

University of Southampton

AN ANALYSIS OF THE FOLDING OF
DNA THREE-WAY JUNCTIONS IN
RELATION TO A NOVEL ASSAY FOR
DNA DETECTION

Kerensa J. Jones

A thesis submitted for the degree of Doctor of
Philosophy

School of Biological Sciences

November 2005

UNIVERSITY OF SOUTHAMPTON
ABSTRACT
FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
SCHOOL OF BIOLOGICAL SCIENCES
Doctor of Philosophy
AN ANALYSIS OF THE FOLDING OF DNA THREE-WAY JUNCTIONS IN
RELATION TO A NOVEL ASSAY FOR DNA DETECTION
By Kerensa J. Jones

Three-way DNA junctions (3WJ) are composed of three mutually complementary strands of DNA. The strands form three double stranded helical arms which exchange at the branch point. These structures are important in recombination, slipped-strand DNA, RNA architecture and, more recently, for DNA detection in an isothermal assay (Signal Mediated Amplification of RNA Technology or SMART). In the presence of divalent metal ions, three-way junctions can stack with two arms forming a quasi-continuous helix with a third arm extended. There are two possible conformations this stacking can take depending on the sequence of the junction. This relationship is, however, not yet fully understood.

Understanding the dynamics of folding of DNA junctions, in particular the relationship between sequence and folding, allowed improvements to be made to the SMART assay. Sequences from the clinical setting of the SMART assay have been used to investigate three-way junction behaviour and to give an analysis of the folding of junctions in complex and simple systems.

By using gel retardation assays to look at the relative angles between arms in the junction, the effect of changing the sequence on the dominant conformation of the junction can be analysed. This technique has been used to show that sequences far from the junction branch point as well as those adjacent to it can have a large effect on the conformation of a 3WJ. Adding chemical linkers into the branch point can also have a dramatic effect, causing a junction to switch its dominant conformation despite no change in sequence. The extension of a primer within a 3WJ context has also been shown to be very dependent on the sequence and folding of a junction. This is thought to be due to the accessibility of the 3WJ branch point to DNA polymerase. This appears to be the major cause of a difference in signal strength produced in the SMART assay.

A novel technique was developed using redundant bases to generate every sequence combination around the branch point for a large-scale analysis. Junctions were separated by their mobility on a polyacrylamide gel into different conformers. The DNA in each band on the gel was purified, cloned and sequenced to give the identity of sequences which fall into each conformation. This analysis showed a number of sequences common to each band on the gel. When two sequences from different bands were synthesised, they separated well from each other and as predicted on a gel. They showed a strong tendency to correctly base pair even under rapid annealing conditions.

Melting studies were performed using the Roche LightCycler and junction stability was measured by a variety of different fluorescence methods; fluorophore and quencher pairs on same arm, SyBr Green (intercalates with double stranded DNA) and FRET donor and acceptor pairs on different arms. While both SyBr Green and FRET pairs provided useful data on the three-way junctions, placing the fluorophore and quencher on different pairs of arms provided the most useful melting data.

Data from these different techniques have been linked together to provide an analysis of the 3WJ in terms of sequence, buffer conditions, interaction with other enzymes, equilibrium and stability. Much of this work has been directly relevant to the SMART reaction. Novel techniques have been laid down for the future study of similar junctions.

Contents

Abstract	I
Contents	II
List of figures	VI
List of tables	XI
Abbreviations	XII
Acknowledgements	XIV
1. Introduction	1
1.1 DNA Junctions	2
1.1.1 The structure of DNA	2
1.1.2 Branched DNA Molecules	2
1.1.3 Naturally occurring three and four-way junctions	3
1.1.4 Three-way DNA junction structure	6
1.1.5 Introducing flexibility to the three-way junction	10
1.1.6 Conformational preference of the three-way junction	13
1.1.7 The effect of sequence on the conformation	13
1.2 DNA based diagnostics	18
1.2.1 Genetic diversity	18
1.2.2 Pathogen detection	19
1.2.3 Clinical DNA diagnostics tools	20
1.3 Signal Mediated Amplification of RNA technology	25
1.4 Split Promoter Amplification Reaction	28
1.5 Aims of the project	31
2. Materials and Methods	32
2.1 Reagents	33
2.1.1 Buffers	33
2.1.2 Oligonucleotides	34

2.2	5' end labelling of oligonucleotides	42
2.2.1	Gel electrophoresis	42
2.2.2	Gel purification of oligonucleotides	42
2.2.3	Elution and precipitation of DNA from a polyacrylamide gel	43
2.3	Annealing of oligos	43
2.3.1	Non-radiolabelled junctions	43
2.3.2	Radiolabelled junctions	43
2.4	Long-short arm assay	43
2.4.1	Staining with ethidium bromide	44
2.5	Chemical probing of three-way junctions	44
2.5.1	Diethyl pyrocarbonate	44
2.5.2	Osmium tetroxide	44
2.5.3	Cleavage of DNA with piperidine	45
2.5.4	Preparation of a GA ladder	45
2.6	DNase I protection	45
2.7	SMART reaction	45
2.7.1	Enzyme Linked Oligosorbant Assay (ELOSA)	47
2.7.2	DNA polymerase extension assays	47
2.7.3	Transcription assay	48
2.8	RNA purification	48
2.8.1	Quantification of RNA	49
2.9	Redundant junctions	50
2.9.1	Annealing and purification	50
2.9.2	Restriction digest with BamHI	51
2.9.3	Plasmid ligation	51
2.9.4	Transformation into <i>E. coli</i> (TG2)	52
2.9.5	Selection of transformants	52
2.9.6	Sequencing of the plasmid and insert	52
2.9.7	Analysis of the sequencing results	53

3.	Analysis of the conformational differences between three-way junctions	54
3.1	The conformation of DNA three-way junctions	54
3.2	MecA and Coag	56
3.3	Probing the junction for single stranded regions	59
3.4	Altering the central base pairs of the junction	65
3.5	Exploring the effect of sequence	71
3.6	Protection from cleavage around the branch point	77
3.7	Discussion	83
4.	Accessibility of the three-way junction to enzymes	87
4.1	Review of the SMART reaction	88
4.2.1	Effect of inserting extra linkers on the conformation of the 3WJ	96
4.2.1	Chemical probing of junctions 1:1 and 3:3	99
4.3	Re-designing the extension assay to SMART conditions	102
4.3.1	Extension under SMART assay conditions	106
4.3.2	Adding octanediols to prevent non-desired extension	114
4.3.3	Using alternative RNA polymerases	114
4.3.4	Changing the overlap region to prevent concatemerisation	115
4.4	Transcription reactions	117
4.5	Discussion	118
5.	Separating conformers in DNA three-way junctions	122
5.1	Designing the redundant junctions	123
5.2	Confirming the method	130
5.3	Other junctions	132
5.4	Discussion	135

6.	Melting studies of DNA junctions	137
6.1	Measuring the melting temperature of DNA junctions	138
6.2	Measuring the melting of individual arms in the junction -- fluorophore and quencher	139
6.2.1	Melting profile of a three way junction	142
6.2.2	Effect of the identity of the linker on junction stability	144
6.2.3	Stability of partial junctions	146
6.2.4	The effect of placing the fluorophore and quencher on a different arm of the junction.	148
6.2.5	The behaviour of a three-way junction under varying salt conditions	151
6.2.6	The effect of pH on the melting profile of the three-way junction.	154
6.2.7	The ability of the three and two-and-a-half-way junctions to recognise mismatches	156
6.3	SYBR Green binding to DNA	159
6.3.1	Melting profile of a three-way junction	160
6.3.2	Comparison of the melting of different junctions	163
6.4	Fluorescence Resonance Energy Transfer	166
6.4.1	Analysing the melting curves for the positions of the arms relative to each other.	168
6.5	Discussion	173
7.	General Discussion	176
	References	183

