 Aims

It is known that the expanded GAA.TTC repeat sequence associated with FRDA adopts non-B DNA structures (triplexes and sticky DNA) and that sticky DNA requires two GAA.TTC repeats, each at least 59 repeats long, in the same direct orientation within the same plasmid. This work aims to probe the structure and properties of these GAA.TTC trinucleotide repeat sequences. A series of plasmids have been generated containing different length of the GAA.TTC repeats in two regions of pUC18, either singly or in combination, and these have been used to explore the formation of unusual DNA structures. Previous studies suggest that sticky DNA is a bi-triplex, thus chemical and enzymatic probes diethylpyrocarbonate, potassium permanganate and S1 nuclease were used to examine intramolecular triplex formation in supercoiled and linear DNA fragments. The effects of GAA.TTC repeats on the structural behaviour of the plasmids were also examined by using different gel electrophoresis experiments. Finally, the effect of these repeats on gene expression was assessed using luciferase reporter assay

# 2 Materials and Methods

## 2.1 Materials

### 2.1.1 Oligonucleotides

The oligonucleotides that were used to generate expanded GAA.TTC repeats to form triplex and sticky DNA are listed below (Figure 2.1). These were synthesised by Professor Tom Brown (Department of Chemistry, Southampton University). They were used as supplied at stock concentrations 31.02 µM for GAA oligo and 52.51 µM for TTC oligo and stored at -20 ºC until required.

i) 5´- pAAGAAGAAGAAGAAGAAGAAGAAGAAG -3´ (30 mer)

ii) 5´- pTTCTTCTTCTTCTTCTTCTTCTTCTTC -3´ (30 mer)

**Figure 2.1:** Sequence of the oligonucleotides that were used to prepare the expanded GAA.TTC repeats. These oligos contain a 5´-phosphate.

### 2.1.2 Chemicals and Enzymes

The majority of the enzymes used in this thesis were purchased from Promega Corporation, UK. T4 DNA ligase, GoTaq polymerase and Mung Bean Nuclease were purchased from New England Biolabs. Luciferase reporter vector (pGL3-control) and the dual luciferase assay system kit were purchased from Promega. Tryptone, carbenicillin, IPTG and X-gal were purchased from Melford Labs. Yeast extract and blood agar from Difco (Becton, Dickinson &Co.). Agarose and Ficoll were from Sigma-Aldrich. Sodium chloride and ethanol were from Fisher Scientific. GelRed$^{TM}$ (nucleic acid gel stain) from Biotium. SURE 2 supercompetent cells were purchased from Agilent Technologies. Wheat Germ Topoisomerse I was purchased from Inspiralis. The QIAquick gel extraction kit and QIAprep spin miniprep kit were obtained from Qiagen (Crawley, UK). [α-$^{32}$P]dATP used to label the 3´-end of DNA fragments was purchased from Perkin Elmer, at a concentration of 3,000 Ci/mmol and stored at 4 °C. AMV reverse transcriptase was purchased from Sigma-Aldrich (Poole, UK) and stored at -20 °C. Klenow Fragment (3´→ 5´ exo-) was purchased from New England Biolabs. Both Accuagel (40%

acrylamide;bisacrylamide 19:1) and Sequagel (25% acrylamide;bisacrylamide 19:1 containing 8M urea) were purchased from National Diagnostics (Hull, UK).

## 2.2 Protocols

### 2.2.1 Preparation of *E.coli* competent cells

In this project, two types of *E. coli* host strains were used in recombinant DNA experiments. The first cells were *E. coli* TG2 cells which were prepared by picking a single colony of these cells from a fresh agar plate and growing the cells overnight at 37ºC in sterile Luria Broth (LB) media (16 g tryptone, 10 g yeast extract and 5 g sodium chloride per litre). 1 ml of this culture was then transferred to 100 ml of LB media and grown until an optimal density of 0.6-0.8 at 600 nm was reached. The cells were then pelleted by centrifugation at 3000 rpm for 10 minutes at 4 ºC in a sterile 25 ml tube. The media was then decanted from the cells pellet, which was then resuspended in 20 ml of sterile transformation buffer (50 mM calcium chloride, 10 mM Tris-HCl pH 7.5). The resuspended cells were then placed on ice for 30 minutes after which this suspension was again centrifuged and finally resuspended in 5 ml of transformation buffer and stored at 4 ºC for up to two weeks.

The second type of *E. coli* strain was SURE supercompetent cells which were purchased from Stratagene, UK and were stored at -80 °C until required. These cells are defective in several DNA repair pathways and are used for cloning unstable DNA. Since the *recA* gene is defective in these cells, this results in an increased stability of DNA clones that contain long repeats.

### 2.2.2 Cloning

#### 2.2.2.1 Cloning GAA repeats into the pUC18 vector

In this study, the multiple cloning site (MCS) of vector pUC18 (Figure 2.2) was used for the first cloning procedure. Two different strategies were used to insert GAA repeats into the vector.

(A) Strategy 1:

The complementary oligonucleotide strands were annealed by mixing equimolar amounts (30 µg/ml) in the presence of NaCl (100 mM) and were incubated at 100 °C for 2 minutes.

The mixture was then cooled slowly to room temperature and oligomers of the sequence were prepared by ligation by adding 2 µl 10x ligase buffer followed by 1 µl (400 units) T4 DNA ligase (Figures 2.3 and 2.4).

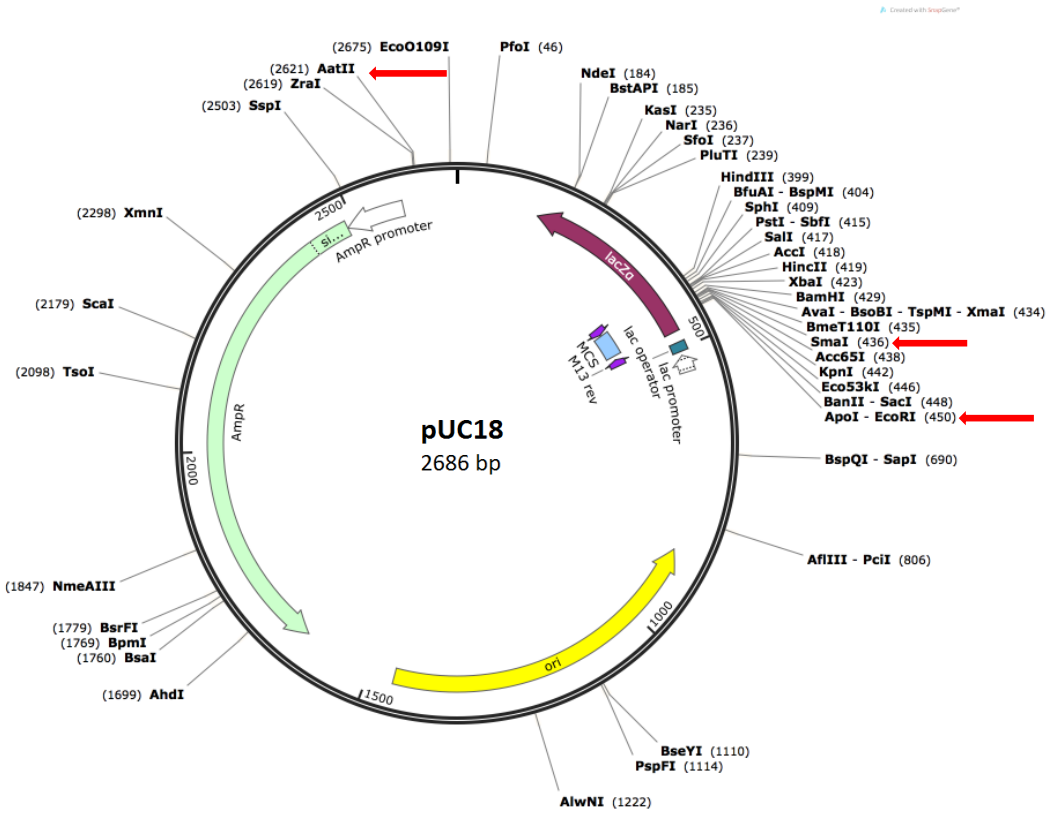



**Figure 2.2:** pUC18 cloning vector. EcoRI, SmaI and AatII cloning sites are highlighted with red arrows. Taken from SnapGene.

**A)**

5´-TTCTTCTTCTTCTTCTTCTTCTTCTTC-3´
3´-GAAGAAGAAGAAGAAGAAGAAGAAGAA-5´

**B)**

5´-TTCTTCTTCTTCTTCTTCTTCTTCTTC-3´
3´-GAAGAAGAAGAAGAAGAAGAAGAAGAA-5´

**Figure 2.3:** Two possible forms of duplex GAA.TTC repeat would be generated by ligation. Duplex (A) that contains an AA overhang at one end and a TT overhang at the other. The oligos can also anneal, staggered the other way with a C overhang at one end and a G overhang at the other as in (B).

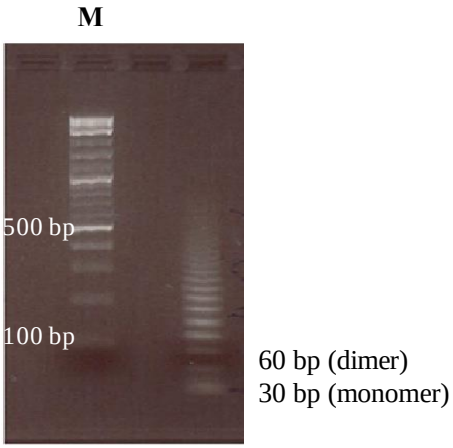



**Figure 2.4:** 1% agarose gel showing different lengths of the duplex DNA (monomer, dimer and multimers) which were generated by self-ligation of the annealed oligonucleotides (right lane). Left lane (M) represents 100 bp DNA ladder (marker).

Two different methods were attempted for directly cloning this ligated GAA.TTC repeat sequence.

(I) Method 1:

1 µg of pUC18 vector was digested with EcoRI, followed by filling in the recessed EcoRI 3′ termini with 5 µM dATP using Go Taq DNA polymerase (1 unit) and 1x Taq reaction buffer. This should generate a vector with 5'-AA overhangs, to complement the TT overhang at one end of the ligated oligonucleotides. Multimers of the ligated GAA.TTC repeat oligonucleotide, which contain a TT overhang were attempted to insert in the vector using T4 DNA ligase (400 units) and 1x ligase buffer, followed by addition 5 µM of dTTP and 1 unit of Go Taq DNA polymerase. The blunt ended fragments were ligated by T4 DNA ligase (Figure 2.5).

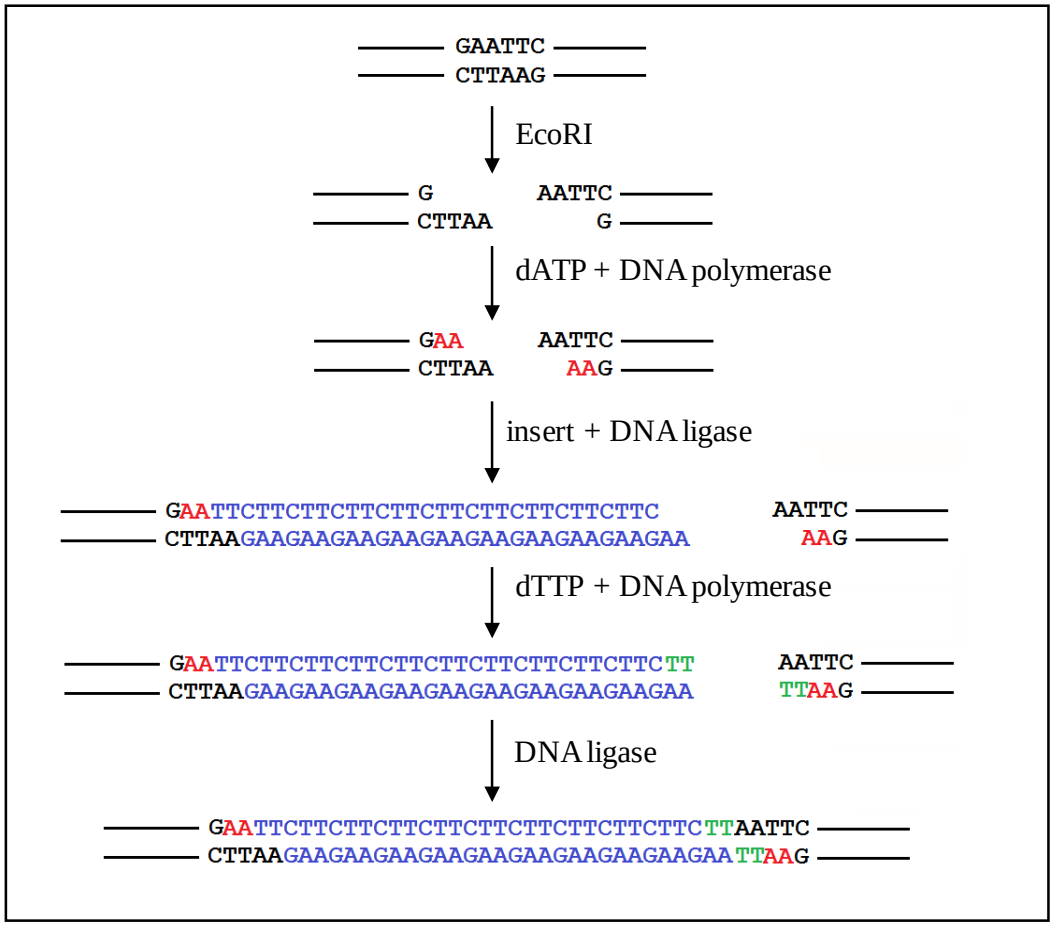



**Figure 2.5:** Model of the possible ligation of the duplex GAA.TTC repeat into EcoRI site of pUC18. The bases in red indicate extra nucleotides inserted into EcoRI region which were needed to complete the ligation.

(II) Method 2:

A mixture of ligated DNA oligonucleotides was blunt ended by treating with mung bean nuclease and ligated into the SmaI site of pUC18 (Figure 2.6A). For this the oligonucleotide ligation ladder was first precipitated by adding 3 volumes of absolute alcohol and left on dry ice for 10 minutes and then centrifuged for 10 minutes at 13,000 rpm. The supernatant was removed and the pellet was dried under a vacuum in a speed vacuum. The pellet was then dissolved in 17µl distilled water, 2µl of 10x mung bean nuclease reaction buffer and then 10 units of mung bean nuclease was added. This mixture was then incubated at 30 °C for one hour to produce blunt ends. This mix was then precipitated by the addition of 3 volumes ethanol, centrifuged and dried as previously described. The pellet was then redissolved in 10µl of distilled water. Cloning vector pUC18 was digested with 10 units of SmaI and 2µl of 10x buffer J in a total reaction volume of 20µl for 3 hours at room temperature. The blunt ended DNA fragments (10µl)

and digested pUC18 (2µl) were mixed and ligated with T4 ligase (1µl of 400 units/ml) and 10x ligase buffer (2µl) were added and incubated at room temperature overnight. The resultant plasmids were transformed into *E. coli* TG2 cells which were then spread on an agar plate containing carbenicillin and grown overnight at 37 °C. Colonies containing the transformed cells were detected by blue-white selection on agar plates containing 100 µg/ml carbenicillin, 1mM IPTG and 5 mM X-gal; this allowed selection between recombinant plasmids and non-recombinant plasmids. Transformed cells containing an insert appear as white colonies whereas colonies which contain unmodified pUC plasmid were blue (this will be further explained in Chapter 3, Section 3.1). Colonies containing the successful clones were grown in broth media and plasmids were isolated by QIAprep spin miniprep kit. More details of the transformation and purification processes are presented in the next sections.

(B) Strategy 2:

The GAA.TTC repeats were extended and amplified by PCR (Figure 2.6B). Since this is a highly repetitive sequence, the two oligos do not necessarily exactly pair up during the annealing step and the overhangs were then filled in by polymerase, ready for a similar reaction in the next round of PCR. The reaction compositions and PCR program used are as shown in Tables 2.1 and 2.2, respectively.

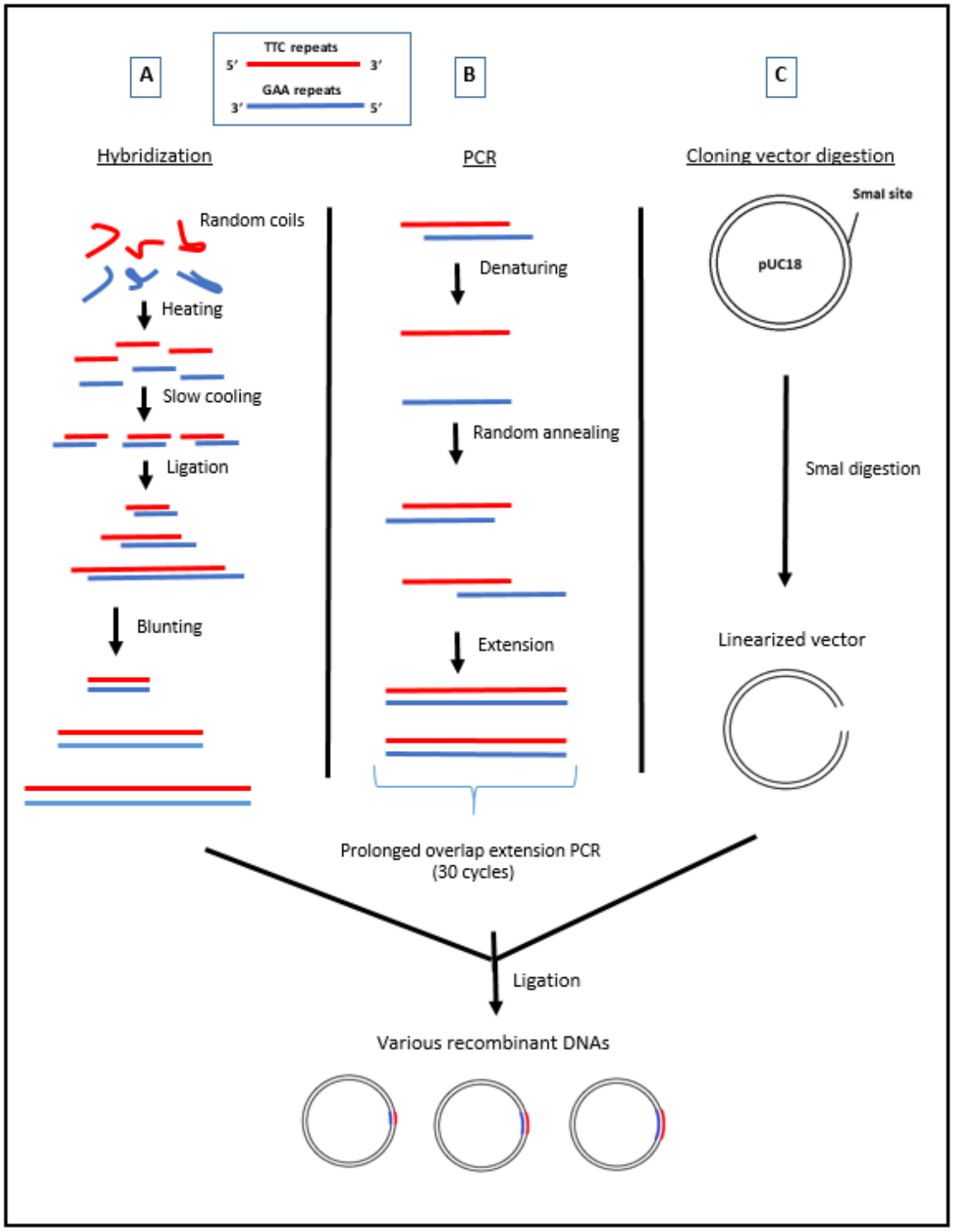



**Figure 2.6:** A scheme showing two methods (A and B) used for cloning various lengths of GAA.TTC repeats into SmaI site of the cloning vector pUC18 (C).

**Table 2.1:** Composition of the PCR reaction sample

| | | |
|---|---|---|
| Oligos (31µM): | 0.5 µl | (5′-AAGAAGAAGAAGAAGAAGAAGAAGAAG-3´) |
| Oligos (31µM): | 0.5 µl | (5´-TTCTTCTTCTTCTTCTTCTTCTTCTTC-3´) |
| dNTPs (10 mM each): | 1 µl | |
| DMSO: | 1.5 µl | |
| Reaction buffer 5x: | 10 µl | |
| Taq polymerase (500 units/ml): | 0.3 µl | |
| $ddH_2O$ (to make up volume to 50 µl) | | |

The PCR product was precipitated by the addition 3 volumes ethanol, centrifuged and dried. The blunt ended DNA fragments were then ligated into the SmaI site of pUC18. The recombinant plasmid was transformed and purified as described in sections 2.2.3 and 2.2.4.

**Table 2.2:** Indicated below are the cycling conditions and PCR programs for cloning GAA repeats into four different sites in two different vectors. All programs were run for 30 cycles in stage 2. Hold is the final step of PCR until samples taken from the PCR machine.

| | pUC18 | | pGL3-Control | |
|---|---|---|---|---|
| **Steps** | SmaI site | AatII site | SacI site | PciI site |
| **Step 1** | | | | |
| Initial Denaturation | 98 °C<br>5 minutes | 98 °C<br>5 minutes | 98 °C<br>5 minutes | 98 °C<br>5 minutes |
| **Step 2** | | | | |
| Denaturation | 98 °C<br>30 seconds | 98 °C<br>30 seconds | 98 °C<br>30 seconds | 98 °C<br>30 seconds |
| Annealing | 50 °C<br>30 seconds | 64 °C<br>30 seconds | 52 °C<br>30 seconds | 64 °C<br>30 seconds |
| Extension | 72 °C<br>1 minute | 72 °C<br>1 minute | 72 °C<br>1 minute | 72 °C<br>1 minute |
| **Step 3** | | | | |
| Final Extension | 72 °C<br>7 minutes | 72 °C<br>7 minutes | 72 °C<br>7 minutes | 72 °C<br>7 minutes |
| **Step 4** | | | | |
| Hold | 4 °C | 4 °C | 4 °C | 4 °C |

### 2.2.2.2 Subcloning of GAA repeats from the multiple cloning site into the AatII site of pUC18

To study the properties of sticky DNA in pUC18, the AatII site (5′-GACGT/C-3’) was selected for inserting a second set of GAA repeats into the pUC18 plasmids containing GAA repeats in the SmaI site. In this study, most primers which were used to subclone the insert from one position to another, have an extra few bases as an overhang at the 5'-end of one or both of the amplification primers to introduce a new restriction site which was cleaved off with the enzyme as shown in Figure 2.7.

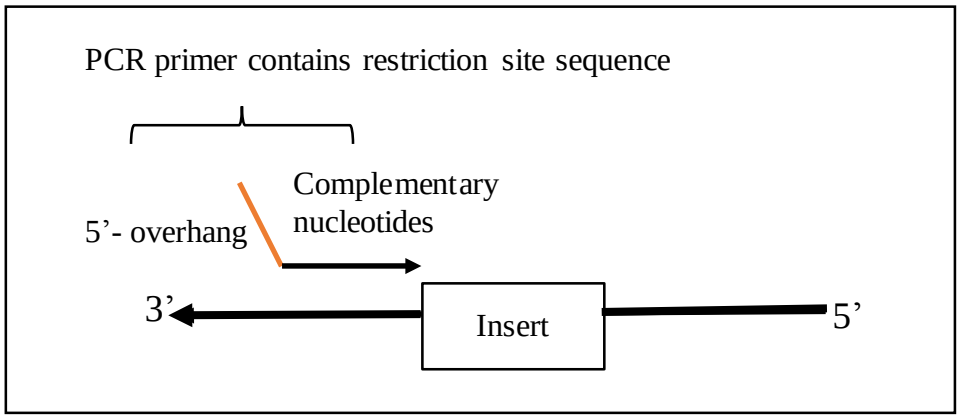



**Figure 2.7:** Incorporation of extra sequence at the 5’-end of the primer (indicated with the red line) in order to introduce new restriction site sequences into PCR products.

To subclone GAA repeats into the AatII site of pUC18 clones, a pair of primers (forward and reverse) containing an AatII recognition site was designed, as shown in Figure 2.8, along with the DNA template sequence. The PCR mixture, including the DNA template, dNTPs, DMSO, primers, buffer and Taq polymerase, and the volumes of each component are shown in Table 2.3. The PCR conditions used for amplification reaction are shown in Table 2.2. After running the PCR, 10 µl of PCR products was transferred to an Eppendorf tube. 10 units of restriction enzyme AatII and 2 µl buffer were added in a total volume of 20 µl and incubated at 37 °C for 3 hours to produce sticky ends that could be ligated into the AatII site of the pUC18 clones. The mixture was then precipitated with ethanol as previously described. The AatII-AatII fragments containing (GAA) repeats were ligated into AatII site of pUC18 clones by T4 DNA ligase. The resultant plasmids were then transformed into SURE competent cells and plated out on carbencillin-containing agar plates.

**Flanking sequence of SmaI site of vector pUC18 containing the insert:**

5'-CTCTAGAGGATCCCC**(GAA)**$_n$GGGTACCGAGCTCGAATTC-3'

**AatII site in vector pUC18:**

**Forward:** 5′-GGCTGCGACGTCTAGAGGATCCCCGAAGAAG-3′

**Reverse:** 5′-GGCAGCGACGTCGAATTCGAGCTCGGTACCCTTC-3′

**SacI site in vector pGL3-Control:**

**Forward:** 5'- GGCTCGCGAGCTCTAGAGGATCCCCGAAGAAG-3'

**Reverse:** 5'-GAATTCGAGCTCGGTACCCTTC-3'

**PciI site in vector pGL3-Control:**

**Forward:** 5′-GCTGCGACATGTCGACTCTAGAGGATCCCCGAAG-3′

**Reverse:** 5′-CCAGCACATGTACGAATTCGAGCTCGGTACCCTTC-3′

**Figure 2.8:** Sequences of primers used for subcloning GAA repeats into three different sites in two different vectors. The red bases indicate an extra sequence at the 5' end of the primer to form recognition site of the restriction enzyme. The underlying bases indicate the restriction sites of the enzymes.

**Table 2.3:** Composition of the PCR reaction sample

| | |
|---|---|
| Plasmid DNA 100 ng/µl: | 0.5 µl |
| Primer (50µM each): | 0.3 µl |
| dNTPs (10 mM each): | 1 µl |
| DMSO: | 1.5 µl |
| GoTaq® Reaction Buffer (5x): | 10 µl |
| GoTaq® DNA Polymerase (5 units/µl): | 0.3 µl |
| dd$H_2O$ (to make up volume to 50 µl) | |

#### 2.2.2.3 Subcloning of GAA repeats into the SacI site of pGL3-Control vector

To examine the effect of GAA.TTC repeat tracts (either one or two in number and of varying lengths) on gene expression, the GAA.TTC repeats from the MCS of pUC18 were inserted into the SacI site of pGL3-Control (Figure 2.9). The repeat sequences from the pUC18 clone were amplified by PCR, using primers that contained SacI restriction sites for insertion into the pGL3-control vector. The primers and DNA template sequences are shown in (Figure 2.8). The composition of the PCR mixture and PCR program are presented in Tables 2.2 and 2.3, respectively. 10 µl of PCR products were transferred to an Eppendorf tube, 10 units of SacI and 2 µl buffer J were then added in a total volume of 20 µl and incubated at 37 °C for 2 hours to produce sticky ends that could be ligated into SacI site of pGL3 vector. The SacI-SacI DNA fragments containing (GAA) repeats were ligated into the SacI site of pGL3-Control vector and the resultant plasmids were then transformed into SURE competent cells.

#### 2.2.2.4 Subcloning of GAA repeats into PciI site of pGL3-Control Vector

Subcloning the GAA repeats from the MCS of pUC18 into the PciI site of pGL3-Control vector (Figure 2.9) was very similar to cloning described above. Briefly, a pair of primers containing a PciI recognition site was designed as shown in Figure 2.8. The composition of the PCR mixture and the PCR program are also presented in Tables 2.2 and 2.3, respectively. The PCR products were digested with PciI to produce sticky ends that could be ligated into the PciI site of the pGL3 clone and the empty pGL3 vector. The fragments containing (GAA) repeats were ligated into the PciI site and the resultant plasmids were then transformed into SURE competent cells.

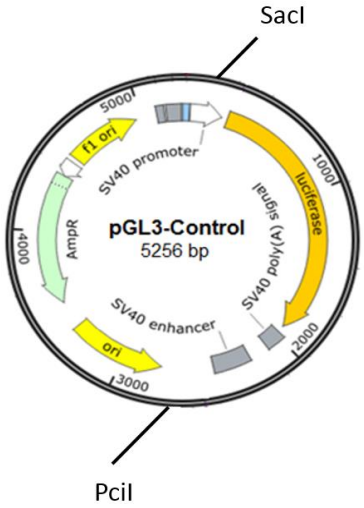



**Figure 2.9:** pGL3-Control vector showing positions of SacI and PciI cloning sites. Adapted from SnapGene.

### 2.2.3 Transformation

5 µl of the ligation mixture was mixed with 100 µl of TG2 competent cells in a 1.5 ml micro-centrifuge tube and placed on ice for at least 30 minutes. This was then heat shocked at 45 ºC for 1 minute and placed immediately back on ice for 10 minutes. 500 µl of broth medium was added and then incubated at 37 °C for 1 hour. The mixture was then spread on agar plates containing carbencillin, which were incubated overnight at 37 ºC in an inverted position and then stored at 4 ºC until required. For transformation into SURE cells, 1 µl of the ligation mixture was mixed with 20 µl of SURE competent cells in a sterile 1.5 ml micro-centrifuge tube, kept on ice then heat shocked, replaced on ice, and broth medium was added as described previously. The cells were then transferred to agar plates containing carbencillin and incubated overnight at 37 ºC.

### 2.2.4 Plasmid Purification

Colonies were picked from the agar plates using a sterile metal loop and grown overnight in 5 ml of LB media containing 100 µg/ml carbencillin. Each culture was split into two 1.5 ml Eppendorf tubes and centrifuged at 3500 rpm for five minutes to pellet the cells. The supernatant was discarded and the plasmids were purified using a Qiagen QIAprep kit. This involved resuspending the cells in 250 µl of buffer P1 containing a ribonuclease that degrades any RNA that would affect the purity of plasmid DNA (208). The suspension was transferred to a sterile 1.5 ml Eppendorf tube. Cells were then lysed by

alkaline lysis with addition of 250 µl of buffer P2 containing sodium hydroxide. The suspension was mixed by inversion. After 5 minutes, the suspension was neutralised with 350 µl of buffer N3 containing guanidine hydrochloride and acetic acid. The solution was then centrifuged at 13000 rpm for 10 minutes and the supernatant then transferred to a Qiagen spin column and centrifuged three times for 60 seconds at 13000 rpm, discarding the flow through each time. Nucleases were removed by addition of 500 µl of buffer PB containing high concentration of guanidine hydrochloride and isopropanol. The spin column was centrifuged for 60 seconds and the supernatant removed. Plasmid DNA was desalted by washing with 750 µl of buffer PE containing ethanol, centrifuged for 60 seconds and the flow through was discarded. The remaining PE buffer was removed by centrifugation for an additional 60 seconds. The spin column was placed in a 1.5 ml Eppendorf tube and 50 µl of elution buffer EB was added; the tube was then spun for 1 minute to elute the plasmid. The concentration of the purified plasmids was measured by using Nano Drop spectrophotometer. The plasmids were stored in Buffer EB at -20 ºC until required.

### 2.2.5 Screening for successful clones

#### 2.2.5.1 Detection of inserts cloned into SmaI site of vector pUC18

Blue-white screening was used to check the presence of an insert in the MCS of pUC18. This method allows for rapid and easy identification of successful ligation of cloning experiments through the colour of the bacterial colony. In this method of screening, only white colonies were picked from the agar plate using a sterile tip and grown in 5 ml of LB media containing carbincillin overnight at 37 °C. The plasmid was purified as described in section 2.2.4. Plasmids were then sent to Eurofins MWG Operon for sequencing as shown in the Appendix.

#### 2.2.5.2 Detection of inserts cloned into AatII site of vector pUC18 and SacI site and PciI of vector pGL3-Control

The presence of DNA insert in a cloning vector was determined by screening the *E.coli* colonies using ‘colony PCR’. Two primers were designed to be complementary to sequences flanking the insert region as shown in Figure 2.10. A number of colonies were randomly selected from the plate and a portion of each one transferred to tubes containing the PCR mixture using a sterile pipette tip and stirred to mix. Preparation of the PCR reaction mixture and conditions are shown in Tables 2.4 and 2.5, respectively.

Table 2.6 lists a summary of the clones obtained.

**Flanking sequence of AatII site in vector pUC18 containing the insert:**

5′-CGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAGAGGATCCCC**(insert)**GGGTACCGAGCTGGAATTC
GACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCG-3′

**Primers:**

**Forward:** 5′-CGCACATTTCCCCGAAAAGTGC-3′
**Reverse:** 5′-CGAAAGGGCCTCGTGATACGCC-3′

**Flanking sequence of SacI site in vector pGL3-Control containing the insert:**

5′-CTAGCAAAATAGGCTGTCCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCTATCGATAGGTACCGAGCTCT
AGAGGATCCCC**(insert)**GGGTACCGAGCTCTTACGCGTGCTAGCCCGGGCTCGAGATCTGCGATCTAAGTAAGCTT
GGCATTCCGGTACTGTTGGTAAAGCCACCATGGAAGACGCCAAAAACATAAAG-3′

**Primers:**

**Forward:** 5′-CTAGCAAAATAGGCTGTCCC-3′
**Reverse:** 5′-CTTTATGTTTTTGGCGTCTTCCA-3′

**Flanking sequence of PciI site in vector pGL3-Control containing the insert:**

5′-CACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTCGACTCTAGAG
GATCCCC**(insert)**GGGTACCGAGCTGGAATTCGTACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAA
AGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCG-3′

**Primers:**

**Forward:** 5′-CACTCAAAGGCGGTAATACGG-3′
**Reverse:** 5′-GGCGTTTTTCCATAGGCTCCG-3′

**Figure 2.10:** List of flanking regions of the inserts and primers used for colony PCR. The restriction sites are highlighted in red.

**Table 2.4:** Composition of the PCR reaction sample

| | |
|---|---|
| Primer (50µM each): | 0.3 µl |
| dNTPs (10 mM each): | 1 µl |
| GoTaq® Reaction Buffer (5x): | 10 µl |
| GoTaq® DNA Polymerase (5 units/µl): | 0.3 µl |
| dd$H_2O$ (to make up volume to 50 µl) | |

**Table 2.5:** Shown below are 'colony PCR' reaction conditions. All programs were run for 30 cycles in stage 2. Hold is the final step of PCR until samples are taken from the PCR machine.

| | pUC18 | pGL3-Control | |
|---|---|---|---|
| **Steps** | AatII site | SacI site | PciI site |
| **Step 1** | | | |
| Initial Denaturation | 98 °C<br>10 minutes | 98 °C<br>10 minutes | 98 °C<br>10 minutes |
| **Step 2** | | | |
| Denaturation | 98 °C<br>30 seconds | 98 °C<br>30 seconds | 98 °C<br>30 seconds |
| Annealing | 52 °C<br>30 seconds | 52 °C<br>30 seconds | 50 °C<br>30 seconds |
| Extension | 72 °C<br>1 minute | 72 °C<br>1 minute | 72 °C<br>1 minute |
| **Step 3** | | | |
| Final Extension | 72 °C<br>7 minute | 72 °C<br>7 minute | 72 °C<br>7 minute |
| **Step 4** | | | |
| Hold | 4 °C | 4 °C | 4 °C |

**Table 2.6:** List of the cloned plasmids obtained.

| Plasmid | pUC18 | | pGL3-Control | |
|---|---|---|---|---|
| Location | SmaI | AatII | SacI | PciI |
| Number of (GAA.TTC) repeats | 10 | - | 13 | - |
| | 19 | - | 29 | - |
| | 22 | - | 52 | - |
| | 29 | - | 72 | - |
| | 29 | 29 | 85 | - |
| | 29 | 29 inv | 85 | 85 |
| | - | 28,29 | - | 85 |
| | 48 | 83 | | |
| | 70 | - | | |
| | 85 | - | | |
| | 85 | 62 | | |
| | 85 | 85 | | |
| | 94 | - | | |

### 2.2.6 Probing assays for intramolecular triplex DNA

Previous study suggest that sticky DNA is a combination of two intramolecular triplexes in one closed plasmid (3). Thus, triplex structures formed by various lengths of GAA.TTC repeats were investigated first. Intramolecular triplex structures (H-DNA) contain single strand regions and those were detected by two different techniques. The first experimental method used chemical and enzymatic probes which are known to be appropriate tools for investigating unusual DNA structures; detecting the presence of exposed or single strand nucleotides. The chemical probing method used in this study relies on the ability of some chemical agents to modify specific unpaired DNA bases, followed by piperidine treatment to cleave DNA at the modified sites, while enzymatic probing used a single-strand specific nuclease (S1) to cleave unpaired regions of DNA. The two main chemical probes employed in this thesis (Chapter 4) are diethylpyrocarbonate (DEPC) and potassium permanganate ($KMnO_4$). These reagents are often used to detect DNA distortions, particularly unstacked regions (209-212). DEPC is used to identify exposed purines especially adenines within DNA loops and $KMnO_4$ reacts with unpaired thymines in single-stranded DNA. The second technique used S1 nuclease to locate the single stranded regions that accompany the formation of intramolecular triplexes within supercoiled DNA.

#### 2.2.6.1 Chemical probing assay

##### *2.2.6.1.1 DEPC modification*

DEPC reacts with exposed purines (Adenosine >> Guanosine) in single-stranded DNA but not with the triplex itself, since the N7 position is protected by third strand binding within the DNA major groove, and it has very low reactivity with duplex DNA. The reaction of DEPC with $(GAA.TTC)_n$ inserts was therefore examined, comparing supercoiled with linear DNA. 15 µl of supercoiled $(GAA)_n$-containing plasmids (approximately 150 ng/µl) was incubated with 50 µl of (50 mM sodium acetate buffer pH 5, 5 mM $MgCl_2$) or (40 mM Tris-acetate pH 7, 5 mM $MgCl_2$) at 37 °C for two minutes to allow production of intramolecular triplex structure. 4 µl of DEPC (162.14 g/mol) was added to the tube and then incubated again at 37 °C for 30 minutes, quickly shaken every 10 minutes. The reactions were terminated by precipitation with ethanol, followed by washing with 70 % ethanol and dried in a Speed Vac. The purified DEPC-treated DNA was dissolved in distilled water and double digested with EcoRI and PstI restriction

enzymes to study the GAA strand as shown in Figure 2.11. Restriction digested DNA fragments containing insert were labelled with [α-$^{32}$P]-dATP at the 3'-end as described in section 2.2.6.3. The linearised DNA form was prepared by double digestion of the recombinant plasmid using EcoRI and PstI to release the insert region, before labelling DNA fragments at the 3'-end with α-$^{32}$P-dATP.

#### *2.2.6.1.2 $KMnO_4$ modification*

Potassium permanganate oxidises pyrimidine bases (Thymidines >> Cytidines) with a high preference for those nucleotides that are located within single-stranded DNA. 50 µl of (50 mM sodium acetate buffer pH 5.0, 5 mM $MgCl_2$) or (40 mM Tris-acetate pH 7.0, 5 mM $MgCl_2$) was added to 15 µl of recombinant supercoiled plasmid (approximately 150 ng/µl) and incubated at 37 °C for two minutes to produce an intramolecular triplex structure. 0.3 µl of 100 mM $KMnO_4$ was added to the plasmid solution (to give a final concentration 0.5 mM) and left for 2 minutes at room temperature. The reaction was stopped by adding 0.2 µl of 2-mercaptoethanol followed by ethanol precipitation as described previously. The purified $KMnO_4$-treated DNA was dissolved in distilled water and double digested with HindIII and SacI restriction enzymes to study TTC strand as shown in Figure 2.11. The modified plasmid was then labelled with $^{32}$P at the 3'-end as described in section 2.2.6.3. Linear DNA was generated by double digestion of the recombinant plasmid using Hind III and SacI to release a short segment of DNA containing the insert. The DNA fragments were then 3'-end labelled with $^{32}$P-dATP as described in section 2.2.6.3.

**PstI site** **EcoRI site**

5'-CTGCA/GGTCGACTCTAGAGGATCCCC(GAA)$_n$GGGTACCGAGCTCG/AATTC-3'

3'-G/ACGTCCAGCTGAGATCTCCTAGGGG(CTT)$_n$CCCATGGCTCGAGCTTAA/G-5'

**HindIII site** **SacI site**

5'-A/AGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCC(GAA)$_n$GGGTACCGAGCT/C-3'

3'-TTCGA/ACGTACGGACGTCCAGCTGAGATCTCCTAGGGG(CTT)$_n$CCCATGGC/TCGAG-5'

**Figure 2.11:** Sequences of PstI-EcoRI and HindIII-SacI polylinker DNA fragments containing insert used in probing assays. The red bases indicate the restriction sites of the enzymes.

### 2.2.6.2 Enzymatic probing assay

Intramolecular triplex formation by GAA.TTC repeats was investigated by examining their susceptibility to reaction with S1 nuclease, which should cut the displaced single strand and the loop at the turn of the third strand. 17 µl of recombinant DNA (200-250 ng/µl) was digested with 1 µl (1 unit) S1 nuclease (Promega) in 2 µl 10x S1 buffer in a final volume of 20 µl. These were incubated at 37 °C for 1 minute. The plasmid was precipitated by adding 3 volumes of ethanol as described previously. The pellet was re-dissolved in water (51 µl). 6 µl of the appropriate 10x Promega buffer was added, followed by 1 µl (10 units) of restriction enzymes (EcoRI and PstI or HindII and SacI depending on which strand was being studied) and the tube was incubated for at least 1 hour at 37 °C. Linear DNA was generated by double digestion of the recombinant plasmid using EcoRI and PstI or Hind III and SacI to release a short segment of DNA containing the insert. The DNA fragments were then 3'-labelled as described in section 2.2.6.3. All digested plasmid was also labelled.

### 2.2.6.3 Radiolabelling DNA with α-$^{32}$P-dATP

After digestion with the appropriate restriction enzymes, as described above, the modified DNAs were radiolabelled at the 3'end using 0.3 µl of [α-$^{32}$P]-dATP (3,000 Ci/mmol), 2 units of Klenow fragment DNA polymerase I (3′→ 5′ exo-) (NEB) and appropriate volume of DNA polymerase buffer and then left for 1 hour at 37 °C. In the initial experiments, AMV (avian myeloblastosis virus) reverse transcriptase was used for labelling but Klenow fragment polymerase I proved more effective. End-labelled fragments were run on a 0.3 mm-thick polyacrylamide gel to purify it from the remainder of the plasmid, enzymes and other components. 10 µl of loading dye (20% Ficoll, 10 mM EDTA and 0.005% w/v bromophenol blue) was added and the sample loaded onto a 6.5% native polyacrylamide gel (20 cm long) and run at 400 V for approximately one hour (until the dye had reached the bottom of the gel). After running, the wet gel was exposed to X-ray film for about 10 minutes to locate the position of the labelled fragments. The gel contained 2 bands corresponding to the restriction fragment and the remainder of the plasmid DNA and the required band was cut out and eluted in 500 µl of elution buffer containing 10 mM Tris-HCl pH 7.5 and 1 mM EDTA and left on shaker overnight. The eluted DNA was separated from the gel slice and precipitated by adding approximately 1 ml of ethanol; this was left on dry ice for 20 minutes, then centrifuged at 13000 rpm for 10 minutes. The supernatant was removed and checked with a hand-held Geiger counter

to confirm that it contained very little radioactive DNA. The pellet was washed with 70 % ethanol and dried. The unmodified DNA samples were treated with chemical or enzymatic reagents to modify linear DNA as described in sections 2.2.6.1 and 2.2.6.2.

#### 2.2.6.4 Piperidine cleavage at the modified bases

All the modified DNA samples, except nuclease-treated DNAs, were re-dissolved in 100 µl 10% piperidine and then heated in sealed tubes at 95 °C for 30 minutes; the piperidine was subsequently removed under vacuum. All the resultant pellets, including chemical and enzymatic-treated DNAs, were re-dissolved in 4 µl TE buffer and 4 µl DNase I stop solution (80% formamide, 10 mM EDTA, 10 mM NaOH and 0.1%(w/v) bromophenol blue). These were boiled for 3 minutes and then rapidly cooled on ice, ready to load onto the gel.

### 2.2.7 Denaturing polyacrylamide gel electrophoresis

To identify the uniqueness of the bases that had been modified, 3'-labelled single-stranded DNA fragments were loaded onto a denaturing polyacrylamide gel (40 cm long, 0.3 mm thick) containing 8 M urea. The gel was run at 1400 V until the blue dye reached the bottom of the gel (typically about 90-120 minutes) and then fixed in 10% (v/v) acetic acid. The gels were then transferred to Whatman® 3MM paper and dried under vacuum at 80 °C for 1-2 hours. The dried gels were exposed to a phosphorimager screen for analysis of reactive sites by the Image Quant software (Molecular Dynamics).

### 2.2.8 Preparation of GA tract

The sample of supercoiled DNA was digested by two restriction enzymes (either: EcoRI plus PstI or HindIII and SacI depending on the strand being examined), labelled and extracted as described previously in preparation of linear DNA. 1-10 µl of 3'-end labelled DNA was mixed with 5 µl DNase I stop solution and made up to 25 µl with distilled water. The mixture was heated at 95 °C for 30 minutes with the cap open, and then chilled in ice. The sample was loaded to a denaturing polyacrylamide gel alongside the modified samples.

### 2.2.9 S1 mapping

S1 nuclease mapping is a very useful technique to map any single stranded regions that are generated by intramolecular triplex formation, as shown in Figure 2.12. When supercoiled vector pUC18 (control) or pUC18 containing an insert in the SmaI site are

exposed to S1 nuclease, linearisation will occur if the plasmid contains any S1 sensitive regions. The position of these cleavage sites can then be mapped by further digestion with a second enzyme (ScaI) which cuts once in vector pUC18, and results in the appearance of two fragments. If the plasmids are first linearised by ScaI and then exposed to S1 nuclease, this will remove any supercoil-dependent single-stranded regions and only one band will appear on the agarose gel. The products of digestion were compared with plasmids cut with EcoRI and ScaI each of which has a unique cutting site. Figure 2.12 illustrates a scheme of the expected results.

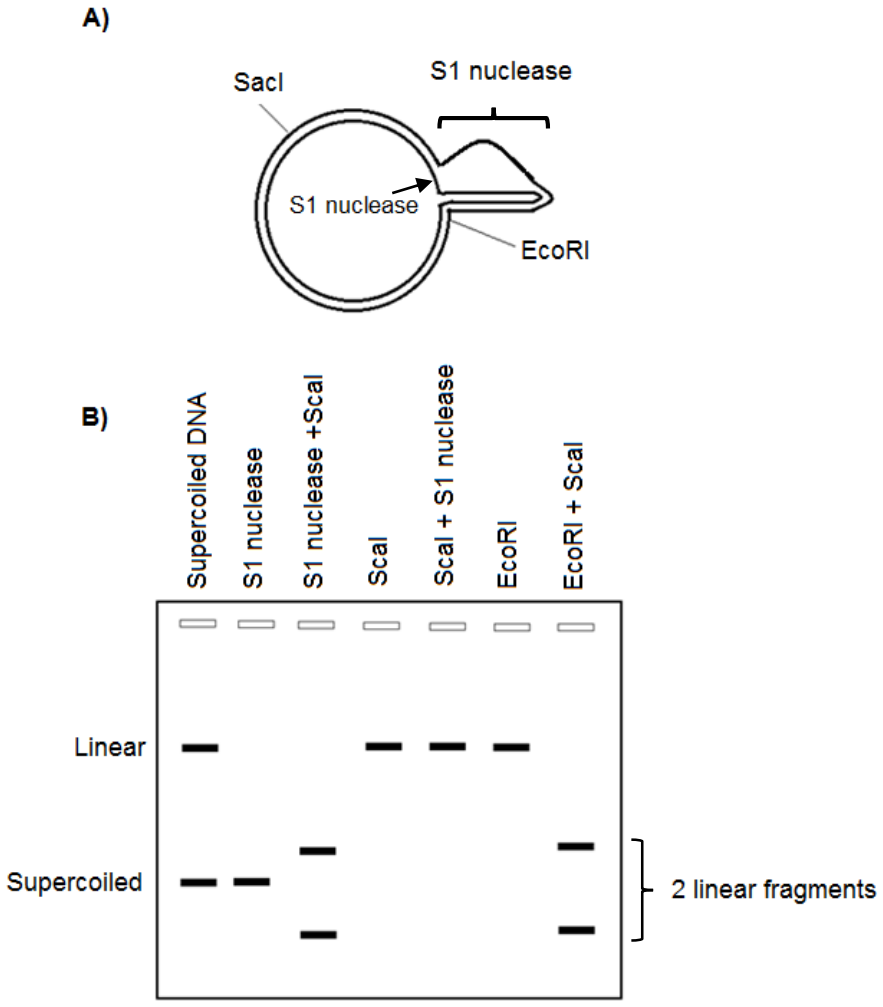



**Figure 2.12:** Diagram illustrates the sensitive regions of S1 nuclease in the intramolecular triplex DNA. **A)** Model of intramolecular triplex formation in closed circular plasmid. Two main single strand regions formed by GAA.TTC repeats (long and short strand). EcoRI restriction site is located in close proximity to the SmaI cloning site, thus EcoRI enzyme was used with ScaI to map the location of the intramolecular triplex. **B)** A scheme represents the map of predicted intramolecular triplex region within the pUC18 cloning vector.

#### 2.2.9.1 Digestion of supercoiled and linear plasmids

A sample of supercoiled plasmid (17 µl, 200-250 ng/µl) was treated with 5 units of S1 nuclease in the presence of 1x S1 nuclease buffer in a final volume of 20 µl then incubated at 37 °C for 5 minutes. DNA was precipitated with ethanol as described previously and re-dissolved in 54 µl of distilled water. The DNA was then cut with 1 µl (10 units) of ScaI restriction enzyme in the presence of 6 µl 10x ScaI buffer for 1 hour at 37 °C.

To linearise the plasmid first, 1 µl (10 units) of ScaI restriction enzyme was added to the supercoiled plasmid in the presence 2µl of 10x ScaI buffer in a total reaction of 20 µl and was then incubated for at least 1 hour at 37 °C. The plasmid was ethanol precipitated as previously described and re-suspended in 17 µl distilled water. The linear DNA was then treated with 5 units of S1 nuclease in the presence of 1x S1 nuclease buffer at 37 °C for 5 minutes.

Plasmids were also cut with EcoRI and ScaI. The plasmids were then run on 1% agarose gel containing GelRed$^{TM}$ (0.1 µl/ml) and visualized under UV light.

### 2.2.10 Relaxation with topoisomerase I

The formation of sticky DNA in plasmids can be observed by its slow mobility in agarose gel electrophoresis. Thus, topoisomerase I was used to relax the supercoiled DNA in order to assess the formation of any gross supercoil-dependent structures. Exposure of the supercoiled plasmid containing sticky DNA to topoisomerase I will yield a DNA.DNA associated region with two closed non-supercoiled loops (Figure 2.13), while treatment of the supercoiled non-sticky DNA with the same enzyme will produce relaxed covalently closed plasmid. To achieve this, 0.5 µl of the supercoiled form of each construct plasmid (100-150 ng/ µl) was treated with 0.2 µl (1 unit) topoisomerase I (Inspiralis) in the presence of 5 µl 2X assay buffer in total reaction 10 µl. This mixture was incubated at 37 °C for 4 hours. Each sample was then divided into two tubes, and 4 µl of Ficoll loading dye was added to each tube. The first DNA sample was then electrophoresed on 0.7% agarose gel containing the fluorescent tag GelRed at 70 volt/cm, whilst the second sample was electrophoresed on agarose gel in the absence of the GelRed. More details are shown in Chapter 3, Section 3.4.

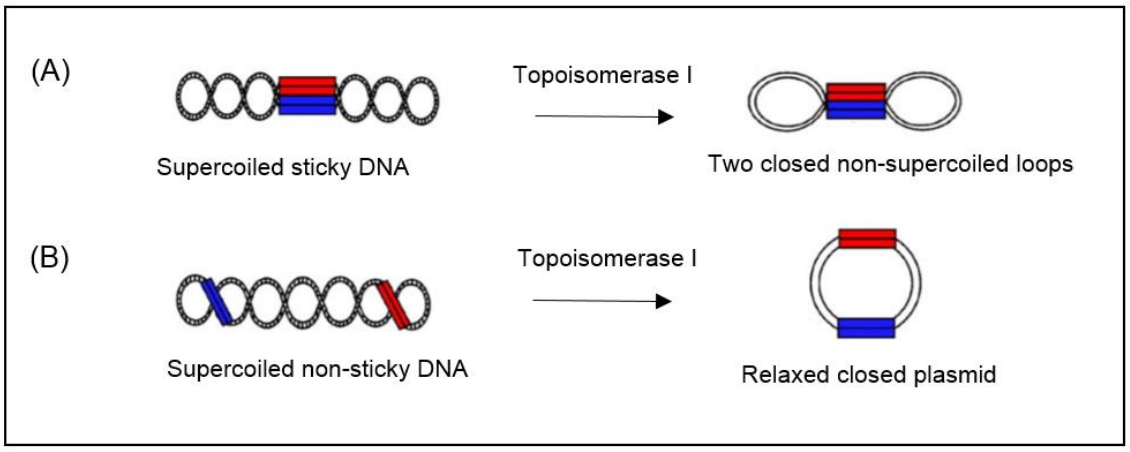



**Figure 2.13:** Models of the supercoiled sticky and non-sticky DNA after treatment with topoisomerase I. **(A)** The DNA.DNA associated region is still stable even after treatment with topoisomerase I. The formation of two closed non-supercoiled structures will yield a band with an alternative migration to the relaxed plasmid. **(B)** The result of treatment of the non-sticky DNA is relaxed covalently closed DNA.

