Intrinsic DNA abnormalities and their effect on plasmid processing

James Ross Cooke BEng MSc

Submitted in accordance with the requirements for the
degree of Doctor of Philosophy

University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Not all DNA has the familiar Watson-Crick, right-handed double helical structure; Left-handed helixes, triplexes, quadruplexes, bent DNA and regions of increased stiffness have all been encountered in nature. Some of these DNA structures are formed as a result of protein binding, while others are intrinsic. Such intrinsic structures may be incorporated into future plasmid constructs for gene therapy and DNA vaccine products.

Intrinsic DNA structures were included at a defined point in a 2.9 kb plasmid, and their effects on cell growth rate, total plasmid yield, and topology (i.e. the relative proportions of supercoiled plasmid, open circular and linear forms), were determined. The stability of the inserted sequences was assessed using gel electrophoresis.

Results suggest that Z-DNA is unstable in a batch *Escherichia coli* DH1 production system grown in complex medium. Encouragingly other sequences studied (triplex, bend and quadruplex) did not cause spontaneous deletions, and no detrimental effect was found on growth rate or on total plasmid yield; indicating that such sequences could be included in future DNA products without any detrimental effect on plasmid yields. Although the intramolecular triplex studied significantly decreased the proportion of supercoiled species.

The effect of different topological forms on transcription of a DNA vaccine in a cell free system was investigated, in order to determine the implications of this on specifications of plasmid products. Open circular plasmid was found to be expressed around 3.5 fold more than the supercoiled form. Current guidelines suggest that plasmid products should contain a minimum of 90% supercoiled species. The supercoiled form is not a single isoform, but contains plasmids with different numbers of supercoils (linking number). Measurement of the levels of supercoiled species in comparison to the other plasmid forms allows monitoring of the production process, however it is recommended that the guidelines be set on individual product basis i.e. drug efficacy will depend on delivery method, longevity and transcription levels.

Contents

Table of contents

Acknowledgements

I gratefully acknowledge the support of GlaxoSmithKline, the Biotechnology and Biological Sciences Research Council and the Centre for Scientific Enterprise London.

I would like to thank all those at GSK Beckenham who made me so welcome during my extended stay there. In no particular order: my industrial supervisor, Liz McKie especially for her encouragement when the cloning wacky structures just wasn't happening and her guidance for the first two years, and reading through the thesis (hope its not quite so boring as its about 2/3rds the size now!); Ronan O'Kennedy for his invaluable advice, discussions and encouragement; Sally Hay(es) and Paul Morrison for teaching me how to be a fermentalist; Pete Steward for fixing fermenters when they needed it; Mark Uden for his cynical and amusing outlook on life (and advice on gel based assays); Sarah Howard, Kathy Griffin, Kate Faulkes, Welcome Ndlouv and especially Val Hall for guidance in the lab and making it a good place to be; Chimmy Salgado for his outstanding knowledge and assistance with all aspects of PCR; Craig Little for saving my thumbs from RSI damage by automating the picogreen assay; Graham Allen for his hard work as lab technician; Drs Anna Hills and John O'Hara (congratulations!) for letting me stay over; all the security staff – sorry for the all the false alarms; and finally the football massive for putting up with my legendary step overs and point blank misses!

Thanks also to those at UCL who assisted, my supervisor and advisor, Eli Keshavaraz-Moore who hopefully will have read this, and John Ward who thought up the project to start with and has read it all; Dunstan Cooke and Haitham Hussain, for their cloning tips during the first six months of the project.

Thanks to my project team at LBS, Paul, Helen and Tamara...the prize gave me lots of extra funded writing up time, which came in handy!

The University of London Sub Aqua Club BSAC69 has been a big part of my life since I arrived in London and thanks must go to all my friends there. Special thanks to all those who let me stay over after I moved north but still needed to get to Beckenham. Having been treasurer of the club in the second year of my studies I didn't want to have to do it again, so no thanks at all to the 2003-2004 treasurer who went to Japan and left the books with me!

Derek Fry, the best physics teacher in the world and my friend.

My grandfather, for his huge experience and thorough advice on thesis writing.

I would like to thank my very special lady friend Catherine, who one day will be doctor of all things moss like, and has loved and supported me throughout. Sorry I got annoyed trying to explain ΔLk - I love you sweetie!

And finally, my mum and dad, for their love, support and occasional inspiration (Barnes *et al* 1965). I am finally going to be a doctor, just a bit different to what you once expected! I love you both very much.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less travelled by,
And that has made all the difference.

(Frost R 1920)

It's not a sad poem I think...

August 2004

Chapter 1:

DNA structure and plasmids

1 DNA structure and plasmids

This chapter includes a description of both basic DNA structure and some of the intrinsic abnormalities that occur in nature. It also discusses plasmid properties, the importance of plasmids as biotherapeutics, industrial plasmid manufacture and processing.

1.1 Basic DNA structure

In the fifty years since basic DNA structure was first discovered and its implications understood, great advances in biotechnology have been made. The nature of DNA now lies at the heart of molecular biology.

1.1.1 Discovery of DNA structure

The discovery and understanding of basic DNA structure can be regarded as one of mankind's greatest achievements. Full historical accounts are available from several sources and for interest a brief summary of certain individual's contributions is given below.

Examination of isolated DNA using X-ray diffraction began in 1943 by Astbury and while inconclusive it was soon postulated that DNA molecules had a 'preferred orientation'. In 1950 Wilkins proved this to be true by obtaining diffraction patterns that showed a crystalline structure. For this to occur DNA molecules must have very similar structures that allow tessellation (fitting together very closely with few gaps). Three years later Franklin obtained diffraction patterns of single DNA fibres and these contained dominant cross-like patterns - characteristic of a double helix. Analysis of Franklin's photographs made available to Watson and Crick by Franklin's colleague Wilkins, and use of Chargaff's work enabled them to elucidate an energetically favourable DNA structure. The model of DNA structure that Watson and Crick proposed in 1953 was hence shown to be "essentially correct".

1.1.2 Classic Watson Crick Model

"Although some refinement might be made, I feel that it is very likely that the Watson-Crick structure is essentially correct." Pauling 1953.

DNA is composed of subunits denoted nucleotides. The nucleotides consist of a deoxyribose sugar, phosphate groups and a base. There are four different bases, adenine (A), cytosine (C), guanine (G) and thymine (T). A and G are purines and C and T are pyrimidines.

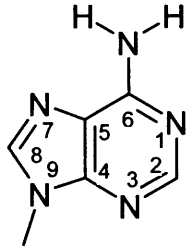
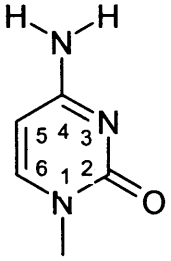
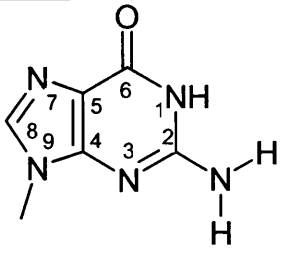
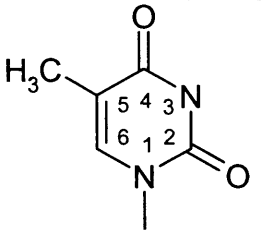
Adenine	Cytosine	Guanine	Thymine
			

Figure 1.1 The four bases of DNA

The "structure has two helical chains each coiled round the same axis...Both chains follow right handed helices...the two chains run in opposite directions. ..The bases are on the inside of the helix and the phosphates on the outside... the two chains are held together by the purine and pyrimidine bases, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side...One of the pair must be a purine and the other a pyrimidine for bonding to occur. ...Only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine)" (Watson and Crick 1953).

As can be seen below the Watson Crick base pairs produce near indistinguishable base geometeries, and it is these that allow such a regular structure to be formed.

A-T G-C

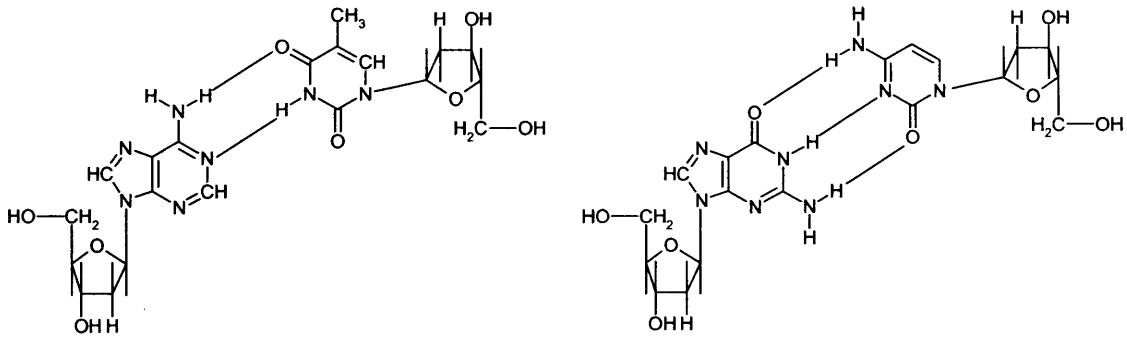


Figure 1.2 The Watson-Crick base pairs

This specificity is due to steric and hydrogen binding factors. There are two and three hydrogen bonds in between base pairs A-T and G-C respectively. This

together with stacking energy considerations accounts for the lower 'melting' temperatures of A-T rich sequences. 'Melting' of DNA is a term used to describe double stranded DNA becoming single stranded.

The major features of the Watson-Crick double helix are:

- The diameter of the helix is 20Å,
- adjacent bases are separated by 3.4 Å along the helix and are related by a rotation of 36 °. Hence the helical structure repeats after ten residues of the chain, i.e. at intervals of 34 Å.

1.1.3 The three main forms of DNA

There are three main conformations of DNA. These are summarised in table 1.1. Fibre and solution methods and more recently single-crystal x-ray analyses have confirmed the properties (Dickerson *et al* 1982).

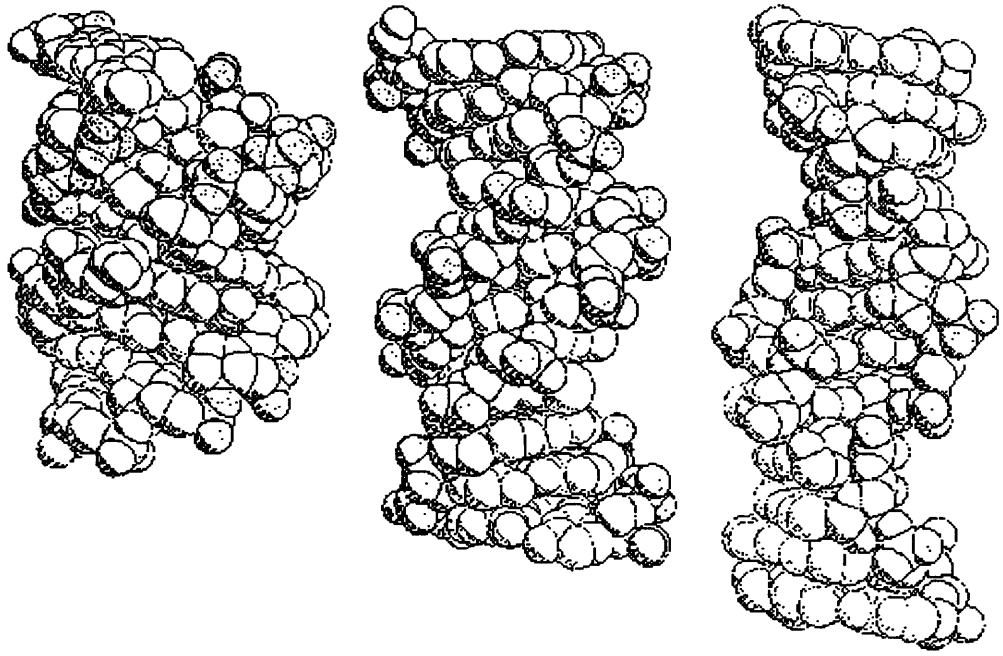


Figure 1.3 Space filling models for comparison of the three main types of DNA - 12 bp each. Most DNA lies in the range between the A- and B- forms. A-DNA (left) has around 11bp per helical turn and is common in high salt concentrations, B-DNA (middle) is the most common form *in vivo* and has 10.5 bp per helical turn, and Z-DNA (right) which has 12 bp per helical turn and has a left-handed sense in contrast the other main (right-handed) forms, formation of Z-DNA is favoured by alternating purine pyrimidine sequences. Reprinted with permission from Dickerson et al., SCIENCE 216:475-85 (1982) . Copyright AAAS.

1.1.3.1 A – and B – DNA and transitions between the two

B – DNA is the most common form of DNA found in vivo. B – DNA has a right-handed sense and has a wide and a narrow grove. Bases are found to be perpendicular to the helical axis. The helical structure repeats after ten and a half residues of the chain.

A – DNA also has a right-handed sense, the major groove is found to be deep and the minor grove is shallow.

Shifts between the A – and B – forms of DNA are possible because of changes of conformation of the ribose ring according to the solution conditions:

B-DNA

In low salt conditions DNA can be considered hydrated, and the most stable form of the ribose ring is in the C2' *endo* pucker. This means that the distance between phosphate groups is around 7Å.

A-DNA

In high salt conditions, the reduced number of water molecules means that the ribose ring has a C3' *endo* conformation, hence a more compact DNA structure is produced when the phosphate groups are closer together than in B-DNA, being 5.9Å apart.

The A – conformation is the most common form of double stranded RNA due to the oxygen atom on the 2' position of the ribose ring stabilizing the C3' *endo* conformation.

1.1.3.2 Z – DNA

As can be seen from figure 1.3, the main difference between Z and the other two major forms of DNA (A and B) are that Z – DNA has a left-handed sense. In addition it is more elongated than the other two types. Formation of Z – DNA is favoured by conditions of high salt and in sequences containing alternating purine-pyrimidine tracts of 12 to 15 base pairs (Hoheisel and Pohl 1987). Cytosine methylation has also been found to enhance formation of Z – DNA in plasmids in *Escherichia coli* (*E. coli)* (Zacharias *et al* 1990).

Repeated sequences of CG have been shown to form Z-DNA without supercoiling, while other repeats, such as TG, do require a threshold level of super helical density (Albert *at al* 1994). There has been some investigation into the minimum number of purine.pyrimidine repeats that are required in linear DNA for Z-DNA to be formed in the middle a long B-DNA form. It has been shown that the minimum number of

repeats of CG is 12, of which the four central base pairs form Z-DNA and the two flanking four base pair repeats exist in the B-form (Dai *et al* 1992). The transitions between the Z- and B- are interesting as the DNA must convert from the right handed form to the left handed form and back again. It is thought that the swap in handedness between the two forms is very rapid, such that the double helix flips from one form to the other without any base pairs at the junction (Dai *et al* 1992).

Purine pyrimidine tracts are over represented, occurring up to 40 fold times more frequently than they would be expected in some organisms (Ussery *et al* 2002). Z-forming sequences are commonly found near the 5' end of genes and in promoter regions (Schroth *et al* 1992). As described in section 1.2.3.1, these regions are subject to changes in super helical density that can provide the energy used to convert B-DNA to Z-DNA.

The form that DNA takes is highly sequence dependent and several papers have presented their findings of the most stable forms produced with different sequences, for example three base pair sequences (Packer *et al* 2000) and more recently eight base pair sequences (Gardiner *et al* 2003).

1.1.4 C –, D – , 'E - ', H –, P – and T – DNA

A, B and Z are far from the only prefixes that can be found in scientific literature describing different DNA forms. For completeness several of the other forms are described in Appendix 2. Only the H form is of interest to the specific nature of this project and is further described in a later section.

A – DNA	B – DNA	Z – DNA
• Right-handed sense. • 11bp/helical turn • 26 Å diameter • 2.6 Å rise/residue • 33° helical twist • Sugar pucker C3' endo • Shorter, wider helix than B. • Deep, narrow major groove not easily accessible to proteins. • Wide, shallow minor groove accessible to proteins. • Favoured conformation at low water concentrations. • Base pairs tilted to helix axis and displaced from axis. • Generally tilt is 20°	• Right-handed sense. • 10.5 bp/helical turn • 20 Å diameter • 3.4 Å rise/residue • 36° helical twist • Sugar pucker C2' endo • Most common DNA conformation *in vivo*. • Narrower, more elongated helix than A. • Wide major groove easily accessible to proteins. • Narrow minor groove. • Favoured conformation at high water concentrations (hydration of minor groove seems to favour B-form). • Base pairs nearly perpendicular to helix axis. • Tilt is ~6°	• Left-handed sense. • 12bp/helical turn • 18 Å diameter • 3.7 Å rise/residue • Sugar pucker alternates C3' endo (G), C2' endo (C) • Can be formed in vivo, given proper sequence and superhelical tension, but function remains obscure. • Narrower, more elongated helix than both A and B. • Major "groove" is not really groove. • Narrow minor groove. • Conformation favoured by high salt concentrations. • Conformation favoured by alternating purine-pyrimidine sequence. • Base pairs nearly perpendicular to helix axis. • Zigzag backbone due to C sugar conformation compensating for G glycosidic bond conformation

Table 1.1 Properties of the three main DNA forms.

1.2 Deviations from traditional DNA model

1.2.1 Sequences inducing change when bound to proteins

1.2.1.1 Eukaryotic

For the purposes of this project, sequences that induce change via eukaryotic protein interactions are of little interest. Since the plasmid DNA is grown in prokaryotic organisms, proteins producing such changes (e.g. histone proteins) will not be present.

DNA wraps around histone octamers forming nucleosomes. DNA in nucleosome form takes up around six times less space, allowing the genome of a cell to fit inside the compartment in which it is contained

In future it is envisaged that eukaryotic cells may be used to produce plasmid DNA for use in gene therapy. This is because methylation of DNA differs between eukaryotes and prokaryotes. If plasmid products are ever produced using eukaryotic production systems then investigation into possible effects of eukaryotic proteins may be required.

1.2.1.2 Prokaryotic

A variety of prokaryotic proteins also bind to DNA with several different effects. One of the most dramatic examples is the Catabolite Activator Protein (CAP) that bends DNA by 90° (Schultz *et al* 1991).

The HU protein in prokaryotes seems to have similar function to histone proteins (Rouviere-Yaniv and Gros 1975). The HU protein has been shown to bend DNA quite dramatically. In the presence of the HU protein 80 bp sequences can be circularised, whereas without the HU protein no ligation occurs (Hodges-Garcia *et al* 1989).

Another interesting example are sites recognised by DNA gyrase (Topoisomerase I). DNA gyrase is an enzyme that adds supercoils to DNA in an ATP dependent manner by passing one DNA segment through another by means of a reversible break at specific sites (Morrison and Cozzarelli 1981).

If it was seen as desirable, by altering the number of these sites in a plasmid, it might be possible to increase the level of supercoiling. An alternative strategy to increase supercoiling levels, might be to over-express the gyrase gene. Such a strain is described by the Westerhoff group (Jensen *et al* 1999). A strain over-expressing DNA gyrase was requested from the Westerhoff group both from the academic department and by GSK Process Development group, but was not

forthcoming. It would have been very interesting to see the effects of increase gyrase expression on plasmid topology and processing. Since strain development takes considerable time and the strain developed by the Westerhoff group did not arrive this work was not attempted. It seems a logical step to see the effect of such a strain on the plasmid production process. Induced expression would probably be required so as to prevent alterations of the host DNA expression.

1.2.2 Non – Watson Crick base pairing

Watson Crick base pairs are not the only stable way that bases can be held together by hydrogen bonds. Reverse Watson Crick, Hoogsteen and Reverse Hoogsteen bonding, shown in figures 1.4 and 1.5, can all form and their stability accounts for some of the unusual structural features discussed hereafter (Hoogsteen 1963).

Reversed Watson Crick A-T

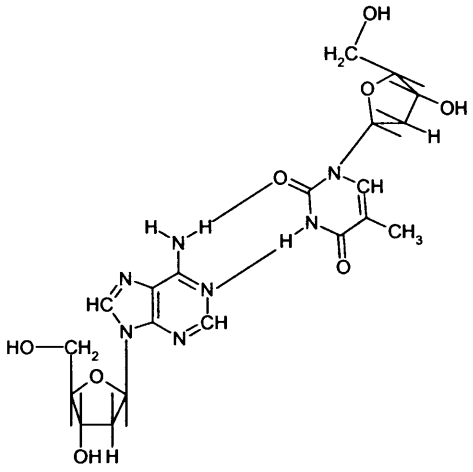


Figure 1.4 A Reverse Watson-Crick base pair. Compared to the Watson-Crick bond, one of the nucleotides is rotated 180°. The glycosidic bonds are in the trans orientation, rather than cis in the Watson-Crick bond.

Hoogsteen A-T

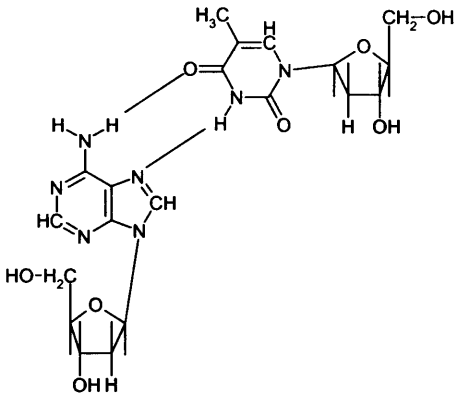


Figure 1.5 A Hoogsteen base pair

Many of the possible hydrogen binding conformations that can be formed between the DNA bases were considered before the Watson-Crick structure was established. The Watson-Crick structure was by no means obvious because the stability of many of the different base pairing arrangements are similar. Indeed it is because of the comparative stability of this array of possible base pair conformations that structures other than the Watson-Crick exist *in vivo*.

1.2.3 Sequences with intrinsic differences from the Watson Crick double helix.

1.2.3.1 The effect of supercoiling

Briefly, supercoiling is torsional stress on double stranded DNA that is maintained either by protein interaction or by virtue of the DNA being circular, for example in a plasmid. Supercoiling in plasmids is described in more detail in a later section 1.3.1.7. The energy from unwinding supercoils is often converted into producing non-Watson Crick type structures. *In vivo* it has been shown that, in both eukaryotes and prokaryotes, during transcription the movement of RNA polymerase along DNA, generates negative supercoils upstream and positive supercoils downstream of the transcription complex (Rahmouni and Wells 1989, Giaever and Wang 1988).

1.2.3.2 Effects in nature.

It has been reasoned that DNA exhibiting a higher order of structure than predicted by the Watson Crick model has biological function (Wells *et al* 1988). One of the main indications for this hypothesis is that specific sequences are required for

unusual DNA structures to form, and these occur more frequently than would be expected if the sequence of nucleotides were random. For example, homopurine·homopyrimidine sequences with the potential to form intramolecular triplexes, appear in eukaryotes three to four times more frequently than they would be expected (Behe 1995, Schroth and Ho, 1995). Purine-pyrimidine tracts, which are able to form Z-DNA, are also over represented up to 40 times more often than would be expected in some organisms (Ussery *et al* 2002). DNA with non-Watson Crick structure may provide recognition sites for protein binding and several roles have been suggested (Ciotti et al 2001, Kiyama and Camerini-Otero 1991, Gagana *et al* 1999, Lee *et al* 1984, Simonsson *et al* 1998). DNA hence can be said to convey more information than might be predicted by the translation of the sequence of bases alone. Additionally there are instances of unusual DNA structures that have detrimental effects.
Examples of all of these are given below.

1.2.3.3 Bending

DNA is not, as was once supposed, a naturally straight rod, and in fact it can bend anisotropically. Intrinsic bends and curves appear when special base sequences are repeated in phase with the DNA helical repeat. "A tracts" are responsible for the vast majority of DNA bends encountered. A tracts are a run of A-Ts around half a helical turn long (5 to 6 bp) repeated in phase, that is at 10 to 11 bp intervals. The degree of bending produced by other sequences has been found to be small in comparison (Crothers *et al* 1990).

Marini and co workers first encountered bent DNA in 1982, when certain DNA fragments were found to exhibit anomalously slow gel electrophoretic mobilities (Marini *et al* 1982). Anomalous gel mobility is caused by DNA molecules of a bent conformation have increased difficulty in passing through gel pores that are not much larger than the DNA helix diameter (Cooper 1977).

An important technique known as "Cyclically permuted gel electrophoresis" was developed using the kinetoplast DNA studied by Marini and co workers. Cyclically permuted fragments of the kinetoplast DNA were produced. It was found that a section of DNA with the bend in the middle ran much more slowly than a cyclically permuted variant containing the bend at one end. Gel electrophoresis of these fragments allowed the site of the bend to be found (Wu and Crothers 1984). This has since become a powerful tool in locating regions of bent DNA.

1.2.3.3.1 Modelling of DNA bending

A number of models have been proposed that attempt to model and predict DNA bending given the sequence information. For the purposes of this project, existing methods to predict deviation from 'normal' DNA structure are important. Should any of the sequences studied produced difficulties in production and processing it would be important to know if sequences planned for inclusion in plasmid products may contain such sequences.

1.2.3.3.2 Examples of bends and purpose in nature.

Intrinsic DNA bends have been found, by using circular permutation polyacrylamide gel electrophoresis and computer modelling, to flank transcription start sites (Schroth *et al* 1992).

Predictions of DNA bending in the *E. coli* genome has shown that promoter regions are more likely to be bent than coding or random sequences (Gabrielian *et al* 1999-2000, Bolshoy and Nevo 2000).

It has also been hypothesized that intrinsic DNA curvature is a common feature in eukaryotic promoters (Marilley and Pasero 1996, Schatz and Langowski 1997) and recent experimental work supports this finding with the study of bends in the human cdc2 promoter (Nair 1998), yeast GAAL1-10 and GAL80 genes (Bash *et al* 2001) and the TATA box (Davis *et al* 1999).

These predictions indicate some role for the bent sequences in promoter function. What precisely this role, however, is unclear, and was the subject of considerable research in the 1990s, with several review papers published on the topic (Hagerman 1990, Harrington 1992, Perez-Martin *et al* 1994). Current opinion remains divided about the role of DNA bends in promoter sequences, and the different theories are summarized below:

1) The curved DNA forms large loops around RNA polymerase enhancing affinity to the complex (Matthews 1992, Perez-Martin *et al* 1994).

2) Small intrinsic curvature can enhance protein-DNA contact, hence intrinsically curved DNA may fine-tune the interaction of promoters and regulatory factors (Suzuki and Yagi 1995, Starr *et al* 1995).

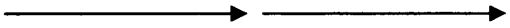
3) DNA curvature may bring together components of the transcriptional complex that are distant (Matthews 1992).

4) Curvature results in helical stress, affecting the energy required to melt DNA and unwind the double helix, therefore assisting (or perhaps functioning in place of) initiation proteins (Natale *et al* 1993).

1.2.3.4 Symmetry elements

DNA often contains repeated elements, generally these are confined to non-coding regions, but this is not always the case. Symmetry elements are direct repeats, mirror repeats and inverted (palindrome) repeats:

Direct repeat



5' AATGTGTCAACG AATGTGTCAACG 3'
3' TTACACAGTTGC TTACACAGTTGC 5'

Mirror repeat

5' AATGTGTCAACG GCAACTGTGTAA 3'
3' TTACACAGTTGC CGTTGACACATT 5'

Inverted (palindrome) repeat

5' AATGTGTCAACG CGTTGACACATT 3'
3' TTACACAGTTGC GCAACTGTGTAA 5'

The repeated regions may or may not be perfectly symmetrical and may be or may not be adjacent.

Non-adjacent inverted (palindrome) repeat

5' AATGTGTCAACG TTGAG CGTTGACACATT 3'

3' TTACACAGTTGC AACTC GCAACTGTGTAA 5'

Triplet repeat sequences are also common and have been found to be responsible for several inherited genetic disorders in humans. It is thought that they are produced by slippage at the replication fork.

These symmetry elements are responsible for many of the intrinsic structures discussed below.

1.2.3.5 Triplex formation (The H form of DNA)

The formation of triplexes in nucleic acids was first discovered in 1957 with the binding of PolyU to PolyA·Poly U. (Felsenfeld *et al* 1957). Early studies involved interactions between single stranded and duplex DNA. In particular PolyT with PolyA·PolyT. In *vivo*, a triplex structure in DNA is formed when a region becomes single stranded, and its complimentary strand hydrogen bonds with double stranded DNA nearby. This produces a portion of DNA in which three strands are intertwined and one strand is without its complimentary sequence. This 'H form' of DNA was described by Mirkin and co-workers, and they deduced from a series of designed plasmids that a homopurine·homopyrimidine mirror repeat is required (Mirkin *et al* 1987). A helical representation of H-DNA is shown in figure 1.6 a.
Supercoiling is a requirement for triplexes and several studies have been concerned with correlating the number of base pairs in a triplex to the number of supercoils relaxed (Glover *et al* 1990). On average one supercoil is unwound per 10.5 base pairs, i.e. one per helical turn in B – DNA.
In a homopurine·homopyrimidine mirror repeat, it is most common for the half of the pyrimidine strand to Hoogsteen bond with a neighbouring duplex, while the other half of the pyrimidine strand is left unpaired. As shown in figure 1.6 b, depending on whether the 5' or 3' half of the pyrimidine strand is incorporated in the triplex the structure is termed Hy5 or Hy3. (Htun *et al* 1984).

Triplexes have been observed using both NMR (Radhakrishnan and Patel 1994) and more recently by X-ray crystallography (Rhee *et al* 1999) and the research findings indicate that the predicted and actual structures are closely matched.

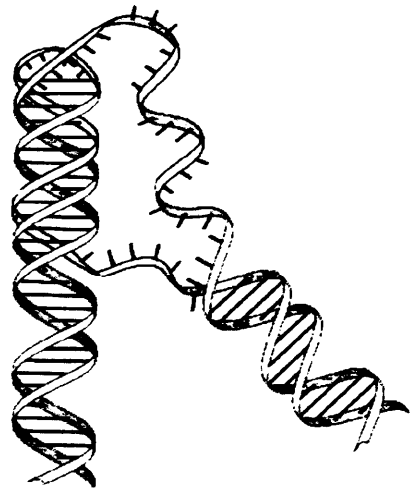
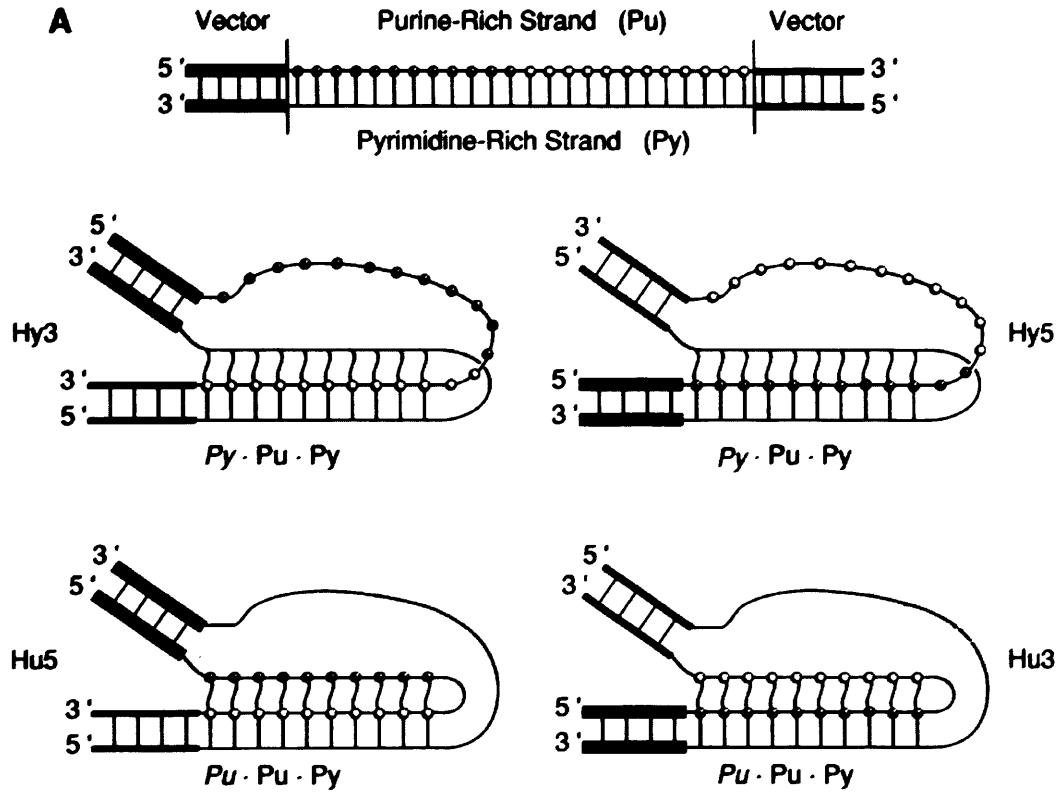


Figure 1.6 Intramolecular DNA triplexes.

a) A helical representation of intramolecular triplex DNA. The third strand lays in the major groove, whereas its complementary strand exists as a single stranded region. Analysis of intramolecular triplex structures on ploy acrylamide gels suggests that DNA containing the intramolecular triplex is significantly bent.

Reprinted from DNA STRUCTURE AND FUNCTION, R R Sinden, Copyright (1994), with kind permission from Elsevier.



b) Four possibilities for intramolecular triplex formation. Intramolecular triplex structures can form within regions of DNA with predominately purines (shown as unfilled circles) in one strand in which the Pu.Py region has mirror repeat symmetry.

Base 'pairing' in triple stranded DNA:

All Watson-Crick base pairs are conserved in triplex structures. In *Pyrimidine*·Purine·Pyrimidine base triplets, the third base (*Pyrimidine*) forms a Hoogsteen pair with the central Purine base:

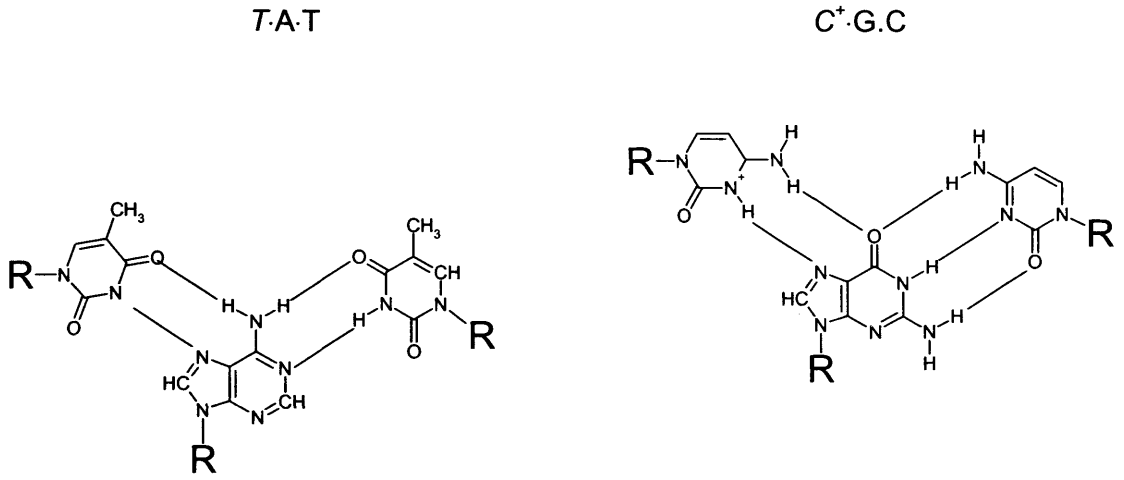


Figure 1.7 Base 'pairing' in *Pyrimidine*·Purine·Pyrimidine base triplets.

By contrast in *Purine*·Purine Pyrimidine base triplets, the third base (*Purine*) forms a Reverse Hoogsteen pair with the central Purine base:

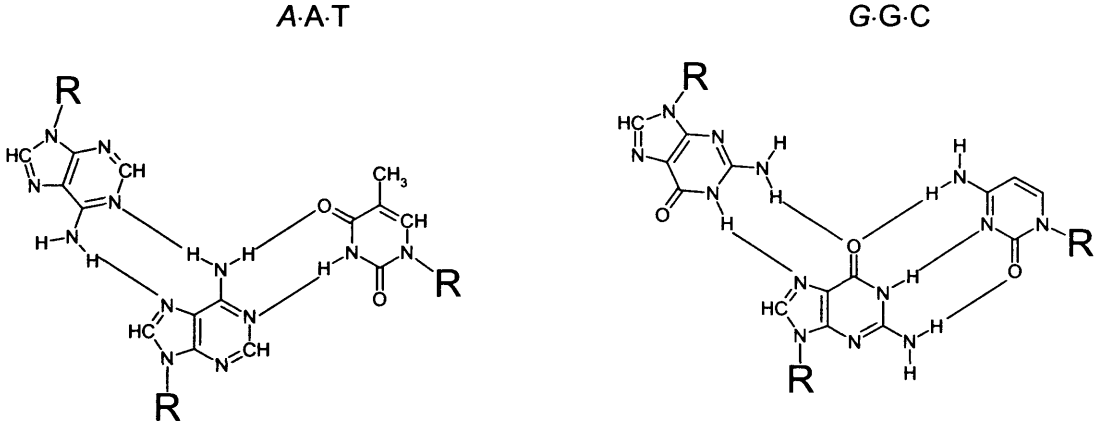


Figure 1.8 Base 'pairing' in *Purine*·Purine Pyrimidine base triplets

The homopurine·homopyrimidine sequences required for triplex formation appear three to four times more often than would be expected and are often found upstream or within genes (Behe 1995, Schroth and Ho 1995). As yet there is no clear understanding of their function or biological importance. Given their prevalence and

proximity to genes, triplexes may be involved with transcriptional regulation (Miller and Sobell 1966).

A number of proteins that bind triplex DNA in preference to duplex DNA, have been shown to exist. (Kiyama and Camerini-Otero 1991, Ciotti *et al* 2001). If H-DNA occurs in nature and has biological function, there would almost certainly also exist proteins that specifically recognise triplex DNA. Hence this is taken as good proof for the formation of triplexes *in vivo*.

There is much potential for triplex DNA to be used in gene therapy applications. This generally involves oligonucleotides designed to form triplexes with duplex DNA. Such a triplex upstream of a gene has the potential to block transcription and prevent a particular protein from being synthesised. This is called 'the antigene strategy' (Cooney *et al* 1988) and a recent review discusses the progress made in this field (Chan and Glazer 1997).

1.2.3.6 Cruciform-like structures (Holliday junctions, Four-Way Junctions)

Cruciform structures are formed from palindromic sequences. Formation requires the central region of the sequence of DNA to unwind. This has an energy requirement, which has been shown to come from negative supercoiling. Cruciform structures only exist in supercoiled DNA and not in either linear or open circular forms (Lilley 1980).

Since the central region of the DNA must unwind, the base composition of this region affects the formation of cruciforms. Palindromes with central regions containing high amounts of A-T require less energy to form cruciform structures than those with high amounts of C-G. This is because there is one less hydrogen bond within A-T than C-G base pairs, and additionally a lower stacking energy between adjacent bases. This was proven to be the case by Zheng and Sinden (Zheng and Sinden 1988).

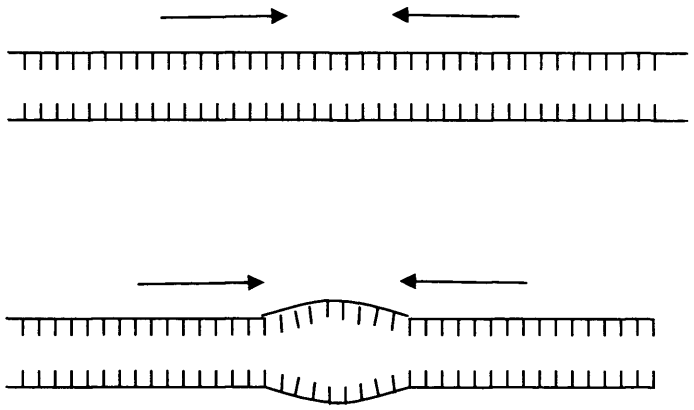


Figure 1.9 Cruciform formation

Inverted repeats occur non-randomly in the DNA of most organisms and several roles have been suggested for them in *vivo* (Pearson *et al* 1996). Among these is control of the initiation in plasmid replication (Noirot *et al* 1990). A sequence with potential to form a cruciform has also been found close to the origin of replication of human mitochondrial DNA (Crews *et al* 1979).

More recently a sequence likely to form a cruciform structure was found in the

promoter region of several human genes. This implies some role in initiating transcription, particularly as these promoters do not have the normal TATA box (Gadhavi *et al* 2001).

1.2.3.7 DNA quadruplexes / tetraplexes and telomeres

DNA quadruplexes were first discovered at the end of chromosomes, in telomeres. Several different sequences have been shown to be responsible for quadruplexes, falling broadly into two groups: G quadruplexes (Williamson 1994) and i-motif quadruplexes (Gehring *et al* 1993). The quadruplex structures of telomeres have been well studied and some of these are given below.

Until recently it was not thought that DNA quadruplexes existed outside of telomere regions, however, several other sequences have now been shown capable of forming G-quadruplexes. These include immunoglobin switch regions (Sen and Guilbert 1988), insulin polymorphic region (Catasti *et al* 1996), Retinoblastoma susceptibility genes (Murchie and Lilley 1992), and in the c-myc control region (Simonsson *et al* 1998).
Their function in the immunoglobin switch region are probably to serve as recombination sites in order to produce extensive antibody diversity.
The HIV virus has also been shown to produce quadruplex structures, but further research is required in this area (Skripkin *et al* 1994).

Indirect evidence for quadruplex function in nature has been obtained in the form of proteins shown to bind them. The yeast Rap1 protein has been shown to accelerate G quadruplex formation (Giraldo and Rhodes 1994).

1.2.3.7.1 G-quadruplexes

Telomeres consist of simple tandem repeat sequences, which are generally extremely G rich, for example $d(NG_4N)$.
The basic unit of G quadruplexes is the G tetrad (Arnott *et al* 1974, Zimmerman *et al* 1975). The G tetrad is four G bases bonded by both Watson-Crick and Hoogsteen bonds in a square planar arrangement as shown below.

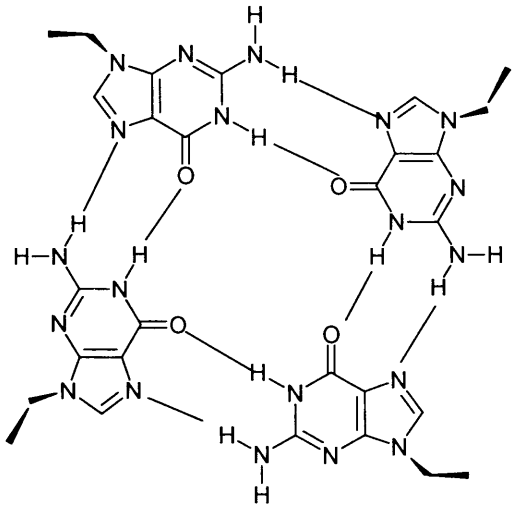


Figure 1.10 The G tetrad

There are several possible arrangements for G quadruplexes, depending on relative strand directions and the *syn*/*anti* distribution around each G tetrad. The figure above shows all glycosidic bonds in the *anti* conformation.

NMR studies of many repeated sequences have taken place, and combining information from molecular dynamics the structures of a number of quadruplexes has been resolved (Smith and Feigon 1992, Wang and Patel 1993, Wang and Patel 1994, Wang and Patel 1995, Smith *et al* 1995)

1.2.3.8 Triplet repeats

Triplet repeats have been shown to be responsible for more than twelve human genetic diseases, including fragile X syndrome, myotonic dystrophy and Friedreich's ataxia (Chastain and Sinden 1998). Hairpin structures, specifically those associated with fragile X syndrome $(GCC)_n$ and $(GGC)_n$, have been shown to produce slippage at replication forks while DNA is duplicating (Chen *et al* 1998). These slippage events, due to the unusual DNA structures, help to explain both the occurrence of triplet repeats and also the varying length polymorphisms associated with some triplet repeat inherited diseases (Chen *et al* 1998, Catasti *et al* 1999). Different triplet repeats have been shown to produce a wide variety of stable DNA structures. Using NMR it has been shown that the repeat responsible for Friedreich's ataxia, (GAA / TTC) forms triplex structures (Mariappan *et al* 1998), that responsible for myotonic dystrophy (CTG) forms hairpin structures (Mariappan *et al* 1996), and that

a repeat responsible for fragile X syndrome (GCC) forms quadruplex structures (Fry and Loeb 1994).

Triplet repeated sequences either interrupt gene-coding sequences or form unusual structures that effect gene expression.

For the purposes of this project triplet repeats will not be considered. Previous studies have shown highly repetitive sequences to be unstable in *E. coli* (Hashem *et al* 2002) and additionally it is unlikely that triplet repeats would be included in future plasmid products.

1.2.3.9 Centromeric DNA

The human centromeric repeat is (AATGG)n and forms a doubly folded hairpin structure (Catasti *at al* 1999). Such a sequence is unlikely to be included in future plasmid products and so was not considered for the purposes of this project.

1.2.3.10 "Slipped" Mispaired DNA (SMP-DNA)

Regions of DNA containing direct repeats can form slipped, mispaired structures. In order to form, the entire region must untwine and one strand of one copy of the direct repeat must pair with the other strand of the other copy. In supercoiled DNA this formation might be energetically favourable as supercoils would be lost, however there would also be a loss in stability due to a reduction in both base stacking and hydrogen bonding (Sinden 1994).

The first suggestion of SMP-DNA came from a group who found sequences sensitive to S1 nuclease (Hentschel 1982). Conclusive proof of their existence took some time longer however as the sequences examined by Hentschel may have been able to form triplex structures, themselves sensitive to S1 nuclease.

Long direct repeats have already been shown to be genetically unstable in *E. coli* (Williams and Müller 1987), and so have not been further considered.

1.2.4 'Levels' of structure, modelling and prediction

References to primary, secondary, tertiary and quaternary levels of structure have, as far as possible, been avoided in this thesis, the exception being when referring to problematic PCR reactions, where secondary structure refers to any non B-DNA form. The reason for avoiding other references to levels of structure is because different definitions for each have been widely used in the literature. The author's personal view is that there are 5 levels of structure that can be defined as follows.

1) Primary – The sequence of bases (e.g. CAGGTAC)
2) Secondary – The basic conformation (e.g. B-DNA, A-DNA, Z-DNA)

3) Tertiary – Intrinsic bending (e.g. straight, strong bend)
4) Quaternary – Non Watson-Crick structures present with no supercoiling
5) Quinary – The effect of supercoiling and its associated structures.

Prediction of unusual DNA structures from sequences alone would be advantageous as difficulties with plasmid processing might be avoided or foreseen.