List of figures

Figure 1-1: A three and four way junction.	4
Figure 1-2: The formation and progression of a Holliday junction.	4
Figure 1-3: Helices in the VS ribozyme.	7
Figure 1-4: A four-way DNA junction.	9
Figure 1-5. Models of the unstacked three-way junction.	9
Figure 1-6: NMR structure of a folded three-way junction.	12
Figure 1-7: The model for the co-axially stacked, bulged three-way junction.	14
Figure 1-8: Isomers of the 3WJ.	14
Figure 1-9: Gel retardation studies of a bulged 3WJ.	15
Figure 1-10: FRET studies of 4WJ.	17
Figure 1-11: Binding of the probes in the presence and absence of the target.	27
Figure 1-12: Extension and transcription of the 3WJ.	27
Figure 1-13: SMART schematic showing the assay, amplification and detection steps.	29
Figure 1-14: SMART 3WJ and SPAR 2.5WJ junctions.	30
Figure 1-15: Transcription from the 2.5WJ.	30
Figure 2-1: SMART junctions.	34
Figure 2-2: Full length LSA junction.	37
Figure 2-3: Outline of the LightCycler junctions.	39
Figure 2-4: Light cycler junctions.	39
Figure 2-5: Position of mutations in the target strand.	40
Figure 2-6: Positions of the donor and acceptor molecules on the three-way junctions.	41
Figure 2-7: Sequences of the two FRET junctions.	41
Figure 2-8: Redundant junction.	51
Figure 3-1: The MecA and Coag junctions displayed with the position of the restriction sites and octanediol (blue circle).	57

Figure 3-2: The long-short arm assay of three-way junctions MecA and Coag with 1mM magnesium.	58
Figure 3-3: Control gels showing the effect of exposure of single stranded DNA to DEPC and osmium tetroxide.	60
Figure 3-4: 10 minute exposure of junctions MecA and Coag to DEPC.	62
Figure 3-5: 10 minute exposure of junctions MecA and Coag to osmium tetroxide.	63
Figure 3-6: Control gels showing the effect of using unpaired bases at the branch point rather than a chemical linker.	64
Figure 3-7: Comparison of the central ten base pairs of MecA, J1V6, MecAV1 and MecAV2.	67
Figure 3-8: Long-short arm assay using MecA and MecAV1.	68
Figure 3-9: Long-short arm assay using MecAV1 and MecAV2.	68
Figure 3-10: DEPC probing of MecAV1 and V2.	69
Figure 3-11: Exposure of MecAV1 and MecA V2 to osmium tetroxide.	70
Figure 3-12: The full sequences of junctions J1V6 and J1V6-2.	72
Figure 3-13: Radioactive version of the LSA assay for J1V6-2.	74
Figure 3-14: The full sequence of J1V6-3.	75
Figure 3-15: Non-radioactive LSA for J1V6 3.	75
Figure 3-16: Chemical probing of the J1V6-3 junction with DEPC and osmium tetroxide.	76
Figure 3-16: The cutting of double stranded DNA by DNaseI.	78
Figure 3-18: DNaseI digestion of junctions MecA and Coag.	79
Figure 3-19: Comparison of the cleavage of the MecA junction and equivalent duplex by DNaseI.	80
Figure 3-20: Exposure of junctions MecA V1 and MecAV2 to DNaseI.	82
Figure 3-21: Models of the three-way junctions suggested by conformation and chemical probing studies.	86
Figure 3-22: Model of the protection from DNaseI in a folded 4WJ.	86
Figure 4-1: Review of the extension and transcription steps of SMART.	89
Figure 4-2: Schematic representation of the DNA extension assay.	90

Figure 4-3: Sequence of the MecA SMART 3WJ.	92
Figure 4-3: Sequence of the Coag SMART junction.	93
Figure 4-5: Position of the octanediols in the modified MecA junctions.	94
Figure 4-6: Extension of MecA, Coag and the modified MecA junctions.	94
Figure 4-7: Diagram of the regions of DNA into which each lane is divided.	95
Figure 4-8: Fully complementary versions of junctions MecA, 1:1 and 3:3.	97
Figure 4-9: Long-short arm assay using MecA and junctions 1:1 and 3:3 containing extra octanediols at the branch point.	98
Figure 4-10: Model of the averaged conformations of junctions MecA, 1:1 and 3:3.	98
Figure 4-11: Chemical probing of junctions 1:1 and 3:3 with DEPC.	100
Figure 4-12: Chemical probing of junctions 1:1 and 3:3 with osmium tetroxide.	101
Figure 4-13: Outline of the SMART reaction from transcription of the first RNA to detection.	103
Figure 4-14: Sequence of junction JO plus facilitators.	104
Figure 4-15: Sequence of Junction J2.	105
Figure 4-16: Extension of Junction JO and J2 under full SMART conditions.	107
Figure 4-17: Extension of the <i>Salmonella</i> junction under varying assay conditions.	109
Figure 4-18: Extension of oligo 1848 (J2 extension) by RNA polymerase using rNTPs.	110
Figure 4-19: Hairpin loop structure formed by a short DNA template (Krup, 1989)	112
Figure 4-20 Possible dimers formed in oligos 1848 and 1734.	112
Figure 4-21: Extension of an oligo unrelated to the SMART sequences, by T7 RNA polymerase.	113
Figure 5-1: Design of junctions containing redundant bases.	126
Figure 5-2: The two possible conformers resulting from the folding of the	

junction in Figure 5-1.	126
Figure 5-3: Separation of the pool of junctions by polyacrylamide gel Electrophoresis	127
Figure 5-4: Cloning of the purified three-way junction into the pUC18 plasmid	127
Figure 5-5: Separation of junctions TA-CT and TA-GG.	131
Figure 5-6: Testing the ability of complementary junctions to base pair accurately when other similar partners are available.	131
Figure 5-7: Conformers formed by placing different bases at position N ₁ .	134
Figure 6-1: Illustration of how the melting of the DNA is measured by fluorescence.	140
Figure 6-2: Outline of the light cycler junctions.	141
Figure 6-3: Light cycler junctions.	141
Figure 6-4: Melting profile of a three-way junction and duplex of equivalent sequence.	143
Figure 6-5: The rate of change in fluorescence over increasing temperatures for a 3WJ and duplex.	143
Figure 6-6: Comparison of the melting profiles and melting temperatures of three-way junctions with different linkers.	145
Figure 6-8: Comparison of the melting profile and melting temperatures of a three-way junction, partial junctions and duplex.	147
Figure 6-9: Positions of the fluorophore and quencher at position 1 on arm A and position 2 on arm B.	149
Figure 6-10: Melting profiles of junctions measured from arm A and arm B.	150
Figure 6-11: The effect of varying the concentration of sodium and magnesium on the melting temperature and melting profile of a three-way junction.	153
Figure 6-12: Position of the mismatches in target strands Mut (mutation) 1,2 and 3.	157
Figure 6-13: Position of the mismatches in the target strand.	157
Figure 6-14: Melting temperatures of the 3WJ containing the hexaethylene glycol, the 2.5 and 2WJ.	158

Figure 6-14: Melting profiles, peaks and melting temperatures of MecA, a duplex and Coag measured using SYBR Green.	162
Figure 6-16: Melting profiles of related sets of junctions using SYBR Green.	164
Figure 6-17: Melting temperatures of junctions in 50mM sodium phosphate plus and minus magnesium.	165