### 2.2.11 Reporter gene expression assay

Reporter gene technology has become a powerful tool in studies of gene expression. The firefly luciferase reporter is one of the most important reporters and it is frequently used to investigate the regulatory potential of a cloned DNA sequence on gene expression because of its high sensitivity. This property comes from its lack of any endogenous luciferase activity in most cell types. Firefly luciferase is a monomer protein and does not require post-translational modification for its activity, thus upon synthesis it is secreted out of the cell. Briefly, the basic steps of the firefly luciferase assay are follows: first, a luciferase expression plasmid is introduced into appropriate mammalian cells; second, incubation for 24-48 hours, then lysis of the cells to release luciferases into the solution; third, addition of a limited amount of substrate (luciferin) and oxygen; finally, measuring the level of luciferase expression as a light emission by a luminometer (213, 214). The dual-luciferase reporter assay in combination with the pRL Vector which express Renilla luciferase as the second reporter, offers the exceptional speed, sensitivity and convenience of two luciferase reporter assays in one tube. The firefly luciferase is used as the experimental reporter and *renilla* gene expression serves as an internal control (214).

#### 2.2.11.1 Cell culture

Cell culture is a technique that grows specific living cells under appropriate conditions which mimic their natural environment. Human Hela cells (cervical carcinoma cells) were used as a host for pGL3-control plasmids carrying GAA.TTC repeats to examine their effects on gene expression. Hela cells were kindly supplied by Dr Mark J. Coldwell (Centre of Biological Sciences, University of Southampton) and stored frozen at -176 °C in liquid nitrogen. The Hela cell lines were cultured at 37 °C with 5% $CO_2$ in Dulbecco's Modified Eagle's Medium (DMEM) containing 4.5 g/L D-Glucose, 25 mM HEPES and 10% heat inactivated fetal calf serum (FCS). The following day, the media was aspirated off and the cells were then washed twice in pre-warmed calcium-magnesium free phosphate buffered saline (CMF-PBS). Trypsin-EDTA was added to the flask which were then placed at 37 °C in a 5% $CO_2$ incubator for 5 minutes to suspend the cells. After the cells were detached from the plate, pre-warmed culture medium was added to inhibit the trypsin activity. The cells were counted by pipetting 10 µl from the cell suspension solution to the haemocytometer counting chamber to accurately quantitate and standardize experimental conditions.

#### 2.2.11.2 Transfection

One day before transfection, 5000 Hela cells were mixed with 200 µl growth media in each well of a 96-well plate and kept at 37 °C in a 5% $CO_2$ incubator overnight. For each well of cells to be transfected, 1.8 µl of transfected reagent (Gene juice, Novagen®) was added to 60 µl of serum-free media in a sterile Eppendorf tube, mixed gently with a vortex mixer and then incubated at room temperature in the hood for 5 minutes to form a transfection-complex mixture. 3 µl of 1µg/µl pGL3-control plasmid which contains GAA.TTC repeat was mixed with 3 µl of 1µg/µl phRL-CMV plasmid containing the *Renilla* luciferase gene and then added to each sample. The samples were then incubated at room temperature for 10-20 minutes in the hood. After this period of incubation, 5.65 µl of the transfection-complex mixture was loaded into each well of the overnight 96-well plate and the plate was placed in the 5 % $CO_2$ incubator at 37 °C for 24 hours. After 24 hours of incubation, the media was removed and replaced with new growth media and the cells were incubated for another 24 hours in the incubator before assaying gene expression.

#### 2.2.11.3 Dual luciferase assay

The luciferase activity in each reporter vector was assayed according to the manufacturer's protocol (Dual-Glo® Luciferase Assay System, Promega). After 48 hours of transfection, the media was aspirated and the cells were washed with 100 µl 1x PBS buffer. 20 µl of 1x passive lysis buffer was added into each well to release firefly and *Renilla* luciferases. The plate was incubated on a plate rocker for 15 minutes. After this period of incubation, 5 µl of each cell lysate was transferred into a new well of a white 96-well plate. 25 µl of luciferase assay reagent (LarII) and Stop & Glo reagent were added into each well. Measurement of luciferase activity was done automatically with a GloMax® Multi Detection system.

# 3 Cloning and Characterisation of Plasmids Containing GAA.TTC Repeats

## 3.1 Introduction

In order to study the effects of GAA.TTC repeats on the structural and biological behaviour of plasmids, we constructed multiple plasmids that contained various lengths of the triplet repeat (either single or in combination). In this work, the pUC18 plasmid was used to study the structural properties of GAA.TTC sequence repeats, through the use of chemical and enzymatic probing methods, as well as its behaviour during agarose gel electrophoresis. The pUC18 vector is one of the most popular cloning vectors in *E.coli*. It was engineered to have many different unique restriction sites in a small space. This polylinker or multiple cloning site (MCS) is located within the *lac*Z gene, so that the insertion of DNA fragments into this region inactivates the gene. This provides a sereenable marker to identify and distinguish plasmids containing insert from plasmids without insert based on colour differences of colonies on agar plate. The *lac*Z gene of pUC18 plasmid codes for the α-subunit of β-galactosidase, which converts lactose into glucose and galactose. Therefore, the β-galactosidase gene in some strains of *E.coli* is unable to synthesise complete β-galactosidase protein without complementation from the *lacZ* gene produced from the pUC18 plasmid (215). In our cloning experiments with pUC18, a sequence of GAA repeats was first inserted into MCS region and, after the transformation process, cells containing the recombinant plasmid were plated on agar plate that contained lactose analogue X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) which is cleaved by β-galactosidase to generate coloured blue product. IPTG (isopropylthiogalactoside) was added as an inducer of the enzyme. After incubation overnight at 37 °C, successful clones were detected as white colonies on plates containing X-gal and IPTG.

The second main work described in this chapter was to prepare constructs of the pGL3 luciferase reporter vector containing various lengths of GAA.TTC repeats. Investigation into whether the luciferase expression level of these constructed vectors is affected significantly by presence or absence of single or double sequence of GAA.TTC repeats will allow us to better understand the effect of unusual DNA structures formed from this

triplet repeat on the regulation of gene expression (explained in more detail in chapter 5). The pGL3 vector used in this study contains appropriate promoter and enhancer sequences, resulting in high expression of the luciferase gene in mammalian cells.

Two stretches of GAA.TTC repeats, which are at least 59 repeats in length, in a single supercoiled DNA are required for the formation of sticky DNA (125). Thus, this chapter focuses on experimental methods used to clone different lengths of GAA.TTC repeats into pUC18 and pGL3 control vectors. In addition, a number of techniques were employed to study conformational behaviour of plasmids containing this triplet repeat during agarose gel electrophoresis.

## 3.2 Cloning inserts into pUC18 and pGL3-control vectors

Prior to insertion of expanded GAA.TTC triplet repeats into the SmaI site of vector pUC18, the synthetic oligonucleotides $(GAA)_{10}$ and $(TTC)_{10}$ were extended in two ways. The first depended on the ability of DNA ligase to join the cohesive ends of the annealed duplex triplet repeats to generate multimers of this sequence $(GAA.TTC)_{10}$. The second utilised PCR to synthesise increased lengths of the GAA.TTC repeats. The DNA fragments generated from both strategies were treated with mung bean nuclease to give blunt ends. In subcloning experiments, moving the inserts from one position to another, either in the same vector or in a different vector, PCR was employed to amplify the GAA.TTC using primers that included appropriate restriction sites. The desired DNA was then inserted into the vector that had been cut with the same enzyme using DNA ligase as described previously. To confirm correct vector recombination, random colonies were selected and analysed by colony PCR (see chapter 2 for more details). After colony PCR had confirmed the presence of a DNA insert the plasmids were sent to Eurofins MWG Operon for sequencing to confirm that each construct had no mutations and contained the correct sequence.

### 3.2.1 Experimental design

The required feature of sticky DNA is the association of two long GAA.TTC repeat sequences that are distal to each other in a supercoiled plasmid. We chose the SmaI and AatII sites as the best positions to insert these triplet repeats into the pUC18 vector, where the distance between them is 501 bp. In the initial cloning experiments, many attempts failed to clone the second GAA repeat into AlwNI and SspI sites of cloned pUC18 plasmid, but we realised later that these positions were situated in critical regions of the

plasmid. The AlwNI site is located in the origin of replication and the SspI site in the promoter region of the antibiotic resistance gene. However, cloning of the first insert into SmaI site was described in Chapter 2, Section 2.2.2.1. The second stage in the cloning procedure was subcloning the insert from the SmaI region into the AatII site of pUC18 and the SacI and PciI sites of pGL3 control vector. To achieve this, the insert in the SmaI region was amplified by PCR, using primers containing appropriate restriction sites (if required) for each subcloning process as shown in chapter 2. All these primers were designed to be between 18 to 28 nucleotides with a base composition of between 50-60 % G/C with a G or C at the 3' end to prevent DNA breathing.

### 3.2.2 Results

#### 3.2.2.1 Cloning of GAA.TTC repeats into pUC18 and pGL3 vectors

After several unsuccessful attempts, various lengths of GAA.TTC repeats, ranging from 10 to 94, were successfully cloned into pUC18 and pGL3 vectors as shown in Table 2.6, chapter 2. For the cloning into the pUC18 vector, although we would have liked to generate some longer repeats (more than 94) we were unable to generate plasmids with longer than this length and most of the plasmids contained shorter inserts with less than 50 GAA.TTC repeats. This is considered in the discussion section at the end of this chapter.

As mentioned previously, the synthetic sequence $(GAA.TTC)_{10}$ was extended by self-ligation with T4 DNA ligase or self-PCR. For oligonucleotides that had been extended by ligase, a mixture of the ligated DNA fragments (monomers, dimers and multimers < 95 repeat) were cloned into the SmaI site of pUC18. We also attempted to clone much longer stretches of the GAA.TTC repeat, using only high-order multimers of ligated duplex oligonucleotides. For this protocol we separated the ligated oligonucleotides on a 1.5% agarose gel (Figure 3.1), in which higher order oligomers are clearly visible. The bands corresponding to the multimers with at least 110 triplet repeats, were excised from the gel and eluted using QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's recommendations. However, the concentration of the purified DNA fragments as measured by NanoDrop 2000 (Thermo Scientific), was very low (3-8 ng/µl). Although the mixture of the ligated DNA was loaded in 5 wells and the multimers were extracted and combined into one sample, the concentration was not sufficient for cloning and only blue colonies were seen in the agar plate.

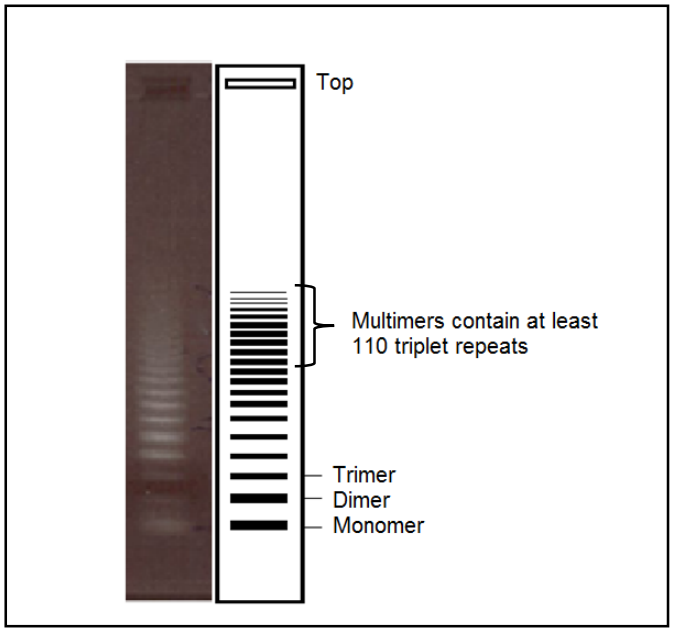



**Figure 3.1:** 1.5% agarose gel showing a mixture of ligated DNA fragments. Right lane is schematic reproduction of the left lane (for clarity), showing the individual bands in multiples of the triplet repeat.

Long and short inserts (94 and 29 GAA repeats) in the SmaI site were selected for subcloning into the AatII site of pUC18, and the SacI and PciI sites of pGL3. PCR was employed to amplify the inserts using flanking primers designed to incorporate additional unpaired bases at the 5' terminus that included restriction sites required for cloning. The amplified DNA products and vector DNA were digested with appropriate restriction enzymes in order to produce compatible ends for ligation. After transformation of ligated plasmids, random selections of colonies were examined by colony PCR to check that the ligation had been successful since blue/white selection was not available. The relevant forward and reverse primers were used for each position as presented in Chapter 2, Figure 2.10. Figure 3.2 shows an example of an agarose gel representing the products of colony PCR to determine the presence or absence of the insert DNA in the colonies. Successful cloning into the AatII site of the construct pUC18 was shown by the appearance of a new band at approximately 220 base pairs (lane 5), whereas the amplified size of this region from the native vector was approximately 95 base pairs (lanes 1-4 and 6-8).

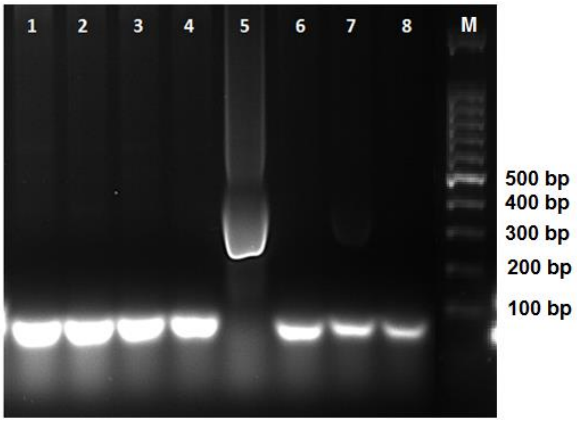



**Figure 3.2:** Example of agarose gel electrophoresis of the colony PCR products to screen the inserted DNA in the AatII site of the pUC18 vector. Lanes (1-8) correspond to results of colony PCR from seven different bacterial colonies. Lanes (1-4, 6-8) show negative results, whereas lane 5 represents successful positive clone (approximately 220 base pairs amplified from 29 triplet repeats). Lane (M) represents the DNA marker.

After we had identified which colonies contained inserts of the appropriate length, the positive clone was sent for sequencing using the relevant forward primer (5'-CGCACATTTCCCCGAAAAGTGC-3') as shown in Chapter 2, Section 2.2.5. The sequencing results (Figure 3.3) confirmed that the cloned insert has the correct sequence with no mutations.

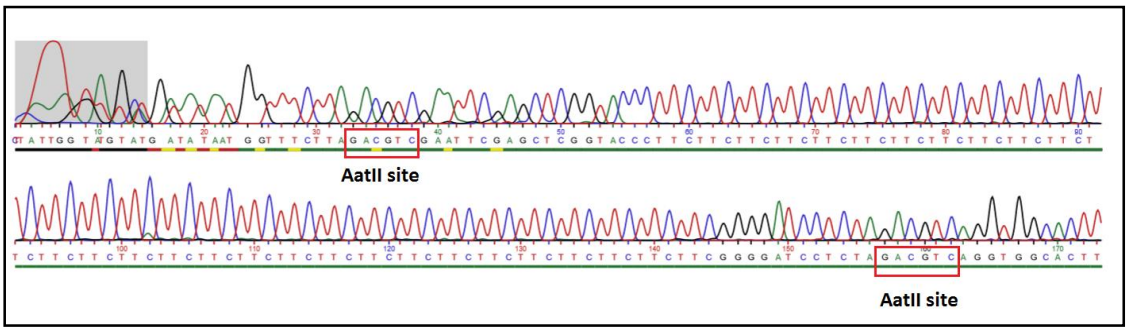



**Figure 3.3:** Sequencing analysis of 29 repeats of GAA.TTC sequence which was subcloned into AatII site of construct pUC18. The cloned sequence is indicated between the AatII restriction sites (red boxes).

29 repeats of GAA.TTC sequence were subcloned successfully into AatII and SacI sites of pUC and pGL3 vectors, respectively. However, we could not subclone the full length of the 94 triplet repeats and all clones contained much shorter lengths of the sequence. To help alleviate this problem, we used super competent SURE (Stop Unwanted Rearrangements Events) cells instead of TG2 cells for the transformation process. The

clones with these cells also did not generate longer than 94 repeats in the SmaI site. Although long inserts were more stable with SURE cells than TG2 cells, we failed to subclone 94 GAA repeats. Thus, we decided to subclone the shorter length (85 repeats) into the AaII site of the pUC18 vector and into the SacI and PciI sites of pGL3 vector using SURE cells. The results obtained from DNA sequencing confirmed that the full length of 85 repeats were subcloned successfully into SacI and PciI sites of pGL3 vector (either in single or in combination) but not into the AatII site. However, we were able to generate pUC18 plasmid containing two long repeats of GAA.TTC sequence located at positions SmaI (85 repeats) and AatII (62 repeats), and this will be shown in the next section.

#### 3.2.2.2 Generation and Characterization of plasmid containing 85 and 62 GAA repeats

The main aim of the work described in this chapter was to generate a set of pUC18 plasmids harbouring long and short lengths of GAA.TTC repeat in one or two positions. The minimum length of the triplet sequence required to adopt sticky DNA conformation is 59 repeats, under superhelical torsional stress. However, in the initial cloning experiments, we were unable to obtain plasmids harbouring more than 85 (GAA.TTC) repeats in two positions in a single pUC18 vector; most of the plasmids contained short inserts (less than 50 triplet repeats). A possible explanation is that this could be due to instability of the long triplet repeats (deletion) during replication in *E.coli* (this is further explained in the discussion of this chapter). During routine studies of the cloned plasmids by agarose gel electrophoresis, we noted that all construct plasmids (either harbouring one or two tracts of GAA.TTC sequence) were prone to form retarded bands. We first considered that these represented nicked plasmids, but it subsequently became clear that these were multimers (dimers and trimers) of the plasmids (this phenomenon will be expanded in section 3.5).

In addition, we observed that one of the cloned plasmids, which contained two stretches of the triplet repeats, exhibited an unusual mobility on agarose gels (Figure 3.4, lane 3). It can be seen that the bottom band in lane 3 (recombinant plasmid with two inserts) has a much slower mobility than the bottom band in lane 2 (recombinant plasmid with one insert) and the band in lane 1 (native supercoiled pUC18), whereas DNA in the bottom band of lane 2 moved slightly slower than native pUC18 (lane 1). The initial sequencing of this plasmid showed that the SmaI site contained 85 triplet repeats, with 62 triplet repeats in the AatII site. When this abnormal plasmid was retransformed into SURE cells

it gave the same gel mobility in the agarose gel, but the sequencing results showed that, while SmaI site still contained 85 triplet repeats, different sizes were obtained for the insert in the AatII site. The 3'-ends of these sequences were often difficult to read and appeared to show overlapping bands suggesting that the preparation contained more than one type of plasmid.

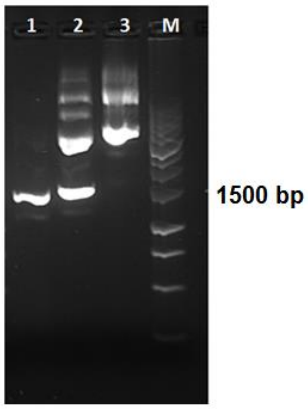



**Figure 3.4:** 1% Agarose gel showing native supercoiled pUC18 plasmid (lane 1), pUC18 plasmid harbouring 85 (GAA.TTC) repeats in the SmaI site (lane 2), and unusual recombinant pUC18 plasmid (lane 3). Lane (M) represents DNA marker.

We considered the possibility that this supercoiled plasmid containing two inserts could have the ability to form sticky DNA, thereby generating the retarded band (lane 3). To investigate this possibility, the unusual recombinant plasmid was cleaved with different restriction enzymes (SspI, NdeI, double digestion with SspI and NdeI, BamHI, HindIII, EcoRI and SacI), and then analysed by agarose gel electrophoresis as shown in Figure 3.5. The native supercoiled pUC18 and its cleaved products with various restriction enzymes were used as a control. By this means we serendipitously discovered that the abnormal plasmid contained multiple DNA inserts. As seen in Figure 3.5, this plasmid is cleaved by SspI and NdeI into a long linear DNA fragment and two short fragments (521 bp and 590 bp). By comparing the sizes of these short fragments with the band at 367 bp, which is generated by the same enzymes with the control pUC18, we concluded that the plasmid contained two different length AatII inserts in two different regions. The plasmid also contained three BamHI sites, one HindIII site and two sites each for EcoRI and SacI. Based on the sizes of the cleaved DNA fragments in the agarose gel, we generated a possible restriction map for this plasmid, as shown in Figure 3.6, which allowed us to

identify the number and locations of the restriction sites on the plasmid. This unusual plasmid appears to be a dimer of the recombinant plasmid and contains two sites of SspI and NdeI, four sites of BamHI, EcoRI and SacI, two sites of HindIII, as indicated in the map below.

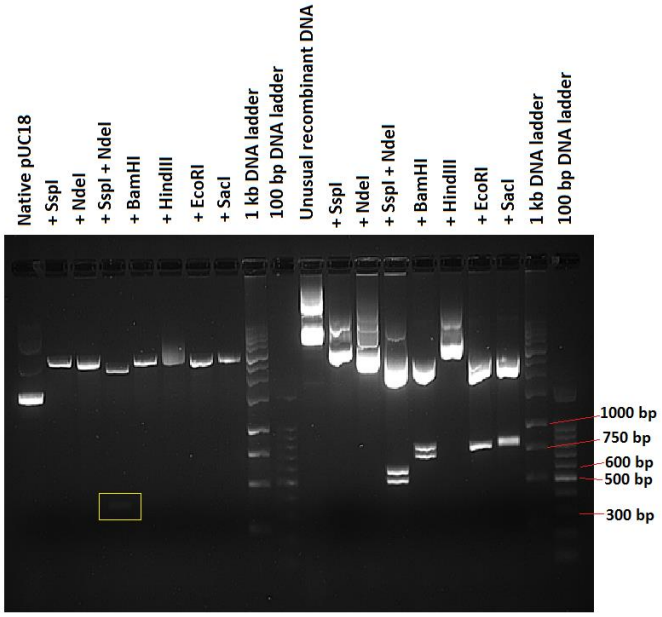



**Figure 3.5:** Agarose gel electrophoresis of the unusual DNA plasmid after digestion with various restriction enzymes compared with the pUC18 control. The first lane on the left side is the native (empty) supercoiled pUC plasmid, followed by seven restriction enzyme digestions, as shown on the gel. The products of cleavage by double digestion (SspI and NdeI) are a short fragment (367 bp, as indicated in yellow box) and a long linear fragment.

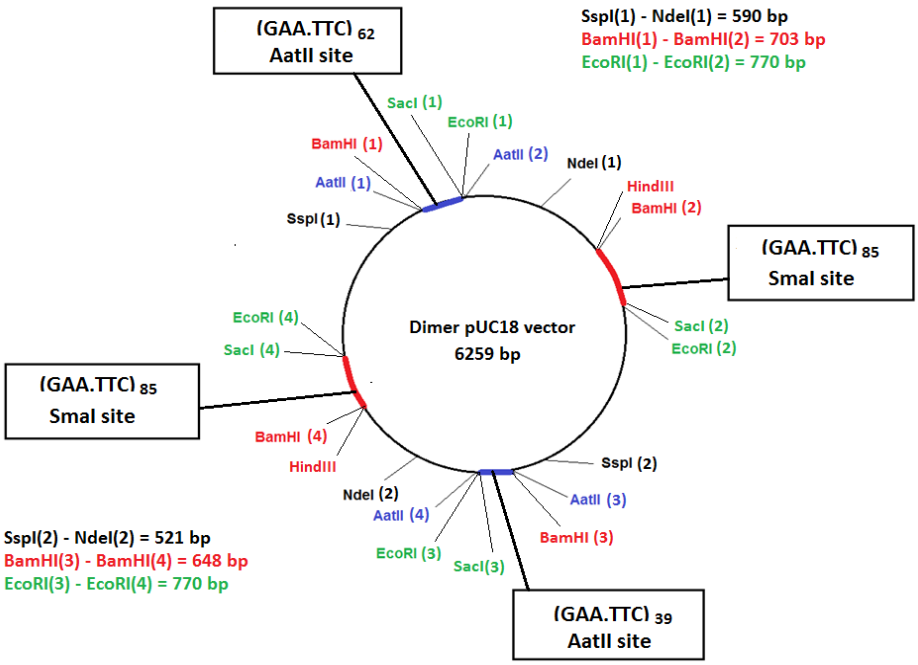



**Figure 3.6:** Map of the dimer recombinant plasmid.

Since the insertion of two long stretches of GAA repeats (>59 repeats) within a single pUC18 vector took considerable time, we decided to cut the dimer plasmid with NdeI, and then religate it to generate the two monomer plasmids (Figure 3.7). One of these new plasmids will harbour 85 GAA repeats in the SmaI site and 62 GAA repeats in the AatII site, and the other plasmid will harbour 85 GAA repeats in the SmaI site and 39 GAA repeats in the AatII site. The ligated plasmids were transformed into *E.coli* SURE cells and three different plasmids were isolated from three different colonies and digested with the enzymes SspI and NdeI. The cleaved products were analysed by agarose gel electrophoresis as shown in Figure 3.8. Plasmids 1 and 2 produced short DNA fragments corresponding to 590 bp and digestion of plasmid 3 generated shorter fragment corresponding to 521 bp, as well as the remainder linear DNA in each product. By comparing with cleaved products of the dimer plasmid, it clearly appears that the double construct plasmid successfully produced two new plasmids harbouring 62 and 39 GAA repeats in the AatII site in each one. These plasmids were then sequenced by Eurofins MWG Operon and the results confirmed these predictions as shown in Appendix.

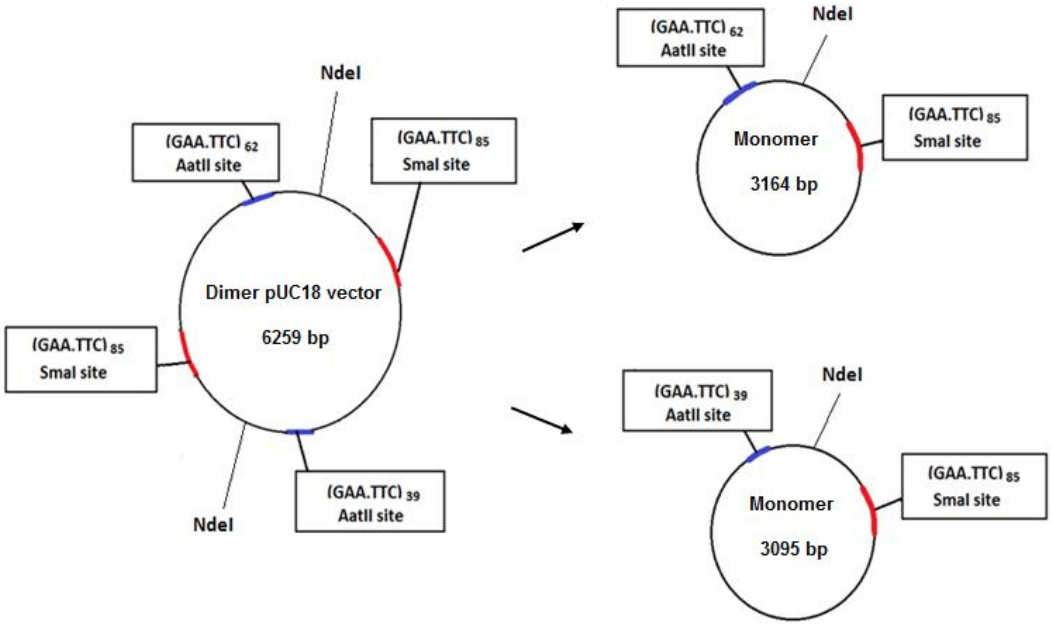



**Figure 3.7:** Illustrates generation of two monomer recombinant plasmids from the parental dimer plasmid. This was carried out by digestion with NdeI enzyme and ligated again with ligase.

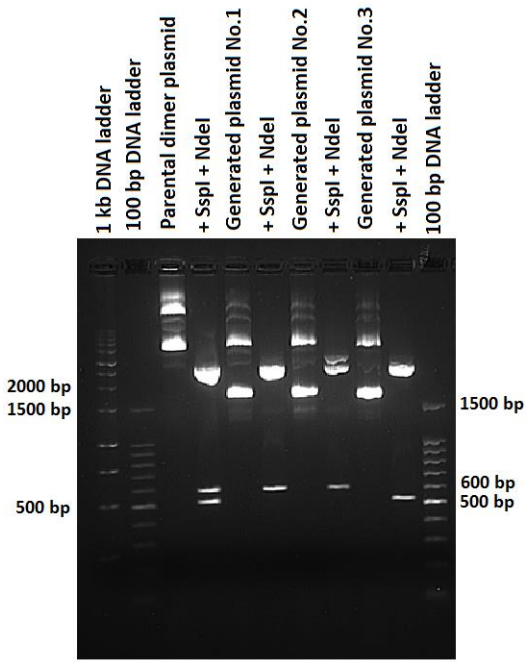



**Figure 3.8:** Products of three different transformed plasmids after double digestion with SspI and NdeI, were run on 1% agarose gel. The parental dimer plasmid and its cleaved product were used as a control.

Regarding the sequencing results, it is worth noting that peak intensities of many cloned inserts dropped off early. Figure 3.9 shows an example of the sequencing results of TTC strands using reverse primer. In this Figure, it is obvious that strong sequencing signals appeared at the beginning of the repeat then gradually tapered off to a difficult reading region. This reading problem may be due to formation of unusual DNA structures (*i.e.* hairpins structure) by the triplet repeat that arrest DNA polymerase and attenuate the chain elongation.

In addition, DNA sequencing of some the cloned inserts appeared as multiple peaks with the same height or in different heights that overlap each other. An example is shown in Figure 3.10. Sequencing of this plasmid produced clear peaks for the first 46 GAA repeats (red box) after which the sequencing became highly jumbled with at least two overlapping peaks at each position. We are confident that the plasmid was isolated from a single colony, and therefore it seems likely that it contained two different plasmids, each with different length inserts (one with 46 GAAs and one with a longer insert). It appears that some of the plasmids had suffered a partial deletion of the triplet repeat during replication. This must have occurred shortly after transformation as the two species appear in similar quantities.

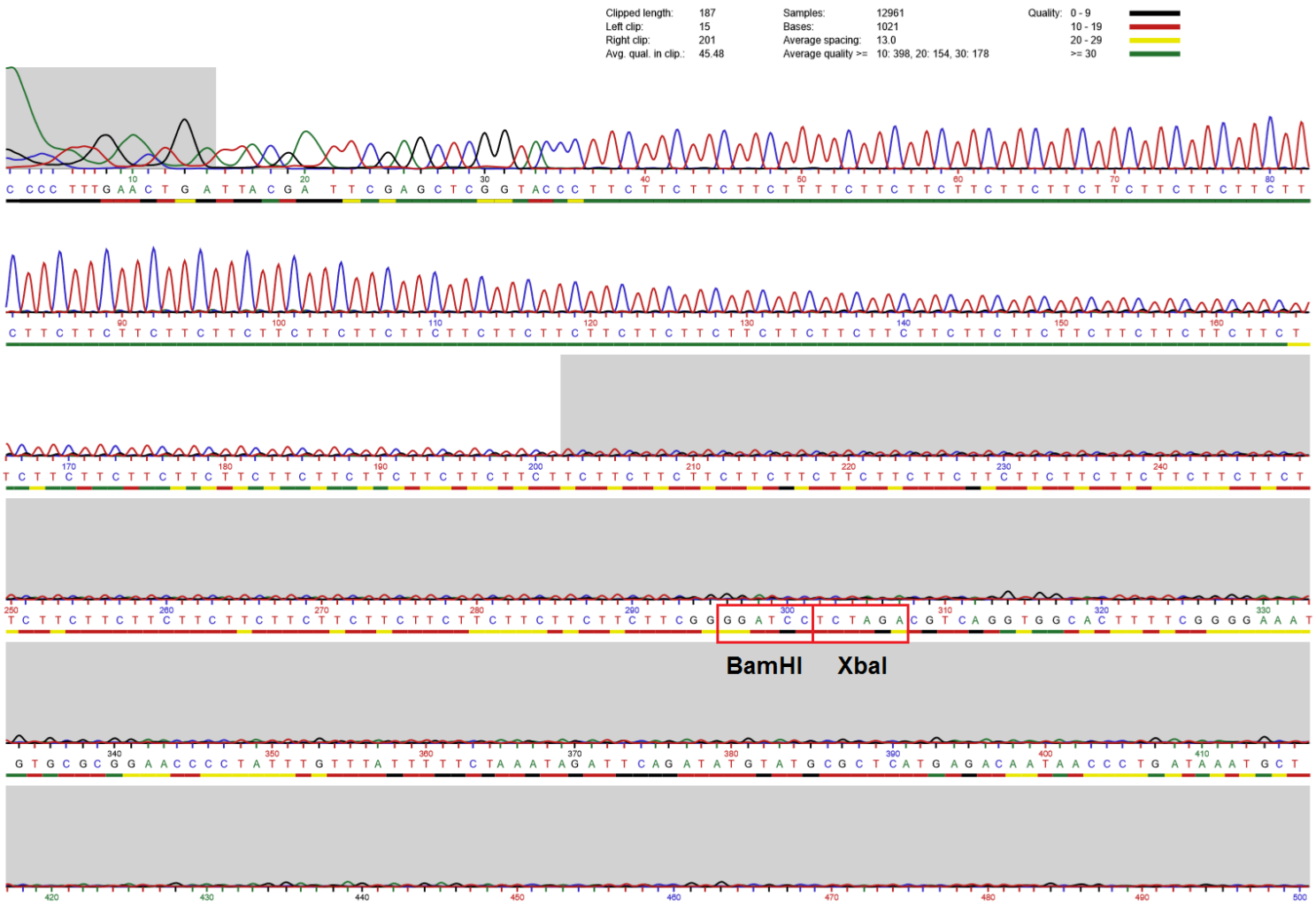



**Figure 3.9:** Sequencing analysis of GAA.TTC repeats in the SmaI site of pUC18 using universal reverse primer. The result shows gradual attenuation of DNA sequencing within the repetitive stretches of DNA sequence.

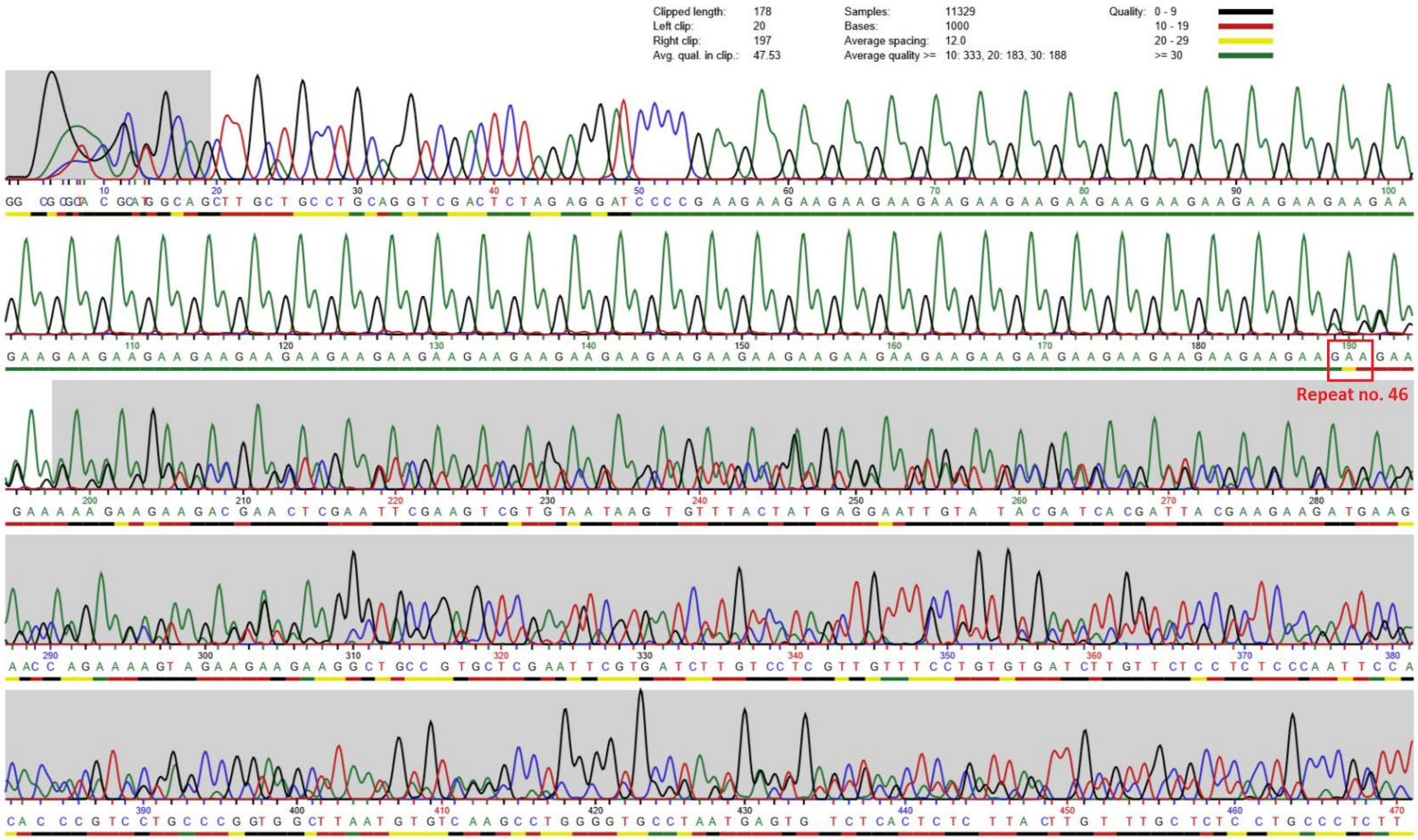



**Figure 3.9:** Sequencing analysis of GAA.TTC repeats in the SmaI site of pUC18 using universal forward primer. The result shows overlapping peaks appeared from repeat number 46 in the sequencing data, indicating presence of heterologous DNA template in this region.

## 3.3 Detection of sticky DNA by gel mobility shift assay

It is known that DNA.DNA associated sticky DNA regions within supercoiled DNA are extremely stable and require a high temperature (80°C) to convert the adhered region back to a duplex conformation. The first investigation of the presence of sticky DNA caused by long GAA repeats used gel electrophoresis mobility shifts (3), in which it was observed that plasmids containing at least 59 GAA repeats generated a retarded digestion band in addition to the expected linear fragment. This was discussed in Chapter 1, Section 1.4.1. We therefore examined the presence of sticky DNA in these plasmids by cleavage with different restriction enzymes followed by agarose gel electrophoresis.

### 3.3.1 Basic Experimental design

Two assays were used to detect the formation of sticky DNA by agarose gel electrophoresis: restriction enzyme digestion and restriction enzyme accessibility (Figure 3.11). EcoRI site was selected as an appropriate cleavage site (Figure 3.12) for detecting the appearance of any EcoRI-cleaved sticky DNA as a retarded band in the gel. The BamHI cleavage site is immediately adjacent to the GAA tracts and this was therefore used to examine the accessibility of the DNA.DNA associated region to cleavage, on the assumption that the presence of a sticky DNA confirmation would attenuate restriction enzyme cleavage in the surrounding region. In this investigation, plasmids that were used for detection of sticky DNA are called pUC18(X,Y) where X is the repeat number in the SmaI site and Y is the number in the AatII site. These plasmids are empty pUC18(0,0), pUC18(29,0), pUC18(29,29), pUC18(85,0), pUC18(85,39), pUC18(85,62).

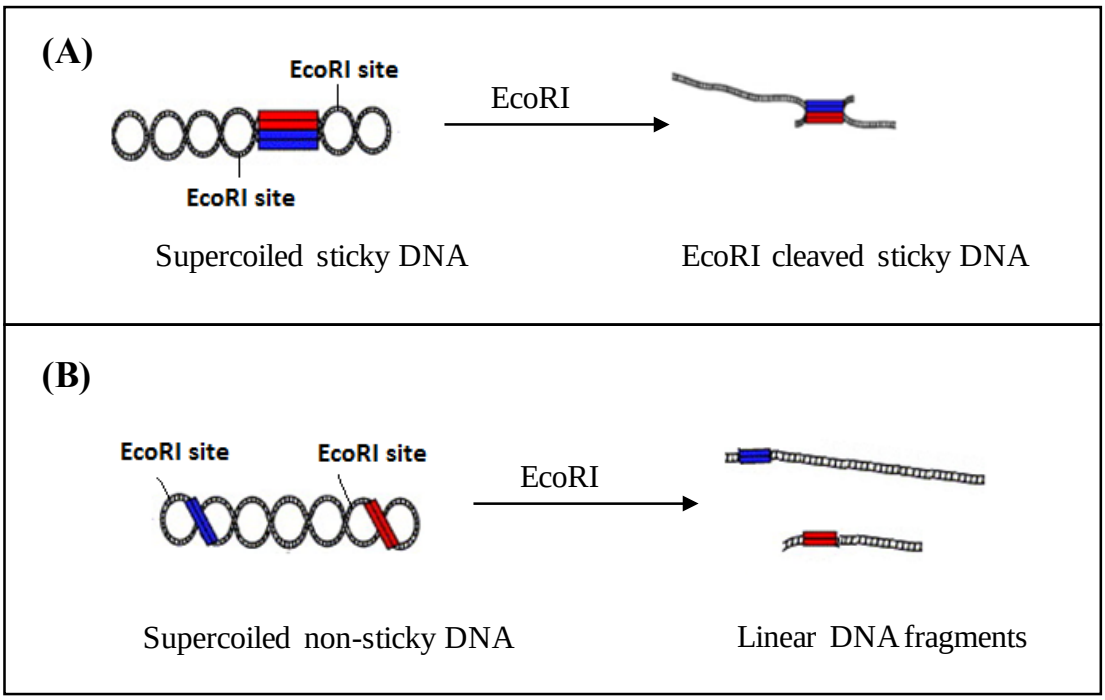



**Figure 3.10:** Models of the supercoiled sticky and non-sticky DNA after digestion with EcoRI enzyme. **(A)** The DNA.DNA associated region is still stable even after enzymatic digestion. The linear duplex arms of the EcoRI cleaved sticky DNA will affect the mobility of the DNA on the agarose gel, generating a retarded band compared to the linear DNA. **(B)** The result of cleavage of the non-sticky DNA is only linear DNA fragments.

```
5'-GGATCCCC(insert)GGGTACCGAGCTCGAATTC-3'
```

**Figure 3.11:** The DNA sequence of the cloned SmaI or AatII sites region between BamHI (blue) and EcoRI (red) sites.

### 3.3.2 Results

Previous studies show that the two long tracts of GAA.TC repeats must be within the same supercoiled plasmid and in a direct orientation with each other to form the sticky region. We expected pUC18(85,62) to generate EcoRI cleaved sticky DNA and produce a retarded band on the gel after digestion by EcoRI, compared to the remainder of the vector, which migrated as a linear DNA. However, no detectable retarded band was found in pUC18(85,62) and only linear DNA fragments were observed on the agarose gel as shown in Figure 3.13. The other four construct plasmids, which were not expected to form sticky DNA, were also analysed in parallel for comparison. It can be seen that the results

of cleavage of these plasmids by the EcoRI or BamHI enzyme gave the same product on the gel; only linear fragments were observed, with no evidence of any retarded bands. Plasmids that contained one tract of GAA repeats generate a single linear DNA fragment after digestion with either EcoRI or BamHI, while plasmids that contained two GAA repeat tracts gave two linear DNA fragments as there are additional EcoRI and BamHI sites in the cloned AatII region that were created by the subcloning process (see Materials and Methods, Chapter 2). It should also be noted that all the plasmid preparations (either having one or two tracts of the triplet repeats) appeared as more than one band in the control (uncut) lanes. These are not nicked or linear DNA, but correspond to multimeric plasmid structures; this will be considered further in the next sections.

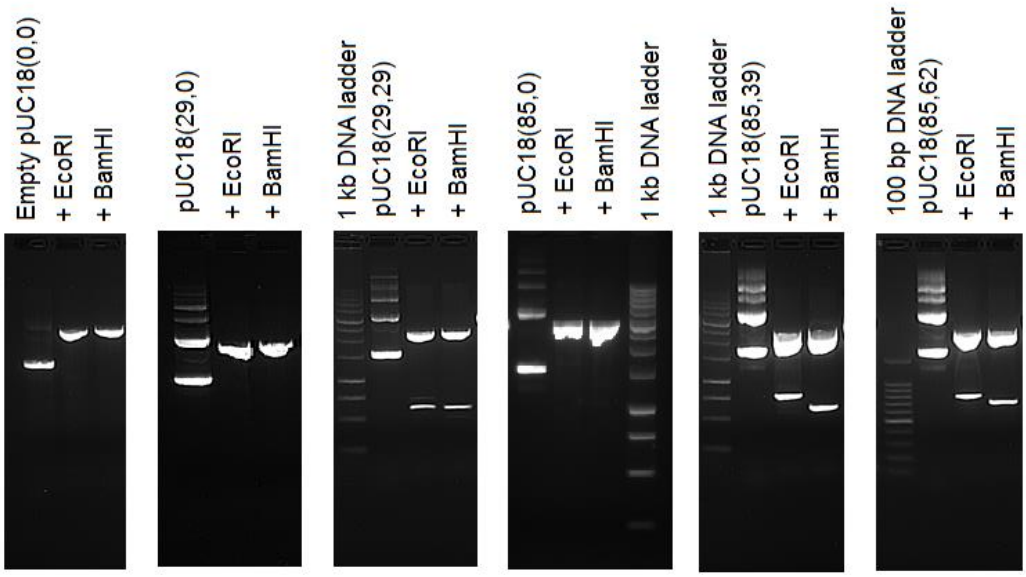



**Figure 3.12:** Agarose gel electrophoresis showing EcoRI and BamHI digested plasmids.

The closest restriction site to the inserts is for BamHI, and this was therefore used to test for the presence of sticky DNA by examining the accessibility of this site to digestion. We reasoned that the presence of an unusual DNA structure might alter the accessibility of these BamHI cleavage sites, and that plasmid pUC18(85,62) would be cleaved less easily than the other plasmids. The results of this experiment are shown in Figure 3.13, in which it can be seen that BamHI is able to completely cleave all the plasmids.

In the light of this negative result we examined the effect of different buffer conditions on the plasmid containing two GAA inserts pUC18(85,62). This was incubated with

(50 mM sodium acetate buffer pH 5.0 containing, 5 mM $MgCl_2$) or (40 mM Tris-acetate pH 7.0, containing 5 mM $MgCl_2$) at 37 °C for 18 hours to allow metal ions to promote and stabilise the sticky DNA formation. The DNAs were then digested with EcoRI to test the formation of the sticky DNA retarded band, and also digested with BamHI to examine its accessibility to the digestion. The results are shown in Figure 3.14 and show no EcoRI retarded bands, with only simple linear digestion products. BamHI also cleaved the plasmids easily. These negative results might suggest that longer lengths of GAA repeats are required for sticky DNA formation. Note that the plasmid controls in these experiments contain several bands, which are thought to correspond to plasmid multimers as described below.

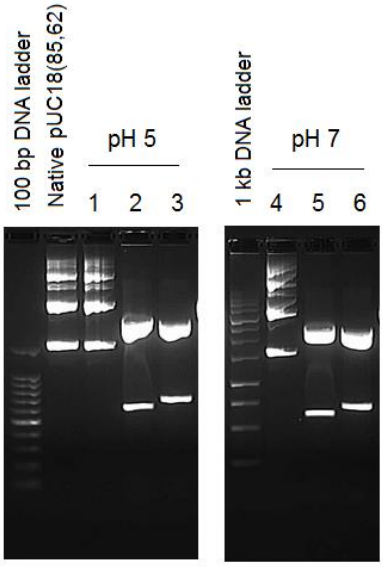



**Figure 3.13:** Agarose gel electrophoresis showing BamHI (lanes 2 and 5) and EcoRI (lanes 3 and 6) digested pUC18(85,62) after incubation at pH 5 and pH 7; lanes 1 and 4 correspond to the native plasmid, in the presence of $MgCl_2$.

## 3.4 Relaxation with topoisomerase I

Most DNA in its native state is a negatively supercoiled, *underwound*, meaning that there is a reduction in the number of helical turns compared to relaxed DNA and more base pairs per helical turn, resulting in a decrease in the angle between adjacent base pairs. During transcription, the movement of RNA polymerase along DNA generates negative supercoils upstream and positive supercoiling downstream of the transcription bubble. Negative and positive supercoiling states create torsional tension in the winding of the

DNA double helix, and this stress is relieved by cellular enzymes known as topoisomerases (89). However, the torsional energy which is generated from DNA unwinding can destabilise the Watson-Crick bonds of the double helix and subsequently causes localised topological changes in the DNA molecule. The formation of non-B DNA structures is often sensitive to variations in superhelical tension, DNA binding proteins and environmental conditions. Thus, the aim of this section is to assess the formation of any gross supercoil-dependent structures, which are generated by GAA repeats, using topoisomerase I.

DNA topoisomerases are enzymes that regulate the topological state of DNA through their ability to remove the torsional stress of supercoiled DNA by transient breakage of one (type I topoisomerase) or two strands (type II) of the double helix (216). Native bacterial plasmid DNA is negatively supercoiled and can adopt a mixture of coiled states, termed topoisomers, when treated with topoisomerase. In this experiment, agarose gel analysis was used to separate DNA topoisomers, which are produced by topoisomerase I. Supercoiled DNA molecules adopt a more compact shape and migrate rapidly through agarose gel matrix, while relaxed DNA migrates more slowly.

### 3.4.1 Experimental design

In this experiment, the empty pUC18 plasmid was used as a control and a set of construct plasmids, having one or two inserts of GAA repeats, were tested to examine the role of negative supercoiling on any conformational changes. This change can be detected by changes in the electrophoretic mobility of the plasmid in agarose gel. The supercoiled plasmids were incubated with wheat germ topoisomerase I to relax the supercoils, as described in Chapter 2. Each sample was then divided equally into two tubes and then electrophoresed on 0.7% agarose gel in the presence and absence of GelRed. The GelRed molecule intercalates between adjacent base pairs of double strand DNA resulting in the unwinding of the helix which, in turn, causes it to relax. If the DNA sample is completely relaxed, the GelRed will convert it to positive supercoiling DNA (Figure 3.15). Thus, the GelRed agent was used as a component of the agarose gel in order to examine the relaxation of our plasmids which had been catalysed by topoisomerase I. In this investigation, plasmids are described in the same way as in the preceding section. These plasmids are empty pUC18(0,0), pUC18(29,0), pUC18(85,0), pUC18(29,29), pUC18(29,29inv.), pUC18(85,62), pUC18(85,62)* and pUC18(85,39), where (inv.)

indicates inverted orientation of the inserts and (*) indicates unusual form of this construct.