A large number of models exist to model both single stranded RNA and single stranded DNA. As discussed in the earlier section on DNA bending, limited models of double stranded DNA are available to establish the tertiary structures of B-DNA; however, few modelling tools exist for the prediction of double stranded DNA structure. Possible reasons for this are the difficulty of modelling such large molecules, lack of computing power and the lack of perceived need. Some modelling work has been performed on the structure of supercoiled DNA, including tertiary and quaternary structure. Modelling of tertiary structure (DNA bends) is described in section 1.2.3.3.1. These models assume DNA exists only in the B- form however, and they do not take account of any of the other structures formed by DNA such as triplexes, nor of any interactions that may occur between sequences in different regions of a plasmid.

The current position is that there are no modelling tools available to accurately predict DNA structure. As such intrinsic abnormalities described are generally considered curiosities, especially as their biological functions are, as yet, relatively unknown.

Using simple software it is possible to detect DNA repeats. As mentioned in section 1.2.3 repetitive sequences are responsible for many of the unusual DNA structures that have been observed in nature. It ought to be relatively simple to detect homo purine / pyrimidine regions that might produce triplex structures, purine-pyrimidine tracts that encourage Z-DNA formation, and direct repeats that can cause SMP-DNA.

1.3 Plasmid Properties

1.3.1 Plasmid properties in nature

The term 'plasmid' was first coined in 1952 to describe all extrachromosomal hereditary elements (Lederberg 1952). Plasmids were shown to consist of DNA in 1961 when it was shown that bacterial mating is accompanied by a transfer of DNA (Marmur *et al* 1961) and this was confirmed using radiolabelled DNA in 1962 (Silver and Ozeki 1962).

1.3.1.1 General properties

Bacterial plasmids are extrachromosomal and consist of double stranded DNA. Plasmids may be covalently closed circles, open circles or linear in shape. They range in size from between a few kilo bases up to 1.7 mega bases.

1.3.1.2 Plasmids confer host cell advantages

Plasmids may carry genes essential for host cell growth and division, or genes that give the host a particular advantage within the bacterial population. Such advantages include antibiotic resistance, heavy metal ion resistance, virulence, nitrogen fixation, ability to degrade unusual compounds (metabolism functions) and 'fertility'. 'Fertility' will be discussed later, but is involved in one method of plasmid transfer.

1.3.1.3 Host range

Plasmids tend to have a narrow host range and this is probably due to their reliance upon particular host genes for their replication. Some plasmids belong to 'incompatibility groups', these are groups of similar plasmids that cannot coexist within the same cell. Many plasmids exist with a lot of copies of the plasmid per cell. The number of plasmids per chromosome is termed the 'copy number' and is regulated. Copy numbers vary between different plasmids; generally, large plasmids have low copy numbers, and small plasmids have high copy numbers.

1.3.1.4 Copy number control

Plasmid genes altering the rate of initiation of replication control the copy number. There are two major methods of initiation control:

- RNA primers are one method of initiating DNA replication and this RNA production can be regulated using antisense RNA. A typical and well-understood example of this kind of control is the ColE1 plasmid. The rate of production of the replication primer called RNA II is manipulated:

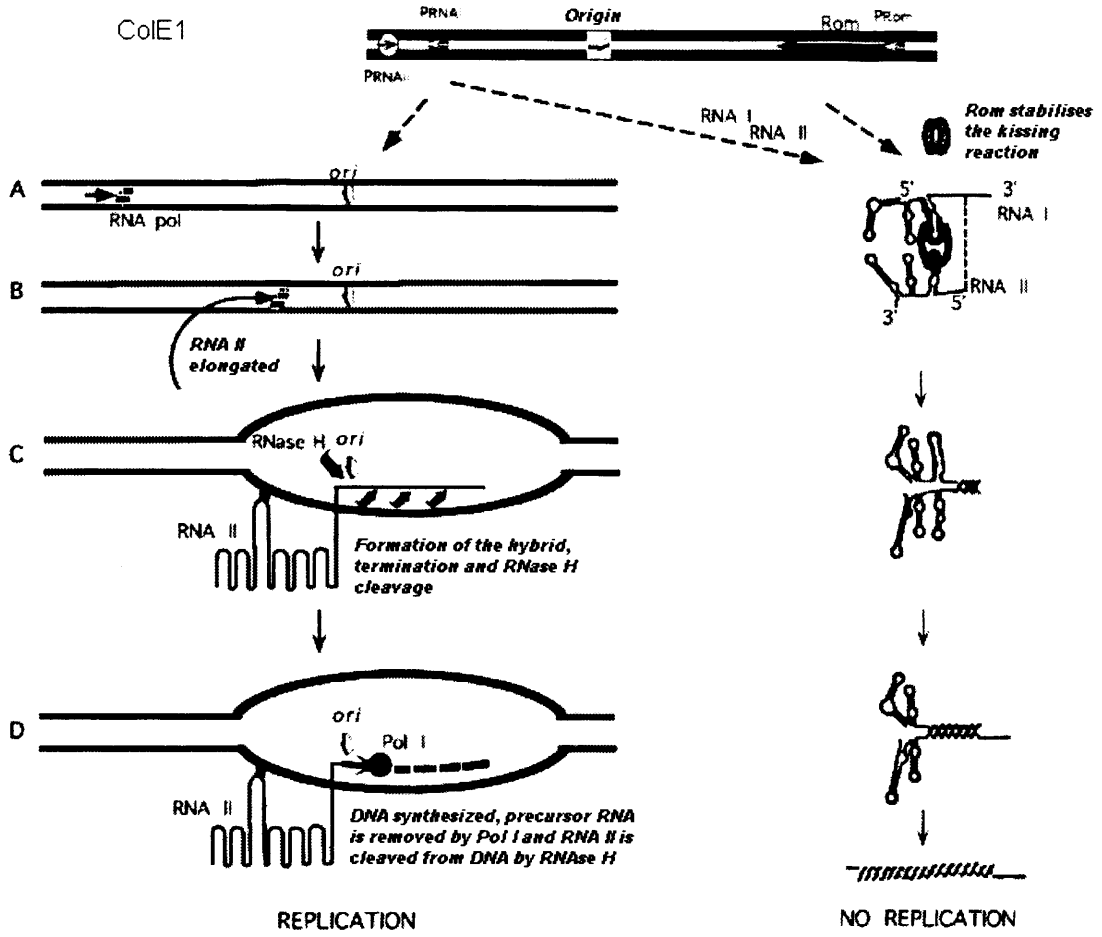


Figure 1.11 ColE1 plasmid number control mechanism. Figure adapted from Actis *et al* 1998.

The left-hand side of the diagram depicts the mechanism of initiation of replication. The arrows inside black or white circles indicate initiation of transcription locations for RNA II and RNA I. The origin of replication (*ori*) and the site of action of RNase H are indicated. The steps involved in the initiation of replication are denoted by diagrams A to D. The interaction between RNA I and RNA II that leads to the inhibition of DNA synthesis is shown on the right hand side of the figure. The first interaction between RNA I and RNA II (kissing) is reversible and stabilised by the Rom protein. The concentrations of all RNAs and Rom protein depend upon the number of plasmids within the cell and hence copy

number is controlled.

- Alternatively the function of essential replication proteins may be regulated by binding to repeated 18-22 base pair sites called iterons.

Examples of both types of regulation are encountered in both large and small plasmids.

1.3.1.5 Horizontal plasmid transfer

Horizontal plasmid transfer can occur in three ways, conjugation, transformation and transduction. In nature the most important method is conjugation.

Conjugation is the transfer of genetic material from one bacterial cell to another via cell-to-cell contact. Plasmids can be classified into four groups: Those that code for their own transfer are called Conjugative and are either Repressed or Derepressed; plasmids not coding for their own transfer are deemed Non-conjugative and are either Mobilisable or Non-mobilisable.

Gram positive and Gram negative (double membrane) bacteria have different conjugating methods. Gram negative conjugative bacterial plasmids produce sex pili. The 'fertility' function referred to earlier, encoded by the F-plasmid, is such an example. There are three types of pili, thin flexible, thick flexible and rigid. Each type of pilus may be more advantageous than another under different conditions. The pilius is used to bring the bacteria close together. Genetic transfer then occurs through the 'conjugation bridge' which may or may not be the sex pilus:

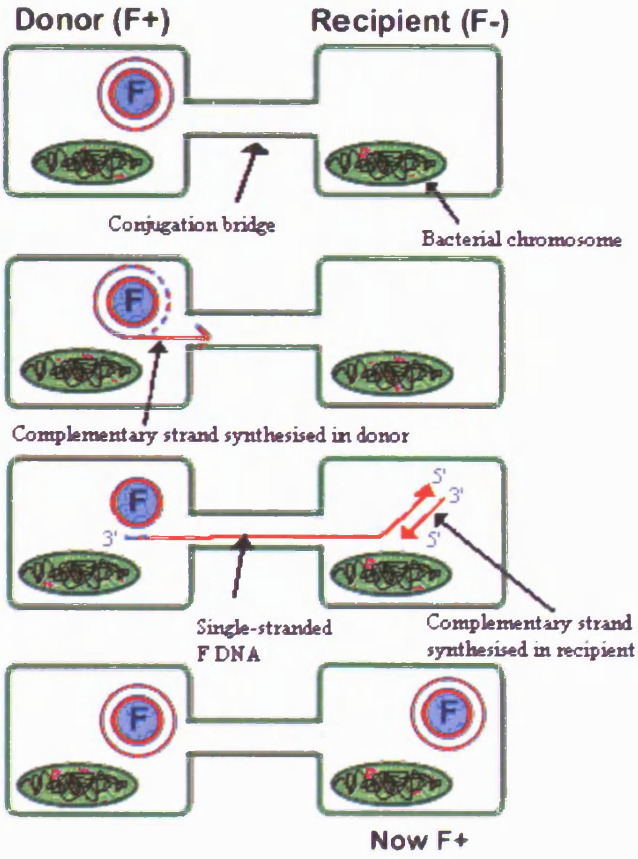


Figure 1.12 Conjugative plasmid transfer

Such conjugative plasmids tend to be large, low copy number ones.

Derepressed plasmids are always ready to be transferred, whereas the transfer functions in repressed plasmids are almost always switched off. Transfer of repressed plasmids occurs either because the repression does not always work or when the repression is stopped due to susceptible cell detection. In *Streptococcus* for example the recipient produces a sex pheromone that induces formation of a special surface protein making the cell sticky and so the donor and recipient adhere together.

Mobilisable plasmids are not self transmissible, but use the transfer apparatus of other plasmids to mediate their transfer. This is possible as they contain an 'origin of transfer' sequence and mobilising genes.

In nature plasmids are either conjugative or mobilisable.

Transformation involves uptake of naked DNA from the environment. It is possible

that plasmids could be released into the environment (for example from dead cells) and these could be taken up by other bacterial cells.

Plasmids could also be transferred by Transduction. Faulty packaging of bacteriophage DNA occurs infrequently and in its place a plasmid could be wrapped. Since the attachment and injection functions of bacteriophages are contained in the protein coat and not the genetic material that they contain, such incorrectly packed virion particles containing plasmid DNA would be active. Whilst this is possible it is unlikely to occur in nature.

1.3.1.6 Plasmid stability

Plasmids may be lost from cells either via structural or segregational instability:

1.3.1.6.1 Structural instability

Recombination, transposition and slippage during replication can all result in large-scale DNA rearrangements. Homologous and illegitimate recombination generate deletions, inversions, duplications and fusions.

In order to prevent new variants of cloning vectors from proliferating, the processes by which they arise are blocked. Cloning vectors are "stress-tested" and bioinfomatics is used to find repeat sequences, which ought to be removed. Regions of homology have been found to result in loss of parts of the plasmid. This can be a result of homologous recombination in recombinant proficient cells – so called 'crossing over' occurring between sections of DNA containing similar sequences. Deletions may also occur in cells that are not recombinant proficient due to slippage during DNA replication. It is also important that vectors place modest metabolic demands on their hosts in order to minimise the potential changes in fitness associated with changes in plasmid structure (Summers 1994).

1.3.1.6.2 Segregational instability

When plasmid-containing cells divide there are three possible ways in which the plasmids may be distributed:

a) Random

b) Some plasmids may be actively partitioned

c) All plasmids are actively partitioned.

Naturally occurring low copy number plasmids have previously been show to exhibit active partitioning. If the distribution was random then after several generations a low copy number plasmid would be contained in only a small percentage of the population.
High copy number plasmids inheritance is likely to be described by random distribution.

Assuming a random distribution the probability of a cell division producing a plasmid free cell is $2 \times (0.5)^n$, where n is the plasmid copy number, assumed to be constant over the cell life.
For copy numbers over 20 this theory predicts very stable populations of plasmid containing cells. In fact experimental results show that the populations are not quite as stable as predicted.

1.3.1.6.3 Effect of plasmid load

Plasmids in cells can produce significant metabolic burden. Cells containing fewer plasmids than the general population will have a lower metabolic burden than those having more plasmids. These cells thus have a selective advantage. This can lead to a "selective drift" resulting in lower copy numbers in the general population. This does not alter the rate at which plasmid free cells arise; however, plasmid free cells will not have such a burden and will therefore outgrow the plasmid containing cells. This leads to plasmid free cells accumulating quickly.
Vectors with low copy number that can be induced to be of high copy number have been suggested as a method to prevent plasmid free cells having a notable advantage.

1.3.1.6.4 Effect of copy number variance:

Measurement of copy number gives only mean copy number values and it is difficult to determine the standard deviations of these. A few cells with low copy number would inevitably produce many more plasmid free cells than predicted by random distribution. This is the most likely cause of the disparity between predicted and observed stability in high copy number plasmids mentioned above.

1.3.1.7 Supercoiling of plasmids

Supercoiling affects plasmid structure and as mentioned previously herein, many unusual DNA structures will only form when DNA is supercoiled.

In vivo closed loops of double stranded DNA are ubiquitous. Plasmids, bacterial chromosomes (Bauer 1978) and viral genomes (Vinograd *et al* 1965) form single closed loops, while eukaryotic DNA is organised into domains by DNA binding proteins (Section 1.2.1.1).

Plasmids exist in three main forms, linear, open circular and supercoiled. These forms are sometimes referred to as Type III, Type II and Type I respectively.

1.3.1.7.1 Linking number

Linking number, described by White, has been used to describe the degree of supercoiling for closed loops of DNA (White 1969):

Supercoils added to a loop of DNA are manifested in two different ways:

- Decreasing the number of base pairs per helical turn and hence the number of helical turns in the loop.
- Causing the DNA helix to form a helix of higher order.

These two properties of circular DNA are described by the terms Twist number (Tw) and Writhing number (Wr). Tw gives the number of helical turns in the DNA, while Wr describes the supercoiling of the DNA in space. The Writhing number and the number of helical turns can vary continuously as the shape of the DNA duplex is not fixed; however, their sum will stay constant. This quantity is the Linking number and will be unchanged so long as the DNA strands are not broken.

Linking number is defined as the number of times one strand crosses the other when the DNA is made to lie flat on a plane. The only ways in which linking number can change are by chemical / enzymatic action, or by formation of unusual DNA structures (Sections 1.2.3.5 and 1.2.3.6).

1.3.2 Plasmid properties in biotechnology

1.3.2.1 Horizontal transfer

To allow control over the plasmid containing cells, any natural mobilisable and conjugative features are either removed or disabled. This also helps prevent uncontrolled spread of genetically engineered DNA that is unacceptable to the general public.

1.3.2.2 Antibiotic resistance.

Antibiotic resistance is used as a selective marker to select for plasmid containing cells. Selection for several plasmids in the same cell can be achieved if each confers resistance to a different antibiotic. As discussed in the next chapter, there is a choice of antibiotic resistance marker in plasmid products, and there may be no need for them at all given recent technological developments.

1.3.2.3 Inheritance stability.

Oligomer formation

Plasmid multimers are formed by homologous recombination. Oligomers are maintained at low copy number. This has been shown to be due to the copy number control mechanism depending on the number of origins of replication:

A series of pUC8 derivatives containing tandem repeats of the origin region produced a progressive drop in copy number (Summers and Sherratt 1984).

The naturally occurring plasmid ColE1 is very stable, particularly when compared to similarly sized cloning vectors. Cloning vectors are commonly lost at frequencies of between 10^{-2} and 10^{-5} per cell per generation under non-selective conditions (Summers and Sherratt 1984).

By using deletion derivatives of the plasmid it was found that ColE1 stability is due to a 240 bp *cis*-acting site called *cer*. Oligomers of *cer*$^{+}$ plasmids are converted to monomers by intramolecular recombination between the *cer* sites (Summers 1991).

The inclusion of *cer* into cloning vectors is therefore advantageous because the number of plasmids available at partitioning is increased. The use of *cer* in industry is widespread as it provides the additional benefit of reducing levels of multimers.

More recent work has shown that the *cer* sequence produces a short mRNA that somehow inhibits cell division, ensuring that plasmid multimers are resolved before

the cell divides (Chatwin and Summers 2001). This discovery has great possibilities in terms of applications such as whole cell biotransformations. In the context of genetic vaccines and gene therapy, it gives cause for concern however as the sequence might be expressed and effect mammalian cells. The safety implications of the use of the *cer* sequence in plasmid products therefore need further consideration.

1.3.2.4 Multiple cloning sites

All commonly used vectors contain closely arranged synthetic cloning sites. These are made up of sequences that are recognised by restriction enzymes. These sites ought to occur only once on the plasmid. The sites can be used by themselves or in tandem to allow a great variety of DNA fragments to be cloned.

1.3.2.5 Size

For the requirements of cloning, it has been desirable to use plasmids that are small in order that as large as possible fragments of DNA of interest can be inserted. This is because cloning using small plasmids is easier than with larger ones. In particular transformation efficiency is inversely related to plasmid size. Size becomes the biggest factor after 15Kb. Large plasmids are also more difficult to characterise by restriction mapping.

1.4 Plasmids as Biotherapeutics

1.4.1 Gene Therapy

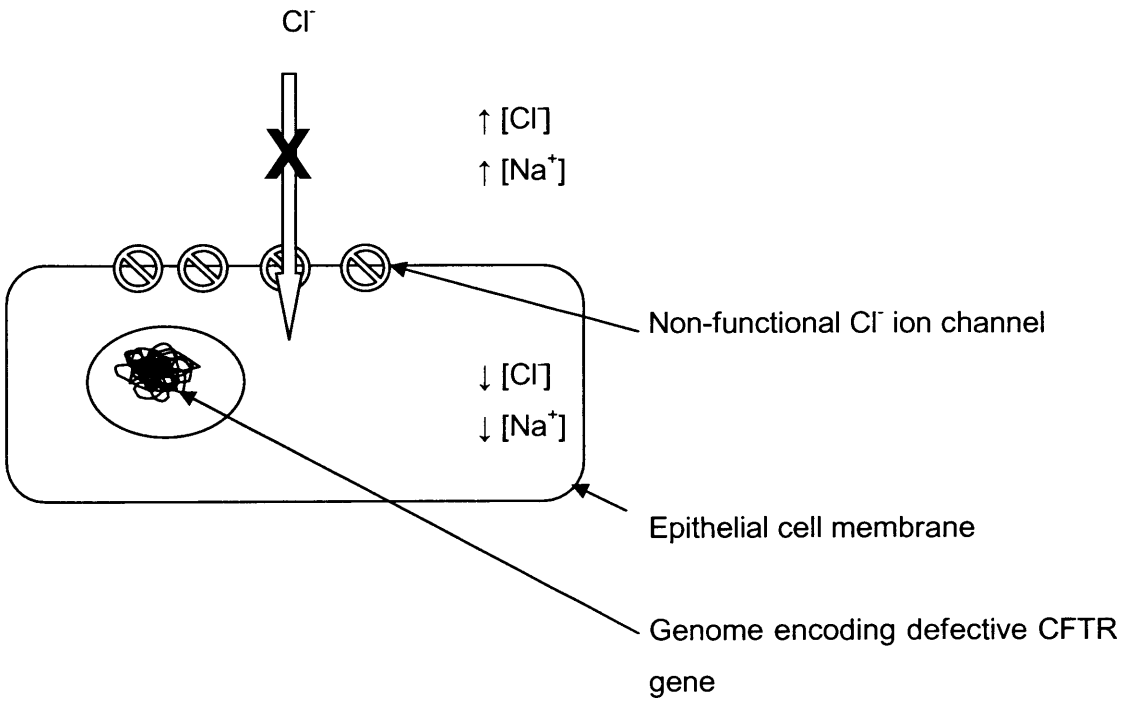
Since it was first coined, the meaning of the term Gene Therapy has evolved significantly as the technology has advanced and been applied to different problems. Simplistically gene therapy is replacing or fixing a defective gene. The most modern and widely accepted definition is 'the treatment or prevention of disease by gene transfer' (Mountain 2000).

Over 4,000 human diseases are caused by single gene defects (Gottschalk and Chan 1998). These include sickle cell anaemia, cystic fibrosis, Hypercholesterolaemia, Duchenne muscular dystrophy and some forms of dwarfism. The rationale behind corrective gene therapy is straightforward: A functional wild type gene is introduced to cells that are unable to produce the given protein due to a mutation in the gene. The introduced gene is expressed by the cell, producing the

desired protein. A large amount of excitement was induced when researchers revealed that skeletal muscle cells injected with plasmid DNA were able to express plasmid-encoded genes (Wolf *et al* 1990).

A better understanding of gene therapy and its implications is obtained by illustration with an example:

Cystic Fibrosis is the most common life shortening hereditary monogenic autosomal recessive disease in Europeans, effecting 1 in every 3,200 births and resulting in around 400 deaths per year. Its study has received much attention in the Gene Therapy field: A single defective gene, normally producing a Cl^- channel protein (Cystic Fibrosis Transmembrane Regulator – *CFTR*), causes the recessively inherited disease (Bradley 1999, Crytsal *et al* 1994, Davies *et al* 1998, Koehler *et al* 2001, Matsue and Teramoto 2000, Rosenstein and Zeitlin 1998). Life expectancy with conventional treatments of physiotherapy, antibiotics and pancreatic supplements of cystic fibrosis sufferers is 30 years (Bradley 1999, Geddes and Alton 1998). Soon after the discovery of the *CFTR* gene, it was shown that by providing a normal copy of the *CFTR* cDNA sequence, the genetic defect could be corrected, hence increasing life expectancy (Crystal *et al* 1994, Davies *et al* 1998).



The non-functional *CFTR* protein results in osmotic imbalance.

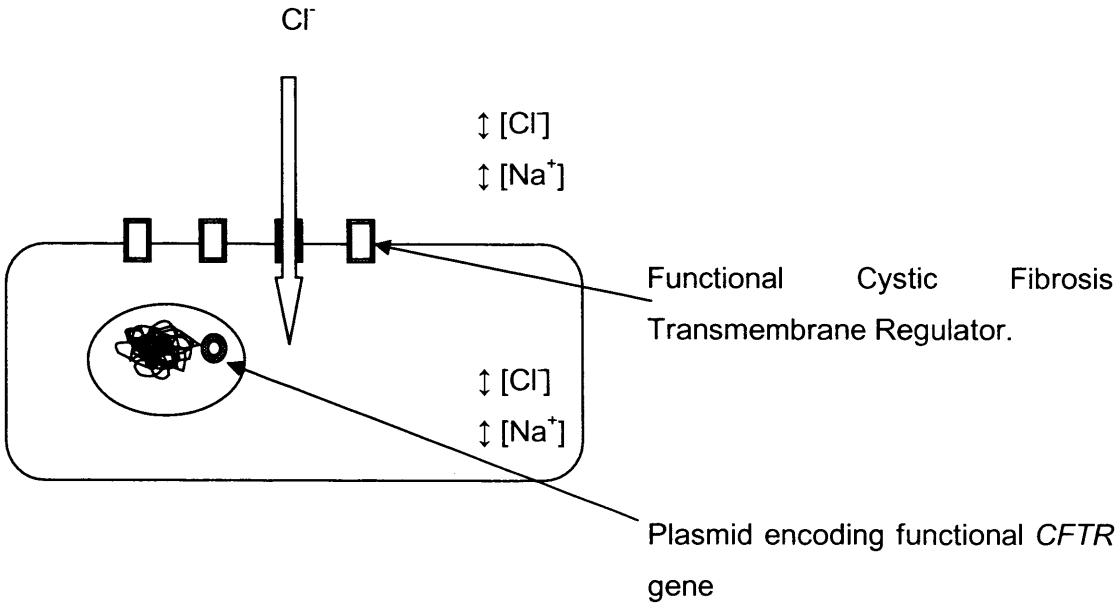


Figure 1.13 Somatic Gene Therapy concept. Schematic showing how a plasmid encoding a functional *CFTR* gene could result in defective cells having working Chloride ion channels and, hence alleviating the symptoms of Cystic Fibrosis.

The concept behind gene therapy is that a gene encoding the *CFTR* protein could be introduced to the cell. The gene would then be transcribed and translated by the host machinery producing active *CFTR* protein. This would allow the balance of Cl^- to be maintained and so alleviate the symptoms of cystic fibrosis.

Unfortunately, as yet, the promise of gene therapy has not been realised. Over the last decade around 12 independent cystic fibrosis clinical trials have been reported, trying different gene delivery methods and targeted sites. None of these can be counted as a success, and correction of cystic fibrosis has not been achieved and successful gene delivery is estimated to be another decade away (Caplen 2001, Geddes and Alton 1998, Koehler *et al* 2001).

1.4.2 DNA Vaccines

Following the finding that plasmid-encoded genes could be expressed in mammalian cells, it was found that mice produced antigen-specific antibody responses to proteins expressed by plasmids (Tang *et al* 1992). Mice have been successfully vaccinated against the influenza virus by means of plasmids containing influenza genes; the mice were shown to both express antibodies and cytotoxic T-lymphocyte responses were produced (Ulmer *et al* 1993).

There is significant interest in DNA vaccines as they have several advantages, in terms of economic, environmental effects and safety (Donnelly *et al* 1996, Hassett and Whitton 1996, Davis 1997) over more conventional treatments: Their production does not require cultivation of dangerous or infectious agents, and unlike some viral vectors that have been used for gene delivery there is no risk of an attenuated virus back mutating and recreating the virulent form. DNA vaccines may also provide treatment against diseases for which the production of attenuated virus has so far proven unsuccessful e.g. HIV. There is also the potential to co-administer treatments against a number of diseases, on a single or multiple plasmid vectors.

1.5 Plasmid DNA manufacture and processing

1.5.1 Shear damage to plasmids during processing

One of the major difficulties associated with production of plasmid DNA is damage to plasmids due to shear. Shear forces encountered during plasmid processing affect the structure of the plasmid, which is important for its proper functioning (Levy *et al* 1999). The sensitivity to shear has been shown to increase with the size of the plasmids studied.

1.5.2 Cell recovery

Due to the dramatic effects that shear has on plasmid DNA, mechanical methods are generally not used to recover plasmids from bacterial cells. In preference, chemical lysis is used. The most common method used is alkaline lysis, essentially as originally described by Birnboim and Doly (Birnboim and Doly 1979).

1.5.3 Alkaline lysis

This involves the addition of an alkaline – SDS mixture. The SDS solubilizes the phospholipid and protein components of the cell membrane, causing lysis of the bacterial cells. This is generally performed in the presence of RNAse in order to degrade the host RNA, although alternatives to this exist. The chromosomal DNA and host proteins are irreversibly denatured by the sodium hydroxide. Long incubations with sodium hydroxide have also been show to irreversibly denature the plasmid DNA and so are avoided.

Plasmid DNA is able to renature upon neutralisation: The alkaline mixture of cell contents is neutralised by addition of acidic potassium acetate. This has a very high salt and K^+ ion content and causes the SDS to precipitate. These salt-detergent complexes trap the denatured proteins, chromosomal DNA and other cellular debris and forming a delicate gel matrix of SDS-protein, called a flocculating suspension, or 'floc'.

The low molecular weight plasmid DNA renatures correctly and remains in solution, the floc is hence removed.

1.5.4 Clarification

The choice of unit operation to remove the floc is made between centrifugation, filtration or this step may be totally bypassed and the alkaline lysate may be purified directly by chromatography.

As the floc and plasmids are delicate, high shear forces must be avoided. Small-scale plasmid preparations generally use fixed-angle centrifuges (Lahijani *et al* 1996). However, for large-scale preparations, centrifuges are avoided. This is due to the semi-continuous mode of operation of industrial type centrifuges: the centrifugal acceleration acting on the liquid entering the centrifuge causes high shear forces, which can break up the floc (Theodossiou *et al* 1998).
On a large scale, filtration is considered to be the better option than centrifugation, since large shear forces are not present and so the floc is not broken up. (Horn *et al* 1995, Theodossiou *et al* 1998). Scale up of filtration operations is also widely considered easier than scale up of centrifugation. Filtration allows the solution state of the retentate to be easily adapted by using buffers in order to modify solution conditions, such as pH and salt concentration, in order to make the required changes before later chromatographic steps.

Some research indicates that the floc may be applied directly from cell lysis to chromatographic steps, without any loss in purity or quality and increasing yield (Ferreira *et al* 1999).

1.5.5 Purification and regulatory aspects

Plasmid DNA prepared as described above will be contaminated with bacterial chromosomal DNA, RNA, proteins and endotoxins. In addition impurities may be introduced to the system from the fermentation process itself. Plasmid required for pharmaceutical purposes must be of substantially higher purity than that required for traditional molecular biology. Plasmid DNA can hence be separated using a number of further filtration and/or chromatographic steps.

General requirements both for current Good Manufacturing Practise (cGMP) for producing finished pharmaceuticals, and of general biological product standards; are in place. (Code of Federal Regulations 2000 Title 21 Parts 211 and 610 respectively). The cGMP requirements cover the methods, processes and facilities and are designed to ensure products are consistently safe, pure and potent. The general standards cover tests required to ensure that all lots of vaccines for human use meet defined specifications over purity, identity, potency and general safety.

Further to these general guidelines, the American Food and Drug Administration (FDA) gives stringent guidelines over the amounts of permitted contaminants in plasmid DNA products. These guidelines are outlined in an FDA points to consider

document published in 1996.

Interpretation of the criteria has been made by a number of research groups. Curling and Smith produced an easy to follow summary of the criteria.

Property	Criteria
Biological activity	Coherent fragments with plasmid restriction digests Expected migration in gel electrophoresis (for size & supercoiling) Transformation efficiency comparable with plasmid standard
Purity Supercoiled isoform Absorbance A_{260} / A_{280}	 >95% >1.8 <2.0
Impurities Protein RNA Genomic DNA Endotoxins	 <0.1 % <2 % Undetectable by gel electrophresis <0.01 µg/µg plasmid by Southern Blot <0.01 EU/µg plasmid

Table 1.2 Summary of FDA points to consider. Taken from Curling and Smith 2003, claiming to quote the FDA Points to consider 1996 (Curling and Smith 2003)

A further requirement has been laid out for pharmaceutical grade plasmid DNA (Smith and Klinman 2001); while it was common practice to verify the sequence of the protein-encoding gene, some research suggests the plasmid backbone itself is of importance (Sato *et al* 1996, Klinman *et al* 1997). Hence the FDA now require a submission of the full plasmid sequence before permission is given to begin phase I clinical trials.

Routinely all steps in the purification process are tested in order to ensure that phage would be removed if it ever entered the system.

The advent of novel stains for gel electrophoresis and qPCR technology means that assays for contaminants may now be considerably more sensitive. It therefore seems likely that some of the criteria given by the FDA will be updated in the near future.

Interestingly, in contrast to the table above (claiming the FDA 1996 guidelines as their source), the FDA points to consider document actually does not specify a minimum percentage of the supercoiled plasmid isoform, rather stating:
'Plasmid-derived DNA species such as linear and relaxed circular DNA may be less effective in expressing the inserted antigen gene. There should be a specification for the minimum amount of supercoiled DNA present.' (FDA Points to consider, December 1996).
Following these guidelines, plasmid DNA manufacturing based research has aimed to produce pure plasmid DNA in the supercoiled form. Whilst plasmid based products have yet to reach market, minimum levels for the amount of supercoiled DNA have been suggested as 90% or 95% (Shamlou 2003, Curling and Smith 2003 respectively) based on levels published elsewhere. Supercoiled plasmid DNA can be converted to open circular form and then linear forms by the action of shear, potentially at any stage of the purification process (Shamlou 2003). By setting criteria for the proportion of supercoiled plasmid in the final product the reproducibility of the purification procedure can be monitored.
This criterion is discussed further in Chapter 6.

1.6 Project aims

Future plasmid products may contain structures such as those described in section 1.2 for two major reasons:

- In addition to containing a correct copy of the coding sequence, future gene therapy products may also include promoter regions, introns, untranslated regions and enhancers.
- Sequences taken from viral genomes for the purposes of genetic vaccination may form non-Watson Crick structures.

This study was to see what effects, if any, inclusion of these structures would have on the properties of plasmid particularly with respect to their production.

A summary of objectives:

- To produce and characterise a plasmid backbone closely resembling those used in the pharmaceutical / biotech industry.
- To clone sequences of DNA known to produce unusual intrinsic structures

into the parental plasmid. The sequences were to be included at a defined point in order to produce a family of plasmids.

- To characterise the family of plasmids.

Characterisation of the plasmids was to include any effects on fermentation yield, host cell growth rate and plasmid stability (defined section 1.7). Dependent on findings further analysis of the effect of particular unit operations upon plasmid production was envisaged.

1.6.1 Previous studies including structural abnormalities in plasmids

1.6.1.1 A large artificial cruciform

A large, 5.5Kb artificial palindrome was reportedly cloned into pUC19 a high copy number plasmid (Ravin and Ravin 1998). This is particularly interesting since it had previously been shown by many research groups that cloning and maintenance of perfect inverted repeats of longer than 150 bp was probably not possible (Lilley 1981, Mizuuchi *et al* 1982). Long inverted repeats are often deleted via intramolecular recombination (Warren and Green 1985) and by slippage at replication forks (Sinden *et al* 1991).

The large palindrome reported by Ravin and Ravin, was shown to reduce copy numbers in supercoiled plasmids. Palindromes of this size have not been found in nature however, and in addition, because inverted repeated sequences are unstable in *E. coli* (see above) it is unlikely that structures of this kind would ever be incorporated into plasmids for gene therapy or genetic vaccine use.

Ravin and Ravin also reported an increase in multimer formation. Unfortunately the paper is in Russian and due to difficulties in translation, it has been difficult to establish if these 'multimers' are true multimers, or simply smaller plasmids generated because the large amount of repetitive sequence increases the chances of homologous recombination.

1.6.1.2 Random *BamHI* DNA fragments from human lymphocytes

Of more interest are the random clones that Ravin and Ravin produced later (Ravin and Ravin 1999). They produced 70 clones containing random *BamHI* DNA

fragments from human lymphocytes. Of these, three had very low copy numbers. It was reasoned that these three clones probably contained DNA exhibiting unusual secondary structure. By using a novel linear vector the same sequences were not found to be problematic. As described in section 1.2 many of the unusual intrinsic structural features are only produced when DNA is supercoiled. Hence by using a linear vector system, potential problems can be avoided. Interestingly Ravin and Ravin state that they are not sure which of the properties of their novel plasmid is responsible for cloning 'abnormal' sequences. It seems likely that the lack of supercoiling in the linear vector, allows stable retention of 'abnormal' sequences, as without supercoiling the DNA is likely to remain in a simple form without higher levels of structure.

1.6.1.3 Intrinsic DNA bends

The theoretical effects of localized bending on plasmid structure when supercoiled has previously been examined. Results from computerized modelling of short intrinsic bends in supercoiled plasmids, predict major differences in the most stable structures that can be formed (Yang et al 1995).

1.6.1.4 Direct repeats

A direct repeat has been shown to be genetically unstable in *E. coli* (Williams and Müller 1987), possibly due to SMP-DNA formation and subsequent deletion.

1.6.1.5 Z-Triplex in Eukaryotic cells

A sequence associated with the Chinese hamster *dhfr* gene can produce Z-DNA or triplex DNA structures under different experimental conditions (Wells *et al* 1990). This sequence was found to reduce plasmid replication after 48 hours in COS-7 cells by 20-50% when a single copy was cloned in either orientation on either side of the SV40 origin of replication. When copies of the motif were placed on both sides of the SV40 origin, replication was reduced by 85-90%. Two-dimensional gel electrophoresis indicated that the Z-triplex region causes replication intermediates to accumulate during the late phases of replication. Yield of plasmids containing the Z-triplex produced in *E. coli* were also reported to be reduced although this was not measured in a rigorous way (Brinton *et al* 1991).

1.6.1.6 Genetic instability due to transformation into *E. coli*

A recent paper has indicated that the genetic stability of plasmid DNA is highly dependent upon transformation into the cell. Certain sequences have been shown to be lost significantly more frequently than others (Hashem *et al* 2002).

1.7 Project outline

A suitable plasmid backbone was to be developed, to which intrinsic structural abnormalities could be added at a defined point. Any effect upon host growth could then be determined in fermentation, along with any changes in yield. Plasmid stability was also examined. There are three forms of plasmid stability: Segregational stability, Genetic stability and Structural stability.

The family of plasmids to be studied contains plasmids with the following features: an intrinsic bend, a purine•pyrimidine tract capable of forming Z-DNA, an intramolecular triplex and a four-stranded (quadruplex) structure.
Passing references in literature are made as to effects that unusual structures might have on plasmid DNA production (Brinton *et al* 1991, Levy *et al* 1998).
Due to different definitions of the word 'stability' when used in both fields of engineering and biochemistry, working descriptions used for the purposes of this project are given below.

1.7.1 Cell growth

Cells must harbour plasmid in order to produce sufficient antibiotic resistance. With a truly selective medium system plasmid free cells would not survive. Additionally in truly selective media there may be some form of titration system in that cells containing many plasmids might grow more quickly than cells with fewer plasmids that do not produce sufficient antibiotic resistance.

Plasmids containing unusual structural features might be more difficult for bacterial cells to replicate. Possible reasons for this include stalling of replication forks when the unusual structure is encountered as has been shown to occur with one sequence in eukaryotic cells (Brinton *et al* 1991). This would result in reduced copy number, which might adversely affect cell growth due to a reduced resistance to the antibiotic used.

1.7.2 Segregational stability

This is most likely to be the biologist's view of the term 'stability'. When bacterial cells containing plasmids divide, the amount of plasmid entering each of the daughter cells may not be equal. Since plasmids are associated with metabolic burden, cells containing fewer plasmids have a competitive advantage. This is what can lead to selective drift – a reduction in average copy number with fermentation time. The competitive advantage due to harbouring fewer plasmids is balanced against the need for cells to harbour plasmid in order to produce sufficient antibiotic resistance. By examining plasmid yields from fermentations, containing the family of plasmids, any differences in segregational stability could be inferred.

1.7.3 Genetic stability

This is most likely to be the biochemist's view of the term 'stability'. The sequences included in the plasmid family have unusual structures. Structures such as these have not been found in the *E. coli* genome to date. It was therefore unknown if the host DNA replication machinery would faithfully replicate the sequences added.

A previous study has indicated that Z-DNA is genetically unstable in *E. coli*. Deletions were shown to remove the Z forming sequence to a size no longer able to adopt the Z conformation at normal levels of superhelical density (Freund et al 1989). Simple dinucleotide repeats were used however, and these can produce several other secondary structures pertaining to deletion as discussed in a later publication by the same research group (Bichara et al 2000) and shown by another research group (Hashem *et al* 2002). It is hence difficult to conclude that problems encountered previously were due to the Z-form or structures formed because of the highly repetitive sequences used.

1.7.4 Structural stability

This is most likely to be the engineer's view of the term 'stability'. Structural stability refers to the tendency of a plasmid to be damaged. As is suggested in a recent paper concerning the effects of shear on plasmid production (Levy *et al* 1998) unusual structural features may greatly alter the level of shear at which the supercoiled plasmid form may be converted to open circular or linear form.

The theoretical effects of localized bending on to the most stable supercoiled plasmid structure are dramatic (Yang *et al* 1995). Despite the included sequences being so small in length, in comparison to the plasmid as a whole, the most stable three-dimensional forms of plasmids containing the sequences will be substantially

different from those of supercoiled plasmids that do not contain unusual structures. It would hence seem intuitive that they will have different sensitivities to shear stresses.

Implications of plasmids containing intrinsic unusual DNA structures

Possible consequences of their inclusion include:

- Structural instability.
- Higher amounts of recombination.
- Altered shear sensitivity.
- Greater amount of open circular DNA.
- Reduction in copy number and hence yield.
- Reduction in the growth rate of plasmid containing cells.

1.7.5 A description of the sequences chosen for inclusion in this study

Incorporating sequences exhibiting unusual properties found around human genes gives a more realistic representation of what may occur in future gene therapy plasmid products. Much literature has indicated that sequence context has major influence over the formation of DNA structures. Hence the use of PCR was thought to be the best way to generate sufficient quantities of target sequences to allow them to be cloned. Viral sequences would have been more difficult to obtain in the time available for the study, and at the time the study started, few intrinsic structures had been found in double stranded DNA generated from viral sequences.

1.7.5.1 DNA quadruplexes

1.7.5.1.1 The Insulin –linked polymorphic region.

Discovery of a quadruplex structure in the insulin-linked polymorphic region was made in 1984 (Owerbach and Aagaard 1984). The nucleotide sequence of a long polymorphic region (positions 134-2096) located 365 bp upstream of the human insulin gene is composed of 139 repeating sequences whose consensus structure is related to 'acaggggtgtgggg'. These large numbers of G bases are responsible for the quadruplex structure (Castasti et al 1997, Lew et al 2000). The formation of inter- and intramolecular G-quartets is thought to modulate insulin transcription (Lew et al 2000).

As the repeated sequences were spread over a 2 kB region, it was planned to clone the entire 2 kB region into one of the larger parental plasmids. In the event of unusual plasmid properties being found, a 2 kb piece of 'control DNA' would have been added to the same parental plasmid to allow direct comparison without the influence of size related effects.

PCR of this region was attempted several times, but was unsuccessful. Due to time constraints and progress in cloning other unusual structures cloning of the region was abandoned.

1.7.5.1.2 G-Quartet – Nuclease Hypersensitivity Element

The c-myc control region has been shown to contain a DNA quadruplex (Simonsson *et al* 1998), and direct evidence that the structure involved is a G-quadruplex has been obtained (Siddiqui-Jain *et al* 2002). A nuclease hypersensitivity element (NHE) has been shown to control 85-90% of c-myc transcription and it's structure and its effect on c-myc transcription has been extensively studied. The G-quadruplex formed acts as a negative regulator of the c-myc gene, and conversion to unstructured single stranded forms is required before transcription can occur (Postel *et al* 2000).

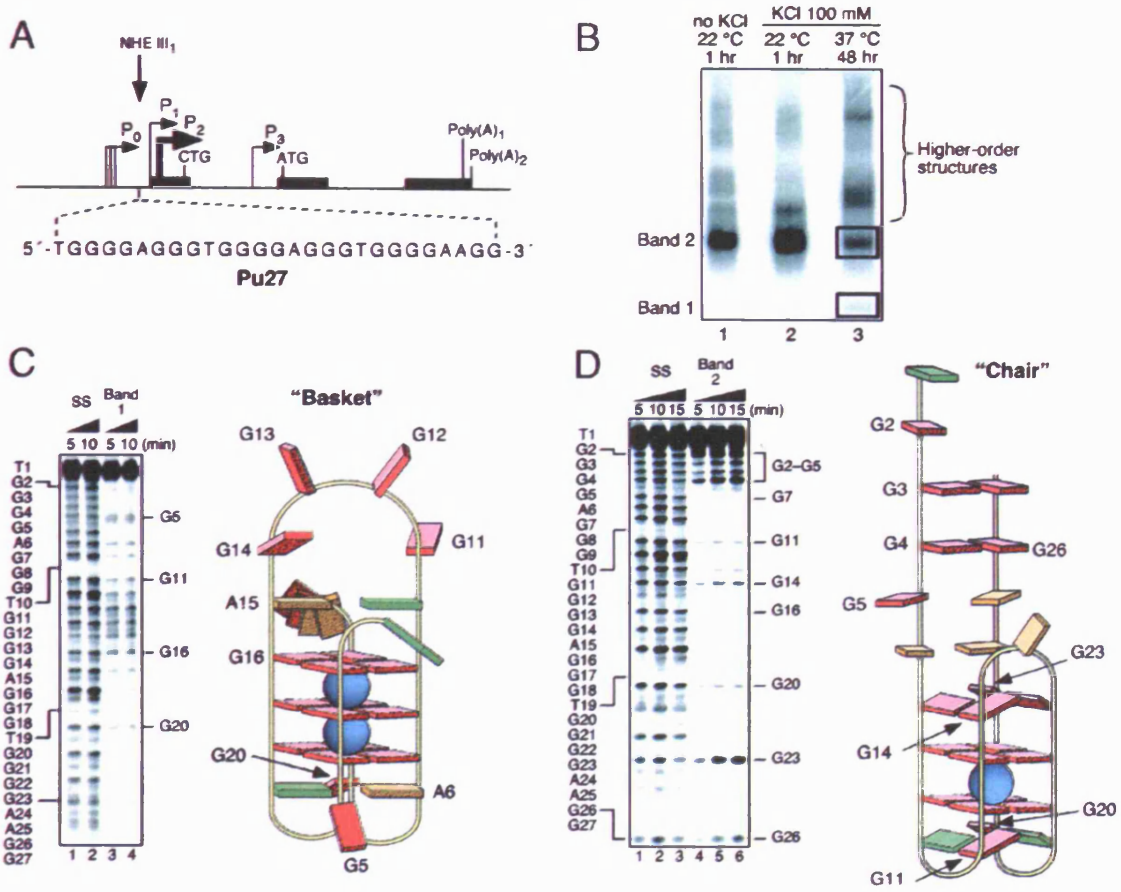


Figure 1.14 Quadruplex structures of the c-myc region. Determination of the structures of the unimolecular G-quadruplexes formed after incubation of the Pu 27 strand (Pu27) in 100 mM KCl for 48 h at 37°C. (A) Promoter structure of the c-MYC gene; shown in Inset is the 27-mer sequence of the purine-rich strand upstream of the P1 promoter (3). (B) Nondenaturing gel analysis (15% polyacrylamide/12.5 mM KCl/NaCl, 4°C, 16 h) of Pu 27 preincubated under the conditions specified in the figure at a strand concentration of 25 μM. (C) DMS footprinting of band 1 in B. (Left) DMS treatment of the denatured Pu27 (lanes 1 and 2) and the isolated band (lanes 3 and 4). The Pu27 base sequence is shown to the left. (Right) Proposed structure based upon the footprinting pattern. Base colors: red, guanine; green, thymine; and orange, adenine. (D) As in C, but for band 2 in B. (Figure used with permission from Siddiqui-Jain *et al* 2002).