List of tables

Table 2-1: Sequences of all the SMART junctions.	36
Table 2-2: Modified extension sequences for J2.	36
Table 2-3: Sequence of the long-short arm assay junctions.	37
Table 2-4: Sequences of the Light Cycler junctions.	38
Table 2-5: Sequences of the target strands in the Light Cycler junction.	40
Table 2-6: Dilutions of the SMART oligos.	46
Table 2-7: Dilution of the stock RNA to create a standard curve.	50
Table 4-1: The percentage of DNA extended after 30 mins using <i>Bst</i> DNA polymerase at 41°C	95
Table 4-2: Analysis of the conditions under which the shorter extension product is seen.	111
Table 4-3: Analysis of the interaction of various oligos with T7 and other RNA polymerases.	116
Table 5-1: Identity of bases N_1 and N_2 obtained from sequencing clones from the R-TA junction pool	129
Table 5-2: Sequencing results (N_1 and N_2) for the pools of junctions with the four different fixed base pairs.	133
Table 6-1: Melting temperatures of junctions measured from arms A and B.	149
Table 6-2: Melting temperatures recorded for 3WJ H over varying magnesium and sodium chloride concentrations.	152
Table 6-3: The effect of pH on the melting temperature of 3WJ H.	155

List of abbreviations

A	Adenine
AFM	Atomic Force Microscopy
AMPS	Ammonium persulphate
ARMS	Allele Refractory Mutation System
ATP	Adenosine triphosphate
C	Cytosine
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
ELOSA	Enzyme Linked Oligosorbant Assay
FRET	Fluorescence Resonance Energy Transfer
G	Guanine
LSA	Long-short arm assay
MALDI-TOF	Matrix Assisted Laser Desorption – Time of Flight mass spectrometry
MRSA	Methicillin Resistant <i>Staphylococcus aureus</i>
NMR	Nuclear Magnetic Resonance Imaging
Oligo.	Oligonucleotide
PCR	Polymerase Chain Reaction
R-Enase	Restriction endonuclease
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribose nucleic acid
rNTP	Ribonucleotide triphosphate
rpm	revolutions per minute
rRNA	Ribosomal RNA
SSCP	Single Strand Conformation Polymorphism
SMART	Signal Mediated Amplification Reaction
SNP	Single Nucleotide Polymorphism
SPAR	Split Promoter Amplification Reaction
TBE	Tris Borate EDTA

TBM	Tris Borate Magnesium
TE	Tris EDTA
T	Thymine
UTP	Uridine triphosphate
2.5WJ	Two and a half way junction
3WJ	Three-way junction
4WJ	Four-way junction

Acknowledgements

I would like to express my sincere gratitude to the following people:

Professor Keith Fox for his supervision and advice. The Fox lab past and present for making the long hours of lab work bearable, even if we couldn't agree on music taste.

Dr. Jo Seal and Dr. Don Cardy at British Biocell International for the funding and constant interest in the work. Dr. Emma Wright and Dr. Pete Marsh for the supervision at Cytocell.

Dr. Carly Brooks and Dr. Nadia Azzi for all the many scientific discussions about each other's work at home, over many bottles of wine. It has been an education in transgenics and organic chemistry as well as DNA junctions.

My parents for their inspiration and understanding. I am particularly grateful for all the financial and emotional support. My sisters Beth and Tara who now have a precedent set of becoming the fourth and fifth Dr Jones's in the family.

Lastly, I am forever grateful to Christian, for his love and patience. Thank you for helping with all the presentations and writing, and for letting me chat for hours about something that I'm sure made no sense to you. I am particularly grateful for all the computer help in my hour of need.

The work in this thesis was supported by a BBSRC CASE studentship with Cytocell and then with British Biocell International.

Chapter 1

INTRODUCTION

This chapter details the importance and structure of three-way DNA junctions and looks at the techniques used to study them. It also looks at DNA detection methods, including the SMART assay used during the research for this thesis. The aims and objectives of this project are outlined.

1 Introduction

1.1 DNA Junctions

1.1.1 The structure of DNA

The importance of DNA first became apparent after it was identified as the long sought after heritable particle by Oswald Avery in 1944 (Avery *et al.*, 1944, from 1979 reprint). Originally dismissed for its deceptively simple composition of just four bases, DNA became the centre of attention up to the present day.

Since the identification of the right-handed double helix of DNA by Watson and Crick in 1953, there have been a number of other structures identified. DNA has a very flexible backbone, and within the three billion base pairs in the human genome there are many possibilities for the formation of other structures. Examples of these different forms are the various helical forms of A and Z DNA, as opposed to the B form originally identified. These differ in the number of bases per turn (10 in B form, 11 in A form) and handedness (Z DNA is left handed). Not only are there variations on the helical form, there are folded structures which emerge from standard Watson-Crick base pairing and other connections such as Hoogsteen base pairing (reviewed in Ghosh and Bandel, 2003). These include junctions, quadruplexes and triplexes

1.1.2 Branched DNA molecules

Branched DNA structures are examples of folding of the DNA involving Watson-Crick base pairing. The folding produces branch type structures and can involve looped out DNA. They occur in a number of situations in both DNA and RNA. In DNA they are best seen in Holliday junctions, formed during homologous recombination of the chromosomes (Holliday, 1964). Like sequences exchange strands of DNA forming a four-way X type structure. The three-way junction, which

is the focus of this thesis, is found in DNA as a recombination intermediate similar to the Holliday junction. (Minagawa *et al.*, 1983; Pearson and Sinden, 1996; Pearson *et al.*, 1998). It is also more commonly found in the architecture of RNA species, which adopt much more complicated folding patterns (Steitz, 1992; Rettberg *et al.*, 1999; Lafontaine *et al.*, 2001). These are all discussed later in this chapter.

The nucleic acid strands involved in branched DNA structures are mutually complementary, base pairing with each other along their length. They run antiparallel to each other as they would in a simple duplex of DNA (Figure 1-1) and each arm forms a right-handed helix (Leontis *et al.*, 1994; Eichman *et al.*, 2000). Junctions have been constructed from short oligonucleotides, with anything from three to six arms (Wang *et al.*, 1991) but only three and four way junctions have been characterised *in vivo*.

1.1.3 Naturally occurring three and four-way DNA junctions

Four-way DNA junctions (4WJ), also known as Holliday junctions are involved in meiotic recombination between homologous chromosomes (Holliday, 1964). Strand exchange occurs between the two chromosomes and the structure moves along the DNA by a process called branch migration. Eventually, the junction is resolved into two chromosomes having interchanged a region of sequence (Figure 1-2). A similar event occurs during double strand break repair, where a sequence from the other chromosome can be used as a template to fill in the gap.

Branch migration and resolution are tightly controlled by a number of enzymes (reviewed in White *et al.*, 1997). The RuvABC complex is fundamental to meiotic recombination in *E.coli*. Structure is a key issue in the recognition of the 4WJ. RuvA binds to the junction as a tetramer. The surface of the protein is concave and electrostatically positive, allowing the DNA duplex arms to sit in grooves on the protein surface. It selects the face of the junction with the major groove at the crossover (Hargreaves *et al.*, 1998). Ruv B, an ATP-dependent motor then bind to the complex, driving branch migration. The junction is finally resolved by the

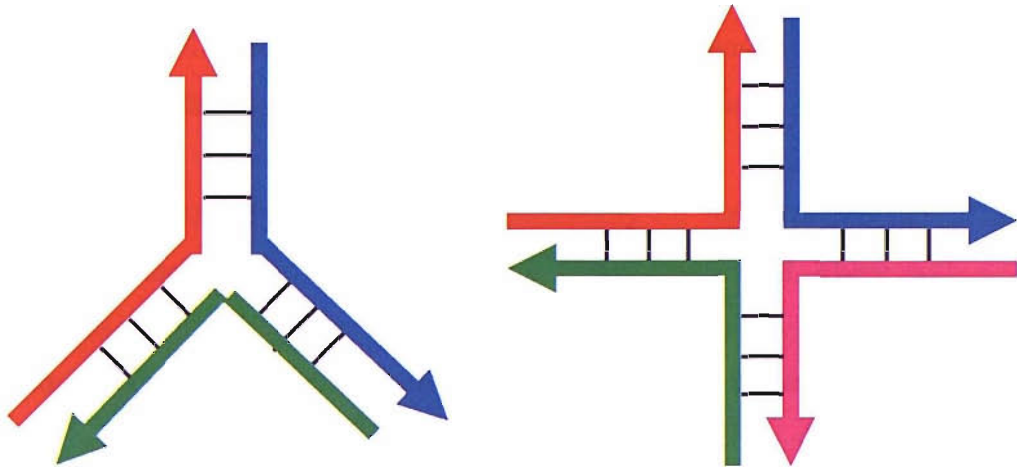


Figure 1-1: A three and four way junction. Branched DNA junctions are composed of mutually complementary strands running antiparallel to each other. A three and four way junction are shown with each strand involved a different colour. The direction of the arrows indicates the 5' to 3' direction of the strand. The black lines represent base pairing.