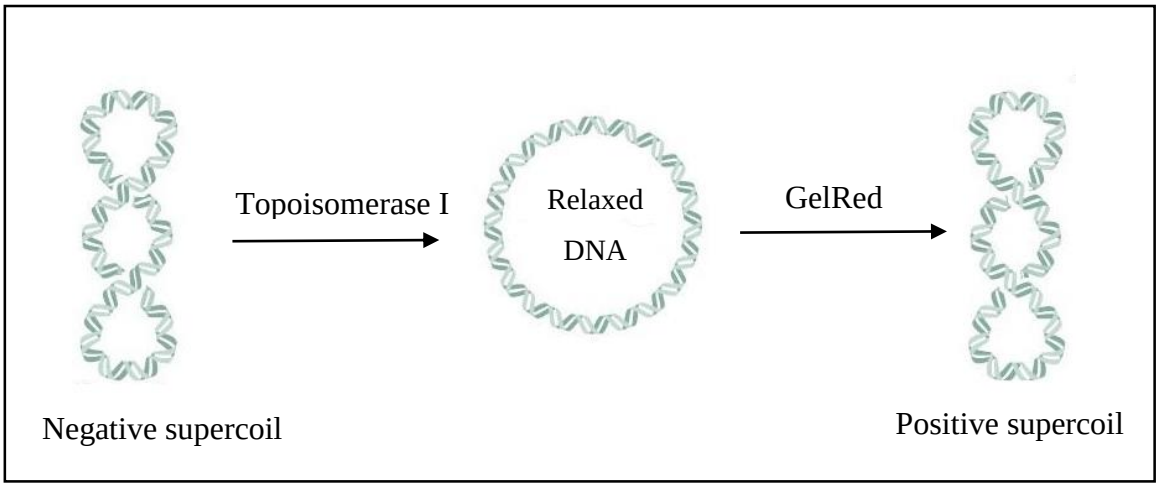



**Figure 3.14:** Illustrates the effect of topoisomerase I on negative supercoiling and, GelRed on the relaxed DNA.

### 3.4.2 Results

A representative agarose gel showing topoisomerase I relaxation of the different supercoiled plasmids is shown in Figure 3.16. Lanes 1 and 15 correspond to the native supercoiled pUC18 and appear as one band that was relaxed completely by the enzyme (lanes 2 and 16) as seen in the gels A and C which run in the absence of the intercalating agent GelRed. It can be seen that the native construct plasmids appeared as multiple bands (lanes 3, 5, 7, 9, 11, 13, 17) and each construct plasmid was affected by the activity of topoisomerase I, generating relaxed DNA which migrated more slowly (lanes 4, 6, 8, 10, 12, 14 and 18). This observation indicates that the relaxation activity of all the construct plasmids was not influenced by the presence of GAA repeats in these vectors.

It is interesting to note that in gels A and C the relaxed DNA appeared as a series of bands, a mixture of topoisomers, instead of one single band (completely relaxed DNA). This is due to an equilibrium between the topoisomers that is achieved by topoisomerase I relaxation. In addition, the appearance of two main regions of relaxed topoisomers can be observed in each lane of plasmids that are treated with the enzyme (but not in the pUC18 control). This observation indicates that each construct plasmid contains different sizes (multimers) of supercoiled DNA, which will be investigated in section 3.5.

The products of these reactions were also run on gel containing GelRed (which will fully unwind the plasmids) and the results are shown in Figure 3.16B and D. Untreated plasmids with topoisomerase I were run in the presence of the GelRed dye (lanes 1, 3, 5, 7, 9, 11, 13, 15 and 17), and exhibit one band for the empty pUC18 and multiple bands for the construct plasmids, as for the reactions that were run in the absence of the GelRed. The mobility of DNA bands after topoisomerase treatment is shown in lanes (2, 4, 6, 8, 10, 12, 14, 16 and 18); these seem to be identical to DNA bands of untreated plasmids, since GelRed intercalates into the DNA and so alters its structure and migration through the agarose gel.

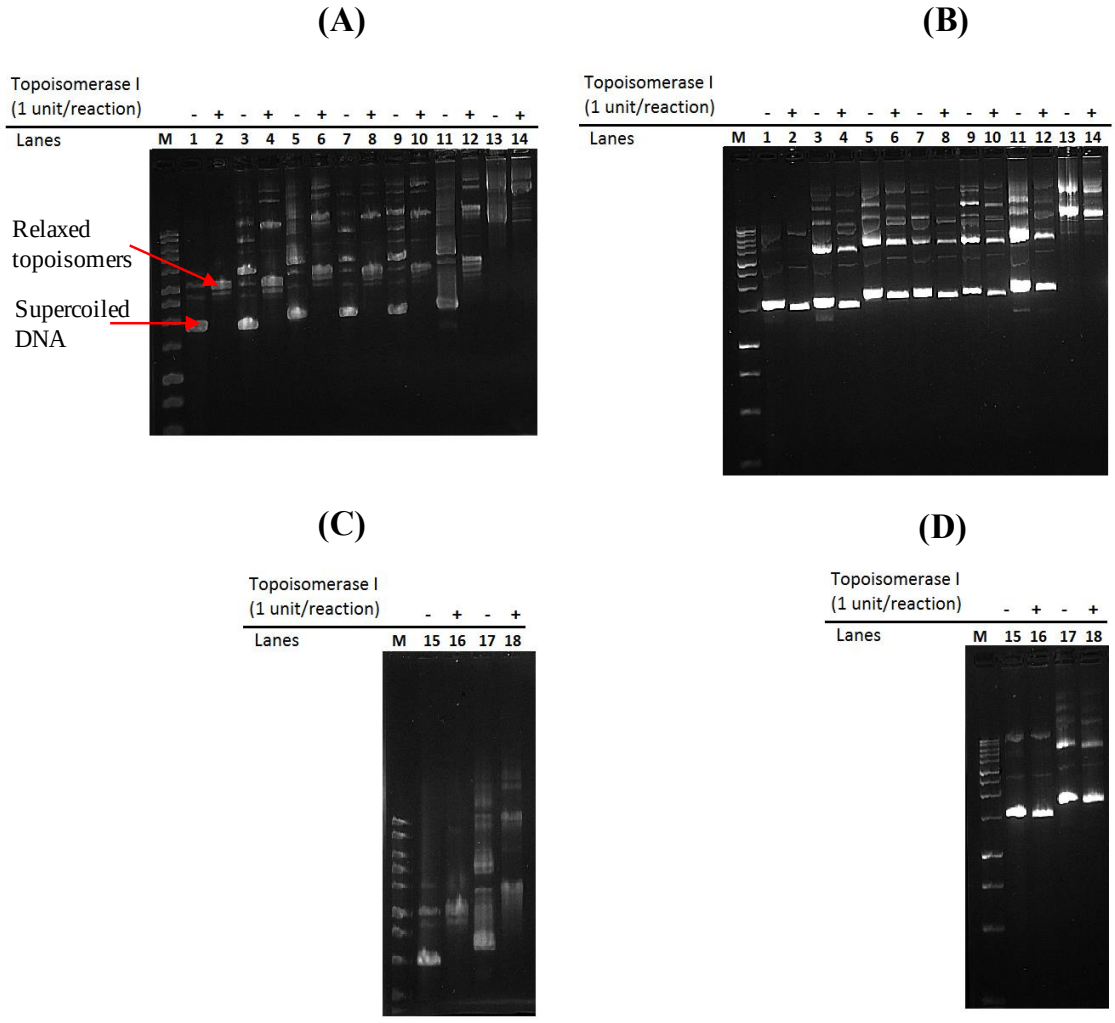



**Figure 3.15:** 0.7% agarose gels of the plasmid supercoil relaxation assay by topoisomerase I. In **(A)** and **(C)**, the reaction products were stained with GelRed after running the gels, whereas in **(B)** and **(D)**, GelRed was included in the gels. Lanes 1, 2, 15 and 16 represent the empty pUC18(0,0). Lanes 3 and 4, represent pUC18(29,0). Lanes 5 and 6 represent pUC18(85,0). Lanes 7 and 8 represent pUC18(29,29). Lanes 9 and 10 represent pUC18(29,29inv.). Lanes 11 and 12 represent pUC18(85,62). Lanes 13 and 14 represent pUC18(85,62)* and lanes 17 and 18 represent pUC18(85,39). Treatment with and without topoisomerase I are indicated by (+) and (-), respectively. Lane M indicates 1 kb DNA ladder.

## 3.5 Investigation of the DNA conformations of the $(GAA)_n$-containing plasmids

Instability of construct plasmids in *E.coli* hosts has been frequently observed, especially with plasmids harbouring multiple repeats. This phenomenon is still poorly understood. However, plasmid instability can result from a number of factors including the expression of toxic gene products, metabolic burden effects, plasmid copy number, genotype of the *E.coli* host strain and genetic recombination across specific sequences in the plasmid (217). Homologous recombination may produce multimeric forms of plasmid that do not segregate properly at bacterial cell division.

It is well known that plasmids can exist in different shapes and forms: closed circular with different degrees of supercoiling, multimeric forms as catenates or concatamers, linear DNA, or other DNA structures such as sticky DNA. We noted that, after isolation, plasmid pUC18(85,62) occasionally migrated much slower than the original construct. As already noted, each plasmid construct appeared as several bands in the agarose gels, and we suggest that these correspond to different forms of the plasmid in the sample (dimers, trimers etc.). In this section, time course digestions with different enzymes (ScaI, BamHI and Acc651) were used to analyse the different plasmid form for each band in plasmids pUC18(85,0) and pUC18(85,62). Limited amounts of each enzyme were added to samples of the plasmid and incubated for different digestion times. The products were then analysed on 0.7% agarose gels. The results were compared with native pUC18 and an isolate of pUC18(85,62) which has unusually slow gel mobility.

### 3.5.1 Results

When small closed circular plasmid DNA is run in agarose gels only one band, corresponding to the supercoiled DNA, is observed. Two other slower migration bands may be observed corresponding to linear DNA and open circular (relaxed) forms. The open circular form is produced by breaking one strand and linear DNA is obtained by cutting two strands at the same position. In addition, plasmid DNA may appear as multimeric forms as mentioned previously. However, in the case of multimeric plasmid forms, it is not at all easy to assign all bands to plasmid forms, since the electrophoretic mobility of different DNA shapes is affected by the electrophoresis operating conditions (89, 218).

### 3.5.1.1 Analysis of the unusual DNA bands seen with plasmid pUC18(85,62)

From the results of time course digestion as shown in the following figures, we identified the prominent DNA forms that always associated with our construct plasmids, and their distribution based on agarose gel electrophoresis. Figure 3.17 shows the time course of digestion of empty pUC18 and pUC18(85,62) with ScaI (2 units/µl). The undigested pUC18 plasmid consists of one clear band (supercoiled DNA) while construct pUC18(85,62) appears as two major bands. As expected, when the pUC18 was digested with ScaI, the supercoiled form gradually disappeared and was replaced with the linear plasmid at approximately 2500 bp (Figure 3.17, lanes 1-5). When pUC18(85,62) was digested under the same conditions, the upper (red arrow) and lower (yellow arrow) bands both gradually disappeared and were replaced with several other bands, though the final product was a single band at about 3,000 base pairs corresponding to the expected linear fragment. At short digestion times a slower-migrating band at approximately 6,000 bp is evident, though this disappears at longer incubation times or higher concentrations of the enzyme (lanes 6 and 7). These results are consistent with the suggestion that the band of lower mobility corresponds to a plasmid dimer, which is first cut to a double length linear fragment that is subsequently cut a second time to produce the usual length monomer.

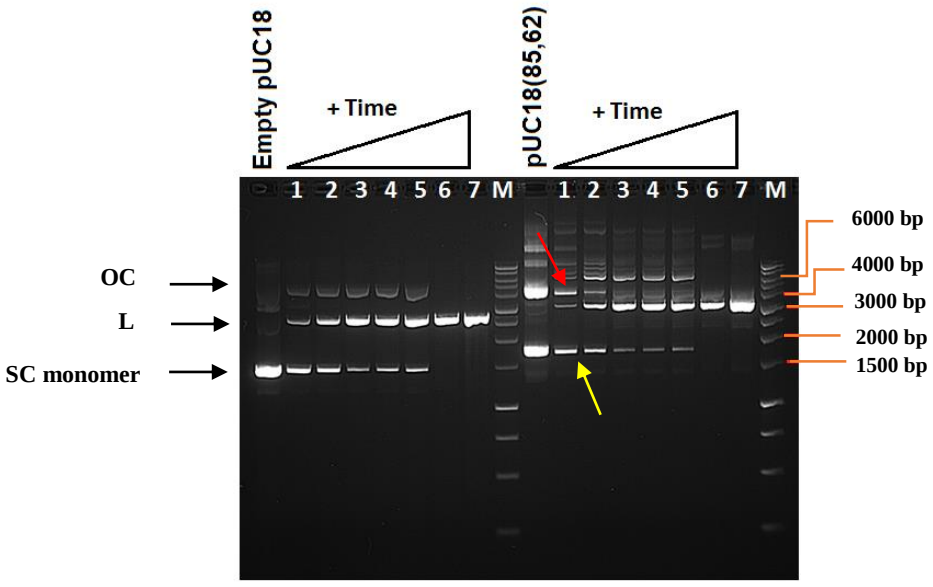



**Figure 3.16:** Time course of cleavage of pUC18 and pUC18(85,62) with ScaI. The products were separated on a 0.7% agarose gel. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively. Lanes 1-6 correspond to digestion by the enzyme for 5 sec, 1, 3, 5, 10 and 20 minutes. Lane 7 was obtained by digesting the plasmid with a high concentration of ScaI.

To confirm this theory, the DNA in the upper and lower bands was extracted from the gel using a QIAquick gel extraction kit (Qiagen). These plasmids were then digested with ScaI in a similar set of experiments and the results are shown in Figure 3.18.

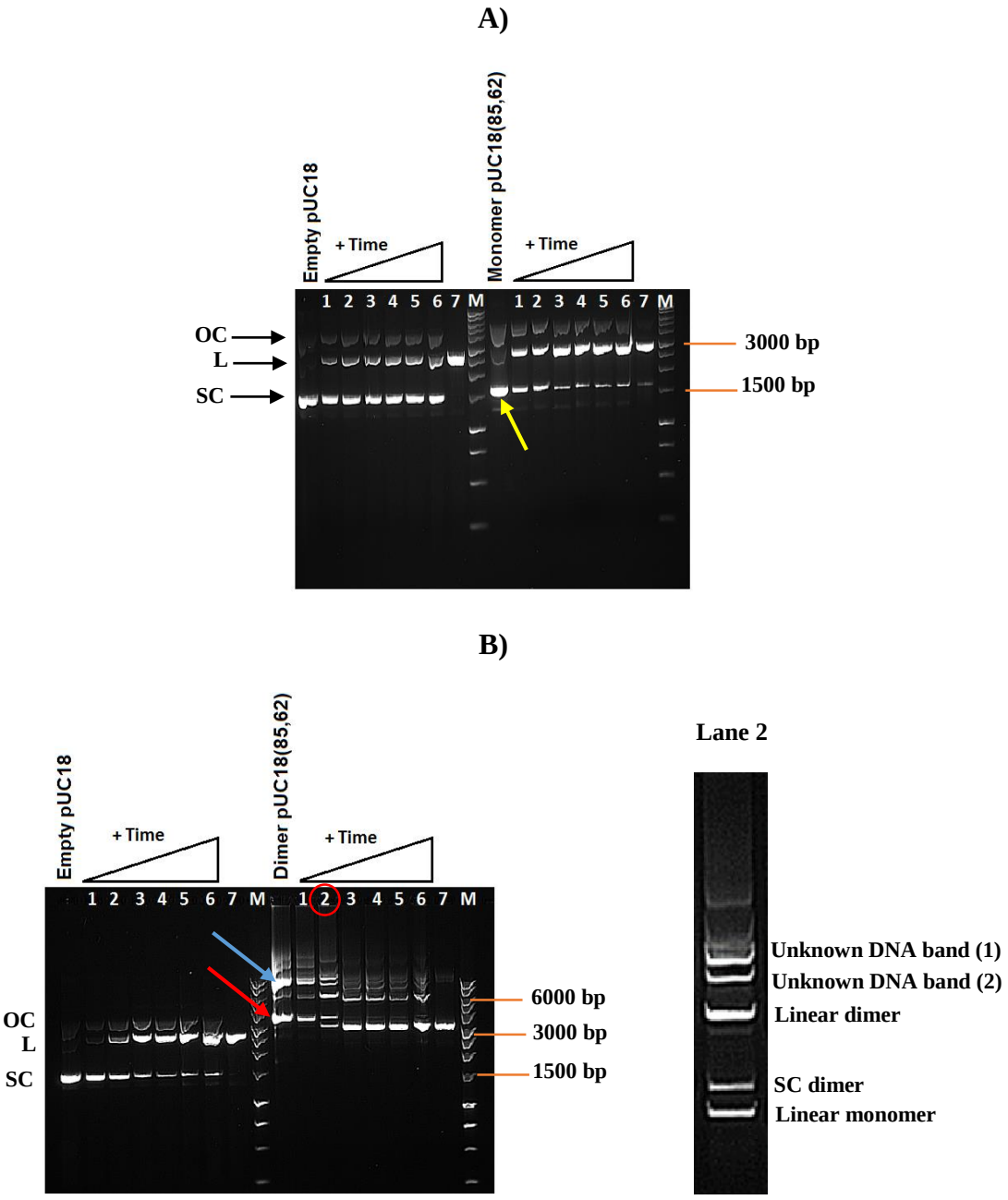



**Figure 3.17:** Illustrates time course digestion of monomer **(A)** and dimer form **(B)** of construct pUC18(85,62) with ScaI. Yellow arrow in **(A)** indicates monomer plasmid. In **(B)** red arrow represents dimer form of the plasmid and blue arrow indicates higher order multimer form (trimer or tetramer). The bands in lane 2 of **(B)** (red circle), were enlarged for clarity as shown in the right lane. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

In Figure 3.18A, it is clear that supercoiled monomer form of the construct pUC18(85,62) (indicated with a yellow arrow) is converted to a linear monomer form at 3000 bp (lanes 1-7). In contrast, the lower mobility dimer plasmid (indicated with a red arrow) in Figure 3.18B, generates further slower bands after extraction (indicated with blue arrow), and is converted to a linear dimer and linear monomer (lane 2), while on longer digestion the only species corresponds to the linear monomer (lane 7). A simple diagram for this cleavage is shown in Figure 3.19. This finding supports our suggestion that the upper band of the original construct pUC18(85,62) (Figure 3.17) corresponds to a supercoiled dimer. The additional slower mobility bands probably correspond to higher order multimers (trimers or tetramers).

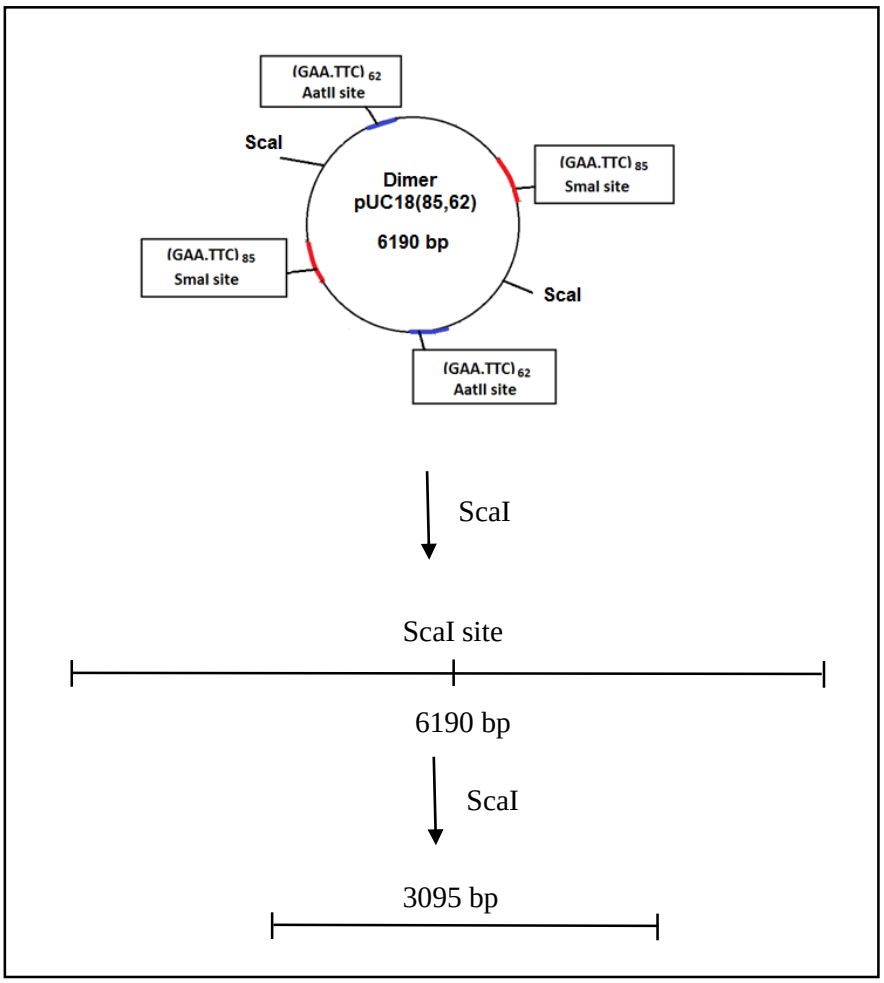



**Figure 3.18:** Schematic representation of two linear DNA fragments that would have been produced by digestion of dimer pUC18(85,62) with ScaI.

Another preparation of plasmid pUC18(85,62) was also obtained, which displayed an even slower gel migration; this also was investigated by digestion with ScaI in the same manner as the previous experiment and the result is shown in Figure 3.20A. The undigested construct pUC18(85,62)* consists of two distinct bands which disappear gradually as the digest proceeds (lanes 1, 2 and 3; right side of the gel). Again this digestion produces multiple bands, which finally resolve into a band that corresponds to the monomeric linear fragment (about 3,000 bp). In this case the bands appear to correspond to monomers (3000 bp), dimers (6000 bp) and trimers (9000 bp). This finding indicates that this version of construct pUC18(85,62) corresponds to a plasmid trimer.

We repeated this experiment using a new sample of this construct and ran the gel as slowly as possible to get better separation. The results are shown in Figure 3.20B. In this instance bands with the mobility corresponding to linear trimers can be clearly seen at 9000 bp (lanes 1-5), moving slower than the uncut plasmid (arrow).

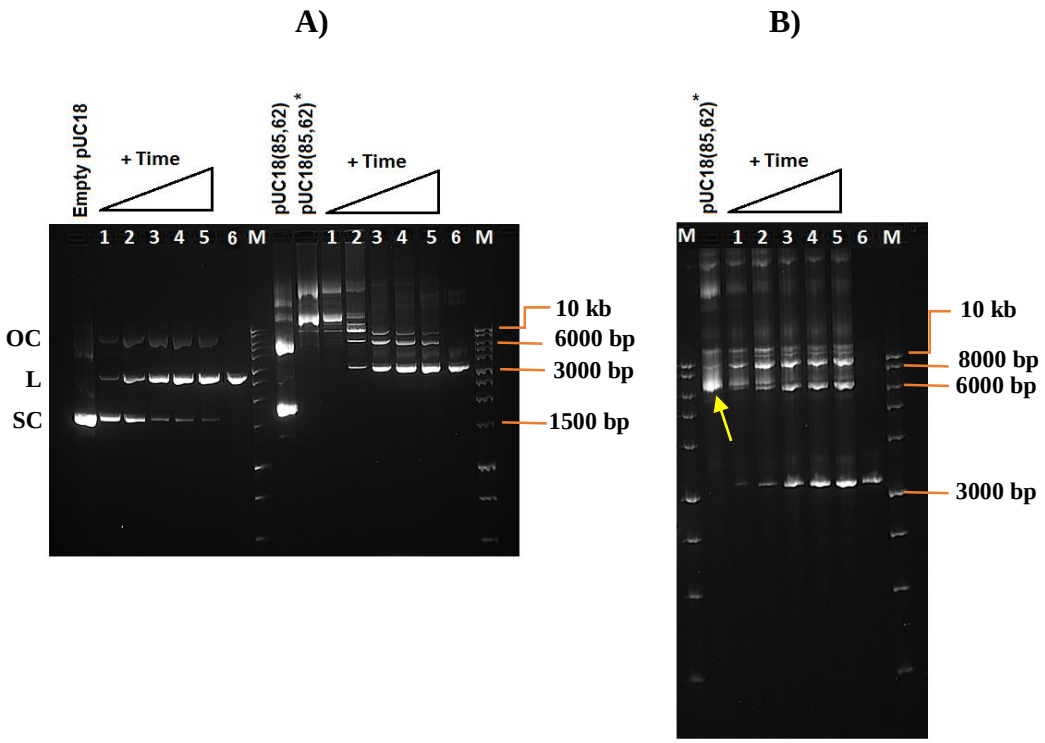



**Figure 3.19:** Time course digestion of two different samples of construct pUC18(85,62); native construct pUC18(85,62) in **(A)** and its unusual version pUC18(85,62)* in **(B)**. Lanes 1-5 in both gels correspond to digestion by the enzyme for 5 sec, 1, 3, 5 and 10 minutes. Lane 6 represent the complete digestion of the plasmid with a high concentration of ScaI. Lane M represents 1 kb DNA ladder. Yellow arrow indicates native construct pUC18(85,62)*. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

Further time course experiments of DNA cleavage were performed with these preparations of pUC18(85,62) using restriction enzymes BamHI and Acc651 to confirm these findings and the results are presented in Figures 3.21 and 3.22. The restriction enzymes BamHI and Acc651 have recognition sequences GGATCC and GGTACC respectively, which are located near at the edge of the insert. We included these enzymes in this analysis since two inserts of GAA repeats can adopt unusual DNA structures under superhelical stress, especially intramolecular triplexes and sticky DNA, which might affect access to their recognition sites.

Looking at Figure 3.21 (A and B) it can be seen that the cleavage patterns of the higher mobility form of pUC18(85,62) with BamHI and Acc651 is similar to that seen with ScaI (Figure 3.17). As the digestion proceeds, slower-migrating bands appear (lanes 1-6) in (A) and (lanes 2-5) in (B) at approximately 6000 bp, in addition to the appearance of faster bands at approximately 750, 2500 and 3000 bp, which increase in intensity at longer digestion times. This observation is consistent with the suggestion that this plasmid is a dimer; a list of the potential cleavage products by these two enzymes is shown in Figure 3.23. Complete digestion by each enzyme (with higher concentration) is shown in (lane 7), in which two clear bands at approximately 750 and 2500 bp can be seen since the monomer of pUC18(85,62) contains two restriction sites for each of these enzymes. This result supports our previous observation that this preparation of plasmid pUC18(85,62) contains a mixture of monomers and dimers. It is also worth noting that BamHI and Acc651 cleavage does not seem to have been affected by the presence of the two GAA repeats in this construct.

A similar experiment was performed with the preparation that displayed an even slower gel mobility, and the results are shown in Figure 3.22. Partial cleavage with BamHI and Acc651, (in lanes 3, 4, 5 and 6 in A and lanes 2, 3 and 4 in B), generated different lengths of faster-migrating bands (multilinear DNA), in addition to a slower band which appeared only in A. These results are consistent with the suggestion that this plasmid preparation consists of a trimer of pUC18(85,62) as shown in Figure 3.24, though we cannot exclude the possibility that this is a tetramer.

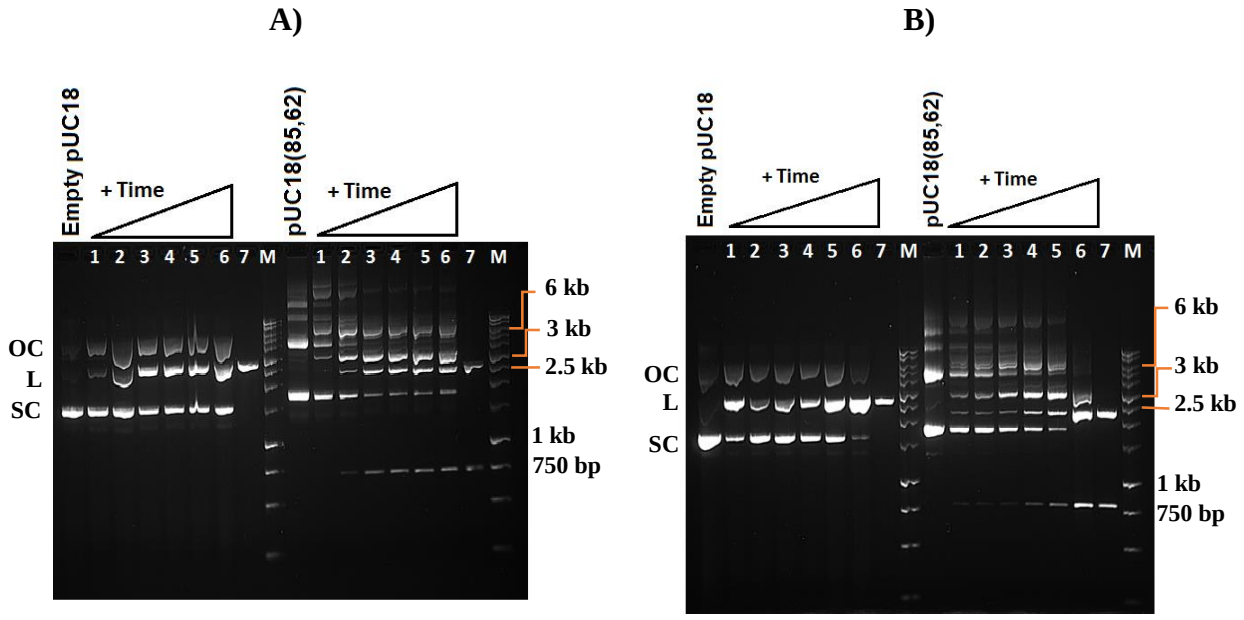



**Figure 3.20:** Time course of digestion by BamHI **(A)** and Acc651 **(B)** of plasmid constructs pUC18(85,62). Digestion of pUC18 was used as a control. Lanes 1-6 correspond to digestion for 5 seconds, 1, 2, 3, 5 and 10 minutes, respectively. Lane 7 represent the complete digestion of the plasmid with a higher concentration of the enzyme. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

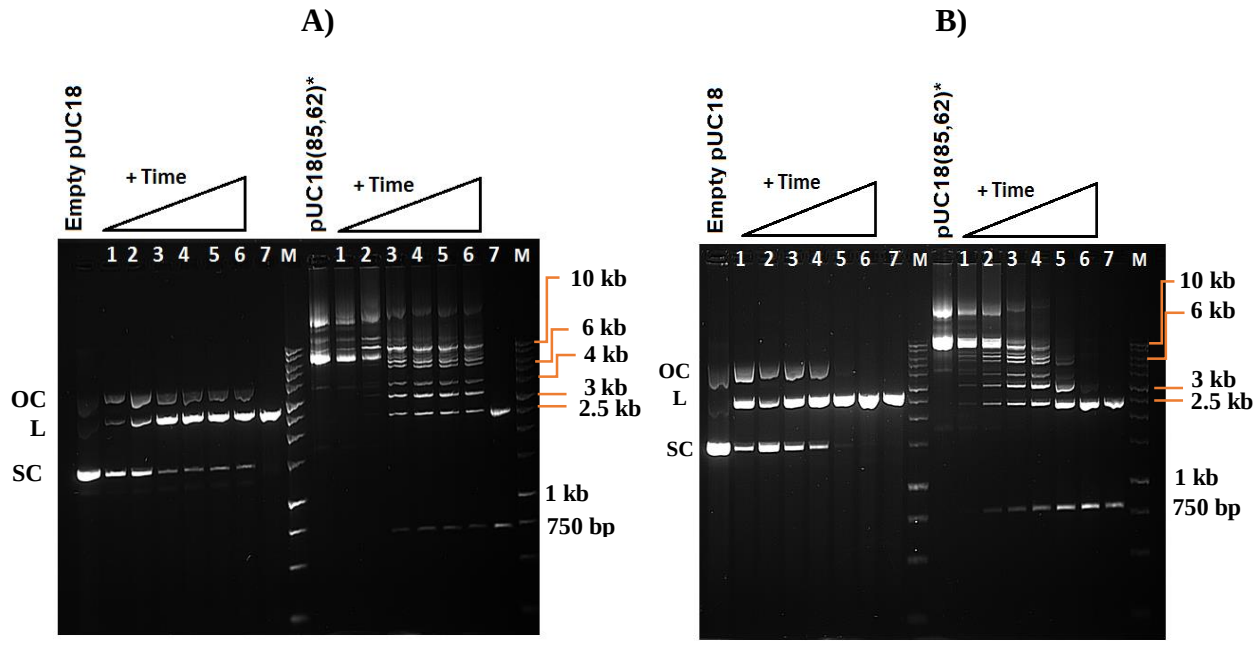



**Figure 3.21:** Time course of BamHI **(A)** and Acc651 **(B)** digestion of the pUC18(85,62) preparation with unusually slow mobility. Lanes 1-6 correspond to digestion for 5 seconds, 1, 2, 3, 5 and 10 minutes. Digestion of pUC18 was used as a control. Lane 7 represent complete digestion of the plasmid with a high concentration of the enzyme. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

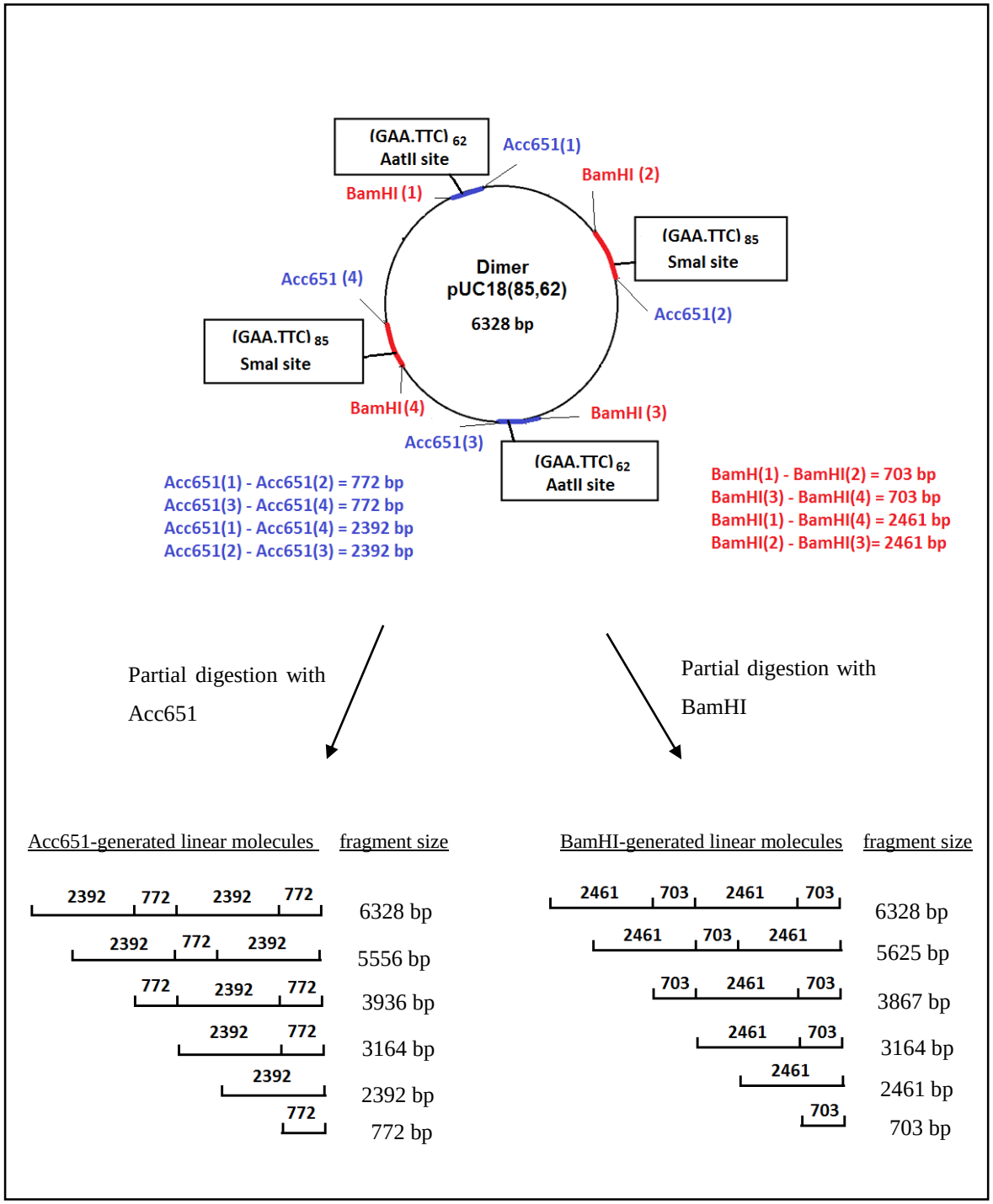



**Figure 3.22:** Restriction map of dimer plasmid construct pUC18(85,62) showing the four cleavage sites for BamHI and Acc651. In addition, schematic representation illustrates lengths of six fragments that result from partial digestion with BamHI or Acc651.

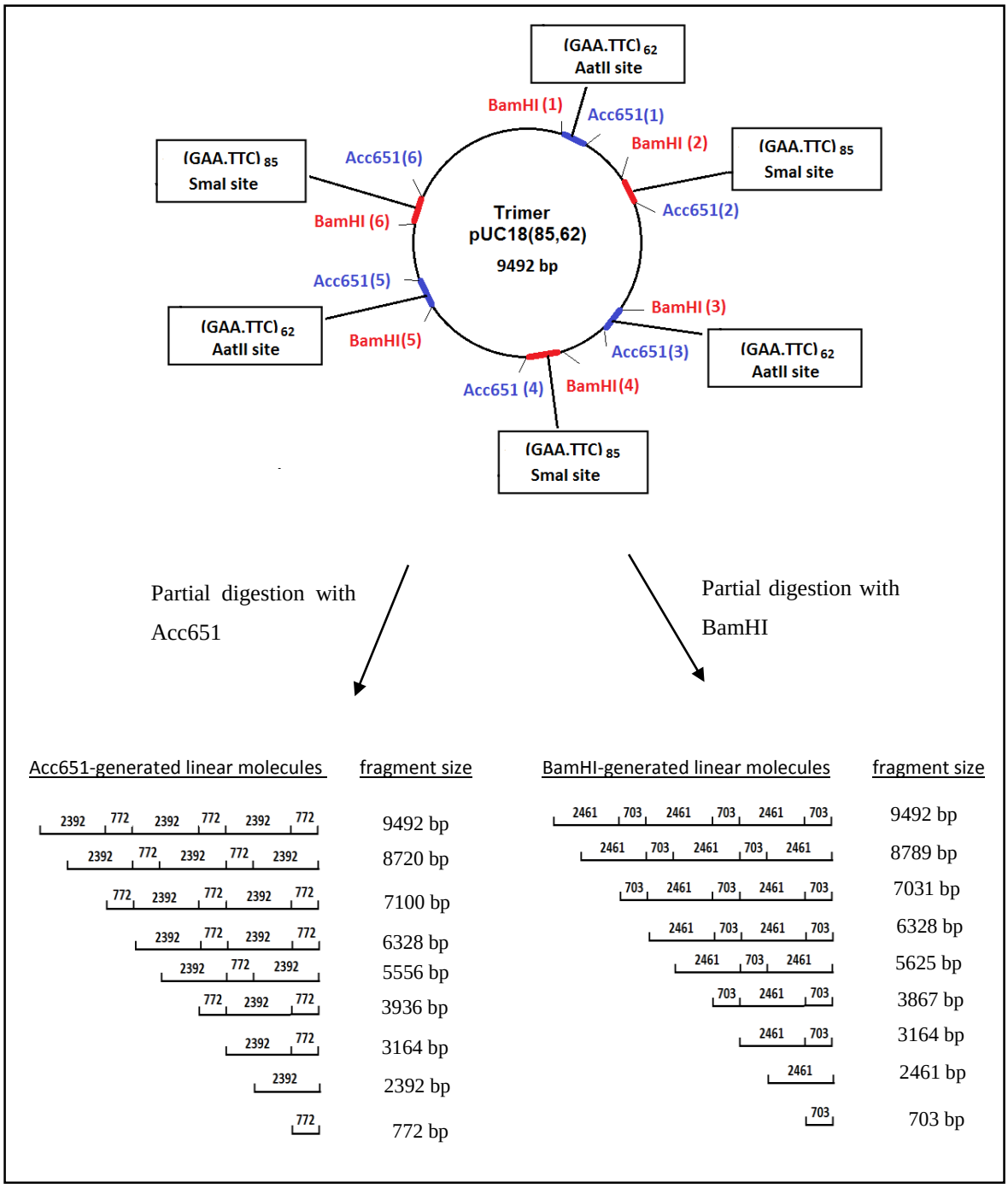



**Figure 3.23:** Restriction map of trimer plasmid construct of pUC18(85,62) showing the six sites for BamHI and Acc651. In addition, schematic representation illustrates lengths of nine fragments that result from partial digestion with BamHI or Acc651.

Plasmid pUC18(85,62) was incubated with 50 mM sodium acetate buffer, pH 5 with 5 mM $MgCl_2$ or 40 mM Tris-acetate buffer, pH 7 with 5 mM $MgCl_2$ at 37 °C overnight to promote sticky DNA formation under superhelical stress. The plasmids were then digested with BamHI (1 unit/µl) at different incubation times to examine whether there was any inhibition of the cleavage due to the formation of an intramolecular triplex or sticky DNA. As shown in Figure 3.25 (A and B), no significant changes were observed between the cleavage patterns at pH 5 and pH 7.

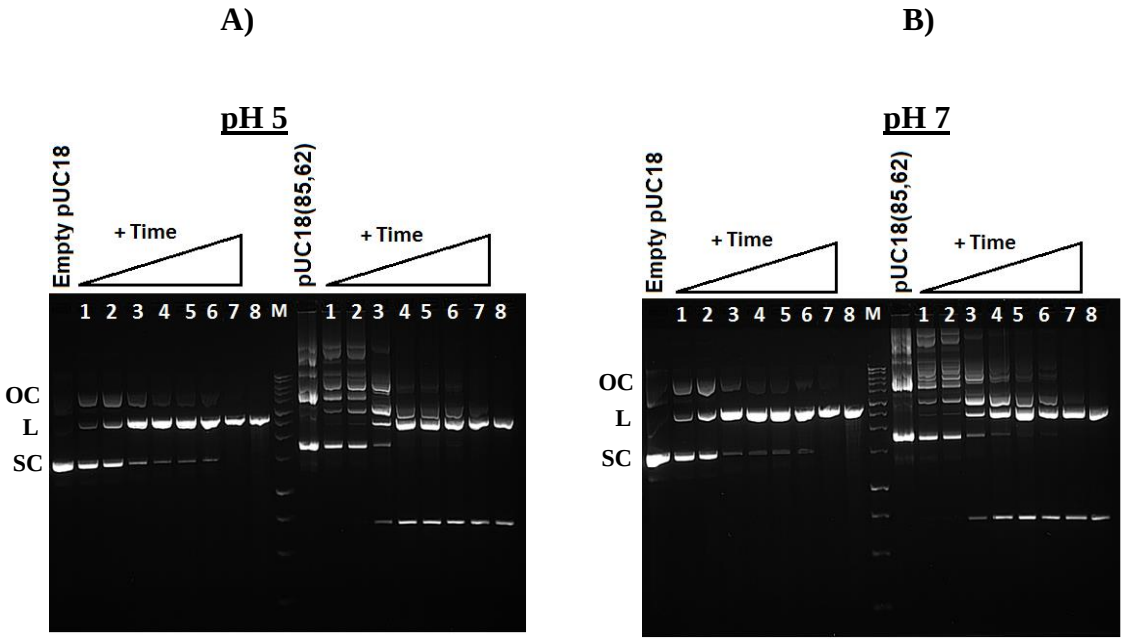



**Figure 3.24:** Time course digestion of construct pUC18(85,62) with BamHI at pH5 **(A)** and pH7 **(B)**. Lanes (1-7) indicates incubation times of 5 seconds, 1, 2, 3, 4, 5 and 10 minutes, respectively. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

We further investigated the oligomeric plasmid forms of plasmid pUC18(85,62), by retransforming each of the plasmid forms (monomer, dimer and oligomer) into SURE cells. Plasmids were isolated from four different colonies from each transformation and analysed by gel electrophoresis, the results are shown in Figure 3.26. It appears that the monomer construct generated different plasmid forms after retransformation (monomer, dimer, trimer and multimers). The dimer and oligomeric DNA forms generated retransformants with the same or slower mobility as the parent plasmid, and did not revert to the monomeric form. We suggest that lanes 5-8 consist of dimers and multimers; lanes

9-12 consist of trimers and multimers, while lanes 1-4 contain all these plasmid forms. This finding suggests that these plasmids (dimers and higher oligomeric forms) cannot revert to smaller plasmids after retransformation, but can further oligomerise.

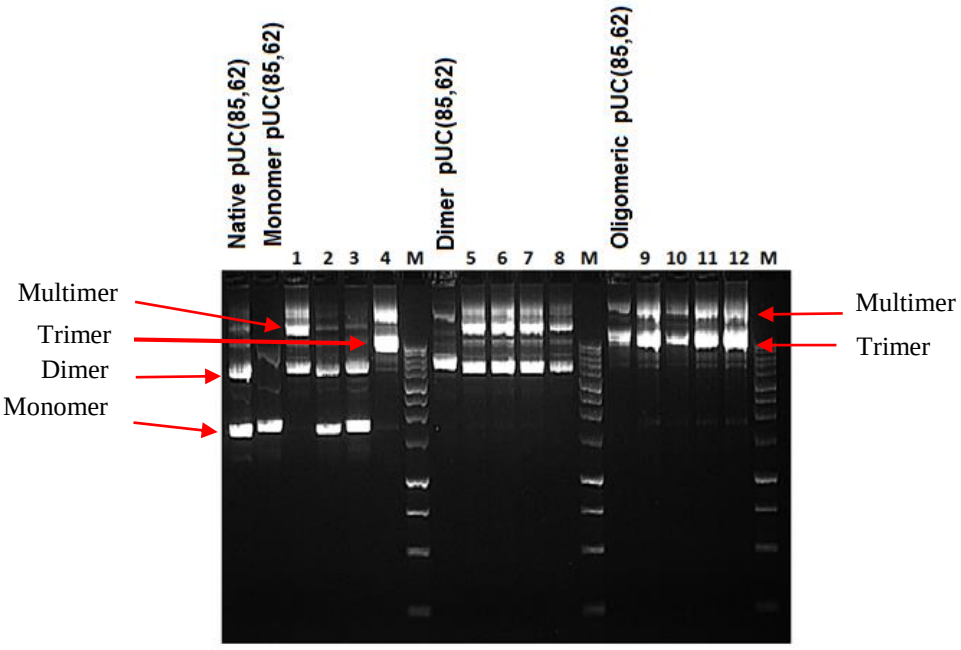



**Figure 3.25:** 0.7% agarose gel electrophoresis of different plasmid forms after retransformation of different forms of pUC18(85,62) plasmids into SURE cells. Lanes (1-4), (5-8) and (9-12) represent migration of four different plasmids after retransformed the monomer, dimer and oligomeric forms in SURE cells, respectively.

### 3.5.1.2 Analysis of the unusual DNA bands seen with pUC18(85,0)

Similar restriction enzyme cleavage mapping experiments were performed with plasmid pUC18(85,0), which only contains one long GAA insert. The undigested construct plasmid consisted of two distinct bands indicated by the blue arrows in Figure 3.27A. When this was cleaved with ScaI for increasing times, the upper and lower bands disappeared gradually. As the digest proceeded, intermediate bands appeared at approximately 6000 bp (lanes 2, 3 and 4), then disappeared after longer incubation. Lane 7 in A and B and lane 6 in C show plasmid DNA completely digested with ScaI, producing one clear band of linear DNA, at approximately 3000 bp. To determine the plasmid form in the upper and lower bands, the DNA molecules were extracted and then examined

separately in the same manner as the previous experiment. Briefly, as shown in Figure 3.27B, the supercoiled monomer form of the undigested construct pUC18(85,0) (yellow arrow) is converted to monomeric linear DNA. The dimer form (indicated with a red arrow) in Figure 3.27C, which contains an additional slower band (indicated with blue arrow), is converted to a linear dimer (lane 2). With high concentration of the enzyme, all the DNA bands are converted to one clear band. This result indicates that the lower band must be supercoiled monomer DNA and the upper band is a dimeric form of the same plasmid. However, we could not identify exactly the plasmid form in the additional slower band (blue arrow), though this probably corresponds to a trimer or tetramer.

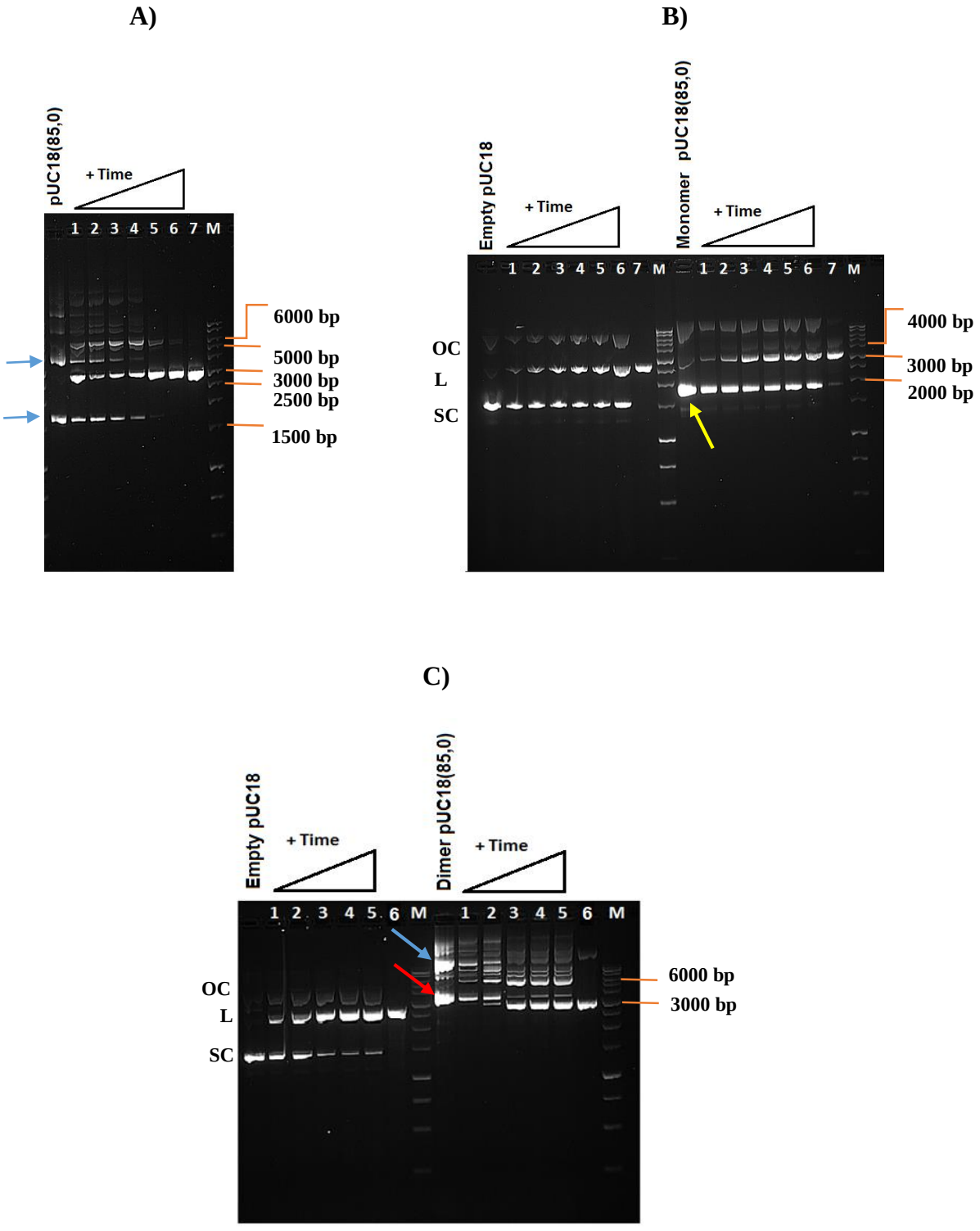



**Figure 3.26:** Time course digestion of native **(A)**, monomer **(B)** and dimer forms **(C)** of construct pUC18(85,0) with ScaI for different times: lanes 1-6 in **(A)** and **(B)** correspond to digestion for 5 seconds, 1, 3, 5, 10 and 20 minutes respectively. Lanes 1-5 in **(C)** correspond to digestion for 5 seconds, 1, 3, 5 and 10 minutes respectively. In **(B)** yellow arrow indicates monomer form of the construct pUC(85,0) and in **(B)** red and blue arrows indicate dimer and higher order oligomeric forms of the plasmid, respectively. Lane M represents 1 kb DNA ladder. OC, L and SC indicate open circular, linear and supercoiled DNA, respectively.

## 3.6 Discussion

GAA.TTC trinucleotide repeats have been shown to adopt abnormal DNA structures such as triplex DNA, sticky DNA and hairpins which may interfere with DNA replication, transcription and recombination (219-222). The results presented in this chapter describe the preparation and characterisation of pUC18 plasmids containing single or double tracts of GAA repeats which were created by cloning extended duplex oligonucleotides of GAA.TTC repeats. In addition, the insert was subcloned into a pGL3-control vector which was used in gene expression assays in Chapter 5. Although the synthetic oligonucleotides $(GAA.TTC)_{10}$ were extended by T4 DNA ligase and PCR, there was limited success in cloning of the triplet repeat using the PCR technique. In general, it is recommended to avoid sequences such as these in PCR experiments as they are known to be prone to forming unusual DNA structures. (223).

GAA repeats were inserted into two different locations in plasmids pUC18 at the SmaI and AatII sites, and in pGL3-control at the SacI and PstI sites. Detection of insertion in the SmaI site was relatively straightforward as this is within the polylinker and disrupts the *lacZ* gene allowing simple blue-white selection. Insertion into the AatII, SacI and PstI sites is not accompanied by any simple detection marker and so many colonies were screened by colony PCR to detect the presence of inserts.