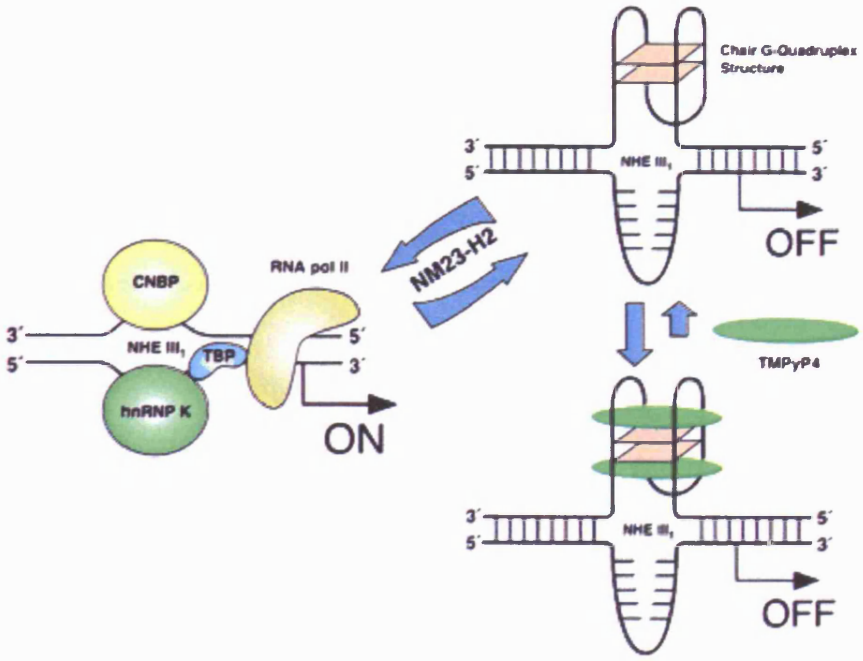


Figure 1.15 Model for the activation and repression of gene transcription involving the accessory role of NM23-H2 in interconversion of the unstructured purine and pyrimidine single-stranded DNA forms to the paranemic secondary DNA structures. Interaction of the G-quadruplex structure with TMPyP4 stabilizes the gene-off form by inhibition of conversion to the single-stranded gene-on forms Figure reprinted from Siddiqui-Jain *et al* 'Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription' PNAS 2002, with permission from the National Academy of Sciences Copyright 2002.

Formation of the quadruplex in plasmids in *E. coli* might effect plasmid replication, and during processing may make the plasmid more susceptible to damage.

1.7.5.2 Triplexes

Eukaryotic sequences have been found to be S_1 sensitive and so probably form triplex structures (Mirkin *et al* 1987). As the presence of triplex structures was thought to be context sensitive, cloning was designed to include the surrounding sequences. The promoter region of the human platelet-derived growth factor (PDGF) A-chain gene contains three S_1 sensitive regions (Lin *et al* 1992).

The PDGF has potential gene therapy applications for a novel approach to periodontal tissue engineering (Giannobile *et al* 2001). Hence if difficulties were found to arise in production of a plasmid containing the promoter region of this gene, then this would directly relevant as the sequence may occur in future gene products.

1.7.5.3 Z-DNA

For this study, a region with the potential to form Z-DNA was taken from an intron of the Human coagulation factor IX gene (Yoshitake *et al* 1985). This contains a long alternating purine-pyrimidine sequence. This was seen as a better model than insertion of a long repeated CG sequence since it not only exists in nature, but also would be less likely to form other unusual structures, such as cruciforms.

Simple dinucleotide repeats have been shown to be unstable (Bichara *et al* 2000), however if a clone was made with a long $(CG)_n$ repeat, it could have been tested in a range of host cells to see if the insert would be maintained with higher fidelity.

1.7.5.4 Intrinsic bends

Bends have been shown to have a great effect on structure by theoretical models of supercoiled plasmids. (Yang *et al* 1995)

1.7.5.4.1 - cdc2 promoter

This is a human promoter sequence that is intrinsically bent and does not contain a TATA box (Nair 1988).

1.7.5.4.2 A Highly Curved Sequence

Using freely available DNA bending modelling software (bent.it ®) an extremely bent sequence of DNA was designed. This featured phased A-tracts interspersed with sequences bending the DNA even further.

AAAAATGGCCCAAAAAGGGCCCAAAAATCTCCAAAAATATATAAAAAGGCCTAAA
AACCCGGTAAAAATATATAAAAAGGGCCC

This is a highly curved sequence (18.0 degrees per helical turn). This sequence will hence bend the DNA by around 140 degrees. It was decided to use a relatively long sequence rather than a shorter repeated one. A shorter repetitive sequence (such as $(TCTCTAAAAAATATATAAAAA)_n$) would be more likely to produce other DNA secondary structure, rather than simply induce curvature, and additionally it would have been difficult to obtain a clone in which the two oligonucleotides correctly annealed due to the highly repetitive sequence.

1.8 The next steps

With several unusual sequences identified, a suitable parent plasmid was to be constructed and characterised, into which the unusual sequences could be cloned.

Chapter 2:

Materials and Methods

2 Materials and methods

2.1 Cultures

2.1.1 *Escherichia coli* bacterial strain genotypes

The *E.coli* K-12 strains used as plasmid hosts were:

DH5α [F'/endA1 hsdR17 ($r_K^- m_K^+$)supE44 thi-1 recA1 gyrA (Nal^r) relA1 Δ(lacZYA-argF) U169 (ϕ80dlacΔ(lacZ)M15] (Woodcock *et al* 1989).

DH1 [F^-recA1 endA1 gyrA96(Nal^r) thi-1 hsdR17 supE44 relA1] (Woodcock *et al* 1989).

2.1.2 Growth and maintenance of bacterial strains

E. coli strains were maintained in 20 % glycerol stocks and stored at -70 °C. Glycerol stocks were prepared from selective agar plates using 5 mL sterile 20 % (v/v) glycerol. Multiple colonies were resuspended from the surface of the agar with the sterile pipette tip used to measure the 5 mL 20% glycerol solution. The cell suspension was aliquoted in 1.8 mL volumes, and stored in 2 mL Nunc tubes at -70 °C. A single wire-loopful of such a stock was sufficient to inoculate a selective agar plate.

Bacteria were grown at 37 °C on nutrient agar, or in LB medium using a rotary shaker (New Brunswick Scientific). Plasmid harbouring strains were maintained by supplementing liquid/solid media with selective antibiotics as appropriate, according to plasmid phenotype: ampicillin sodium salt (500 μg mL^{-1}) for ampicillin resistance, kanamycin sulphate (50 μg mL^{-1}) for kanamycin resistance.

In order to produce plasmid, small-scale cultures were grown overnight in 5 mL selective nutrient broth, using 20 mL universal bottles. 1.5 mL aliquots were spun down in sterile eppendorf tubes using a bench top benchtop centrifuge (13,000 rpm, 10 minutes).

To produce larger amounts of plasmid, cultures were grown in 50 mL selective nutrient broth, in 1 L shake flasks overnight and cell paste harvested by centrifugation.

2.2 Media buffers and solutions

Unless otherwise stated chemicals and media components were supplied by BDH Laboratory Suppliers, Dorset, UK. In general water from a Milli-Q system (Millipore,

Billerica, Massachusetts, USA), was used for electrophoresis and cell cultivation. Sterile-filtered cell culture tested water (Sigma, Gillingham Dorset, UK), was used for manipulations using enzymes.

2.2.1 Growth Media

Nutrient agar (Oxiod, Hampshire, UK) was produced as recommended using 28 g L^{-1}. LB medium (Sigma, Dorset, UK) was produced as recommended using 32 g L^{-1}. Fermentation media consisted of 48 g L^{-1} yeast extract, 25 g L^{-1} HY SOYA, 10 mL L^{-1} glycerol, 6 g L^{-1} Na_2HPO_4, 3 g L^{-1} KH_2PO_4, 1 g L^{-1} ammonium chloride and 100 µL L^{-1} poly(propylene glycol). Media was autoclaved for 20 minutes at 121 °C, 15 psi.

2.2.2 Antibiotic stock solutions

Both Ampicillin and Kanamycin were produced at stock concentrations of 50 mg mL^{-1} in sterile-filtered cell culture tested water, and were filter sterilised (0.22 µm filters, Acrodisc, Gelman Sciences, Ann Arbor, USA), and stored at -20 °C.

2.2.3 Buffers and other solutions

(i) TE buffer
10 mM Tris-HCl (pH 7.5); 1 mM Na_2EDTA

(ii) EB buffer (Qiagen)
10 mM Tris-HCl (pH 8.5)

(iii) Loading buffer (normal)
0.25 % (v/v) bromophenol blue; 0.25 % (v/v) xylene cyanol; 30 % (v/v) glycerol; 0.1 M Na_2EDTA

(iv) Loading buffer (for linking number gels)
0.25 % (v/v) bromophenol blue; 2.5 % (v/v) xylene cyanol; 30 % (v/v) glycerol; 0.1 M Na_2EDTA

(v) Restriction enzyme buffers
All restriction enzymes used were purchased from NEB and the commercially supplied buffers were used.
BSA was added to restriction digests from a stock solution of 10 mg mL^{-1}, to a final concentration of 100 µg mL^{-1} as required.

(vi) 1 x DNA gyrase buffer (John Innes, Norfolk, UK)

35mM Tris-HCl (pH 7.5), 24 mM KCl, 4mM $MgCl_2$, 2mM dithiothreitol, 1.8mM spermidine, 1mM ATP, 6.5% (w/v) glycerol, 0.1 mg mL^{-1} albumin

(vii) 1X T4 DNA ligase buffer

50 mM Tris-HCl (pH 7.5); 10 mM $MgCl_2$; 10 mM dithiothreitol; 1mM ATP; 25 µg mL^{-1} BSA

(viii) 1X *Taq* PCR buffer (Roche)

200 mM Tris-HCl (pH 8.4); 500 mM KCl; 1.5 mM $MgCl_2$

(ix) Tris-Borate electrophoresis buffer (TBE)

A 10X stock of TBE (0.45 M Tris-Borate; 0.01 M EDTA) was produced by dissolving 54 g Trizma base, and 27.5 g Orthoboric acid in distilled water, adding 20 mL of 0.5 M Na_2EDTA (pH 8.0), and adding further distilled water to a final volume of 1 L. The stock solution was then autoclaved, and subsequently diluted 1:10 with distilled water prior to use.

(x) TBE/Ethidium Bromide staining buffer

10 mg mL^{-1} EtBr (Sigma) was added to 1X TBE (final concentration of 0.2 µg mL^{-1})

(xi) SYBR Gold Staining buffer

10,000 X concentrate SYBR Gold (Molecular Probes, Leiden, The Netherlands) was diluted 10,000 fold in 1 X TBE

(xii) Phenol/Chloroform

Phenol:Chloroform:Isoamyl alcohol (25:24:1), saturated with 10 mM Tris-HCl (pH 8.0); 1 mM Na_2EDTA, was obtained from Sigma. The phenolic phase was pH 6.7 ± 0.2.

(xiii) Alkaline lysis buffers for plasmid DNA extraction (Qiagen)

P1 (Resuspension buffer) 50 mM Tris-HCl (pH 8.0); 10 mM Na_2EDTA; 100 µg mL^{-1} RNase A

P2 (Lysis buffer) 200 mM NaOH; 1 % (w/v) SDS

P3 (Neutralisation buffer) 3.0 M potassium acetate (pH 5.5)

(xiv) TSB Buffer

100 mL LB broth, 1 mL 1M $MgCl_2$, 1 mL 1 M $MgSO_4$, 10 g PEG 3500, 5 mL DMSO

(xv) HeLa Nuclear Extract Transcription buffer (Promega)
20mM HEPES (pH 7.9 at 25°C); 100mM KCl; 0.2mM EDTA; 0.5mM DTT; 20% glycerol

(xvi) Acrylamide monomer
A 30 % (w/v) Acrylamide; 0.8 % (w/v) Bis-Acrylamide stock solution (EC-890) was obtained from Protogel.

(xvii) Tris-HCl resolving gel buffer
A 1.5 M Tris-HCl (pH 8.8); 0.384 % (w/v) SDS stock solution (EC-892), was obtained from Protogel.

(xviii) Tris-HCl stacking gel buffer
A 0.5 M Tris-HCl (pH6.8); 0.4 % (w/v) SDS stock solution (EC-893), was obtained from Protogel.

(xix) Coomassie Brilliant Blue stain solution
2 g Coomassie Brilliant Blue R 250 (Sigma) was dissolved in 400 mL distilled water, before the addition of 500 mL methanol and 100 mL glacial acetic acid.

(xx) Destain solution
40 % (v/v) methanol; 10 % (v/v) glacial acetic acid

2.3 Molecular biology

2.3.1 DNA preparation

2.3.1.1 Preparation of alkaline lysate for yield determination by PicoGreen assay

Bacterial pellets containing approximately 2.5 x 10^9 cells were resuspended in 250 µL of buffer P1 (as above). 250 µL of buffer P2 (see above) was then added, and the cell solution was mixed gently yet thoroughly by inverting 4 - 6 times. 350 µL chilled buffer P3 (see above) was then added, and mixed as before. The lysate was cleared by centrifugation in a bench top centrifuge (13,000 rpm, 10 min). Clarified lysate was placed into fresh tubes for the PicoGreen assay.

2.3.1.2 Preparation of samples for topology assay

Bacterial pellets containing approximately 2.5 x 10^9 cells were resuspended in 125 µL of buffer P1 (as above). 125 µL of buffer P2 (see above) was added, and the cell solution was mixed gently yet thoroughly by inverting 4 - 6 times. 175 µL chilled buffer P3 (see above) was then added, and mixed as before. The lysate was cleared by centrifugation in a bench top centrifuge (13,000 rpm, 10 min). The clarified lysate was added to 42.5 µL chilled 3M Sodium Acetate and 850µL –20°C 100% Ethanol. After incubation for 1 hour at –20°C the DNA was recovered by centrifugation in a bench top bench top centrifuge (13,000rpm, 10min, 4°C) and subsequently resuspended in 50 µL elution buffer.

2.3.1.3 Small scale preparation for cloning, yield determination by OD_{260} and linking number experiments

Bacterial pellets containing approximately 2.5 x 10^9 cells were resuspended in 250 µL of buffer P1 (as above). 250 µL of buffer P2 (see above) was then added, and the cell solution was mixed gently yet thoroughly by inverting 4 - 6 times. 350 µL chilled buffer P3 (see above) was then added, and mixed as before. The lysate was cleared by centrifugation in a bench top centrifuge (13,000 rpm, 10 min), and the clarified supernatant instantly applied to the anion-exchange resin of a QIAprep spin column, and bench top centrifuged for 1 min. The column was washed with 0.75 mL buffer PE, before being centrifuged for a further minute to remove any remaining traces of the ethanolic wash buffer. Finally the plasmid DNA was eluted into a fresh eppendorf tube by the addition of 50 µL of EB buffer, and centrifuging.

2.3.2 Enzymatic manipulation

Restriction digests and other DNA modifications were carried out using the conditions recommended by the enzyme suppliers (Promega, New England Biolabs, John Innes). Typically plasmid restriction digests were performed using 1 µg of DNA using 10 Units of restriction enzyme, in a total volume of 20 µL with 2 µL of 10 x buffer, as supplied with each enzyme.
Ligations were incubated at 16°C overnight.

2.3.3 PCR

PCR was used in order to amplify sequences for cloning, both from chromosomal and plasmid templates.

PCR was performed in 200 μL tubes, generally with a total reaction volume of 25 μL, this comprising of suitably dilute template (5 ng – 200 ng per reaction), 2.5 mL 10 x Taq buffer, 0.5 mL 10 U mL^{-1} Taq Polymerase, 0.5 mL 10 mM each dNTPs, 1.5 mL 25mM $MgCl_2$ and 1mL of each 10 pM Primer. Reactions were prepared on ice. A Perkin-Elmer GeneAmp Thermal Cycler was used, typically for 25 cycles: 95 °C, 5 minutes (1st cycle only) 95 °C, 45 s; 60 °C, 1 minute; 72 °C, 1.5 minutes; 72 °C 15 minutes (last cycle only).

2.3.4 *In vitro* transcription

A commercial HeLa nuclear extract *in vitro* transcription system was used (Promega, Madison, USA). Essentially the same reaction conditions were used as recommended by the manufacturers, however the mRNA was purified differently.

Each reaction used 7.4 μL of HeLa Nuclear Extract, corresponding to eight units of activity, as determined by Promega. Each unit of activity being defined as the amount of extract required for the incorporation of 50 fmol of nucleotides into a 363-nucleotide runoff transcript generated from the CMV immediate early promoter per hour at 30 °C, under standard conditions.

Each reaction contained 7.4 μL HeLa cell nuclear extract, 2.6 μL HeLa Nuclear Extract transcription buffer, 1.5 μL 50 mM $MgCl_2$, 1 μL 100 mM (each) rNTPs, 2 μL 50ng μL^{-1} treated plasmid made up to a volume of 25μL with Nuclease-free water giving a concentration of $MgCl_2$ of 3 mM and rNTPs of 0.4 mM.

Reactions were incubated at 30 °C for 60 minutes.

After 60 minutes incubation the plasmids were spiked with 175 μL of 1.34 ng μL^{-1} (172 μL 50 ng μL^{-1} mixed with 6228 μL water) human RNA (Applied Biosystems). This allowed any discrepancies in the RNA recovery procedure to be normalized.

2.3.5 RNA purification

RNA was recovered using an automated MagNApure isolation kit for total RNA (Roche). Briefly, under chaotrophic salt conditions RNA is bound to the silica surface of Magnetic Glass Particles. DNase digests any DNA present. Unbound substances such as proteins (including nucleases), cell membranes, and PCR inhibitors are removed by washing with buffer. Remaining cell debris is removed and chaotrophic salt concentration is lowered by a second washing buffer. Purified RNA is eluted at elevated temperature.

2.3.6 Reverse Transcriptase

The reverse transcriptase reaction was carried out in duplicate to check for experimental error. RNA was converted to DNA using reverse transciptase and random hexamers. The priming process was triggered by random hexamers binding to the RNA. DNA extension was then produced by reverse transcriptase producing a complementary DNA strand to the RNA. An ABI prism 7700 q-PCR machine was used 20 °C, 10 minutes, 37 °C, 1 hour; 90 °C, 5 minutes.

2.3.7 qPCR

Semi-quantitative PCR was performed on an ABI prism 7700 q-PCR machine (Applied Biosystems, CA, USA).

A high magnesium concentration was used in order to generate large quantities of product, and a sequence specific probe was used. dUTP was used instead of dTTP. Uracil – n – gycosylase was added before commencing qPCR. This nicks at Uracil bases present, destroying any previous contamination. q-PCR was performed for 50 cycles: 55 °C, 10 minutes (first cycle only when Uracil – n – glycoslalase nicks any contaminating products), 95 °C, 5 minutes (1st cycle only, denaturing Uracil – n – glycoslalase), 60 °C, 1 minute; 95 °C, 1 minutes.

Analysis was performed using the delta delta Ct method as described elsewhere (Livak and Schmittgen 2001).

2.4 Preparation of competent cells and transformation

Competent cells were prepared in one of two ways.

2.4.1 Method 1:

A single colony from a freshly streaked LB agar plate of the E. coli strain to be made competent was used to inoculate 5 mL of sterile LB broth supplemented with 20 mM MgSO4 in a 20 mL universal bottle, and grown overnight. An approximate 1 in 100 dilution of this was made by adding 1 mL of this culture to inoculate 100 mL of LB broth in a 0.5 L conical flask, which had been pre-warmed to 37 °C. The culture was grown at 37 °C, in a rotary shaker, for approximately 2 hours monitoring the broth optical density until A600 of between 0.35 and 0.45 was reached, indicating that the cells were at the early log phase of their growth curve. The culture was incubated on ice for 10 min, before being transferred into chilled, sterile 50 mL Falcon centrifuge

tubes. The cells were subsequently harvested using a bench top centrifuge (2800 g, 10 min, 4°C), and resuspended in 10 mL ice cold 0.1 M CaCl2; 15 % (v/v) glycerol. After incubating for 30 min on ice, the cells were harvested by centrifugation as before, and resuspended in 1 mL of the same solution. Bacteria were assumed to be competent after incubating on ice for a further 30 min. At this point the cells were stored in aliquots of 100 mL at -70 °C for future use.

Transformations were performed by adding 100 μL of the competent cell mixture to appropriate amount of plasmid (~1 μg for moving plasmids between strains, and 100 ng for sub cloning) and incubating on ice for 45 min. Cells were then heat shocked for 90 seconds at 42 °C, and added to 5 mL sterile LB broth in a 20 mL universal, and grown for 1 hour at 37 °C. Transformants were selected by plating 100 μL of the recovered transformation culture onto appropriately selective nutrient agar plates. For more difficult cloning reactions, the remaining transformation culture was spun down, re-suspended in a volume of 100 μL and plated out as before.

2.4.2 Method 2:

This method produced extremely competent cells, however was much more labour intensive as it required the preparation of fresh cells each time. The method of producing competent cells was similar to method 1, but with three differences.. Firstly the starter culture was inoculated with 5 μL glycerol stock rather than using a colony from a fresh plate. Secondly TSB was used instead of 0.1 M $CaCl_2$; 15 % (v/v) glycerol. Thirdly heat shock was not required for transformation.

2.5 Fermentation

2.5.1 Fermentation glycerol stocks

Glycerol stocks for fermentation were prepared as follows:

A colony was picked off a freshly streaked plate and grown in 5 mL of medium overnight at 37°C on a shaker-incubator set at 230 rpm. The optical density at 600 nm of overnight culture was determined using a DU70 spectrophotometer (Beckman, High Wycombe, Bucks., UK). The volume required to produce an optical density at 600 nm of 0.02 OD units was then calculated and used to inoculate 50 mL of medium in a 500 mL shake flask. This 500 mL shake flask was incubated at 37 °C on a shaker incubator as before and the optical density at 600 nm (OD_{600}) was monitored,

using a DU70 spectrophotometer, until it reached 2.0 OD units. 50 mL of sterile glycerol was then added and the stock frozen at –70 °C in 1.8 mL aliquots.

2.5.2 Inoculum preparation

A 1.8 mL aliquot of fermentation glycerol stock was added to 50 mL of medium in 100 mL shake flasks. The inoculum culture was grown overnight at 37 °C on a shaker incubator set at 230 rpm for around 11 hours, generally reaching an OD_{600} of around 6.

2.5.3 Fermentation

Batch fermentations were performed at 37 ^{0}C in 1-litre-working-volume SGI bioreactors (Inceltech, Toulouse, France). The OD_{600} of inoculum was determined, and the volume required to produce an OD_{600} of 0.02 OD units was used to inoculate the fermentation vessels. Oxygen levels were monitored using an Ingold polarographic probe (Mettler-Toledo Ltd, Beaumont Leys, Leicester, U.K.) and maintained at 30% O_2 saturation, initially using cascade control by adjustment of stirrer speed and, after this reached 1000 rpm, by adjustment of the sparging rate. pH was maintained at 7.00 ± 0.1 by the automatic additions of 1M sodium hydroxide and 1M phosphoric acid.

Both optical density and Wet Cell Weight (WCW) were used to monitor growth. Optical densities of cultures were taken at 600nm using a DU70 spectrophotometer as before. For WCW determination, duplicate 1mL samples of culture were spun at 13,000 rpm for 10 min in a bench-top centrifuge in pre-weighed tubes. The supernatants were carefully removed by first pouring, and then using stretched cotton wool swabs to wick away any remaining liquid. The tubes were then re-weighed and the WCW of the cell pellet determined.

2.5.4 Samples for plasmid analysis

Using a correlation of 5 OD units at 600nm equating to 2.5×10^9 cells, the volume of culture required to produce samples containing this number of cells was estimated and five samples were taken at regular (~hourly) intervals. Cell pellets were obtained and frozen at –20 °C.

2.6 Electrophoresis

2.6.1 General electrophoresis to establish size of linear DNA species

Depending on the size of DNA fragment to be examined agarose concentration between 0.4 and 4 % (w/v) were used. For cloning purposes 1 x TAE buffer was used, otherwise 1 x TBE buffer. Between 0.4 and 4 g of agarose was added to 100 mL of the relevant buffer and dissolved by boiling. After cooling to around 60 °C agarose was poured into casting trays and wells formed using combs containing the appropriate number of teeth. Once set, the casting blocks were removed and the gel submerged in the appropriate buffer (1 x TAE or 1 x TBE). Electrophoresis was performed at suitable voltages and times (100V for 1-2 hours to 15V overnight). Gels were subsequently stained using Ethidium Bromide and irradiated using a short wave UV light transilluminator.

2.6.2 Gel electrophoresis for small (<150 bp) DNA species

Polyacrlyamide gel electrophoresis was used in order to visualise small DNA species. Pre cast 2 – 20 % gels (catalogue number 161-1237) were used (BioRad, Hercules, USA) with a BioRad Ready Gel Mini-PROTEAN 3 electrophoresis cell.
Gels were run at 120 V (~12 V cm^{-1}) for around 100 minutes. Gels were subsequently stained using Ethidium Bromide and irradiated using a short wave UV light transilluminator.

2.6.3 Plasmid topology assay to determine proportion of supercoiled, open circular and linear plasmids.

2.6.3.1 Staining with SYBR Gold

This method in essence is similar to densitometry methods using Ethidium Bromide as a dye (O'Kennedy *et* al 2003). A much more sensitive densitometry technique was used, pictures were generated containing several hundred thousand shades of grey, rather than the 256 shades of grey associated with methods published previously which require several scans for accurate analysis of each band. There are also a number of advantages in the use of the SYBR Gold stain over ethidium bromide:
Unlike ethdium bromide it does not bind different plasmid isoforms differently (Even-Chen and Barenholz, 2000)

It is considerably more sensitive exhibiting greater than 10,000 fold increase in fluorescence when binding nucleic acids rather than a corresponding 30 fold increase associated with ethidium bromide (Tuma *et al* 1999, Le Pecq 1966).
Safety data is not currently available about the SYBR gold stain and as an intercalating dye (Personal communication Dr Joanne Thorndike, Biowhittaker, Berkshire, UK) it must be treated with the same care as Ethidium bromide.

Clarified lysates (prepared as above) were purified by ethanol precipitation by adding 1/10 volume of chilled 3 M sodium acetate and 2 volumes of -20 °C 100 % ethanol (Sambrook et al., 1989).
Electrophoresis of approximately 500 ng and 50 ng (1 in 10 dilution) of plasmids was performed in 0.6 % (w/v) agarose, alongside 2, 5 and 10 fold dilutions of MassRuler DNA ladder, High Range (Fermentas, Maryland, US). This mass ladder consists of 9 different sized DNA fragments, from 1500 to 10,000bp, of concentration 1.6 to 10 ng mL-1. The gel was stained using SYBR Gold (Molecular Probes, Leiden, The Netherlands), for 20 minutes in the dark, and destained for 60 minutes in distilled water before image capture. The effect of staining time on the dynamic range of the assay was studied and this is outlined in Chapter 5.

Gels were scanned using the ProXpress Proteomic Imaging System (Perkin-Elmer, Wellesley, MA, USA). Excitation of 480 nm and Emission of 530 nm were used with an exposure time of 30 ms. Band quantification was performed using ImageQuant software (Amersham Biosciences, Amersham, UK). Band quantification was performed in a similar manner to that previously described (O'Kennedy *et* al 2003). The mass ladders used (10 µL Fermentas High Range Mass ladder diluted 2,5 and 10 fold respectively) produced known bands of DNA between 1.6 and 50 ng, which were used in order to calibrate the dynamic range of the assay.

2.6.3.2 Non-radioactive labelling using DIG labelling and detection

Before the SYBR Gold staining method and densitometry method was established (above) an accurate non-radioactive method of determining the percentage of different species was to use Digoxigenin (DIG) labelling. DNA probes are labelled with DIG-11-dUTP. These probes bind to target DNA on Southern blots. An anti-digoxigenin antibody • alkaline phosphatase conjugate is allowed to bind to the hybridised probe. A signal is then detected by adding chemiluminescent alkaline phosphatase substrates that can be detected using X-ray film.

Gel electrophoresis and transfer of DNA to membrane

DIG reagents were obtained from Boehringer Mannheim (now Roche Diagnostics). Plasmids were purified as described in section 2.3.1.3, and samples were run in 0.6% agarose gel for 20 to 24 hours at around 15 V cm^{-1}. Gels were then stained using ethidium bromide and the length and width of gel to be blotted was measured and marked. The gel was then soaked in 0.25 M HCl for 20 minutes at room temperature and then rinsed with distilled water. The gel was then denatured in 0.4M NaOH for 30 minutes at room temperature. A blotting platform was set up in a shallow plastic box and covered with one sheet of Whatmann 3MM paper of size 20 x 20 cm and covered with 0.4M NaOH until liquid was just below the level of the blotting plate. A piece of DIG membrane and 3 pieces of Whatmann 3MM paper the same size as the gel to be blotted were prepared. The membrane was soaked in distilled water until wet, soaked in 0.4 M NaOH and laid over the gel in the appropriate position. Three sheets of Whatmann 3MM paper wetted in 0.4M NaOH were laid over the membrane, and on top of these 5 sheets of blotting paper. A 1 Kg weight was then applied to the lid of the blotting platform and left overnight. The next day the membrane was crosslinked using a UV crosslinker. The membrane was rinsed in 2X SSC and dried on 3MM Whatmann paper and then hybridised with the DIG non-radioactive system.

DIG probe preparation

1 µL of pSVβ (approximately 20 ng) was added to 16 µL distilled water and denatured by heating to 100 °C for 10 minutes and crash cooled in an ice / ethanol bath. 4 µL DIG-High Prime was added, mixed briefly and incubated for 20 hours at 37 °C. The reaction was stopped by adding 2 µL 0.2M EDTA (pH 8.0) and heating to 65 °C for 10 minutes. 2.5 µL LiCl and 75 µL chilled (-20 °C) ethanol was added and left for 30 minutes before centrifuging for 15 minutes are 13,000 rpm. The supernatent was carefully removed and the pellet washed with 50 µL cold 70 % ethanol (v/v). The supernatant was removed. The pellet was dried under vacuum and dissolved in 50 µL TE buffer.

Hybridisation

The membrane was pre-wetted in 2 x SSC and placed in a glass roller bottle with an appropriate volume (5 mL) of pre warmed (37 °C) DIG Easy Hybridisation Buffer and incubated at 37 °C for 30 minutes. 5 µL of the probe prepared above was denatured by boiling for 5 minutes and then placed on ice for 5 minutes. 5 mL of fresh pre

warmed (37 °C) DIG Easy Hybridisation Buffer and filtered through a 0.22 mm filter. The pre-hybridisation buffer was removed from the roller bottle and replaced with fresh filtered hybridisation buffer containing the labelled probe and incubated at 37 °C overnight.

Washing and detection

Blots were washed for two lots of 20 minutes in 2 x SSC + 0.1 % SDS at room temperature with agitation, and again for two lots of 20 minutes in 2 x SSC + 0.1 % SDS at 65 °C. After equilibrating the blot in washing buffer for one minute the membrane was transferred to 1 x freshly prepared Blocking Buffer for 30 minutes with gentle agitation. The membrane was then put into a solution containing a 1:10,000 dilution of Anti-digoxigenin AP conjugate in Blocking Buffer and agitated very gently for at least 30 minutes. The blot was washed with two lots of Washing Buffer for 15 – 60 minutes, before equilibrating for at least 5 minutes in a small volume of 1 x Detection Buffer with agitation. The membrane was placed in a hybridization bag sealed on three sides and a 1:100 dilution of CDP STAR in 1 x Detection buffer was added to the membrane before sealing the fourth side of the bag and incubating for 5 minutes. The membrane was removed from the bag, placed in a fresh bag and sealed. The blot was then exposed to X-ray film for sufficient time to obtain good exposures, and the X-rays were further analysed by densitometry.

2.6.4 Protein separation in SDS-polyacrylamide gel

Vertical gel electrophoresis was performed using a Sigma-Aldrich unit (catalogue number Z33,957-1). Gel plates cleaned with warm water and polished with ethanol were assembled with 1.0 mm spacers into the gel-casting base supplied with the unit. A 15 % (w/v) resolving gel was made by mixing a 30 % (w/v) acrylamide stock with resolving gel buffer. Gel polymerisation was induced by the addition of 200 mL freshly prepared 10 % (w/v) ammonium persulphate, and 20 mL TEMED, per 10 mL gel solution, approximately 6 mL of the gel solution was carefully pipetted into the casting apparatus. A thin layer of water was carefully overlaid to avoid the formation of a meniscus, and to aid polymerisation by sealing the gel from the atmosphere, and the gel was left to polymerise for 1 hour.

A 5 % (w/v) stacking gel solution was prepared by mixing 1 mL 30 % (w/v) acrylamide stock with 0.5 mL stacking gel buffer and 4.5 mL sterile distilled water. After the addition of 100 mL 10 % (w/v) ammonium persulphate, and 20 mL TEMED, 2 mL of stacking gel solution was pipetted onto the resolving gel (after decanting the

water layer). Wells were formed by the insertion of a 1.0 mm comb, and the gel left to polymerise for no more than 1.5 hours, to avoid diffusion across the pH gradient of the stacking and resolving gels. After assembling the gel in the running tank, the comb was carefully removed, before submerging the exposed ends of the gel in 1X Tris-glycine running buffer. Traces of acrylamide were removed from the wells, and it was ensured that no air bubbles would obscure the flow of electrical current through the system.

Electrophoresis was performed at 150 V; 50 mA, until the tracking dye had eluted from the end of the gel.

Following electrophoresis, protein bands were visualised by staining with Coomassie Brilliant Blue stain solution for 2 hours, and subsequently destained with destain solution, until the background intensity was suitably reduced relative to the protein species.

2.7 Yield determination

2.7.1 Spectrophotometric measurement

Plasmids prepared using Qiagen columns as described above and their absorbance at 260 nm (A_{260}) measured using a Heλios Spec (Thermo Spectronic, Madison, USA). A concentration of 50μg mL^{-1} ds DNA gives an A_{260} of 1. Purity of samples was checked by the ratio of absorbance at 260 and 280 nm.

2.7.2 Fluorescence-based method

PicoGreen (Molecular Probes, Leiden, The Netherlands) has been shown to be a sensitive dye that can be used to detect double stranded DNA in the presence of single stranded DNA, RNA and other compounds commonly found to contaminate nucleic acid preparations (Noites *et al* 1999).

2.7.2.1 Manual method

Samples were manually prepared to the clarified lysate step, as described above and were transferred to fresh benchtop centrifuge tubes. Standards were produced using Lamdba DNA as shown in the table below:

Cuvette	Total DNA (ng)	Volume (µL) of Lambda DNA	Volume (µL) of 1x TE	Volume (mL) of working concentration PicoGreen Reagent
1	2000	1000	0	1
2	200	100	900	1
3	10	10	990	1
4	1	1	999	1
5	Blank	0	1000	1

Table 2.1 Standard preparation for manual PicoGreen assay

The contents of each cuvette were mixed using a disposable Pasteur pipette before being incubated for 5 minutes in the dark before reading. Readings were taken using a VersaFlour fluorometer (BioRad) using Excitation filter EX 480/20 (470-490 nm) (BioRad) and Emission filter EM 520/10 (505-515 nm)

5 µL of each clarified lysate sample was added to 995 µL 1 X TE with 1 mL of diluted PicoGreen reagent and fluorescence measured as before. Two reference standards were also used to check the assay validity.

2.7.2.2 Automated method

Samples were manually prepared to the clarified lysate step, as described above, while the necessary dilutions, standard curve and addition of PicoGreen reagent were all performed by a Tecan RSP150 Robotic Sample Processor (Tecan UK Ltd, Reading, UK).

Standard curves were produced in duplicate from a reference plasmid of known concentration, 4 – 800 ng mL^{-1}, and also duplicate Lambda DNA samples at 250 ng mL^{-1}, were included as a control. Clarified lysate samples were diluted 200 fold. Duplicate 100µL aliquots of the standards and samples were mixed with 100 µL working concentration PicoGreen (Molecular Probes). After 5 minutes incubation in the dark, plates were read using a SpectraMax Fluorescent plate reader (Molecular Devices, Wokingham, UK) with excitation and emission filters of 485nm and 539nm respectively.

Chapter 3:

Construction and characterisation of a suitable plasmid backbone

3 Construction and characterisation of a suitable plasmid backbone

In order to determine the effects of sequences exhibiting intrinsic structural abnormalities, it was decided to produce a family of plasmids, differing only at a defined point, where the sequence of interest would be inserted. In order to do this a well characterized parental plasmid was produced. The aim at the outset was to produce a plasmid as similar as possible to those used by GSK in their production process.

3.1 Perceived industrial requirements

3.1.1 Antibiotic resistance / selectable marker

Without positive selection for cells containing plasmids, it is likely that a culture containing a high copy number plasmid (with no special partitioning system) will generate plasmid free cells (Summers 1991). Since plasmids produce a metabolic burden on the cell, cells containing fewer plasmids proliferate more quickly than those containing more plasmids. This leads to selective drift as the fast growing fermentation environment favours cells containing fewer (or no) plasmids.

As a result of this, some form of selection pressure must be applied throughout the fermentation in order to ensure plasmids are harboured in all cells and hence ensure high yield. A commonly used antibiotic in biochemistry is ampicillin. However this is an unsuitable choice for use to generate plasmids for biotherapeutics for two reasons.

The mechanism of antibiotic resistance to ampicillin can lead to plasmid free cells, particularly in high-density cultures. The ampicillin resistance gene produces an enzyme that is secreted into the periplasmic space. It hydrolyses the β-lactam ring of the ampicillin destroying its activity (Sykes and Mathew 1976). This means that during a fermentation ampicillin levels in the broth will become depleted. Cells with fewer plasmids may therefore be able to survive and have an advantage over cells with high copy numbers of the plasmid. This can lead to the selective drift, lowering copy numbers and possibly leading to plasmid free cells.

Ampicillin is widely used in medicine. Use of antibiotics in non-medical applications has lead to "new" resistance genes and multi-resistant pathogens with increased pathogenicity in food animals (Young 1994). In addition some patients have been shown to exhibit hypersensitivity to ß-lactams and residuals that may be found in products. For these reasons and in order to try and reduce this spread of antibiotic resistance, the view of the World Health Organisation is that the non-essential use of antibiotics is to be avoided outside of the medical industry.

Kanamycin is the most favoured antibiotic used in the production of biopharmaceuticals. Kanamycin is not widely used in medicine as it has been found to produce unwanted side effects including ototoxicity (Humes 1984) and foetal damage. Unlike ampicillin, the kanamycin antibiotic resistance mechanism does not reduce the levels of antibiotic (Davies and Smith 1978).

Some novel alternative methods of maintaining plasmids in cells are available, but these are subject to licence fees for commercial explotation. The operator/repression titration system removes all need for antibiotics, without harbouring plasmid the host cells are unable to proliferate (Cranenburgh *et al* 2001). The FDA stipulates that antibiotic remaining in plasmid product must be at a very low level, and consideration should be made to the lowest level that would generate an unintentional clinical effect. This means that substantial care and monitoring of processing steps is required when antibiotic selection markers are used. The novel system described by Cranenburgh has obvious advantage over more conventional production systems.

3.1.2 *cer* sequence

The *cer* sequence is included in several of GSK's vectors. *cer* increases plasmid segregational stability and hence reduces the likelihood of generation of plasmid free cells in production.

3.1.3 Size

It was decided that a realistic size for future plasmid products would be up to 15 Kb. Though plasmid products undergoing clinical trials in 2000 were smaller than this, it was expected that, in the future, the size of plasmids would increase (Levy *et al* 2000). Plasmid products may encode more than one gene sequence. For example, in addition to the therapeutic or prophylactic molecule, genes may be included in plasmids coding for signalling molecules, like cytokines, in order to increase DNA vaccine effectiveness (Cohen *et al* 1998).

3.2 Starting point [pSVβ].

Biochemical Engineering studies at UCL concerned with plasmid DNA manufacture have mainly been concerned with the high copy number (pUC based) plasmid [pSVβ], which is 6.894 Kb (Fig 3.1). This plasmid is carries an ampicillin resistance gene.

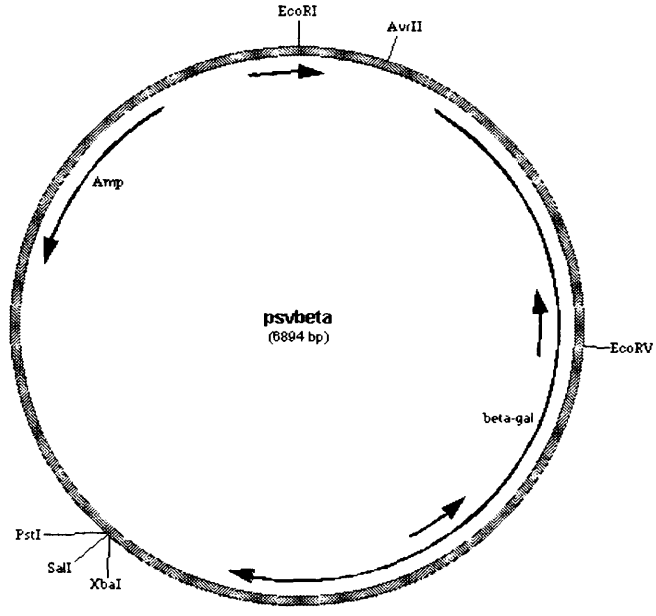


Figure 3.1 Plasmid [pSVβ] (Promega Corp., Madison, WI) comprises an ampicillin resistance gene, a β-gal gene and a pUC origin of replication.

3.3 Kanamycin resistant plasmid

A kanamycin resistance gene, originally from transposon TN903 (Oka A *et al* 1981), was taken from plasmid pUC4K (Viera J and Messing J 1982) via restriction with *PstI*, and was ligated into the *PstI* site of pSVβ. This produced plasmids containing the kanamycin resistance gene in two different orientations, [pQR235] and [pQR236]. Shake flask growth studies showed that there were no differences between these plasmids in terms of yield and host cell growth rates (Fig 3.2). It was decided to pursue using only one, [pQR235].

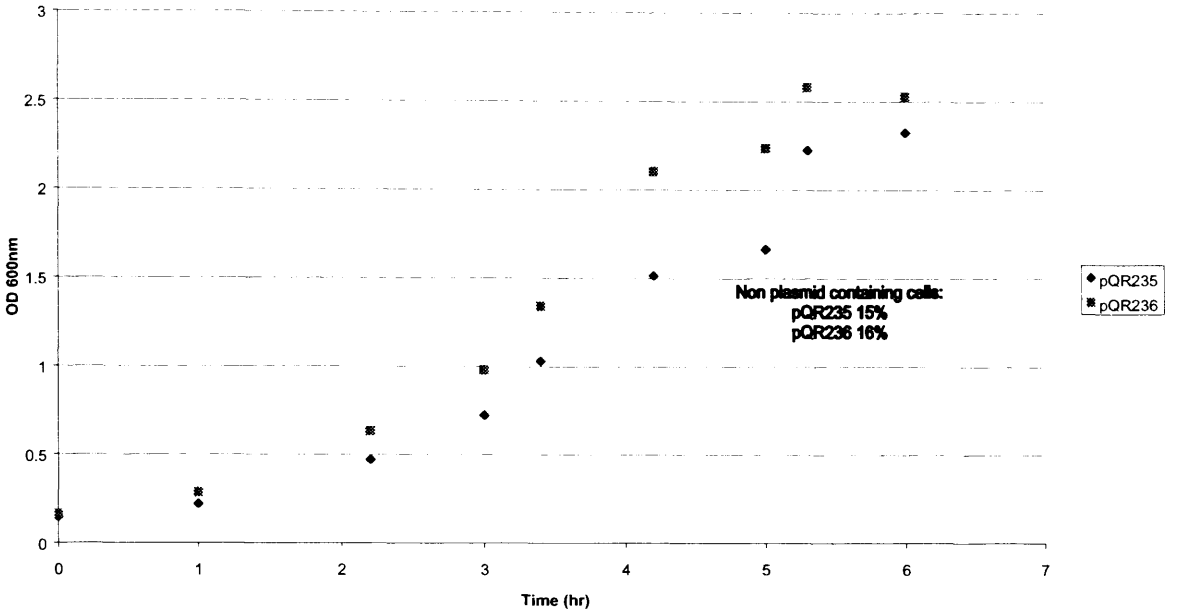


Figure 3.2 Graph comparing growth, in shake flasks, of DH5α cells containing plasmid with the Kanamycin gene in one of two orientations.

3.4 Cer sequence insertion

The *cer* fragment, originally a 279 bp *Hpa*II-*Taq*I fragment from ColE1, was amplified from [pKS492], a derivative of pUC9. Amplification was performed by polymerase chain reaction (PCR) using primers designed to incorporate the *Sal*GI restriction site and to amplify the 278 bp of the ColE1 fragment known to be required for *cer* action. [pQR235] was cut using SalGI, dephosphorylated, and ligated with the PCR product, also cut with *Sal*GI. The presence of the fragment was confirmed by restriction analysis. Growth curves of [pQR235] and [pQR332] were obtained and numbers of plasmid free cells determined after a number of non-selective growth cycles (Fig 3.3).

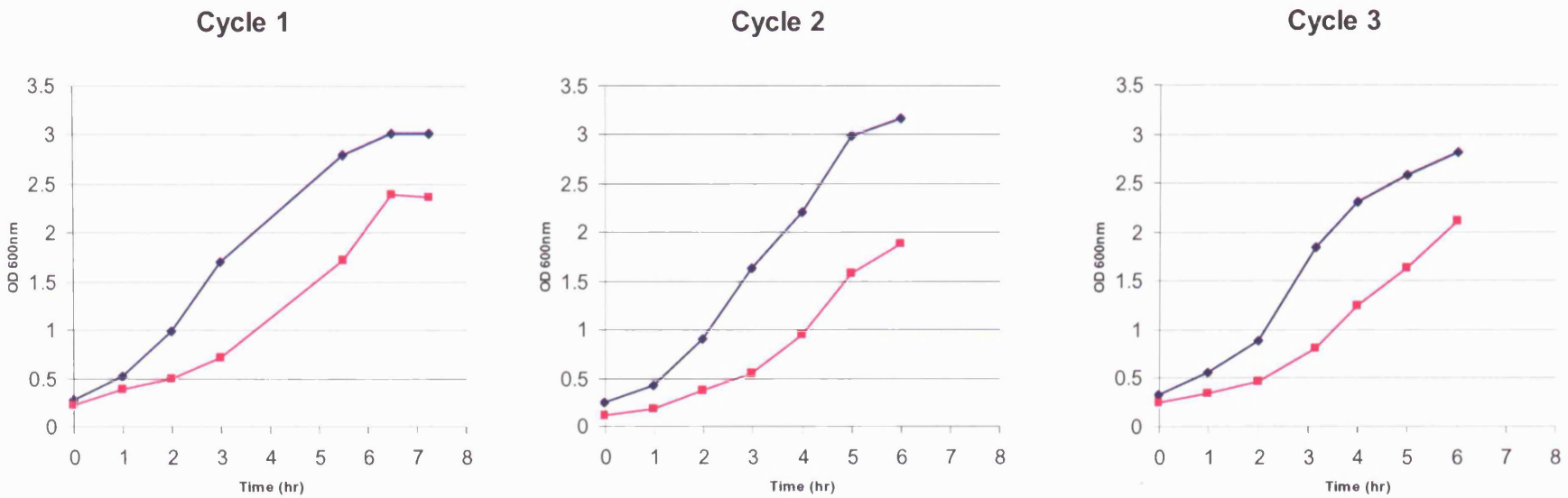


Figure 3.3 Growth comparison of DH1 containing [pQR235] (blue) and [pQR332] (pink) in three successive growth cycles.