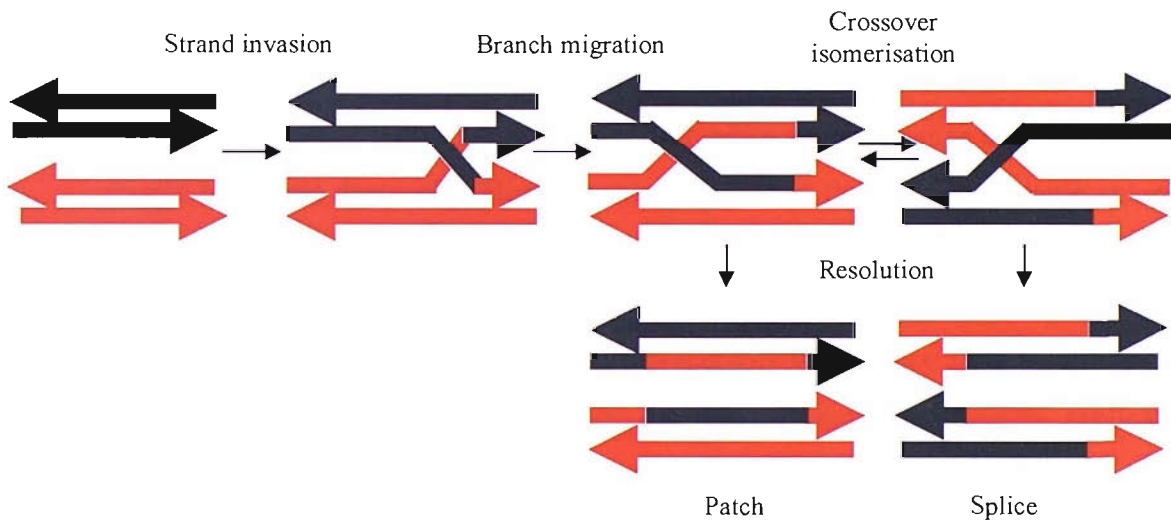


Figure 1-2: The formation and progression of a Holliday junction. A Holliday junction is created when two similar sequences are in close contact for example two homologous chromosomes paired together on the spindle during meiosis or the use of an identical sequence as a template to repair a break. Strand invasion occurs when one strand binds to the like sequence on the other duplex. The crossover structure moves along the DNA, swapping sequence in its wake. As an X-shaped structure, the Holliday junction can be resolved in one of two planes resulting in two different patterns of swapped sequence. When a patch occurs, each duplex is left with one original stand and one strand containing a short region from the other duplex. The splice results in a mosaic of sequence on all strands.

endonuclease RuvC. The intervention of these enzymes speeds up an otherwise very slow process and thereby promotes homologous recombination.

Other examples of enzymes which act specifically on the 4WJ are the junction resolving enzymes T4 endonuclease VII and T7 endonuclease I. These both act as homodimers to bind both sides of the junction then make a cut on either side. The two subunits do not appear to interact with each other to achieve this cutting but bind to the DNA for long enough for both cuts to occur in the lifetime of the complex (Parkinson and Lilley, 1997).

Three-way junctions (3WJ) are formed as intermediates in recombination as well as from slipped out DNA that occurs in long sections of trinucleotide repeats (Pearson and Sinden, 1996; Pearson *et al.*, 1998). These naturally occurring hairpin structures have been shown to migrate slowly on an acrylamide gel due to kinking of the DNA (confirmed by atomic force microscopy, Oussatcheva *et al.*, 1999). 3WJs and other structures which constitute slipped strand DNA are of interest because of their mutagenic potential. Trinucleotide repeat instability at certain sites is responsible for fragile Myotonic dystrophy, Huntington's disease and Fragile X syndrome amongst others (Cummings and Zoghbi, 2000; Nadel *et al.*, 1995). Slipped out secondary structures, stable 3WJs and hairpins among them, are usually recognised by DNA repair proteins. They may be mutagenic precursors where they accumulate and are not repaired, leading to expansion or deletion of these sites during replication (Pearson *et al.*, 2002).

DNA repair mechanisms do exist to combat this type of mutagenesis. The *E.coli* excision repair protein UvrA excises such structures from the DNA and may recognise the kink formed by hairpin loops and 3WJs (Oussatcheva *et al.*, 2001). Endonuclease VII from phage T4 and Endo X3 from *Saccharomyces cerevisiae*, are both high specificity for DNA junctions (Jensch and Kemper, 1986; Jensch *et al.*, 1989). The tumour suppressor protein, p53, has been shown to bind to Holliday and 3WJs with high specificity (Lee *et al.*, 1997). Binding increases the efficiency of resolution by T4 endonuclease VII and T7 endonuclease I. It is likely that this is part of a checkpoint control to prevent progression through the cell cycle without resolution of branched DNA molecules.

RNA has far more use for three- and four-way junctions than DNA. It often forms complex architecture in order to fulfil its more functional role in the cell. Rettberg *et al.*, (1999), demonstrated a role for a 3WJ in frameshifting in the bacterial insertion sequence IS911. The regulation of gene expression is tightly controlled at many levels. Specific sequences or folding patterns are often used to signal a change in regulation. The 1250 base pair transposable element has two partially overlapping open reading frames (ORFs), A and B. Orf A, and OrfAB are required for transposition. OrfB alone is also expressed but its function remains unknown. 10-15% of ribosomes switch to the -1 reading frame before the end of OrfA to produce the combined product. Internal initiation at the 2nd site is enhanced by a Shine-Dalgarno sequence and an element 3' to the start site. Deletion studies revealed the presence of the 3' element dependent on the SD sequence. Secondary structure prediction backed up by chemical and enzymatic cleavage experiments showed this region to be a 3WJ. The terminator of Orf A lies within the structure. The authors speculate that the precise structure of the 3WJ is needed for precise control of translation. It may also have evolved because the thermodynamic stability of the structure was right for the situation.

At the core of the VS ribozyme lie five helices, organised by two three-way junctions (Figure 1-3). Small mutations in sequence of junction 2-3-6 can lower the cleavage activity of the ribozyme, suggesting that the folding of the junction is very important and is controlled by the sequence (Lafontaine *et al.*, 2001). Four-way junctions also often occur in RNA architecture. The 10-23 DNA enzyme, which consists of two strands of DNA and two of RNA, contains two four-way junctions co-ordinating five double helical domains (Nowakowski *et al.*, 1999). There is also speculation that a branched DNA molecule mimicking a Holliday junction may be involved in bringing together the splice sites in pre-messenger RNA together (Steitz, 1992).