Since at least 59 GAA repeats in two positions of a single plasmid are required for sticky DNA formation (125), it was hoped to construct a family of plasmids harbouring much longer number of triplet repeats to get a better understanding of sticky DNA structure. Although we used synthetic oligunucleotides of $(GAA.TTC)_{10}$ to prepare expanded repeats, we failed to generate very long sequences. Various lengths of $(GAA.TTC)_n$ repeats, where n ranged from10 to 94, were successfully cloned into pUC18 and pGL3 vectors, though the majority of the triplet tracts had less than 50 repeats. The difficulty in obtaining construct plasmids with longer inserts may be attributed to the ability of a single stranded GAA or TTC repeat to form slipped DNA (hairpin) structures during replication in *E.coli,* which may be involved in deletion of nucleotides (Figure 3.29). Longer inserts of GAA repeats are more unstable (more deleted) in *E.coli* compared to shorter inserts. Therefore, only plasmids containing a maximum of 85 GAA repeats were studied.

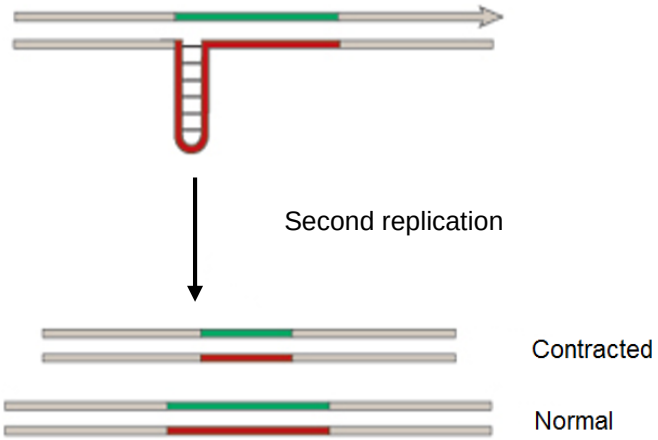



**Figure 3.28:** Schematic representation of hairpin formation during replication. Formation of a hairpin on the template strand can lead to the deletion of the GAA repeat. Adapted from reference (224).

The structural and thermodynamic properties of this hairpin can be assessed using 2-aminopurine substitutions for adenine at various positions in GAA repeat sequence (Figure 3.30). 2-Aminopurine is a fluorescent analogue of adenine that has been shown to only minimally perturb DNA structure and stability, thus it can be used as a convenient probe for investigating triplet repeat sequences. 2-Aminopurine can also pair with cytosine as a guanine-analogue. The local structure of the hairpin can be monitored using the fluorescence intensities of the 2-aminopurines and the changes in the intensity relative to the denatured state (225, 226). Previous studies have shown that the fluorescence from 2-aminopurine in the $(CAG)_n$ repeat loop differs significantly from 2-aminopurines in single strands and 2-aminopurines within a duplex DNA as well as the local positions of the probe within the loop show variations in the emission intensities of the substituted 2-aminopurines. This suggests that stacking and hydrogen-bonding interactions induce significant order within the loop (227).

(A) (B)

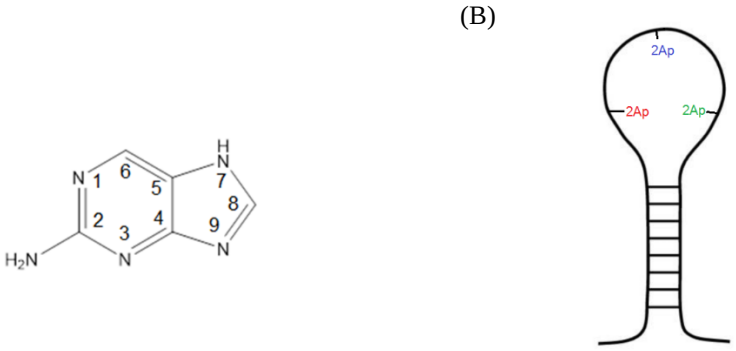



**Figure 3.29:** Illustrates structure of 2-aminopurine (2Ap) **(A)** and substitution of 2Ap in the hairpin DNA formed by GAA repeats **(B)**.

In the initial cloning experiments, instability (deletion) of the inserts in *E. coli* TG2 cells was observed, especially for long inserts (>50 repeats). Therefore, we attempted to transform the cloned plasmids into *E. coli* SURE competent cells that carry mutations which abolish many of the *E. coli* repair genes and therefore prevent rearrangement and deletion of unwanted DNA arrangements. Although *E. coli* SURE was the best strain for stably maintaining the plasmids, deletion products were still observed on some occasions, indicating that SURE cells are not perfect for cloning and maintaining unstable inserts. According to this observation, samples of plasmids that were used for later experiments were routinely re-sequenced to confirm that no deletions had occurred.

Previous studies have demonstrated that sticky DNA can be formed in plasmids containing two $(GAA.TTC)_{60}$ tracts cloned in a direct repeat orientation if they are separated by as little as only 88 bp. This was observed as a retarded band on agarose gels after linearisation of the plasmid (126). However, in contrast, the results presented in this chapter do not provide any evidence for sticky DNA formation in plasmids containing two tracts of 85 and 62 GAA repeats. Since the distance between these tracts of 85 and 62 repeats (501 bp) was much longer than the distance between the tracts in the previous study (88 bp), this should make it easier for our plasmids to form sticky DNA. We have no explanation for these negative results, but maybe longer inserts of GAA repeats are required to form sticky DNA. In addition, plasmids containing two tracts of 85 and 62 GAA repeats were examined by relaxation with topoisomerase I in parallel with different construct plasmids, yielding large relaxation of each plasmid. Each reaction product

showed an almost identical pattern to that of the pUC18 control. This result indicates absence of any gross supercoil-dependent structures, especially sticky DNA.

The time course restriction digestion experiments confirmed that the low mobility plasmid species is a plasmid dimer that results from two monomer plasmids covalently ligated together. We considered whether it might be an interlocked intermolecular dimer in which two plasmid molecules are associated, but this is not consistent with the digestion patterns. This unusual behaviour of the plasmids containing GAA.TTC sequence may be due to the higher recombinogenic capacity of the GAA repeats. In addition, recombination between the GAA.TTC tracts is the main cause of plasmid instability leading to deletions or expansions of sequences. These events might occur during the replication process in which the replication fork transiently pauses at GAA repeats, which could increase DNA break and induce homologous recombination, leading to the formation of larger co-intergrants (220, 228, 229). Figure 3.30 shows a model of homologous recombination between two GAA.TTC repeats from two separated plasmids. According to this model, recombination begins after DNA replication. Stalling of replication forks at GAA repeat region may result in replication fork collapse, followed by introduction of a single nick into each duplex DNA. Next, symmetric exchange of single DNA strands between duplexes occurs to form heteroduplex DNA, forming a Holliday junction. This structure has four-way symmetry and can move down the paired plasmids by a process called branch migration, forming more heteroduplex DNA. The Holliday junction can be resolved in a vertical direction, generating dimer plasmid or in a horizontal direction giving two monomer plasmids. The Holliday model of recombination can demonstrate some of the features of the recombination in the plasmids containing GAA.TTC repeats.

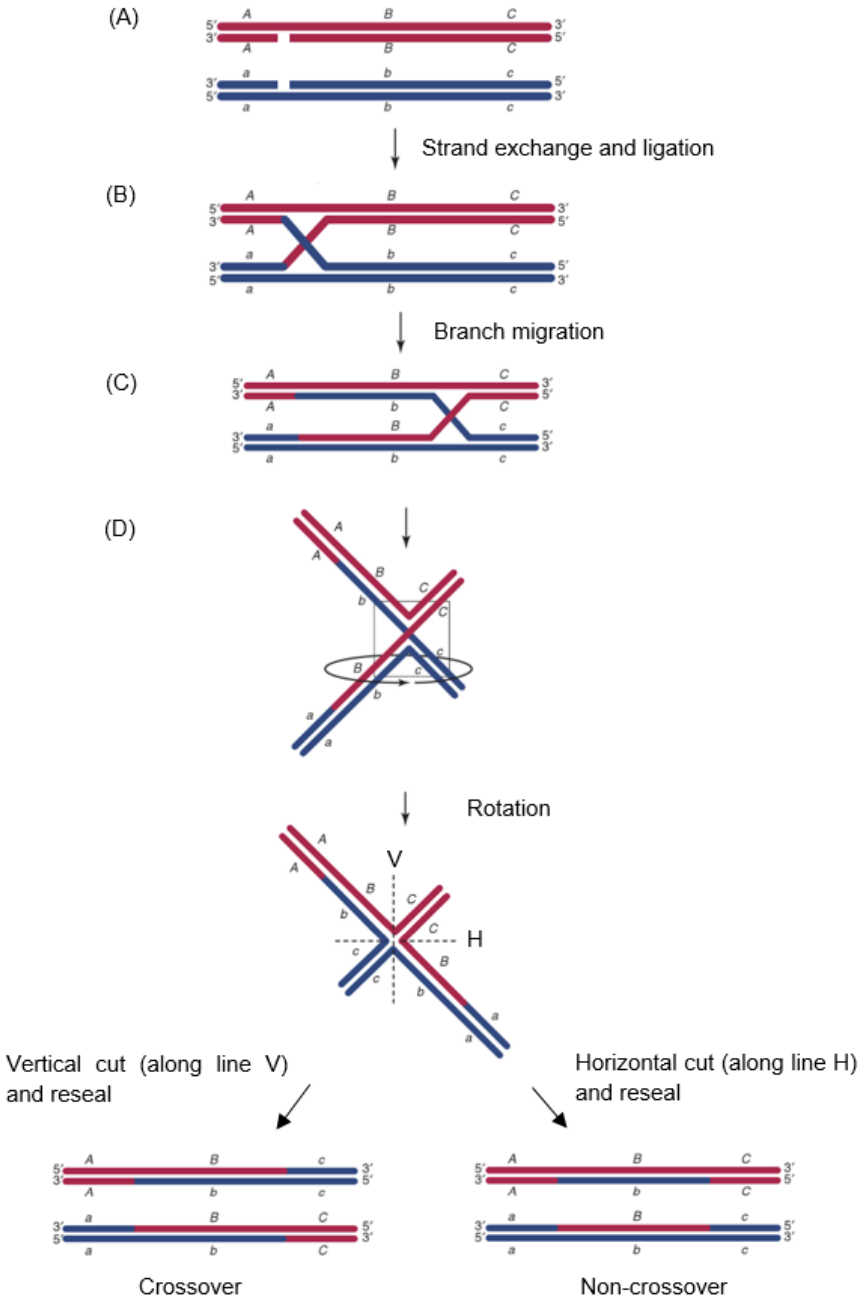



**Figure 3.30:** Pathway of homologous recombination between two GAA repeat regions. **(A)** Symmetric nicks are made in the two plasmids. **(B)** Strand exchange between the plasmids and ligation results in a crossed strand or Holliday junction structure. **(C)** Branch migration of Holliday junction extends the heteroduplex DNA region, which is symmetric and formed in both plasmids. **(D)** The Holliday junction can undergo isomerisation, a rotation of the DNA strands around the crossed strand junction as indicated, to produce the open structure. Resolution of the Holliday junction in a north-south (vertical) direction yields the splice product plasmids on the left, which are crossover for the markers A and C, generating dimer plasmid. Resolution of the Holliday junction in an east-west (horizontal) direction yields the patch plasmids on the right, which are noncrossover of the markers A and C, generating two monomer plasmids. Adapted from reference (230).

# 4 Chemical and Enzymatic Probing of Plasmids Containing GAA repeats

## 4.1 Introduction

It is well known that stretches of polypurine.polypyrimidine (Pur.Pyr) sequences in plasmids can adopt several non-canonical B-DNA structures, depending on their sequences, pH and superhelical density (88, 92, 231-233). These sequences are abundant in eukaryotic genomes and may play important roles in regulating gene function; thus the structural behaviour of poly(Pur.Pyr) sequences in supercoiled plasmids has attracted much interest. GAA or TTC sequences, have two important sequence symmetry elements; as direct repeats they can form slipped-strand DNA structures, though this is most likely to occur on linear DNA (234, 235); $(GAA.TTC)_n$ sequences have mirror repeat symmetry and so have the capacity to form base triplets that can convert B-form (duplex DNA) to non-B-forms (intramolecular triplexes, H-DNA) under superhelical tension (9, 236).

For the formation of an intramolecular triplex the complementary DNA strands of the GAA.TTC sequence must first separate to allow half of the pyrimidine or purine strand to fold back and bind as a third strand to the purine strand in the other half of the Watson-Crick duplex via Hoogsteen base pairing (Chapter 1). The other half of the displaced complementary strand is assumed to be left in an unstructured, single-stranded configuration. The displaced single strand and the loop at the turn of the third strand are therefore hypersensitive to chemical modification and cleavage by single strand specific nucleases. Previous studies have suggested the formation of sticky DNA, which is a combination of two intramolecular triplexes in one closed plasmid (3). Thus, in this work, triplex structures formed by various lengths of GAA.TTC repeats were investigated by using the chemical probes diethylpyrocarbonate (DEPC) and potassium permanganate ($KMnO_4$) and the enzyme S1 nuclease which react with single stranded DNA regions.

## 4.2 Investigation of intramolecular triplex structures adopted by GAA.TTC repeats using chemical and enzymatic probes

Conformational changes of helical DNA, including triplexes, induced by negative supercoiling can be identified and characterised by either chemical reactivity or enzyme cleavage (237). Chemical probes, which react with the DNA bases without causing strand cleavage, are the most effective reagents for identifying local structural transitions in supercoiled DNA, while enzyme nuclease probes are less effective due to the release of the topological constraints with the first cleavage event (237). The most important advantage of using chemical probes is their specificity for reacting with specific bases. In addition, they can be used in varying temperature, pH and ionic conditions. Chemical modification by DEPC or $KMnO_4$ does not generate strand breaks but modifies the bases, and subsequent treatment with hot piperidine is required to produce cleavage, while enzymatic probing uses a single-strand specific nuclease (S1) to cleave unpaired regions of DNA. The reaction products are then analysed in sequencing gels according to Maxam and Gilbert reactions, and this approach has been widely used to probe the formation triplexes *in vitro* (93, 180, 193, 195, 198).

## 4.3 Experimental design

This chapter will focus on the two different techniques employed to examine intramolecular triplex formation in GAA.TC repeats. The first technique used chemicals (DEPC and $KMnO_4$), and enzyme S1 nuclease, as probes to examine the location of exposed or single strand nucleotides that accompany the formation of intramolecular triplexes in supercoiled DNA. The second technique used S1 nuclease to map the single stranded regions resulting from any structural transitions. S1 mapping is a low resolution technique, while DEPC, permanganate and S1 cleavage give base specific resolution. By combining the information obtained from these techniques, we are able to provide strong evidence for the presence of intramolecular triplex formation and identify the structural details. These techniques are fully described in Chapter 2, Section 2.2.6. A set of plasmid constructs, as described in earlier chapters, having one or two inserts of GAA repeats (Table 4.1), was used for these experiments. These are also called pUC18(X,Y) where X is the GAA-repeat number in the SmaI site and Y is the number in the AatII site.

**Table 4.1:** List of construct plasmids used in chemical and enzymatic probing assays.

| Plasmid | Number of repeats in the SmaI site | Number of repeats in the AatII site |
|---|---|---|
| pUC18(19,0) | 19 | 0 |
| pUC18(22,0) | 22 | 0 |
| pUC18(29,0) | 29 | 0 |
| pUC18(29,29) | 29 | 29 |
| pUC18(29,29inv) | 29 | 29 |
| pUC18(85,0) | 85 | 0 |
| pUC18(85,62) | 85 | 62 |
| pUC18* (85,62) | 85 | 62 |
| pUC18(85,39) | 85 | 39 |

inv = inverted repeat

* = unusual version of construct pUC18(85,62)

### 4.3.1 Chemical probing assay

Various lengths of (GAA.TTC) repeats were investigated at single nucleotide resolution using the chemical probes DEPC and $KMnO_4$ to examine the existence of intramolecular triplexes under superhelical stress *in vitro*. DEPC reacts with unpaired purines, especially adenosines, while $KMnO_4$ reacts with pyrimidine bases, primarily thymidine residues that are located in single-stranded regions (238, 239). Both of these reagents have very low reactivity with native B-DNA. The use of both DEPC and $KMnO_4$ provided a complete analysis of both GAA- and TTC-containing sequences in DNA.

For these experiments, the native supercoiled pUC plasmids containing different lengths of GAA repeats were purified by alkaline lysis using Qiagen minipreps. The construct plasmids were then incubated in buffer at pH 5 or pH 7 (see Materials and Methods) to allow formation of intramolecular triplex structures under native superhelical density, followed by treatment with the chemical reagent. These were then cleaved with EcoRI and PstI or Hind III and SacI, (depending on which strand was being examined), to release a short fragment of DNA containing the insert. These fragments were uniquely labelled

with $^{32}P$ at one end (filling in the EcoRI or HindIII sites), followed by cleavage with 10% piperidine. The cleavage products were resolved on denaturing polyacrylamide gels. Since intramolecular triplex formation within (GAA.TTC) insert is supercoiled-dependent, we also released linear DNA of the insert prior to treatment with the chemical reagents to compare the supercoiled with linear DNA (for further details see Chapter 2, Section 2.2.6).

#### 4.3.1.1 Chemical modification with DEPC

DEPC is a small molecule that can be used as a chemical probe to detect structural changes in a DNA due to its ability to attack nucleophilic centres in purines (especially N7 adenine > guanine) when they are not in a B-DNA conformation. Purines in double stranded B-DNA are not accessible to modification by DEPC because of the close stacking of neighbouring bases. When this stacking is altered and the DNA conformation deviates from B-form (such as strand separation or bending) the purines become more susceptible to DEPC reaction. DEPC reacts with N-7 atom of unpaired purine ring or purines in the *syn* conformation as in Z-DNA. This leads to a ring-opened product and scission of the glycosidic bond between the sugar and the base, generating a piperidine-sensitive site (Figure 4.1). Strand cleavage of DEPC-modified bases is then achieved by treatment with hot piperidine which breaks the phosphodiester bond displacing the modified bases (212, 240, 241).

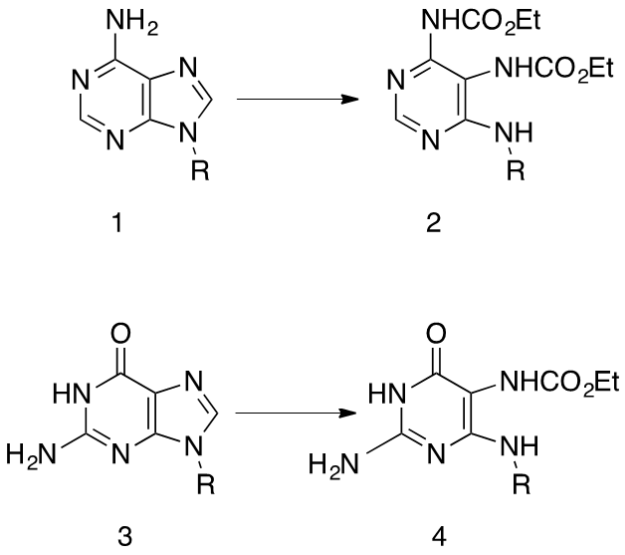



**Figure 4.1:** Diethylpyrocarbonate modification of adenine and guanine. Unpaired adenine (1) reacts with DEPC to give the ring-opened dicarbethoxylated derivative (2) as the major product. Guanine (3) reacts with DEPC to give the ring-opened product (4) with a several fold lower yield than adenine (240).

#### 4.3.1.2 Chemical modification with $KMnO_4$

$KMnO_4$ is one of the most effective and versatile chemicals to study sequence-dependent variations in B-DNA structures (190-192). It oxidises pyrimidine bases, especially unpaired thymines >> cytosines > guanines > adenines. Thus, thymines in the looped portions of intramolecular triplexes are more susceptible to $KMnO_4$ modification. $KMnO_4$ attacks exposed thymine rings at the C5-C6 double bond by out-of-lane attack, oxidising them to vicinal diols, to produce piperidine-labile 5-hydroxy-6-keto derivative (Figure 4.2). Although the ring remains intact, there is a loss of aromaticity due to insertion of the hydroxyl groups on the 5 and 6 carbons. Exposing the vicinal diol to hot piperidine leads to ring opening and cleavage of the phosphodiester backbone (211, 242).

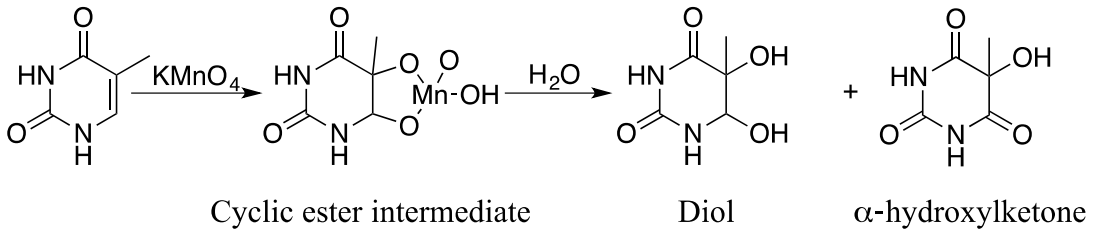



**Figure 4.2:** A schematic reaction of thymine with potassium permanganate (192). This reaction is sterically inhibited by stacking interactions in B-DNA.

### 4.3.2 S1 nuclease

S1 nuclease is a single strand specific endonuclease isolated from *Aspergillus oryzae* which can be used to detect and characterise small variations in the secondary structures of nucleic acids (243-245). The main advantage in using S1 nuclease to explore local altered DNA structures is its high specificity for single-stranded DNA, even at high concentrations of the enzyme (245). Another advantage is that this enzyme is sequence–independent while other endonucleases, such as restriction enzymes cut DNA at specific sequences (237). However, the drawback of this enzyme is that S1 works only in acidic media (pH optimum 4.5 – 5.0). Although acidic pH is ideal to form protonated triplexes, this property is not relevant to physiological conditions. In this research S1 nuclease was employed for two different purposes. Firstly, as a probe to determine single strand regions of the H-DNA and analysed on denaturing polyacrylamide gels. Secondly, S1 was used to map the single strand regions that accompany H-DNA in supercoiled DNA and analysed on agarose gel.

## 4.4 Results

### 4.4.1 Identification of the intramolecular triplex regions within supercoiled DNA by using S1 nuclease

The ability of S1 nuclease to cleave single stranded DNA provides strong evidence for the presence of any single stranded regions that are generated by intramolecular triplex formation within our construct plasmids. As described in Chapter 2, each supercoiled plasmid was treated with S1 nuclease and then cut with enzyme ScaI which has a unique site on the plasmid. Identical samples of these plasmids were first linearised with ScaI (to eliminate any supercoil-dependent single-stranded regions) and then exposed to S1 nuclease. These products were compared with plasmid samples that had been digested with both ScaI and EcoRI; as the EcoRI restriction site is close to the location of the GAA repeats, this should give similar length fragments to those produced by S1 and ScaI.

#### 4.4.1.1 Digestion of native pUC18 with S1 nuclease

Before examining the properties of plasmids containing GAA-repeats the S1 nuclease technique was tested on the native pUC18, as a control, and to aid mapping the S1 nuclease cleavage sites GAA-containing constructs. S1 nuclease first recognises the single stranded region and nicks the plasmid at this site. In the second step, S1 recognises the nick position and cleaves the opposing strand producing linear DNA (246). The results of this experiment are shown in Figure 4.3. When supercoiled vector pUC18 (lane 1) was exposed to S1 nuclease, a mixture of nicked and linear DNA fragments was observed, in addition to the presence of the supercoiled DNA (lane 2). The cleaved products were then digested with ScaI, resulting in the appearance of two fragments at approximately 750 bp and 1900 bp. After commencing these studies we realised that pUC18 contains a palindromic sequence around position 1397 bp, which can adopt a cruciform structure under superhelical stress (94, 244, 247, 248). The banding pattern observed after S1 cleavage of native pUC18, therefore probably reflects cruciform formation within this plasmid. When the supercoiled plasmid was first linearised with ScaI, followed by digestion with S1 nuclease, the expected single linear DNA was observed at approximately 2700 bp (lane 4); no additional bands were observed when this was treated with S1 nuclease. This result indicates that vector pUC18 itself contains a supercoil-dependent S1 nuclease sensitive site that is removed after linearisation of the plasmid. Although the presence of this cruciform site will complicate analysis and interpretation

of the S1 mapping experiments on the GAA-containing plasmids, the cruciform cleavage products have different fragment sizes to those predicted for the GAA-plasmids. It will therefore be interesting to see whether the plasmid constructs demonstrate H-DNA or cruciform formation (or both), as extrusion of an intramolecular triplex might reduce the superhelical stress and so prevent cruciform formation and *vice versa*. It should also be noted that one nick with S1 nuclease will remove the supercoiling. If this occurs first at one or other structure then it will remove the formation of the other one.

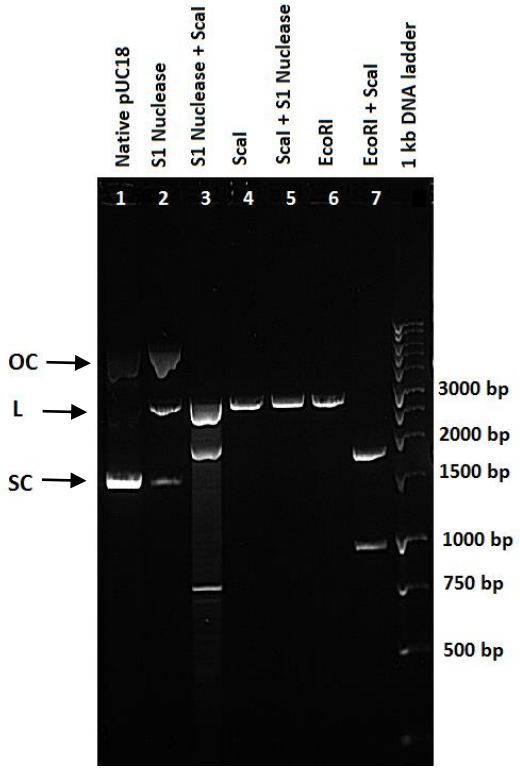



**Figure 4.3:** Mapping of S1 nuclease cleavage sites in pUC18. Lane 1 corresponds to the intact plasmid. Lane 2 corresponds to digestion of the plasmid with S1 nuclease. In lane 3, the S1 nuclease treated plasmid is cut with ScaI. Lane 4 is a ScaI restriction digest. Lane 5 is the ScaI-linearised DNA followed by exposure to S1 nuclease. Lane 6 is an EcoRI restriction digestion product and lane 7 is the digestion product of EcoRI and ScaI. OC, L and SC indicate open circular, linear and supercoiled DNA respectively.

#### 4.4.1.2 Detection of intramolecular triplex region in plasmids containing short or long insert of (GAA.TTC) repeat.

It is known that intramolecular triplexes are usually sensitive to cleavage by S1 nuclease since they contain regions of single stranded DNA. Figure 4.4A shows S1 nuclease cleavage of plasmid pUC18(29,0) containing 29 GAA repeats in the SmaI site. When the supercoiled plasmid (lane 1) was treated with S1 nuclease, the supercoiled DNA

disappeared and linear species were observed (lane 2). Following digestion with ScaI, two distinct DNA fragments could be seen at approximately 1000 bp and 1800 bp (lane 3). In contrast, when the construct plasmid was first linearised by cutting with ScaI (lane 4) and then treated with S1 nuclease (lane 5), only one band was seen, implying that, as expected, negative supercoiling was required to generate the region-specific single stranded DNA. If the S1 site is in the region expected to form an intermolecular triplex, the products of S1 and ScaI cleavage should generate DNA fragments of 943 and 1743 bp, which is compatible with the sizes of the bands seen in lane 3. This indicates that S1 nuclease was able to locate an unusual DNA structure containing single-stranded regions within our construct plasmids. Furthermore, that bands corresponding to cleavage of the pUC18 cruciform are absent.

The S1 nuclease experiment with plasmid pUC18(85,0) displayed similar results to those presented above for the plasmid harbouring 29 GAA repeats. As shown in Figure 4.4B supercoiled construct plasmid (lane 1) was cut with S1 nuclease and produced linear DNA (lane 2), again, indicating the presence of S1 nuclease sensitive site (s). After digestion with ScaI, two smeared bands can be seen at approximately 1000 bp and 1800 bp (lane 3). This confirms that S1 cuts the plasmid in the correct location but suggests that there are several available cleavage sites within this region. As the $(GAA.TTC)_{85}$ insert covers 250 base pairs a single stranded loop might be able to form in more than one precise position. Thus, random double stranded breaks would result in smear of DNA fragments, as is seen after digestion with ScaI. When the plasmid was first cut with ScaI and then cleaved with S1 nuclease only one band was visible, confirming that formation of the single-stranded region requires superhelical stress.

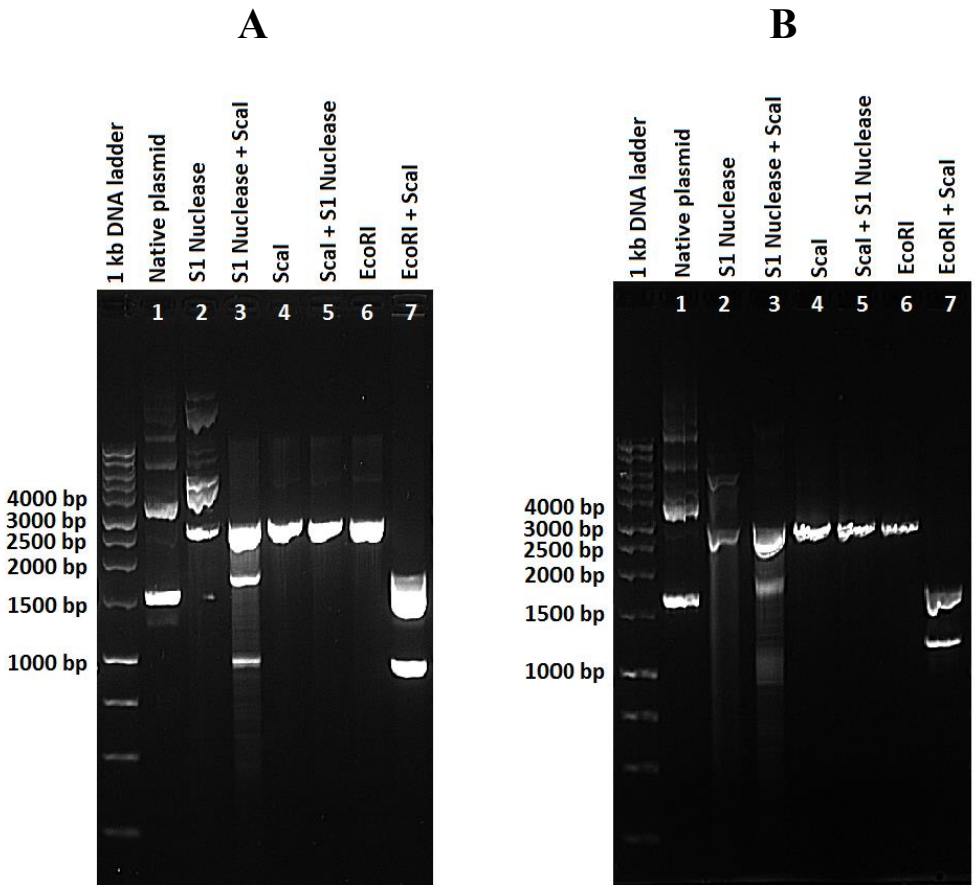



**Figure 4.4:** Mapping of S1 nuclease cleavage sites in plasmids containing 29 GAA repeats **(A)** and 85 GAA repeats **(B)**. Lane 1 corresponds to the intact plasmid. Lane 2 corresponds to digestion of the plasmid with S1 nuclease. In lane 3, the S1 nuclease treated plasmid was cut with ScaI. Lane 4 is a ScaI restriction digest. Lane 5 is the ScaI-linearised DNA followed by digestion with S1 nuclease. Lane 6 is EcoRI restriction digestion and lane 7 is the digestion product of EcoRI and ScaI. OC, L and SC indicate open circular, linear and supercoiled DNA respectively.

### 4.4.2 Chemical modification with DEPC and $KMnO_4$

The reactions with DEPC and $KMnO_4$ were used to confirm intramolecular triplex formation in the GAA.TTC repeat-containing plasmids. The chemically modified supercoiled DNAs were cleaved by EcoRI and PstI (GAA strand) or HindIII and SacI (TTC strand) to study H-DNA formation at the insert in the SmaI site. H-DNA formation in the AatII site of the plasmids pUC18(29,29) and pUC18(29,29inv) were determined by digestion with EcoRI and AatII (GAA strand), since one of the primers used in subcloning the insert contains an EcoRI restriction site as shown in Chapter 2, Section 2.2.2.2. For short $(GAA.TTC)_n$ repeats (n = 19, 22, 29), this technique confirmed the formation of intramolecular triplex structures at pH 5 and pH 7 at single nucleotide resolution, but longer repeats ($n \geq 39$) were too large for a definitive chemical probe

characterisation. Before considering the results, it is worth noting that the GAA and TTC strands run in opposite orientation; the 5'-end of one triplet repeat is aligned with the 3'-end of the other. As we labelled the 3'-end of either strand, the bands of one strand (GAA) in each gel were numbered from bottom to top (low to high numbers) and the complementary strand (TTC) from bottom to top (high to low numbers). The intensities of these bands were quantified by densitometer scans using Image Quant software. Some of double adenines (AAs) and double thymines (TTs) produced overlapping peaks for which it was not possible to resolve individual bases; in these instance we used DNA histograms in which the intensities of each double adenine (AA) or double thymine (TTs) were combined into a single bar. These histograms are presented above (DEPC) and below ($KMnO_4$) the DNA sequence. In addition, any bars corresponding to Gs in the (GAA) strand or Cs in (TTC) strand were ignored as they are much smaller and probably correspond to non-specific cleavage during the experiments, as these probes are specific for A>>G or T>>C, respectively.

Figure 4.5A shows the DEPC and $KMnO_4$-modification patterns of the strands containing $(GAA)_{19}$ and $(TTC)_{19}$, respectively, at two different pH values. Looking first at DEPC modification of the GAA strand, it can be seen that while there is very little reaction with the linear DNA (lane 2) there is considerable cleavage at adenines in the central part of the GAA tract with the supercoiled DNA at both pH 5 and pH 7 (lanes 3 and 4). The most reactive adenines are located at $6^{th}$, $7^{th}$, $8^{th}$ and $9^{th}$ repeats from the 3' end, indicating predominant sites of folding of the purine strand in the intramolecular triplex. The neighbouring adenines are less reactive towards DEPC at both pH 5 and pH 7. $KMnO_4$ probing of the TTC strand at pH 5 and pH 7 also shows one main modification region of thymine residues as seen in lanes 7 and 8, which is more obvious in lane 7. In lane 7, the centre of the sequence was more exposed to permanganate modification, especially at $9^{th}$, $10^{th}$, $11^{th}$ and $12^{th}$ repeats. Additional hyperreactivity can be seen around position 15 which indicates the presence of structural distortion at the junction between the triplex and duplex DNA (for further details see Discussion). Furthermore, it can be seen that the number of bases that react with DEPC (lanes 3 and 4) are more than those exposed to $KMnO_4$ reaction (lanes 7 and 8), suggesting that half of the TTC stretch folds back to form the Hoogsteen-paired third strand of the triplex, leaving nearly half of the GAA strand displaced. The strong modification at the centre of the TTC strand may reflect the formation of a short single stranded loop (hairpin) in the triplex structure.

In lane 7 at positions 1 to 7, it can be seen clearly that the first thymine residue of each TTC repeat is more sensitive to $KMnO_4$ modification than the second one. This observation can be attributed to the methyl group of the first thymine residue, which diminishes the accessibility of the 5-6 double bond of the neighbouring thymine. Previous studies have demonstrated that the thymine residue at 3'-terminus of double stranded DNA is more sensitive to chemical probing than the other thymines in the same strand; this is due to shielding of the reactive site by the methyl group of the 3'-neighbouring thymine (249). Thus, the first thymine residue at repeat 19 in lane 7 close to 3'-end is more sensitive to permanganate than thymines at repeats 18-15.

Figure 4.5B shows the densitometric scans of the DEPC and $KMnO_4$ modified lanes. For the DEPC trace, it runs 5'- to 3'- from left to right, corresponding to bottom to top of the gel (lanes 3 and 4), while $KMnO_4$ runs 3'- to 5'- from left to right, corresponding to bottom to top of the gel (lanes 7 and 8). Analysis of these data indicates that $(GAA.TTC)_{19}$ adopts an intramolecular triplex (class PyPu.Py and isomer Hy5), since the 3'-half of the GAA sequence shows extensive DEPC modification and specific $KMnO_4$ modification of thymines is localised at the centre of the sequence. A schematic representation of a model for this H-DNA is presented in Figure 4.5C. In this model, the 5'-half of the TTC sequence (third strand) binds to the target purine strand in a parallel orientation forming Hoogsteen base pairs. This type of triplex is characterised by the formation of TA.T and $CG.C^+$ triplets which requires base protonation to stabilise this triplex.

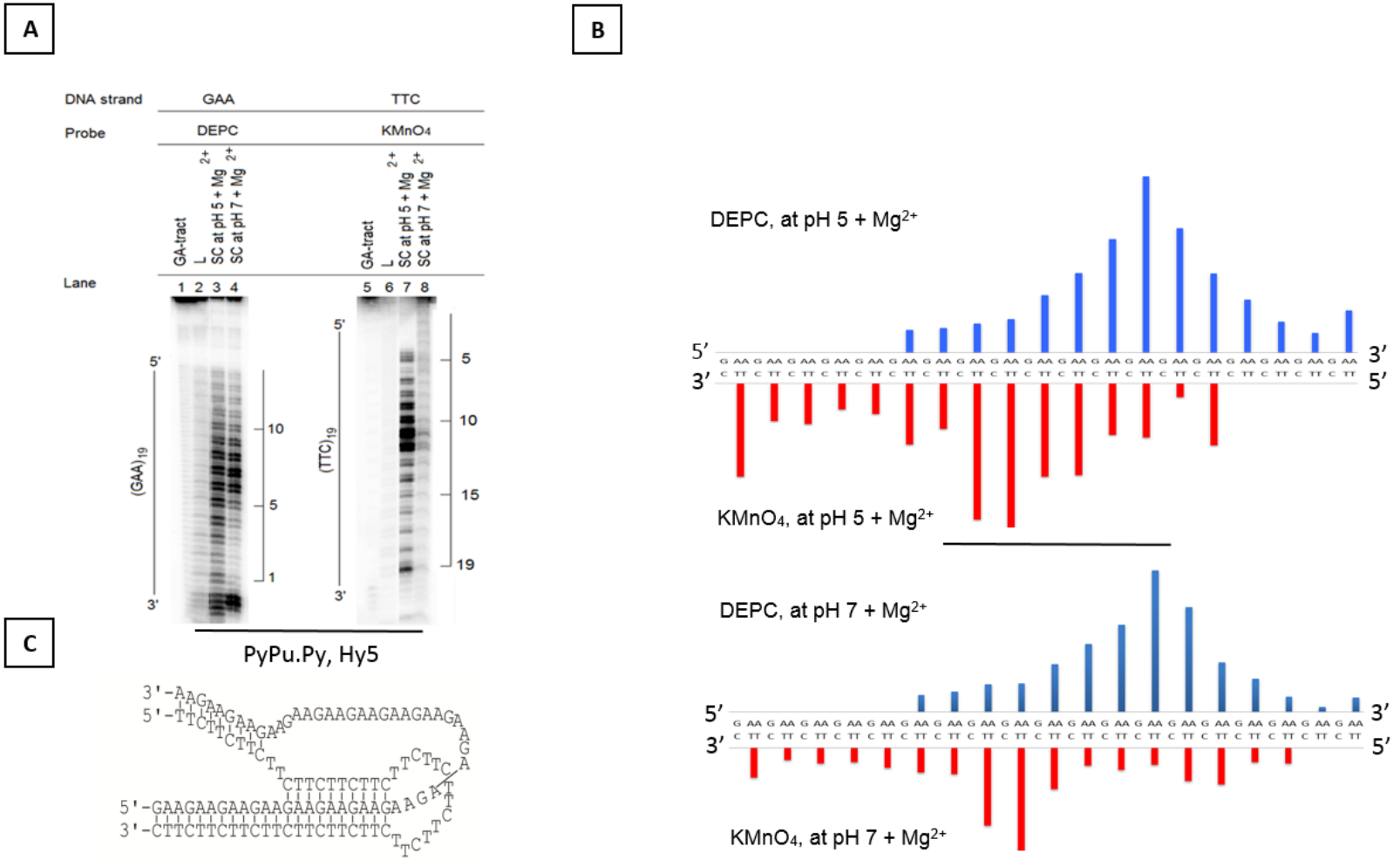


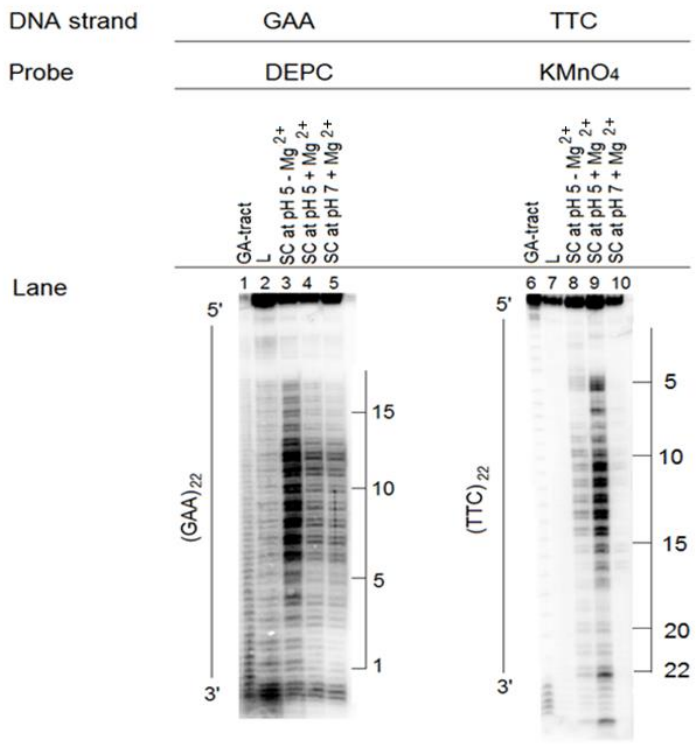

**Figure 4.5:** Chemical probing of pUC18(19,0) with DEPC and $KMnO_4$. **(A)** DEPC and $KMnO_4$ modification patterns of the GAA and TTC strands, respectively, at pH 5.0 (50 mM sodium acetate, 5 mM $MgCl_2$) and pH 7 (40 mM Tris-acetate, 5 mM $MgCl_2$). The track labelled 'GA' is a Maxam-Gilbert marker specific for guanines > adenines. SC and L represents supercoiled and linear plasmids, respectively. **(B)** Summary of the chemical cleavage data shown in (A). The height of each bar represents the average of the relative cleavage of each double (AAs) or (TTs) that was determined by densitometer scans of lanes 3, 4, 7 and 8 from panel (A). **(C)** Model of H-DNA that best fit the data shown.

Similar reactions were carried out with the plasmid containing $(GAA.TTC)_{22}$ repeats in the SmaI site (pUC18(22,0)) at pH 5 and pH 7. Figure 4.6A shows the cleavage patterns of the DEPC and $KMnO_4$ reactions on the supercoiled and linear samples after incubation at pH 5 (in the absence and presence of magnesium) or pH 7. Looking at DEPC modification of the GAA strand (lanes 3-5), it can be seen that the reactivity pattern is similar to that seen with the plasmid carrying 19 GAA repeats (Figure 4.5). DEPC probing of the GAA strand in all conditions reveals one clear region of modification in the centre of the GAA strand, especially around positions 6-12, with approximately equal intensity of each band. Very little reaction is observed in the neighbouring adenines and in linear DNA (lane 2). The cleavage pattern does not change much with pH. The $KMnO_4$ reactivity patterns with this construct also show one large modification region of thymine residues as seen in lanes 8 and 9 (though this is less clear in lane10), which is situated in the middle of the TTC strand. Modified sites are also observed near the top and bottom of the lanes. No reactivity to $KMnO_4$ is observed in the linear DNA (lane 7). This is similar to the pattern of reactivity seen with the plasmid carrying 19 GAA repeats (Figure 4.5, lane 7), with the exception that both thymine residues in each repeat are equally sensitive to the permanganate reaction. However, at pH 5 in the presence of magnesium, more thymine residues appear to react with $KMnO_4$, as seen by comparing lane 9 with lane 8. This is due to $Mg^{2+}$ playing a significant role in the complete formation of intramolecular triplex structures (21).

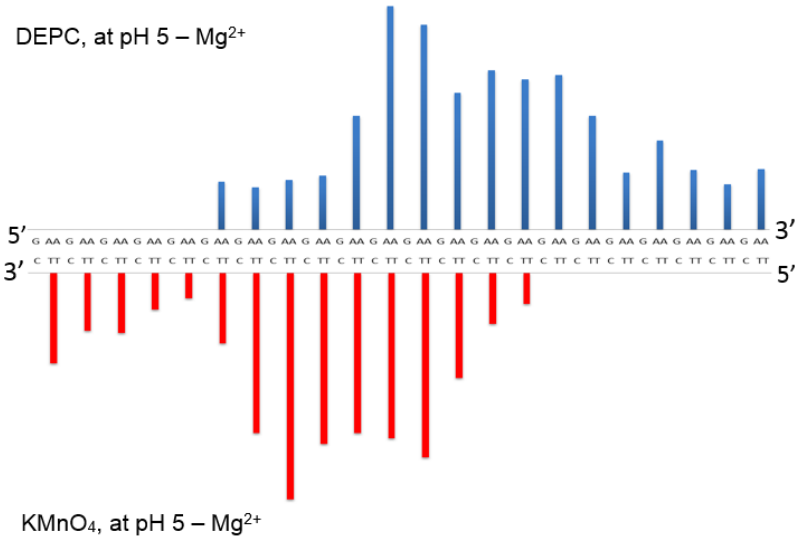
The results of densitometer traces of the data for DEPC and $KMnO_4$ modification of the $(GAA.TTC)_{29}$ insert are summarised in Figure 4.6B. These data are not consistent with the formation of a single unique H-DNA structure since both sequences exhibited large modified regions. This might suggest that it can adopt several different structures with similar energies; for instance a mixture of parallel and antiparallel forms as shown in Figure 4.6C. The patterns of DEPC modification obtained at acidic and neutral pH are consistent with a triplex model for the protonated H-form of the sequence (PyPu.Py). In this conformation (Figure 4.6C, left panel), the 5'-half of the pyrimidine stretch (TTC strand) forms a triplex with the 5'-half of the purine central strand. This model is similar to that designed for the plasmid carrying 19 GAA repeats (Figure 4.5C) which is stabilised by $CG.C^+$ and TA.T base triplets. However, the modification patterns obtained with $KMnO_4$ give a different pattern, and suggesting that *H-triplex (PyPu.Pu) might be formed, in which the 5' region of the purine strand (GAA) folds back upon itself to bind

antiparallel to the GAA strand using reversed Hoogsteen hydrogen bonds, leaving part of the 3' region of the pyrimidine strand (TTC) unpaired (Figure 4.6C, right panel). This class of triplex is stabilised by TA.A and CG.G.

**A**


DNA strand
GAA
TTC
Probe
DEPC
KMnO4
GA-tract
L
SC at pH 5 - Mg 2+
SC at pH 5 + Mg 2+
SC at pH 7 + Mg 2+
Lane
1 2 3 4 5
6 7 8 9 10
5'
3'
(GAA)22
(TTC)22
15
10
5
1
5
10
15
20
22

**B**


DEPC, at pH 5 – Mg2+
5′
3′
3′
5′
KMnO4, at pH 5 – Mg2+

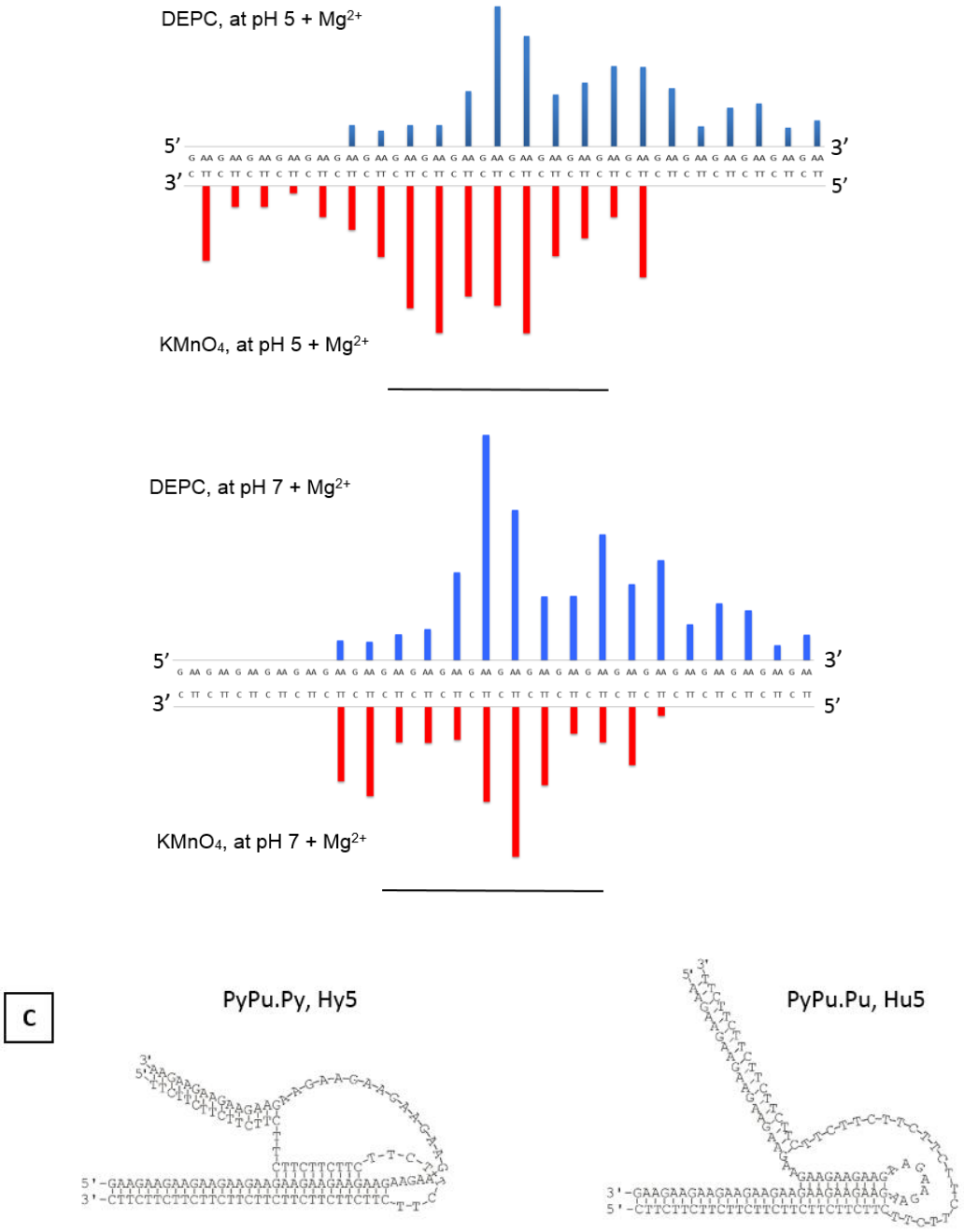
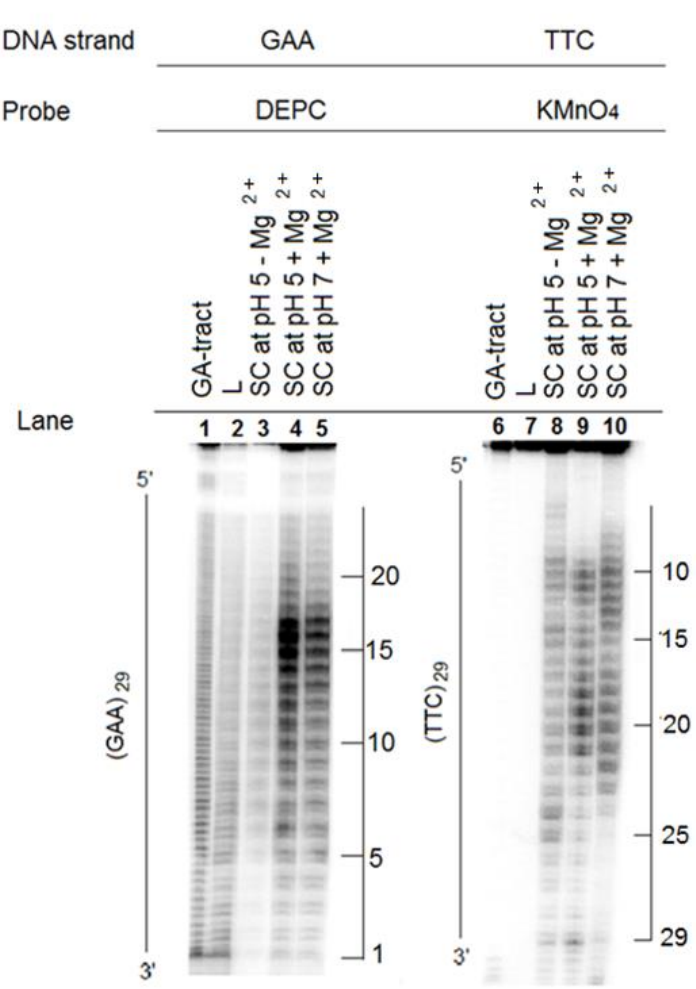


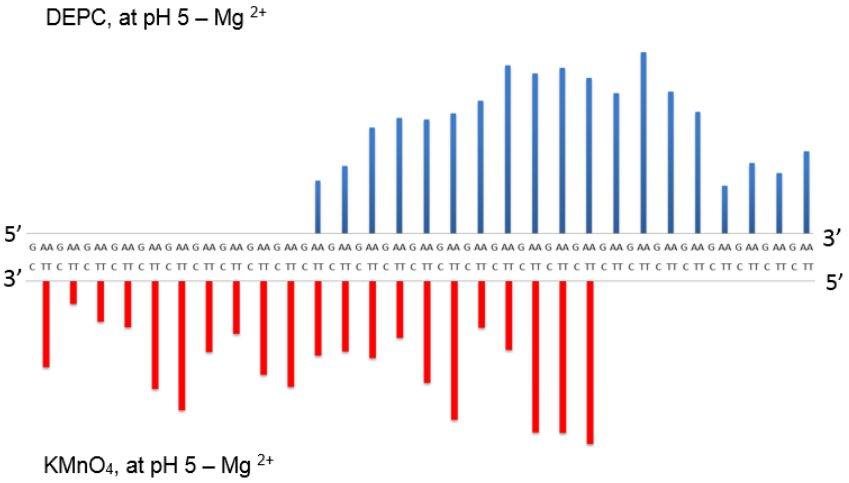

**Figure 4.6:** Chemical probing of pUC18(22,0) with DEPC and $KMnO_4$. **(A)** DEPC and $KMnO_4$ modification patterns of the GAA and TTC strands, respectively at pH 5.0 – $Mg^{+2}$ (50 mM sodium acetate), pH 5.0 + $Mg^{+2}$ (50 mM sodium acetate, 5 mM $MgCl_2$) and pH 7 (40 mM Tris-acetate, 5 mM $MgCl_2$. The track labelled 'GA' is GA marker. SC and L represents supercoiled and linear plasmids, respectively. **(B)** Summary of the chemical cleavage data shown in (A). The length of each bar corresponds to the average of the relative cleavage of each double (AAs) or (TTs) that determined by densitometer scans of the lanes 3, 4, 5, 8, 9 and 10 from panel (A). **(C)** Mixture of both parallel (PyPu.Py) and antiparallel (PyPu.Pu) triplex can adopted by $(GAA.TTC)_{22}$.