3.5 Plasmid retention

Well-defined colonies of DH5α were picked from freshly streaked nutrient agar plates containing kanamycin and used to inoculate 5 mL of nutrient broth containing kanamycin, in sterile 20 ml glass universal bottles. These cultures were grown overnight at 37 °C on a lateral shaker. These flasks were used to inoculate 50 mL of nutrient broth in a 0.5 L conical flask, which was then incubated at 37 °C in an orbital shaker (200 rpm) for several hours, the optical density (OD) being recorded at intervals.

For the comparison between [pQR235] and [pQR236] kanamycin was added to the broth. In the case of comparisons between [pQR235] and [pQR332] no kanamycin was added. Once the cultures reached stationary phase (OD ~ 7.0 A600nm) 5 mL was taken and added to another 50 mL of nutrient broth in a 0.5 L conical flask, which was treated as before. This was repeated so as three non-selective growth cycles occurred.

Samples at the end of each non-selective growth cycle were examined to determine percentages of plasmid free cells. Serial dilutions were made and plated out on non-selective and selective nutrient agar plates.

Cycle	1 (±15%)	2 (±15%)	3 (±15%)
[pQR235] (without cer)	85	81	72
[pQR332] (with cer)	100	100	100

Table 3.1 Percentage of plasmid containing cells after unselective growth

The *cer* sequence can be seen to have increased the inheritance stability of the plasmid (Fig 3.4). Summers, credited with the discovery of the *cer* sequence, in 1996, demonstrated how the sequence increased Segregational stability of plasmids by using a strain that produces a large number of oligomers.

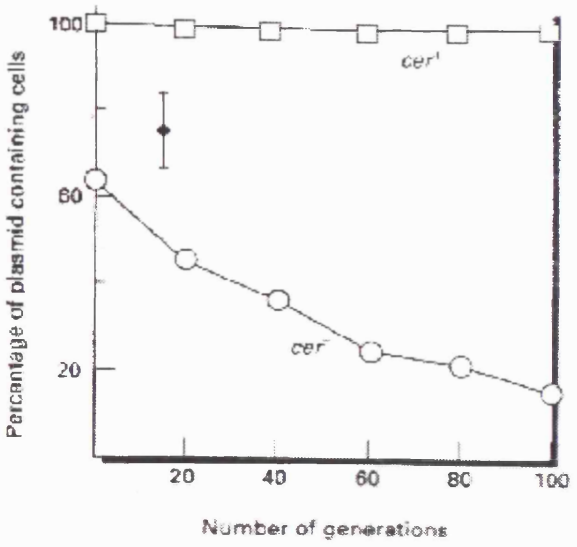


Figure 3.4 Inheritance stability of ColE1 derivatives with and without a *cer* site, in an *sbc*A strain that produces plasmid oligomers. The *cer*- derivative is so unstable that nearly 40 % of cells lost the plasmid during growth of the culture, which was used to start the experiment (Figure adapted from 'The Biology of Plasmids' Summers DK 1996). The diamond (♦) marks the position of *cer* ⁻ plasmid after approximately 14 generations in DH1.

In order to compare the loss of *cer* ⁻ plasmid from DH1 and the *sbc*A strain used by Summers, the total number of generations produced in the three successive growth flask experiments was determined.

Estimation of cell generation is not accurate as neither the work of Summers nor that work presented here was from cultures in steady state.

Since $x = x_o e^{\mu t}$ (Where x = cell number)

When the number of cells doubles at t_d (doubling time), the cell concentration will be $2x_0$.

Hence $2x_0 = e^{\mu td}$

$x_0 = e^{\mu td}$

$t_d = \ln 2 / \mu$

By assuming a linear relationship between biomass and the optical density readings at 600nm a μmax for DH1 [pQR332] was found at approximately 0.44, hence Td = 1.45 hours. Hence after 20 hours, there would have been 14 generations. After approximately

14 generations 72 % of cells were found to contain the *cer* plasmid. For interest, this point has been marked on the figure 3.6 with a diamond (♦).

3.6 Increasing plasmid size

In order to increase the size of the plasmid it was decided to add a piece of DNA of around 10kB. The selection criteria for this piece of DNA were that it should not be expressed in *E.coli* and that the DNA should be sequenced. In addition it is desirable that the inserted DNA should not contain regions of DNA with inherent structural abnormalities, such as those described in Chapter 1.

Obtaining such a piece of DNA ought to have presented little problem as artificial chromosome libraries that are fully sequenced are easily available. However ϕC31 was chosen as the source as it is has been fully sequenced and it was thought that Streptomyces promoters are not recognised by *E.coli* RNA polymerase (Smith *et al* 1999).

Even if the sequence from ϕC31 contained any unusual structural features not picked up in the screening of the sequence, comparisons between a family of plasmids known to include DNA abnormalities at a defined point would still have been possible.

Producing large plasmids can be difficult due to a much lower rate of transformation (Suzuki *et al* 1997).

Analysis of ϕC31 revealed several suitable *Avr*II fragments. The choice of *Avr*II was made in an attempt to optimise the usefulness of restriction sites. By insertion of one of the screened *Avr*II fragments into [pQR332] itself cut with *Xba*I, a further *Avr*II fragment could have been added later using the *Avr*II site present in [pQR332].

The 7836 bp *Avr*II fragment of ϕC31 (Smith *et al* 1999) was isolated by gel electrophoresis, and ligated with [pQR332], cut using *Xba*I. Several strains harbouring the plasmid grew well, in shake flasks (data not shown).

The *Avr*II fragments of ϕC31 did not resolve well by conventional agarose gel electrophoresis. Since there was a limited supply of the ϕC31 DNA the fragments were inserted into [pQR332] without knowledge of precisely which fragments had been inserted. Only one transformant was produced. Using a variety of restriction enzymes a restriction map was produced without aid of plasmid manipulation software, showing that

the 7836bp *Avr*II fragment of ϕC31 had been incorporated and its orientation was elucidated.

Many attempts were made to add another ϕC31 *Avr*II fragment to [pQR333] and also to add larger ϕC31 fragments cut with a range of enzymes to [pQR332], however all were unsuccessful.

[pQR333] was the subject of experiment at a UCL pilot plant experiment using both a novel host (Cooke *et al* 2001) and novel media (A Kay PhD Thesis, Department of Biochemical Engineering, UCL, 2004). Very low yields were obtained in a 70L fermentation (data not shown). At this early stage it was not possible to determine if this was due to the plasmid itself, novel strain, novel media or fermentation conditions. It was hence decided to carry on using the plasmid, check that no significant expression of ϕC31 proteins was occurring and then to determine its fermentation characteristics.

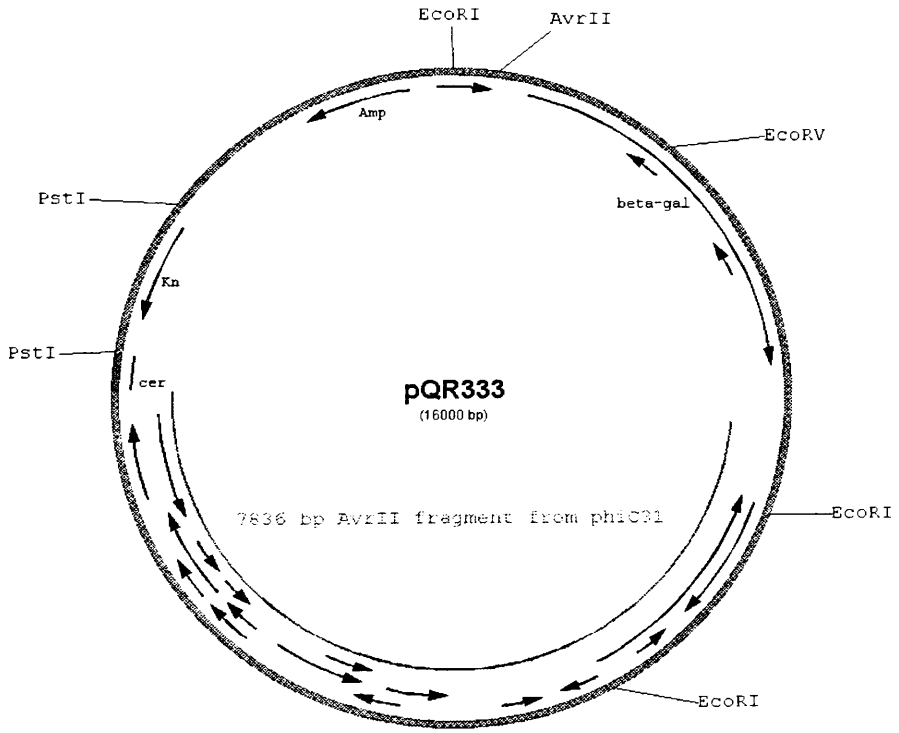


Figure 3.5 [pQR333]

3.7 Multiple cloning site

A multiple cloning site (Fig 3.6) was designed and added to [pQR333]. The design of the multiple cloning site allowed a check to be made that only one copy of the site was incorporated.

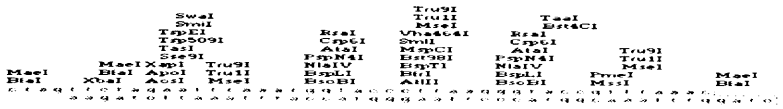


Figure 3.6 Multiple cloning site

Two oligonucleotides were produced and diluted to 20 μM solutions. The two solutions were mixed and heated to 75 °C and allowed to cool slowly. This produced a double stranded section of DNA containing several restriction sites and sticky CTAG ends, this comprised the multiple cloning site.

[pQR333] was cut with *Avr*II and dephosphorylated. This was ligated with the multiple cloning site in the presence of *Spe*I. The *Spe*I was added in order to prevent multiple cloning sites from annealing.

Presence of the multiple cloning site was confirmed by comparing electrophoresis gels of [pQR333] and [pQR334] cut with enzymes contained in the multiple cloning site. Additionally it was checked that [pQR334] could not be cut by *Spe*I.

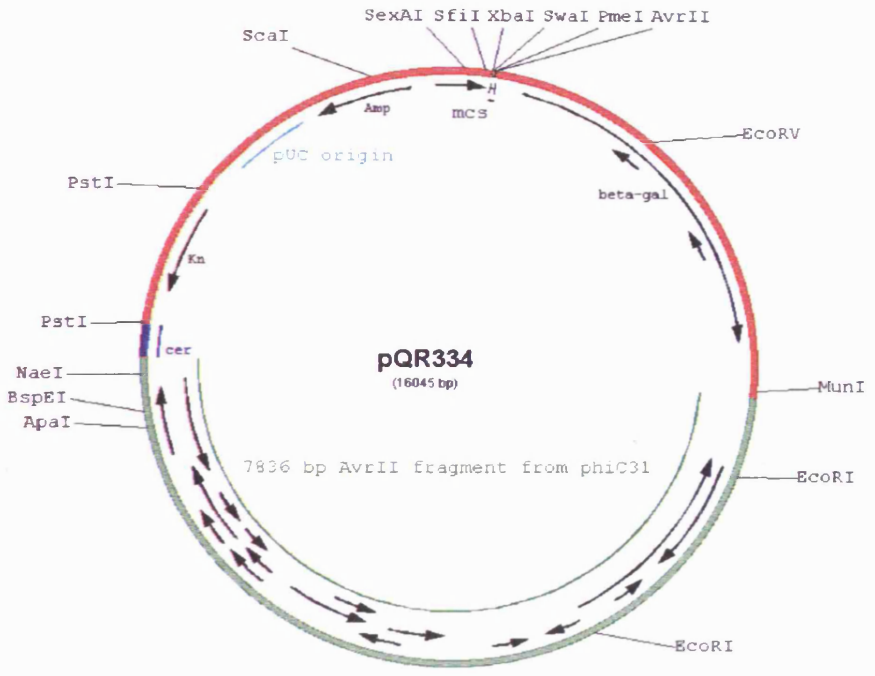


Figure 3.7 [pQR334]

3.8 Results from fermentations.

A summary of the fermentation data obtained with DH1 housing different plasmids is given below.

3.8.1 Growth rates

Construct	Antibiotic present	Initial growth rate (hr^{-1})	Late phase growth rate (hr^{-1})	OD 600nm at 9 hours	Wet cell weight at 9 hours ($g\ L^{-1}$)
[pSVβ]	-	0.661, 0.686	0.249, 0.332	11.6, 16	43, 80 *1
[pSVβ] *2	Ampicillin	-0.028, 0.249	0.531, 0.422	7, 12	6, 12
[pQR235]	-	0.780, 0.784	0.355, 0.420	17.2, 18.2	74.9, 67.2
[pQR235]	Kanamycin	0.417, 0.392	0.240, 0.249	8.6, 10.0	26.9, 29.1
[pQR332] *3	Kanamycin	0.728, 0.725 0.462, 0.391	0.149, 0.227 0.283, 0.283	8.0, 6.0 10.3, 11.6	49.4, 32.6 34.2, 37.8
[pQR334]	Kanamycin	0.536, 0.399 0.462, 0.438	0.158, 0.153 0.243, 0.248	4.5, 3.5 5.9, 6.7	24.0, 21.8 19.9, 19.9

Table 3.2 Summary of growth rate data from batch fermentations in DH1.

***1 contained almost exclusively plasmid free cells.**

***2 Strange lag for [pSVβ] with ampicillin fermentations. They went on to much higher optical densities and wet cell weights.**

***3 The two sets of [pQR332] + Kanamycin give different initial growth rates. There were some problems with the dissolved oxygen tension probes at the beginning of this fermentation.**

It was found that the inserted ϕC31 DNA caused a 35% reduction in growth rate based on absorbance at 600nm.

3.8.2 Copy numbers

The pico-green assay and phenol-chloroform extraction followed by absorbance at 260nm has been performed on some of the fermentation samples. There was a great deal of fluctuations in the results of these, but copy numbers were in the order of 1000 from the pico-green assay and higher from the absorbance assay.

No significant differences in the average copy numbers of plasmids was found.

3.8.3 Protein Expression

It was thought that the reason for the slower growth of the plasmid containing DNA from ϕC31 might be that genes from the *Streptomyces* phage were expressed during bacterial fermentations. As mentioned earlier, one of the reasons for the selection of ϕC31 as a source of DNA to increase the size of the plasmid was that the promoters from *Streptomyces* ought not be expressed in *E.coli.* However low levels of expression of a few *Streptomyces* promoters in *E.coli* has been observed previously. (Deng *et al* 1986).

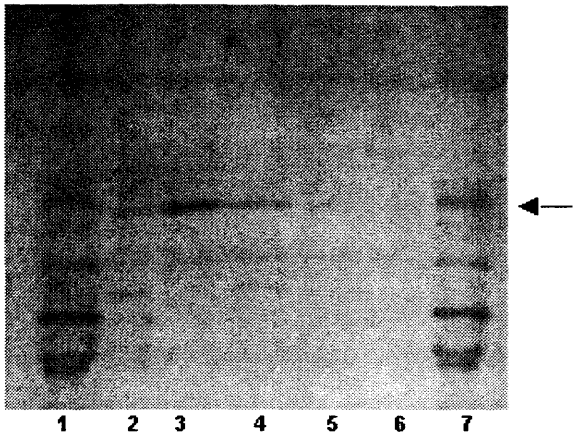


Figure 3.8 Protein expression gel. Lane 1 contained protein standard, lane 2 pQR334, lane 3 pQR332, lane 4 pQR235, lane 5 pSVβ, lane 6 DH1 and lane 7 protein standard. Arrow indicates position of Kanamycin reistance protein. No significant differences other than the Kanamycin resistance gene protein in lanes 2, 3 and 4 can be seen.

No large amounts of ϕC31 genes were seen to be expressed from the above gel. While it is possible that low levels of expression occur it is unlikely that these would add significant metabolic burden.

3.8.4 Plasmid topology

Plasmids were isolated using small-scale commercially available chromatographic spin prep columns (Qiagen) from samples taken at the beginning and end of fermentations. The topology of these plasmids was examined using Southern blotting and labelling with a DIG probe.

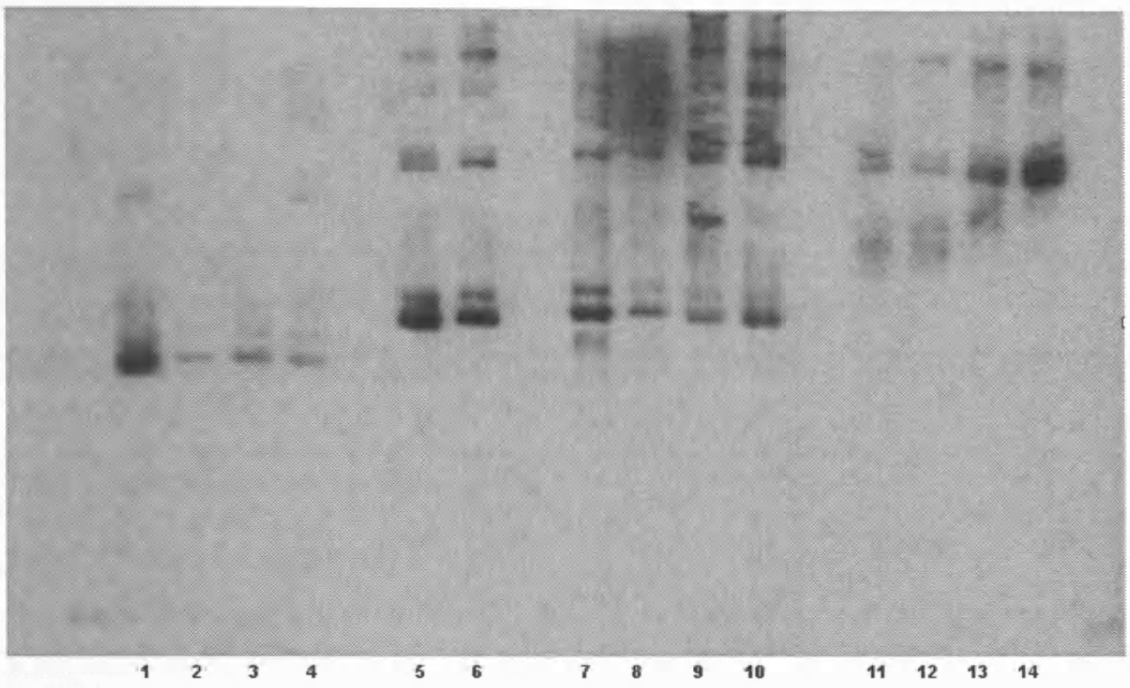


Figure 3.9 Picture of topology assay, labelled with [pSVβ] DIG probe, lanes 1 and 2 contained [pSVβ] from inoculums, lanes 3 and 4 contained [pSVβ] obtained at the end of fermentation, lane 5 contained [pQR235] from innoculum, lane 6 contained [pQR235] obtained at the end of fermentation, lanes 7 and 8 contained [pQR332] from inoculums, lanes 9 and 10 contained [pQR332] obtained at the end of fermentation, lanes 11 and 12 contained [pQR334] from inoculums, lanes 13 and 14 contained [pQR334] obtained at the end of fermentation.

Construct	% Supercoiled and band above supercoiled at beginning of fermentation excluding concatemers	% Supercoiled and band above supercoiled after 8 hrs of fermentation excluding concatemers
pSVβ	88.5	84.7
pQR235	60.5	77.7
pQR332	78.0	74.1
pQR334	37.9	7.4

Table 3.3 DIG based topology assay results

3.8.5 Choice of parent plasmid

The difficulties experienced in cloning the 16 Kb plasmids, the low transformation efficiencies encountered in comparison to the smaller plasmids and problems associated with larger plasmids reported in literature (Levy *et al* 2000, Griffiths *et al* 2000, Siguret *et al* 1994) indicated that directly cloning into [pQR334] might be problematic.

It was therefore decided to capture the unusual structures in more suitable smaller cloning vectors before subcloning into [pQR334]. It was thought that cloning of unusual structures would be easier performed using a small plasmid parent. The fermentation characteristics of cells containing these plasmids and characterisation of the plasmids themselves could be compared to the parent plasmid. After cloning the sequences into [pQR334] further characterisation could take place in order to determine if effects observed in the smaller plasmids would occur in larger plasmids, to a greater or lesser extent.

All plasmids studied generated reasonably high yields, but cells containing plasmids with the Kanamycin resistance gene grew slowly. It seemed therefore that either [pSVβ] or [pQR332] would be reasonable choices to use as the small parental plasmid. However neither of these plasmids are designed for cloning and in particular [pQR332] contained limited cloning sites.

[pBS840], known to contain an unusual DNA structure and described in more detail in the next chapter, was obtained from another research group. [pBS840] was based on

[pBluescript KS+] a cloning vector with considerable advantages, in terms of multiple cloning sites and blue white selection, over both [pSVβ] and [pQR332]. It was already known that cells containing [pBluescript II KS +] (differing from [pBluescript KS +] only by the addition of two *BssHII* restriction sites), grow rapidly producing good plasmid yields. It was therefore decided to pursue cloning unusual sequences into [pBluescript II KS +]. Given sufficient time sequences would then be cloned into the 16 Kb [pQR334] to see if they had any effect in the production and characteristics of a larger plasmid.

Chapter 4:

Construction of plasmids with intrinsic structural features

4 Construction of plasmids with intrinsic structural features

4.1 Introduction

It was intended to find sequences already shown to exhibit unusual DNA structure and to clone them at a defined point into a well-characterised vector. A number of suitable target sequences were found.

It was decided that these should be amplified by PCR together with their surrounding sequences, since sequence context may influence the formation of secondary structures (Dlakic and Harrington 1995, Canella and Seidman 2000, Vallone and Benight 2000, Movileanu *et al* 2002). However; a flaw in this philosophy was highlighted when PCR failed. When performing PCR it is generally recommended to avoid regions containing unusual structures (Dieffenbach *et al* 1995). Consequently, there was limited success in the cloning of unusual structures in this way.

An alternative strategy, which was used in order to clone a quadruplex structure, was to use two long complementary oligonucleotides with overlapping sticky ends. This method could have been enhanced for longer sequences by using a number of oligonucleotides that overlapped with one another.

The generous gift of a triplex containing plasmid [pBS840] from Dr Z-Y Wang (Washington University Medical Center and the Jewish Hospital, St Louis, Missouri) was gratefully received, as attempts to PCR this sequence resulted only in the amplification of surrounding sequences.

4.1.1.1 A sequence without any unusual structural features

Any differences exhibited by the family of plasmids containing unusual structural features compared to the parent plasmid may be accounted for by two different explanations:

- the unusual intrinsic structures themselves
- the additional size of inserts

In order to demonstrate that the effects are caused by plasmid structure, a sequence of DNA was selected to add to the parent plasmid [pBluescript II KS+] in order to produce a plasmid of more similar size to those with interesting inserts.

There were a number of criteria for the selection of this piece of DNA:

- known sequence
- size (between 550 and 900 bp) and suitability for cloning
- no known promoter regions included
- no repeats
- no propensity to form A or Z DNA
- small amount of intrinsic bending

Lambda DNA is a good source of many different sized fragments as it is large (49 kb), cut by many enzymes and so a fragment of the desired size with restriction sites appropriate to the parent vector can be obtained. As expected a BLAST search comparing Lambda and [pBluescript II KS+] showed no significant similarities between the two sequences.

Using the plasmid manipulation package Clone Manager, suitable sized fragments without start codons nor untranslated 5' regions were identified (Table 4.1).

Single digests

Bss HII	604
Eco RV	618, 588
Hinc II	728, 570
Hind III	564

Double digests

	Bam HI	*Cla* I	*Eco* RI	*Eco*RV	*Hind* III
Bam HI	*	*	*	*	*
Cla I	N/A	*	*	*	*
Eco RI	N/A	N/A	*	*	*
Eco RV	604	N/A	719	*	*
Hind III	784	N/A	N/A	721	*
Pst I	N/A	N/A	828	N/A	705
Sac I	N/A	736	N/A	942	N/A

Table 4.1 Fragments of Lambda DNA obtained by digestion with the indicated enzymes.

The 721bp fragment *Eco* RV - *Hind* III, the 705bp fragment *Hind* III – *Pst* I, and the 604 *Bam* HI – *EcoRV*, all contain the end of coding sequences and hence may contain unusual 5' sequences.

The two *BssHII* sites in [pBluescript II KS+] mean that addition of the 604bp *BssHII* fragment, would only increase plasmid size by 431bp.

942	*Eco* RV – *Sac* I	contains intrinsic bends
828	*Eco* RI – *Pst* I	contains intrinsic bends
784	*Bam* HI – *Hind* III	1 direct repeat (15 bp one mismatch)
		+ GTATTGGTTTATTtG (+) 285-299
		+ GTATTGGTTTATTgG (+) 555-570
736	*Cla* I – *Sac* I	contains intrinsic bends
728	*HincII*	contains intrinsic bends

721	*EcoRV – HindIII*	contains end of a coding sequence
719	*EcoRI – EcoRV*	contains intrinsic bends
705	*HindIII – PstI*	contains end of a coding sequence
618	*EcoRV*	no obvious bends or repeats
604	*BssHII*	sites in bluescript mean plasmid would only increase by 431bp

After a number of unsuccessful attempts touchdown PCR, use of betaine, careful primer design, altering the concentrations of reagents and using different templates achieved some success.

Primers redesign involved locating non-repetitive regions on either side of the region of interest and also ensuring that pyrimidines rather than purines were included at the 3' ends of the primers. Primer design was performed at GSK Beckenham using the Primer Express package (Applied Biosystems, CA, USA) under the guidance of Mr Chaminda Salgado. Primer design heuristics, as outlined in literature (e.g. Dieffenbach *et al* 1995) were followed, with the obvious exclusion that the amplification targets contained secondary structure. Primers were designed to achieve optimum amplicon melting temperature, not to form strong secondary structures or primer-dimers themselves.

Suitable primers could not be designed for the PDGF intramolecular triplex as the amplicon melting point for specific primers was around 88°C. Z-Y Wang's research group, who had previously cloned the triplex, generously donated the plasmid they created containing a DNA triplex. The sequence was originally obtained from the human promoter of the platelet-derived growth factor A-chain gene and cloning described by the group (Wang *et al* 1992).

The very high percentage of GC bases, linked by three rather than two hydrogen bonds, of the pBS840 insert is probably responsible for such a high amplicon melting temperature.

DNA sequence	*Base Pairs*	*% GC*
618 EcoRV	(618)	52.75
PBluescript II KS+	(2961)	50.42
CG(6) Insert	(395)	41.85
Cdc2 Insert	(745)	40.76
pBS840 Insert	(818)	80.82

Table 4.2 Comparison of GC content of inserted sequences

4.2 Primer design

The primers designed for the insulin-linked polymorphic region were:
GATCTAGATCTTGGGCCATCCGGGACT and
GATCTAGACTTCTGATGCAGCCTGTCCTGGA

For the *cdc2* promoter:
GATCTAGAGCGCAACATAATGAGACCCA and
GATCTAGAAGTTTCAAACTCACCGCGCT

For the intron from the human coagulation factor IX gene:
GATCTAGACACTGTCGTATAATGTGGTCCATCA and
GATCTAGATCAGATCAACAGCACCTTTGGTT

4.2.1 Cloning the human coagulation factor IX gene

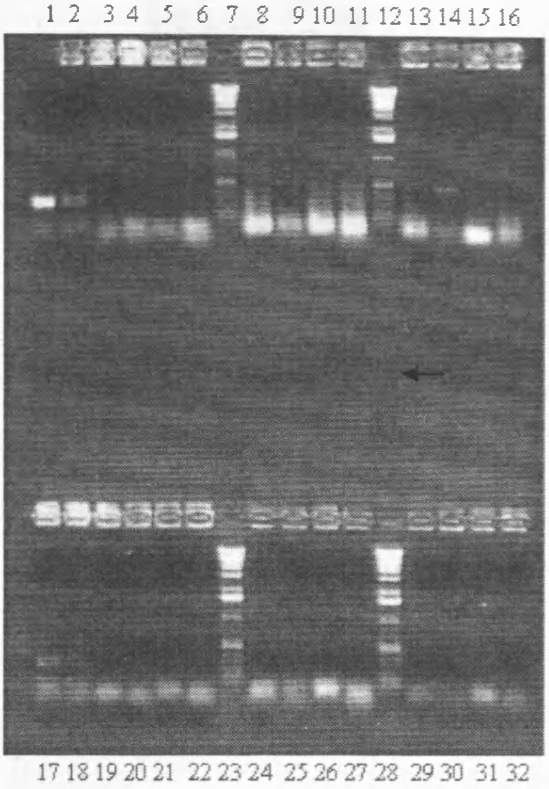


Figure 4.1 PCR of unusual DNA sequences. Lanes 1, 2, 17 and 18 β-actin controls (300bp) using buccal template (lanes 2, 17 and 18) and commercially prepared human genomic DNA (lane 1). Lanes 7, 12, 23 and 28 contained 1KB plus ladder. Gel shows the results of PCRs under six different sets of conditions of the four targets. Lane 14 shows the only successful PCR reaction and another 23 failed PCR reactions can be seen. Lane 14 contained the PCR product of the human coagulation factor. Without the successful CG PCR reaction, this gel is typical of many obtained.

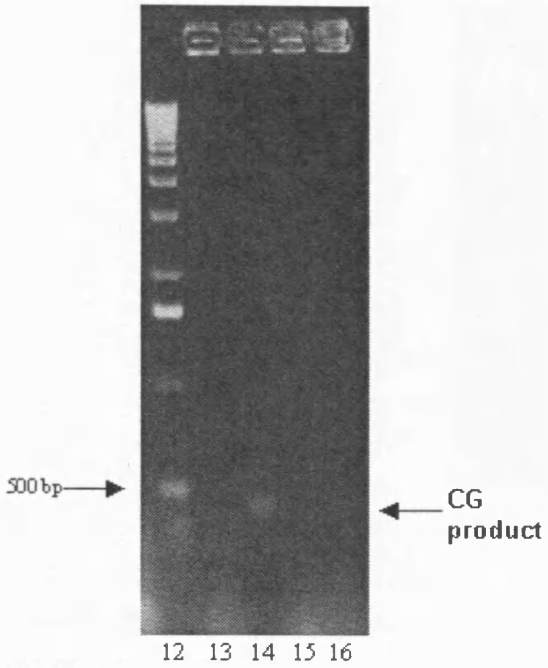


Figure 4.2 The same gel run further showing the size of the CG PCR product running as expected at around 400bp. Lane 12 contained 1KB plus ladder, lane 14 contained product from the CG PCR reaction.

The CG PCR product was captured in pCR4, a TOPO-TA vector, following the manufacturers instructions (Invitrogen). Clones were tested by cutting with *Pme* I and *Not* I, restriction sites located on either side of the insert (Fig 4.5).

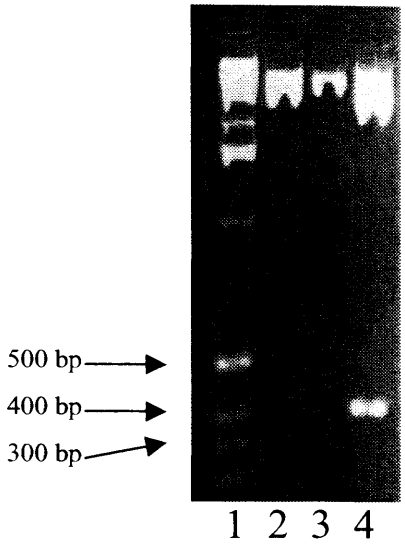


Figure 4.3 Gel to show TOPO TA clone containing CG insert. Lane 1 1KB plus ladder, lanes 2 and 3 clones with no insert, Lane 4 Coagulation factor insert cut out from TOPO TA vector.

Sequencing of the clone by Lark Technologies showed that the correct sequence had been amplified by the PCR and captured in the TOPO vector.
This was shown to contain the correct DNA sequence. During the restriction endonuclease analysis, several other clones were identifies, which contained 100 – 200 bp inserts. These were also sequenced and found to have a deletion in the purine-pyrimidine tract. A hypothesis as to why this occurred is presented in Chapter 5.

The 430 bp *Not* I – *Pst* I fragment of the TOPO vector containing the full-length sequence was ligated with [pBluescript II KS+], (Stratagene) itself cut using *Not* I and *Pst*I (Fig 4.6).

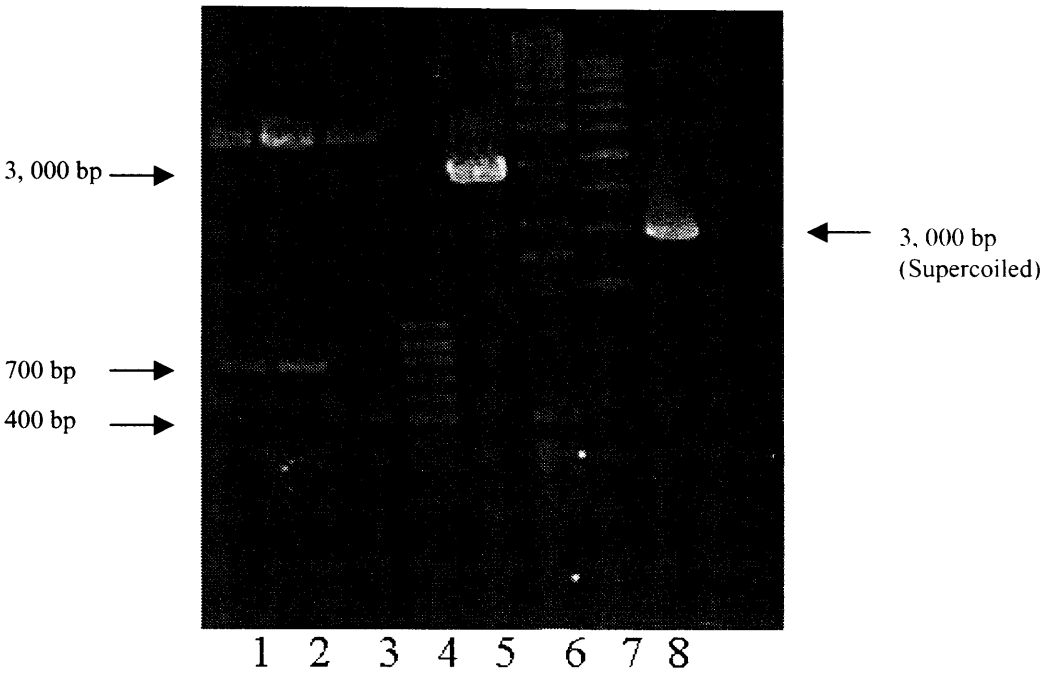


Figure 4.4 Gel to show CG and *cdc2* inserts. Lanes 1 and 2 contained the *cdc2* insert (~700bp) cut from TOPO vectors, lane 3 contained the coagulation factor insert (~400bp), lane 4 contained a low range mass ruler (Fermentas), lane 5 contained [pBluescript II KS +] cut with *Pst* I and *Not* I, lane 6 contained a 1KB plus ladder, lane 7 contained a supercoiled mass ladder and lane 8 contained uncut [pBluescript II KS +].

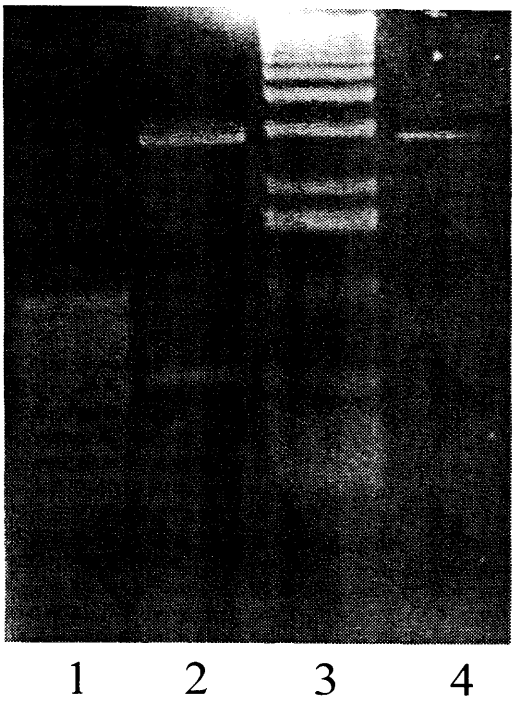


Figure 4.5 Gel showing full length CG insert cloned into [pBluescript II KS +]. Lane 1 Fermentas low range ladder, Lane 2, clone cut with *Not* I and *Pst* I, Lane 3 1 KB ladder plus, Lane 4 Linear [pBluescript II KS+]

Sequencing confirmed the presence of the full length CG sequence in the plasmid.

4.2.2 Cloning the *cdc2* promoter region

After several more unsuccessful attempts at PCR of the *cdc2* region, using the new primer set, some imaginative methods were used. One of these was to use 'Q-solution' as part of an old GC cloning kit (Roche). Q solution has been shown to consist of the amino acid analogue betaine, by use of NMR (Frackman *et al* 1998). Betaine improves amplification of DNA by PCR by helping reduce the formation of secondary structures (Weissensteiner and Lanchbury 1996, Henke *et al* 1997). There are two theories as to how betaine does this: The contact of A T base pairs in the major grove (Rees *et al* 1993). The second theory is that betaine changes the hydration of the minor groove (Mytelka and Chamberlin 1996) altering DNA flexibility.

The addition of betaine to the *cdc2* PCR reaction worked and generated large quantities of product (Fig 4.8).

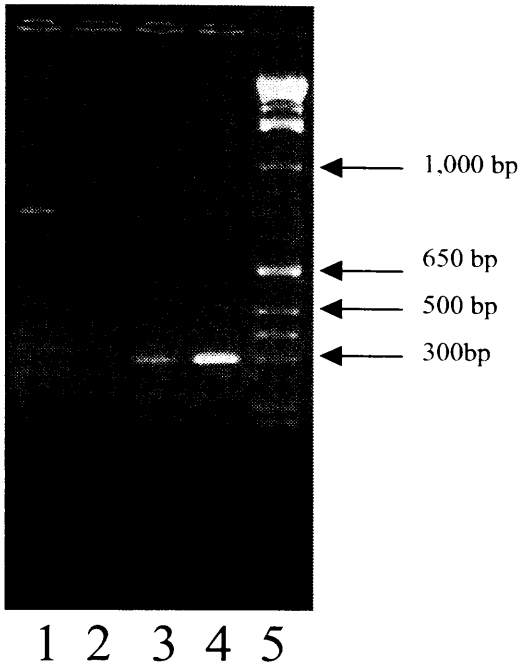


Figure 4.6 PCR of unusual sequences using betaine. Lane 1 contains the *cdc2* amplicon of 702bp, Lane 2 (typical of the rest of the gel which is not shown) contained a failed PCR reaction product, Lanes 3 and 4 show the Beta-actin control without (3) and with (4) addition of betaine. Lane 5 contained a 1KB ladder.

Attempts to clone the *cdc2* PCR product directly into [pQR338] failed. This might have been due to the size of the plasmid and hence associated low transformation efficiency and shear sensitivity.

More *cdc2* product was generated as before, and also by using the last of the previous successful PCR as a template. This was extracted from a gel, and single 5' adenine overhangs were added to the PCR product by incubation at 74°C for 10 minutes with Taq polymerase in 1xTaq buffer and dATPs. The sequence was then captured in the TOPO TA vector as before. The presence of the correct sequence was confirmed by restriction digest and sequencing (Lark Technologies).

The 772 bp *Not* I – *Pst* I fragment of this TOPO vector was isolated by gel electrophoresis, and ligated with [pBluescript II KS+], (Stratagene) itself cut using *Not* I and *Pst* I. The presence of the fragment was confirmed by restriction analysis.

4.3 Cloning using oligonucleotides

4.3.1 G-quartet - NHE

The sequence of the NHE in the *c-myc* control region had previously been shown to form from only a short oligonucleotide sequence. Following the difficulties encountered using PCR with the other unusual structures, a different strategy was attempted using complementary oligonucleotides.

A 100-mer and a 109-mer oligonucleotides were designed and manufactured to be complimentary with overhanging sticky ends suitable for cloning into *Bam* HI – *Hind* III sites.

NHEforward:

GATCCGGCTCTCTTACTCTGTTTACATCCTAGAGGTAGAGTGCTCGGCTGCCCGGC
TGAGTCTCCTCCCCACCTTCCCCACCCTCCCCACCCTCCCCATAAGCGCCCTCA

and NHEreverse:

AGCTTGAGGGCGCTTATGGGGAGGGTGGGGAGGGTGGGGAAGGTGGGGAGGAG
ACTCAGCCGGGCAGCCGAGCACTCTACCTCTAGGATGTAAACAGAGTAAGAGAGC
CG.

5µg of each oligo was dissolved in 50µL 1xNE buffer 2 (10mM Tris-HCl, 10mM $MgCl_2$, 50mM NaCl, 1mM dithiothreitol) (New England Biolabs, Hitchin, UK). This solution was heated to 100°C for 5 minutes and allowed to cool slowly.

The fragment was cloned into [pBluescript II KS+] cut with *Bam* HI and *Hind* III using a ratio of insert to vector molar ends of 10:1.

The presence of the fragment was confirmed by restriction analysis. Sequencing proved to be more of a problem however, and electropherograms from the [pBluescript + NHE] sequencing reaction flat line at the point where the DNA quadruplex was expected to form. This is characteristic of strong secondary structure. Adequate sequencing data was only obtained after using a number of different additives known to remove DNA secondary structure. DMSO and betaine were tried before Lark Technologies managed to read through the sequence using proprietary additives.

4.3.2 Highly bent sequence

Two oligonucleotides 89mers were designed and manufactured to be complimentary with overhanging sticky ends. A similar method to that followed for the NHE was followed, but cloning of the highly curved sequence was unsuccessful.

The core sequence was:

AAAAATGGCCCAAAAAGGGCCCAAAAATCTCCAAAAATATATAAAAAGGCCTAAAAA

CCCGGTAAAAATATATAAAAAGGGCCC

Using the bend.it ® server, this sequence was predicted to bend DNA by around 18° per helical turn, generating a bend of around 140°.

4.3.3 $(CG)_n$

It has previously been shown that $(CG)_n$ inserts form a variety of unusual DNA structures, and are unstable. Deletions were shown to remove an even number of base pairs and "remove the Z forming sequence to a size no longer able to adopt the Z conformation at normal levels of superhelical density" (Freund *et al* 1989). Despite this, attempts were made to clone 400-500bp $(CG)_n$ purchased from Sigma into [pBluescript II KS+]. It was envisaged that following successful cloning, studies could be undertaken to see if the sequence was more stable in different host cells lines.

The vector was prepared in a variety of ways with overhanging C and G bases and also blunt ended as Sigma were unable to confirm the state of their product.

Failure may have been due to toxic effects of such a long sequence, or the unknown quality of the product and hence poor ligation conditions.

4.4 The next steps

With a small family of plasmids constructed containing unusual features at a defined point, the next step was to characterise them, and see what, if any, their impact upon a plasmid production system would be.

Chapter 5:

Fermentation and characterization of plasmids with intrinsic structural features

5 Fermentation and characterization of plasmids with intrinsic structural features.

5.1 Introduction

The insertion of unusual features may occur in future gene therapy and in DNA vaccine plasmids. In the case of gene therapy products, this may be as a result of inclusion of sequences surrounding the coding sequence of interest, such as promoter regions and enhancers. In earlier chapters the possible biological relevance that intrinsic structures may have, particularly with regard to gene regulation has been covered. Viral sequences converted to double stranded DNA sequences for DNA vaccines might also produce non-Watson Crick structures.

In order to determine the implications that the inclusion of such sequences might have on the production of plasmid DNA, a family of plasmids was produced containing unusual structural features at a defined point. As mentioned in Chapter 1, several plasmid properties were studied, effect on host cell growth, yield, segregational stability, genetic stability and structural stability.

5.2 Fermentation results: Effect on cell growth and yield

5.2.1 Experiment to examine effect of initial clone choice on maximum cell growth rate

It is desirable to determine if the unusual structural features incorporated into plasmids affect host cell growth. In order to do this several fermentations were to be carried out, allowing measurement of cell growth in two different ways. Concerns with this methodology however are potential differences between the host cell state at the start of the fermentations. In addition a distribution of copy numbers will be present in a given population of cells. This might result in different maximum cell growth rates. If the initial average copy number of cells in inoculum was high, then the metabolic burden on the cells might slow down cell division (Bailey *et al* 1986). Host cells with less plasmid would have a lower metabolic burden and hence replicate more rapidly.

In order to minimise possible disparities inocula for fermentations were all prepared in the same way. In order to test to see if inocula prepared using different transformants

exhibited similar maximum growth rates an experiment was carried out. If initial cell state from inocula of cells containing the parental plasmid, all prepared in the same way significantly affected cell growth, then studies on the effect of non-Watson Crick structures in plasmids on growth rate would have to ensure that initial copy number distributions were similar.

DHI cells were transformed with [pBluescript II KS+] and plated on selective media. 6 colonies were picked off and shaken overnight at 37°C in 20mL universal containing 5mL LB broth with the appropriate ampicillin concentration. These were then used to seed 6 x 500mL shake flasks containing 50mL LB broth with the appropriate ampicillin concentration to a density of around 0.1 OD units at 600nm. These 500mL flasks were shaken at 37°C. Growth was monitored by OD measurement at 600nm at regular intervals.