1.1.4 Three-way DNA junction structure

In order to study the structure of 3WJs, models can be made using three short, mutually complementary oligonucleotides. There are a number of techniques

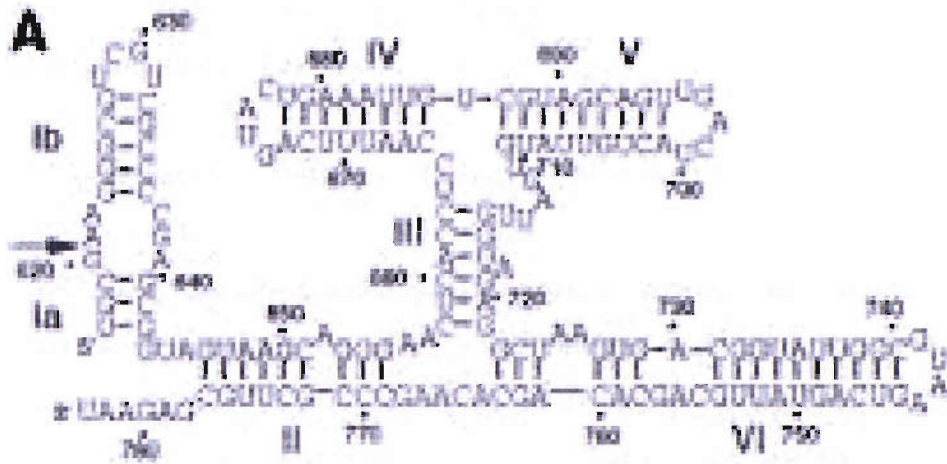


Figure 1-3: Helices in the VS ribozyme. The picture shows the layout of the five helices in the VS ribozyme. They are held together by two 3WJs. The figure is taken from Lafontaine *et al.*, (2001).

available to study the overall structure, including Nuclear Magnetic Resonance (NMR), Atomic Force Microscopy (AFM), Fluorescence Resonance Energy Transfer (FRET), gel electrophoresis and chemical probing. These techniques will be explained in more detail as needed.

One of the most useful techniques used to study 3WJs uses changes in mobility on a polyacrylamide gel to analyse the angles formed between the arms. This technique was originally developed by Cooper and Hagerman (1987) but was later adapted by David Lilley's lab (Duckett and Lilley, 1990) and has been more widely used to study four-way junctions. Restriction sites a short distance away from the branch point on each arm allow selective shortening of each arm of the junction in turn. The mobility of the fragment has been shown by AFM (a high resolution scanning technique), to be directly related to the angle between the remaining two arms (Cooper and Hagerman, 1989). The closer to linear the junction becomes, the easier it is able to pass through the matrix of the gel and so the faster it migrates.

Gel retardation studies of the 3WJ show that all the cut fragments migrate at about the same rate and hence all the angles between the arms are similar (Duckett and Lilley, 1990). Thymines at the branch point are reactive to osmium tetroxide, which modifies unstacked thymines. These data point to an open, unstacked structure, which may be Y-shaped. This structure is independent of the presence or absence of divalent metal ions.

4WJ's likewise form an open X-shaped structure in the absence of metal ions. The accumulation of negative charges from the phosphate groups in the backbone causes repulsion of the bases at the branch point. As a result, the centre remains open and exposed. However, in the presence of divalent cations, such as magnesium, the charges are neutralised and the base pairing continues throughout the junction. The base pairing allows the helices to stack coaxially to form two interconnected duplexes. The arms are crossed at an angle of about 40° (Eichman *et al* 2000) (Figure 1-4).

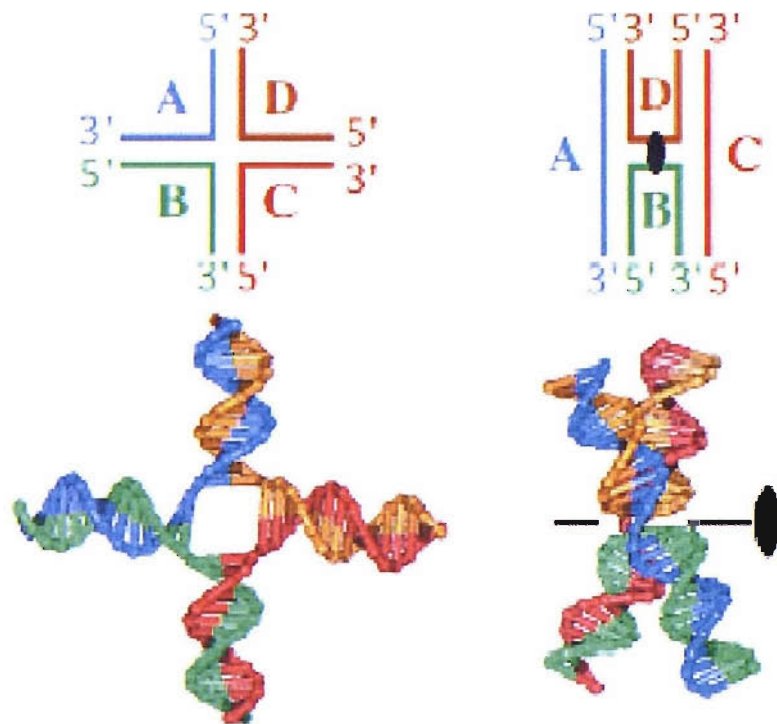


Figure 1-4: A four-way DNA junction. The top and bottom left diagrams show the 4WJ in its open and unstacked conformation as it would be in the absence of magnesium. When the charges at the branch point are neutralised, the junction folds to form pairs of coaxially stacked arms with two-fold symmetry (Eichman *et al.*, 2000).

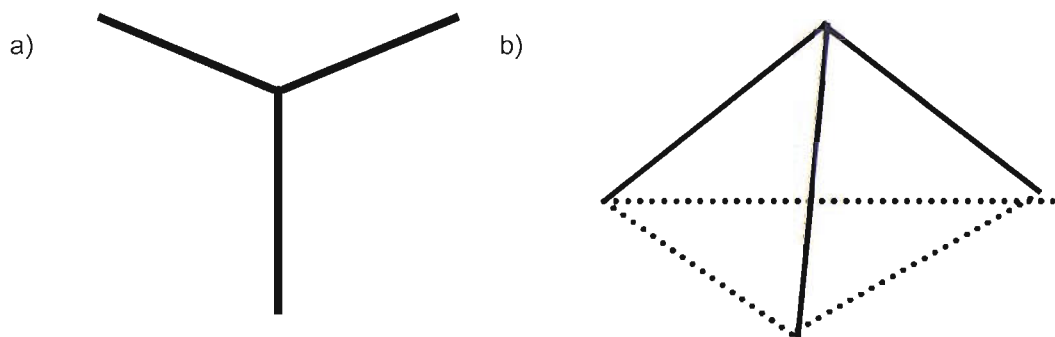


Figure 1-5. Models of the unstacked three-way junction. All angles between the three arms are identical. Figure 1-6a shows a planar representation of this. 1-6b shows the junction as a tetrahedral shape with the three arms forming the edges connecting the apex and base. The angle between each arm is approximately 80° .

To get a more detailed picture of the angles formed by the three-way junction, techniques such as FRET studies and AFM can be used. FRET looks at the distance between two points, for example the tips of two arms, and measures the intensity of the energy transfer between a donor molecule fixed to one point and an acceptor at the other. These studies confirm that the 3WJ structure is open and unstacked (Stühmeier *et al.*, 1997). Unlike 4WJs, 3WJs do not stack in the presence of cations but remain open and reactive to chemical probing with osmium tetroxide (Duckett and Lilley, 1990).

AFM studies of heteroduplex 3WJs (where one oligonucleotide contains an inverted repeat which folds back on itself to form the third arm of the junction), suggest a pyramidal form for the junction. The arms form the edges of the pyramid and the tips form the base (Oussatcheva *et al* 1999; Shlyakhtenko *et al* 2000) (Figure 1-5). The average angle between the arms, regardless of the length of the third arm is about 80° (Shlyakhtenko *et al.*, 2000). If the molecule was planar and Y shaped, angles of 120° would be expected. The angles formed by the arms in the open structure are very variable when measured using AFM and FRET, indicating a high degree of flexibility (Shlyakhtenko *et al.*, 2000; Yang and Millar, 1996).

1.1.5 Introducing flexibility to the three-way junction.

When a 3WJ is made containing unpaired nucleotides at the branch point on one strand, the character of the junction is changed and the junction stacks in the presence of divalent metal ions (Leontis *et al.*, 1991). This stacking involves maximising the interactions of the bases so that pairing occurs right up to the branch point.