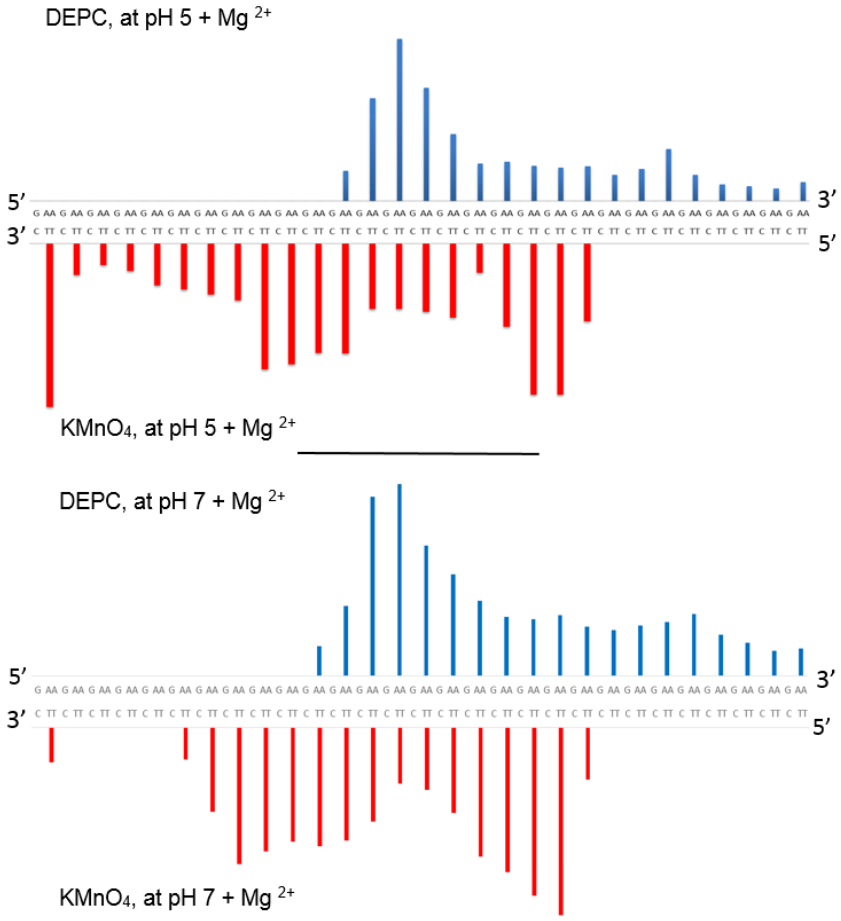
Figure 4.7 shows the results of similar experiments with plasmid pUC18(29,0) containing $(GAA.TTC)_{29}$ repeats in the SmaI site (other results with this plasmid were included in Chapter 3). As shown in Figure 4.7A, the pattern of DEPC reactivity shows considerable cleavage of adenines at the centre of the GAA strand in the supercoiled DNA (lanes 4 and 5). No differences were observed between DEPC modifications at pH 5 (lane 4) and pH 7 (lane 5). The most reactive adenines are located around the fourteenth to seventeenth repeats from the 3' end. No significant reactivity to DEPC is observed with the linear fragment (lane 2). Additional hyperreactivity can be seen around position 6 in lane 4 which may indicate the presence of a structural distortion at the junction between the triplex and duplex DNA. $KMnO_4$ modification of the TTC strand at pH 5 (no $Mg^{2+}$) (lane 8) resulted in the appearance of several modification regions. However, in the presence of magnesium (lane 9) at the same pH, two distinct modification regions (positions 9-12 and 17-21) from the 5' end are evident. Interestingly, some of the less reactive thymines in lane 9 (positions 7, 8, 13, 14, 22 and 23) become reactive to $KMnO_4$ at pH 7. Combining the patterns for both DEPC and $KMnO_4$ modification indicates that $(GAA.TTC)_{29}$ insert adopts an intramolecular triplex (class PyPu.Pu) with a different isomer (Hu3), since the GAA strand shows specific reaction with DEPC at the centre of the sequence and a large region of the TTC strand is exposed to $KMnO_4$ reaction. A schematic representation of a model for H-DNA (PyPu.Pu, isomer Hu3) is presented in Figure 7.4C. However, the modification patterns obtained with $KMnO_4$ give two distinct regions (especially in lanes 9 and 10, Figure 4.7A). This suggests that the proportion of the $(GAA.TTC)_{29}$ containing plasmid molecules forms two short intramolecular triplexes which is almost impossible to prove.

A


DNA strand
GAA
TTC
Probe
DEPC
KMnO4
GA-tract
L
SC at pH 5 - Mg 2+
SC at pH 5 + Mg 2+
SC at pH 7 + Mg 2+
Lane
1 2 3 4 5
6 7 8 9 10
5'
3'
(GAA) 29
(TTC) 29
20
15
10
5
1
10
15
20
25
29

B


DEPC, at pH 5 – Mg 2+
5′
3′
3′
5′
KMnO4, at pH 5 – Mg 2+




C

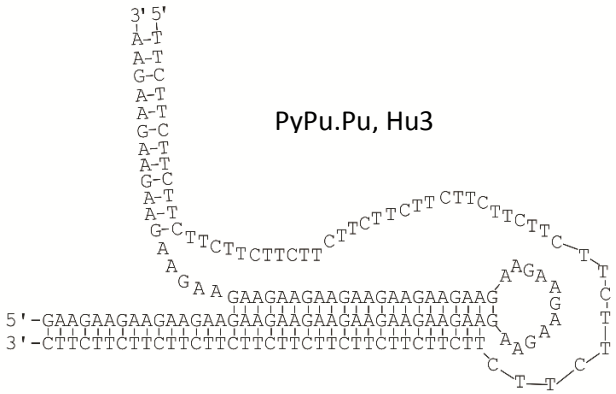
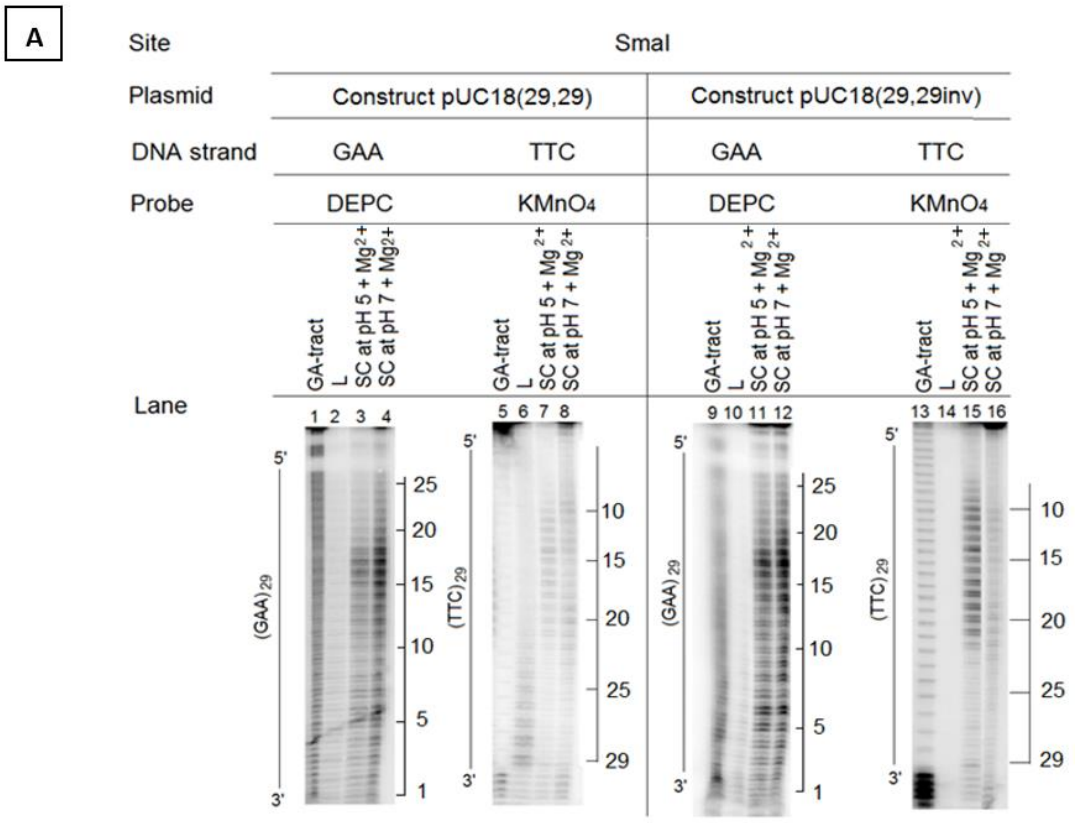
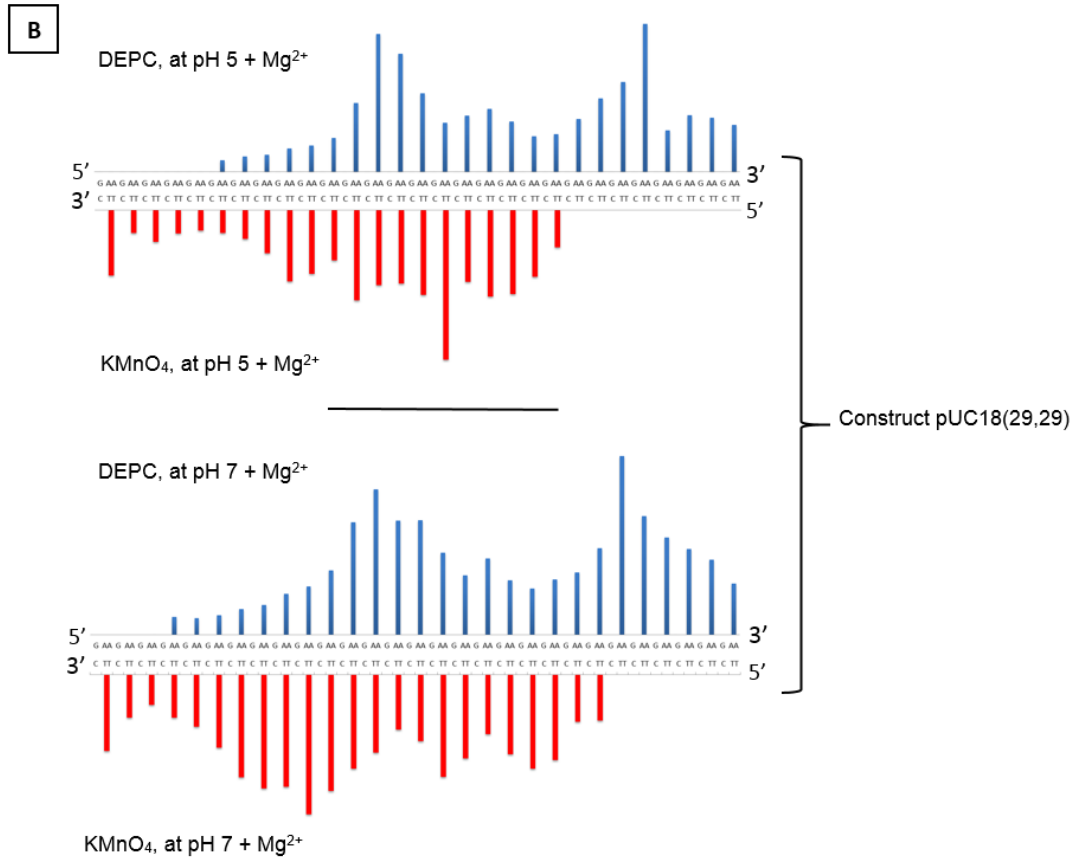


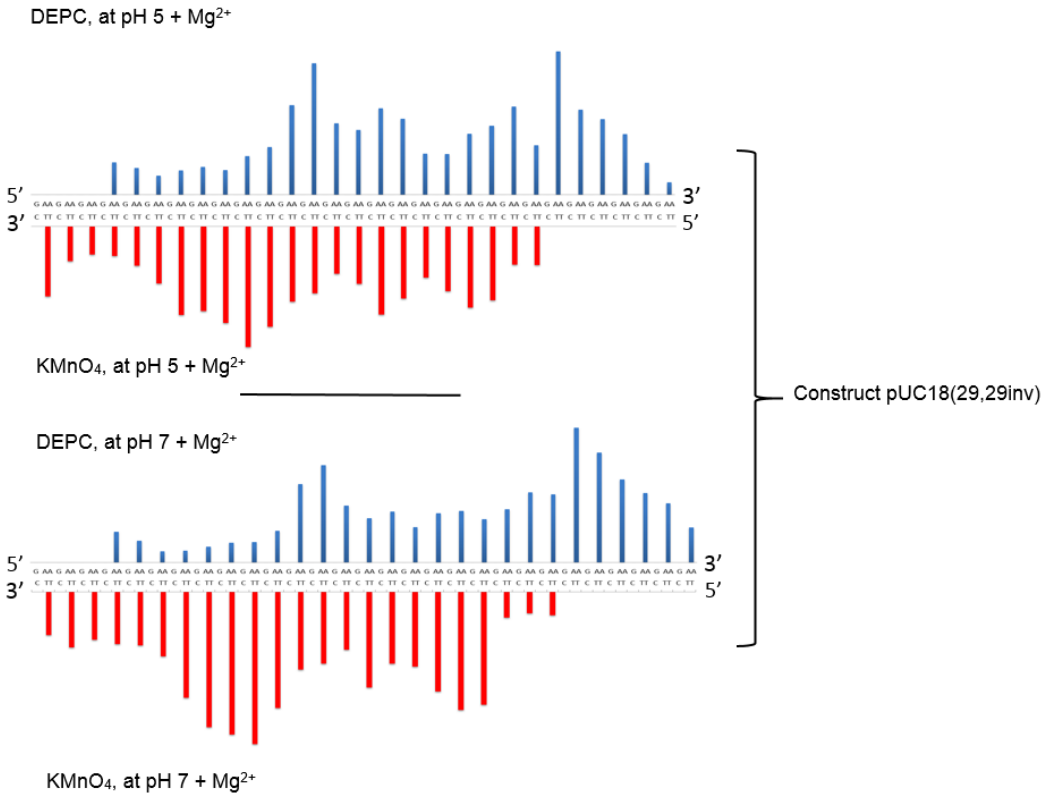

**Figure 4.7:** Chemical probing of pUC18(29,0) with DEPC and $KMnO_4$. **(A)** DEPC and $KMnO_4$ modification patterns of the GAA and TTC strands, at respectively at pH 5.0 – $Mg^{+2}$ (50 mM sodium acetate), pH 5.0 + $Mg^{+2}$ (50 mM sodium acetate, 5 mM $MgCl_2$) and pH 7 (40 mM Tris-acetate, 5 mM $MgCl_2$. The track labelled 'GA' is GA marker. SC and L represents supercoiled and linear plasmids, respectively. **(B)** Summary of the chemical cleavage data shown in (A). The height of each bar represents the average of the relative cleavage of each double (AAs) or (TTs) that determined by densitometer scans of the lanes 3, 4, 5, 8, 9 and 10 from panel (A). **(C)** Model of H-DNA that best fit the data shown.

Reactions identical to those outlined above were conducted at pH 5 and pH 7 with plasmids having two $(GAA.TTC)_{29}$ repeats, either in direct orientation pUC18(29,29) or indirect pUC18(29,29inv). The DEPC and $KMnO_4$ reactivity of the repeats contained in the SmaI site are presented in Figure 4.8A. Looking first at DEPC modification of both constructs, the cleavage patterns (lanes 3, 4, 11 and 12) are similar to those seen with construct pUC18(29,0) (Figure 4.7A, lanes 4 and 5). As we previously observed, the centre of the $(GAA)_{29}$ strand is predominantly modified with DEPC and there is no significant difference between the patterns at pH 5 and pH 7. In addition, adenines at the 3'-end (position 6) also exhibited reactive towards DEPC. The patterns of permanganate reactivity at pH 5 and pH 7 are shown in lanes 7, 8, 15 and 16. In both plasmids, one distinct region of the TTC strand is susceptible to oxidation by permanganate at pH 5 compared with two regions at pH 7. This also suggests that a proportion of the 29 GAA repeats in these plasmids at pH 7 prefer to form two intramolecular triplexes instead of one long structure.

The results of chemical cleavage patterns for DEPC and $KMnO_4$ modification of the $(GAA.TTC)_{29}$ insert in the SmaI site of both constructs (pUC18(29,29) and pUC18(29,29inv)) are consistent with the formation of an antiparallel H-DNA (class PyPu.Pu, isomer Hu3), similar to that obtained with the plasmid carrying one tract of 29 GAA repeats (Figure 4.7C). In this structure (Figure 4.8C), the 3'-half of the GAA strand folds back on itself to form triplex with the 5'-half of the same strand using reverse Hoogsteen base pairs. This suggests that the centre of the purine strand (GAA) was exposed to DEPC reaction as a result of short single stranded loop formation. In addition, adenines at the 3'-end became accessible to DEPC modification indicating the presence of a structural distortion at the junction between the triplex and duplex DNA.


A
Site
SmaI
Plasmid
Construct pUC18(29,29)
Construct pUC18(29,29inv)
DNA strand
GAA
TTC
GAA
TTC
Probe
DEPC
KMnO4
DEPC
KMnO4
GA-tract
L
SC at pH 5 + Mg2+
SC at pH 7 + Mg2+
Lane
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
5'
3'
(GAA)29
(TTC)29
25
20
15
10
5
1
10
15
20
25
29


B
DEPC, at pH 5 + Mg2+
KMnO4, at pH 5 + Mg2+
DEPC, at pH 7 + Mg2+
KMnO4, at pH 7 + Mg2+
5′
3′
Construct pUC18(29,29)




C

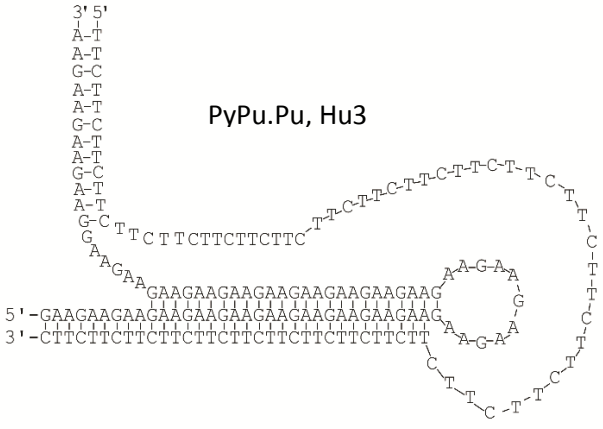
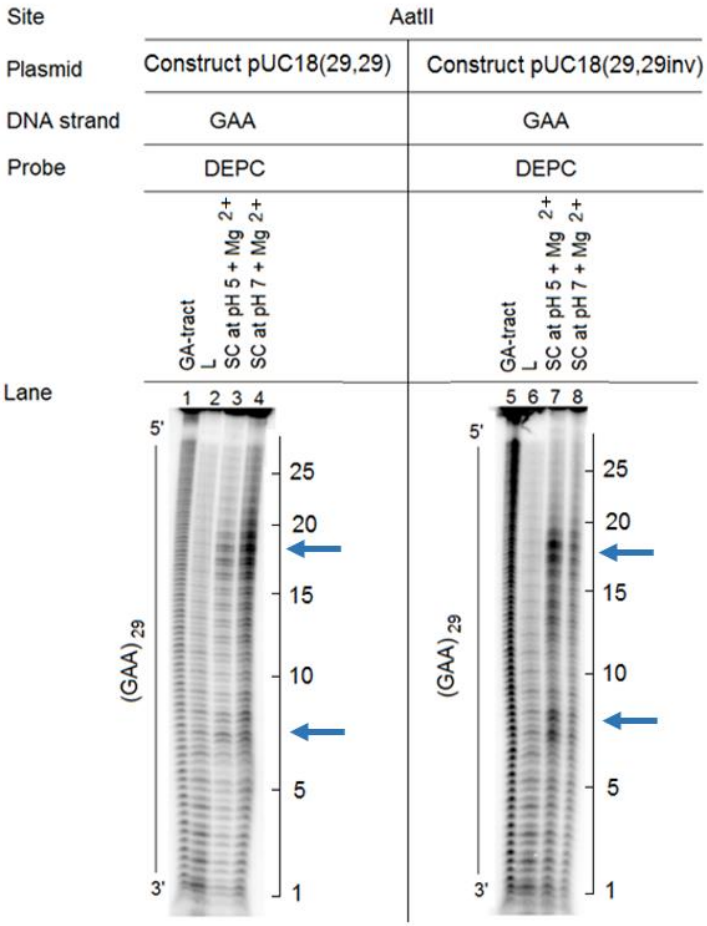



**Figure 4.8:** Chemical probing of the insert in the SmaI site of both pUC18(29.29) and pUC18(29,29inv) plasmids with DEPC and $KMnO_4$. **(A)** DEPC and $KMnO_4$ modification patterns of the GAA and TTC strands, respectively, at pH 5.0 + $Mg^{+2}$ (50 mM sodium acetate, 5 mM $MgCl_2$) and pH 7 (40 mM Tris-acetate, 5 mM $MgCl_2$. The track labelled 'GA' is a GA marker. SC and L represents supercoiled and linear plasmids, respectively. **(B)** Summary of the chemical cleavage data shown in (A). The length of each bar represents the average of the relative cleavage of each double (AAs) or (TTs) that determined by densitometer scans of the lanes 3, 4, 7, 8, 11, 12, 15 and 16 from panel (A). **(C)** Model of H-DNA that best fit the data shown.

Figure 4.9 shows the results of the reaction of DEPC with the $(GAA)_{29}$ sequence in the AatII site of constructs pUC18(29,29) and pUC18(29,29inv). The distributions of reactivity in these constructs are similar to those observed in the SmaI site of the same constructs with DEPC reactivity in Figures 4.7A and 4.8A. The cleavage pattern does not change much with pH; both contain two distinct regions of modification (blue arrows); the first region is mostly at the central of the GAA strand and the second region in the 3'-half of the strand. Again, this result can be consistent with the formation of antiparallel H-DNA (PyPu.Pu), since the strongest modification occurred in the loop of the hairpin (positions 17-19), while the less strongly modified region (positions 7 and 8) might correspond to the junction between triplex and duplex DNA. This configuration is similar to those presented in Figurers 4.7C and 4.8C.

A



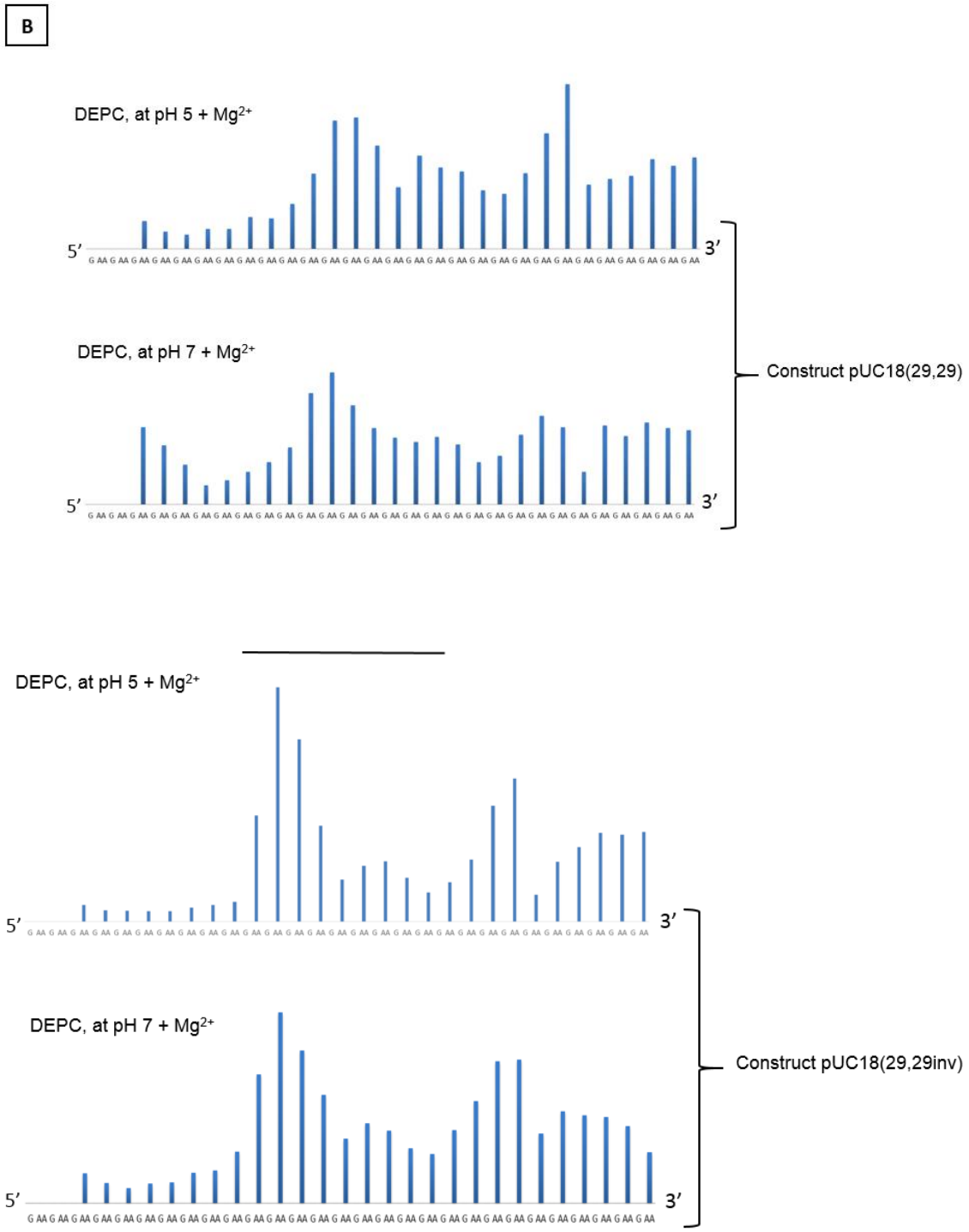



**Figure 4.9:** Chemical probing of the insert in the AatII site of both pUC18(29,29) and pUC18(29,29inv) plasmids with DEPC. **(A)** DEPC modification patterns of the GAA strands at pH 5.0 + $Mg^{+2}$ (50 mM sodium acetate, 5 mM $MgCl_2$) and pH 7 (40 mM Tris-acetate, 5 mM $MgCl_2$). The track labelled 'GA' is a GA marker. SC and L represents supercoiled and linear plasmids, respectively. Blue arrows show the positions of the maximum amount of modification in lanes 3, 4, 7 and 8. **(B)** Summary of the chemical cleavage data shown in (A). The length of each bar represents the average of the relative cleavage of each double (AAs) or (TTs) that determined by densitometer scans of the lanes 3, 4, 7 and 8 from panel (A).

Similar experiments were performed with plasmid pUC18(85,0) containing $(GAA.TTC)_{85}$ in the SmaI site. The results for the reactions with DEPC and $KMnO_4$ are presented in Figure 4.10. Looking first at DEPC modification at pH 5 (lane 3), clear hyperreactivity can be seen in the adenines in the 5' (upper) half of the GAA strand, suggesting that this region of GAA sequence contains locally unwound DNA. This is consistent with a suggestion that the 3'-half of the $(TTC)_{85}$ strand folds back in the middle to form a parallel triplex (PyPu.Py; Hy3), leaving most of the 5'-half of the $(GAA)_{85}$ strand in a single stranded configuration. Similar reactivity was observed at pH 7 (lane 4). It is interesting to observe that the $KMnO_4$ pattern of this construct (lanes 7 and 8) is similar to that seen with DEPC (lanes 3 and 4). In lane 7, the bands in the 5' (upper) half of the TTC strand are much darker than those in the lower (3') part of the strand, suggesting that half of the $(GAA)_{85}$ strand folds back on itself to form an antiparallel triplex (PyPu.Pu; Hu3), leaving the 5'-half of the $(TTC)_{85}$ strand in a single strand configuration (though this would contradict the results with DEPC). One modified thymine is also observed at position 1 (lane 7) as seen previously with permanganate reactions.

Similar experiments were carried out with construct pUC18(85,62) that contains two tracts of 85 and 62 GAA repeats and the version of this plasmid that displays unusually slow gel mobility (pUC18*(85,62)) and construct pUC18(85,39). As shown in Figure 4.11, DEPC reactivity of the strands containing $(GAA)_{85}$ at pH 5 (lane 3) and pH 7 (lane 4), $(GAA)_{62}$ at pH 5 (lane 9) and pH 7 (lane 10) produced uniform modification of adenines in the sequence, whilst clear reactions with $KMnO_4$ were observed on the 5'-side (upper) of the TTC strand at pH 7 (lane 6). This observation demonstrates again that thymine residues in this region are more accessible than those in the lower part of the TTC strand. Looking at the DEPC cleavage pattern of the $(GAA)_{85}$ strand within the construct pUC18*(85,62) (lanes 13 and 14) and $(GAA)_{39}$ strand within construct pUC18(85,39) (lanes 21 and 22), it can be seen that all the adenines in the $(GAA)_{85}$ and $(GAA)_{39}$ strands are equally reactive to DEPC. Similar results are obtained with permanganate reaction with the $(TTC)_{85}$ strand in the construct pUC18*(85,62) (lanes 17 and 18). However, there is no clear-cut boundary for the modified region as typically observed for a defined triplex structure as seen in Figures 4.5-4.9 and all the thymines appear to be equally reactive. Similar results are obtained at both pH 5 and pH 7. As DEPC and $KMnO_4$ have difference in ionic strengths, this could affect the conformational

equilibrium of triplex formation and subsequently would affect the reactivity toward the exposed single strand regions.

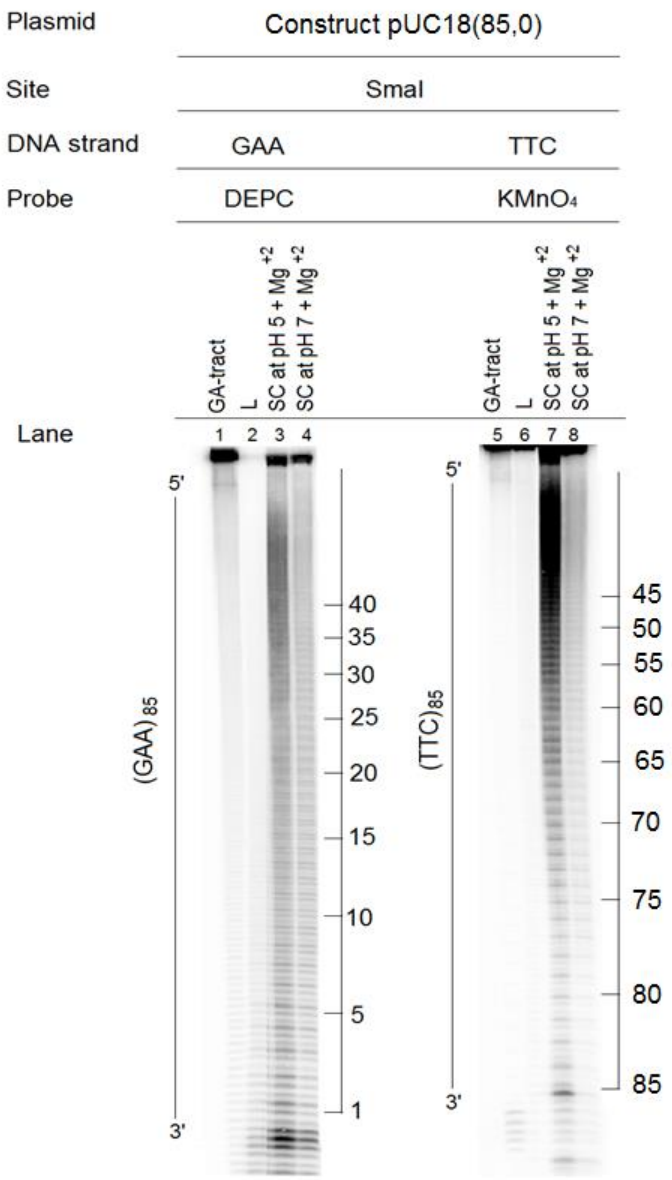



**Figure 4.10:** DEPC and $KMnO_4$ modification patterns of the $(GAA)_{85}$ and $(TTC)_{85}$ strands in the SmaI site of construct pUC18(85,0) at pH 5 and pH 7. The track labelled 'GA' is a GA marker. SC and L represents supercoiled and linear plasmids, respectively.

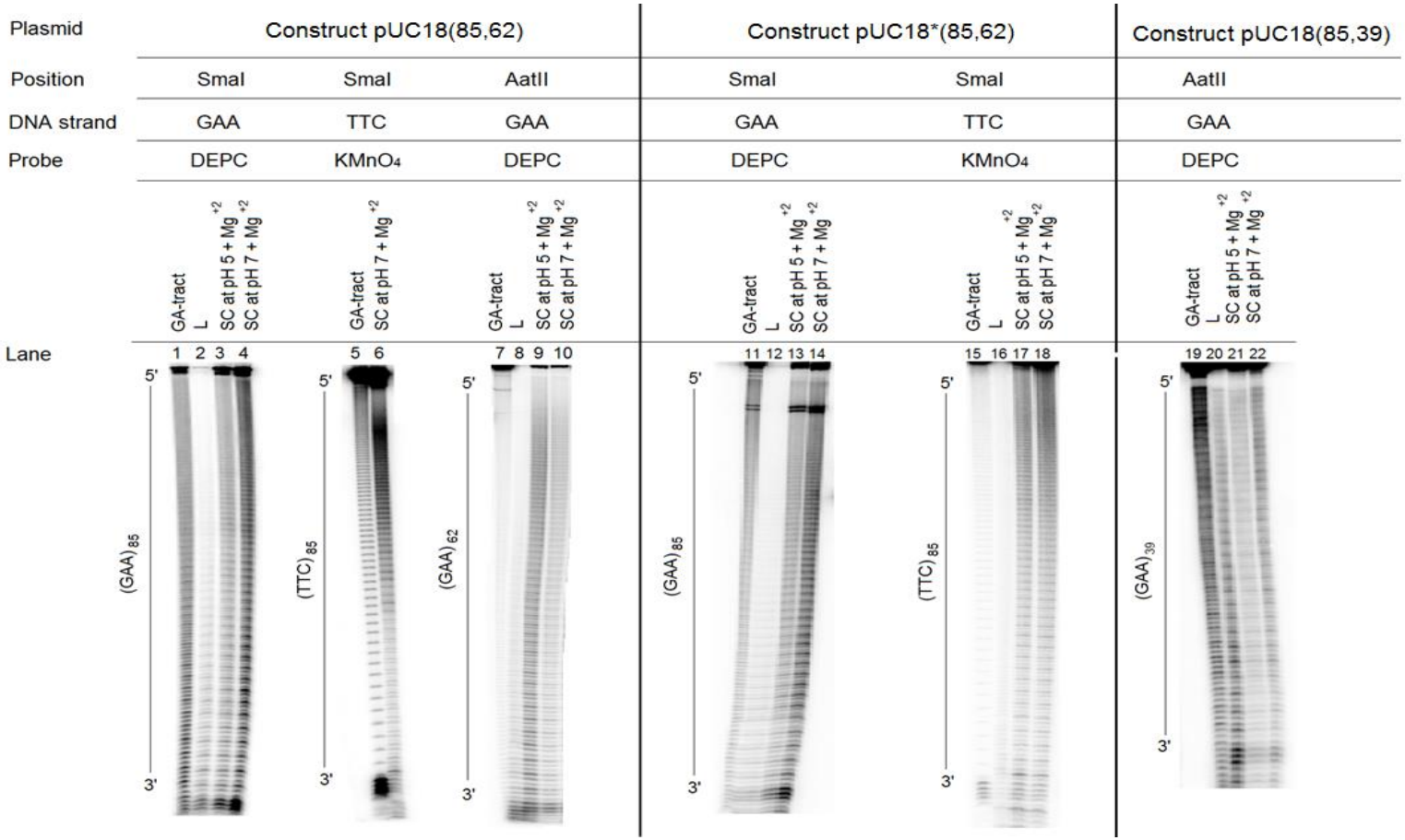



**Figure 4.11:** Chemical probing of the GAA and TTC strands within constructs pUC18(85,62), pUC18*(85,62) and pUC18(85,39). GA is a GA marker. SC and L represents supercoiled and linear plasmids, respectively.

### 4.4.3 Probing with S1 nuclease

S1 nuclease was employed as an enzymatic probe in order to confirm the DEPC and $KMnO_4$ data for detecting intramolecular triplexes. Figure 4.12A presents S1 nuclease digestion of plasmid pUC18(29,0) containing $(GAA.TTC)_{29}$ repeats, for which the $KMnO_4$ and DEPC reactions suggested the formation of an intramolecular triplex under superhelical stress. As can be seen in lanes 3 and 4, the S1 nuclease cleavage occurred between the $6^{th}$ and $16^{th}$ repeat of the $(GAA)_{29}$ strand and the most reactive site is located at the centre of the sequence (position 15). This corresponds to the short single stranded loop formed at the centre of an intramolecular triplex (antiparallel H-DNA). The reason why some bases in the 3' half of GAA sequence were exposed to S1 nuclease cleavage may be because the third strand does not wrap perfectly around the acceptor duplex DNA; these bases are therefore displaced from the helix and become sensitive to attack by S1 nuclease. For the $(TTC)_{29}$ strand (lanes 7 and 8), S1 cleavage is clearly observed in the 5'-(upper) half of the TTC strand suggesting that this region is displaced during H-DNA formation. In both strands, the S1 reactivity occurred only in the supercoiled form. Figure 4.12B shows the DNA histograms of the data shown in panel A. It is worth noting that S1 nuclease cleaves phosphodiester bonds adjacent to adenines and thymines, but little or no cleavage with guanines and cytosines. The reason may due to stacking differences which in turn produce different conformations of phosphodiester backbone, resulting in different susceptible to S1 cleavage. Therefore, bars correspond to Gs for (GAA) strand or Cs for (TTC) strand were ignored. The data for S1 nuclease cleavage presented in Figure 4.12 confirmed that the $(GAA.TTC)_{29}$ insert; either in pUC18(29,0), pUC18(29.29) or pUC18(29,29inv), adopt antiparallel H-DNA (class PyPu.Pu, isomer Hu3) under negative supercoiling.

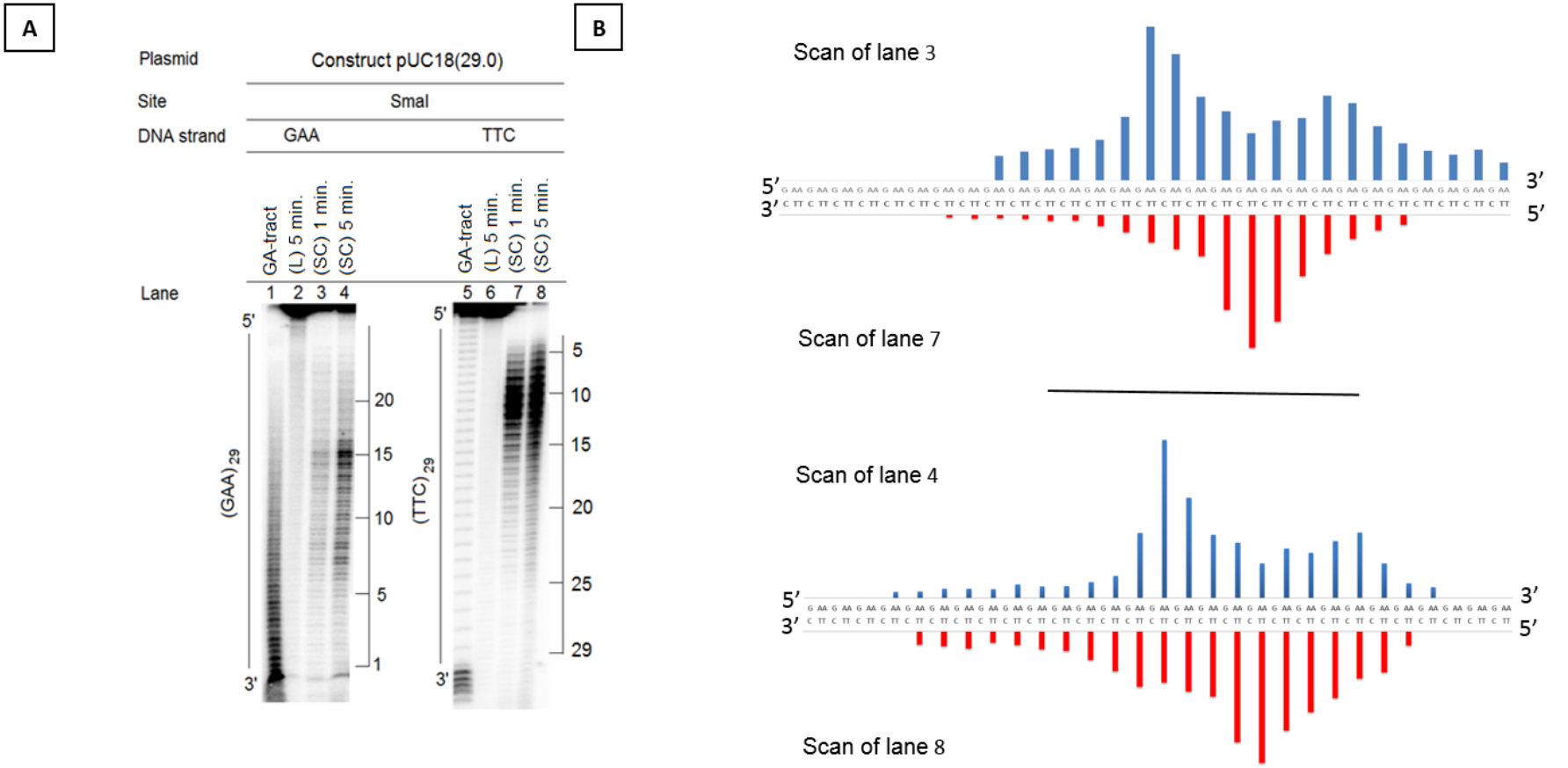



**Figure 4.12: (A)** Patterns of S1 nuclease probing of the $(GAA)_{29}$ and $(TTC)_{29}$ strands of plasmid pUC18(29,0) at 37 °C. GA is a GA marker. SC and L represents supercoiled and linear plasmids, respectively. The supercoiled DNA was digested for 1 and 5 minutes as indicated. Linear forms of the construct were incubated with S1 nuclease for 5 minutes. **(B)** Summary of the enzymatic cleavage data shown in (A).

Similar experiments were carried out with the constructs pUC18(29,29) and pUC18(85,39). As shown in Figure 4.13A, the reaction was examined with the $(GAA)_{29}$ sequence in construct pUC18(29,29), within both the SmaI and AatII sites, (lanes 2 and 5). Regions of enhanced susceptibility are evident, especially in the middle region of each strand. Several bands of strong S1 nuclease reactivity can be seen in the $(GAA)_{39}$ sequence in the construct pUC18(85,39) (lane 8). Weak reactivity is observed between these dark bands especially in the 3'- (lower) half of this strand. In contrast, for pUC18(29,29) several bands of S1 nuclease reactivity can be seen with adenines as well as guanines (lanes 2 and 5), thus in the DNA histograms (Figure 4.13B) bars of guanine residues have been taken into account.

The S1 reactivity patterns obtained with the $(GAA)_{29}$ insert (lanes 2 and 5) are in a good agreement with the formation of antiparallel H-DNA; this was suggested from the chemical probing of the same sequence (Figures 4.7C and 4.8C), since the strongest S1 cleavage occurred in the central region of GAA tract. This again suggests that $(GAA)_{29}$ strand folds back in the middle to form a triplex with the other half of the GAA sequence in an antiparallel orientation. However, part of the 3'- (lower) half of the third strand was also sensitive to S1 nuclease cleavage, as observed with pUC18(29,0) (Figure 4.12A, lanes 3 and 4). This may be because the third strand does not wrap properly around the duplex DNA and it seems that only a small proportion of the third strand GAA forms a triplex using reverse Hoogsteen base pairs. The extensive sensitivity of the 3'-end of the $(GAA)_{39}$ sequence to S1 nuclease cleavage (lane 8), suggests that the 5'-half of $(TTC)_{39}$ may fold back to form a triplex (PyPu.Py; Hy5) with the purine strand, leaving the 3'-half of the GAA strand as single-stranded DNA, although we failed to detect any specific modified regions toward DEPC or $KMnO_4$ reactions of the same construct.

This chemical and enzymatic probing with long inserts (≥39 repeats) was repeated several times and did not always give the same results, thus we presented representative gels obtained in these experiments.

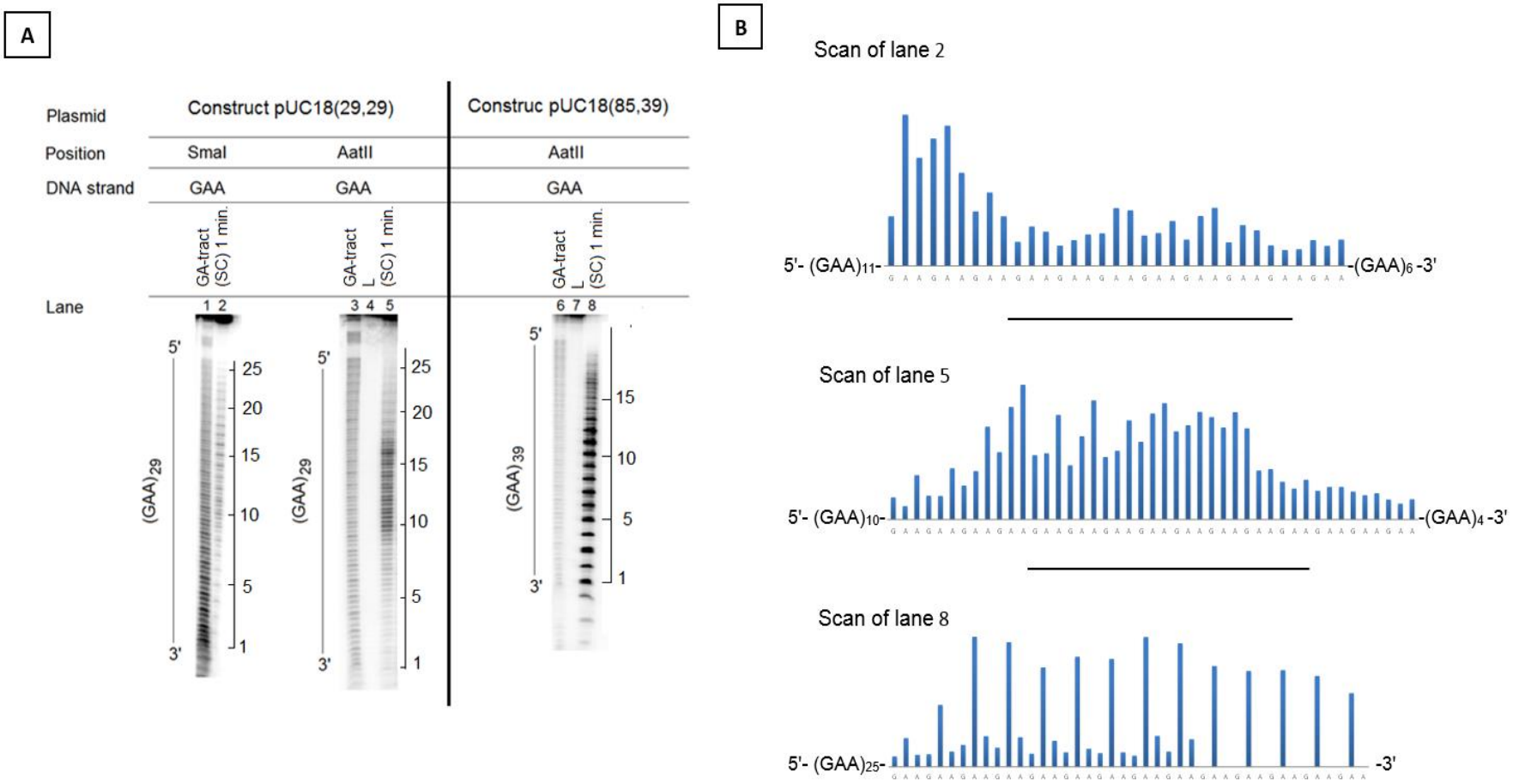



**Figure 4.13: (A)** Patterns of S1 nuclease probing of the $(GAA)_{29}$ and $(GAA)_{39}$ strands within plasmids pUC18(29,29) and pUC18(85,39), respectively. The samples were digested with the enzyme at 37 °C for 1 minute. **(B)** Represents summary of the results shown in (A). Round brackets ($(GAA)_n$) indicate the lengths of the GAA repeats that have very weak reaction with S1 nuclease.

## 4.5 Discussion

DNA is a dynamic molecule and in solution can adopt several different structural conformations depending on the precise nucleotide sequence and environmental conditions. The GAA.TTC tract is a polymer of purine.pyrimidine sequence known to form the following distinctly different DNA structures in plasmids: (i) canonical B-DNA duplex; (ii) intramolecular triplex which can adopt the form PyPu.Pu or PyPu.Py, depending on whether the third strand is polypurine (GAA strand) or polypyrimidine (TTC strand). In addition, two possible triplex isomers can be formed depending whether the 3'- or 5'-end of the mirror repeat is single stranded; (iii) the sticky DNA conformation, a complex triplex structure that has been poorly characterised that is formed by two directly repeating GAA.TTC tracts in one supercoiled plasmid. Since sticky DNA formation has been reported to require at least 59 repeats of $(GAA.TTC)_n$, in this work we analysed the local structures formed by this triplet repeat that fall in two categories of short ($n < 59$) and long ($n \geq 59$) repeats. The experiments performed in this chapter were complex and time consuming, therefore a set of the construct vectors containing short and long were selected for these experiments.