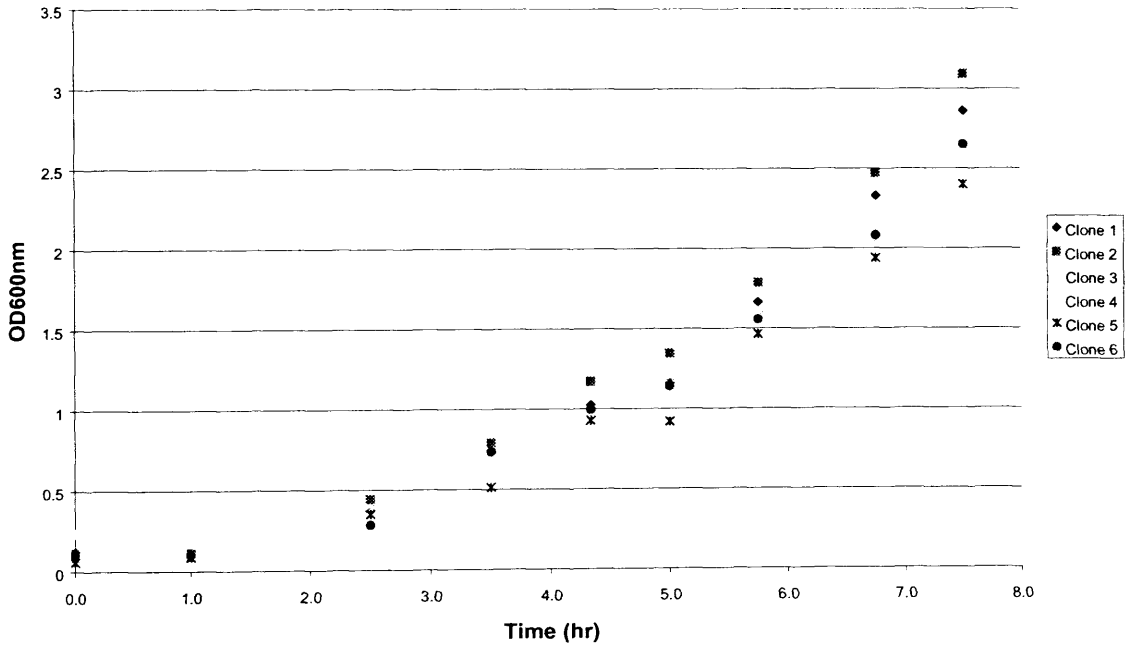


Figure 5.1 Comparison of growth between different inocula of DHI [pBluescript II KS+] in shake flasks

Maximum growth rates were hence determined for each starting culture:

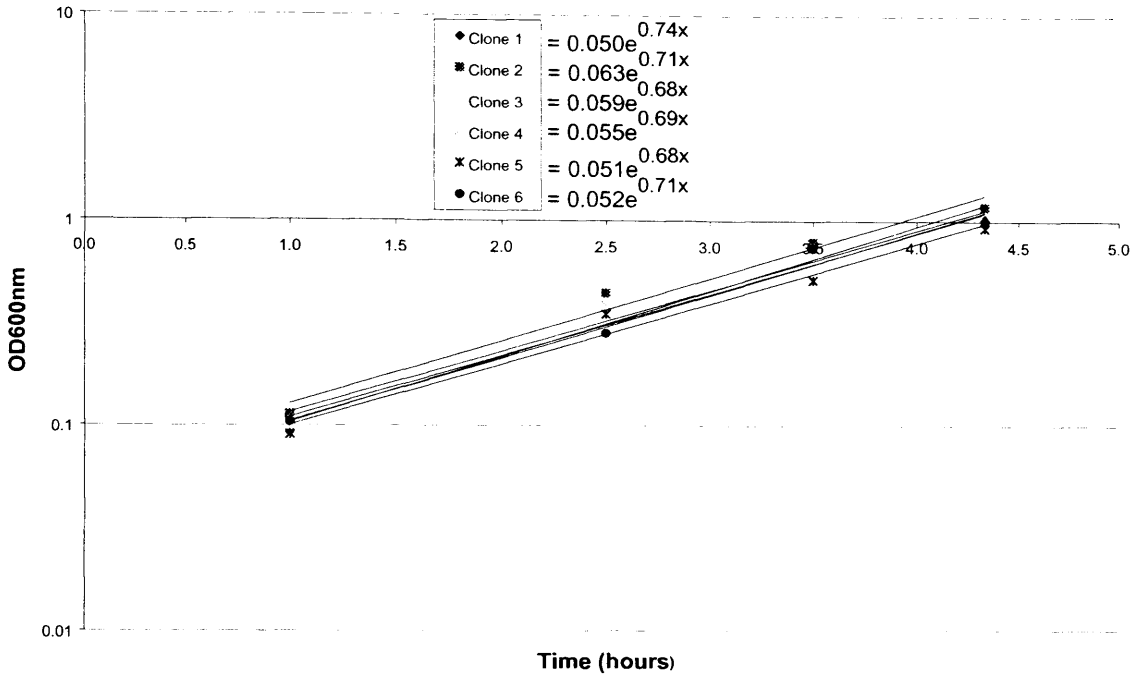


Figure 5.2 Maximum growth rate (μ_{max}) in shake flasks of different clones of DH1 [pBluescript II KS +]

Colony used	Maximum growth rate (hr^{-1}) in shake flask using LB, determined via OD600nm (± 10 %)
1	0.74
2	0.71
3	0.68
4	0.69
5	0.68
6	0.71

Table 5.1 Maximum growth rates in shake flasks of different clones of DH1 [pBluescript II KS +]

The highest and lowest observed maximum growth rates were found to be within 8% of one another (Fig 5.2 and Table 5.1), which is within the bounds of expected experimental error of the growth rate measurement (±10%). These results imply that the distribution of copy numbers in cells produced using the proptocl described in section 2.5.1, do not significantly effect cell growth rate. Additionally, internal research at GSK Beckenham suggests that growth rate is dependent upon the choice of plasmid backbone (data not shown).

5.2.2 Fermentation and cell growth rate

Stocks were prepared of DH1 containing one of the family of pBluescript plasmids described in chapter 4. Fermentations were carried out as described in chapter 2. Growth of DHI cells harbouring the plasmids was monitored both by using optical density at 600nm (A_{600}) (Fig 5.4) and Wet Cell Weight (WCW) (Fig 5.3).

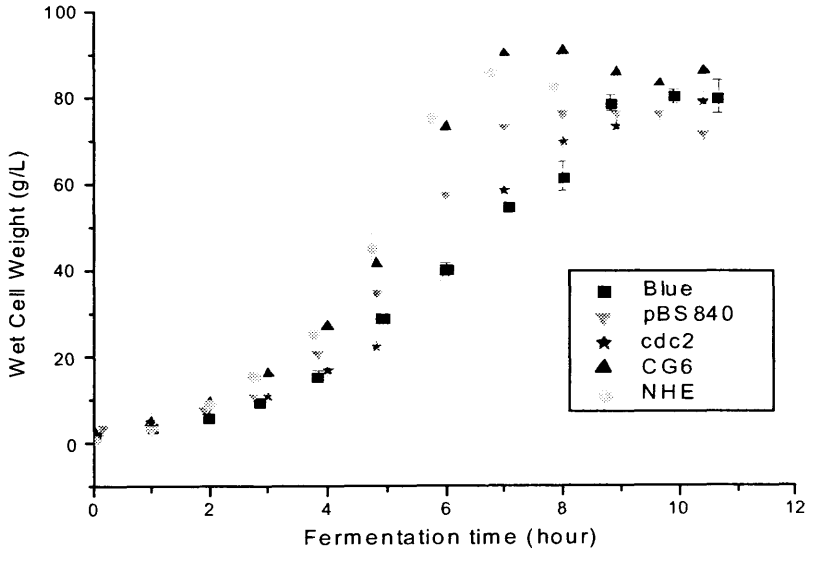


Figure 5.3 Average wet cell weight of DHI cells harbouring the plasmids shown (Blue denotes [pBluescript II KS+]) measured over the course of fermentations. Error bars indicate the standard deviation in data between triplicate fermentations.

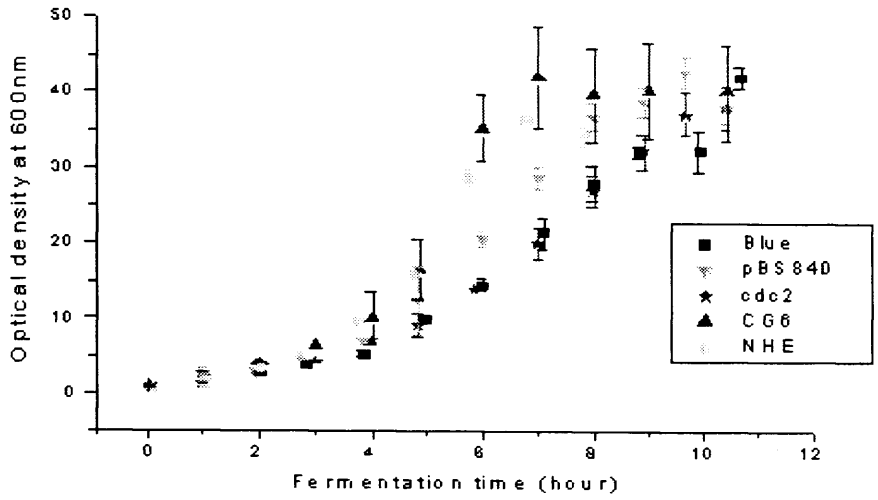
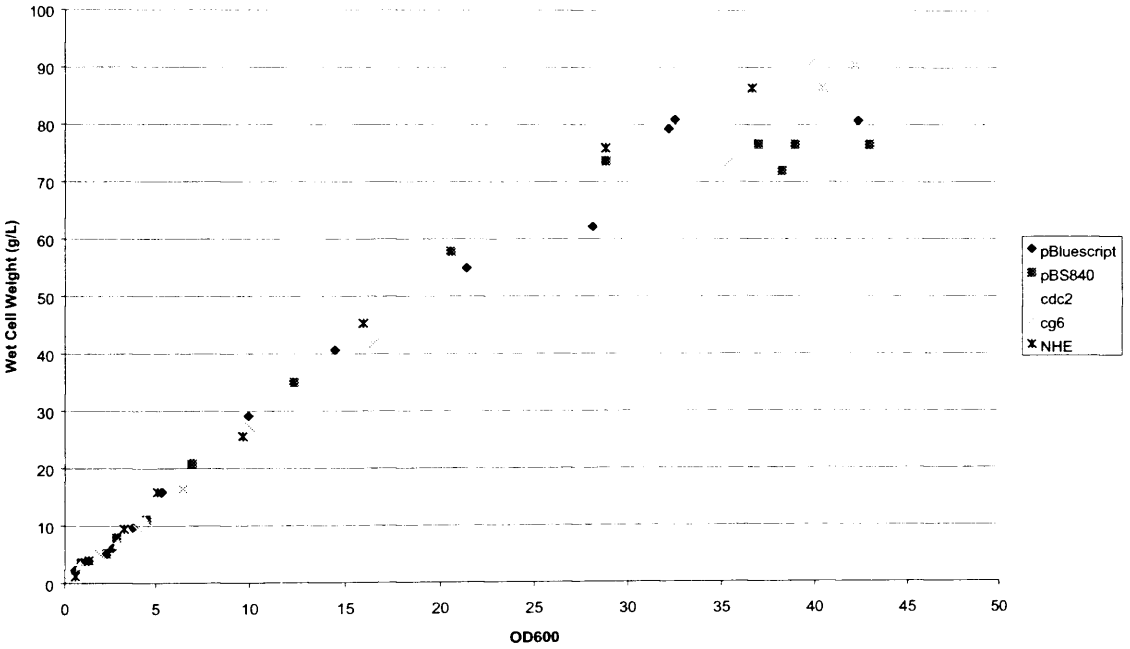


Figure 5.4 Optical density at 600nm measured over the course of fermentation. Error bars indicate the standard deviation in data between triplicate fermentations.

In order to compare the results obtained from both methods, the WCW and optical



densities at 600nm results can be plotted against each other (Fig 5.5).

Figure 5.5 Comparison of OD600 and Wet Cell Weight measurements

Maximum growth rates (μ_{max}) for each fermentation were determined by using both the optical density and wet cell weight data.

The mean value of these maximum growth rates was then determined for each construct. The figure comparing OD600 and WCW measurements shows good correlation between readings. The error involved in the WCW measurement was constant over the course of fermentation, whereas the error associated with measuring optical density increases with fermentation time due to experimental error incurred in the necessary dilution steps.

All the fermentations reached comparable maximum cell densities within 11 hours. As is readily seen from the growth curves, cells containing [pBluescript + CG] and cells containing [pBluescript + NHE] grew significantly faster than cells harbouring the other plasmids. Maximum cell growth rates were determined for each construct as before and are shown in Figs 5.6 and 5.7.

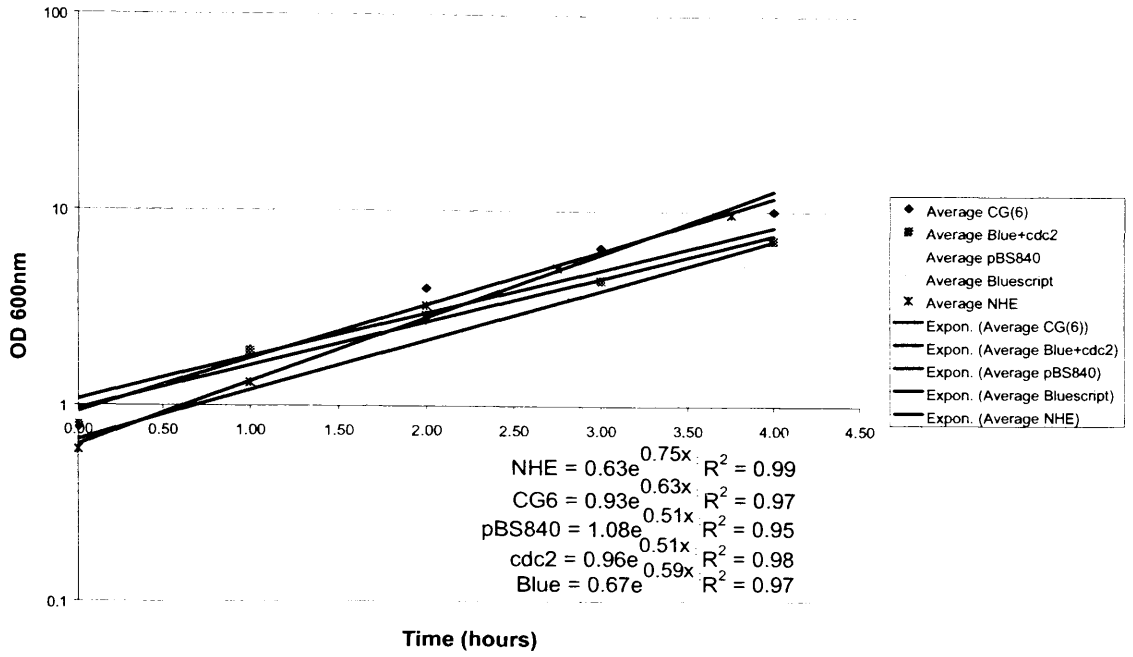


Figure 5.6 Average maximum growth rates (μ_{max}) of pBluescript based constructs in DHI determined by Optical Density measured at 600nm. Points plotted are average ODs from duplicate (or in the case of pBluescript + NHE, triplicate) fermentations.

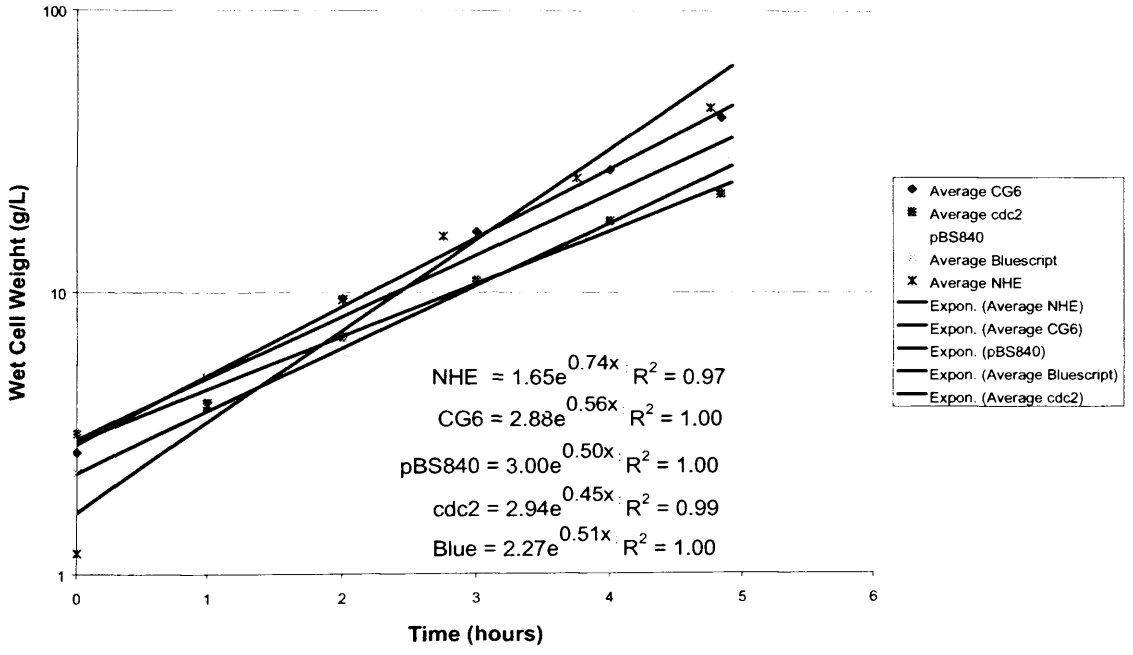


Figure 5.7 Average maximum growth rates (μ_{max}) of pBluescript based constructs in DHI determined by Wet Cell Weight. Points plotted are average WCWs from duplicate (or in the case of pBluescript + NHE, triplicate) fermentations.

Construct in DH1	Average Growth Rate (hr^{-1}), determined via	
	Optical Density at 600nm (± 10 %)	*Wet Cell Weight* (± 5%)
pBluescript + NHE	0.75	0.74
pBluescript II KS+	0.59	0.51
pBluescript + cdc2	0.51	0.45
pBS840	0.51	0.50
pBluescript + CG(6)	0.63	0.56

Table 5.2 Average growth rates of DH1 with different [pBluescript] based plasmids, in complex medium batch culture.

Within the bounds of experimental error, cells containing [pBluescript II KS +], [pBluescript + cdc2] and [pBS840] are seen to exhibit comparative maximum growth rates, while cells containing either [pBluescript + CG] or [pBluescript + NHE] exhibited faster growth.

Analysis of plasmid yield provides some explanation for this.

5.2.3 Yield throughout fermentation

Plasmid yield was determined throughout fermentation by both a spectrophotometric method (A_{260}) (Fig 5.8) and a fluorescence-based method (Fig 5.9) (PicoGreen).

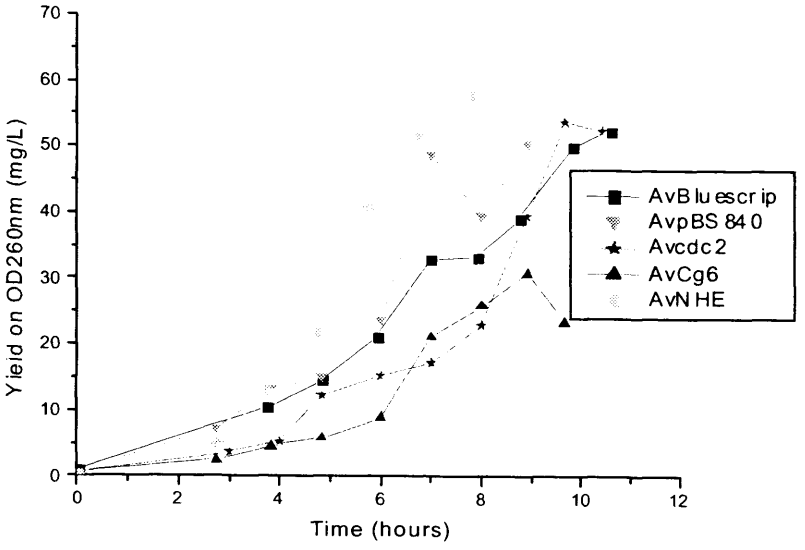


Figure 5.8 Plasmid yield determined by A_{260} throughout batch fermentation in DHI. Data shown averaged over two (or in the case of pBluescript + NHE three) fermentations

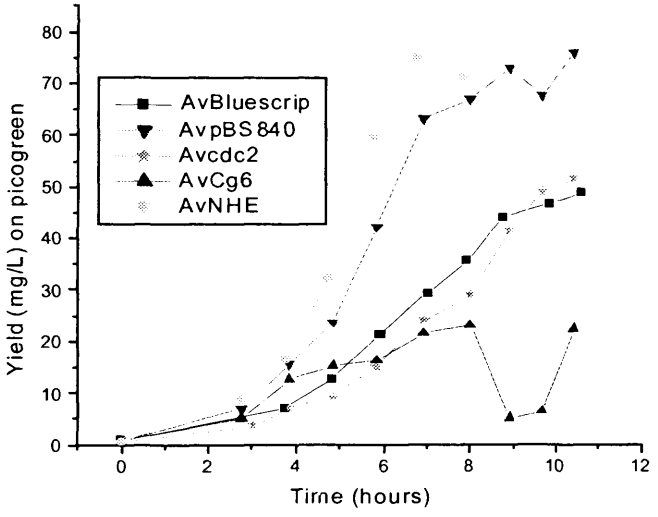


Figure 5.9 Plasmid yield determined by PicoGreen throughout batch fermentation in DHI. Data shown averaged over two (or in the case of pBluescript + NHE three) fermentations

From both sets of plasmid yield data it is shown that the yield of [pBluescript + CG] and hence average plasmid copy number is significantly lower than that obtained with the other plasmids. This may either have been as a result of the generation and proliferation

of plasmid free cells in the cell culture, or a general reduction in copy number throughout the bacterial population. Replica plating, which would have determined which, did not produce statistically significant results; however all 20 of the colonies tested contained plasmid.

The contrast between the results obtained using the spectrophotometric method (A_{260}) and a fluorescence-based method (PicoGreen) are evident from the above graphs. In order to compare the assays the values obtained for yield by both methods can be plotted against one another (Fig 5.10).

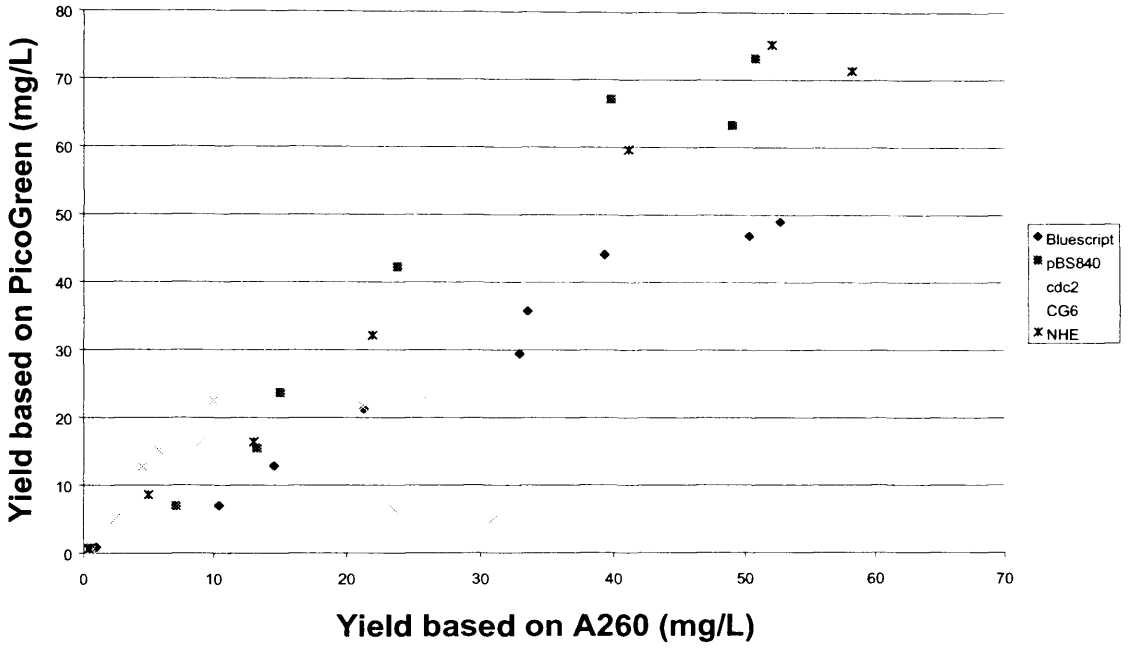


Figure 5.10 Comparison of results obtained for total plasmid yield using A260 and PicoGreen methods

The results presented are averages over several fermentations and this accounts for the apparent differences in yields measured by the two different methods for [pBluescript + CG], as for individual samples the results were similar. The disparity between results obtained for both [pBS840] and [pBluescript + NHE] between the two methods are more difficult to explain. The samples were measured in the same batch of picogreen samples, at a later time than the other samples. One possible explanation for the difference in results is that the standard plasmid used to calibrate the picogreen assay

may have had different proportion of linear species in it at different times. This would result in a shift in the calibration curve.

Overall, the observed differences in growth rates of cells harbouring the plasmids, yields and yield measurement were not considered to be major. Relatively large amounts of plasmid were being produced in all cases apart from [pBluescript + CG]. Small differences in yields have a large cost of goods impact at scale.

5.3 Effect of unusual structures on plasmid structural stability.

As mentioned in Chapter 1, the FDA guidelines state that "There should be a specification for the minimum amount of supercoiled DNA present.' (FDA Points to consider, December 1996), and published guidelines suggest a minimum figure of 90% supercoiled species (Shamlou 2003).

The plasmid topology assay was carried out as described in Chapter 2, section 2.6.3.1. Briefly samples of approximately the same number of cells were taken over time during cell culture. Plasmid was then extracted from these cells by alkaline lysis and ethanol precipitated. Spin prep columns were not used as anecdotal evidence suggested that they may damage plasmids, this was verified later in this section. Samples were then diluted 1/10 and loaded both neat and in their diluted form onto agarose gels containing suitable DNA mass ladders (Fig 5.11). Later experiments confirmed that by staining for 24 hours, and not 20 minutes, the dynamic range of the assay was increased such that dilution of samples was unnecessary. For consistency however, topology assays of the fermentation samples were performed in the same way.

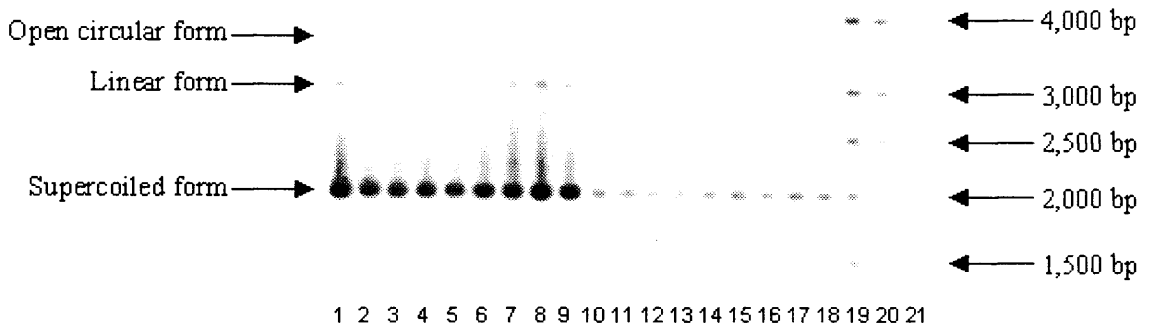
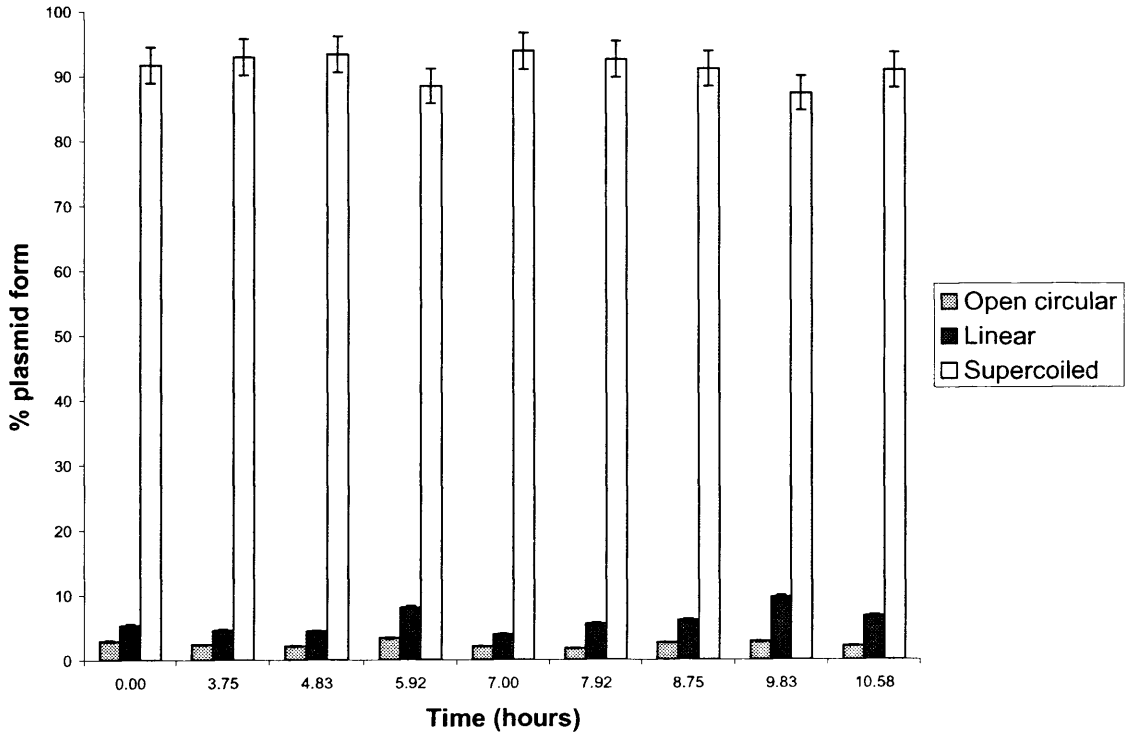


Figure 5.11 Analysis of plasmid topology from samples harvested over cell culture

a) 0.6 % 1 x TBE agarose gel of pBluescript II KS+ over the course of cultivation in bioreactor. Samples were obtained of approximately 2.5 x 10^9 cells, as described in the text. Plasmid was purified via ethanol precipitation following alkaline lysis and resuspended in a volume of 50µL. Lanes 1 to 9 contained 10µL neat plasmid, harvested at time 0, 3.75, 4.83, 5.92, 7.00, 7.92, 8.75, 9.83 and 10.59 hours respectively, lanes 10-18 contained 10µL of plasmid diluted 1 in 10. Lanes 19, 20 and 21 contained 10µL Fermentas High Range Mass ladder diluted 2,5 and 10 fold respectively.



b) Results from the same gel, showing no significant change in the relative proportions of supercoiled, open circular and linear plasmid over the course of fermentation.

Construct	Plasmid form (Average %)		
	Open circular	Linear	Supercoiled
pBS840 (triplex)	6.4±1.3	8.1±2.6	85.5±3.1
pBluescript II KS+ (parent)	2.4±0.5	5.4±1.8	92.2±2.2
pBluescript + NHE (quadruplex)	2.6±0.9	5.4±1.6	92.0±2.5
pBluescript + cdc2 (bend)	2.2±0.7	5.8±2.0	92.0±2.4

Table 5.3 Relative proportions of open circular, linear and supercoiled forms of plasmids.: Plasmids were purified by ethanol precipitation following alkaline lysis from samples obtained throughout cell culture. Data are the mean ± S.D. from two sets of cell culture samples. The differences in the amount of open circular and supercoiled plasmid between [pBS840] and the other plasmids are statistically significant ($P < 0.01$) as determined by the Tukey test.

As can be seen from table 5.3, there is a >5 % reduction in the proportion of supercoiled species obtained with the triplex containing plasmid compared to the other plasmids in the bluescript family. This difference in topology is not a size effect since [pBluescript cdc2] is of comparable size differing by only 73 bp.

This implies that the triplex sequence is responsible for decreasing the relative proportion of supercoiled species. This may be due to either the supercoiled plasmid containing the triplex sequence being more sensitive to shear effects, or there being more plasmid present in the *E. coli* cells in non-supercoiled form possibly due to stalling at replication forks. Due to the increase in the proportion of linear plasmid an increase in shear sensitivity seems the most likely explanation.

Samples of different plasmids for the topology assay were obtained and analysed at different times. There was therefore some concern that systematic error could account for the difference in the relative proportion of different species obtained with the triplex containing plasmid [pBS840]. Systematic errors include the times that plasmids were left in alkaline lysis buffers, and also any possible calibration errors in the densitometry quantification process. These are difficult to account for, and in particular with gel based assays relying on internal calibrations. Hence, in order to verify that the reduced proportion of supercoiled plasmid of [pBS840] in comparison with the other plasmids, a direct comparison was planned. Having the plasmids all on the same gel was desirable not only aesthetically, but also since quantification by agarose gel electrophoresis relies on comparison with the Mass Ladder data, which if loaded differently would yield different results.

Additionally experiments by others at GSK Beckenham (data not shown) had indicated that by increasing the staining time with SYBR Gold, from that recommended by Molecular Probes (between 20 and 40 minutes), to 24 hours, increased the assay linear range. A simple experiment confirmed this, (Fig 5.12) showing that dilution of samples and hence some errors incurred in this were avoidable.

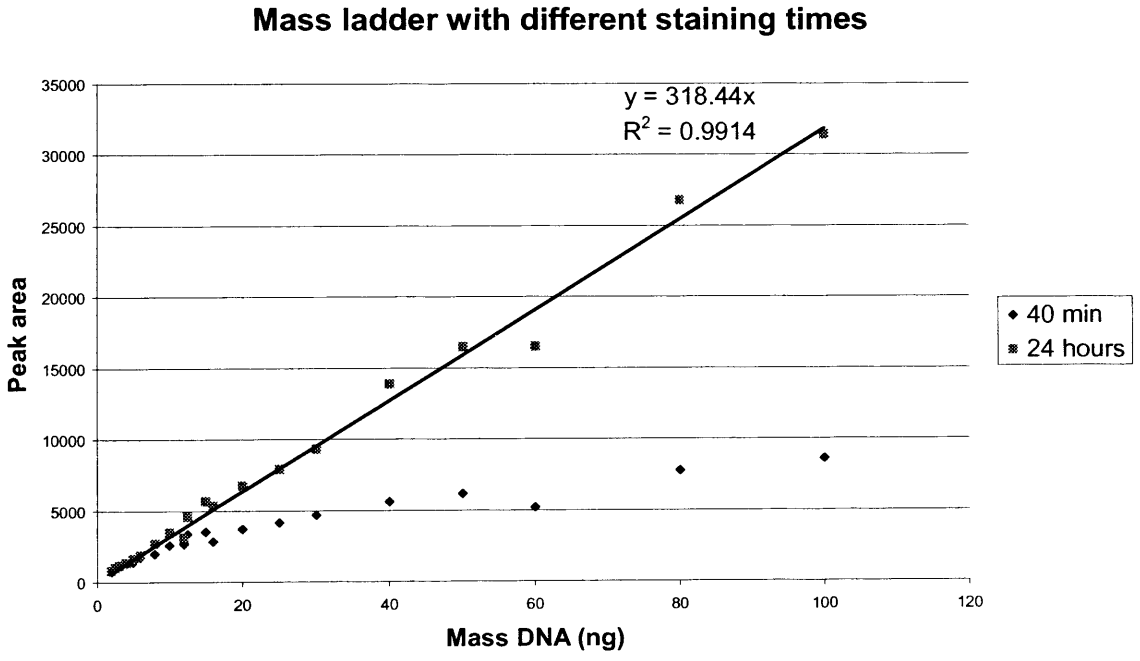
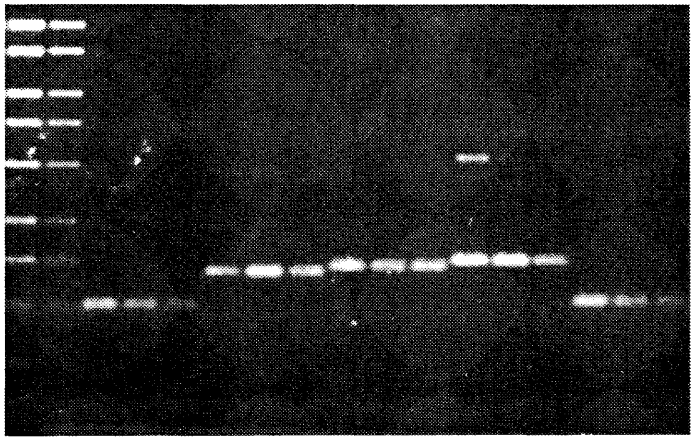


Figure 5.12 Comparison of assay dynamic range with 40 minute and 24 hour staining times.

Plasmid samples were available from fermentations, however they differed in age considerably and plasmids have previously been shown to alter topology with time (Middaugh *et al* 1998, Walther *et al* 2003).

Plasmid containing cells were therefore grown up over night and plasmid extracted as before by alkaline-lysis and ethanol precipitation. Since plasmid solutions prepared in this way are not of high purity, quantification by A_{260} was infeasible and a fixed volume of sample was used (10 µL). This corresponded to approximately 100 ng as the yield of plasmid was considerably lower using the ethanol precipitation rather than commercially available chromatographic methods.

The results obtained from this experiment (Fig 5.13) closely agreed with those generated from the fermentation samples. However three samples (lanes 7, 12 and 13) were loaded outside of the acceptable linear range of the assay.



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Figure 5.13 Gel comparing the relative proportions of open circular, linear and supercoiled plasmid species between plasmids. Lanes 1 and 2 are Mass Ladders, Lanes 3-5 contained [pBluescript II KS+], Lanes 6-8 contained [pBluescript + 618 λ], Lanes 9-11 contained [pBluescript II KS + cdc2], Lanes 12-14 contained [pBS840] and Lanes 15-17 contained [pBluescript II KS + NHE]

Following quantification of the samples using densitometry, the volume of samples required to give readings within the assay linear range were established.

Samples were stored at –20°C overnight and thawed the following day. An equal mass of each plasmid was loaded and the gel was run as before. This produced interesting results as a much larger proportion of non-supercoiled species was observed with

[pBS840] the triplex containing plasmid, than had been observed in the original gel (Figure 5.13).

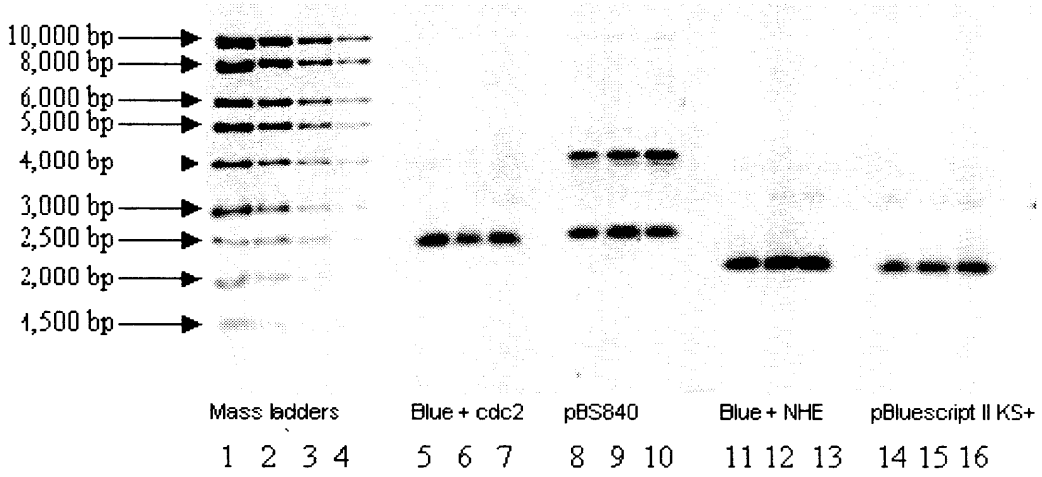


Figure 5.14 Second gel comparing the relative proportions of open circular, linear and supercoiled plasmid species between plasmids, after freeze-thaw. After quantification of the first gel the required amounts of samples were calculated so as to ensure even loading and that bands would be generated in the assay linear range. Lanes 1- 4 contained Mass Ladders, lanes 5-7 contained [pBluescript + cdc2], lanes 8-10 contained [pBS840], lanes 11-13 contained [pBluescript + NHE] and lanes 14-16 contained [pBluescript].

Construct	Average percentage in form	
	Open circular and Linear	Supercoiled
pBS840 (triplex)	50.4 ±9.7 (14.5)	49.6 ±9.7 (85.5)
pBluescript II KS+ (parent)	10.5 ±2.0 (7.8)	89.5 ±2.0 (92.2)
pBluescript + NHE (quadruplex)	7.5 ±1.4 (8.0)	92.5 ±1.4 (92.0)
pBluescript + cdc2 (bend)	8.1 ±1.6 (8.0)	91.9 ±1.6 (92.0)

Table 5.4 Plasmid topology following freeze thaw. Data are means of triplicate samples ± S.D. Figures in parenthesis average figures obtained before freeze-thaw.

It is well documented that freeze-thaw cycles can cause substantial damage to bio molecules. This is due to the open lattice structure of water molecules in ice generating a substantial increase in volume upon freezing.

In order to see if the effect of the triplex sequence present in [pBS840] would produce the same effect in lowering the proportion of supercoiled species in preparations of a larger plasmid, the sequence was cloned into the 16 Kb [pQR334] generating plasmid [pQR338].

Once again no significant effect on plasmid yield or growth rate (Table 5.5) of plasmid containing cells was observed between parental [pQR334] and daughter plasmid [pQR338].

Construct	Average Maximum Growth Rate (hr^{-1}) ± 5%
[pQR334]	0.28
[pQR338]	0.30

Table 5.5 Maximum growth rate of DH1 harbouring plasmid in a complex medium batch system

Plasmids were prepared on a small scale as before, both by ethanol precipitation and using commercially available chromatographic columns (Qiagen). Despite loading samples in higher concentrations than could be accurately quantified, no supercoiled species were observed for [pQR334] nor [pQR338] from samples prepared using chromatographic methods. The effects of shear on large plasmids were described in Chapter 1, section 1.5.1. The results for the ethanol prepared samples showed significantly less supercoiled species than obtained with the smaller plasmids, but with a high degree of variation between samples.

Construct	Plasmid form (Average %)		
	Open circular	Linear	Supercoiled
[pQR334] (16Kb parent)	20.8 ± 5.8	65.0 ± 8.5	18.2 ± 4.8
[pQR338] (triplex)	23.4 ± 4.1	61.6 ± 8.1	19.7 ± 8.1

Table 5.6 Topology of [pQR334] and [pQR338] extracted by ethanol precipitation following alkaline lysis

Note that considerable variation in the proportion of supercoiled species was observed in [pQR334] by the DIG assay reported in Chapter 3, section 3.8.4, with results varying between 7.4 and 37.9 %.

It seems clear that none of the methods (alkaline lysis followed by ethanol precipitation, spin prep or gravity driven chromatographic purification) were adequate to prepare the 16 KB plasmids intact. In order to properly examine the effects of the triplex sequence on large plasmids, a more suitable plasmid purification process would need to be developed.

5.3.1 Attempt to determine site of shear sensitivity

As can be seen in figure 5.14 [pBS840] is shown to have higher proportions of both open circular and linear species than the other plasmids, following freeze thaw. In order to investigate the site of physical damage to the plasmid a number of experiments were planned and executed.

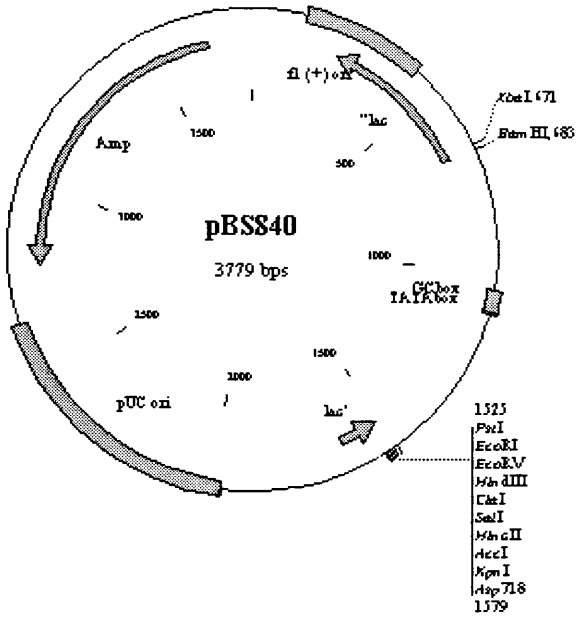


Figure 5.15 [pBS840] showing the position of restriction sites for *Bam* HI (683bp) and *Hind* III (1543bp) either side of the triplex containing region

By linearising an aliquot of the freeze-thawed plasmid sample with *Hind* III and separately linearising an aliquot with *Bam* HI, it was hoped that distinct bands smaller than the linearised plasmid would be visible. This in itself would show that the plasmid had become shear damaged to the linear form more frequently at given sites. If a simple smear had been produced then this would have shown that the site of shear damage was random.

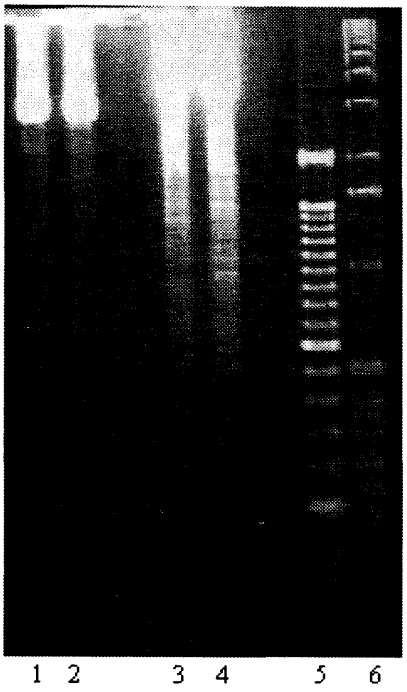


Figure 5.16 Experiment to determine whether shear was more likely to occur at the site of the triplex in [pBS840]. Lane 1 contained [pBluescript II KS +] cut with *Bam* HI, lane 2 contained [pBluescript II KS +] cut with *Hind* III, lane 3 contained freeze-thawed [pBS840] cut with BamHI, lane 4 contained [pBS840] cut with *Hind* III, lane 5 contained a 100bp lader, lane 6 contained a 1KB plus ladder. Gel over-exposed in order to see pattern of smaller gel fragments

Due to the large amount of previously uncut and open circular plasmid the full-length linearised band was very intense in comparison to the bands on the rest of the gel. It was therefore difficult to ascertain if smaller bands, ~500bp with *Hind* III and ~338 with *Bam* HI were truly present.