The introduction of unpaired bases in the junction is perhaps not as unusual as it first seems. Unpaired bases are naturally present in many RNA junctions. There is evidence that in 3WJs, two unpaired bases are sufficient to form a covalent bridge to allow base stacking. Placing them in 4WJs, however, destabilises the structure (Kadrmas *et al.*, 1995). UV melting data shows that 3WJ's with 2-5 unpaired bases

are similar in stability but are all more stable than a 3WJ without any unpaired nucleotides.

In the presence of magnesium, two arms of the bulged 3WJ stack co-axially while the third is extended (Welch *et al.*, 1993; Leontis *et al.*, 1994; Rosen and Patel, 1993a&b). FRET studies support the co-axial stacking model and show that the extended arm forms an acute angle leaning away from the bulge (Figure 1-6) (Stühmeier *et al.*, 1997). Adding one to five unpaired bases at the branch point changes the angles between the arms in the stacked junction (Welch *et al.*, 1993). FRET can be used to show that the addition of two unpaired nucleotides allows increased mobility of one of the helices neighbouring the bulge (Yang and Millar, 1996). This is demonstrated by a wide range of angle values reflecting a Normal distribution.

The crucial concentration of magnesium for folding to occur is between 50-200 μ M (Welch *et al.*, 1993). The composition of the bulge does not appear to influence the stacking arrangements or angles between the arms, the important factor is the length added to the phosphodiester backbone. This releases the steric constraints upon the arms and allows the junction to maximise its stacking interactions by forming a more energetically favourable arrangement.

NMR can be used to provide information about the immediate environment of each base and hence give detailed structural information. It can be used to show that all three of the arms form double helical B form structures right up to the branch point, forming a co-axial duplex with the third helix extended (Leontis *et al.*, 1994; Rosen and Patel, 1993a). The extended arm forms an acute angle leaning away from the bulge of unpaired bases (represented by a circle). The greater the number of unpaired bases in the bulge, the smaller the acute angle becomes and the more the extended arm leans away from the unpaired bases (Rosen and Patel, 1993a&b). The more 5' of the unpaired bases loops out to sit in the minor groove of the preceding helix while the more 3' base remains in the structure. While the remaining base pairs surrounding the junction do form correctly, the base pairs involved in stacking interactions at the branch point are less stable and may be under increased

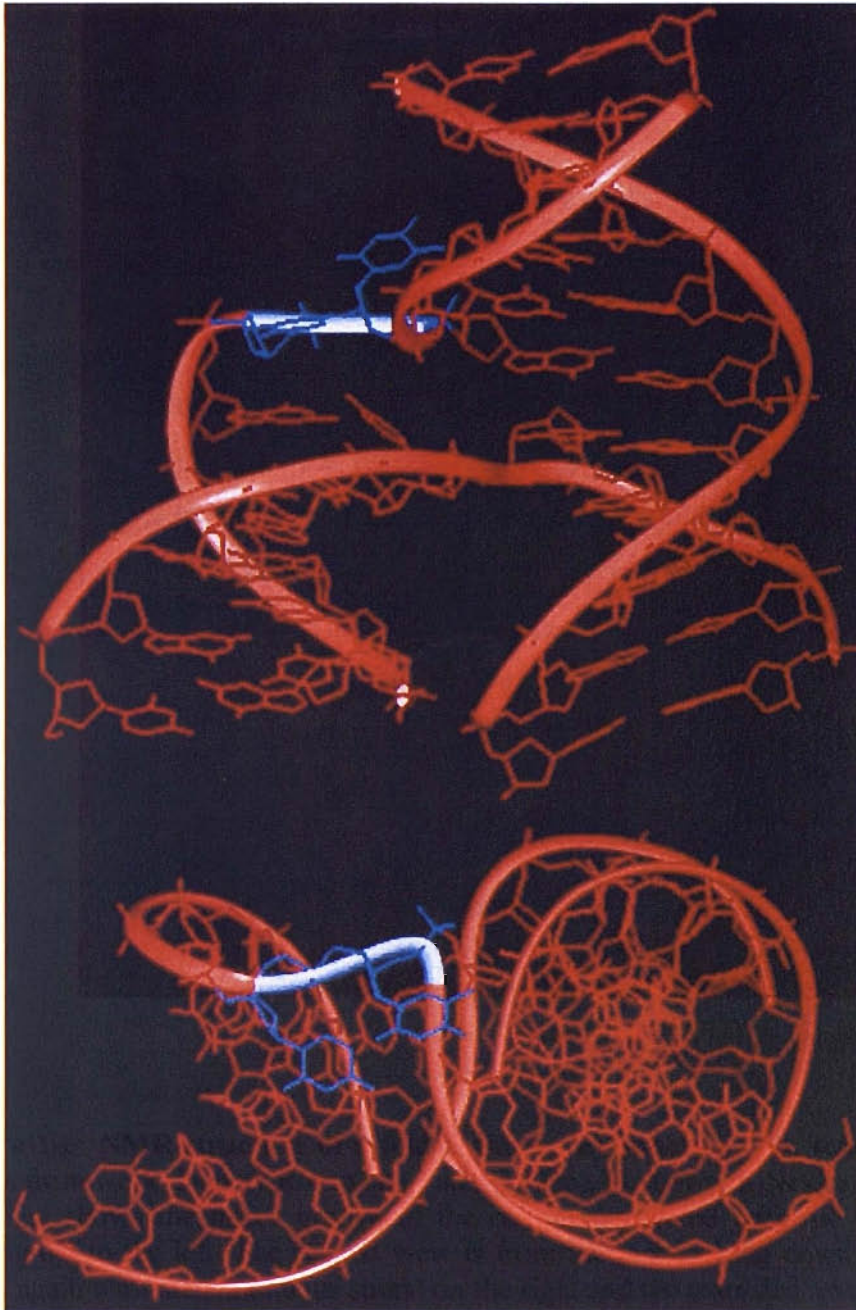


Figure 1-6: NMR structure of a folded three-way junction. The structure above comes from work done by Rosen and Patel (1993b), as shown in Neidle, (1999). The top view shows the major groove of the continuous strand with the third strand coming off to the left. The second view is from the top, looking down the coaxial stack, again with the continuous strand on the right and the extended arm on the left. In this model, there are two unpaired bases to allow flexibility in the junction. These have been picked out in blue.

conformational stress. Partial destabilisation of the Watson-Crick base pairing may be necessary to minimise steric constraints (Rosen and Patel, 1993b). The NMR structure produced by Rosen and Patel can be seen in Figure 1-6, while a simplified version is shown in Figure 1-7.

1.1.6 Conformational preference of the 3WJ.

As with the four-way junction, there are two possible isomers of the stacked 3WJ depending on which two of the three helices stack continuously (Cooper and Hagerman, 1989), (Figure 1-8). Gel mobility studies generate two different patterns of mobility for bulged 3WJs in the presence of magnesium (Welch *et al.*, 1995) (Figure 1-9). Folding is sequence dependent but it is not clear what rules govern this folding although attempts have been made to form some (Welch *et al.*, 1995; Overmars *et al.*, 1996; van Buuren *et al.*, 2000; Assenberg and Fox, 2002; Wu *et al.*, 2004). The sequence of the junction may make it energetically more favourable to be in one conformation than the other and hence have a dominant conformation (Assenberg and Fox, 2002).

1.1.7 The effect of sequence on the conformation

Changing bases around the branch point can cause a 3WJ to alter its pattern of mobility to that of the opposing conformer, or to an intermediate pattern (Welch *et al.*, 1995, Assenberg and Fox, 2002). The existence of the intermediate pattern would suggest that a junction can exist in a dynamic equilibrium between the two conformers and the unstacked intermediate. Since electrophoresis takes place over a period of a few hours, the result is an averaged conformation.