Different regions of single stranded DNA are present within the structure of an intramolecular triplex; (i) the displaced single strand that serves as a hinge between flanking duplexes; (ii) the loop at the turn of the third strand; (iii) a region between the triplex and adjacent duplex at the duplex-triplex junction. To confirm intramolecular triplex formation in our GAA.TTC repeat-containing plasmids and to study the structural properties, chemical and enzymatic probes were employed to detect the presence of exposed or single strand nucleotides. Previous studies have not found any chemical or enzymatic reactivity at the base of the acceptor duplex in an intramolecular triplex (250, 251) and it has also been difficult to detect any discontinuity between the acceptor duplex and the triplex from steric hindrance of the third strand leaving the duplex (37). In this chapter, we present evidence that $(GAA.TTC)_n$ inserts in plasmids can adopt intramolecular triplexes at native superhelical density using DEPC, $KMnO_4$ and S1 nuclease and compare the structures that are formed in plasmids that contain one or two $(GAA)_n$ tracts.

The results reported in this chapter did not show any evidence for ‘sticky DNA’ formation, only there was clear evidence for the formation of H-DNA within the shorter

repeat inserts. Under superhelical stress, two types of H-DNA, (PyPu.Py) and (PyPu.Pu) with different isomers, were adopted by short GAA repeats. The chemical probing data presented in Figures 4.5 show that short $(GAA.TTC)_{19}$ repeats formed one type of H-DNA (PyPu.Py, Hy5) (5'-half of the pyrimidine sequence (TTC) as the third strand), while $(GAA.TTC)_{22}$ repeats formed a mixture of both (PyPu.Py, Hy5) and (PyPu.Pu, Hu5), since large regions of both GAA and TTC strands showed a strong sensitivity to DEPC and $KMnO_4$, respectively, as presented in Figure 4.6. This structural diversity may play a role in their biological process, since perturbation by protein binding or other factors will be relatively easily for a system that is already in equilibrium. Chemical and enzymatic patterns obtained with the $(GAA.TTC)_{29}$ insert, either singly (Figures 4.7 and 4.12) or in combination in the same vector (Figures 4.8 and 4.9) show that this repeat adopts a unique (PyPu.Pu; Hu5) triplex DNA structure with the 5'-half of the purine sequence (GAA) as the third strand; no formation of the other isomer was observed with 29 GAA repeats. The reasons for selection of the different isomers are not exactly known. Htun and Dahlberg (1988; 1989) (88, 251) studied the differential formation of Hy5 and Hy3 triplex isomers. According to their model, the Hy3 isomer forms at higher negative superhelical densities whereas the Hy-5 isomer forms more easily at less negative superhelical densities. Hy3 is thought to be more stable than Hy5 as it relaxes one more negative supercoil than formation of Hy5. However, several Py.Pu sequences have been investigated by analysing chemical modification patterns (180, 183, 252) and it has been suggested that only one isomer of H-DNA is formed by short Py.Pu repeats (<40 bp). In most instance the 3'-half of the pyrimidine strand formed the third strand while the 5'-half of the purine strand was single-stranded DNA. Although previous studies have shown that synthetic Py.Pu sequences exhibit unusual properties in supercoiled plasmids (179, 250, 253, 254), this is the first demonstration that (PyPu.Py; Hy5) and (PyPu.Pu; Hu3) are adopted by GAA.TTC repeats. The DEPC and $KMnO_4$ reactivity results for all the examined constructs above confirmed that bases in these regions are much more reactive to the probes under superhelical stress than in linear DNA. This indicates the significant role of negative supercoiling in the structural transition to an altered DNA conformation in these repeat tracts.

Compared with other mirror-repeated Py.Pu sequences that form protonated H-DNA (PyPu.Py) at pH 5.0, our short $(GAA.TTC)_n$ repeats (n=19 and 22) also appear to form this type of triplex at pH 7.0 in the presence of magnesium. Previous study have also

shown that $(GAA.TTC)_n$ repeats (n = 9, 23) can exist as protonated H-DNA at neutral pH, which is unexpected as it is well above the pK of cytosine. Even the addition of $Mg^{+2}$ did not convert the structure into a (PyPu.Pu) triplex (4). However, these GAA repeats form triplexes that contain 67% T.AT triplets and it is known that this triplet is stabilised by $Mg^{2+}$ (unlike $C^+$.GC). Conversely, 29 GAA repeats formed another type of triplex structure (PyPu.Pu) at both pH 5.0 and pH 7. The Friedreich's ataxia GAA.TTC repeats (flanked by the natural intron 1 sequence) formed a (PyPu.Pu) triplex which is more stable at pH 7.0 and this is probably due to the high content of pH-independent TA.T triads especially at the ends (4). However, we have no clear explanation why the chemical reactions with $(GAA.TTC)_n$ (n=19, 22 and 29) gave different structures for H-DNA and the cleavage pattern of each construct does not change much with pH. For 19 and 22 GAA repeats this is probably because the CG.$C^+$ triplets are more stable than the other triplets, as mentioned in Chapter 1, Section 1.3.1.3, though this greater stability would have to be retained at pH 7. These lengths of triplet repeats may require higher supercoiling to form an antiparallel triplex at neutral pH (255). Therefore, the model shown in Figure 1.11 (Chapter 1) cannot be used as a definitive rule for which type of triplex is formed.

The H-DNA structures formed by short GAA repeats may not involve in the Friedreich ataxia. It was suggested that H-DNA could upregulate transcription in healthy people in two different ways; one possibility is that H-DNA itself has a positive *cis*-acting effect. It has been shown that presence of H-DNA forming tracts in promoter regions can increase transcription efficiency compared to promoters which are absent in these tracts (256). Another way in which the H-DNA structure may upregulate transcription is that the single stranded DNA extruded on triplex formation may provide a target binding site for RNA polymerase (33). RNA polymerase has been shown to bind at single strand regions of supercoiled plasmids (257, 258). This suggests that the single strand extruded from an H-DNA structure could facilitate transcription by promoting the binding of RNA polymerase.

When the $(GAA.TTC)_n$ repeats are sufficiently long ($n \geq 39$), a moderately uniform pattern of modifications was observed with DEPC and $KMnO_4$ in both strands, with the exception of both the GAA and TTC strands in construct pUC18(85,0) (Figure 4.10, lanes 3 and 7), and the TTC strand in construct pUC18(85,62) (Figure 4.11, lane 6). The chemical modification patterns of construct pUC18(85,0), show that the 5' (upper) half of

both GAA and TTC strands were sensitive to DEPC and $KMnO_4$ modifications. This also suggests that $(GAA.TTC)_{85}$ forms a mixture of parallel and antiparallel forms of H-DNA with similar energies and that both forms are present. Constructs pUC18(85,62), pUC18*(85,62) and pUC18(85,39) (Figure 4.11), show no evidence of single or mixture of triplexes although the 5' (upper) half of the TTC strand in the SmaI site of pUC18(85,62) became more reactive to $KMnO_4$ in the supercoiled DNA (lane 6). In addition, long tracts of GAA repeats ($n \geq 62$) failed to produce any clear reaction with S1 nuclease. The failure of DEPC, $KMnO_4$ and S1 nuclease to identify any specific modified regions within long GAA and TTC sequences could be attributed to the dynamic instability of such triplexes which may reflect the inability of the third strand to fit properly into the major groove of an acceptor duplex DNA. In the meantime, the long loop of the displaced strand may not be purely single-stranded but may interact with the third strand. The stability of DNA triplexes is known to be sensitive to the ionic conditions, as triplex formation require the assembly of three polyanionic strands. Thus, increasing the ionic strength from 5 mM to 10 mM of $MgCl_2$ may provide a stabilizing effect of long H-DNA. In addition, lower temperatures may also help long GAA repeats to form stable triplex as some reports showed that the rate of triplex association decreases with temperature (259).

Other studies have shown that long Py.Pu tracts, including GAA.TTC repeats form two short triplexes instead of one long structure (4, 195, 254, 260) (Figure 4.14). This preference may be related to destabilisation of a long triplex and its rearrangements into two shorter structures (37).

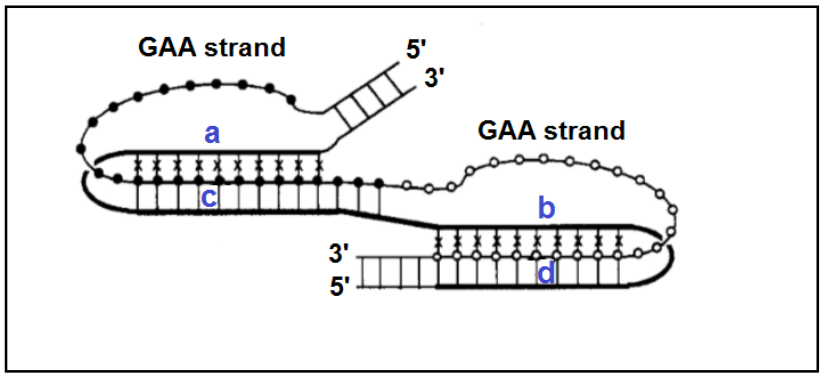
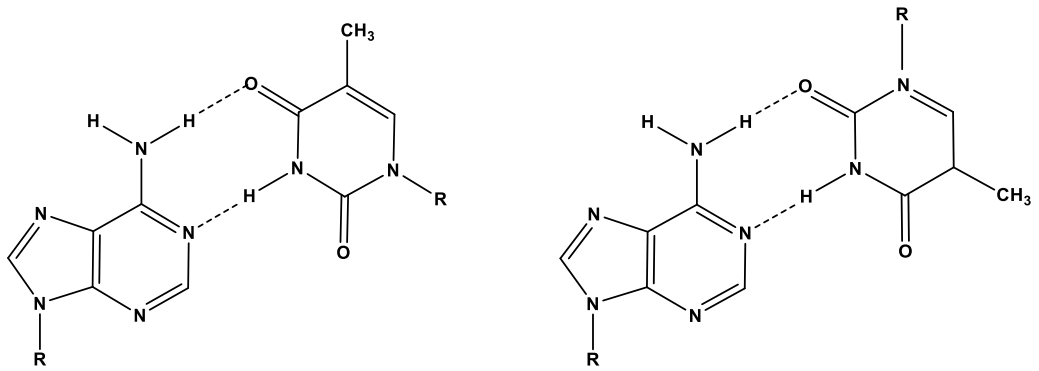



**Figure 4.14:** Possible bi-triplex DNA structure for long $(GAA.TTC)_n$ sequence ($42 \leq n \leq 75$). Two halves of the GAA sequence are unpaired. Regions a and b of the TTC strand form triplex with regions c and d of the acceptor duplex DNA. Adapted from reference (4).

DNA oligonucleotides with appropriate sequences can adopt a parallel stranded duplex conformation and stabilized by reverse Watson-Crick G.C and A.T hydrogen bonded base pairs. Parallel stranded DNA can be also formed by Hoogsteen A:T base pairing (261). In the reverse Watson-Crick orientation, as shown in Figure 4.15, the base and ribose sugar are extended in a *trans* position compared with the *cis* configuration found in B-DNA. In a reverse Watson-Crick base pair, the thymines are hydrogen bonded through the O2 and N3 positions, whereas the N3 and O4 positions are hydrogen bonded in the normal Watson-Crick base-pairing scheme (89). Therefore, additional secondary structure may be formed by long GAA repeat as shown in Figure 4.16. LeProust *et al*., in 2000 (262) showed that a mixture of the isolated GAA and TTC oligonucleotides has the ability to adopt various structures including Watson-Crick antiparallel duplex, triplex DNA and parallel stranded DNA; the parallel (GAA.TTC) duplexes were in equilibrium with the antiparallel Watson-Crick (GAA.TTC) duplexes. Since there is more than one structure could be formed within our long duplex (GAA.TTC) repeats, the chemical probes may not be able to give a unique structural signature.




Watson-Crick A.T base pair

Reverse Watson-Crick A.T base pair

**Figure 4.15:** B-form DNA contains A.T base pairs (left) with strands oriented in an antiparallel fashion. Parallel DNA requires the formation of reverse Watson-Crick base pairs in which one base is oriented 180 °C with respect to the Watson-Crick base pair orientation (right).

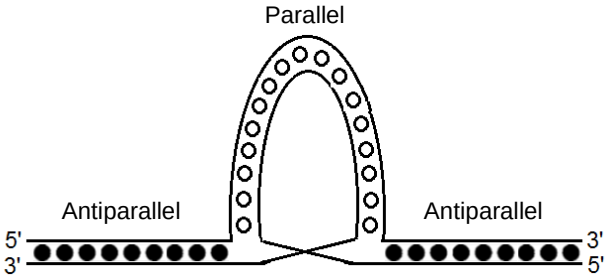



**Figure 4.16**: Possible secondary structure formed by long GAA repeats; a parallel loop embedded in an antiparallel DNA. The region with reverse Watson-Crick base-pairing is designated by open circles bridging the two strands (light and dark) of the helix, whereas the symbols in the antiparallel strands regions with conventional Watson-Crick base-pairs are solid.

The S1 nuclease was also employed to map the single strand regions that accompany H-DNA in supercoiled DNA and analysed on agarose gel. The data presented for S1 nuclease mapping in this chapter (Section 4.4.1.2) of plasmids pUC18(29,0) and pUC18(85,0) revealed that S1-cleavage sites were located in the MCS of these plasmids indicating the presence of local unwinding DNA in the short and long tracts of GAA repeats. These regions were cleaved by S1-nuclease only under negative supercoiling which is consistent with the DEPC and KMnO4 chemical probing experiments. These findings confirm that superhelical stress is required to induce the formation of the intramolecular triplex structure.

# 5 The Effects of GAA.TTC Repeats on Gene Expression

## 5.1 Introduction

Long $(GAA.TTC)_n$ repeats ($n \geq 59$) in the first intron of the FXN gene are known to severely reduce the levels of mature frataxin mRNA (140, 263) and the frataxin protein (145), indicating supressed gene expression. This reduction may be due to either reduced transcription or abnormal post-transcriptional processing (140, 145, 264). Previous studies have shown that the size of the GAA.TTC repeats is inversely correlated with the amount of the frataxin protein, an effect that has been observed in cultured cells from FRDA patients (145, 163). It is known that long stretches of (GAA.TTC) have the ability to form triplexes and sticky DNA in supercoiled plasmids (3, 4, 136). Since there is a correlation between the length of the GAA.TTC repeat and the formation of sticky DNA, it has been hypothesised that this type of non B-DNA structure may be involved in the repression of the FXN gene in FRDA patient cells. Ohshima *et al*. (1998) (221) first demonstrated the effect of GAA repeats on transcription, using a reporter vector whose expression was inhibited by an intronic GAA repeat insertion. The experiments described in this chapter examine the effect of the hyperexpanded GAA repeats on gene expression *in vivo*. To achieve this, we used short and long inserts of $(GAA.TTC)_n$ repeats ($n = 29$ or 85) in the pGL3 luciferase reporter vector, either singly or in combination. These were prepared as described in Chapter 2.

## 5.2 Basic experimental design

The previous chapters discussed how DNA molecules that contain GAA repeats can undergo structural perturbations under negative supercoiling. The existence of non-B-DNA structures formed by GAA repeats may interfere with transcription in a length-dependent manner. To examine this effect we cloned various lengths of GAA repeats into the SacI site of pGL3 control vector upstream of the luciferase start codon, and into the PciI site (Figure 5.1). As well as examining these effects within supercoiled plasmids, these constructs were also linearised to remove the torsional stress that is thought to be required for secondary structure formation.

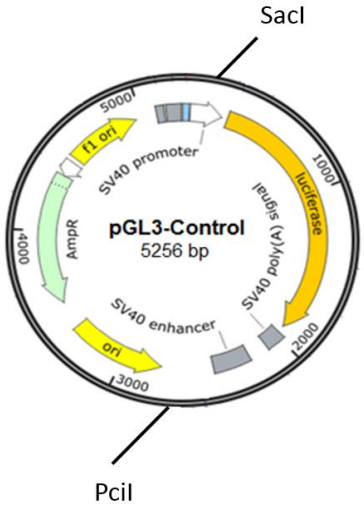



**Figure 5.1**: pGL3-Control vector showing positions of SacI and PciI cloning sites. Adapted from SnapGene.

The level of luciferase expression of these constructed vectors was determined using a dual luciferase reporter assay, comparing expression of these constructs containing the firefly luciferase gene with that of *Renilla* luciferase on a separate co-transfected plasmid. Briefly, each construct containing the experimental *Firefly* luciferase gene was co-transfected into HeLa cells along with a second vector phRL-CMV (expressing the *Renilla* luciferase, serving as a transfection efficiency control in the experiment) at ratio of 5:1. The cells were harvested 2 days after transfection and the *Renilla* and *Firefly* luciferase activities were measured using the dual luciferase reporter assay system (Promega). In this system, *firefly* and *Renilla* luciferases are simultaneously expressed but individually measured. The *firefly* luciferase reporter was measured first by adding the Luciferase Assay™ Reagent II to generate a luminescent signal, after which the reaction was quenched and the *Renilla* luciferase reaction was initiated by adding Stop and Glo™ reagent to the same sample. The *firefly* luciferase requires beetle luciferin, ATP, magnesium and molecular oxygen to emit light while *Renilla* luciferase utilises only coelenterate luciferin (coelenterazine) and molecular oxygen to catalyse the luminescent reaction (Figure 5.2). For each sample, the measurements were performed in triplicate and each experiment repeated three times. Empty pGL3-control and pGL3-basic reporter vectors were included in the luciferase assay and served as positive and negative controls, respectively.

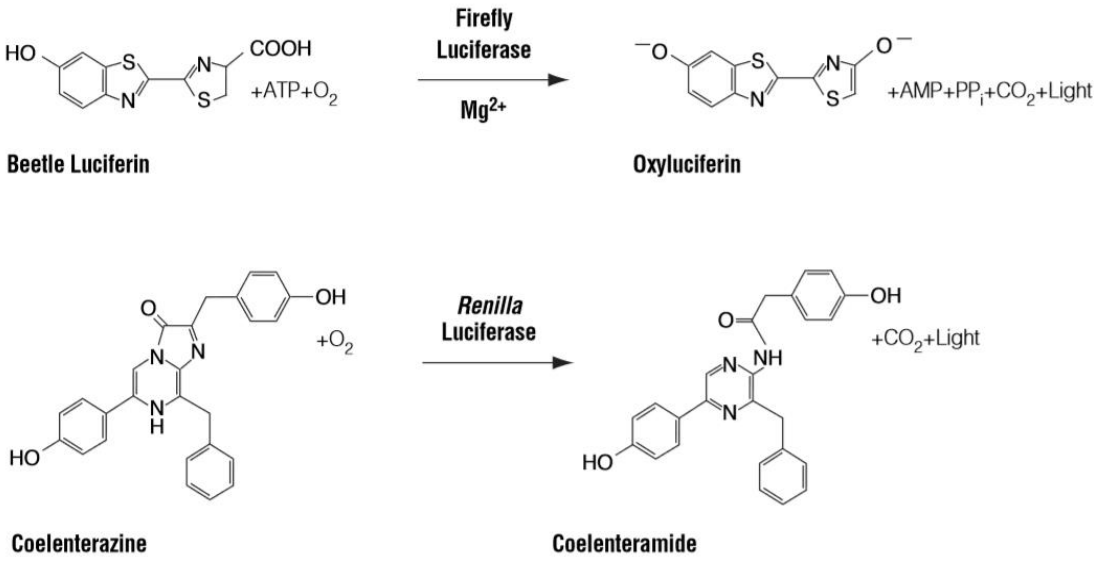



**Figure 5.2:** Bioluminescent reactions catalysed by *firefly* and *Renilla* luciferases. Taken from Promega catalogue, part no.TM058.

## 5.3 Data analysis

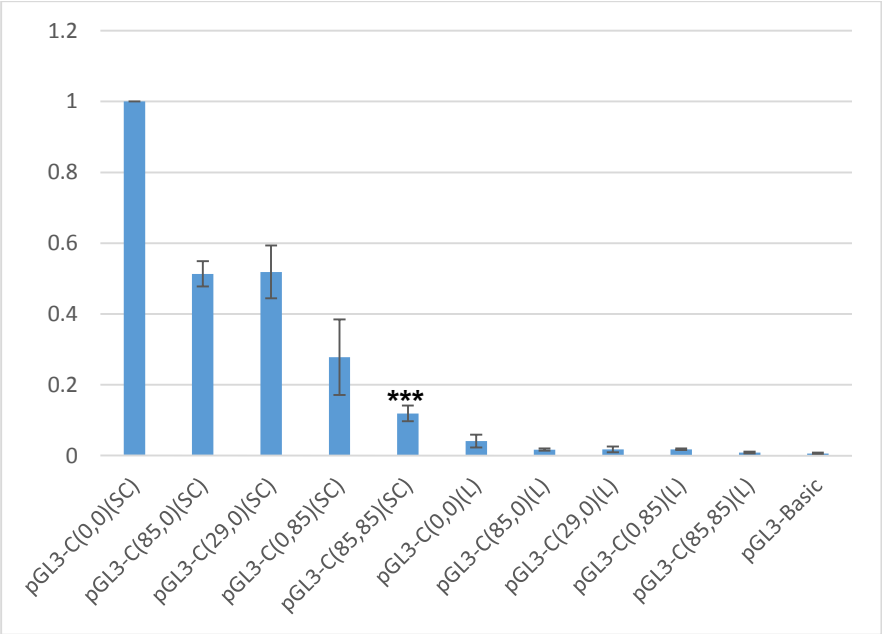
The ratio of the average *firefly* luciferase activity to *Renilla* was calculated and normalised relative to the transfected empty pGL3 control. Data were reported as means ± standard deviations of 3 independent experiments, in triplicate in each experiment. The means between the control and construct plasmids were compared by applying the *t* test. (*) denotes $P < 0.05$. Analysis was performed using SPSS 11.5.

## 5.4 Results

In order to assess the effect of GAA repeat expansions on gene expression, we constructed five reporter plasmids containing different lengths of GAA repeats (see Chapter 2, Section 2.2.2.4). These five reporter constructs are called pGL3-C(X,Y) where X is the repeat number in the SacI site and Y is the number in the PciI site. These plasmids are empty pGL3-C(0,0), pGL3-C(85,0), pGL3-C(29,0), pGL3-C(0,85), pGL3-C(85,85). We then investigated the effects of these construct reporters on the luciferase gene expression when transfected into HeLa cell lines. In addition, these constructs were digested with AlwNI enzyme to produce linear plasmids. These linear forms of the construct vectors were examined alongside the supercoiled plasmids. The results are shown in Figure 5.3.

It can be seen (Figure 5.3A) that linear forms of the construct vectors produced much lower luciferase expression relative to the supercoiled forms. This suggests that expression of transfected plasmid depends on DNA topology. Supercoiling DNA gives higher levels of transcription than linears because DNA linears preferentially degrade. The supercoiled plasmids, pGL3-C(85,85) (SC) showed the greatest decrease in luciferase activity relative to the other supercoiled constructs and showed ( ~ 89%; $P \leq 0.001$) decrease compared to the positive control. The pGL3-C(85,0)(SC) and pGL3-C(29,0)(SC) constructs showed slightly lower levels of luciferase expression as compared to the positive control (~49% decrease; $P \leq 0.001$). Interestingly, pGL3-C(0,85)(SC) revealed lower expression relative to the pGL3-C(85,0)(SC). Figure 5.3B shows the relative luciferase activities of the linear constructs compared with linear control pGL3-C(0,0)(L). These constructs showed very little reduction in luciferase activity compared to the linear control. This again indicates that supercoils express better than linears.

**(A)**



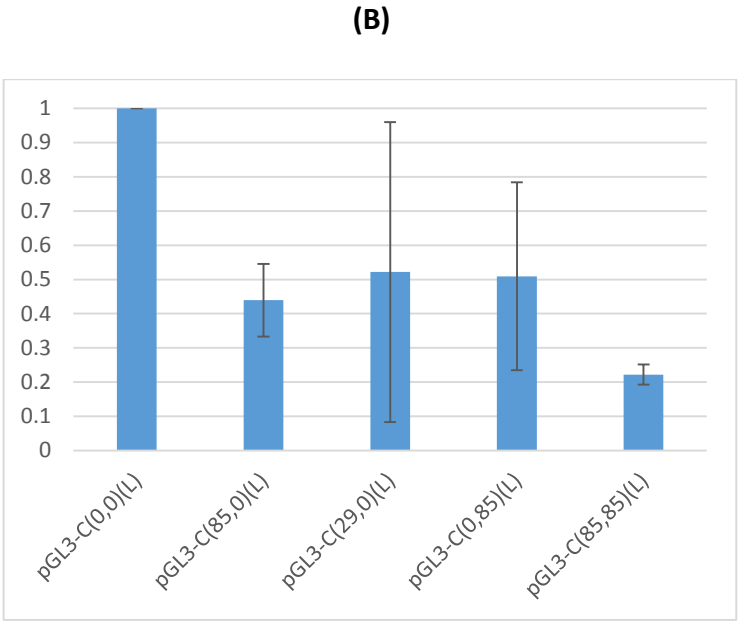



**Figure 5.3:** Relative *firefly* luciferase activities of the five vector constructs in Hela cells. **(A)** Shows the relative luciferase activities of the supercoiled and linear constructs compared with supercoiled control pGL3-C(0,0)(SC). **(B)** Shows the relative luciferase activities of the linear constructs compared with linear control pGL3-C(0,0)(L). The results are average values obtained from three independent experiments, each performed in triplicate. The error bars represent the standard deviation from triplicates of three separate experiments. SC and L represent supercoiled and linear forms of DNA, respectively. The asterisks indicate results that are significantly different from the positive control; reporter pGL3-C(0,0)(SC).

## 5.5 Discussion

It is well known that oligopurine.oligopyrimidine sequences that are capable of forming H-DNA can play a significant role in regulating the transcription of many genes (265). Several studies have shown that H-DNA can turn on or turn off gene expression depending on a number of factors, such as the position of H-DNA in a gene and the adjacent sequences (33, 34). On the basis of a cell-based reporter assay, when a potential H-DNA sequence was inserted in the promoter region of the β-*lactamase* gene of pUC19 vector, expression of the *lacZ* gene was increased significantly in *E.coli* relative to native pUC19 plasmid (256). In contrast, H-DNA formed within the coding region or between the promoter and coding sequence enhanced gene transcription (266, 267).

Although there is considerable evidence that H-DNA forming sequences within or near genes may affect transcription, the mechanisms that are involved are undefined and remain a significant challenge. In *E.coli*, when H-DNA forming sequences were inserted into the *lacZ* reporter gene, no inhibition of gene expression was observed in mammalian cells (267). However, the presence of similar sequences upstream of a *lacZ* reporter gene caused strong inhibition (dysfunction) of transcription in mammalian cells, which was in contrast to the results from similar studies in *E. coli* (268). Several studies have shown that H-DNA in eukaryotic systems can either up-regulate (269, 270), down-regulate (267, 271-273), or have no effect on gene transcription (274, 275). These different results may be due to different mechanisms involved under each specific condition (32). For example, different lengths of H-DNA forming sequences were inserted into mouse albumin promoters and examined for transcriptional activity and for H-DNA formation. The transcriptional activity predicted the ability of the specific sequences to form H-DNA but not by repeat number, position, or the number of mutant base pairs (251).

In this chapter, we analysed the effect of GAA.TTC repeats on gene expression by transfecting HeLa cells with five luciferase reporter constructs harbouring different lengths of GAA.TTC repeat. The data presented in Figure 5.3 show that expanded GAA.TTC repeats, either in the SacI site or the PciI site, inhibited gene expression in a length-dependent manner. The construct vector containing two long inserts of GAA repeats (pGL3-C(85,85)(SC)) exhibited a clear and pronounced reduction in luciferase activity compared to other constructs. The simplest explanation is that this triplex-forming sequence can adopt a stable secondary DNA structure, such as H-DNA for

constructs containing one insert, or sticky DNA for constructs containing two tracts of 85 GAA repeats, thereby inhibiting transcription. It can also be noted that insertion of 85 GAA repeats close to the enhancer region of the *firefly* luciferase gene (pGL3-C(0,85)(SC)) (*i.e.* remote from the actual gene and its promoter) led to reduced transcription efficiency of about half that in the construct containing the same length insert in the upstream region (pGL3-C(85,0)(SC)). There is no clear explanation for this difference but it is possible that the presence of H-DNA forming sequence in the PciI site decreases the binding of transcription factors to the enhancer DNA, while insertion of the same insert into the SacI site may impede the process of RNA polymerase. Another possible reason could arise from the position of PciI site in the pGL3-control vector, which is very close to the origin of replication region that affect proliferation of the plasmids into the HeLa cells. This suggests that insertion of GAA.TTC repeats within the PciI site has more influence than those in SacI site on the luciferase activity of pGL3-control vector. A study conducted by Kohwi *et al.* (1991) (35), showed the influence of $(G.C)_n$ stretches of varying length on gene expression. Their findings also revealed a remarkable length dependence between the ability of triplex-forming sequence *in vitro* and its ability to enhance gene expression *in vivo.* They concluded that short repeats of $(G.C)_n$ sequence ($n < 30$) serve as a target for transcription activators, while longer repeats ($n \geq 32$) can prevent activators binding by adopting an intramolecular triplex structure.

In prokaryotes and eukaryotes, the topological state of the DNA substrate often plays a significant role in processes including replication, transcription and recombination (276-278). In order to examine the importance of DNA supercoiling in gene expression, our reporter constructs were transfected into Hela cells as linear forms, alongside their supercoiled forms. By comparing the expression of *firefly* luciferase gene of transfected supercoiled and linear DNAs (Figure 5.3A), we are able to demonstrate that supercoiled DNA has a profound effect on transcription efficiency. Previous studies showed that the influence of supercoiling is most profound in plasmids that contain enhancer sequences (279). This suggests that supercoiling DNA affects the ability of transcription factors to bind with enhancer sequences. Other (Py.Pu) repeats may have similar mechanisms such as GAAGGA.TCCTTC repeats in the human tumour necrosis factor receptor gene, and GAA.TTC and GAG.CTC triplets in the cardiac α-myosin heavy chain gene, whose biological roles are still unknown.

Our analysis of the negative effect of GAA.TTC expansions on gene expression is consistent with the observations in FRDA. In FRDA, GAA.TTC repeats are usually much longer and substantial additional investigation is still required to evaluate the effect of the intronic hyperexpanded GAA.TTC repeats on frataxin expression.

# 6 General conclusions

## 6.1 Conclusions

DNA in nature mostly adopts the double helix B-form configuration, which consists of two antiparallel DNA strands which are bonded by the complementarity of Watson-Crick base pairs. However, DNA has the capability of adopting several other conformations as dictated by its sequence and the conditions. These alternative structures include left-handed Z-DNA which occurs at alternating purine-pyrimidine sequences (13), cruciforms at inverted repeat sequences (94), and triplexes at mirror repeats of polypurine.polypyrimidine sequences (9). Several of these unusual DNA structures are implicated in a number of biological processes (10, 11). One such alternative DNA structure is associated with long repeats of the trinucleotide sequence GAA, which causes the genetic disease Friedreich's ataxia, which belongs to a large group of trinucleotide repeat expansion diseases (280, 281). Long GAA.TTC repeats form several types of unusual DNA structures that have potential to affect gene expression (137, 155, 220, 235, 282, 283). The study of the structural behaviours and biophysical properties of long stretches of GAA.TTC repeats is important with respect to the etiology of FRDA. Therefore, the principle aim of this thesis was to examine and characterise the formation of non-B-DNA structures in the FRDA-associated GAA.TTC repeats that were cloned in suitable vectors.

The studies presented in this thesis have been categorised into three main chapters. The first chapter (Chapter 3), describes construction and characterisation of pUC18 plasmids containing single or double tracts of GAA repeats. In addition, the inserts were subcloned into a pGL3-control vector which was used in gene expression assays in Chapter 5. According to the literature, sticky DNA formation requires at least 59 GAA repeats in two positions of a single plasmid (125), thus a family of pUC18 plasmids harbouring short (< 59) and long (≥ 59) lengths of the GAA.TTC repeats was prepared. We failed to produce more than 85 repeats of the cloned GAA.TTC sequence in the pUC18 plasmid and short cloned inserts (< 50 repeats) were the majority in the cloned plasmids. The reason may be due to instability (deletions) of the long GAA repeats compared to the shorter sequence during replication in *E.coli* as mentioned in the discussion section of Chapter 3.

Sticky DNA is the only non-B DNA conformation that requires two tracts of triplet repeats such as GAA.TTC sequence in one plasmid. The association of these two tracts within supercoiled plasmid is very stable and requires high temperature (80°C) as well as the removal of divalent cations with EDTA to convert the adhered region back to a duplex conformation (6). Therefore, we used different gel-based techniques to detect the formation of sticky DNA (or other non-B structures), including a gel mobility shift assay, restriction enzyme accessibility and relaxation with topoisomerase I. Based on the gel results in Chapter 3, there was no evidence for sticky DNA formation in all the pUC18 clones, even those containing two tracts of long repeats pUC18(85,62), which would have been expected to form this novel conformation. We do not have a clear explanation for these negative results, but offer some possibilities and suggestions. All previous investigations of sticky DNA formation demonstrated that this novel DNA was formed only when the two tracts of long GAA.TTC repeats were equal in size; no previous study has shown that sticky DNA can form with different lengths of the triplet repeats in one plasmid (6, 125, 137). This may be interpreted as suggesting that each of the GAA repeats within our pUC18(85,62) can form different unusual DNA structures as demonstrated by Potaman *et al*., 2004 (4). It is known that formation of a mixture of two triplexes is possible within 42 repeats of a GAA.TTC sequence. Thus, if a mixture of small triplexes forms within one or both inserts of 85 and 62 repeats, the size and region of these non-B DNA structures would be different and might not permit interaction with each other to form stable sticky DNA. However, to counteract this possibility, the dimer form of our construct pUC18(85,0) contained two equal lengths of the GAA.TTC sequence (85 repeats) and this also failed to detect any sticky DNA conformation. These findings suggest that longer tracts of GAA.TTC sequence (>85 repeats) are required to form sticky DNA. In addition, plasmid incubation at room temperature or increasing $Mg^{+2}$ concentration may help long GAA repeats to form more stable triplexes, which in turn, generates sticky DNA conformation.

All the pUC18 plasmids appeared as multiple bands in the agarose gels and, by using limited restriction digestions, we confirmed that the low mobility plasmid species is a dimer that consists of two monomer plasmids covalently ligated together. This unusual behaviour of plasmids containing GAA.TTC sequences may be due to the higher recombinogenic capacity of the GAA repeats. This is in agreement with previous reports in which the recombinational hot spot characteristics may be a common feature of all

triplet repeat sequences (220, 284). In addition, H-DNA formed by GAA.TTC repeats may cause stalling of the DNA polymerase during DNA replication and then replication fork arrest. This impediment may cause replication errors in which these inserts interfere with the normal termination process and subsequently the polymerase keeps tracking around the plasmid leading to the dimerisation of the plasmid. The duplex part of the DNA triplex prefer to adopt A-like helix, thus A-form helices could play a role in some biological process including recombination (89). Because single-stranded region of the H-DNA structure is more susceptible to damage, our findings suggests that H-DNA is a source of genetic instability resulting from double-strand breaks and, in turn, enhance homologous recombination between plasmids. The plasmid containing two long tracts of 85 and 62 GAA repeats was examined by treating with topoisomerase I, in parallel with the other plasmid constructs, and yielding large relaxation of each plasmid. Each reaction product showed an almost identical pattern to that of the pUC18 control. This result indicated an absence of any gross supercoil-dependent structures, especially sticky DNA.

The experiments described in Chapter 4 investigated the formation of intramolecular triplexes *in vitro* with short (< 59) and long (≥59) repeats of the GAA.TTC sequences. Two different chemical agents DEPC and $KMnO_4$ as well as S1 nuclease were used to identify local structural transitions under native superhelicity. The use of DEPC and $KMnO_4$ together provided us with a complete analysis of both the GAA- and TTC-containing DNA strands. On the basis of chemical and enzymatic results, there was strong evidence for H-DNA formation in short $(GAA.TTC)_n$ sequences ($n \leq 29$) and some evidence in longer inserts. Two types of H-DNA, (PyPu.Py and PyPu.Pu) with different isomers were adopted by short GAA repeats (≤ 29). However, we failed to detect any unique structural perturbations within the longer repeats which could be due to formation of several multiple short intramolecular triplexes. Previous investigations support this result, in which long Py.Pu tracts, including GAA.TTC repeats, were shown to form two short triplexes instead of one long structure (4, 195, 254, 260). This preference may be related to destabilisation of a long triplex and its rearrangements into two shorter structures (37). In addition, parallel-stranded DNA, in which a central parallel-stranded segment was flanked by normal antiparallel-stranded regions, has been proposed to form by GAA repeats (285). Thus, it is likely that a parallel-stranded loop could be formed within our long duplex (GAA.TTC) repeats. The major conclusion from this work is that chemical and enzymatic probing analysis of sufficiently long GAA.TTC repeats (≥39

repeats) is complicated and unable to confirm the type of H-DNA that is formed. Unusual DNA structures such as parallel duplexes and multiple short triplexes might explain the difficulty in obtaining a unique analysis by chemical or enzymatic probes. However, structural models for long GAA repeats may need additional experimental support. The chemical and enzymatic probing experiments on supercoiled plasmids that are described in Chapter 4 are time consuming; this limited the number of sequences that could be investigated.

Long GAA repeats in Friedreich's ataxia patients which have the potential to adopt H-DNA induce large deletions of the repeats, likely resulting from slippage events during DNA synthesis (286). Sakamoto *et al*., in 1999 (3) found that the instability in this region was initiated by interaction of two triplex structures forming sticky DNA. Plasmids containing short GAA repeats, which does not form sticky DNA, no heterogeneous fragments were detected, suggesting that the sticky DNA structure play a significant role the genetic instability. However, in our work, the H-DNA-forming GAA.TTC repeats examined are short (19–29 triplet repeats) and are not capable of adopting secondary structures other than H-DNA.

In healthy people, H-DNA formation by GAA repeats could play a significant role in transcription efficiency which affect nucleosome organization and nucleosome phasing from the triplex region (256). In addition, triplex formation could provide a target binding site for RNA polymerase (33). Various structures could be formed by expanded GAA repeats including one long intramolecular triplex, two or more short intramolecular triplexes, sticky DNA and a parallel stranded loop. These structures can lead to a simple loss of frataxin protein in Friedreich ataxia patients by interfering with the transcription of the frataxin gene.

Finally, Chapter 5 examined the effect of GAA.TTC repeats on gene expression *in vivo*, using a set of luciferase reporter constructs harbouring different lengths of GAA repeats in two different positions, under native superhelicity and linear DNA forms. This revealed a significant decrease in gene expression in supercoiled plasmid constructs compared to the supercoiled control plasmid. Their linear counterparts exhibited a very weak effect compared to the linear control. However, linear DNA forms of the constructs `produced much lower luciferase expression relative to the supercoiled forms, indicating that supercoiling is important for expression as linear DNA is much more prone to degradation

by exonucleases than closed circular DNA. These results indicated that H-DNA-forming GAA.TTC repeats can modulate gene expression at the transcriptional level. This correlates with previous studies which have shown that potential H-DNA forming sequences can suppresses gene expression (145, 163, 221, 271-273).

## 6.2 Future considerations

Although great progress has been made into the understanding the molecular mechanisms involved in FRDA, significant aspects need further clarification and investigation, particularly those concerning the regulation of the FXN gene and the dynamics of GAA repeat tracts at the FXN gene. The capacity of long GAA.TTC repeats to form non-B DNA structures in *E.coli* has marked biological consequences, and supports the possible existence of this conformation in FRDA patient cells (125, 137). However, several questions remain unanswered *i.e.* Does sticky DNA really exist and function in humans? If it forms, what is the lifetime of this structure? What is the detailed molecular structure of sticky DNA *in vitro* and *in vivo*? In addition, an understanding of the structural behaviour of the GAA.TTC repeat is important to develop new therapeutic approaches for FRDA. Therefore, new experimental methodologies are needed to solve this issue.