In order to try and overcome this problem several aliquots of freeze-thawed [pBS840] were run on a TAE gel alongside linear and open circular markers, the linear bans from the samples were gel extracted. The sheared linear [pBS840] was split into two aliquots and digested with either BamHI or HindIII as before. Insufficient DNA was recovered however and it was not possible to see any DNA on the resulting gel.

5.4 Chromatographic effects on plasmid topology

As mentioned at the beginning of this section, plasmids were purified by ethanol precipitation rather than using chromatographic methods in order to avoid mechanical shear damage to plasmids.

In order to examine the effects of miniprep columns, plasmid samples were purified both by use of the columns and by ethanol precipitation. These were then run on a single gel (Fig 5.17). Figure 5.14 above is actually a small portion of this gel.

The chromatographic effect of the spin prep columns on plasmids is dramatically demonstrated in the gel. With all constructs examined there was a substantially lower percentage of supercoiled plasmid when prepared by the spin prep method in comparison to more traditional ethanol precipitation. Hence while commercially available columns are suitable for rapidly producing clean plasmid suitable for cloning, they should not be relied upon to give a true representation of the topological state of plasmid in cells.

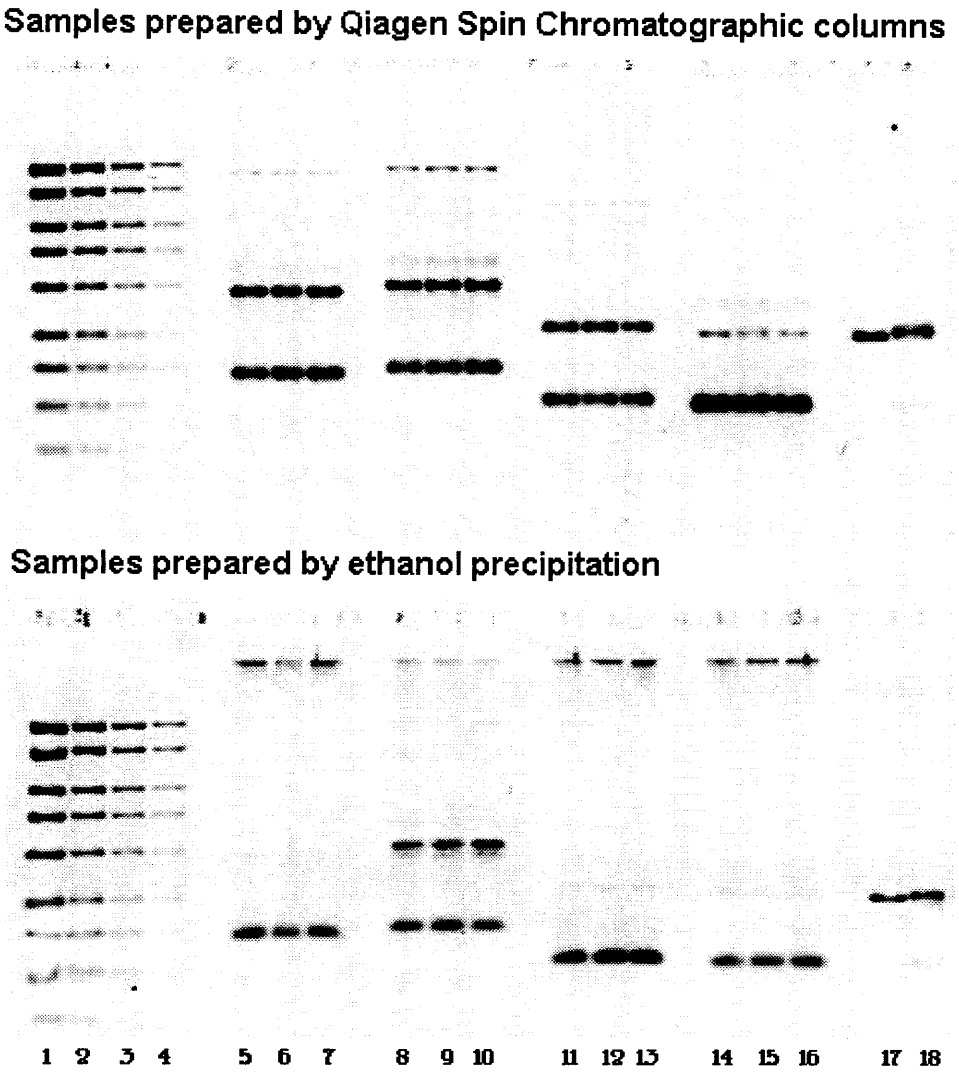


Figure 5.17 Gel showing comparative topologies of plasmids isolated by spin prep columns (top) and ethanol precipitation (bottom). Lanes 1 to 4 contained Fermantas High Range Mass Ladder, diluted and loaded as described in the text, lanes 5-7 contained [pBluescript + cdc2], lanes 8-10 contained [pBS840], lanes 11-13 contained [pBluescript + NHE], lanes 14-16 contained [pBluescript II KS+], lane 17 contained linear [pBluescript II KS +] and lane 18 contained nicked open circular [pBluescript II KS+].

While more plasmid can be seen to be loaded in the samples prepared by spin prep columns, differences between the two preparations in the proportion of each species are clear to see.

5.5 Effect of unusual structures on plasmid genetic stability.

As mentioned in chapter one, plasmids containing mirror repeats capable of forming cruciform structures, and plasmids containing direct repeats capable of forming slipped structures (SMP-DNA) have previously been shown to be genetically unstable (Lilley 1981, Mizuuchi *et al* 1982, Williams and Müller 1987). It has also previously been claimed that Z-DNA is genetically unstable (Freund *et al* 1989).

It is not unreasonable therefore to have expected some of the plasmids containing unusual structural features to exhibit genetic instability. In particular the coagulation factor insert containing over 100bp of alternating purine pyrimidines, capable of forming Z-DNA was a prime candidate to exhibit such instability.

Having obtained a single clear band from successful PCR of the coagulation factor (see figure 4.5 in Chapter 4) several different sized TOPO plasmids containing the PCR product were obtained. As described in the previous chapter it was thought that this may have been due to anomalous gel electrophoretic mobility of Z-DNA, however subsequent sequencing showed that the insert sequence was substantially smaller than the PCR product. It was thought that this may have been due to small quantities of shorter PCR amplicons in the PCR product. A full-length coagulation factor sequence was found in a TOPO clone, and this was used to produce [pBluescript CG] as previously described. Subsequently cells containing this plasmid were fermented and samples taken throughout fermentation.

Analysis of topology gels to determine the relative proportion of open circular, supercoiled and linear plasmid forms produced some unusual results for the [pBluescript CG]. It appeared that more than one form of the plasmid was present. In order to find if this was due to the presence of Z-DNA in a proportion of the plasmid population, or due to spontaneous deletions a number of experiments were carried out.

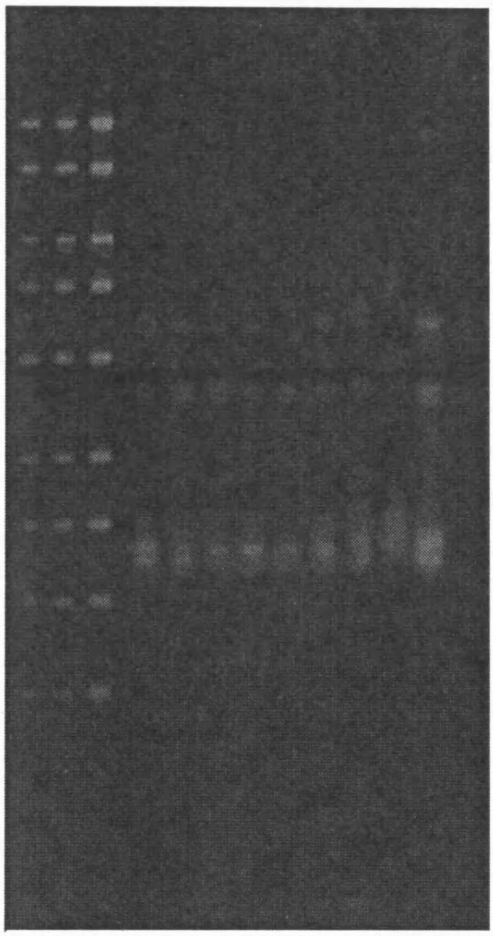


Figure 5.18 Topology gel of [pBluescript CG] fermentation samples taken throughout fermentation. Samples diluted 1 in 10 have been removed from the gel for clarity. Image capture time increased to 2 seconds in order to show the different species of [pBluescript CG] more clearly. Lanes 1 to 3 contained mass ladder as before, lanes 4 to 12 contained fermentation samples.

In order to study the cause of the unusual topology, the original plasmid preparation and a plasmid sample from the end of fermentation were retransformed at low concentration, cells were grown up, miniprepped, and the plasmid cut on either side of the insert.

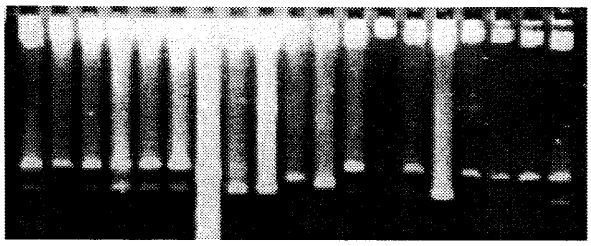


Figure 5.19 3% Agarose gel containing restriction enzyme digests from miniprepped DNA produced from retransformed [pBluescript CG] from original miniprep (lanes 1 - 6) and from the end of a fermentation (lanes 8 - 19). Plasmids were cut with XhoI and SacII to release the insert. Lane 7 contained a 1KB plus ladder. Gel stained with ethidium bromide, and over exposed to show additional inserts, otherwise not clearly visible.

As shown in figure 5.19, several different sized inserts are clearly visible in most samples. The smaller plasmid insert seems to be less prevalent in the retransformed original plasmid preparation, than in the fermentation samples. Given the transformation conditions, which should have introduced only one plasmid per cell, this suggests that the plasmid is genetically unstable.

The gel was not particularly clear and so a polyacrlamide gel (figure 5.20) was used in order to separate and size inserts in a more effective and accurate way.

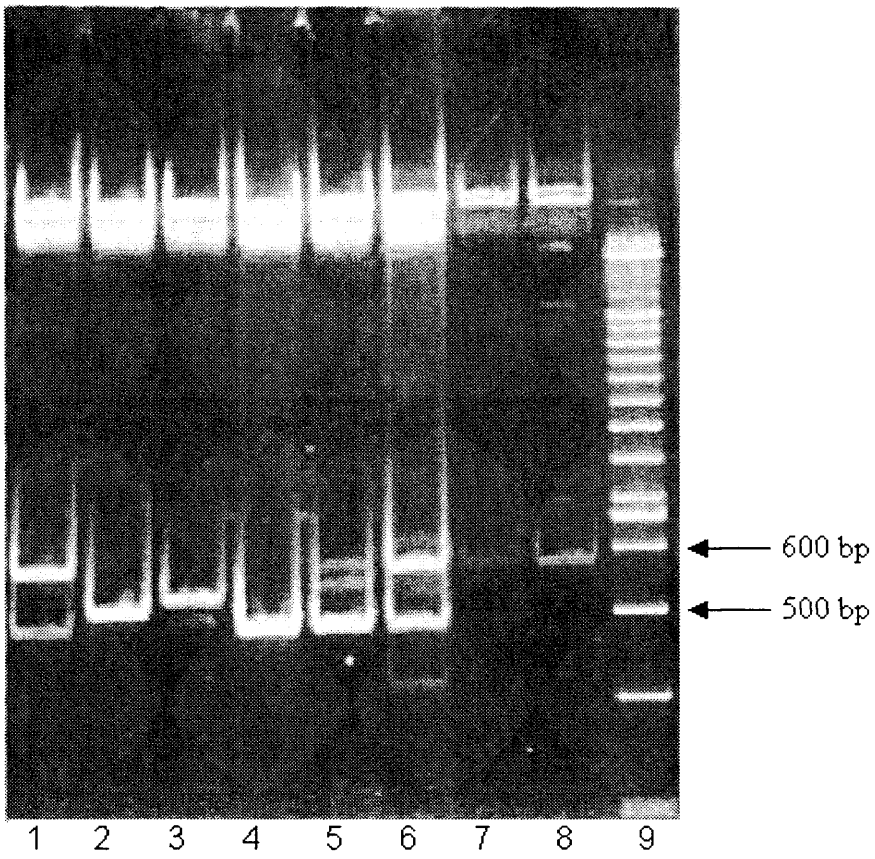


Figure 5.20 [pBluescript CG] insert instability shown in gel. Inserts from Z-DNA containing plasmid cut out using XhoI and SacII run on 5 % polyacrylamide gel for 100 minutes at 120 V, stained with SYBR gold. Lane 9 contained a 100bp ladder. Lanes 8 contained the insert from the original plasmid preparation (543 bp). Lane 7 contained the original plasmid retransformed and grown for a short time in an attempt to produce more of the full length insert, only small quantities of plasmid were produced. Lane 6 shows the different sized inserts obtained from a sample at the end of cell cultivation in bioreactor. Lanes 1-5 show different sized inserts obtained from retransforming plasmid obtained at the end of cell cultivation.

Lane 6 in figure 5.20 clearly shows that after cell culture several different sized inserts were present. Once again some of the retransformed preparations contained two different sized plasmids e.g. lane 1.

Subsequent sequencing of the smallest clones and comparison with the sequence obtained from the original preparation showed a reduction in length of the alternating purine•pyrimidine sequence and removal of the shorter flanking poly purine region.

There are two possible explanations:

1) The transformation process produces deletions.

2) The coagulation factor gene is genetically unstable in *E. coli*.

Certain sequences have been shown to be lost significantly more frequently than others during the transformation process (Hashem *et al* 2002). This could account for the differences in plasmid size when plasmids have been retransformed. It does not however appear to account for the difference between the plasmid at the start and end of fermentation. Only a single sized insert of plasmid is visible in lane 8 of figure 5.19, while several different sized inserts are clear in the sample taken at the end of fermentation shown in lane 6. Clearly less plasmid has been loaded in lane 8, (all the remaining plasmid available - retransforming and growing up plasmid overnight led to deletions shown in figure 5.18) however the SYBR gold stain is very sensitive and so ought to have picked up other species. It might however be the case that the population of plasmids with deleted Coagulation factor sequences was so small as to be undetectable at the start of fermentation, and only after several growth cycles became detectable.

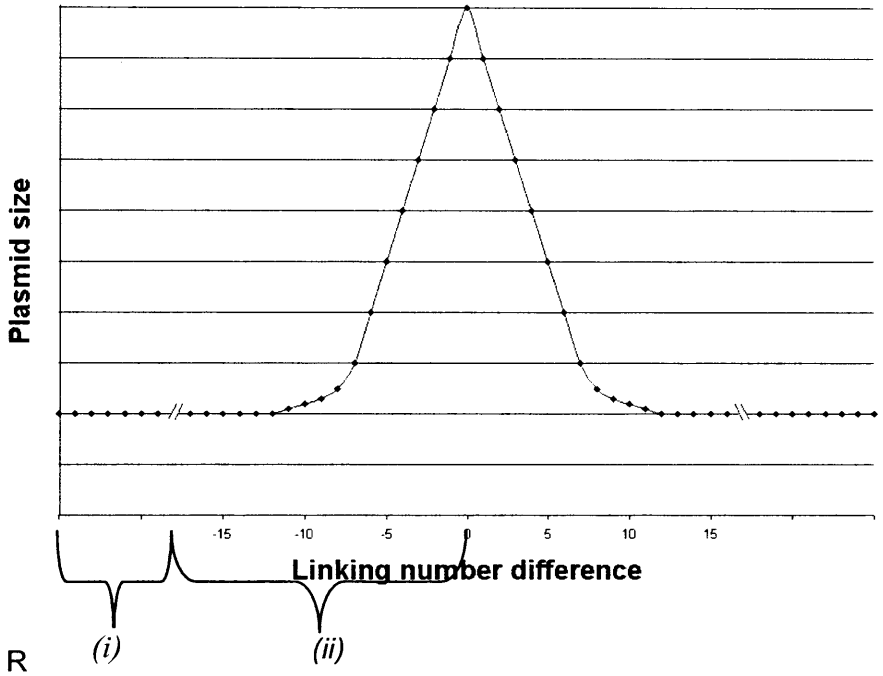
It has already been published that Z-DNA is unstable in *E. coli*, however the sequences chosen for the study were simple dinucleotides, which have been shown to be genetically unstable themselves (Bichara *et al* 2000, Freund *et al* 1989). Hence the Coagulation factor was chosen for the project since it contains a long alternating purine•pyrimidine sequence, which seemed likely to form Z-DNA under appropriate conditions, but was not so highly repetitive as to form other structures prone to spontaneous deletion (Bichara *et al* 2000). Given the problems encountered during cloning, the presence of one species at the start of fermentation and several species at the end, and the presence of more than one species of plasmid in retransformed cells, this seems a good explanation.

The mechanism of deletion of the Z-DNA sequence is not really of great importance in the context of biopharmaceutical production. Any unstable plasmid constructs would be picked up at an early stage in product development by existing methods of screening that generally involve several repetitive growth cycles and topology monitoring by gel electrophoresis. At present there seems no need to include Z-DNA in plasmid products, however should this change, the drug efficacy would have to be tested with different length inserts, and/or alternative strategies for production found.

5.6 Linking number analysis

Some linking number theory was outlined in chapter one. With appropriate intercalater concentrations, supercoiled plasmids can be separated by gel electrophoresis.

Supercoiled plasmids isolated from *E. coli* generally have a large number of supercoils and so are not easily resolved from each other. The addition of a low amount of intercalater increases the writhe, this increases the effective size of the plasmid and hence decreases the electrophoretic mobility.



(i) R Range of linking number differences normally encountered with supercoiled plasmids isolated from *Escherichia coli.*

(ii) Range of linking numbers required for separation in one dimension.

Figure 5.21 Probable relationship between plasmid size and linking number difference

The concentration of the intercalating agent, chloroquine, and required electrophoresis conditions were empirically found at GSK in order to separate plasmids produced in *E. coli*, on the basis of linking number.

It was anticipated that the nicked open circular species would run at approximately the same position as the covalently closed open circular species, as it does in gels

containing no intercalating agents (Uden *et al* 2002). This would have allowed bands to be easily assigned linking number difference values by counting down from the top band on the gel. In a similar way to how this is possible for plasmids with low linking number differences that can be separated in the absence of intercalaters:

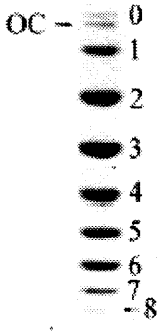


Figure 5.22 Gel electrophoretic separation of topoisomers of pUC19 DNA, in the absence of intercalaters. The mixture of topoisomers covering the range of ΔLk from 0 to -8 were electrophoresed from a single well in 1% agarose from top to bottom. The topoisomer with ΔLk = 0 has the lowest mobility it moves slightly slower than the open (nicked) circular DNA (OC). The value of (- ΔLK) for each topoisomer is shown. Figure reprinted from Vologodskii 2000 with kind permission from the Society of Biophysics Copyright.

However in fact, the relaxed open circular plasmid becomes strongly positively supercoiled in the presence of intercalator molecules and consequently runs faster than the other species. Nicked plasmid is unaffected by agents intercalating since superhelical stress can be removed simply by the rotation of one DNA strand around the other.

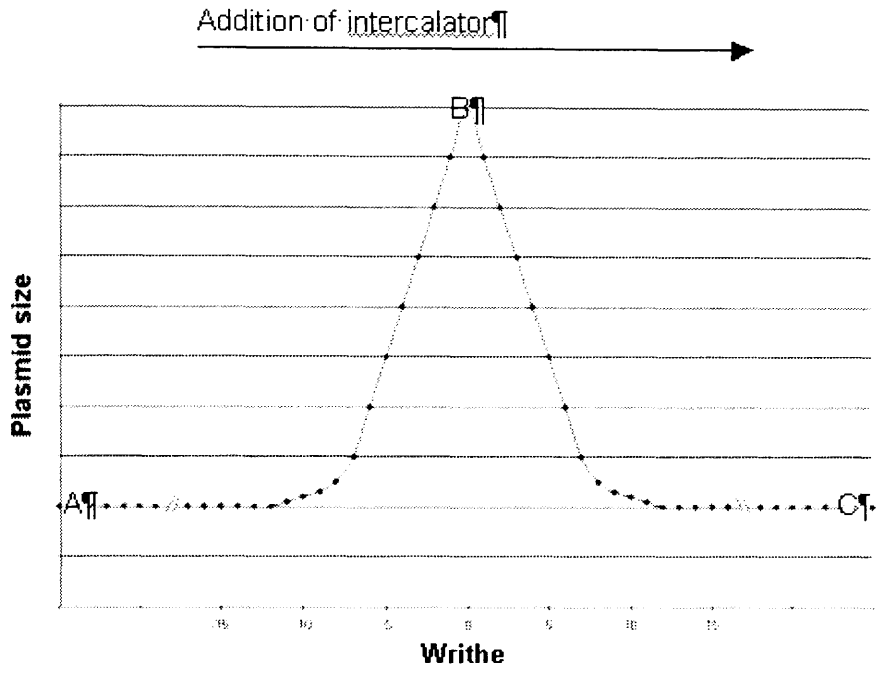


Figure 5.23 Probable effect on plasmid size and writhe upon addition of intercalator.

As the intercalater molecules bind to the DNA, they reduce the helical rotation angle between the base pairs. As more and more molecules bind, the superhelical turns within the DNA molecule are gradually removed and the conformation of the double helix becomes less compact. However once a certain number of ligands bind per base pair (ν) further ingtegration results in an increase in tension and so superhelical turns are added to the molecule.

$\nu = 360\ \Delta Lk / (\varphi N)$ (Vologodskii 2000)

φ denotes the change in angle, in degrees, between adjacent pairs upon integration of a ligand molecule between them.
N denotes the number of bound ligands, which can be determined by spectral methods.

As discussed in Chapter 1, relaxed plasmids have a linking number difference of zero, hence have the maximum size and hence lowest mobility. If the plasmid is nicked, addition of intercalator molecules will not alter the linking number since one end of the DNA is free to rotate, allowing the removal of any superhelical stresses generated. If the

plasmid is covalently closed however, addition of ligands makes a relaxed plasmid become positively supercoiled. It must be emphasised that the binding power of the ligands is insufficient to unwind the DNA double helix however, and the DNA will remain in the B-form in the positively supercoiled plasmid.

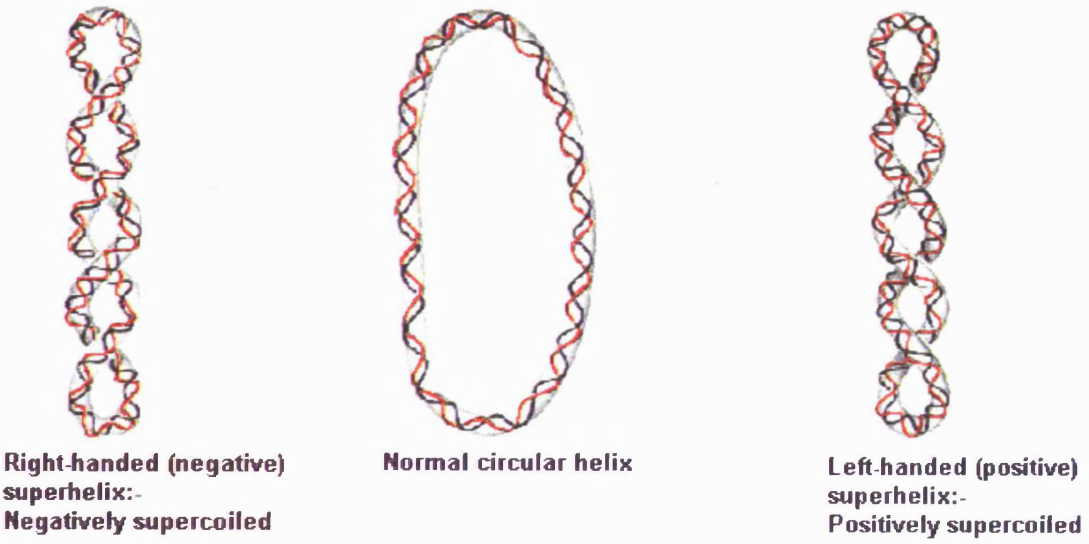


Figure 5.24 Negatively supercoiled, relaxed and positively supercoiled plasmids. Note that the DNA remains in the B-form. (Figure adapted from Stanford University, USA www.stanford.edu).

With addition of a small amount of intercalater, the tightly negatively supercoiled plasmids become sufficiently unwound to allow separation. Relaxed open circular plasmid would become strongly positively supercoiled, and nicked open circular plasmid would be unaffected.

In order to establish the mean average linking number difference, it is therefore necessary to know how many topoisomers are present between the (originally) relaxed open circular plasmid, and each plasmid species.

In order to do this, plasmid with an overlapping range of topoisomers could be run on a gel containing the intercalating agent, alongside plasmid samples. The linking number difference of each topoisomer band could then be established simply by counting back from the open circular plasmid. Bands could further be quantified allowing the mean linking number difference of plasmid samples to be calculated, making some account for changes in ambient conditions as Lk is temperature sensitive (Bauer *et al* 1980).

The figure below shows an envisaged system of working out the linking number difference of each topoisomer present in a plasmid sample:

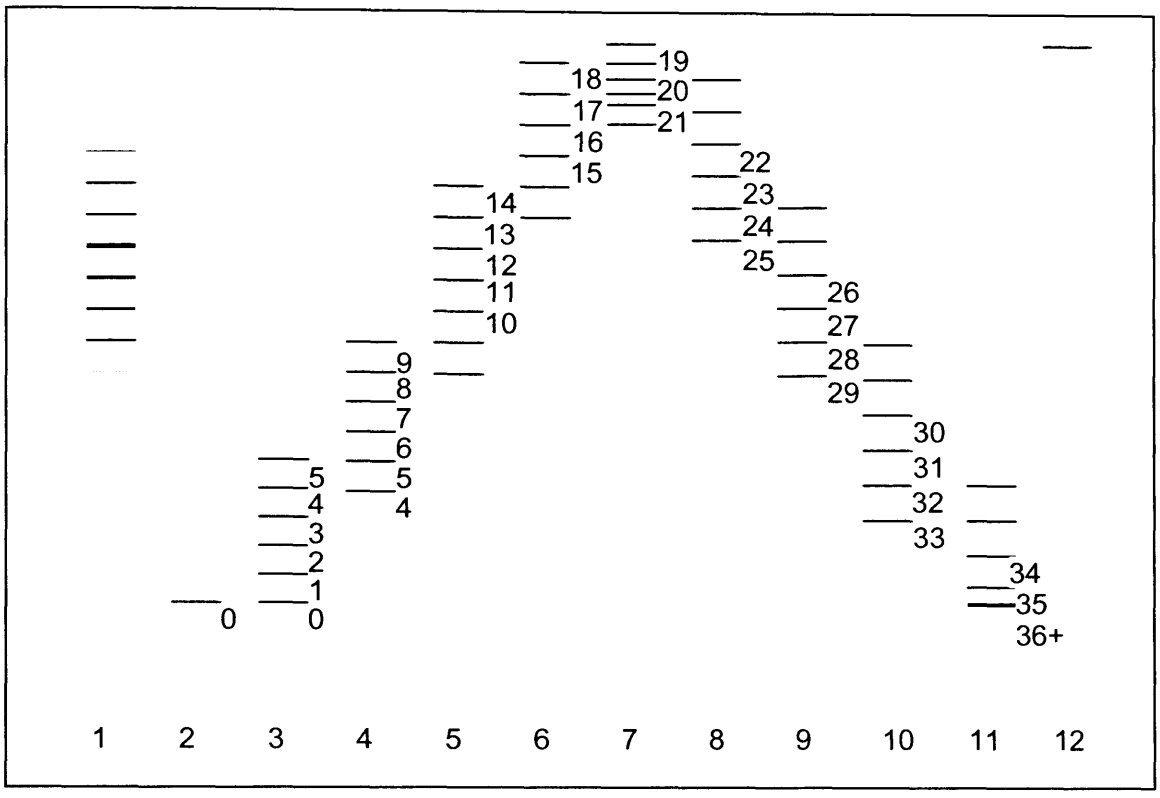


Figure 5.25 Envisaged method of topoisomer separation

Gel following staining. Both the gel and running buffer would contain chloroquine at suitable concentration. Lane 1 would contain a plasmid sample to be analysed (several plasmid samples would probably be run side by side for direct comparison), lane 2 would contain plasmid completely relaxed by topoisomerase, lanes 3 to 11 would contain distributions of plasmid topoisomers prepared by incubation with topoisomerase enzyme in the presence of an intercalater, such as ethidium bromide, after preparation the aliquot for lane 3 would have the smallest linking number difference (least supercoiled) and the aliquot for lane 11 would have the largest linking number difference (most supercoiled). Lane 12 shown to contain nicked relaxed plasmid for interest. If quantification of each topoisomer present in samples was desired then mass ruler dilutions could also be added as for the gels to determine the percentage of plasmid in the supercoiled, open circular and linear forms as described earlier. A linear plasmid sample would also be applied, and the corresponding bands present in each lane would then be ignored. While the band separation will not be the same between sucessive topoisomers, for the sake of illustration ease it has been shown so.

Linking number differences of each topoisomer have been marked. It would hence be a simple task to assign values for the linking number difference to each band present in the supercoiled sample. If mass ladders were also applied to the gel then each topoisomer band in the sample could be quantified. Hence the percentage of each topoisomer could be calculated and the mean linking number difference calculated for each plasmid sample. In this example the sample analysed clearly contains topoisomers with linking number differences between 29 and 22. Assuming B-DNA and using a correlation of 10.5bp per helical turn, an estimate of linking number could additionally be obtained, although this would be purely academic and would serve no useful purpose.

Unfortunately attempts to produce the range of topoisomers failed, and so it was not possible to obtain results such as those envisaged above. A proper set of topoisomers could have been used to calibrate the linking number gels previously obtained.

Using the linking number gels obtained, neither the linking number difference of particular topoisomers, nor the mean linking number, nor actual linking number can be established or estimated.

What was clear from the linking number gels was the linking number distribution was greatly effected by harvesting time. See the figure 5.26 below for an example of one of these gels. For all plasmids examined the average linking number was found to be significantly lower during log phase than in stationary phase. These findings can be explained by the energy state hypothesis described by Jensen and co workers (Jensen et al 1995). This may have implications for harvest times in batch cell culture in order to ensure a homogenous, reproducible drug product.

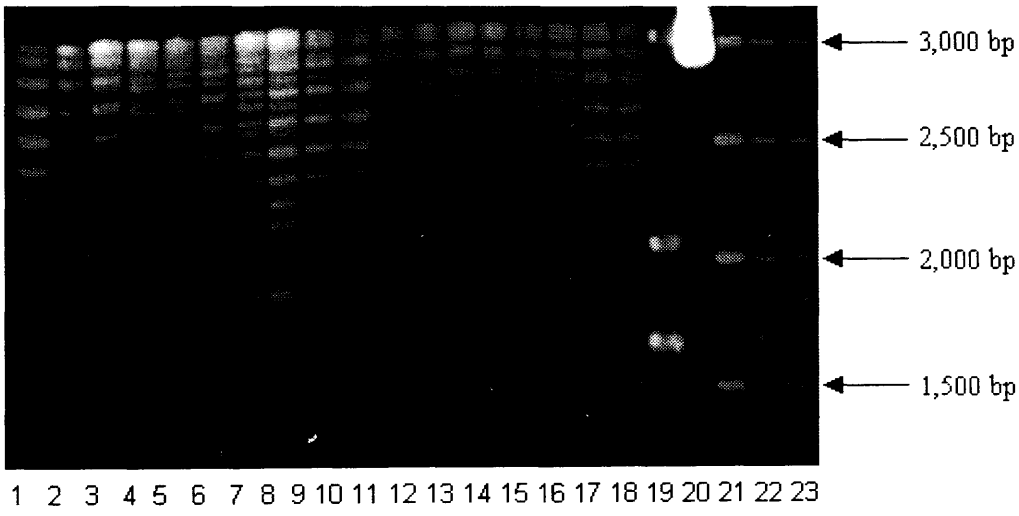


Figure 5.26 Linking number variation with cultivation time. Lanes 1 - 9 contained 500ng (quantified by A_{260} nm) [pBluescript II KS +] purified using commercially available kits (Qiagen) from samples taken during batch cell culture (Inoculum, 3.75, 4.83, 5.92, 7.00, 7.92, 8.75, 9.83, 10.58 hours). Lanes 10 - 18 contained 200ng as before. Lane 19 1 kb ladder. Lane 20 nicked open circular plasmid. Lanes 21 – 23 contained 10μL Fermentas High Range Mass ladder diluted 2,5 and 10 fold respectively.

Comparison of linking number differences between plasmids is possible with the gels produced; however, since the linking number distribution was shown to shift with cell

state, only samples at the same sate can be compared. This was performed on the inoculum samples, using the nicked open circular plasmid as a reference point. Since a comparison was made between different plasmids, it is not required to determine the actual value of the linking number difference, as the relative position of the nicked open circular plasmid would have been the same.

The inoculum culture of plasmid [pBS840] was found to have an average relative linking number difference 1.5 less than that of the parent plasmid [pBluescript II KS+]. This suggests the formation of an intramolecular triplex of around 16bp since the energy of removing supercoils is transferred to formation of the intra molecular triplex (Glover et al 1990).

In order to verify the effects (shown in figure 5.27) on linking number of batch cell culture, and also of the presence a DNA triplex, several experiments were carried out to try and separate the larger 16KB plasmid [pQR334] on the basis of linking number. The largest plasmid previously shown to have been separated on the basis of linking number is less than 10KB. A method to separate plasmids of large size on the basis of linking number was seen as a desirable goal since in time biopharmaceutical plasmids may increase in size and the effects of linking number upon drug efficacy are not yet known (O'Kennedy *et al* 2003). Despite many alterations of electrophoresis conditions, including use of Field Inversion Gel Electrophoresis, no separation of supercoiled samples was achieved. However, relaxed [pQR334] was separated suggesting that a higher order of structure than DNA supercoiling may have been present which prevented plasmids from being separated previously.

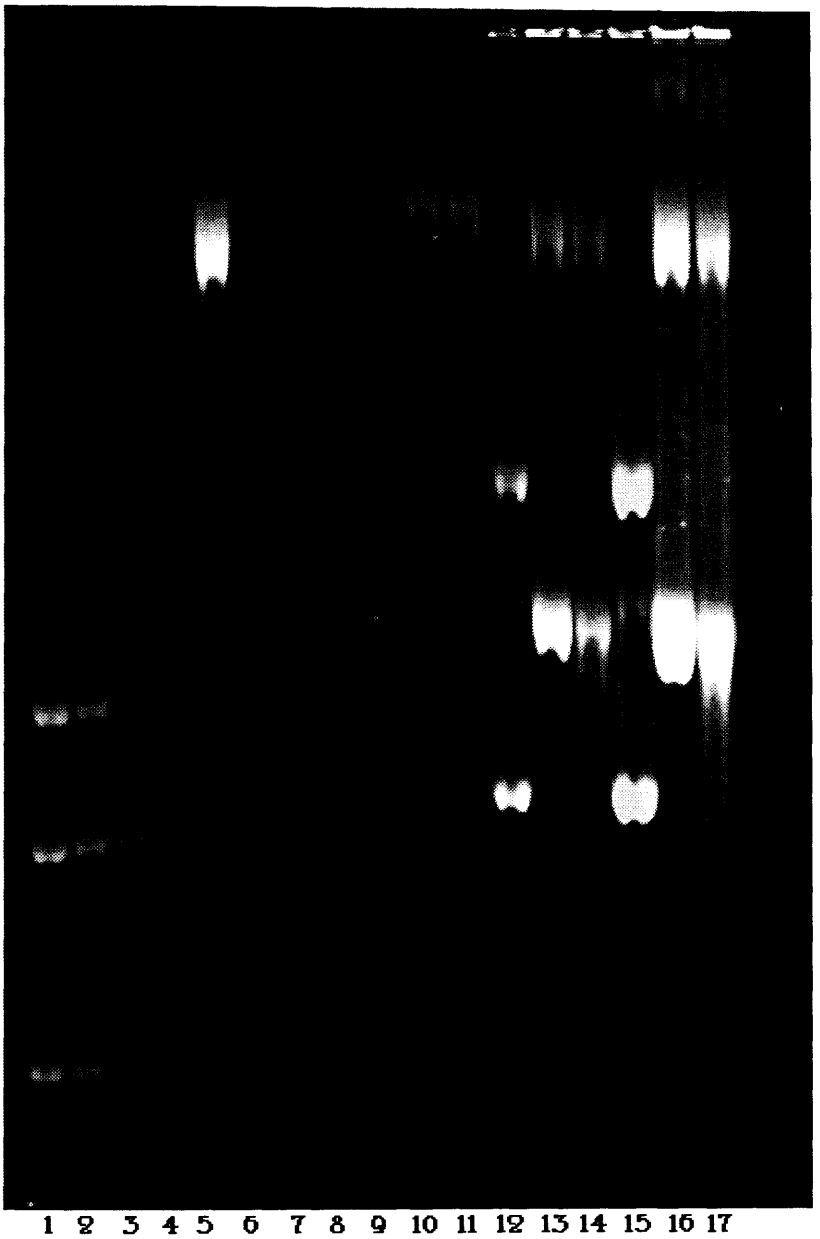


Figure 5.27 Chloroquine field inversion gel of [pQR334] in an attempt to separate plasmid on the basis of linking number. Lanes 1 to 4 contained dilutions of Fermantas mass ladder as before, lane 5 contained a λ mono cut ladder, lanes 6 to 11 contained fermentation samples of [pQR334]. Lanes 12 and 15 contained nicked open circular [pQR334], lanes 13 and 16 contained linear [pQR334] and lanes 14 and 17 contained [pQR334] relaxed with topoisomerase.

It was suggested that the reason for the lack of separation of the [pQR334] supercoiled plasmid on the basis of linking number, was the purity of samples. In particular the use of spin prep columns may have substantially sheared the plasmid. A further fermentation was carried out taking large samples at four time intervals, which were maxi prepped using gravity fed columns in order to reduce potential shear. However the problems with assay remained, clear bands not obtained and no differences between the samples could be observed.

5.7 Further topological analysis

Whilst the sequences added to [pBluescript II KS +] had previously been shown to produce unusual DNA structure, for completeness it was seen as desirable to demonstrate these. The most comprehensive method of doing this described in the literature is 2D gel electrophoresis (Bowater *et al* 1992, Martín-Parras *et al* 1998, Trigueros *et al* 2001, Vologodskii 2000, Wang *et al* 1983).

Successful 2D gel electrophoresis would have also allowed calibration of the linking number gels produced from plasmid samples isolated throughout fermentations.

An easy to follow and satisfactory explanation of the theory behind 2D gel electrophoresis of plasmids is sadly lacking from the literature, and regrettably as found, so too are reliable methodologies.

The principles behind 2D gel electrophoresis are:

A range of plasmid topoisomers is produced from supercoiled plasmid by incubation with different concentrations of ethidium bromide together with DNA topoisomerase. Following heat deactivation of the enzyme a number of phenol-chloroform extractions are carried out to remove all traces of ethidium bromide.

The ethidium bromide intercalates in the DNA and the topoisomerase enzyme is able to break and rejoin the DNA backbone. After the enzyme is denatured and the intercalating agent removed, the resulting super helical tension produces plasmids with given linking numbers. This means that under the correct conditions it should be possible to produce plasmids with a given linking number distribution.

Once such a distribution is established, electrophoresis of the plasmid is carried out in the first dimension with either no or little intercalater present, and then in the second dimension, perpendicular to the first, with a high concentration of intercalater.

In the first dimension the most supercoiled plasmid runs faster than the less supercoiled plasmid. There is no resolution of plasmids with more than a certain number of

supercoils however since they are of comparative size. See figure 5.28 below for a schematic demonstrating how plasmid size may change with linking number. As is intuitive, smaller species will move more rapidly through the gel. Electrophoresis in the first dimension hence separates only species with few supercoils (and any linear or open circular DNA present).

The absolute linking number of plasmids is unknown, an estimate of LK_0 is the size of the plasmid in base pairs divided by 10.5bp.

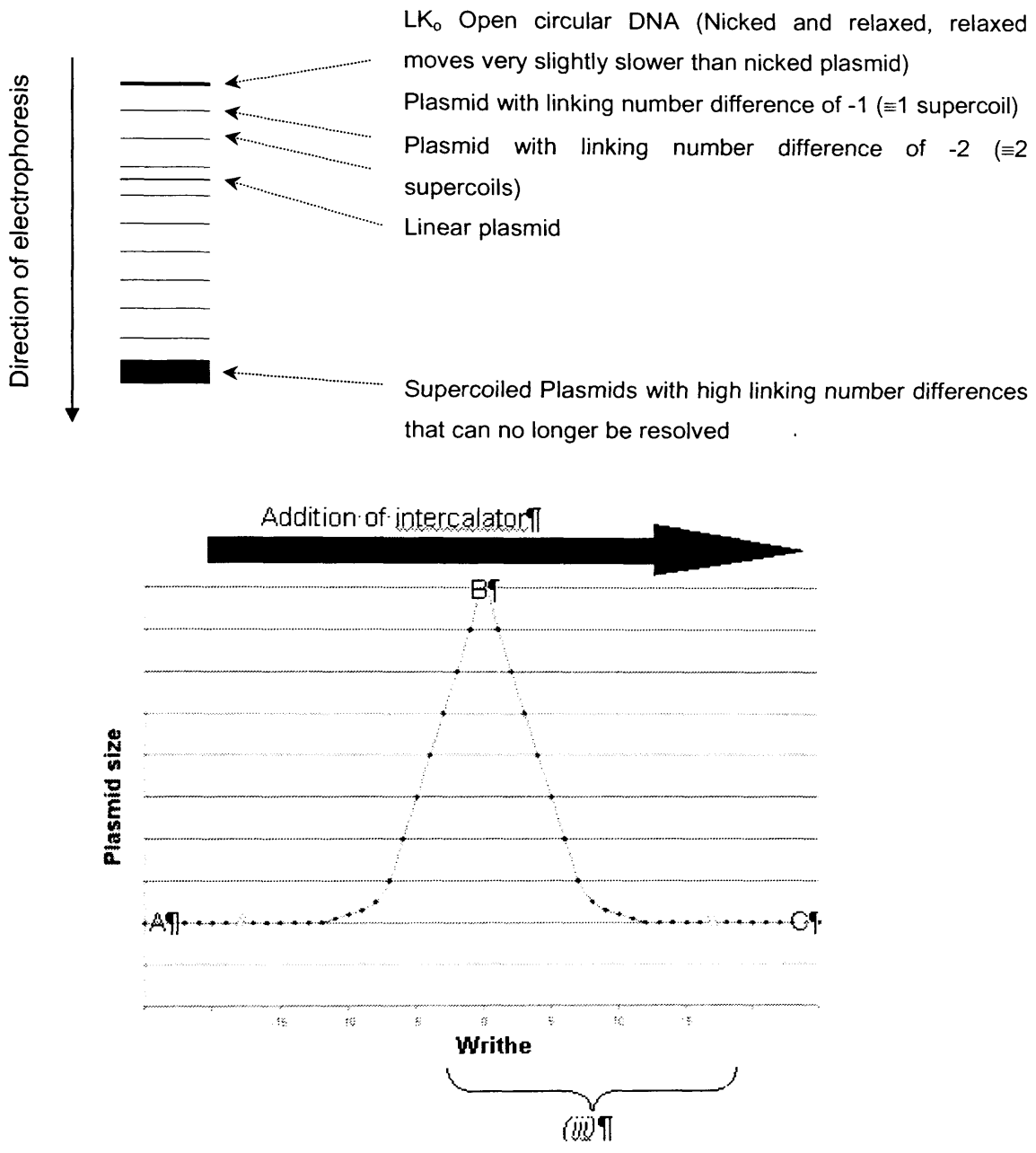


Figure 5.28 Writhe and plasmid size

a) Theoretical resolution of plasmid in first dimension by electrophoresis in the presence of no or little intercalater. Different supercoiled species differ by one linking number. At natural levels of supercoiling plasmid isolated from *E. coli* will not be resolved into species on the basis of writhe under these conditions.

b) Suggested relationship between plasmid size and writhe, showing the range of writhe required for separation in the second dimension.

The addition an intercalating agent for the electrophoresis in the second dimension alters the amount of supercoiling of plasmids by increasing the linking number. The amount of intercalater added should be such that the most supercoiled plasmid isolated from the fermentation ends up being of the largest size. Referring to figure 5.8 above, this means that if the most supercoiled plasmid isolated was at position A it would end up at position B.
Plasmids originally possessing fewer supercoils than the plasmid as position A, would hence become positively supercoiled.
Nicked open circular plasmids will remain at position B, whereas relaxed open circular plasmids would become highly positively supercoiled and move to position C.

After addition of the correct amount of intercalator, the mobility of topoisomers is no longer dependent upon ΔLk but becomes dependent on ($\Delta Lk - N\nu\varphi / 360$), the effective linking number difference.