There is a great deal of evidence that four-way junctions exist in a dynamic equilibrium. Grainger *et al.*, (1998) studied four-way junctions with little bias towards either conformer. By complete digestion of the junctions with enzyme Mbo I, which cleaves across the branch point in one orientation, they demonstrated that all junctions in solution must at some point pass through a given conformer, and therefore must be in an equilibrium. This was true even for junctions with a strong bias towards one conformer. Overmars and Altona (1997), showed using chemical

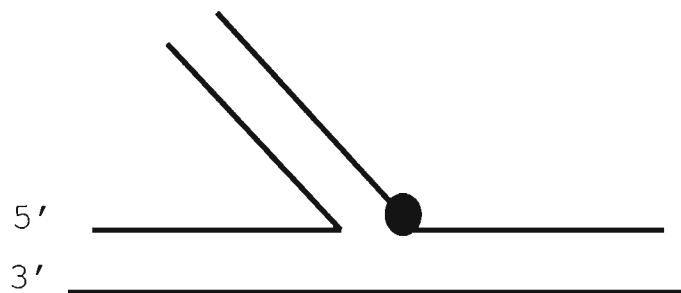


Figure 1-7: The model for the co-axially stacked, bulged three-way junction. With the addition of unpaired bases to one strand at the branch point, the steric constraints upon the junction are released. In the presence of magnesium, this allows it to pair the bulge.

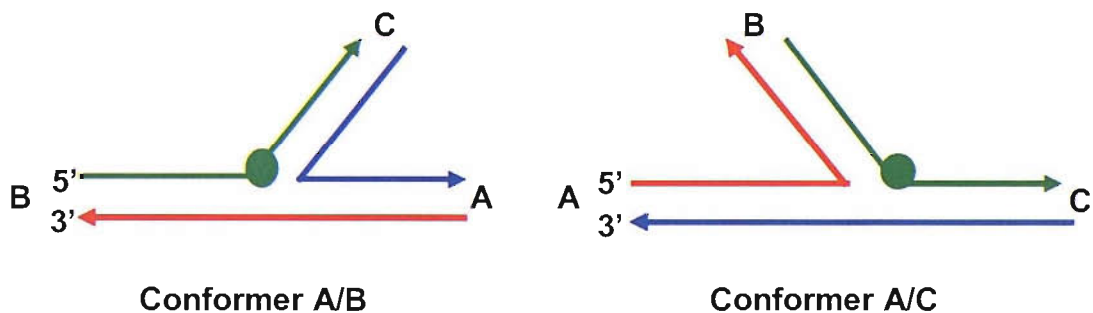


Figure 1-8: Isomers of the 3WJ. The 3WJ can form two non-equivalent isomers, conformers AB and AC (based on the nomenclature of Altona, 1996). The letters refer to the identity of the two arms that are co-axially stacked, so in conformer AB, arms A and B stack to form the continuous helix while arm C is extended. One of the arms containing the bulged strand is always extended so it appears that conformer BC does not form. The individual strands are coloured to show the orientation of each conformer and the direction of the arrows indicates the 3' polarity of the strand.

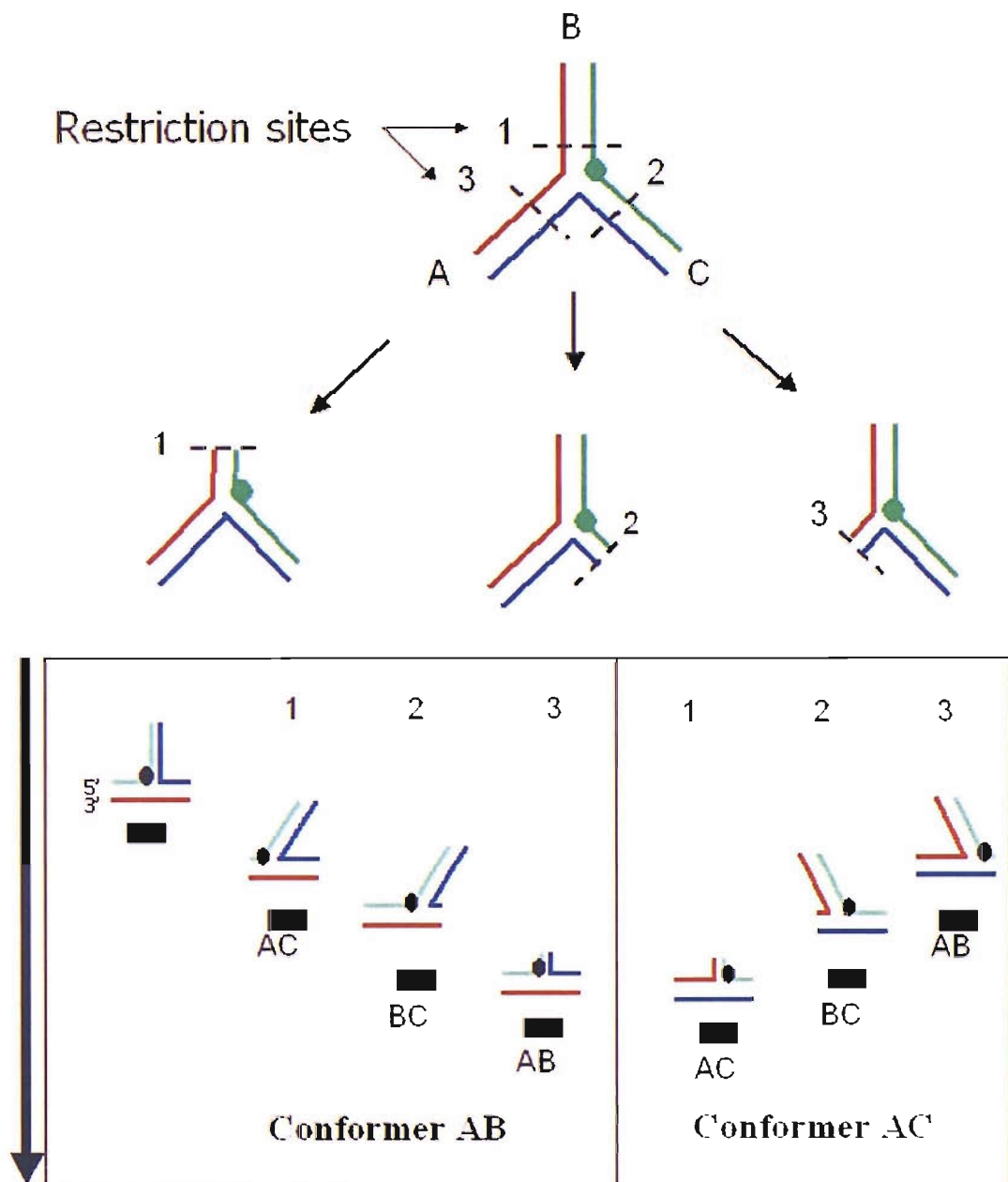


Figure 1-9: Gel retardation studies of a bulged 3WJ. A 3WJ is designed with a unique restriction site in each arm, a short distance away from the branch point. By cutting the junction with each restriction enzyme, a set of three fragments can be generated. Each fragment has two arms intact and the third arm shorter but still present. The angles between the arms should be the same in the cut junction as in the original but the mobility of the junction through a gel should be altered. If the angles between the arms differ, the junction with the greatest angle between its long arms moves through the matrix of the gel most easily.

The figure shows the results from two junctions, one heavily biased towards conformer A/B and the other towards conformer A/C. In reality, few junctions show a pattern that is so biased. The electrophoresis takes place over a few hours and so gives an averaged conformation. Many junctions exist in a much more balanced equilibrium leading to smaller mobility differences between the bands or an intermediate pattern.

exchange NMR spectroscopy that a junction they were studying had a 70% bias towards one conformer and measured the rate of exchange between conformers.