# Appendices

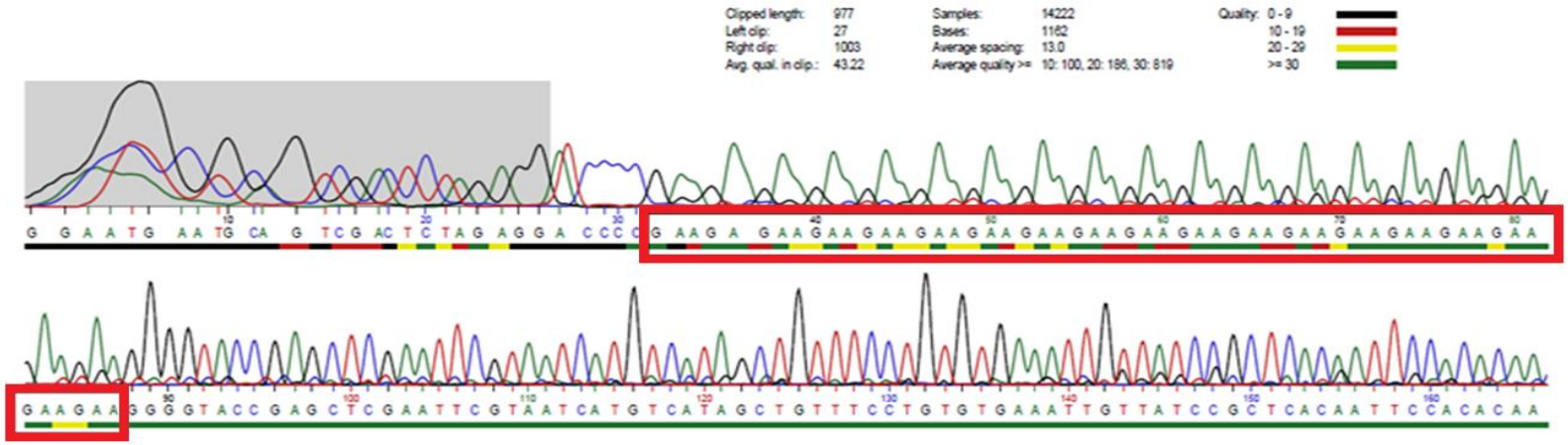



**Figure I:** 19 GAA.TTC repeats in SmaI site of plasmid pUC18

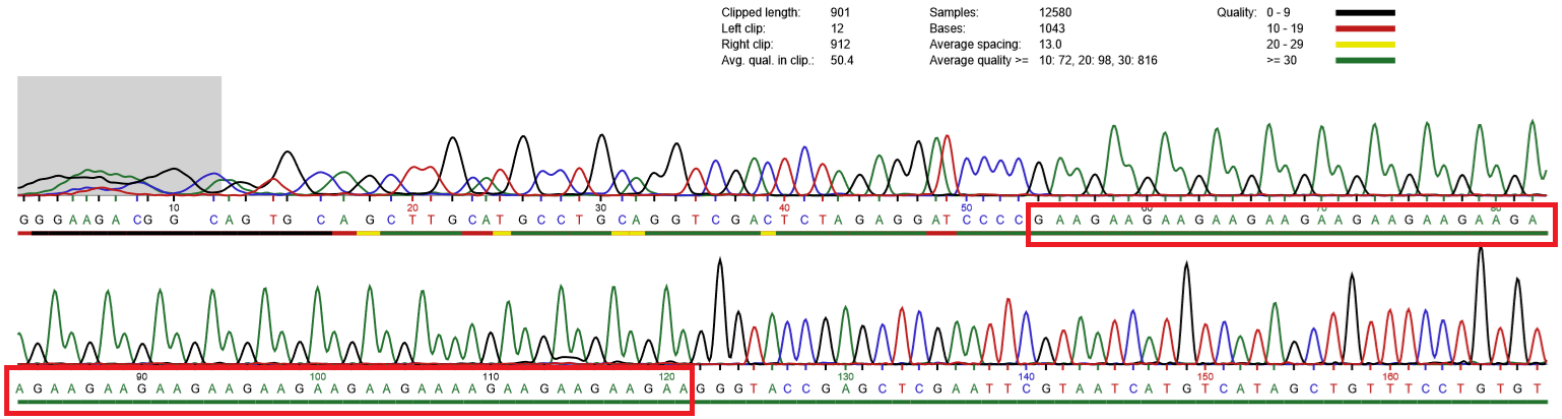



**Figure II:** 22 GAA.TTC repeats in SmaI site of plasmid pUC18

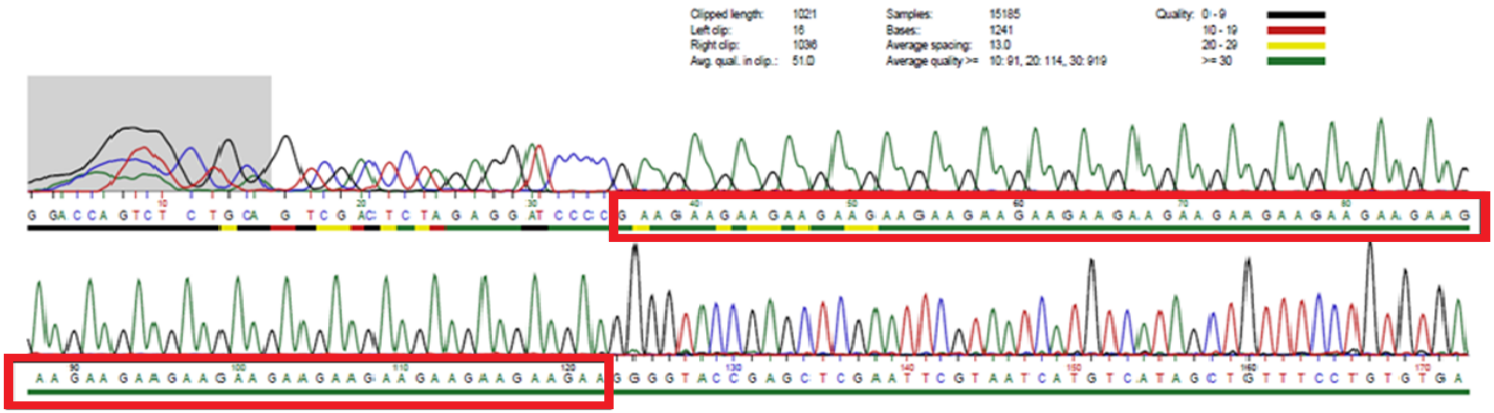



**Figure III:** 29 GAA.TTC repeats in SmaI site of plasmid pUC18

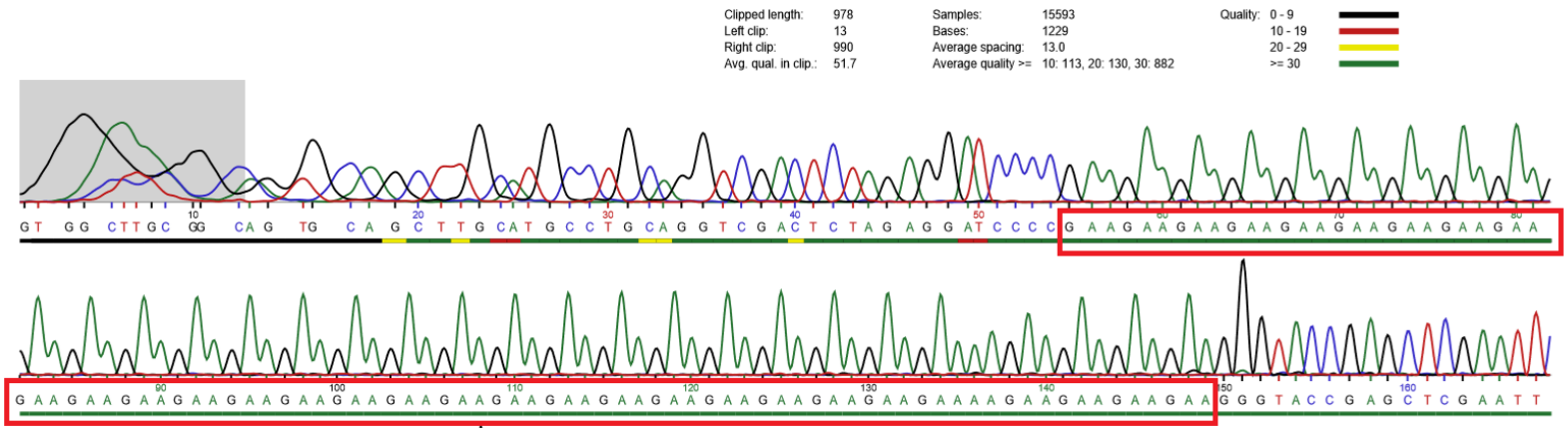



**Figure IV:** 31 GAA.TTC repeats in SmaI site of plasmid pUC18

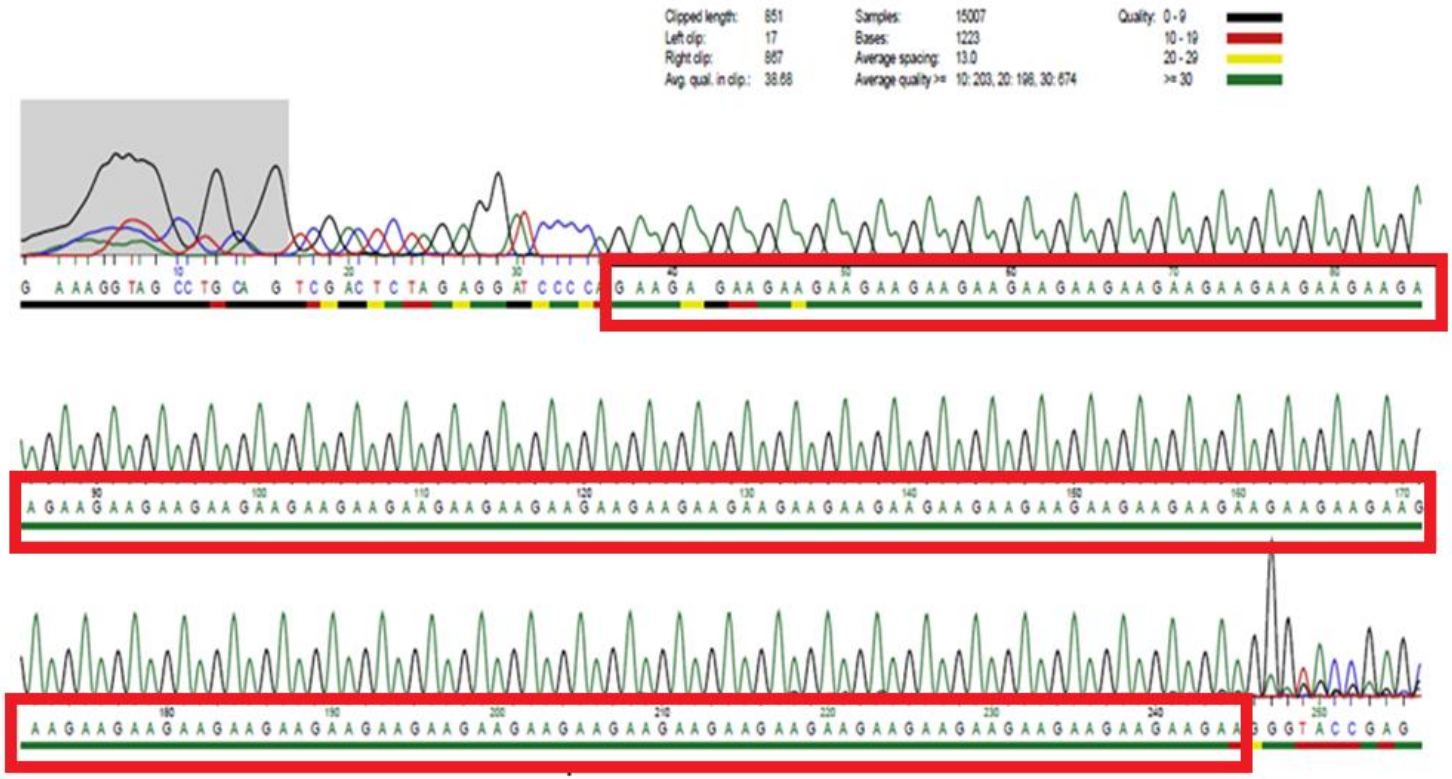



**Figure V:** 70 GAA.TTC repeats in SmaI site of plasmid pUC18

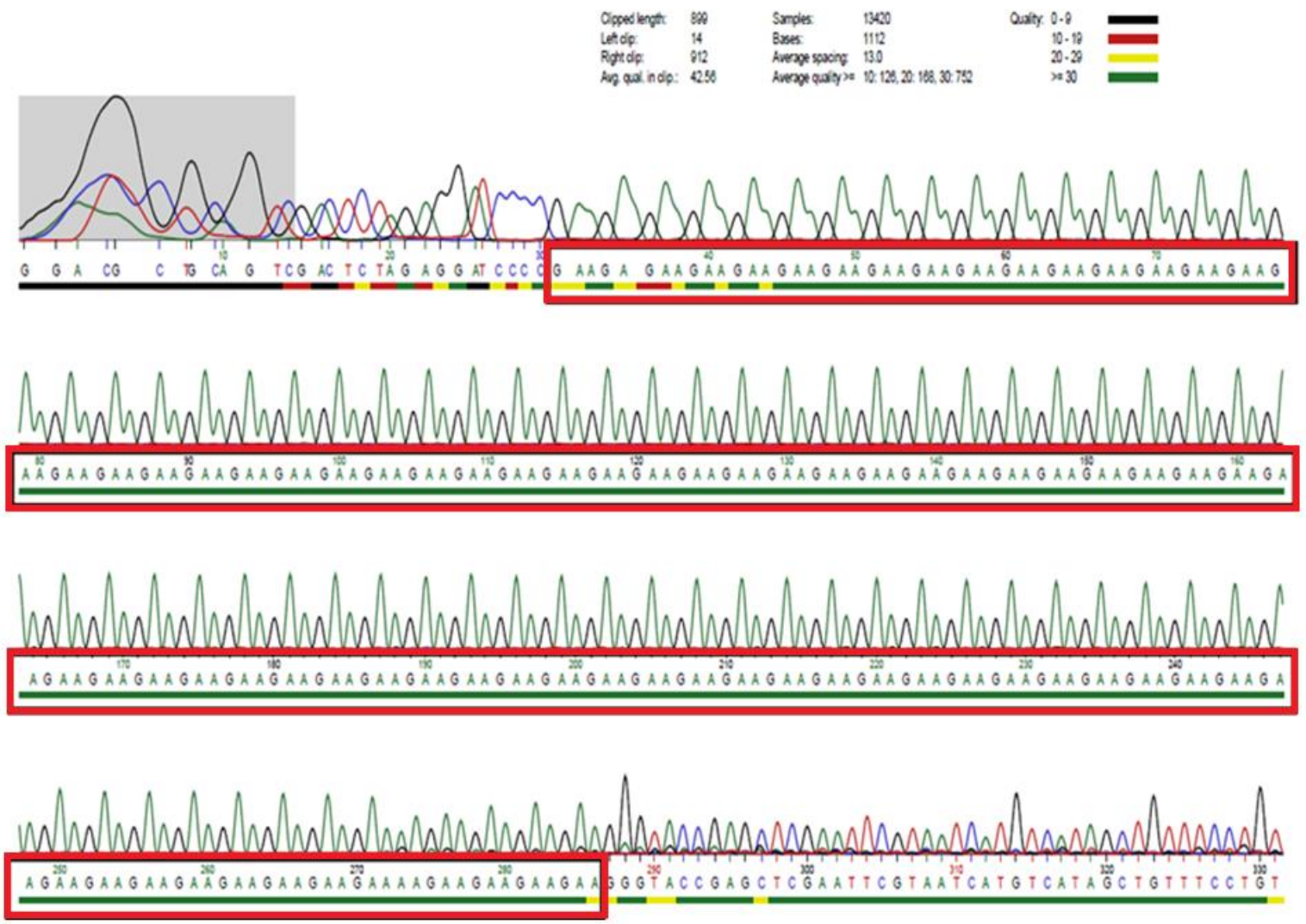



**Figure VI:** 85 GAA.TTC repeats in SmaI site of plasmid pUC18

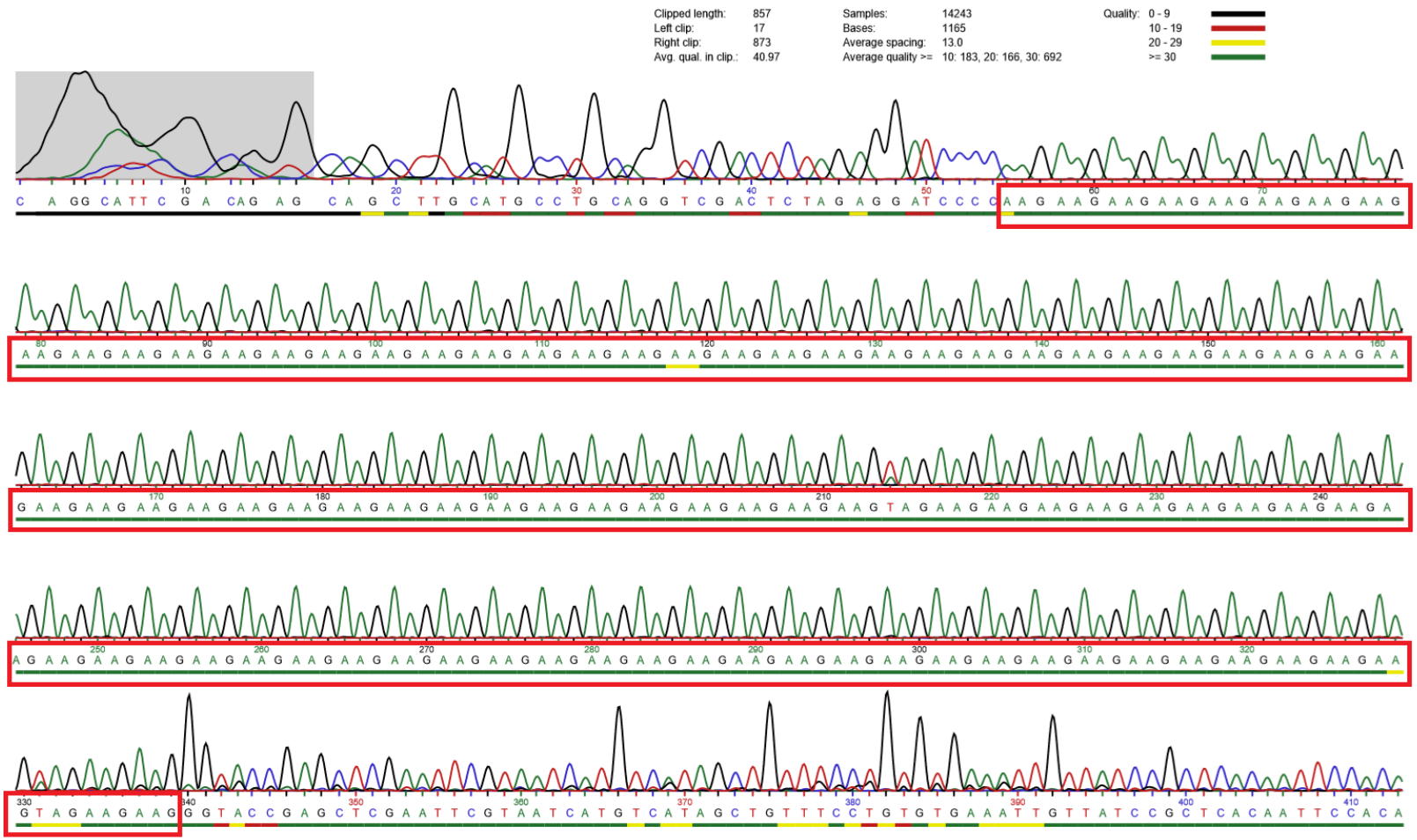



**Figure VII:** 94 GAA.TTC repeats in SmaI site of plasmid pUC18

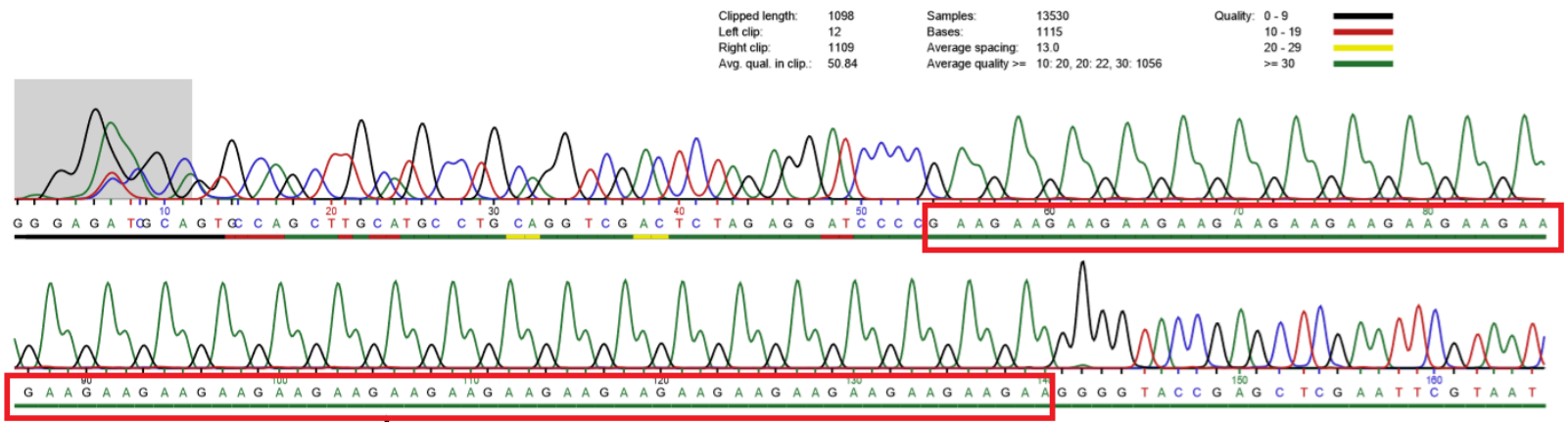



**Figure VIII:** 29 GAA.TTC repeats in SmaI site of (plasmid pUC18 + 29 GAA.TTC repeats in AatII site)

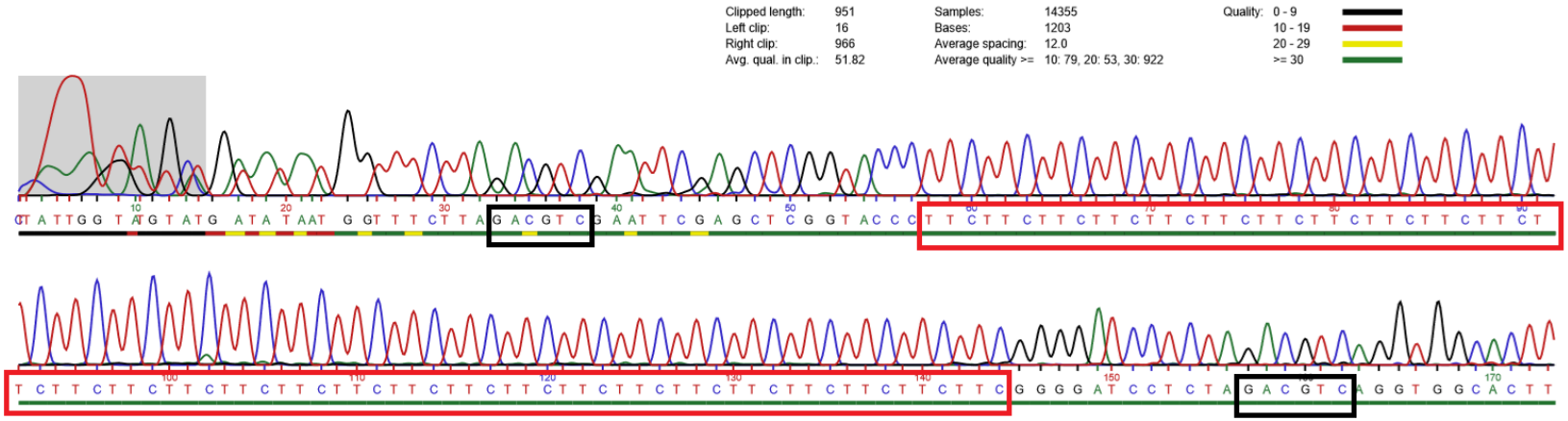



**Figure IX:** 29 GAA.TTC repeats in AatII site of (plasmid pUC18 + 29 GAA.TTC repeats in SmaI site), black boxes indicate AatII sites

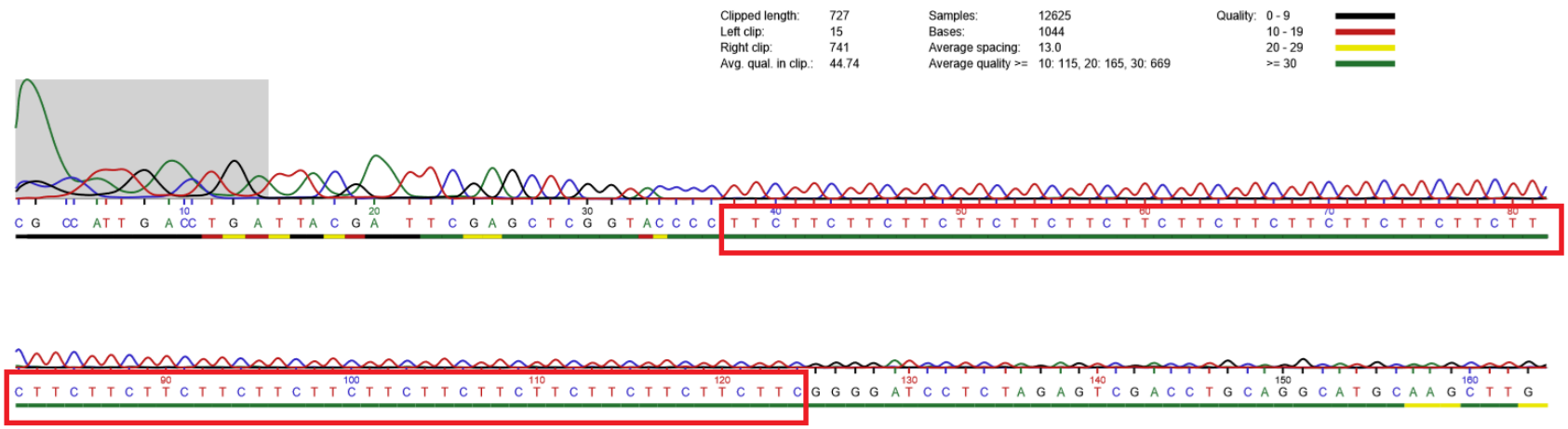



**Figure X:** 29 GAA.TTC repeats in SmaI site of (plasmid pUC18 + 29 GAA.TTC repeats in AatII site, indirect orientation)

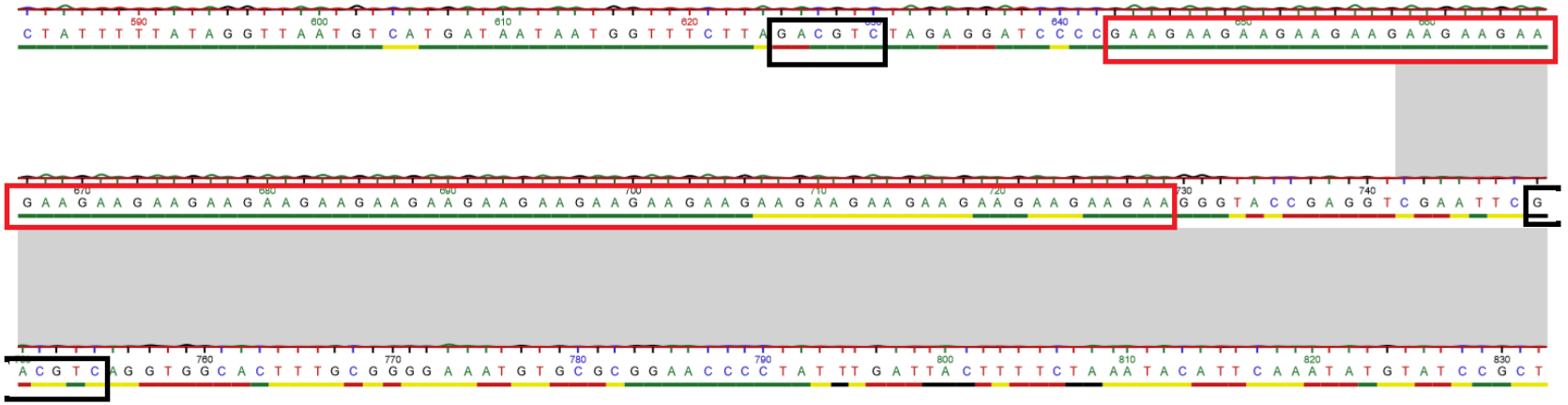


**Figure XI:** 29 GAA.TTC repeats in AatII site of (plasmid pUC18 + 29 GAA.TTC repeats in SmaI site, indirect orientation), black boxes indicate AatII sites

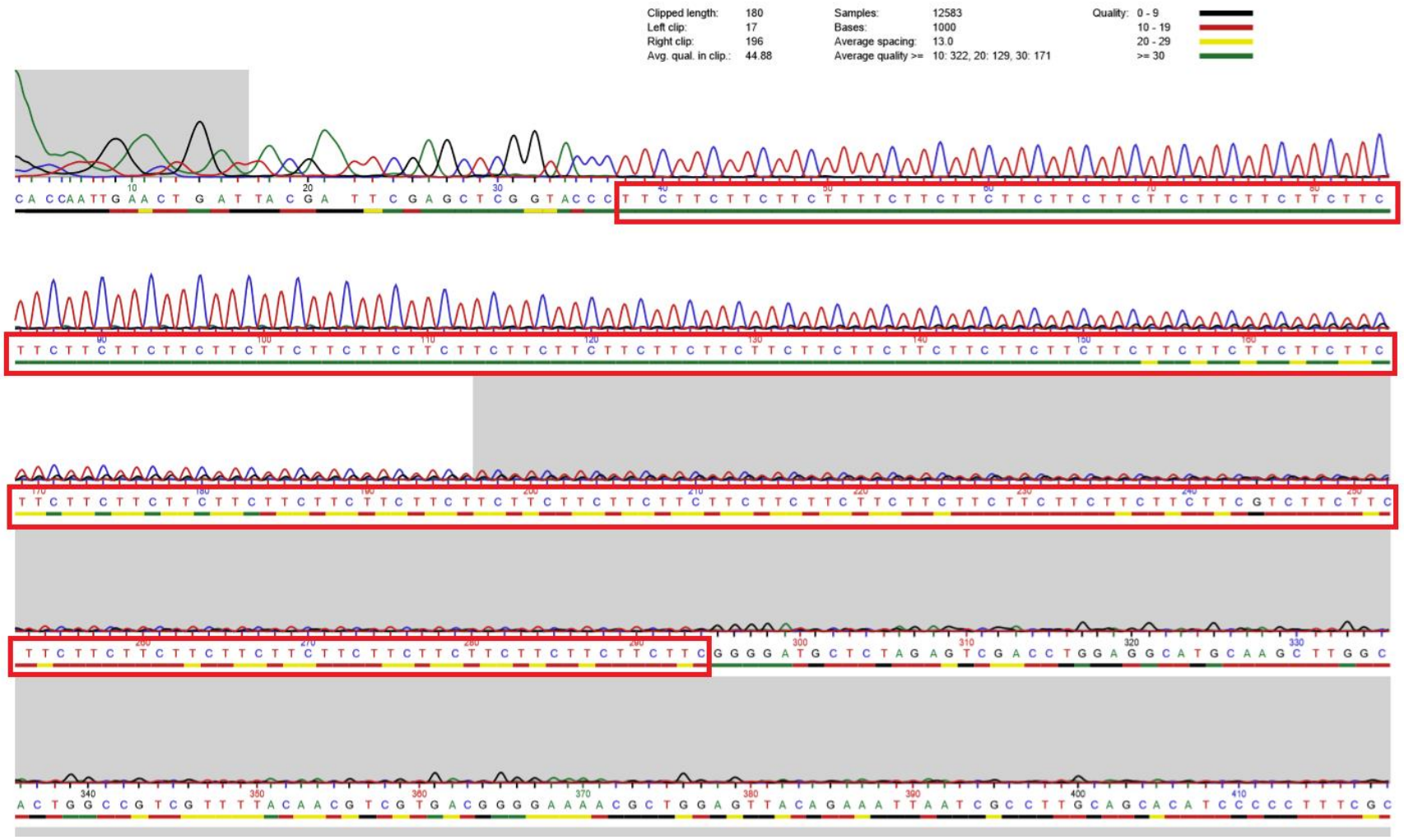



**Figure XII:** 85 GAA.TTC repeats in SmaI site of (plasmid pUC18 + 39 GAA.TTC repeats in AatII site)

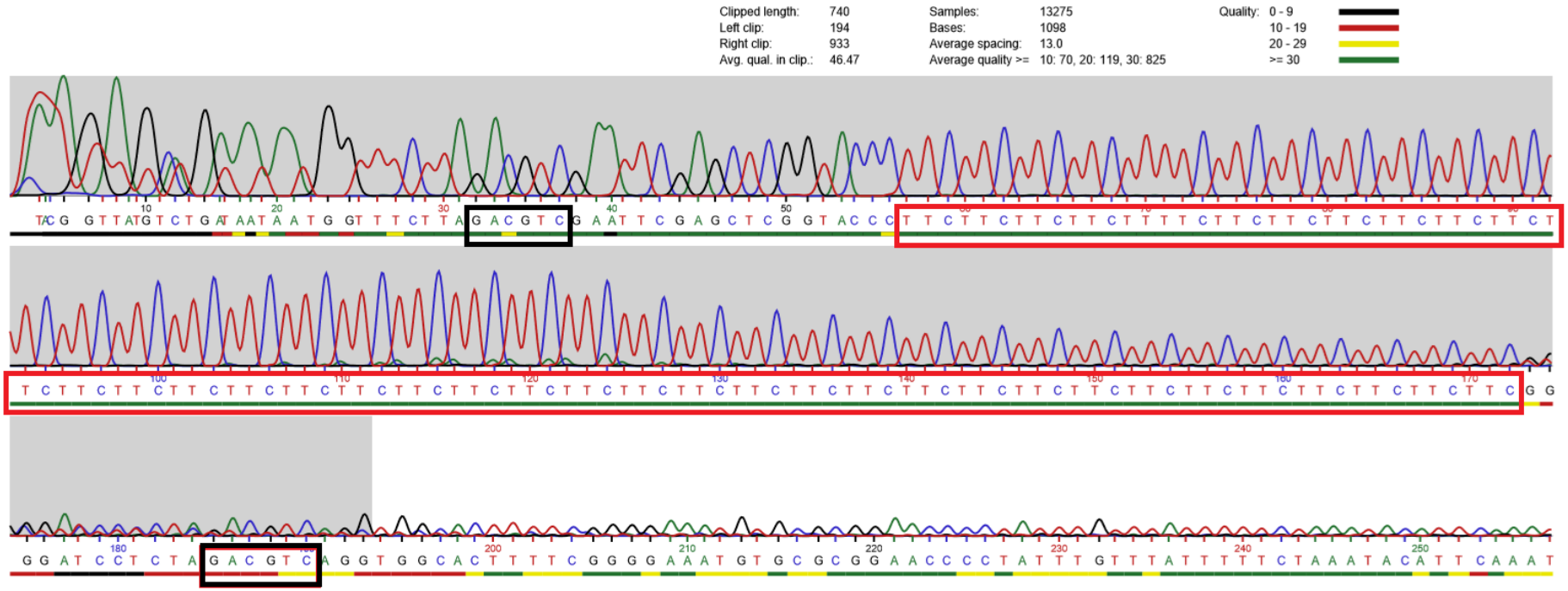



**Figure XIII:** 39 GAA.TTC repeats in AatII site of (plasmid pUC18 + 85 GAA.TTC repeats in SmaI site), black boxes indicate AatII sites

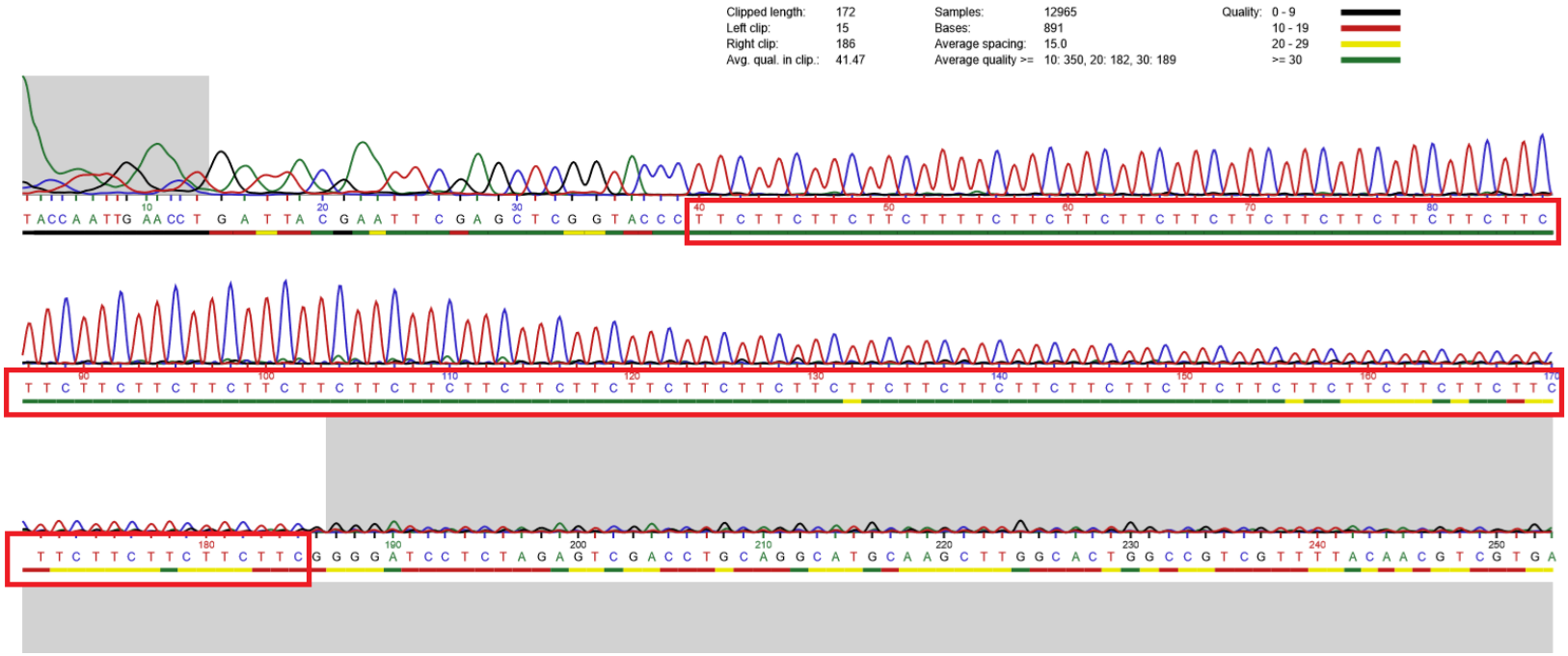



**Figure XIV:** 43 GAA.TTC repeats in SmaI site of (plasmid pUC18 + 84 GAA.TTC repeats in AatII site)


**Figure XV:** 84 GAA.TTC repeats in AatII site of (plasmid pUC18 + 43 GAA.TTC repeats in SmaI site), black box indicate AatII site

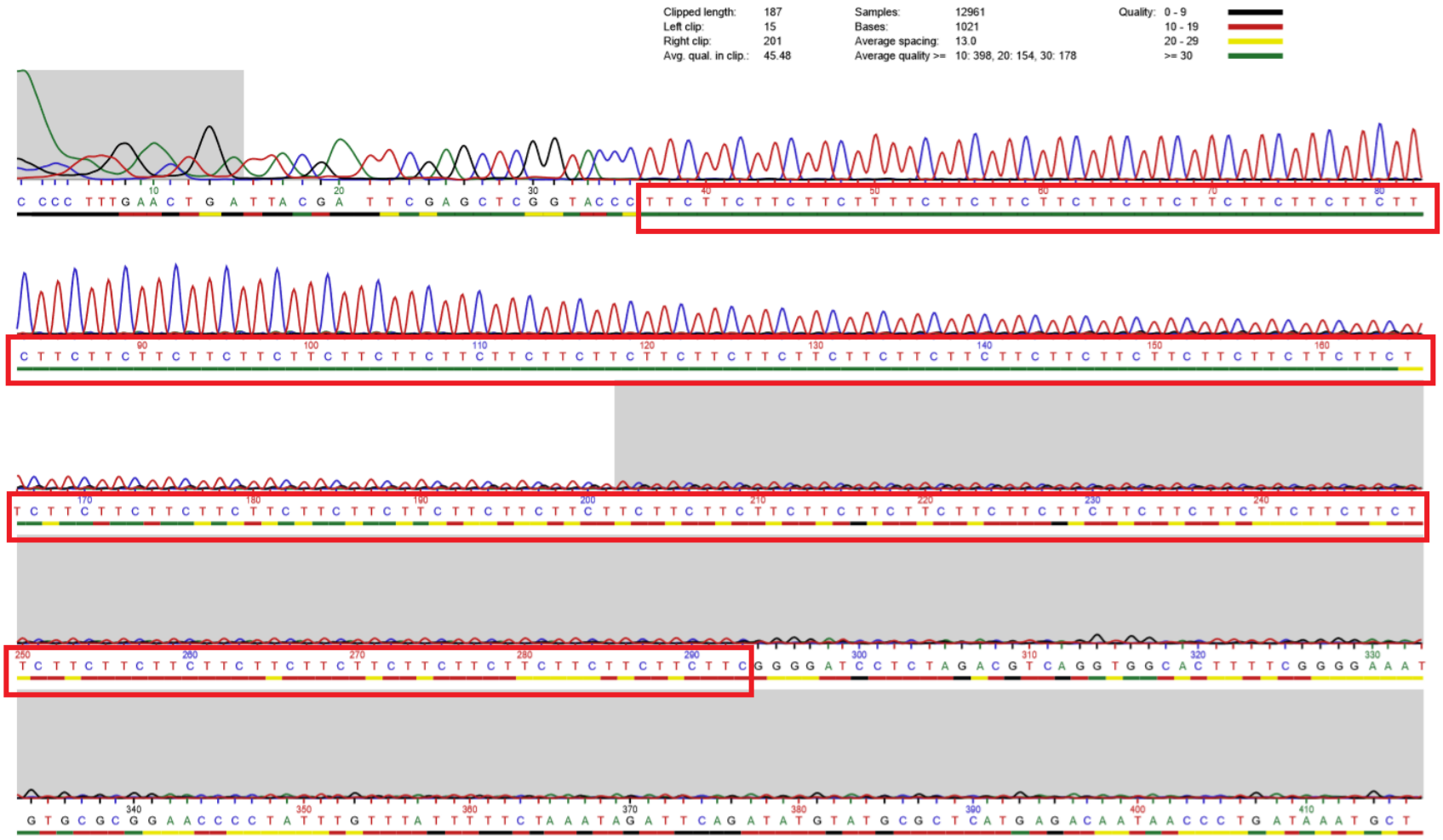



**Figure XVI:** 85 GAA.TTC repeats in SmaI site of (plasmid pUC18 + 62 GAA.TTC repeats in AatII site)

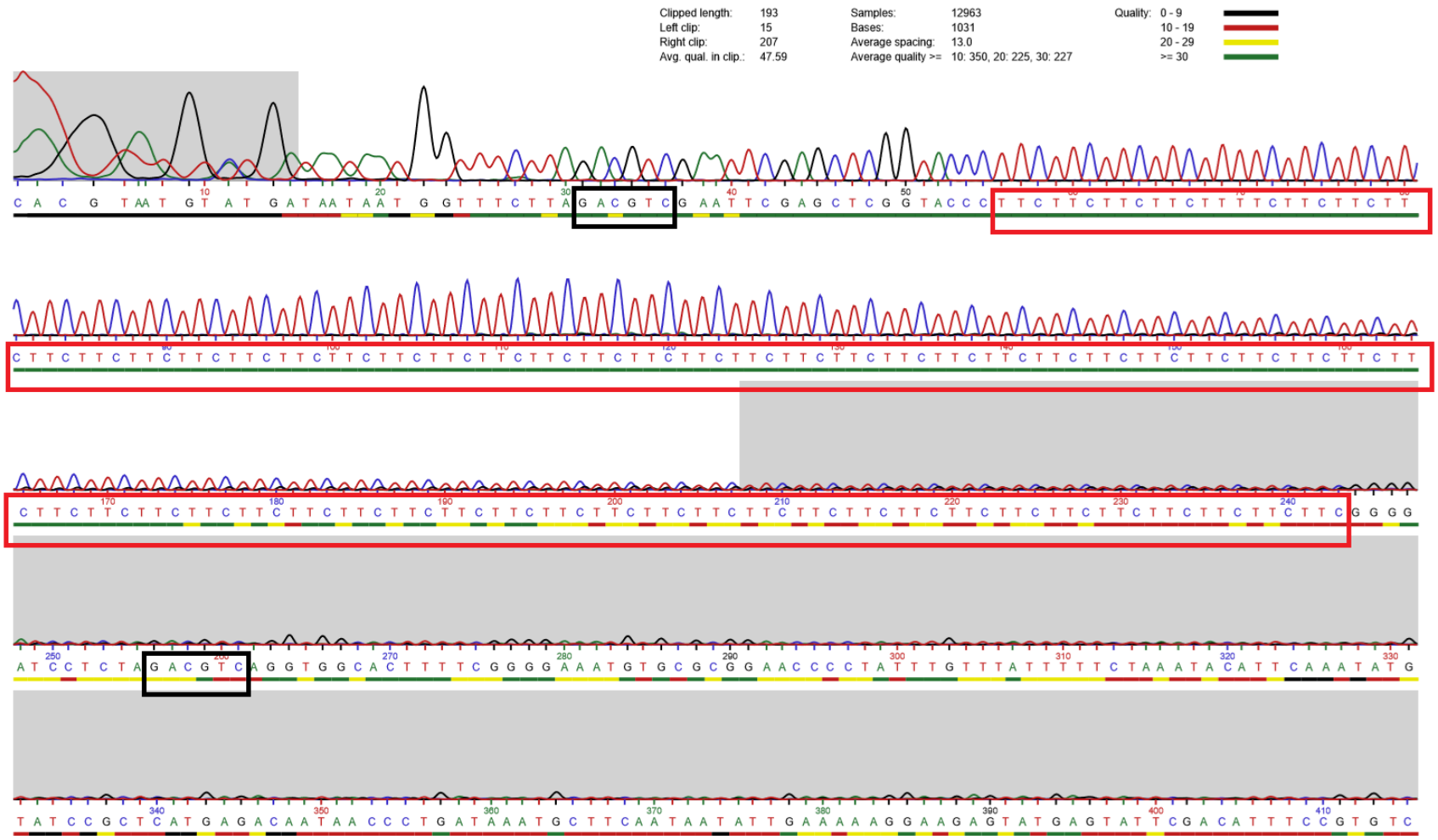



**Figure XVII:** 62 GAA.TTC repeats in AatII site of (plasmid pUC18 + 85 GAA.TTC repeats in SmaI site), black boxes indicate AatII sites

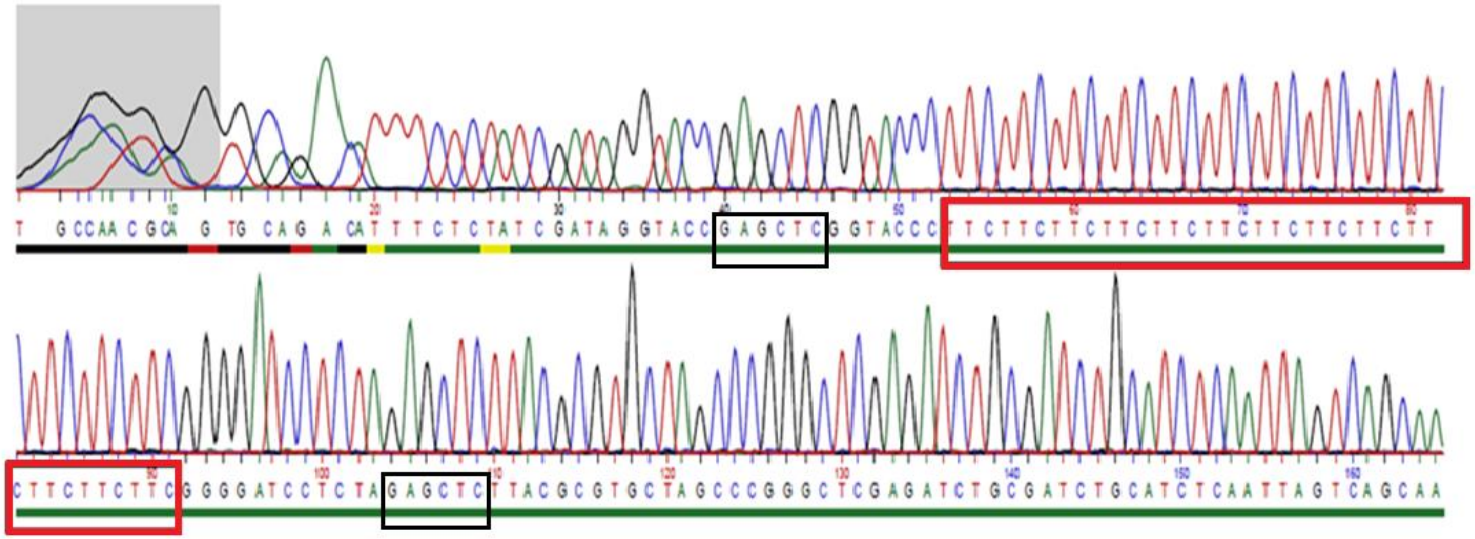


**Figure XVIII:** 13 GAA.TTC repeats in SacI site of plasmid pGL3-control, Black boxes indicate SacI sites

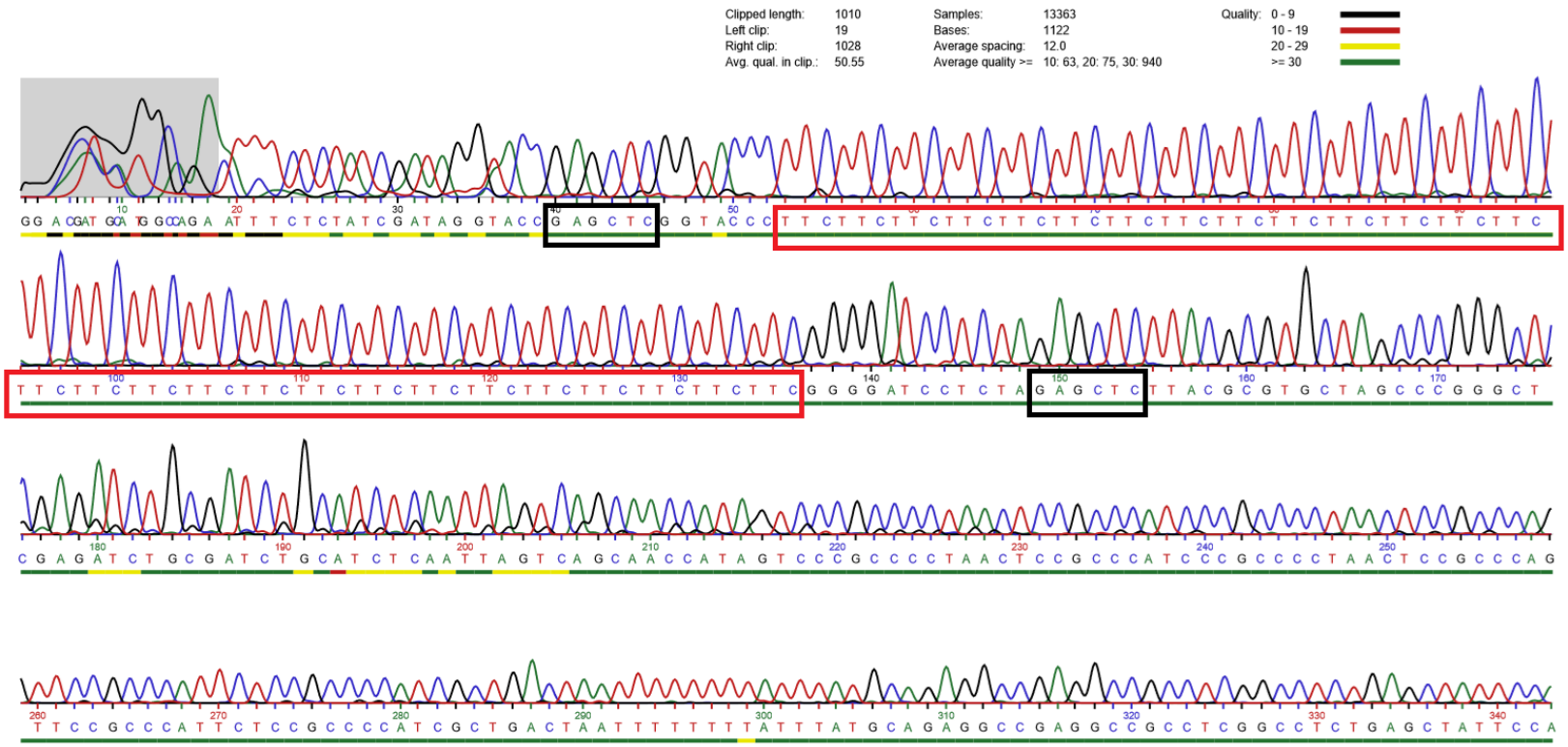



**Figure XIX:** 28 GAA.TTC repeats in SacI site of plasmid pGL3-control, Black boxes indicate SacI sites

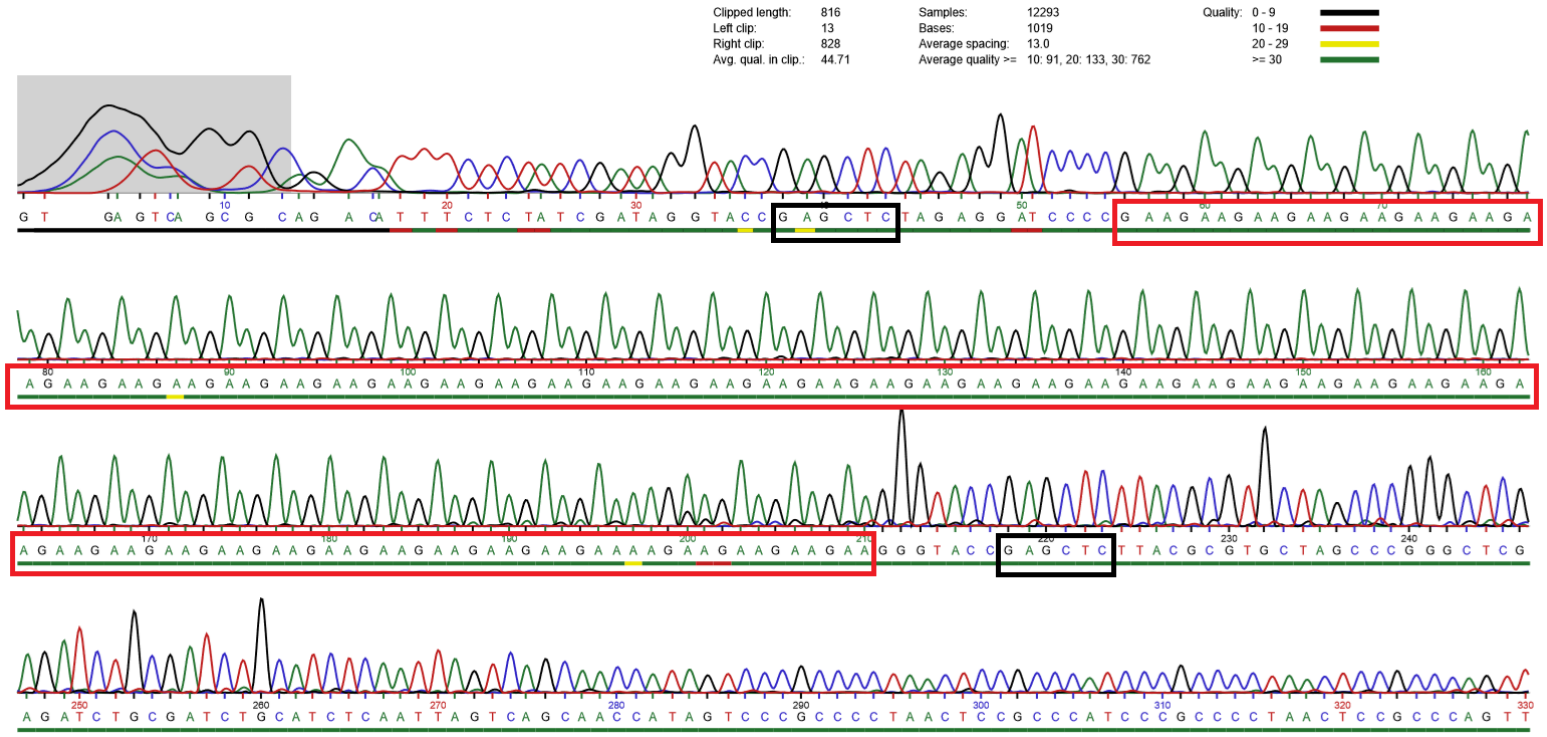



**Figure XX:** 51 GAA repeats in SacI site of plasmid pGL3-control, Black boxes indicate SacI sites

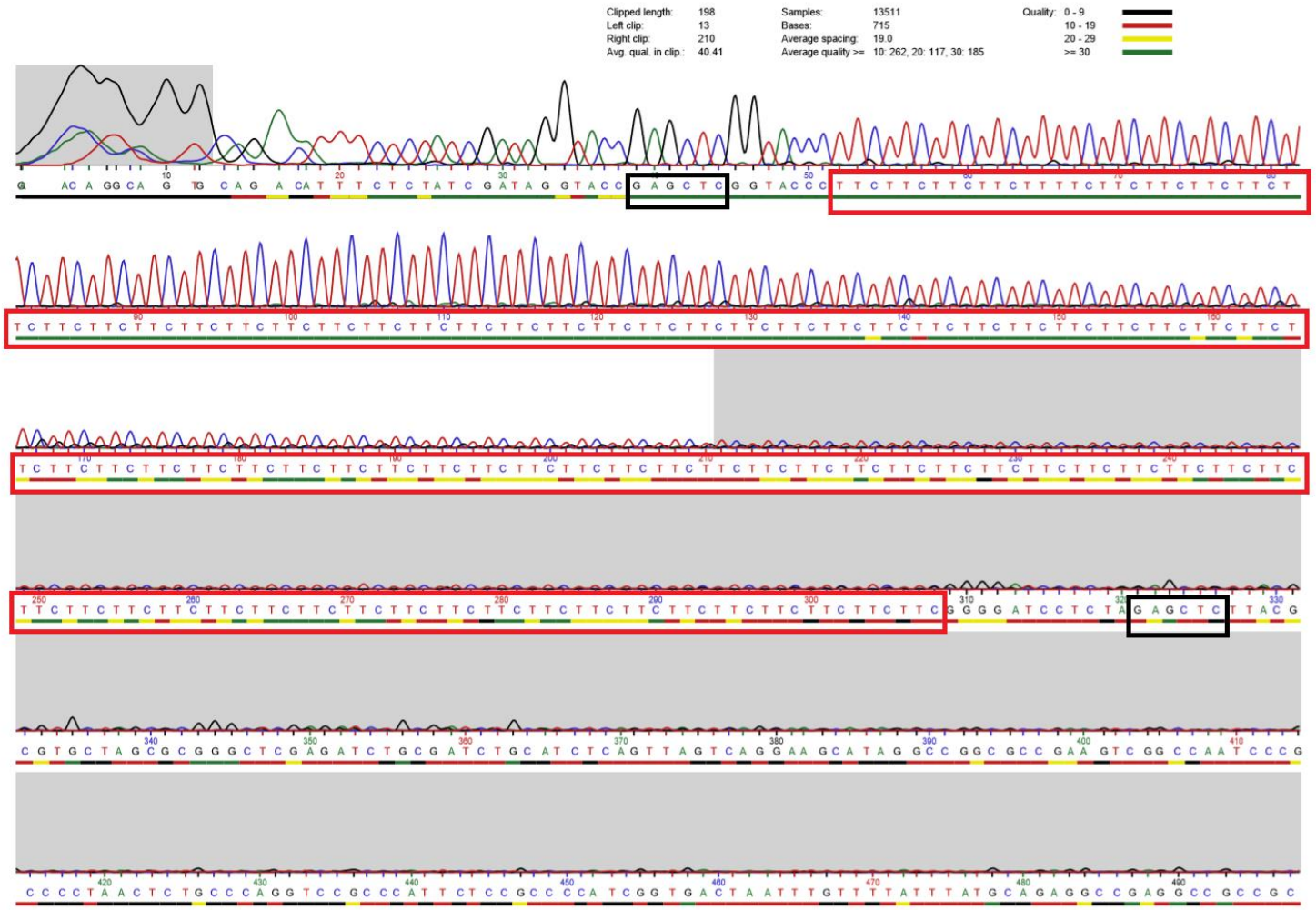



**Figure XXI:** 85 TTC repeats in SacI site of plasmid pGL3-control, Black boxes indicate SacI sites

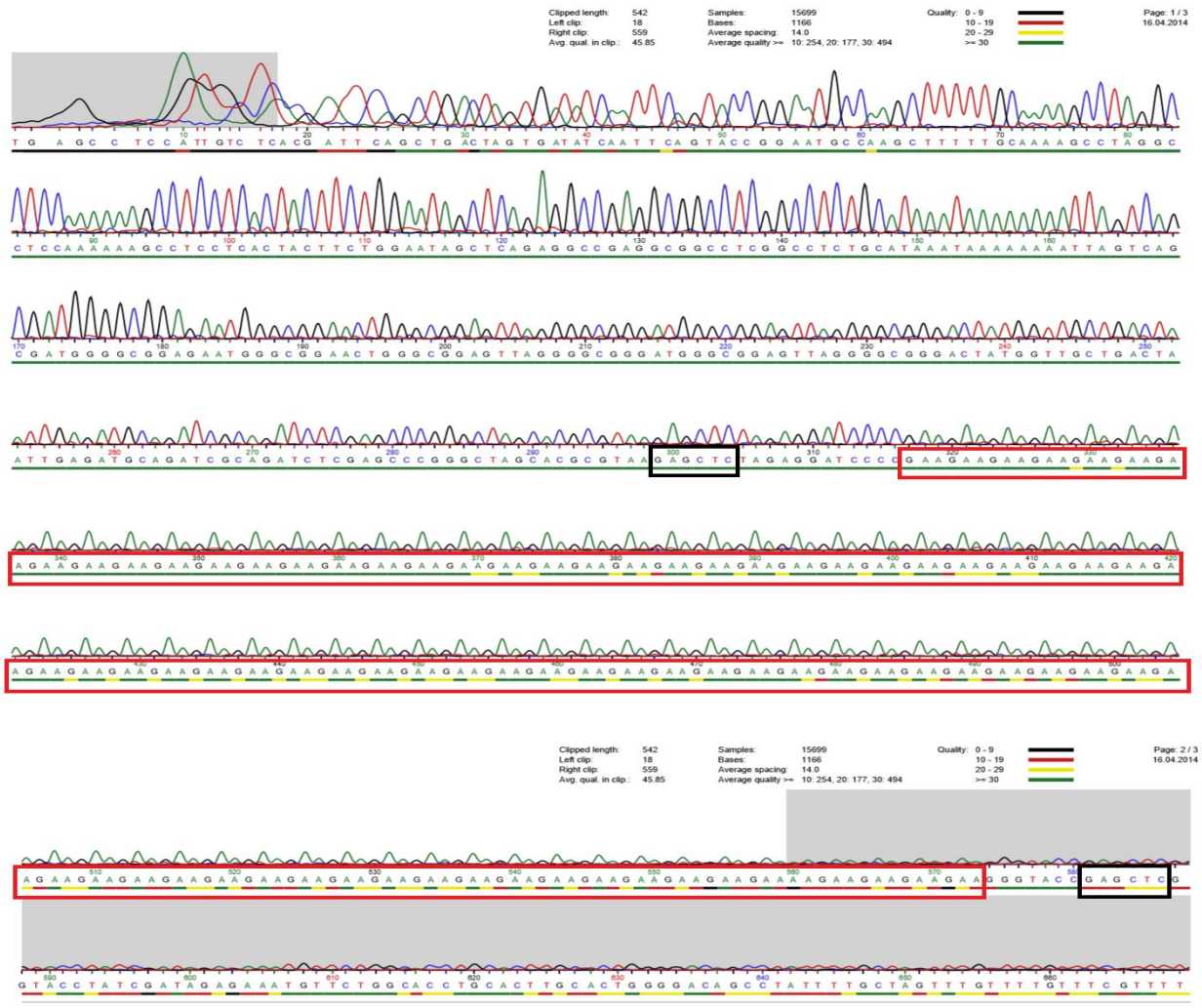


**Figure XXII:** 85 GAA.TTC repeats in SacI site of (plasmid pGL3-control + 85 GAA.TTC repeats in PciI site), Black boxes indicate SacI site

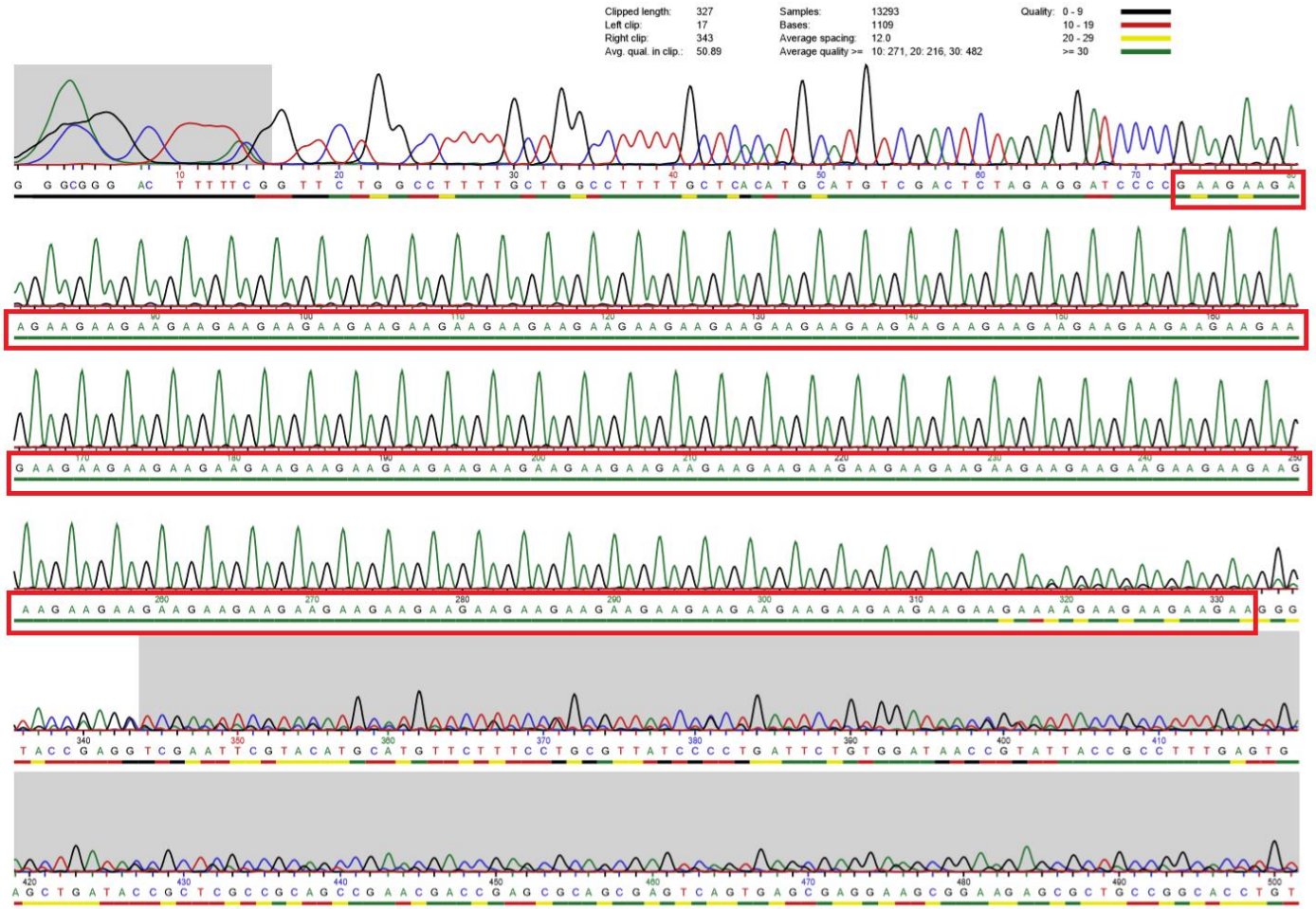



**Figure XXIII:** 85 GAA.TTC repeats in PciI site of (plasmid pGL3-control + 85 GAA.TTC repeats in SacI site)

# References


1. Bowater RP, Wells RD. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog Nucleic Acid Res Mol Biol*. 2001;66:159-202.