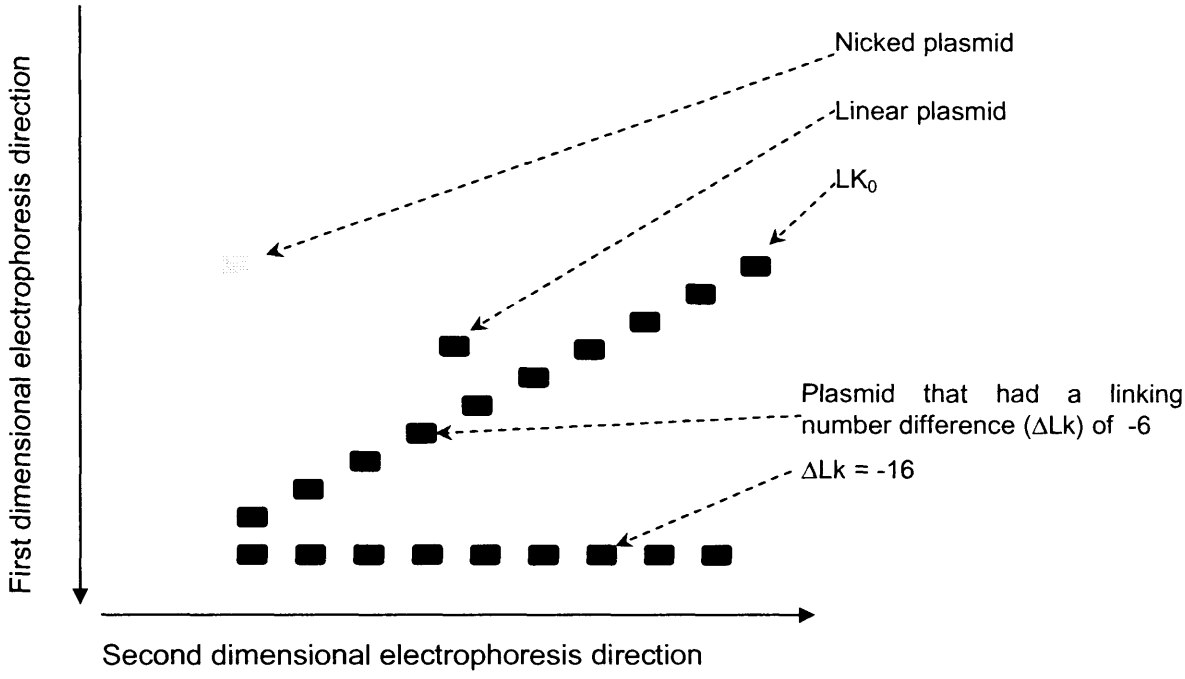


Figure 5.29 Diagram of two dimensional electrophoresis, separating plasmids on the basis of linking number. Intercalator used in the second dimension. Figure 5.24b shows what the gel would look like if stained after running the mixture of topoisomers in the first dimension. The presence of unusual DNA structures can be detected using this method, as at a given superhelical density, the unusual structure will be formed. This would result in a jump in linking number and hence obvious differences in the pattern produced between plasmid containing unusual structures and the parent plasmid.

5.8 The next steps

Further discussion of the results presented in this chapter can be found in Chapter 7. Some of the findings were that the triplex sequence significantly reduced the percentage of supercoiled plasmid, and that the degree of supercoiling of plasmids altered with fermentation time.

The topology of plasmids had been of interest to the pharmaceutical industry for some time, and as mentioned in Chapter 1, substantial research has investigated ways in which the proportion of supercoiled plasmid can be maximised. There is however little evidence to support the claim that supercoiled plasmids are better than other forms.

There are several methods that could be employed in order to test the efficacy of different plasmid topological forms. Methods delivering plasmids into cells would require some kind of normalisation of the amount of plasmid delivered, in order to measure transcription levels. Different plasmid forms may be delivered differently.

In order to investigate the effect of different plasmid forms on transcription it was decided to attempt a number of experiments in a cell free system. Some of these results are presented in the next chapter.

Chapter 6:

Effect of plasmid topology on transcription in a cell free system

6. Effect of plasmid topology on transcription in a cell free system

6.1 Guidelines on plasmid topology

Guidelines from the FDA state that: 'Plasmid-derived DNA species such as linear and relaxed circular DNA may be less effective in expressing the inserted antigen gene. There should be a specification for the minimum amount of supercoiled DNA present.' (FDA Points to Consider, December 1996). Following these guidelines, plasmid DNA manufacturing based research has aimed to produce pure plasmid DNA in the supercoiled form. Whilst plasmid based products have yet to reach market, a minimum for the amount of supercoiled DNA has been suggested as 90% (Levy *et al* 2000, Shamlou 2003) based on levels published elsewhere. Supercoiled plasmid DNA can be converted to open circular form and then linear forms by the action of shear, potentially at any stage of the purification process (Shamlou 2003).

Whilst there is some evidence that the risk of integration into the host genome is higher with the linear form (Nichols *et al* 1995) the chances of this occurring are still very low.

Perhaps following the lead of the FDA, the World Heath Organization recommends that in order to avoid chromosomal intergration, plasmid products should contain 'supercoiled, monomeric ccc DNA.' The idea of single forms of drug species is not new and is generally regarded as a good idea by both the pharmaceutical and biotech industries. The supercoiled form of plasmid is not a single isoform however, and consists of plasmids with different numbers of super helical turns. These can be separated using intercalating agents such as chloroquine. At any particular instant during cell culture there will be a range of linking numbers of the plasmids within the bacterial cells. Linking number distribution changes throughout batch fermentation as demonstrated in Chapter 5 and associated publication (Cooke *et al* 2004).

The degree of supercoiling of plasmids in bacterial cells is suspected to alter transcription levels in bacteria (Dorman *et al* 1996), it is hence a logical step to expect there to be differences in transcription of more or less supercoiled plasmids in eukaryotic cells.

6.2 Plasmid product efficacy

The efficacy of a plasmid product depends on the delivery method used, the longevity of the plasmid and on the expression of the given form by the host transcription machinery.

6.2.1 Plasmid delivery methods

Several different methods to deliver plasmids for therapeutic and vaccination purposes have been developed.

Early DNA vaccine and gene therapy studies using plasmid DNA, generally involved injection of plasmid in saline solution directly into muscle tissue or skin using a syringe and needle (Wolf *et al* 1990, Raz *et al* 1994). The transfection efficiency of naked supercoiled plasmid compared to other forms is significantly higher. Hence early criteria for plasmid products included a specification that plasmids ought to be >90% supercoiled.

The transfection efficiency of naked DNA is however very low in comparison to other delivery methods that are currently in trial. Several delivery methods of plasmids are available:

- Electroporation
- Liposomes
- Polyethylenimine complexes (Oh *et al* 2001)
- Gene gun

One such method involves precipitating plasmid DNA onto gold particles and propelling them into the skin. Plasmid treated in such a way often becomes nicked producing an increase in the relative amount of open circular plasmid in the preparation (GSK internal data).

6.2.2 Plasmid longevity

The topological form of the plasmid may have influence on the length of time that DNA vaccine or gene therapy products remain in human cells. To date no studies have been published investigating this. It is likely, however, that the linear form of plasmid will be degraded faster than circular forms of plasmid due to the action of endonucleases. Studies show that the half-life of plasmids in the blood is less than 5 minutes and that

plasmids are undetectable 1 hour post injection, supercoiled species having a half life of 1.2 minutes, before being converted to the open circular form, having a half life of 21 minutes, before being converted to the linear form with a half life of 11 minutes; retention by cells has been shown to be tissue-dependent, with plasmid detectable by Southern Blot up to 24 hours post-injection (Lew *et al* 1995, Houk *et al* 1999, Kim *et al* 2003). Difference between these experiments may be due to the extraction procedures used and sensitivity of the different analytical techniques.

6.2.3 The effect of topological form on expression

Despite the recommendations of the FDA and WHO, there is a surprisingly small amount of evidence to support the idea that supercoiled DNA is more effective at expressing than the linear forms (Weintraub *et al* 1986). In fact more recent research has shown the open circular form actually exhibits comparable expression levels (Xie and Tsong 1993).

A recent review describes the effects that different contexts have on transcription (Alvarez *et al* 2003).

In order to investigate the effects of different plasmid topological forms on transcription and hence expression efficacy in a DNA vaccine or gene therapy context an HIV vaccine candidate was used. This plasmid was treated in a variety of ways in order to generate a range of plasmid topologies, supercoiled, very supercoiled, relaxed open circular, nicked open circular, and linearised in a number of different places, in the plasmid backbone, in the promoter sequence and in the open reading frame of the HIV gene.

As for the plasmids described in the previous chapter, several attempts to produce the plasmid with different linking numbers was also attempted by incubation with different concentrations of ethidium bromide and DNA topoisomerase. If this had been successful, it would have been possible to see if there were any significant influences of linking number distribution (and hence harvest times of batch culture) on transcription.

6.2.4 Plasmid used

p[ABC] a DNA vaccine in clinical trials, and prime boost candidate, consists of a pUC19 origin of replication, the Kanamycin resistance gene transposon TN903 (Oka A et al 1981), the *cer* sequence (Summers and Sherrat 1984) and a CMV promoter expressing a fusion of three viral genes.

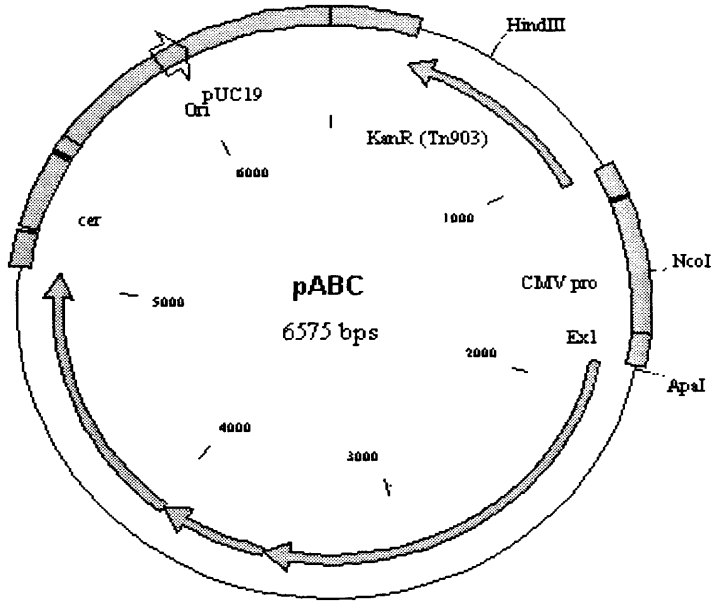


Figure 6.1 Plasmid map of [pABC] showing the pUC19 origin, Kanamycin resistance gene, CMV promoter, viral fusion gene and cer sequence. The *Hind* III site can be seen to be in the plasmid backbone, the NcoI site in the CMV promoter and the *Apa* I site in the open reading frame.

6.2.5 Plasmid treatment

Aliquots of p[ABC] was treated with a number of different enzymes in order to produce a series of plasmids with different topological forms.

Treatment	Plasmid form
NBstNBI	Nicked relaxed
HindIII	Linearised in backbone
HindIII SAP	Linearised and dephosphorylated in backbone
Gyrase	Supercoiled
Topoisomerase	Unwound relaxed
NcoI	Linearised in CMVpromoter
NcoI SAP	Linearised and dephosphorylated in CMV promoter
ApaI	Linearised in open reading frame
ApaI SAP	Linearised and dephosphorylated in open reading frame
Heat denatured SAP	Supercoiled

Table 6.1 Plasmid treatments

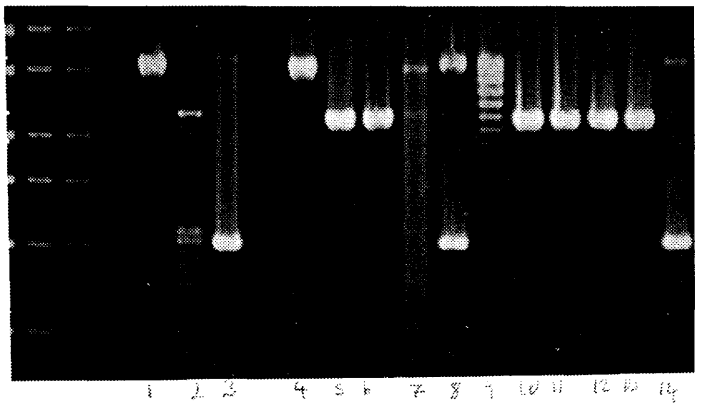


Figure 6.2 Gel of treated plasmids. 3 Mass ladders are on the left hand side of the gel. Lane 1 contained nicked open circular standard, lane 2 contained linear standard, lane 3 contained supercoiled standard, lane 4 contained *NBstNBI* treated nicked open circular plasmid, lane 5 contained plasmid cut by *Hind III*, lane 6 contained plasmid cut by *Hind III* and subsequently dephosphorylated by SAP, lane 7 contained plasmid incubated with BSA, lane 8 contained plasmid treated with DNA Gyrase, lane 9 contained relaxed plasmid produced by incubating with Topoisomerase, lane 10 contained plasmid cut by *NcoI*, lane 11 contained plasmid cut by *NcoI* and subsequently dephosphorylated by SAP, lane 12 contained plasmid cut by *ApaI*, lane 13 contained plasmid cut by *ApaI* and subsequently dephosphorylated by SAP, lane 14 contained plasmid incubated with heat denatured SAP.

Following treatment with modifying enzymes, these enzymes were heat denatured. The plasmids were not further purified. This ensured that each plasmid preparation was of

the same concentration. Any purification would have inevitably resulted in different yields between samples and hence each sample would have required quantification. It has previously been shown that plasmids of different morphologies exhibit different absorbancies at 260nm and bind intercalating dyes to different degrees (Sambrook *et al* 1989, Rock 2003). Dealing with such small quantities of plasmid any differences in concentration of differently treated samples may have made the experiment void.

In order to account for protein present in the reactions, deactivated enzyme was also added to the supercoiled reference plasmid, which was also heated at 80°C for 20 minutes.

Shrimp Alkaline Phosphatase was used in preference to Calf Intestinal Phosphatase since it can be heat deactivated and also does not remove additional bases, which might prevent religation.

6.2.6 Transcription

The HeLaScibe® Nuclear Extract in vitro Transcription System was used (Promega Corporation, Madison, WI, USA).

Essentially the recommended conditions and reagent concentrations were used for transcription as Promega has optimized these for the CMV promoter. Briefly, 100ng of each plasmid were incubated in triplicate for 60 minutes at 30°C, with 8 Units of HeLa cell nuclear extract in Transcription buffer (8.8mM HEPES, 44mM KCl, 0.088mM EDTA, 0.22 mM DTT, 8.8% glycerol) with 3mM $MgCl_2$ and 0.4 mM rNTPs.

After 60 minutes incubation the plasmids were spiked with 175µL of 1.34 ng/µL (172µL 50ng μL^{-1} mixed with 6228µL water) human RNA (Applied Biosystems). This allowed any discrepancies in the RNA recovery procedure to be normalized.

RNA extraction was performed with an automatic preparation device (MagnaPure; Roche) according to the manufacturer's preparation protocol. RNA was converted to DNA using reverse transciptase. This was performed in duplicate for each of the samples. Semi-quantitative PCR was performed on an ABI prism 7700 q-PCR machine (Applied Biosystems,

The RT-PCR reaction exploits the 5′ nuclease activity of the AmpliTaq Gold® DNA Polymerase in cleaving the TaqMan probe during PCR. The TaqMan probe contains a reporter dye at the 5′ end of the probe. There is also a quencher dye at the 3′end of the

probe. The proximity of the reporter and quencher in the intact probe results in suppression of the reporter fluorescence as described elsewhere (Förster 1948, Lakowicz 1983). During the reaction, the reporter dye and quencher dye become separated, resulting in increased fluorescence of the reporter. The accumulation of PCR products is detected by examining a rise in fluorescence produced by the reporter.

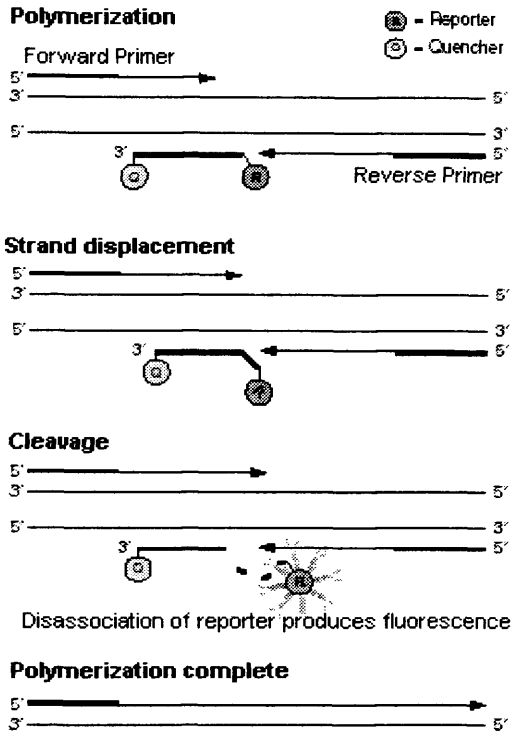


Figure 6.3 qPCR reaction. As the PCR reaction progresses more reporter probes become disassociated from quenchers and so an increase in fluorescence is observed. Figure adapted from Applied Biosystems (www.appliedbiosystems.com).

Analysis of 18S results allowed difference in yields from the RNA purification to be normalized. This assumes that the yield from purification of mRNA and rRNA is the same.

The delta delta Ct method was used to find the levels of transcription to be calculated. This method is comprehensively described elsewhere (Livak and Schmittgen 2001).

6.3 Transcription results

All negative controls produced the expected results:

Samples of RNA extracted, and not converted to DNA by reverse transcriptase, were included in the qPCR reaction and gave no signal. This showed that the extraction of the mRNA was not contaminated by plasmid DNA, which would have given a false indication of the transcription level.

The three negative controls in which no plasmid was included with HeLa cell nuclear extract also produced no signal demonstrating that no contamination occurred.

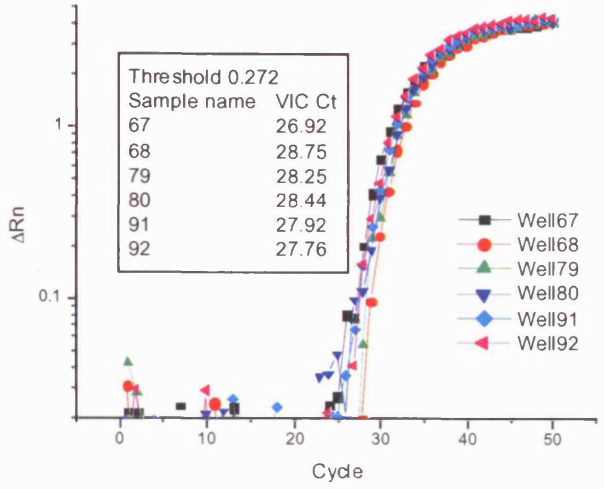


Figure 6.4 VIC signal for negative control

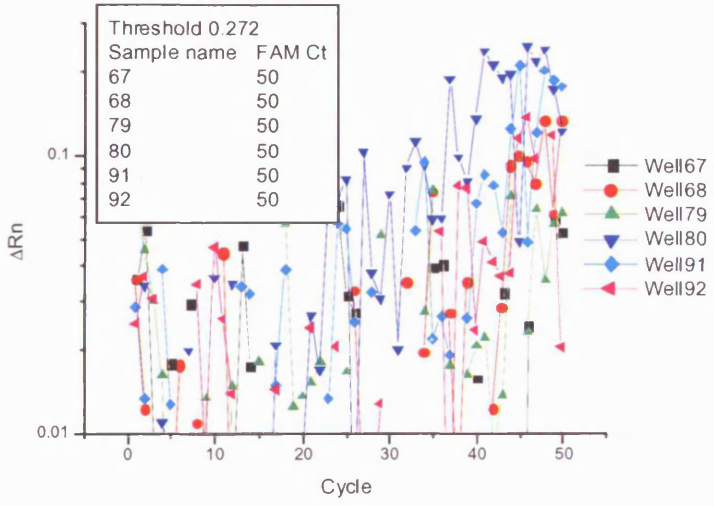


Figure 6.5 FAM signal from negative control indicating no copies of mRNA of interest. Slight probe degradation accounts for small linear increase in observed signal.

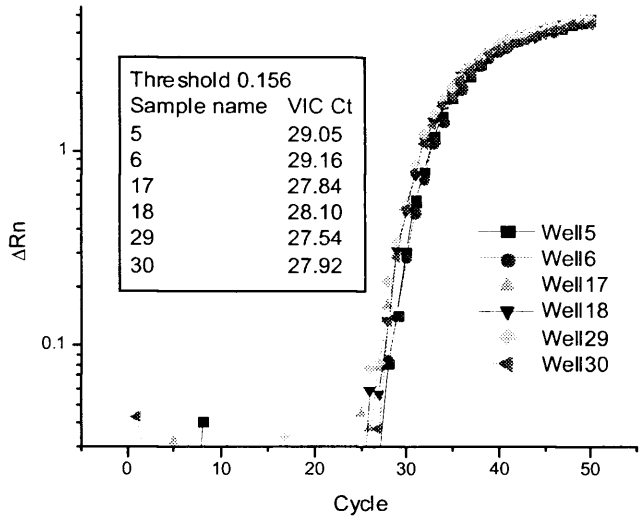


Figure 6.6 VIC signals from three samples obtained from incubating nuclear extract with plasmid cut by NcoI and treated with SAP (in duplicate). The different Ct values indicate differences in total RNA extraction efficiencies.

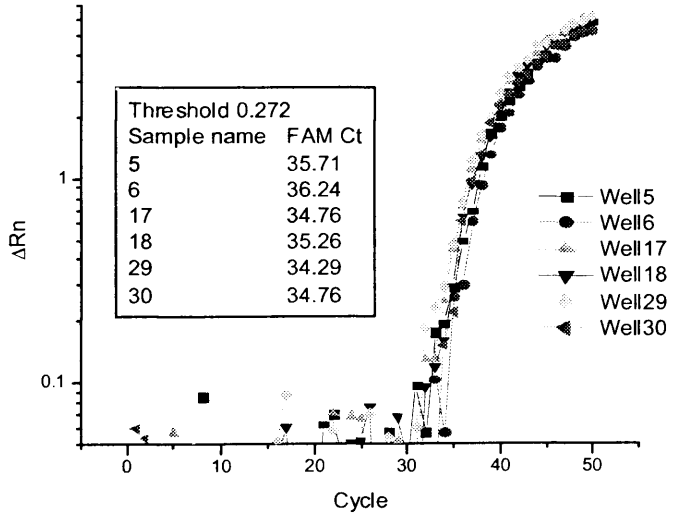


Figure 6.7 FAM signals obtained from three samples obtained from incubating nuclear extract with plasmid cut by NcoI and dephosphorylated by SAP (in duplicate).

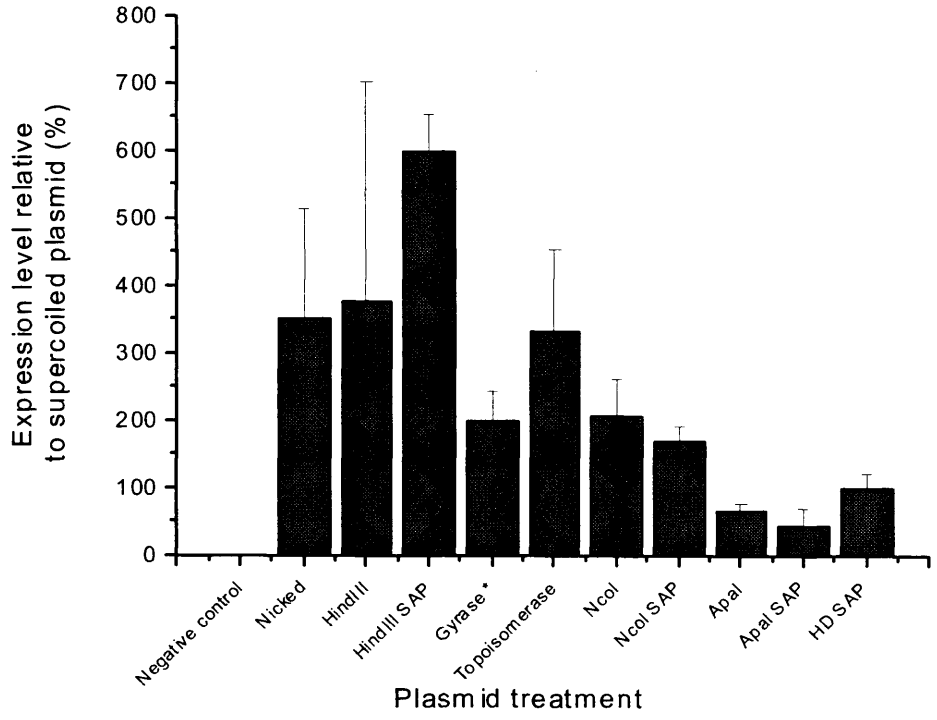


Figure 6.8 Expression level of different plasmid isoforms relative to supercoiled plasmid in the *in vitro* transcription system (%). Three replicates. Error bars denote standard error.

The results indicate that all plasmid isoforms give rise to mRNA of the gene of interest. Both relaxed forms of the plasmid, nicked and unwound, were found to exhibit the same expression level. This was to be expected since physiologically these forms are almost identical. Around 350% (±95% Nicked NBstNI and ±71% Unwound topoisomerase) of the amount of mRNA of the gene of interest was produced compared to that made using the supercoiled plasmid. The plasmid linearised in the vector backbone was also found to express at around this value although there was much greater variation in these results.

Chapter 7:

Discussion and further work

7 Discussion, conclusions and further work

7.1 Project aim

The overall aim of the project was to see if the presence of non Watson-Crick structures would significantly effect plasmid production for the purposes of producing plasmid gene therapy or DNA vaccine products.

7.2 Addressing project aims

In order to determine if unusual structures would effect plasmid production, a vector, [pQR334], was designed and produced that was 16 Kb, close to the size of the envisaged future products.

Cloning with the 16 Kb plasmid proved troublesome due to a lack of cloning sites, lower transformation efficiencies and the lack of a screening mechanism. In the interim unusual sequences were cloned into a smaller vector [pBluescript II KS +]. Cloning of non Watson-Crick type structures was problematic; however a family of plasmids was produced including an intrinsic bend, a quadruplex, a triplex and Z-DNA.

The growth rate of plasmid containing cells, total plasmid yield, and topology (i.e. the relative proportions of supercoiled plasmid, open circular and linear forms), were determined for each construct, and the stability of the inserted sequences were assessed using gel electrophoresis.

After analysis of the Bluescript based family, the triplex sequence was cloned into the 16 Kb parental plasmid [pQR334] creating [pQR338]. This was to see if the effect upon topology that was observed in [pBS840], could be observed in a larger plasmid. Inadequate purification techniques were used for these larger plasmids however and so meaningful comparisons between [pQR334] and [pQR338] were not achieved.

Analysis of the effect of fermentation time on plasmid linking number was also undertaken.

The effect of different plasmid topological forms on transcription in a cell free system was studied using a prospective plasmid product.

7.3 Effect of non- Watson Crick structures on plasmid production

DNA structure and abnormalities are reviewed in Chapter 1. Many non Watson Crick structures are extremely unlikely to occur in plasmid products. Indeed the view could be taken that non Watson Crick structures are encountered so infrequently in coding sequences, that it is unlikely that any unusual structures will be included in plasmid products. While the functions of different non Watson Crick structures remain obscure their deliberate incorporation into plasmid products also seems improbable.

Some non Watson Crick structures have previously been shown to be unstable in *Escherichia coli* production systems. Several non Watson Crick structures were included in a 2.9 kb plasmid at a defined point and their effect upon cell growth rate, total plasmid yield, and topology, were determined. The stability of the inserted sequences were assessed using gel electrophoresis.

Overall, yields, the differences in yield measurement and differences in growth rates of cells harbouring the plasmids, were not considered to be of major importance as relatively large amounts of plasmid were being produced in all cases apart from [pBluescript + CG].

Yields of the Bluescript based plasmids were found to be comparative, the exception being [pBluescript CG] which was significantly lower than that obtained with the other plasmids. This may either have been as a result of the generation and proliferation of plasmid free cells in the cell culture, or a general reduction in copy number throughout the bacterial population. This may account for the increased growth rate of DH1 when attempting to grow these plasmids; as cells containing few or no plasmids have a lower metabolic burden and hence may proliferate more rapidly.

Analysis of plasmid DNA topology over the course of cell cultivation indicated that the time of sampling had no significant effect on the relative proportion of supercoiled to other plasmid forms in the systems studied. If the percentage of the supercoiled form had been shown to be higher at a particular point of cell cultivation, this would have been a factor to consider in deciding harvest times. A previous study (O'Kennedy *et al* 2003) showed that the percentage of supercoiled species varied with the stage of cell growth and therefore had implications for the time of harvest. In that study the stain (DH5 α), the

plasmid [pSV β] and the medium (semi-defined) were different. Different growth media have been shown to produce major differences in the percentage of supercoiled plasmids obtained at the end of cell cultivation (O'Kennedy *et al* 2000). However, reasons for changes in the relative proportion of supercoiled to open circular and linear plasmids, with cell growth state in some media and not others, invites further research.

The triplex containing plasmid [pBS840] exhibited a >5% reduction in the proportion of supercoiled species compared to the other plasmids in the bluescript family. The most likely cause of this reduction is increased shear sensitivity due to non Watson-Crick structure.

A common misconception is that plasmid DNA products are sufficiently stable so as not to require cold storage (Tuteja 1999) and predictions of the stability of the supercoiled form supported this view (Middaugh *et al* 1998). DNA generally requires considerable chemical modifications in order to generate a loss in biological activity (Middaugh *et al* 1998). However, the requirement for cold storage has since been clearly demonstrated by a number of research groups (Evans *et al* 2000, Uden *et al* unpublished data).

The results for [pBS840], which demonstrated a significant decrease in supercoiled species following freeze thaw, indicate that care must be taken in the supply chain of plasmid products if the form of plasmids is shown to effect drug efficacy. The damage incurred to the triplex plasmid seems to indicate a higher propensity to physical damage than with the other plasmids. This may be because some of the DNA is single stranded. If a triplex had some function in a plasmid product then the effect of physical damage due to freeze thaw on the drug efficacy would have to be investigated.

Results suggest that Z-DNA is unstable in a batch DH1 *E.coli* production system grown in complex medium. Encouragingly other sequences studied (triplex, bend and quadruplex) did not cause spontaneous deletions, and no detrimental effect was found on growth rate or on total plasmid yield; indicating that such sequences could be included in future DNA products without any detrimental effect on plasmid yields; although the intra molecular triplex studied significantly decreased the proportion of supercoiled species.

The degree of supercoiling, (linking number) was demonstrated to alter depending on the harvest time in batch cell culture. The effect of different linking number distributions of drug efficacy has not been investigated, but if it does have an effect then the harvest time becomes important for industrial process. In general government agencies require homogeneous drug products, or at least repeatable batches, hence the harvest time in batch cell culture producing plasmids may be of importance.

7.4 Effect of topology on transcription in a cell free system

The transcriptional activity of a plasmid DNA vaccine in several topological forms was investigated *in vitro* using semi quantitative reverse transcriptase PCR.
Given the different distribution of linking numbers encountered from samples taken at different time during batch fermentation, it was desirable to determine if the linking number distribution affected transcription levels. In addition, literature searches showed few studies that compared linear, open circular and supercoiled plasmids.

After some experiments involving attempting to quantify the levels of RNA transcripts produced by sensitive dyes and using DIG labeling (data not shown), semi quantitative reverse transcriptase PCR was used on a plasmid for which the technique had already been optimized.

Relaxed plasmid (both nicked by *NBstNI* and unwound by topoisomerase) produced 350% (±95% and ±71% respectively) mRNA of the gene of interest compared to that produced by the supercoiled form. Plasmid linearised in the vector backbone produced similar mRNA levels to the relaxed plasmid.

All plasmid isoforms studied exhibited transcriptional activity, even those linearised and dephosphorylated within the promoter sequence (206% and 170% of supercoiled plasmid respectively) and those linearised and dephosphorylated in the open reading frame (64% and 43% of supercoiled plasmid respectively).

During processing, plasmids can become linearised in the plasmid backbone, in the promoter sequence or in the open reading frame of the gene of interest. By cutting plasmid with restriction enzymes in each of these three positions the effect of the site of linearisation was investigated. This is of particular interest if unusual DNA structures are

included in promoter or open reading frame regions, which may result in site directed shear events. The potential of site directed shear events to arise is discussed in Chapter 5. Somewhat surprising are the results from the plasmids cut with *NcoI* and *ApaI*. The *NcoI* restriction site is in the middle of the CMV promoter, and *ApaI* site is at the beginning of the open reading frame. A possible explanation for this is a DNA ligase acting in the nuclear extract and also perhaps a phosphorylating enzyme. This would result in the plasmids religating and so would allow transcription to occur at the observed low level. With respect to the plasmids cut with *NcoI*, the CMV promoter may exhibit some basal promoter activity after being cut. Alternatively the assay used may have been insufficiently quantitative.

The supercoiled plasmid was expressed at a lower value than might have been expected in comparison to the other plasmid forms. This is especially interesting given the wide spread belief that supercoiled form is 'better' than the other forms. One explanation for this is the need to unwind the DNA in order for the host transcription machinery to gain access to the genes (Hames and Higgins 1984). While supercoiled plasmid may be more easily inserted into cells, by the time the plasmid arrives at the cell nucleus it may be converted to the open circular form before being transcribed.

Integration of any DNA sequence from plasmid treatments into host chromosomal DNA is undesirable. Such insertions have the potential to form cancers by disrupting the structure or expressions of genes controlling cell growth and division (Robinson et al 1997). Investigations into integration of vaccine DNA into mouse chromosomal DNA failed to detect any integration events. One such study could have detected one integration event in 150,000 nuclei, a mutation rate estimated at 1,000 times less than the spontaneous mutation rate of DNA (Nichols et al 1995). Hence the risk of integration is minimal.

It has been shown that there is a greater chance (although still very small) of the linear form integrating into the host genome (Nichols et al 1995) than other plasmid forms. This is undesirable as it could potentially lead to detrimental gene mutations. Since studies have shown that the half-life of plasmid DNA is short (Lew *et al* 1995, Kim *et al* 2003). It seems likely that plasmid DNA is degraded within the transfected cell, and would become linear, no matter which form it started as. It could therefore be argued that the

greater risk of detrimental gene mutations from linear DNA should not be a reason for deciding on the best form for plasmid DNA.

Plasmid multimers are seen as undesirable since they may potentially produce plasmid inheritance instability. The inclusion of the cer sequence in plasmids has been shown to reduce the amount of multimeric plasmids produced (Summers and Sherrat 1984) and so is recommended for inclusion in plasmid products produced in *E. coli*.
The supercoiled plasmid form is also not one distinct species. Within the band corresponding to supercoiled plasmid is a distribution of plasmids some with more and with some fewer supercoils. These plasmids can be separated on the basis of linking number by using intercalator substances.

The reason given in the FDA guidelines at the start of this chapter concerning the specification for a minimum amount of supercoiled plasmid present in the product have been shown to be inaccurate. Both open circular forms and plasmid linearised in the plasmid backbone are transcribed more than supercoiled plasmid. However, it is clear that there is a need for DNA vaccines to fulfill specific requirements in order to ensure batch-to-batch reproducibility. One of these criteria is 'plasmid quality' the percentage of each form of plasmid. By setting criteria for the proportion of supercoiled plasmid in the final product the reproducibility of the purification procedure can be monitored. Since all plasmid isoforms are transcriptionally active, the criteria for 'plasmid quality' should be set on a case-by-case basis.

7.5 Further work

7.5.1 Evidence of the presence of unusual DNA structures

The major criticism of the paper submitted on the basis of results in Chapter 5 (Cooke *et al* 2004), was that no proof of the unusual structures was provided in the paper. To an extent it can be argued that such evidence was unnecessary since the sequences chosen had all already been shown to exhibit unusual structures; however, several other items of verification were available as discussed below.
As discussed at the end of Chapter 5, two dimensional gel electrophoresis of the full range of topoisomers would have provided comprehensive evidence of the presence of the unusual structures in the system studied. Unfortunately due to time limitations and

several failed experiments this was not possible, but would have provided some sense of scientific completeness.

7.5.1.1 c-MYC

There is direct evidence to showing that the nuclease hypersensitivity element III upstream of the c-MYC P1 promoter forms intramolecular quadruplexes (Siddiqui-Jain et al 2002, Simonsson et al 1998) and the motifs are published therein. Simonsson et al provide evidence of the quadruplex in a 2776 bp pUC based plasmid in *E.coli* DH5a. Great difficulty was encountered in sequencing [pBluescript + NHE]. The electropherograms of the sequencing reaction flat line at the point where the DNA quadruplex is expected to form. This is characteristic of strong secondary structure. Adequate sequencing data was only obtained after using a number of different additives known to remove DNA secondary structure. DMSO and betaine were tried before Lark managed to read through the sequence using proprietary additives.

7.5.1.2 [pBS840]

As described in Chapter 5, [pBS840] was found to have an average relative linking number difference 1.5 less than the parent plasmid [pBluescript II KS+] suggesting the formation of 16bp of triplex DNA (Glover *et al* 1990). [pBS840] has previously been shown to exhibit non-Watson Crick structure (Wang *et al* 1992) suggested to be " a structure similar to an intermolecular triplex ".

7.5.1.3 Coagulation factor

As described in Chapter 1, it has been well documented that alternating purine pyrimidine sequences form Z-DNA at levels of supercoiling encountered in *E.coli*. The choice of sequence was made so as to avoid simple repetitive sequences that have been shown to produce other structures as discussed in Chapter 2.

7.5.1.4 cdc2 promoter

The presence of an intrinsic bend in the cdc2 promoter was discovered using the circular permutation assay (Nair 1998). Nair also provides bioinformatic software predictions of the presence of a bend, but the accuracy of such predictions have yet to be fully tested.

No unusual gel electrophoretic mobility generally shown to accompany bent DNA sequences was observed in the experiments conducted for this project. Nair claims that the bending locus of the cdc2 sequence is at the *TaqI* site. Examining the sequence of [pBluescript cdc2] there are two *TaqI* sites in the cdc2 insert prepared by PCR. It is possible that there is slight variation in the sequence of the cdc2 promoter in the population, or that there was an error in sequencing, or spontaneous mutation in either the sequence cloned by Nair or the author.

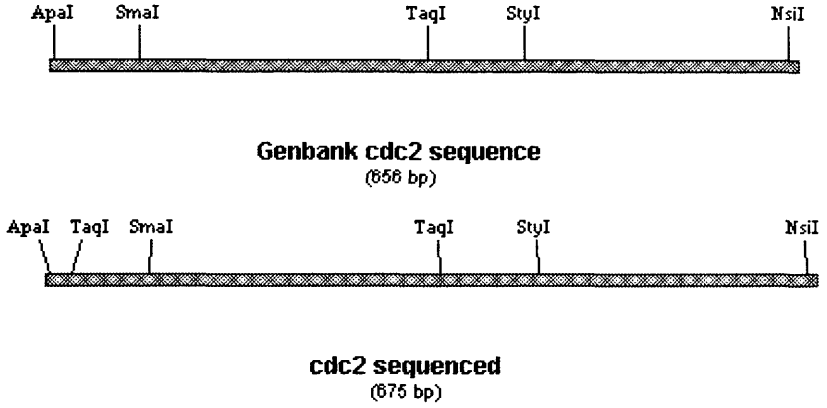


Figure 7.1 Comparison of restriction map between the cdc2 promoter given in Genbank (top) and sequenced from the clone produced in this project (bottom). The presence of two *TaqI* sites in the promoter used in this project may indicate that the promoter is not bent as has been previously reported.

Scrutinizing the circular permutation assay data obtained by Nair, the pattern obtained by cutting with *TaqI* may be a result of the presence of two *TaqI* sites being present in the sequence, rather than a single site in an intrinsically bent sequence.

In order to examine this possibility further PCR of the cdc2 promoter ought to be carried out from different sources and the products sequenced to see if two *TaqI* sites are present between the *NsiI* and *ApaI* sites. If the sequencing data shows the presence of two *TaqI* sites within the region then it would cast doubt over the existence of the bend reported by Nair. Significant differences between cdc2 promoters cloned from different sources would be of interest to biological anthropologists and palaeo geneticists.

In order to verify this finding, the circular permutation assay could also be carried out, perhaps using *StyI*, *HinfI*, *NarI* and *StuI*, with *TaqI* used for interest. The *HinfI* site is adjacent to the *TaqI* site, however it should also be ensured that this enzyme cuts the cdc2 promoter sequence only once. The published sequence shows two *HinfI* sites, while the cdc2 insert cloned for this project had only one site.

7.5.2 Shear damage to the triplex plasmid

As described in Chapter 5, attempts were made to determine the site of shear damage in the triplex containing plasmid [pBS840]. Further experiments would be required to see if the damage occurred at the triplex itself.

The effects on efficacy of the reduced percentage of supercoiled plasmid compared to other constructs would have to be investigated with a plasmid product that included a triplex sequence.

7.5.3 Effect of freeze-thaw on plasmid topology

Effects of freeze-thaw on plasmid topology have been known for some time by molecular biologists but has not been the subject of in depth study. In order to examine the effects in more detail plasmid could be repeatedly freeze-thawed with samples removed after each cycle for testing.

7.5.4 Effect of supercoiling level on plasmid production

As described in Chapter 1, it is possible to manipulate the levels of plasmid supercoiling, by heat shock, cold shock or by altering host DNA gyrase expression levels, perhaps in an inducible manner. Plasmids with different levels of supercoiling might be more or less susceptible to mechanical shear damage.

7.5.5 Effect of cell state on plasmid linking number distribution

In order to see the effects of cell state on linking number, chemostat culture could be used. This would generate plasmid with tight distributions of linking numbers as cells should be in the same physiological state. It has been suggested that linking number could be used as a sensitive assay to any changes in cell culture conditions (Uden and O'Kennedy personal communications).

7.5.6 Effect of plasmid topology on product efficacy

While limited in focus, the investigation into the effect of plasmid topology on transcription levels in a cell free system provided insight into the effects that different plasmid topologies will have on drug efficacy once delivered into the correct cellular compartment.

A series of experiments is being undertaken at GSK Beckenham, in order to determine the influence that plasmid form has on total drug efficacy. Initially experiments are being conducted using mammalian cells with each plasmid isoform, in order to establish transcription and translation levels over time, as well as how long each form remains in the cell. The effect of plasmid topology on the efficacy of different delivery methods could also be investigated *in vivo*.

References

References:

Actis LA, Tolmasky ME and Crosa JH: Bacterial plasmids: Replication of extrachromosomal genetic elements encoding resistance to antimicrobial compounds. Frontiers in Bioscience 3, 1998, d43-62

Albert AC, Roman AM, Bouche G, Leng M, Rahmouni AR: Gradual and oriented B-Z transition in 5'-untranscribed region of mouse ribosomal DNA. J Biol Chem. 1994, 29, 269, 30,19238-44

Alvarez M, Rhodes SJ, Bidwell JP: Context-dependent transcription: all politics is local. Gene, 2003, 313, 43-57

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, 15, 88, 2, 523-533

Arnott S, Chandrasekaran R and Marttila CM: Structures for polyinosinic acid and polyguanylic acid. Biochem J 1974, 141, 537-543

Allemand JF, Bensimon D, Lavery R and Croquette V: Stretched and overwound DNA forms a Pauling-like structure with exposed bases. Biophysics 1998, 94, 24, 14152-14157

Amara RR, Villinger F, Altman JD, Lydy SL, O'Neil SP, Staprans SI, Montefiori DC, Xu Y, Herndon JG, Wyatt LS, Candido MA, Kozyr NL, Earl PL, Smith JM, Ma HL, Grimm BD, Hulsey ML, Miller J, McClure HM, McNicholl JM, Moss B and Robinson HL: Control of a mucosal challenge and prevention of AIDS by a multiprotein DNA/MVA vaccine. Science. 2001, 292, 5514, 69-74

Bailey JE, Da Silva NA, Peretti SW, Seo JH and Srienc F: Studies of host-plasmid interactions in recombinant microorganisms. Ann N Y Acad Sci. 1986, 469, 194-211

Barnes EW, Cooke NJ, King AJ and Passmore R: Observations on the metabolism of alcohol in man. Br J Nutr 1965, 19, 4, 485-489

Bash RC, Vargason JM, Cornejo S, Ho PS and Lohr D: Intrinsically bent DNA in the promoter regions of the yeast GAL1-10 and GAL80 genes. J Biol Chem 2001, 276, 2, 861-866

Bauer WR: Structure and reactions of closed duplex DNA. Annu Rev Biophys Bioeng 1978, 7, 287-313

Bauer WR, Crick FHC and White JH: Supercoiled DNA. Sci Am 1980, 243, 100-113

Behe MJ: An overabundence of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. Nucleic Acids Res 1995, 23, 689-695

Birnboim HC and Doly J: A rapid alkaline extraction procedure for the screening of recombinant plasmid DNA. Nucleic Acids Res 1979, 7, 1513-1523

Bradley D: Cystic fibrosis treatment: a new era? PSTT 1999, 2, 7, 267

Brinton BT, Caddle MS and Heintz NH: Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. J Biol Chem 1991 266, 8, 5153-61

Bolshoy A, McNamara P, Harrington RE and Trifonov EN: Curved DNA without A-A: experimental estimation of all 16 wedge angles. Proc Natl Acad Sci USA 1991, 88, 6, 2312-2316

Bolshoy A and Nevo E: Ecologic genomics of DNA: upstream bending in prokaryotic promoters. Genome Res 2000, 10, 8, 1185-1193

Bossart J and Pearson B: Commercialization of gene therapy--the evolution of licensing and rights issues. Trends Biotechnol. 1995, 8, 290-294

Bowater R, Aboul-Ela F and Lilley DM: Two-dimensional gel electrophoresis of circular DNA topoisomers. Methods Enzymol 1999, 212, 105-120

Calladine CR: Mechanics of sequence-dependent stacking of bases in B-DNA. J Mol Biol 1982, 161, 2, 343-52

Canella KA and Seidman MM: Mutation spectra in supF: approaches to elucidating sequence context effects. Mutat Res. 2000, 450, 1-2, 61-73

Carlson, K., E. Raleigh, and S. Hattman. 1994. Restriction and modification. In J. Karam (ed.), Molecular Biology of Bacteriophage T4, pp. 369-381. American Society for Microbiology, Washington, D.C.

Catasti P, Chen X, Mariappan SV, Bradbury EM and Gupta G: DNA repeats in the human genome. Genetica 1999, 106, 1-2, 15-36

Catasti P, Chen X, Deaven LL, Robert KM, Bradbury EM and Gupta G: Cytosine-rich Strands of the Insulin Minisatellite Adopt Hairpins with Intercalated Cytosine+ ·Cytosine Pairs. J Mol Biol 1997, 272, 369-382

Catasti P, Gupta G,Garcia AE, Ratcliff R, Hong L, Yau P, Moyzis RK and Bradbury EM: Unusual structures of the random repetitive DNA sequences located at human centromeres: Biochemistry 1994, 33, 3819-3830

Caplen NJ: Cystic fibrosis gene therapy trials and tribulations. TRENDS in Molecular Medicine 2001, 7, 11, 488

Chan PP and Glazer PM: Triplex DNA: Fundamentals, advances, and potential applications for gene therapy. J Mol Med 1997, 75, 4, 267-282

Chastain PD and Sinden RR: CTG repeats associated with human genetic disease are inherently flexible. J Mol Bio 1998, 275, 405-411

Chatwin HM and Summers DK: Monomer-dimer control of the ColE1 P(cer) promoter. Microbiology 2001, 147, 11, 3071-3081

Chen X, Mariappan SV, Moyzis RK, Bradbury EM and Gupta G: Hairpin induced slippage and hyper-methylation of the fragile X DNA triplets. J Biomol Struct Dyn 1998, 15, 4, 745-756

Ciotti P, Van Dyke MW, Bianchi-Scarra G and Musso M: Characterisation of a triplex DNA-binding protein encoded by an alternative reading frame. Eur J Biochem 2001, 268, 225-234

Cohen AD, Boyer JD and Weiner DB: Modulating the immune response to genetic immunization. FASEB J 1998, 12, 1611–1626

Cooke GD, Cranenburgh RM, Hanak JA, Dunnill P, Thatcher DR and Ward JM: Purification of essentially RNA free plasmid DNA using a modified Escherichia coli host strain expressing ribonuclease A. J Biotechnol 2001, 85, 3, 297-304

Cooke JR, McKie EA, Ward JM and Keshavarz-Moore E: Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines. *Submitted* J Biotechnol 2004

Cooney M, Czernuszewicz G, Postel EH, Flint SJ and Hogan ME: Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. Science 1988, 241, 456-459

Cooper TB in *The Tools of Biochemistry* 194 – 253, Wiley, New York, 1977

Cothers DM, Haran TE and Nadeau JG: Intrinsically Bent DNA. J Biol Chem 1990, 265, 13, 7093-7076

Crews S, Ojala D, Posakony J, Nishiguchi J and Attardi G: Nucleotide sequence of a region of human mitochondrial DNA containing the precisely identified origin of replication. Nature 1979, 277, 192-198

Crystal RG, McElvany NG, Rosenfeld MA, Chu CS, Mastrangeli A, Hay JG, Brody SL, Jaffe HA, Eissa NT and Danel C: Administration of an adenovirus containing the human CFTR cDNA to the respiratory tract of individuals with cystic fibrosis. Nature Genetics 1994, 8, 42-51

Curling J and Smith L: Chromatogrpahy for Plasmid DNA Manufacture. Genetic Engineering News. 2003, 23, 11, 42

Dai Z, Dauchez M, Thomas G, Peticolas WL: Base sequence criteria and Cartesian coordinates for stable B/Z and B/Z/B junctions in relaxed DNA. J Biomol Struct Dyn 1992, 9, 6,1155-83

Davies J and Smith DI: Plasmid determined resistance to antimicrobial agents. Annu Rev Microbiol 1978, 32, 469-518

Davies JC, Geddes DM and Alton EWFW: Prospects for gene therapy for

cystic fibrosis. Molecular Medicine Today, 1998, July, 292-299

Davis NA, Majee SS and Kahn JD: TATA box DNA deformation with and without the TATA box-binding protein. J Mol Biol 1999, 291, 2, 249-65

Dlakic M and Harrington RE: The Effects of Sequence Context on DNA Curvature. Proc Natl Acad Sci U S A. 1996, 93, 9, 3847–3852

Dlakic M and Harrington RE: Unconventional helical phasing of repetitive DNA motifs reveals their relative bending contributions. Nucleic Acids Res 1998, 26, 18, 4274-4279

Deng Z, Kieser T and Hopwood DA: Expression of a Streptomyces plasmid promoter in Escherichia coli. Gene 1986, 43, 3, 295-300

Dickerson RE, Bansal M, Calladine CR, Diekmann S, Hunter WN, Kennard O, Lavery R, Nelson HCM, Olson WK, Saenger W, Shakked Z, Sklenar H, Soumpasis DM, Tung C-S, Von Kitzing E, Wang AH-J and Zhurkin VB: Definitions and nomenclature of nucleic acid structure components Nucleic Acids Res. 1989, 17, 1797-1803

Dickerson *et al.* in *Cold Spring Harbor Symposium for Quantitative Biology* 1982 v47, 13-24

Dickerson RE, Drew HR, Conner BN, Wing RM, Fratini AV and Kopka ML: The anatomy of A-, B-, and Z-DNA. Science 1982, 216, 4545, 475-85

Dieffenbach CW, Lowe TMJ and Dveksler: General concepts for PCR primer design. 133-142 in *PCR A Laboratory Manual* 1995, Cold Spring Harbour Laboratory Press, New York.

Dlakic M and Harrington RE: Bending and Torsional Flexibility of G/C-rich Sequences as Determined by Cyclization Assays. J Biol Chem 1995, 50, 299945-29952

Dorman CJ: Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. Trends Microbiol 1996 4 214-216

Even-Chen S and Barenholz Y: DOTAP cationic liposomes prefer relaxed over supercoiled plasmids. Biochim Biophys Acta 2000, 1509, 1-2, 176-88

Felsenfeld G, Davies DR and Rich A: Studies on the formation of two- and three- stranded polyribonucleotides. J Am Chem Soc 1957, 79, 2023-2024

Ferreira GN, Cabral JM and Prazeres DM: Development of process flow sheets for the purification of supercoiled plasmids for gene therapy applications. Biotechnol Prog 1999, 15, 4, 725-731

Forster T: Zwischenmolekulare Energiewanderung und Fluoreszenz. Ann Phys (Leipzig) 1948, 2, 55–75

Frackman S, Kobs G, Simpson D and Storts D: Betaine and DMSO: enhancing agents for PCR. Promega Notes, 1998, 65, 27

Frost R: From "The road not taken", Mountain Interval 1920, Henry Holt, New York

Fry M and Loeb LA: The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. Proc Natl Acad Sci USA 1994, 91, 11, 4950-4954

Fujimoto BS and Schurr JM: Dependence of the torsional rigidity of DNA on base composition. Nature 1990, 344, 175-177

Gabrielian AE, Landsman D and Bolshoy A: Curved DNA in promoter sequences. In Silico Biol 1999-2000, 1, 4, 183-196

Gadhavi PL, Greenwood MD, Strom M, King IA and Buxton RS: The Regulatory Region of the Human Desmocollin 3 Promoter Forms a DNA Four-Way Junction. Biochemical & Biophysical Research Communications 2001, 281, 520-528

Gagna CE, Kuo H and Lambert WC: Terminal differentiation and left-handed Z-DNA: a review. Cell Biol Int 1999, 23, 1, 1-5.

Gardiner EJ, Hunter CA, Packer MJ, Palmer DS, and Willett P: Sequence-dependent DNA structure: a database of octamer structural parameters. J Mol Biol. 2003 332, 5,1025-35

Geddes D and Alton E: Cystic fibrosis clinical trials. Advanced Drug Delivery Reviews 1998, 30, 205-217

Gehring K, Leroy JL and Gueron M: A tetrameric DNA structure with protonated cytosine.cytosine base pairs. Nature 1993, 363, 5615-65

Giaever GN and Wang JC: Supercoiling of intracellular DNA can occur in eukaryotic cells. Cell 1988, 2, 55, 5, 849-56

Giannobile WV, Lee CS, Tomala MP, Tejeda KM and Zhu Z: Platetlet-derived growth factor (PDGF) gene delivery for application in periodontal tissue engineering. J Periodontol 2001, 72, 6, 815-823

Glover JN, Farah CS and Pulleyblank DE: Structural characterisation of separated H DNA conformers. Biochemistry 1990, 29, 11110-11115

Goodsell DS and Dickerson RE: Bending and curvature calculations in B-DNA. Nucl Acids Res 1994, 22, 5497-5503

Gottschalk U and Chan S: Somatic gene therapy. Present situation and future perspective. Arzneimittelforschung. 1998, 48, 11, 1111-1120

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC and Gelbart WM (2000): An introduction to Genetic Analysis, Chapter 12 Recominant DNA technology. Seventh Edition, WH Freeman and Company, New York.

Hames BD and Higgins SJ (1984) Transcription and translation a practical approach. IRL Press, Oxford.

Hangerman PJ: Sequence-directed curvature of DNA. Annu Rev Biochem 1990, 59, 755-781

Harrington RE: DNA curving and bending in protein-DNA recognition. Mol Microbiol 1992, 18, 2549 - 2555

Hashem VI, Klysik EA, Rosche WA and Sinden RR: Instability of repeated DNAs during transformation in Escherichia coli. Mutat Res. 2002, 502, 1-2, 39-46

Hashem VI, Rosche WA and Sinden RR: Genetic assays for measuring rates of (CAG).(CTG) repeat instability in Escherichia coli. Mutat Res. 2002, 502, 1-2, 25-37

Henke W, Herdel K, Jung K, Schnorr D and Loening SA: Betaine improves the PCR amplification of GC-rich DNA sequences. Nucleic Acids Res. 1997, 25, 19, 3957-8

Hentschel CC: Homocopolymer sequences in the space of a sea urchin histone gene repeat are sensitive to S1 nuclease. Nature 1982, 295, 714-716

Hodges-Garcia Y, Hangerman PJ and Pettijohn DE: DNA ring closure mediated by protein HU. J Biol Chem 1989, 264, 14621-14623

Hoheisel JD and Pohl FM: Searching for potential Z – DNA in genomic Escherichia coli DNA. J Mol Biol 1987, 193, 3, 447-464

Hoogsteen K: The crystal and molecular structure of a hydrogen bonded complex between 1-methylthymine and 9-methyladanine. Acta Crystallogr 1963, 16, 907-916

Horn NA, Meek JA, Budahazi G and Marquet M: Cancer gene therapy using plasmid DNA: purification of DNA for human clinical trials. Hum Gene Ther 1995, 5, 565-573

Houk BE, Hochhaus G and Hughes JA: Kinetic modelling of plasmid DNA degredation in rat plasma. AAPS PharmSci, 1999, 1, 3, 9

Htun H and Dahlberg JE: Topology and formation of triple stranded H-DNA. Science 1989, 243, 1571-1576

Humes LE: Noise-induced hearing loss as influenced by other agents and by some physical characteristics of the individual. J Acoust Soc Am 1984, 76, 1318-1329

Kabsch W, Sander C and Trifonov EN: The ten helical twist angles of B-DNA. Nucleic Acids Res 1982, 10, 1097-1104

Kim BM, Lee DS, Choi JH, Kim CY, Son M, Suh YS, Baek KH, Park KS, Sung YC and Kim WB: In vivo kinetics and biodistribution of a HIV-1 DNA vaccine after administration in mice. Arch Pharmacal Res 2003, 26, 6, 493-498

Kiyama R and Camerini-Otero RD: A triplex DNA-binding protein from human cells: purification and characterization. Proc Natl Acad Sci USA 1991, 88, 10450-10454

Klinman DM, Yamshchikov G, Ishigatsubo Y: Contribution of CpG motifs to the immunogenicity of DNA vaccines. J Immunol 1997, 158, 3635-3642

Koehler DR, Hannam V, Belcastro R, Steer B, Wen Y, Post M, Downey G, Transwell AT and Hu J: Targeting transgene expression for cystic fibrosis gene therapy. Molecular Therapy 2001, 4, 1, 58-65

Koo HS, HM Wu and D Crothers: DNA bending at adenine-thymine tracts. Nature 1986, 320, 501-506

Lahijani R, Hulley G, Soriano G, Horn HA and Marquet M: High-yield production of pBR322-derived plasmids intended for human gene therapy by employing a temperature –controllable point mutation. Hum Gene Ther 1996, 7, 16, 1971-1980

Ledderberg J: Cell genetics and hereditary symbiosis. Physiological Reviews 1952, 32, 403-440

Lee, J.S, Woodsworth ML, Latimer JP and Morgan AR: Poly(pyrimidine)·poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. Nucleic Acids Res 1984, 12, 6603-6614

Le Pecq JB and Paoletti C: A new fluorometric method for RNA and DNA determination. Anal Biochem 1966, 1, 100-107

Leroy JL, Gueron M, Mergny JL and Helene C: Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. Nucleic Acids Res 1994, 22, 9, 1600-1606

Leslie AGW, Arnott S, Chadraekaran R and Ratliff RL: Polymorphism of DNA double helices. J Mol Biol 1980, 143, 49-72

Levy MS, Collins IJ, Yim SS, Ward JM, Titchener-Hooker N, Shamlou PA and Dunhill P: Effect of shear on plasmid DNA in solution. Bioprocess Eng 1999, 20, 1, 7-13

Levy MS, O'Kennedy RS, Ayazi-Shamlou P and Dunhill P: Biochemical engineering approaches to the challenges of producing pure plasmid DNA. Trends Biotechnol 2000, 18, 7, 296-305

Lew A, Rutter WJ and Kennedy GC: Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1 / MAZ. PNAS 2000, 97, 12508-12512

Lew D, Parker SE, Latimer T, Abai AM, Kuwahara-Rundell A, Doh SG, Yang ZY, Laface D, Gromkowski SH, Nabel GJ, *et al.*: Cancer gene therapy using plasmid DNA: pharmacokinetic study of DNA following injection in mice. Hum Gene Ther. 1995 6, 5, 553-564

Lilley DM: *In vivo* consequences of plasmid topology. Nature 1981, 292, 380-382

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

Lin X, Wang Z, Gu L and Deuel TF: Functional analysis of the human platelet-derived growth factor A-chain promoter region. J Biol Chem 1992,15, 267, 35, 25614-9

Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 2001, 25, 402-408

Marini JC, Levene SD, Crothers DM and Englung PT: Bent Helical Structure in Kinetoplast DNA. Proc Natl Acad Sci USA 1982, 79, 7664-7668; Correction Proc Natl Acad Sci USA 1983, 80, 7678

Mariappan SV, Garcia AE and Gupta G: Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. Nucl Acids Res 1996, 24, 4, 775-783

Mariappan SV, Catasti P, Silks III LA, Bradbury EM and Gupta G: The high-resolution structure of the triplex formed by the GAA/TTC triplet repeat associated with Friedreich's Ataxia. J Mol Bio 1999, 285, 2035-2052

Marilley M and Pasero P: Common DNA structural features exhibited by eukaryotic ribosomal gene promoters. Nucleic Acids Res 1996, 24, 2204-2211

Marmur J and Schildkraut CL: Growth of bacteria labelled with heavy isotopes for the isolation of nucleic acids. Naturwissenschaften 1961, 189, 636-638

Martín-Parras L, Lucas I, Martínez-Robles ML, Hernandez P, Krimer DB, Hyrien O and Schvartzman JB: Topologocal complexity of different populations of pBR322 as visualized by two-dimensional agarose gel electrophoresis. Nucleic Acids Res 1998, 26, 14, 3424-3432

Marvin DA, Spencer M, Wilkins MHF and Hamilton LD: The molecular configuration of deoxyribonucleic acid. III. X-Ray diffraction study of the C form of the lithium salt. J Mol Biol 1961, 3, 547-565

Matsuse T and Teramoto S: Progress in adeno virus-mediated gene therapy for cystic fibrosis lung disease. Current Therapeutic Res 2000, 61, 7, 422-434

Matthews KS: DNA looping. Microbiol Rev 1992, 56, 123-136

Middlaugh CR, Evans RK, Montgomery DL and Casimiro DR: Analysis of plasmid DNA from a pharmaceutical perspective. J Pharm Sci 1998, 87, 2, 130-146

Mirikin SM, Lyamichev VI, Drushlyak KN Dobrynin VN, Filippov SA and Frank – Kamenetskii MD: DNA H form requires a homopurine – homopyrimidine mirror repeat. Nature 1987, 330, 495-497

Mizuuchi K, Mizuuchi M and Gellert M: Cruiciform structures in palindromic DNA are favoured by DNA supercoiling. J Mol Bio 1982, 156, 229-243

Morrison A and Cozzarelli NR: Contacts between DNA gyrase and its binding site on DNA: features of symmetry and asymmetry revealed by protection from nucleases. Proc Natl Acad Sci USA 1981 78, 3, 1416-1420

Mountain A. Gene therapy: the first decade. Trends Biotech 2000, 18, 3, 119-128

Movileanu L, Benevides JM and Thomas GJ Jr.: Determination of base and backbone contributions to the thermodynamics of premelting and melting transitions in B DNA. Nucleic Acids Res 2002, 30, 17, 3767-77

Munteanu MG, Vlahovicek K, Parthasaraty S, Simon I and Pognor S: Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. Trends Biochem Sci 1998, 23, 9, 341-346

Murchie AI and Lilley DM: Retinoblastoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. Nucleic Acids Res 1992, 20, 49-53

Mytelka DS and Chamberlin MJ: Analysis and suppression of DNA polymerase pauses associated with a trinucleotide consensus. Nucleic Acids Res. 1996, 24, 14, 2774-27781

Nair TH: Evidence for intrinsic DNA bends within the human cdc2 promoter. FEBS Letters, 1988, 422, 94-98

Natale DA, Umek RM and Kowalski D: Ease of DNA unwinding is a conserved property of yeast replication origins. Nucleic Acids Res 1993, 21, 555-560

Ng H-L and Dickerson RE: Mildly eccentric 'E – DNA' Nat Struct Biol 2001, 8, 2, 107-108

Noirot P, Bargonetti J and Novick RP: Initiation of rolling-circle replication in pT181 plasmid: Initiator protein enhances cruciform extrusion at the origin. Proc Natl Acad Sci USA 1990, 87, 8560-8564

Oh YK, Kim JP, Yoon H, Kim JM, Yang JS and Kim CK: Prolonged organ retention and safety of plasmid DNA administered in polyethylenimine complexes. Gene Therapy, 2001, 8, 20, 1587-1592

O'Kennedy RD, Baldwin C and Keshavarz-Moore E: Effects of growth medium selection on plasmid DNA production and initial processing steps. J Mol Bio 2000 76, 175-183.

O'Kennedy RD, Ward JM and Keshavarz-Moore E: Effects of fermentation strategy on the characteristics of plasmid DNA production. Biotechnol Appl Biochem 2003 37, 83-90.

Oka A, Sugisaki H, Takanami M: Nucleotide sequence of the kanamycin resistance transposon Tn903. J Mol Biol 1981, 5147, 2, 217-26

Owerbach D and Aagaard L: Analysis of a 1963-bp polymorphic region flanking the human insulin gene Gene 1984, 32, 475-479

Packer MJ, Dauncey MP and Hunter CA: Sequence-dependent DNA structure: tetranucleotide conformational maps. J Mol Biol. 2000 295, 1, 85-103

Paddock GV, Abelson J: Nucleotide sequence determination of bacteriophage T2 and T6 species I ribonucleic acids. J Biol Chem 1975, 10, 250, 11, 4207-19

Pauling L and Corey RB: Structure of the nucleic acids. Nature 1953, 171, 346

Pauling L and Corey RB: A proposed structure for nucleic acids. Proc Natl Acad Sci USA, 1953, 39, 84-97

Pearson CE, Zorbas H, Price GB and Zannis-Hadjopoulos M: Inverted repeats, stem-loops, and cruciforms: significance for initiation of DNA replication. J Cell Biochem 1996, 61, 1-22

Perez-Martin J, Rojo F and de Lorenzo V: Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol Rev 1994, 58, 268-290

Radhakrishnan I and Patel DJ: Solution structure of a pyrimidine.purin.pyrimidine DNA triplex containing T.AT, C+.GC and T.TA triples. Structure 1994 2, 1, 17-32

Rahmouni AR and Wells RD: Stabilization of Z DNA in vivo by localized supercoiling. Science. 1989, 20, 246, 4928, 358-63

Raz E, Carson DA, Parker SE, Parr TB, Abai AM, Aichinger G, Gromkowski SH, Singh M, Lew D, Yankauckas MA, Bird SM and Rhodes GH: Indradermal gene immunization: The possible role of DNA uptake in the induction of cellular immunity to viruses. Proc Natl Acad Sci USA 1994 91 9519-9523

Ravin NV and Ravin VK: [Cloning of large imperfect palindromes in circular and linear vectors]. Genetika 1998, 34, 38-44

Ravin NV and Ravin VK: Use of a linear multicopy vector based on the mini-replicon of temperate coliphage N15 for cloning DNA with abnormal secondary structures. Nucleic Acids Res 1999, 27, 17, e13

Rees WA, Yager TD, Korte J, and von Hippel PH: Betaine can eliminate the base pair composition dependence of DNA melting. Biochemistry. 1993, 12, 32, 1, 137-44

Rhee S, Han Z-J, Liu K, Miles T and Davies DR: Structure of Triple Helical DNA with a Triplex – Duplex Junction. Biochemistry 1999, 38, 16810-16815

Riley S and Hough J: Business Watch. IMS America 1998 (Available on line at http://www.cpsnet.com/reprints/1998/05/buswatch.pdf)

Robinson HL, Ginsberg HS, Davis HL, Johnston SA and Liu MA: The scientific future of DNA for immunization. 1997 A report from The American Academy of Microbiology. (Available on line at http://www.asmusa.org/acasrc/aca1.htm)

Rosenstein BJ and Zeitlin PL: Cystic fibrosis. Lancet 1998, 351, 277-282

Rouviere-Yaniv J and Gros F: Characterization of a novel, low molecular weight DNA-binding protein from *Escherichia coli*. Proc Natl Acad Sci USA 1976, 72, 3428-3432

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning – A Laboratory Manual, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Sato Y, Roman M, Tighe H, Lee D, Corr M, Nguyen M, Carson DA, Raz E: Immunostimulatory DNA sequences necessary for effective intradermal gene immunization. Science 1996, 273, 352-354

Schatz T and Langiwski J: Curvature and sequence analysis of eukaryotic promoters. J Biomol Struct Dyn 1997, 15, 265-275

Schroth GP and Ho PS: Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. Nucleic Acids Res 1995, 23, 1977-83

Schroth GP, Siino JS, Cooney CA, Th'ng JP, Ho PS and Bradbury EM: Intrinsically bent DNA flanks both sides of an RNA polymerase I transcription start site. Both regions display novel electrophoretic mobility. J Biol Chem 1992, 15, 267, 9958-64

Schultz SC, Shields GC and Steitz TA: Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science 1991, 253, 1001-1007

Sen D and Guilbert W: Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 1988, 334, 364-366

Shamlou PA: Scaleable processes for the manufacture of therapeutic quantities of plasmid DNA. Biotechnol Appl Biochem 2003 37 207-218

Shampay J, Szostak JW, Blackburn EH: DNA sequences of telomeres maintained in yeast. Nature 1984, 12-18, 310, 5973,154-7

Shiver JW, Fu TM, Chen L, Casimiro DR, Davies ME, Evans RK, Zhang ZQ, Simon AJ, Trigona WL, Dubey SA, Huang L, Harris VA, Long RS, Liang X, Handt L, Schleif WA, Zhu L, Freed DC, Persaud NV, Guan L, Punt KS, Tang A, Chen M, Wilson KA, Collins KB, Heidecker GJ, Fernandez VR, Perry HC, Joyce JG, Grimm KM, Cook JC, Keller PM, Kresock DS, Mach H, Troutman RD, Isopi LA, Williams DM, Xu Z, Bohannon KE, Volkin DB, Montefiori DC, Miura A, Krivulka GR, Lifton MA, Kuroda MJ, Schmitz JE, Letvin NL, Caulfield MJ, Bett AJ, Youil R, Kaslow DC and Emini EA: Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. Nature 2002, 415, 6869, 331-335

Siddiqui-Jain A, Grand CL, Bearss DJ and Hurley LH: Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. PNAS 2002, 99, 18, 11593-11598

Siguret V, Ribba AS, Cherel G, Meyer D and Pietu G: Effect of plasmid size on transformation efficiency by electroporation of Escherichia coli DH5 alpha. Biotechniques, 1994, 3, 422-426

Silver S and Ozeki H: Transfer of deoxyribonucleic acid accompanying the transmission of colicinogenic properties by cell mating Nature, 1962, 195, 873-874

Simonsson T, Pecinka P and Kubista M: DNA tetraplex formation in the control region of c-*myc*. Nucleic Acids Res 1998, 26, 5, 1167-1172

Sinden RR in *DNA Structure and Function* 102-103, 224-225, 259-261 (Academic Press, London, 1994)

Sinden RR, Zheng G, Brankamp RG and Allen KN: On the deletion of inverted repeated DNA in *Escherichia coli*: Effects of length, thermal stability, and cruciform formation *in vivo*. Genetics 1991, 129, 991-1005

Smith HA and Klinman DM: The regulation of DNA vaccines. Current Opinion in Biotechnology 2001, 12, 299-303

Smith MCM, Burns RN, Wilson SE and Gregory MA: The complete genome sequence of the *Streptomyces* temperate phage ϕC31: evolutionary relationships to other viruses. Nucleic Acids Res 1999, 27, 10, 2145-2155

Smith FW and Feigon J: Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. Nature1992, 356, 6365, 164-168

Smith FW, Schultze P and Feigon J: Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. Structure 1995, 3, 10, 997-1008

Starr DM, Hoopes BC and Hawley DK: DNA bending is an important component of site specific recognition by the TATA binding protein. J Mol Biol 1995, 250, 4, 434-446

Stuhrmann M Dork T: CFTR gene mutations and male infertility. Andrologia 2000, 32, 2, 71-83

Summers DK: The origins and consequences of genetic instability in prokaryotes. Dev Biol Stand 1994, 83, 7-11

Suzuki K, Wakao N, Sakurai Y, Kimura T, Sakka A and Ohmiya K: Transformation of *Escherichia coli* with a Large Plasmid of *Acidiphilium multivorum* AIU301 Encoding Arsenic Resistance. Applied and Environmental Microbiology 1997, 2089-2091

Suzuki M and Yagi N: Stereochemical basis of DNA bending by transcription factors. Nucleic Acids Res 1995, 23, 2083-2091

Sykes RB and Mathew M: The beta-lactamases of gram-negative bacteria and their roles in resistance to beta-lactam antibiotics. J Antimicrob Chemother 1976, 2, 115-57

Tang DC, De Vit M and Johnston SA: Genetic immunization is a simple method for eliciting an immune response. Nature 1992, 356,152–54

Theodossiou I, Collins IJ, Ward JM, Thomas ORT and Dunhill P: The processing of a plasmid-based gene from E.coli. Primary recovery by filtration. Bioprocess Eng 1997, 16, 367-374

Trigueros S, Arsuaga J, Vazquez ME, Witt Sumners D and Roca J: Novel display of knotted DNA molecules by two-dimensional gel electrophoresis. Nucleic Acids Res 2001, 29, 13, e67

Tuma RS, Beaudet MP, Jin X, Jones LJ, Cheung CY, Yue S and Singer VL: Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. Anal Biochem. 1999, 268, 2, 278-88

Tuteja R: DNA vaccines: a ray of hope. Crit Rev Biochem Mol Bio 1999, 34, 1, 1-24

Uden M, Bailey S, Faulkes C, O'Hara J, Howard S, Johnson K, Molloy M, Ndlovu W and McKie E: DNA vaccine characterization. Poster DNA Vaccines Conference 2002

Ulanovsky L, Bodner M, Trifonov EN and Choder M: Curved DNA: Design, synthesis, and circularization. Proc Natl Acad Sci USA 1986, 83, 862-866

Ulmer JB, Donnelly JJ, Parker SE, Rhodes GH, Felgner PL, Dwarki VJ, Gromkowski SH, Deck RR, DeWitt CM, Friedman A, Hawe LA, Leander KR, Martinez D, Perry HC, Shiver JW, Montgomery DC and Liu MA: Heterologous protection against influenza by injection of DNA encoding a viral protein. Science. 1993, 259, 5102, 1745-1749

Ussery, D., Soumpasis, D.M., Brunak, S., Staerfeldt, H.H., Worning, P. and Krogh, A. Bias of Purine Stretches in Sequenced Chromosomes. Comput Chem 2002, 26, 531-541

Vallone PM and Benight AS: Thermodynamic, spectroscopic, and equilibrium binding studies of DNA sequence context effects in four 40 base pair deoxyoligonucleotides. Biochemistry. 2000 39, 26, 7835-46

Vargason JM, Eichman BF and Ho PS: The extended and eccentric E – DNA structure induced by cytosine methylation or bromination. Nat Struct Biol 2000, 7, 758-761

Vieira J and Messing J: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 1982, 19, 3, 259-68

Vinograd J, Lebowitz J, Radcliff R, Watson R and Laipis P: The twisted circular form of polyoma viral DNA. Proc Natl Acad Sci USA 1965, 53, 1104-1111

Vologodskii AV: Circular DNA, (2000), in Bloomfield V, editor, Chapter 13: Supramolecular Assemblies, De Felice, editor-in-chief, *Biophysics Textbook On-Line*, Biophysical Society, Bethesda, MD (www.biophysics.org/biophys/society/btol)

Walther W, Stein U, Voss C, Schmidt T, Schleef M and Schlang PM: Stability analysis for long-term storage of naked DNA: impact on nonviral in vivo gene transfer. Anal Biochem 2003, 318, 2, 230-235

Wang JC Peck LJ and Becherer K: DNA supercoiling and its effects on DNA structure and function. Cold Spring Harbour Symp Quant Biol 1983, 47, 85-91

Wang Y and Patel DJ: Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. Structure,1993, 1, 4, 263-82

Wang Y and Patel DJ: Solution structure of the Tetrahymena telomeric repeat d(T2G4)4 G-tetraplex. Structure, 1994, 2, 12, 1141-56

Wang Y and Patel DJ: Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. J Mol Biol, 1995, 251, 1, 76-94

Wang Z-Y, Lin X-H, Nobyuoshi M, Qiu Q-Q and Deuel TF: Binding of Single – Stranded Oligonucleotides to a Non-B-form DNA Structure Results in Loss of Promoter Activity of the Platelet-derived Growth Factor A-chain gene. J Biol Chem, 1992, 267, 19, 13669-13674

Warren GJ and Green RL: Comparison of the physical and genetic properties of palindromic DNA sequences. J Bacteriol, 1985, 161, 1103-1111

Watson J and Crick F: Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature, 1953, 25, 171, 737-738

Weintraub H, Cheng PF and Conrad K: Expression of transfected DNA depends on DNA topology. Cell, 1986, 46, 1, 115-122

Weissensteiner T and Lanchbury JS: Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. Biotechniques, 1996, 21, 6, 1102-1108

Wells RD, Collins DA, Havey JC, Shimizu M and Wohlrab F: The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. FASEB J 1988, 2, 14, 2939-2949

White JH: Self-linking and the Gauss integral in higher dimensions. Am J Math 1969, 91, 693-728

Widlund HR, Kuduvalli PN, Bengtsson M, Cao H, Tullius TD and Kubista M: Nucleosome Structural Features and Intrinsic Properties of the TATAAACGCC repeat sequence. J Biol Chem 1999, 274, 31847-31852

Widlund HR, Prasad NK, Bengtsson M, Coa H, Tullius D and Kubista M: Nucleosome Structural Features and intrinsic properties of the TATAAACGCC Repeat Sequences. Journal Biological Chemistry 1999, 274, 31847-31852

Williams WL and Müller UR: Effects of palindrome size and sequence on genetic stability in the bacteriophage ϕX174 genome. J Mol Bio 1987, 196, 743-755

Williamson JR: G-quartet structures in telomeric DNA. Annu Rev Biophys Biomol Struc 1994, 23, 703-730

Wolff JA, Malone RW, Williams P, Chong W, Acsadi G, Jani A and Flegner PL: Direct gene transfer into mouse muscle in vivo. Science, 1990, 247, 1465-1468

Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW: Quantitative evaluation of Escherichia coli host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucleic Acids Res 1989, 11, 17, 9, 3469-78

Wu H and Crothers DM: The locus of sequence-directed and protein induced DNA bending. Nature 1984, 308, 509-513

Xie TD and Tsong TY: Study of mechanisms of electric field-induced DNA transfection. V. Effects of DNA topology on surface binding, cell uptake, expression, and integration into host chromosomes of DNA in the mammalian cell. Biophys J, 1993, 65, 4, 1684-9

Yang Y, Westcott TP, Pedersen SC, Tobias I and Olson WK: Effects of localized bending on DNA supercoiling. TIBS 1995, 20, 313-320

Young HK: Do nonclinical uses of antibiotics make a difference? Infect Control Hosp Epidemiol 1994, 15, 7, 484-7

Zacharias W, Jaworski A and Wells RD: Cytosine methylation enhances Z – DNA formation in vivo. J Baceriol 1990, 172, 6, 3278-83

Zheng G and Sinden RR: Effect of base composition at the center of inverted repeated DNA sequences on cruciform transitions in DNA. J Biol Chem 1988, 263, 11, 5356-6461

Zimmerman SB, Cohen GH and Davies DR: X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. J Mol Biol 1975, 92, 181-192

Appendices

Appendix 1: C –, D – , 'E - ', H –, P – and T – DNA

C – DNA should not be confused with cDNA (below). At high levels of humidity of 57-66% C – DNA is formed in fibres and with a helical repeat of 9.3 bp per turn with a diameter of 1.9nm (Marvin *et al* 1961). It seems likely that this form does not occur *in vivo*.

cDNA, meaning complementary DNA, is the term applied to DNA produced from RNA by reverse transcriptase (and hence contains no introns). Initially single stranded, the cDNA is converted to double stranded DNA by DNA polymerase. Depending on sequence and environmental conditions cDNA could adopt any of the forms described herein.

D – DNA may be formed by alternating purine and pyrimidine sequences. D – DNA has a helical repeat of 8.5 bp (Arnott *et al* 1974).

There have been two different reports of 'E – DNA'. The first of these was by Arnott in 1980 and is in fact B – DNA with a 48° twist angle (Leslie *et al* 1980). The second by Vargason *et al* in 2000 was described as a new form of DNA (Vargason *et al* 2000). The analysis program used in determining the crystalline helical parameters for the 'E – DNA' however, uses unconventional reference frames and different definitions than decided upon in the Cambridge meeting of 1988 (Dickerson *et al* 1989). As explained by Ng and Dickerson, the structures described by Vargason *et al* are in the range between A – and B – DNA. DNA often lies within this range due to the ribose ring conformation in different solution conditions, as previously described (Ng and Dickerson 2001).

H – DNA is the term originally used to describe triple stranded DNA (Mirikin *et al* 1987) which is discussed in detail in chapter one.

P – DNA is a very unusual structure in which the phosphate backbone lies within the DNA with the bases on the outside of the structure. This conformation of DNA was first suggested before the Watson and Crick model was published, although involved three and not two strands (Pauling and Corey 1953). Attaching one end of a DNA strand to a flat surface, and the other to a magnetic bead Allemand *et* al (Allemand *et al* 1988) were able to produce some DNA adopting the Pauling structure. This was achieved by using magnets to control the position of the bead, hence stretching and twisting the DNA. A force of 3pN was found to be required for an 18bp sequence. Whether or not DNA will be exerted to forces of this magnitude in nature is open to debate. However, P – DNA

has been found *in vivo* in the bacteriophage Pf1 (Liu and Day 1994). The P – DNA form is believed to be the most elongated in existence.

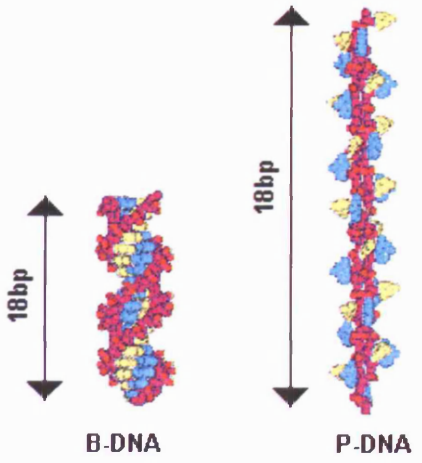


Figure A1.1 Structure of P-DNA deduced from molecular modelling. Space-filling models of a (dG)18·(dC)18 fragment in B-DNA (Left) and P-DNA (Right) conformations. The backbones are coloured purple, and the bases are coloured blue (guanine) and yellow (cytosine). Figure adapted from Allemand *et al* 1998

T – DNA is found from some bacteriophages (T2, T4 and T6) and has a most unusual structure. Instead of the usual cytosine base, T – DNA contains an α-Glucosylated derivative of 5 – Hydroxymethylcytosine that is often glycoslyated (Carlson *et al* 1994). This protects the phage DNA from nucleases that break down host nucleic acids during infection. The modified cytosine base produces a drastic change to the normal B-DNA structure, producing DNA with 8 bp per turn called T-DNA (Paddock and Abelson 1975).

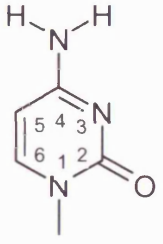
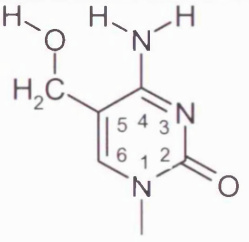
Cytosine	5 – Hydroxymethylcytosine
	

Figure A1.2 Comparison of cytosine and 5-hydroxymethylcytosine

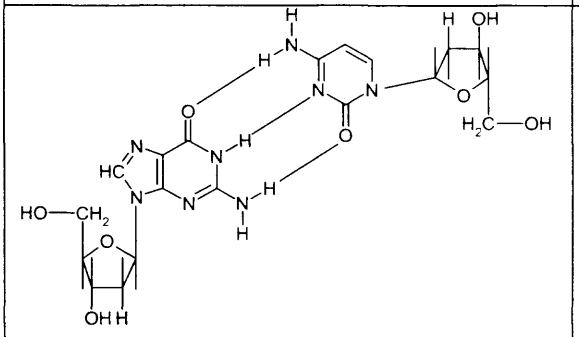
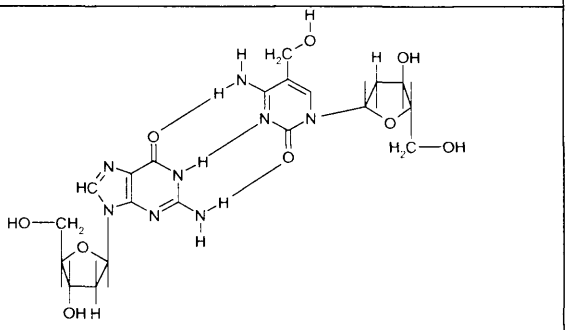
G-C	G-hydroxymethylcytosine
	

Figure A1.3 Comparison of the base pairing of Guanine – Cytosine and of Guanine - 5 hydroxymethylcytosine. Due to the presence of the hydroxymethyl group on the cytosine base, DNA containing 5 hydroxymethylcytosine adopts a structure with 8 bp per helical turn rather than the usual 10.5 bp.

Another use of the term T – DNA is applied to the transfer of DNA for the purposes of generating transgenic plants. As for cDNA, the structure of this would be dependent upon sequence and environment.

Appendix 2: Commercialisation of technology

A2.1 Introduction

During the course of my PhD I was given the opportunity to take the 'New Technology Ventures' second year module from the Master Of Business Administration course at London Business School. The Centre for Scientific Enterprise funded this period of study and this section is a requirement of my attendance on the course. The need to maximise product patent life is discussed with particular emphasis to research into plasmid-based products.

A2.2 The patent system

In order to encourage technological discoveries, world governments generally recognise the patent system. A granted patent provides the holder with a 20 year monopoly on the new invention for commercial application.

Within two months of coming off patent, generics typically reduce sales of the original product by 70 to 80% (IMS data).

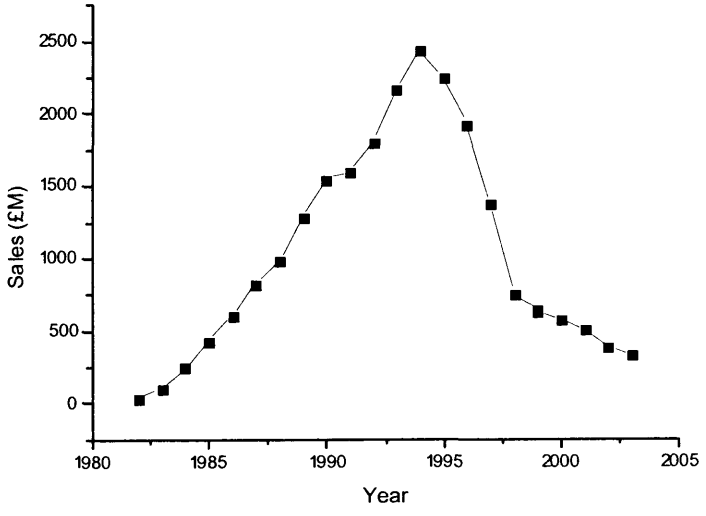


Figure A2.2: Sales of Zantac, anti ulserant drug from 1982 - 2003. Zantac patent expired April 1997. Source Glaxo Holdings, Glaxo Wellcome and GlaxoSmithKline company reports.

With patent life set at 20 years, there is a need to minimise the time between the patent application and drug launch, in order to maximise the monopoly.

In an attempt to compensate for the future patent expiry on key products looming, there has been a significant increase in R&D spending by the industry.

A2.3 R&D investment

Investment in research and development in the pharmaceutical industry as a whole has been rising at considerable rate since the 1980s, with no significant increase in the numbers of new products released annually.

While pharmaceuticals are relatively inexpensive to produce, the costs involved in getting a drug to market are very high. It is estimated that, including the costs of failures, a successful product currently costs over $1000m to develop (Deutsche Bank), although this is set to fall with tax exemptions on some research costs.

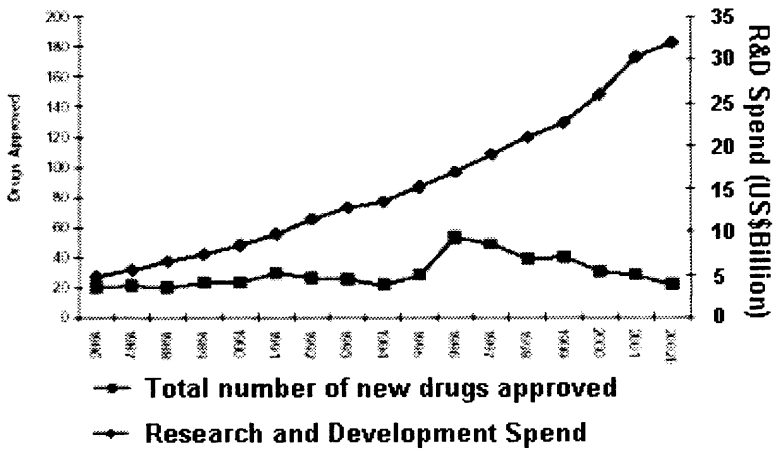


Figure A2.3 R&D expenditures on ethical pharmaceuticals, 1980 - 2002. Deutsche Bank Estimates

Some of the reasons for the rise in spending include:

- large cash flows from previous products and the threat of patent expiry
- an increase in the amount of clinical trial data required before drug approval
- an increase in the costs of producing new entities to test
- investment in new technology platforms

Pharmaceutical companies are no longer solely relying on methods, nor methods of action that produced many of the blockbuster drugs in the past. New drug discovery

methodologies and new drug classes require more upfront investment than with established methods. The view across the investment industry is that as the human genome becomes more fully understood, new drug targets will be discovered.

A2.4 DNA vaccines

One technology platform that has received significant investment is DNA vaccines. These represent a step change in vaccination against viruses. They have several advantages over more conventional treatments. Production of DNA vaccines does not require cultivation of dangerous or infectious agents, and unlike some viral vectors that have been used for gene delivery there is no risk of an attenuated virus back mutating and recreating the virulent form. DNA vaccines may also provide treatment against diseases for which the production of attenuated virus has so far proven unsuccessful e.g. HIV. There is also the potential to co-administer treatments against a number of diseases, on a single or multiple plasmid vectors.

Manufacture of pharmaceutical grade plasmid is envisaged to be a generic process that would be substantially less expensive than for other therapies:

Proteins consist of 22 amino acids with a range of characteristics. Hence proteins are molecules with substantially different physico-chemical properties such as hydrophobicity, temperature stability, pH stability, ionic strength, sensitivity to metal ions and sensitivity to protease. The extensive differences in the physical properties of proteins have lead to the requirement for a wide range of different methodologies to be applied to the purification of individual proteins.

Plasmids on the other hand are made of only 4 nucleotides with very similar properties. It has therefore been assumed that all high copy number plasmids below around 15 kb will be produced using the same method each time. Lower copy number and larger plasmids may require different purification protocols (Shamlou, 2003).

A2.5 Commercial application of this research

A strong incentive for investment into DNA vaccines are that, if they are shown to work, then the same production methods will be able to be employed to produce products vaccinating against a variety of diseases. They also have the potential to work alongside more conventional therapies increasing their effectiveness (Shiver *et al* 2002, Amara *et al* 2001). The results from Chapter 5 show that certain sequences of DNA

should be avoided in plasmids produced in an *E.coli* production system. Sequences put into plasmids could be easily screened using bioinformatics software for regions with the potential to form Z-DNA and cruciform structures. Identifying and avoiding such sequences at an early stage of research ought to allow the resources of pharmaceutical companies to be used more effectively. Following publication of the bulk of the results given in Chapter 5, this information is now freely available across the industry (Cooke *et al* 2004).

It is likely that progress with gene therapy and/or DNA vaccine research will depend upon collaboration between a number of companies. This is because different firms own the rights to many of the core technologies, including delivery mechanisms, production systems, formulation methods, delivery vectors and gene sequences (Bossart and Pearson 1995). By publishing results demonstrating research interests, GSK maintain their reputation as an attractive business partner to prospective collaborators.

Continued research with the sequences used during the project might allow alternative production and purification techniques to be developed, which might be required if future products are to include unusual structural features. Any such developments could be the subject of patents.

As plasmid products are a new class of drug, the FDA will require a large amount of product information before approving products. With particular respect to DNA vaccines and gene therapy products, considerable research has been carried out into producing plasmids in the supercoiled form. The results presented in Chapter 6 however, suggest that this might not be the most effective form. Pre-clinical studies are required to determine the most effective forms with given plasmid delivery methods, in order to give products the best chances of success in later trials.

Appendix 3: Nomenclature

ds - double stranded

μ - cell growth rate

ΔLk - Linking number difference

μ_{max} - maximum cell growth rate

CAP - Calf Alkaline Phosphatase

DIG - Digoxigenin

GSK - GlaxoSmithKline

Kan^R - Kanamycin resistance gene

LBS - London Business School

MCS - Multiple cloning site

NHE - Nuclease hypersensitivity element

OC - Open circular plasmid

OD - Optical density

PDGF - Platelet-derived growth factor

SAP - Shrimp Alkaline Phosphatase

SC - Supercoiled plasmid

t - time

t_d - doubling time

Tw - Twist number

UCL - University College London

WCW - Wet cell weight

Wr - Writhing number

x - cell number

x_o - cell number at time 0

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 114 (2004) 239–254
JOURNAL OF Biotechnology
www.elsevier.com/locate/jbiotec

Impact of intrinsic DNA structure on processing of plasmids for gene therapy and DNA vaccines

James R. Cooke[a], Elizabeth A. McKie[b], John M. Ward[c], Eli Keshavarz-Moore[a,*]

0168-1656/$ – see front matter
doi:10.1016/j.jbiotec.2004.06.011