2. Pandolfo M. The molecular basis of Friedreich ataxia. *Adv Exp Med Biol*. 2002;516:99-118.

3. Sakamoto N, Chastain PD, Parniewski P, Ohshima K, Pandolfo M, Griffith JD, *et al*. Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol Cell. 1999;3(4):465-75.

4. Potaman VN, Oussatcheva EA, Lyubchenko YL, Shlyakhtenko LS, Bidichandani SI, Ashizawa T, *et al*. Length-dependent structure formation in Friedreich ataxia $(GAA)_n.(TTC)_n$ repeats at neutral pH. *Nucleic Acids Res*. 2004;32(3):1224-1231.

5. Krasilnikova MM, Kireeva ML, Petrovic V, Knijnikova N, Kashlev M, Mirkin SM. Effects of Friedreich's ataxia $(GAA)_n.(TTC)_n$ repeats on RNA synthesis and stability. *Nucleic Acids Res.* 2007;35(4):1075-1084.

6. Son LS, Bacolla A, Wells RD. Sticky DNA: *in vivo* formation in *E. coli* and *in vitro* association of long GAA.TTC tracts to generate two independent supercoiled domains. *J Mol Biol.* 2006;360(2):267-284.

7. Crick FHC, Watson JD. The complementary structure of deoxyribonucleic acid. Proceedings of the Royal Society of London, Series A, *Mathematica and Physical Sciences*. 1954;223:80-96.

8. Blackburn GM, Gait MJ, Loakes D, Williams DM. Nucleic acids in chemistry and biology. 3rd ed: *Royal Society of Chemistry*; 2006.

9. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. *Annu Rev Biochem*. 1995;64:65-95.

10. Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem*. 2004;279(46):47411-47414.

11. Belmont PC, J.-F.; Demeunynck, M. . Nucleic acid conformation diversity: from structure to function and regulation. *Chem Soc Rev* 2001; 30: 70-81.

12. Neidle S. Principles of Nucleic Acid Structure. 1st ed: *Oxford University Press*; 2008.

13. Hoheisel JD, Pohl FM. Searching for potential Z-DNA in genomic Escherichia coli DNA. *J Mol Biol*. 1987;193(3):447-464.

14. Harvey SC. DNA structural dynamics: longitudinal breathing as a possible mechanism for the B in equilibrium Z transition. *Nucleic Acids Res*. 1983;11(14):4867-4878.

15. Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G, *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*. 1979;282(5740):680-686.

16. Wang G, Vasquez KM. Z-DNA, an active element in the genome. *Front Biosci.* 2007;12:4424-4438.

17. Rich A. DNA comes in many forms. *Gene*. 1993;135(1-2):99-109.

18. Wells RD, Dere R, Hebert ML, Napierala M, Son LS. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* 2005;33(12):3785-3798.

19. Mirkin SM. Discovery of alternative DNA structures: a heroic decade (1979-1989). *Front Biosci.* 2008;13:1064-1071.

20. Wells RD. DNA triplexes and Friedreich ataxia. *FASEB J.* 2008;22(6):1625-1634.

21. Felsenfeld G, Rich A. Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta*. 1957;26(3):457-468.

22. Beal PA, Dervan PB. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science.* 1991;251(4999):1360-1363.

23. Morgan AR, Wells RD. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J Mol Biol.* 1968;37(1):63-80.

24. Marck C, Thiele D. Poly(dG).poly(dC) at neutral and alkaline pH: the formation of triple stranded poly(dG).poly(dG).poly(dC). *Nucleic Acids Res.* 1978;5(3):1017-1028.

25. Broitman SL, Im DD, Fresco JR. Formation of the triple-stranded polynucleotide helix, poly(A.A.U). *Proc Natl Acad Sci U S A.* 1987;84(15):5120-5124.

26. Vasquez KM, Glazer PM. Triplex-forming oligonucleotides: principles and applications. *Q Rev Biophys*. 2002;35(1):89-107.

27. Moser HE, Dervan PB. Sequence-specific cleavage of double helical DNA by triple helix formation. *Science.* 1987;238(4827):645-650.

28. Le Doan T, Perrouault L, Praseuth D, Habhoub N, Decout JL, Thuong NT, *et al.* Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res*. 1987;15(19):7749-7760.

29. Radhakrishnan I, Patel DJ. Hydration sites in purine.purine.pyrimidine and pyrimidine.purine.pyrimidine DNA triplexes in aqueous solution. *Structure.* 1994;2(5):395-405.

30. Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol Cell Biol.* 2007;27(6):2130-2143.

31. Wasserman SA, Cozzarelli NR. Biochemical topology: applications to DNA recombination and replication. *Science.* 1986;232(4753):951-960.

32. Jain A, Wang G, Vasquez KM. DNA triple helices: biological consequences and therapeutic potential. *Biochimie*. 2008;90(8):1117-1130.

33. Lee JS, Woodsworth ML, Latimer LJ, Morgan AR. Poly(pyrimidine) . poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* 1984;12(16):6603-6614.

34. Ulrich MJ, Gray WJ, Ley TJ. An intramolecular DNA triplex is disrupted by point mutations associated with hereditary persistence of fetal hemoglobin. *J Biol Chem*. 1992;267(26):18649-18658.

35. Kohwi Y, Kohwi-Shigematsu T. Altered gene expression correlates with DNA structure. *Genes Dev*. 1991;5(12B):2547-2554.

36. Radhakrishnan I, Patel DJ. Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure*. 1993;1(2):135-152.

37. Soyfer VN PV. Triple-Helical Nucleic Acids. NewYork1996. 360 p.

38. Lee JS, Johnson DA, Morgan AR. Complexes formed by $(pyrimidine)_n$ . $(purine)_n$ DNAs on lowering the pH are three-stranded. *Nucleic Acids Res*. 1979;6(9):3073-3091.

39. Mirkin SM, Frank-Kamenetskii MD. H-DNA and related structures. *Annu Rev Biophys Biomol Struct*. 1994;23:541-576.

40. Griffin LC, Dervan PB. Recognition of thymine adenine.base pairs by guanine in a pyrimidine triple helix motif. *Science*. 1989;245(4921):967-971.

41. Yoon K, Hobbs CA, Koch J, Sardaro M, Kutny R, Weis AL. Elucidation of the sequence-specific third-strand recognition of four Watson-Crick base pairs in a pyrimidine triple-helix motif: T.AT, C.GC, T.CG, and G.TA. *Proc Natl Acad Sci U S A.* 1992;89(9):3840-3844.

42. Radhakrishnan I, Patel DJ. DNA triplexes: solution structures, hydration sites, energetics, interactions, and function. *Biochemistry.* 1994;33(38):11405-11416.

43. Kohwi Y, Kohwi-Shigematsu T. Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc Natl Acad Sci* U S A. 1988;85(11):3781-3785.

44. Beal PA, Dervan PB. The influence of single base triplet changes on the stability of a pur.pur.pyr triple helix determined by affinity cleaving. *Nucleic Acids Res.* 1992;20(11):2773-2776.

45. Fox KR. Targeting DNA with triplexes. *Curr Med Chem*. 2000;7(1):17-37.

46. Arnott S, Selsing E. Structures for the polynucleotide complexes poly(dA) with poly (dT) and poly(dT) with poly(dA) with poly (dT). *J Mol Biol.* 1974;88(2):509-521.

47. Arnott S, Chandrasekaran R, Hukins DW, Smith PJ, Watts L. Structural details of double-helix observed for DNAs containing alternating purine and pyrimidine sequences. *J Mol Biol*. 1974;88(2):523-533.

48. Rajagopal P, Feigon J. Triple-strand formation in the homopurine:homopyrimidine DNA oligonucleotides $d(G.A)_4$ and $d(T.C)_4$. *Nature*. 1989;339(6226):637-640.

49. Rajagopal P, Feigon J. NMR studies of triple-strand formation from the homopurine-homopyrimidine deoxyribonucleotides $d(GA)_4$ and $d(TC)_4$. *Biochemistry*. 1989;28(19):7859-7870.

50. Rhee S, Han Z, Liu K, Miles HT, Davies DR. Structure of a triple helical DNA with a triplex-duplex junction. *Biochemistry.* 1999;38(51):16810-16815.

51. Sklenar V, Feigon J. Formation of a stable triplex from a single DNA strand. *Nature*. 1990;345(6278):836-838.

52. Sekharudu CY, Yathindra N, Sundaralingam M. Molecular dynamics investigations of DNA triple helical models: unique features of the Watson-Crick duplex. *J Biomol Struct Dyn*. 1993;11(2):225-244.

53. Han H, Dervan PB. Different conformational families of pyrimidine.purine.pyrimidine triple helices depending on backbone composition. *Nucleic Acids Res*. 1994;22(14):2837-2844.

54. Ouali M, Letellier R, Adnet F, Liquier J, Sun JS, Lavery R, *et al*. A possible family of B-like triple helix structures: comparison with the Arnott A-like triple helix. *Biochemistry*. 1993;32(8):2098-2103.

55. Radhakrishnan I, Patel DJ. Solution structure of a pyrimidine.purine.pyrimidine DNA triplex containing T.AT, $C^+$.GC and G.TA triples. *Structure.* 1994;2(1):17-32.

56. Phipps AK, Tarkoy M, Schultze P, Feigon J. Solution structure of an intramolecular DNA triplex containing 5-(1-propynyl)-2'-deoxyuridine residues in the third strand. *Biochemistry*. 1998;37(17):5820-5830.

57. Thuong NTaH, C. . Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angewandte ChemieInternational Edition in English.* 1993;32: 666-690.

58. Paes HM, Fox KR. Kinetic studies on the formation of intermolecular triple helices. *Nucleic Acids Res*. 1997;25(16):3269-3274.

59. Shimizu M, Konishi A, Shimada Y, Inoue H, Ohtsuka E. Oligo(2'-O-methyl)ribonucleotides. Effective probes for duplex DNA. *FEBS Lett.* 1992;302(2):155-158.

60. Durand M, Peloille S, Thuong NT, Maurizot JC. Triple-helix formation by an oligonucleotide containing one $(dA)_{12}$ and two $(dT)_{12}$ sequences bridged by two hexaethylene glycol chains. *Biochemistry.* 1992;31(38):9197-9204.

61. Hampel KJ, Crosson P, Lee JS. Polyamines favor DNA triplex formation at neutral pH. *Biochemistry*. 1991;30(18):4455-4459.

62. Blake RD, Massoulie J, Fresco JR. Polynucleotides. 8. A spectral approach to the equilibria between polyriboadenylate and polyribouridylate and their complexes. *J Mol Biol.* 1967;30(2):291-308.

63. Volker J, Klump HH. Electrostatic effects in DNA triple helices. *Biochemistry.* 1994;33(45):13502-13508.

64. Cheng AJ, Vandyke MW. Monovalent Cation Effects on Intermolecular Purine-Purine-Pyrimidine Triple-Helix Formation. *Nucleic Acids Res*. 1993;21(24):5630-5635.

65. Olivas WM, Maher LJ. Overcoming potassium-mediated triplex inhibition. *Nucleic Acids Res.* 1995;23(11):1936-1941.

66. Floris R, Scaggiante B, Manzini G, Quadrifoglio F, Xodo LE. Effect of cations on purine.purine.pyrimidine triple helix formation in mixed-valence salt solutions. *Eur J Biochem.* 1999;260(3):801-809.

67. Malkov VA, Soyfer VN, Frankkamenetskii MD. Effect of intermolecular triplex formation on the yield of cyclobutane photodimers in DNA. *Nucleic Acids Res*. 1992;20(18):4889-4895.

68. Kang SM, Wohlrab F, Wells RD. Metal ions cause the isomerization of certain intramolecular triplexes. *J Biol Chem*. 1992;267(2):1259-1264.

69. Ussery DW, Sinden RR. Environmental-influences on the *in-vivo* level of intramolecular triplex DNA in Escherichia-Coli. *Biochemistry*. 1993;32(24):6206-6213.

70. Beal PA, Dervan PB. 2nd Structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science*. 1991;251(4999):1360-1363.

71. Thomas T, Thomas TJ. Selectivity of polyamines in triplex DNA stabilization. *Biochemistry*. 1993;32(50):14068-14074.

72. Hampel KJ, Burkholder GD, Lee JS. Plasmid dimerization mediated by triplex formation between polypyrimidine polypurine repeats. *Biochemistry*. 1993;32(4):1072-1077.

73. Pegg AE. Polyamine metabolism and its importance in neoplastic growth and as a target for chemotherapy. *Cancer Res*. 1988;48(4):759-774.

74. Ouameur AA, Tajmir-Riahi HA. Structural analysis of DNA interactions with biogenic polyamines and cobalt(III) hexamine studied by Fourier transform infrared and capillary electrophoresis. *J Biol Chem*. 2004;279(40):42041-42054.

75. Tabor H. The protective effect of spermine and other polyamines against heat denaturation of deoxyribonucleic acid. *Biochemistry*. 1962;1:496-501.

76. Asensio JL, Lane AN, Dhesi J, Bergqvist S, Brown T. The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J Mol Biol*. 1998;275(5):811-822.

77. Roberts RW, Crothers DM. Prediction of the stability of DNA triplexes. *Proc Natl Acad Sci U S A*. 1996;93(9):4320-4325.

78. Soyfer VN, Potaman VN. Triple-Helical Nucleic Acids. New York1996.

79. Asensio JL, Lane AN, Dhesi J, Bergqvist S, Brown T. The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J Mol Biol*. 1998;275(5):811-822.

80. Leitner D, Schroder W, Weisz K. Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry*. 2000;39(19):5886-5892.

81. Sinden RR. DNA structure and Function. Academic Press. 1994.

82. Frankkamenetskii MD, Mirkin SM. Triplex DNA Structures. *Annu Rev Biochem.* 1995;64:65-95.

83. Morgan AR, Wells RD. Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *J Mol Biol.* 1968;37(1):63-80.

84. Gowers DM, Fox KR. Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res*. 1999;27(7):1569-1577.

85. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. Structures of homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn*. 1986;3(4):667-669.

86. Mirkin SM, Lyamichev VI, Drushlyak KN, Dobrynin VN, Filippov SA, Frank-Kamenetskii MD. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature.* 1987;330(6147):495-497.

87. Rajeswari MR. DNA triplex structures in neurodegenerative disorder, Friedreich's ataxia. *J Biosci*. 2012;37(3):519-532.

88. Htun H, Dahlberg JE. Topology and formation of triple-stranded H-DNA. *Science*. 1989;243(4898):1571-1576.

89. Sinden RR. DNA structure and function. 1994.

90. Lipsett MN. Complex formation between polycytidylic acid and guanine oligonucleotides. *J Biol Chem.* 1964;239:1256-1260.

91. Lyamichev VI, Mirkin SM, Frank-Kamenetskii MD. A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn.* 1985;3(2):327-338.

92. Bernues J, Beltran R, Casasnovas JM, Azorin F. Structural polymorphism of homopurine--homopyrimidine sequences: the secondary DNA structure adopted by a $d(GA.CT)_{22}$ sequence in the presence of zinc ions. *EMBO J.* 1989;8(7):2087-2094.

93. Vojtiskova M, Mirkin S, Lyamichev V, Voloshin O, Frank-Kamenetskii M, Palecek E. Chemical probing of the homopurine.homopyrimidine tract in supercoiled DNA at single-nucleotide resolution. *FEBS Lett*. 1988;234(2):295-299.

94. Panayotatos N, Wells RD. Cruciform structures in supercoiled DNA. *Nature.* 1981;289(5797):466-470.

95. Murchie AI, Lilley DM. Supercoiled DNA and cruciform structures. *Methods Enzymol.* 1992;211:158-180.

96. Azorin F, Nordheim A, Rich A. Formation of Z-DNA in negatively supercoiled plasmids is sensitive to small changes in salt concentration within the physiological range. *EMBO J.* 1983;2(5):649-655.

97. Nordheim A, Peck LJ, Lafer EM, Stollar BD, Wang JC, Rich A. Supercoiling and left-handed Z-DNA. *Cold Spring Harb Symp Quant Biol*. 1983;47 Pt 1:93-100.

98. MIRKIN SM. DNA topology: fundamentals. ENCYCLOPEDIA OF LIFE SCIENCES 2001.

99. Vinograd J, Lebowitz J, Radloff R, Watson R, Laipis P. The twisted circular form of polyoma viral DNA. *Proc Natl Acad Sci* U S A. 1965;53(5):1104-1111.

100. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci* U S A. 1987;84(20):7024-7077.

101. Giaever GN, Wang JC. Supercoiling of intracellular DNA can occur in eukaryotic cells. *Cell.* 1988;55(5):849-856.

102. Nadal M. Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature.* 1986;321:256-258.

103. Bates AD, Maxwell A. DNA Topology. 2nd ed. New York: Oxford University Press Inc.; 2005.

104. Lyamichev VI, Mirkin SM, Frankkamenetskii MD. A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn.* 1985;3(2):327-338.

105. Michel D, Chatelain G, Herault Y, Brun G. The long repetitive polypurine polypyrimidine sequence $(TTCCC)_{48}$ forms DNA triplex with Pu-Pu-Py base triplets *in vivo*. *Nucleic Acids Res.* 1992;20(3):439-443.

106. Birnboim HC, Sederoff RR, Paterson MC. Distribution of polypyrimidine . polypurine segments in DNA from diverse organisms. *Eur J Biochem.* 1979;98(1):301-307.

107. Wong AK, Yee HA, van de Sande JH, Rattner JB. Distribution of CT-rich tracts is conserved in vertebrate chromosomes. *Chromosoma*. 1990;99(5):344-351.

108. Bucher P, Yagil G. Occurrence of oligopurine.oligopyrimidine tracts in eukaryotic and prokaryotic genes. *DNA Seq.* 1991;1(3):157-172.

109. Manor H, Rao BS, Martin RG. Abundance and degree of dispersion of genomic $d(GA)_n.d(TC)_n$ sequences. *J Mol Evol*. 1988;27(2):96-101.

110. Wells RD, Collier DA, Hanvey JC, Shimizu M, Wohlrab F. The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. *FASEB J*. 1988;2(14):2939-2949.

111. Boulikas T. Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human beta-globin gene complex. *J Cell Biochem*. 1993;52(1):23-36.

112. Agazie YM, Burkholder GD, Lee JS. Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. *Biochem J.* 1996;316 :461-466.

113. Hampel KJ, Ashley C, Lee JS. Kilobase-range communication between polypurine.polypyrimidine tracts in linear plasmids mediated by triplex formation: a braided knot between two linear duplexes. *Biochemistry*. 1994;33(19):5674-5681.

114. Burkholder GD, Latimer LJ, Lee JS. Immunofluorescent staining of mammalian nuclei and chromosomes with a monoclonal antibody to triplex DNA. *Chromosoma*. 1988;97(3):185-192.

115. Burkholder GD, Latimer LJ, Lee JS. Immunofluorescent localization of triplex DNA in polytene chromosomes of Chironomus and Drosophila. *Chromosoma*. 1991;101(1):11-18.

116. Agazie YM, Lee JS, Burkholder GD. Characterization of a new monoclonal antibody to triplex DNA and immunofluorescent staining of mammalian chromosomes. *J Biol Chem.* 1994;269(9):7019-7023.

117. Gorab E, Amabis JM, Stocker AJ, Drummond L, Stollar BD. Potential sites of triple-helical nucleic acid formation in chromosomes of Rhynchosciara (Diptera: Sciaridae) and Drosophila melanogaster. *Chromosome Res*. 2009;17(6):821-832.

118. Lee JS, Burkholder GD, Latimer LJ, Haug BL, Braun RP. A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. *Nucleic Acids Res*. 1987;15(3):1047-1061.

119. Ohno M, Fukagawa T, Lee JS, Ikemura T. Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma*. 2002;111(3):201-213.

120. Bagasra O, Stir AE, Pirisi-Creek L, Creek KE, Bagasra AU, Glenn N, *et al.* Role of micro-RNAs in regulation of lentiviral latency and persistence. *Appl Immunohistochem Mol Morphol*. 2006;14(3):276-290.

121. Kanak M, Alseiari M, Balasubramanian P, Addanki K, Aggarwal M, Noorali S, *et al.* Triplex-forming MicroRNAs form stable complexes with HIV-1 provirus and inhibit its replication. *Appl Immunohistochem Mol Morphol*. 2010;18(6):532-545.

122. Vojtiskova M, Palecek E. Unusual protonated structure in the homopurine.homopyrimidine tract of supercoiled and linearized plasmids recognized by chemical probes. *J Biomol Struct Dyn*. 1987;5(2):283-296.

123. Palecek E. Probing DNA structure with osmium tetroxide complexes *in vitro*. *Methods Enzymol.* 1992;212:139-155.

124. Karlovsky P, Pecinka P, Vojtiskova M, Makaturova E, Palecek E. Protonated triplex DNA in *E. coli* cells as detected by chemical probing. *FEBS Lett*. 1990;274(1-2):39-42.

125. Vetcher AA, Napierala M, Iyer RR, Chastain PD, Griffith JD, Wells RD. Sticky DNA, a long GAA.GAA.TTC triplex that is formed intramolecularly, in the sequence of intron 1 of the frataxin gene. *J Biol Chem*. 2002;277(42):39217-39227.

126. Vetcher AA, Napierala M, Wells RD. Sticky DNA: effect of the polypurine.polypyrimidine sequence. *J Biol Chem*. 2002;277(42):39228-39234.

127. Kang S, Wells RD. Central non-Pur.Pyr sequences in oligo(dG.dC) tracts and metal ions influence the formation of intramolecular DNA triplex isomers. *J Biol Chem.* 1992;267(29):20887-20891.

128. Ohshima K, Kang S, Larson JE, Wells RD. TTA.TAA triplet repeats in plasmids form a non-H bonded structure. *J Biol Chem*. 1996;271(28):16784-16791.

129. Faucon B, Mergny JL, Helene C. Effect of third strand composition on triple helix formation: Purine versus pyrimidine oligodeoxynucleotides. *Nucleic Acids Res*. 1996;24(16):3181-3188.

130. Marko JF. Short note on the scaling behavior of communication by 'slithering' on a supercoiled DNA. *Physica A*. 2001;296(1-2):289-292.

131. Marko JF, Siggia ED. Fluctuations and Supercoiling of DNA. *Science*. 1994;265(5171):506-508.

132. Jian H, Schlick T, Vologodskii A. Internal motion of supercoiled DNA: brownian dynamics simulations of site juxtaposition. *J Mol Biol*. 1998;284(2):287-296.

133. Huang J, Schlick T, Vologodskii A. Dynamics of site juxtaposition in supercoiled DNA. *Proc Natl Acad Sci U S A*. 2001;98(3):968-973.

134. Son LS, Bacolla A, Wells RD. Sticky DNA: *in vivo* formation in *E-coli* and *in vitro* association of long GAA.TTC tracts to generate two independent supercoiled domains. *J Mol Biol*. 2006;360(2):267-284.

135. Lamont PJ, Davis MB, Wood NW. Identification and sizing of the GAA trinucleotide repeat expansion of Friedreich's ataxia in 56 patients. Clinical and genetic correlates. *Brain*. 1997;120 ( Pt 4):673-680.

136. Sakamoto N, Larson JE, Iyer RR, Montermini L, Pandolfo M, Wells RD. GGA.TCC-interrupted triplets in long GAA.TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. *J Biol Chem*. 2001;276(29):27178-27187.

137. Vetcher AA, Wells RD. Sticky DNA formation *in vivo* alters the plasmid dimer/monomer ratio. *J Biol Chem*. 2004;279(8):6434-6443.

138. Pandolfo M. Friedreich ataxia: Detection of GAA repeat expansions and frataxin point mutations. *Methods Mol Med*. 2006;126:197-216.

139. Schulz JB, Pandolfo M. 150 years of Friedreich ataxia: from its discovery to therapy. *J Neurochem.* 2013;126 Suppl 1:1-3.

140. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, *et al*. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*. 1996;271(5254):1423-1427.

141. Harding AE. Friedreich's ataxia: a clinical and genetic study of 90 families with an analysis of early diagnostic criteria and intrafamilial clustering of clinical features. *Brain.* 1981;104(3):589-620.

142. Sandi C, Al-Mahdawi S, Pook MA. Epigenetics in Friedreich's Ataxia: Challenges and Opportunities for Therapy. Genet Res Int. 2013;2013:852080.

143. Chamberlain S, Shaw J, Rowland A, Wallis J, South S, Nakamura Y, *et al*. Mapping of mutation causing Friedreich's ataxia to human chromosome 9. *Nature*. 1988;334(6179):248-250.

144. Koutnikova H, Campuzano V, Foury F, Dolle P, Cazzalini O, Koenig M. Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. *Nat Genet*. 1997;16(4):345-351.

145. Campuzano V, Montermini L, Lutz Y, Cova L, Hindelang C, Jiralerspong S, *et al*. Frataxin is reduced in Friedreich ataxia patients and is associated with mitochondrial membranes. *Hum Mol Genet*. 1997;6(11):1771-1780.

146. Koeppen AH, Mazurkiewicz JE. Friedreich ataxia: neuropathology revised. *J Neuropathol Exp Neurol*. 2013;72(2):78-90.

147. Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL, *et al*. Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. *Proc Natl Acad Sci* U S A. 1997;94(14):7452-7457.

148. Gibson TJ, Koonin EV, Musco G, Pastore A, Bork P. Friedreich's ataxia protein: phylogenetic evidence for mitochondrial dysfunction. *Trends Neurosci*. 1996;19(11):465-468.

149. Stemmler TL, Lesuisse E, Pain D, Dancis A. Frataxin and mitochondrial FeS cluster biogenesis. *J Biol Chem*. 2010;285(35):26737-26743.

150. Pandolfo M, Pastore A. The pathogenesis of Friedreich ataxia and the structure and function of frataxin. *J Neurol*. 2009;256 Suppl 1:9-17.

151. Dhe-Paganon S, Shigeta R, Chi YI, Ristow M, Shoelson SE. Crystal structure of human frataxin. *J Biol Chem*. 2000;275(40):30753-30756.

152. Musco G, Stier G, Kolmerer B, Adinolfi S, Martin S, Frenkiel T, *et al*. Towards a structural understanding of Friedreich's ataxia: the solution structure of frataxin. *Structure*. 2000;8(7):695-707.

153. Britten RJ. DNA sequence insertion and evolutionary variation in gene regulation. *Proc Natl Acad Sci U S A*. 1996;93(18):9374-9377.

154. Bidichandani SI, Ashizawa T, Patel PI. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am J Hum Genet*. 1998;62(1):111-121.

155. Grabczyk E, Mancuso M, Sammarco MC. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. *Nucleic Acids Res.* 2007;35(16):5351-5359.

156. Parkinson MH, Boesch S, Nachbauer W, Mariotti C, Giunti P. Clinical features of Friedreich's ataxia: classical and atypical phenotypes. *J Neurochem.* 2013;126 Suppl 1:103-17.

157. Montermini L, Richter A, Morgan K, Justice CM, Julien D, Castellotti B, *et al.* Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. *Ann Neurol*. 1997;41(5):675-682.

158. Punga T, Buhler M. Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation. *EMBO Mol Med.* 2010;2(4):120-129.

159. Kumari D, Usdin K. Is Friedreich ataxia an epigenetic disorder? *Clin Epigenetics*. 2012;4(1):2.

160. Grabczyk E, Usdin K. The GAA.TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. *Nucleic Acids Res*. 2000;28(14):2815-2822.

161. Grabczyk E, Usdin K. Alleviating transcript insufficiency caused by Friedreich's ataxia triplet repeats. *Nucleic Acids Res*. 2000;28(24):4930-4937.

162. Baralle M, Pastor T, Bussani E, Pagani F. Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. *Am J Hum Genet*. 2008;83(1):77-88.

163. Greene E, Mahishi L, Entezam A, Kumari D, Usdin K. Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. *Nucleic Acids Res.* 2007;35(10):3383-3390.

164. Al-Mahdawi S, Pinto RM, Ismail O, Varshney D, Lymperi S, Sandi C, *et al*. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum Mol Genet*. 2008;17(5):735-746.

165. Herman D, Jenssen K, Burnett R, Soragni E, Perlman SL, Gottesfeld JM. Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. *Nat Chem Biol.* 2006;2(10):551-558.

166. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137(7):1194-1211.

167. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, *et al*. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*. 2011;479(7371):74-79.

168. De Biase I, Chutake YK, Rindler PM, Bidichandani SI. Epigenetic silencing in Friedreich ataxia is associated with depletion of CTCF (CCCTC-binding factor) and antisense transcription. PLoS One. 2009;4(11):7914.

169. Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol.* 2004;11(11):1068-1075.

170. Burnett R, Melander C, Puckett JW, Son LS, Wells RD, Dervan PB, *et al*. DNA sequence-specific polyamides alleviate transcription inhibition associated with long GAA.TTC repeats in Friedreich's ataxia. *Proc Natl Acad Sci* U S A. 2006;103(31):11497-11502.

171. Urbach AR, Dervan PB. Toward rules for 1:1 polyamide:DNA recognition. *Proc Natl Acad Sci* U S A. 2001;98(8):4343-8.

172. Janssen S, Durussel T, Laemmli UK. Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol Cell.* 2000;6(5):999-1011.

173. Abel T, Zukin RS. Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. *Curr Opin Pharmacol.* 2008;8(1):57-64.

174. Rai M, Soragni E, Jenssen K, Burnett R, Herman D, Coppola G, *et al*. HDAC inhibitors correct frataxin deficiency in a Friedreich ataxia mouse model. PLoS One. 2008;3(4):1958.

175. Barrio JR, Secrist JA, 3rd, Leonard NJ. Fluorescent adenosine and cytidine derivatives. *Biochem Biophys Res Commun*. 1972;46(2):597-604.

176. Yoshioka M, Nakamura A, Iizuka H, Nishidate K, Tamura Z, Miyazaki T. Sensitive fluorimetry of adenine, its nucleosides and nucleotides. *Nucleic Acids Symp Ser*. 1980(8):61-63.

177. Kohwi-Shigematsu T, Enomoto T, Yamada MA, Nakanishi M, Tsuboi M. Exposure of DNA bases induced by the interaction of DNA and calf thymus DNA helix-destabilizing protein. *Proc Natl Acad Sci* U S A. 1978;75(10):4689-4693.

178. Sattsangi PD, Leonard NJ, Frihart CR. 1,N2-ethenoguanine and N2,3-ethenoguanine. Synthesis and comparison of the electronic spectral properties of these linear and angular triheterocycles related to the Y bases. *J Org Chem*. 1977;42(20):3292-3296.

179. Kohwi-Shigematsu T, Kohwi Y. Poly(dG)-poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. *Cell*. 1985;43(1):199-206.

180. Hanvey JC, Shimizu M, Wells RD. Intramolecular DNA triplexes in supercoiled plasmids. *Proc Natl Acad Sci* U S A. 1988;85(17):6292-6296.

181. Ussery DW, Sinden RR. Environmental influences on the *in vivo* level of intramolecular triplex DNA in Escherichia coli. *Biochemistry*. 1993;32(24):6206-6213.

182. Boublikova P, Palecek E. Osmium tetroxide, N,N,N',N'-tetramethylethylendiamine. A new probe of DNA structure in the cell. *FEBS Lett*. 1990;263(2):281-224.

183. Hanvey JC, Klysik J, Wells RD. Influence of DNA sequence on the formation of non-B right-handed helices in oligopurine.oligopyrimidine inserts in plasmids. *J Biol Chem*. 1988;263(15):7386-7396.

184. Johnston BH. The S1-sensitive form of $d(C.T)_n.d(A.G)_n$: chemical evidence for a three-stranded structure in plasmids. *Science*. 1988;241(4874):1800-1804.

185. Malkov VA, Voloshin ON, Soyfer VN, Frank-Kamenetskii MD. Cation and sequence effects on stability of intermolecular pyrimidine-purine-purine triplex. *Nucleic Acids Res*. 1993;21(3):585-591.

186. Malkov VA, Voloshin ON, Veselkov AG, Rostapshov VM, Jansen I, Soyfer VN, *et al*. Protonated pyrimidine-purine-purine triplex. *Nucleic Acids Res*. 1993;21(1):105-111.

187. Courey AJ, Wang JC. Influence of DNA sequence and supercoiling on the process of cruciform formation. *J Mol Biol.* 1988;202(1):35-43.

188. Jayasena SD, Johnston BH. Intramolecular triple-helix formation at (PunPyn).(PunPyn) tracts: recognition of alternate strands via Pu.PuPy and Py.PuPy base triplets. *Biochemistry*. 1992;31(2):320-327.

189. Shimizu M, Hanvey JC, Wells RD. Intramolecular DNA triplexes in supercoiled plasmids. I. Effect of loop size on formation and stability. *J Biol Chem*. 1989;264(10):5944-5949.

190. McCarthy JG, Williams LD, Rich A. Chemical reactivity of potassium permanganate and diethyl pyrocarbonate with B DNA: specific reactivity with short A-tracts. *Biochemistry*. 1990;29(25):6071-6081.

191. Mitas M, Yu A, Dill J, Kamp TJ, Chambers EJ, Haworth IS. Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$. *Nucleic Acids Res*. 1995;23(6):1050-1059.

192. Bui CT, Rees K, Cotton RG. Permanganate oxidation reactions of DNA: perspective in biological studies. *Nucleosides Nucleotides Nucleic Acids*. 2003;22(9):1835-1855.

193. Glover JN, Farah CS, Pulleyblank DE. Structural characterization of separated H-DNA conformers. *Biochemistry*. 1990;29(50):11110-11115.

194. Belotserkovskii BP, Krasilnikova MM, Veselkov AG, Frank-Kamenetskii MD. Kinetic trapping of H-DNA by oligonucleotide binding. *Nucleic Acids Res*. 1992;20(8):1903-1908.

195. Panyutin IG, Wells RD. Nodule DNA in the $(GA)_{37}.(CT)_{37}$ insert in superhelical plasmids. *J Biol Chem*. 1992;267(8):5495-5501.

196. Desai NA, Shankar V. Single-strand-specific nucleases. *FEMS Microbiol Rev*. 2003;26(5):457-491.

197. Nelson KL, Becker NA, Pahwa GS, Hollingsworth MA, Maher LJ, 3rd. Potential for H-DNA in the human MUC1 mucin gene promoter. *J Biol Chem*. 1996;271(30):18061-18067.

198. Pestov DG, Dayn A, Siyanova E, George DL, Mirkin SM. H-DNA and Z-DNA in the mouse c-Ki-ras promoter. *Nucleic Acids Res*. 1991;19(23):6527-6532.

199. Asakura Y, Kikuchi Y, Yanagida M. A cruciform in the direct repeats of the yeast 2 micron DNA: Selective S1 nuclease cleavage at one of the three homologous palindromes. J *Biochem.* 1985;98(1):41-47.

200. Lilley DM, Markham AF. Dynamics of cruciform extrusion in supercoiled DNA: use of a synthetic inverted repeat to study conformational populations. *EMBO J.* 1983;2(4):527-533.

201. Kumar P, Verma A, Maiti S, Gargallo R, Chowdhury S. Tetraplex DNA transitions within the human c-myc promoter detected by multivariate curve resolution of fluorescence resonance energy transfer. *Biochemistry.* 2005;44(50):16426-16434.

202. Blaho JA, Larson JE, McLean MJ, Wells RD. Multiple DNA secondary structures in perfect inverted repeat inserts in plasmids. Right-handed B-DNA, cruciforms, and left-handed Z-DNA. *J Biol Chem.* 1988;263(28):14446-14455.

203. McLean MJ, Lee JW, Wells RD. Characteristics of Z-DNA helices formed by imperfect (purine-pyrimidine) sequences in plasmids. *J Biol Chem*. 1988;263(15):7378-7385.

204. Romier C, Dominguez R, Lahm A, Dahl O, Suck D. Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. Proteins. 1998;32(4):414-424.

205. Christophe D, Cabrer B, Bacolla A, Targovnik H, Pohl V, Vassart G. An unusually long poly(purine).poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic Acids Res*. 1985;13(14):5127-5144.

206. Fowler RF, Skinner DM. Eukaryotic DNA diverges at a long and complex pyrimidine.purine tract that can adopt altered conformations. *J Biol Chem*. 1986;261(19):8994-9001.

207. Beltran R, Martinez-Balbas A, Bernues J, Bowater R, Azorin F. Characterization of the zinc-induced structural transition to *H-DNA at a $d(GA.CT)_{22}$ sequence. *J Mol Biol*. 1993;230(3):966-978.

208. Stadler J, Lemmens R, Nyhammar T. Plasmid DNA purification. *J Gene Med*. 2004;6:S54-S66.

209. McCarthy JG, Rich A. Detection of an unusual distortion in A-tract DNA using $KMnO_4$: effect of temperature and distamycin on the altered conformation. *Nucleic Acids Res.* 1991;19(12):3421-3429.

210. Matyasek R, Fulnecek J, Fajkus J, Bezdek M. Evidence for a sequence-directed conformation periodicity in the genomic highly repetitive DNA detectable with single-strand-specific chemical probe potassium permanganate. *Chromosome Res.* 1996;4(5):340-349.

211. Kahl BF, Paule MR. The use of diethyl pyrocarbonate and potassium permanganate as probes for strand separation and structural distortions in DNA. *Methods Mol Biol*. 2001;148:63-75.

212. Herr W. Diethyl pyrocarbonate: a chemical probe for secondary structure in negatively supercoiled DNA. *Proc Natl Acad Sci* U S A. 1985;82(23):8009-8013.

213. Alam J, Cook JL. Reporter genes: application to the study of mammalian gene transcription. *Anal Biochem.* 1990;188(2):245-254.

214. Sherf B NS, Hannah R and Wood K. Dual-Luciferase ™ Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and *Renilla* Luciferase Assays. *Promega Notes Magazine* 1996:02-08.

215. Brown TA. Gene cloning and DNA analysis: an introduction. 6th. 2010. 79-80.

216. Wang JC. DNA topoisomerases. *Annu Rev Biochem*. 1996;65:635-92.

217. Corchero JL, Villaverde A. Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol Bioeng*. 1998;58(6):625-632.

218. Garner MM, Chrambach A. Resolution of circular, nicked circular and linear DNA, 4.4 kb in length, by electrophoresis in polyacrylamide solutions. *Electrophoresis*. 1992;13(3):176-178.

219. Kim E, Napierala M, Dent SY. Hyperexpansion of GAA repeats affects post-initiation steps of FXN transcription in Friedreich's ataxia. *Nucleic Acids Res*. 2011;39(19):8366-8377.

220. Napierala M, Dere R, Vetcher A, Wells RD. Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *J Biol Chem*. 2004;279(8):6444-6454.

221. Ohshima K, Montermini L, Wells RD, Pandolfo M. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. J Biol Chem. 1998;273(23):14588-14595.

222. Chandok GS, Patel MP, Mirkin SM, Krasilnikova MM. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. *Nucleic Acids Res*. 2012;40(9):3964-3974.

223. Dieffenbach CW, Lowe TM, Dveksler GS. General concepts for PCR primer design. *PCR Methods Appl.* 1993;3(3):S30-7.

224. Mirkin SM. Expandable DNA repeats and human disease. *Nature*. 2007;447(7147):932-940.

225. O'Neill MA, Barton JK. 2-Aminopurine: a probe of structural dynamics and charge transfer in DNA and DNA:RNA hybrids. *J Am Chem Soc*. 2002;124(44):13053-13066.

226. Jean JM, Hall KB. 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proc Natl Acad Sci* U S A. 2001;98(1):37-41.

227. Lee BJ, Barch M, Castner EW, Jr., Volker J, Breslauer KJ. Structure and dynamics in DNA looped domains: CAG triplet repeat sequence dynamics probed by 2-aminopurine fluorescence. *Biochemistry*. 2007;46(38):10756-10766.

228. Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M, Bidnenko V. Rescue of arrested replication forks by homologous recombination. *Proc Natl Acad Sci* U S A. 2001;98(15):8181-8188.

229. Haber JE. DNA recombination: the replication connection. *Trends Biochem Sci*. 1999;24(7):271-275.

230. Tropp BE. Molecular Biology Genes to Proteines. 3rd ed: Jones and Bartlett; 1987.

231. Wells RD. Unusual DNA structures. *J Biol Chem*. 1988;263(3):1095-1098.

232. Antao VP, Gray DM, Ratliff RL. CD of six different conformational rearrangements of poly[d(A.G).d(C.T)] induced by low pH. *Nucleic Acids Res*. 1988;16(2):719-738.

233. Parniewski P, Galazka G, Wilk A, Klysik J. Complex structural behavior of oligopurine-oligopyrimidine sequence cloned within the supercoiled plasmid. *Nucleic Acids Res*. 1989;17(2):617-629.

234. Schlotterer C, Tautz D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 1992;20(2):211-215.

235. Heidenfelder BL, Makhov AM, Topal MD. Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J Biol Chem*. 2003;278(4):2425-2431.

236. Jain A, Rajeswari MR, Ahmed F. Formation and thermodynamic stability of intermolecular (R.R.Y) DNA triplex in GAA.TTC repeats associated with Freidreich's ataxia. *J Biomol Struct Dyn.* 2002;19(4):691-699.

237. Lilley DM. Probes of DNA structure. *Methods Enzymol.* 1992;212:133-139.

238. Leonard NJ, McDonald JJ, Reichmann ME. Reaction of diethyl pyrocarbonate with nucleic acid components. I. Adenine. *Proc Natl Acad Sci* U S A. 1970;67(1):93-8.

239. Iida S, Hayatsu H. The permanganate oxidation of thymidine and thymidylic acid. *Biochim Biophys Acta.* 1971;228(1):1-8.

240. Kohwi-Shigematsu T, Kohwi Y. Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *Methods Enzymol.* 1992;212:155-180.

241. Wang G, Zhao J, Vasquez KM. Methods to determine DNA structural alterations and genetic instability. *Methods.* 2009;48(1):54-62.

242. Rubin CM, Schmid CW. Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res.* 1980;8(20):4613-4619.

243. Yagil G. Paranemic structures of DNA and their role in DNA unwinding. *Crit Rev Biochem Mol Biol.* 1991;26(5-6):475-559.

244. Lilley DM. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci U S A.* 1980;77(11):6468-6472.

245. Wiegand RC, Godson GN, Radding CM. Specificity of the S1 nuclease from *Aspergillus oryzae. J Biol Chem.* 1975;250(22):8848-8855.

246. Schon E, Evans T, Welsh J, Efstratiadis A. Conformation of promoter DNA: fine mapping of S1-hypersensitive sites. *Cell*. 1983;35(3 Pt 2):837-48.

247. Bolivar F, Rodriguez RL, Betlach MC, Boyer HW. Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9. *Gene.* 1977;2(2):75-93.

248. Panayotatos N, Fontaine A. A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. *J Biol Chem*. 1987;262(23):11364-11368.

249. Sverdlov ED, Voloshin ON, Kishko Ia G, Monastyrskaia GS. $Os_4O_4$ reacts primarily with -10 and -35 sites of Escherichia coli lacUV5-promoter. *Dokl Akad Nauk SSSR*. 1987;297(1):225-8.

250. Pulleyblank DE, Haniford DB, Morgan AR. A structural basis for S1 nuclease sensitivity of double-stranded DNA. *Cell*. 1985;42(1):271-280.

251. Htun H, Dahlberg JE. Single strands, triple strands, and kinks in H-DNA. *Science.* 1988;241(4874):1791-1796.

252. Collier DA, Griffin JA, Wells RD. Non-B right-handed DNA conformations of homopurine.homopyrimidine sequences in the murine immunoglobulin C alpha switch region. *J Biol Chem.* 1988;263(15):7397-7405.

253. Evans T, Efstratiadis A. Sequence-dependent S1 nuclease hypersensitivity of a heteronomous DNA duplex. *J Biol Chem*. 1986;261(31):14771-14780.

254. Shimizu M, Hanvey JC, Wells RD. Multiple non-B-DNA conformations of polypurine.polypyrimidine sequences in plasmids. *Biochemistry*. 1990;29(19):4704-4713.

255. Collier DA, Wells RD. Effect of length, supercoiling, and pH on intramolecular triplex formation. Multiple conformers at pur.pyr mirror repeats. *J Biol Chem*. 1990;265(18):10652-10658.

256. Kato M, Shimizu N. Effect of the potential triplex DNA region on the *in vitro* expression of bacterial beta-lactamase gene in superhelical recombinant plasmids. *J Biochem.* 1992;112(4):492-494.

257. Beard P, Morrow JF, Berg P. Cleavage of circular, superhelical simian virus 40 DNA to a linear duplex by S1 nuclease. *J Virol.* 1973;12(6):1303-1313.

258. Germond JE, Vogt VM, Hirt B. Characterization of the single-strand-specific nuclease S1 activity on double-stranded supercoiled polyoma DNA. *Eur J Biochem.* 1974;43(3):591-600.

259. Rougee M, Faucon B, Mergny JL, Barcelo F, Giovannangeli C, Garestier T, *et al.* Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches. *Biochemistry*. 1992;31(38):9269-9278.

260. Kohwi-Shigematsu T, Kohwi Y. Detection of triple-helix related structures adopted by poly(dG).poly(dC) sequences in supercoiled plasmid DNA. *Nucleic Acids Res.* 1991;19(15):4267-4271.

261. Klysik J, Rippe K, Jovin TM. Parallel-stranded DNA under topological stress: rearrangement of $(dA)_{15}.(dT)_{15}$ to a $d(A.A.T)_n$ triplex. *Nucleic Acids Res.* 1991;19(25):7145-7154.

262. LeProust EM, Pearson CE, Sinden RR, Gao X. Unexpected formation of parallel duplex in GAA and TTC trinucleotide repeats of Friedreich's ataxia. *J Mol Biol*. 2000;302(5):1063-1080.

263. Bidichandani SI, Ashizawa T, Patel PI. Atypical Friedreich ataxia caused by compound heterozygosity for a novel missense mutation and the GAA triplet-repeat expansion. *Am J Hum Genet.* 1997;60(5):1251-1256.

264. Rosenberg RN. DNA-triplet repeats and neurologic disease. *N Engl J Med*. 1996;335(16):1222-1224.

265. Praseuth D, Guieysse AL, Helene C. Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim Biophys Acta*. 1999;1489(1):181-206.

266. Sarkar PS, Brahmachari SK. Intramolecular triplex potential sequence within a gene down regulates its expression *in vivo*. *Nucleic Acids Res*. 1992;20(21):5713-5718.

267. Brahmachari SK, Sarkar PS, Raghavan S, Narayan M, Maiti AK. Polypurine/polypyrimidine sequences as cis-acting transcriptional regulators. *Gene*. 1997;190(1):17-26.

268. Eckert KA, Mowery A, Hile SE. Misalignment-mediated DNA polymerase beta mutations: comparison of microsatellite and frame-shift error rates using a forward mutation assay. *Biochemistry*. 2002;41(33):10490-10498.

269. Firulli AB, Maibenco DC, Kinniburgh AJ. Triplex forming ability of a *c-myc* promoter element predicts promoter strength. *Arch Biochem Biophys*. 1994;310(1):236-242.

270. Rustighi A, Tessari MA, Vascotto F, Sgarra R, Giancotti V, Manfioletti G. A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. *Biochemistry*. 2002;41(4):1229-1240.

271. Xu G, Goodridge AG. Characterization of a polypyrimidine.polypurine tract in the promoter of the gene for chicken malic enzyme. *J Biol Chem*. 1996;271(27):16008-16019.

272. Michel D, Chatelain G, Herault Y, Harper F, Brun G. H-DNA can act as a transcriptional insulator. *Cell Mol Biol Res*. 1993;39(2):131-140.

273. Maiti AK, Brahmachari SK. Poly purine.pyrimidine sequences upstream of the beta-galactosidase gene affect gene expression in Saccharomyces cerevisiae. *BMC Mol Biol*. 2001;2:11.

274. Lu Q, Teare JM, Granok H, Swede MJ, Xu J, Elgin SC. The capacity to form H-DNA cannot substitute for GAGA factor binding to a $(CT)_n.(GA)_n$ regulatory site. *Nucleic Acids Res*. 2003;31(10):2483-2494.

275. Pahwa GS, Maher LJ, 3rd, Hollingsworth MA. A potential H-DNA element in the MUC1 promoter does not influence transcription. *J Biol Chem*. 1996;271(43):26543-26546.

276. Wang JC. DNA topoisomerases. *Annu Rev Biochem*. 1985;54:665-697.

277. Smith GR. DNA supercoiling: another level for regulating gene expression. *Cell*. 1981;24(3):599-600.

278. Weintraub H. Assembly and propagation of repressed and depressed chromosomal states. *Cell*. 1985;42(3):705-711.

279. Weintraub H, Cheng PF, Conrad K. Expression of transfected DNA depends on DNA topology. *Cell*. 1986;46(1):115-122.

280. Cummings CJ, Zoghbi HY. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet*. 2000;9(6):909-916.

281. Wells RD, Parniewski P, Pluciennik A, Bacolla A, Gellibolian R, Jaworski A. Small slipped register genetic instabilities in Escherichia coli in triplet repeat sequences associated with hereditary neurological diseases. *J Biol Chem*. 1998;273(31):19532-19541.

282. Mariappan SV, Catasti P, Silks LA, 3rd, Bradbury EM, Gupta G. The high-resolution structure of the triplex formed by the GAA.TTC triplet repeat associated with Friedreich's ataxia. *J Mol Biol*. 1999;285(5):2035-2052.

283. Sakamoto N, Ohshima K, Montermini L, Pandolfo M, Wells RD. Sticky DNA, a self-associated complex formed at long GAA.TTC repeats in intron 1 of the frataxin gene, inhibits transcription. *J Biol Chem*. 2001;276(29):27171-27177.

284. Napierala M, Parniewski P, Pluciennik A, Wells RD. Long CTG.CAG repeat sequences markedly stimulate intramolecular recombination. *J Biol Chem*. 2002;277(37):34087-34100.

285. Van de Sande JH, Ramsing NB, Germann MW, Elhorst W, Kalisch BW, von Kitzing E, *et al*. Parallel stranded DNA. *Science*. 1988;241(4865):551-557.

286. Gacy AM, Goellner GM, Spiro C, Chen X, Gupta G, Bradbury EM, *et al*. GAA instability in Friedreich's Ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol Cell*. 1998;1(4):583-593.