The most elegant study of the rate of conformer interchange comes from McKinney *et al.*, 2003. By measuring time resolved FRET fluctuations between the two conformational states for individual junctions, the rate of flipping was determined. The donor and acceptor molecules are either close together or far apart, showing the two different conformations (Figure 1-10). The kinetics of flipping from one isomer to the other and back again (k_{I-II} and k_{II-I}), accurately shows any bias. Junction 3, which has previously been shown to have a heavy bias towards isomer II, shows a k_{I-II} value of $12s^{-1}$ and k_{II-I} of $6.1s^{-1}$. This confirms the bias and demonstrates that even a very heavily biased conformer exists in quite a dynamic equilibrium. Junction 7, which previously showed near equal populations of each conformer, existed in a much more even equilibrium, k_{I-II} $5.7s^{-1}$ and k_{II-I} $6.1s^{-1}$. The rate of flipping is extremely high, however, the rate of exchanges lowers with increasing magnesium concentration. This represents, not a change in stability, but an alteration in activation energy for conformer exchange. The energy required to displace the magnesium ions from the branch point becomes increasingly higher.

In a Holliday junction, the dominant conformer affects the resolution of the junction by junction resolving enzymes. The sequence and resolution may be particularly important at site-specific recombination events where branch migration is limited (Azaro and Landy, 1997; Miick *et al.*, 1997).

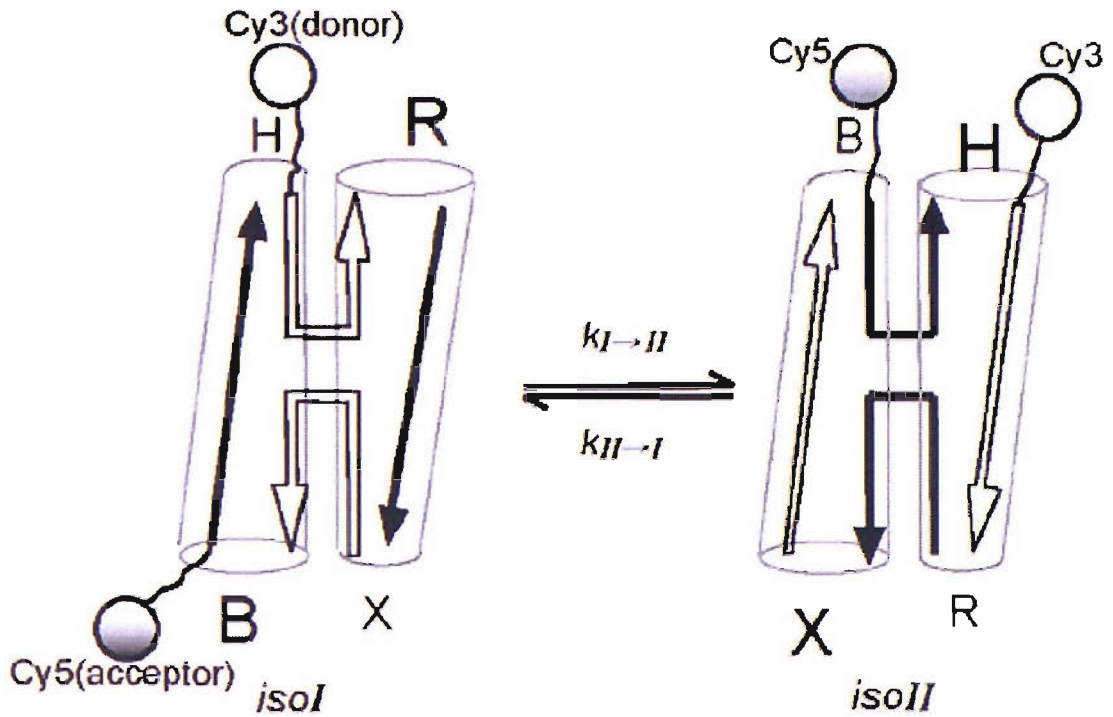


Figure 1-10: FRET studies of 4WJ. A donor and acceptor molecule are attached to different arms in the junction. Depending on the conformation of the folded junction, the donor and acceptor are either close together or far apart. FRET measures the transfer of energy between the two molecules so reads a very different signal for each conformational state. Time-resolved FRET shows the junctions flipping back and forth between the two states very rapidly in a dynamic equilibrium. The kinetics of this flipping can be measured (McKinney *et al.*, 2003).

1.2 DNA based diagnostics

1.2.1 Genetic diversity

Over the last few years there has been a drive to sequence as many genomes as possible. There has been much media focus on the completion of the first draft of the human genome by public and private effort. Sequence information alone, however, is of limited use. Eventually, this data could be used for better understanding of the human body and the improvement of health care. Before that can begin, much work needs to be done connecting variation in terms of polymorphisms and mutations with disease and normal gene function (Collins *et al.*, 1997).

The genetic diversity between individuals can be seen mainly as polymorphisms and mutations. These can occur in both coding and non-coding regions of the DNA. It is estimated that on average there are four single base changes in each gene (Cargill *et al.*, 1999). Of these SNPs found in coding DNA (cSNPs), about 40% will alter the coding sequence of the resulting protein. A site that is polymorphic has a number of different versions or alleles. The differences may be as small as a single base change or consist of a large insertion or deletion at a specific point. As these changes become fixed in the population, they can be seen at differing frequencies. Any site where the rarest allele is present in greater than 1% of the population can be called polymorphic. Other sites contain alleles that are much rarer than that. These are termed mutations because they are not fixed in the population. Single Nucleotide Polymorphisms (SNPs) are the most common form of variation. These are currently being mapped to produce a complete picture of human variation (Altschuler *et al.*, 2000).

Polymorphisms can be very important in the function of genes, whether they cause a change in the sequence of a protein or a change in the regulation of the gene itself. Many diseases have a genetic basis, whether this is Mendelian or complex. For example, mutations in the APP, PSEN1 and PSEN2 genes usually cause fully penetrative Alzheimers disease, whereas other mutations such as $\epsilon 4$ in APOE, only increase the suceptibility to developing Alzheimers (Bertram and Tanzi, 2005).

A large number of both simple and complex genetic diseases are caused by single base changes in the coding regions, resulting in a change in the amino acid sequence of the protein. The mutations leading to Mendelian genetic diseases tend to be at conserved residues in the protein and so are more likely to change its function (Thomas and Kejariwal, 2004). Cystic fibrosis is another example of a Mendelian genetic disease where faulty alleles for a single gene are inherited from both parents. There are a number of specific disease alleles that account for the majority of cystic fibrosis cases, $\Delta F508$ is responsible for 70% of cases (Riordan *et al.*, 1989; Kerem *et al.*, 1989) This is advantageous when running tests for the disease because there are only a few sequences that are responsible for the majority of cases and these can be quickly screened and identified. Screening of parents, for the five alleles responsible for 85% of cystic fibrosis cases is very effective at identifying couples with a 1 in 4 chance of having an affected child (Livingstone *et al.*, 1994). Further steps can then be taken.

Not all of these polymorphisms will have a dramatic effect on the functioning of the protein but many will have subtle effects. These subtle effects can be enough to alter the biochemistry of the body. These changes are shown in susceptibility to disease, and different metabolism of compounds. This can be particularly important in drug metabolism. People may respond very differently to the same drug (reviewed in Evans and Relling., 1999). To the pharmaceutical industry this means that different drugs will work differently on different people. If we can find out in advance who will respond to which drug best, we save a lot of money, time and effort and more importantly, lives.

1.2.2 Pathogen detection

The spread and virulence of pathogens has been an ongoing concern. The most highlighted, certainly in the UK, is the spread of methicillin resistant *Staphylococcus aureus* (MRSA), called the hospital superbug (Feil *et al.*, 2003; Perez-Roth *et al.*, 2004; Wannet *et al.*, 2004). It is well known for its multiple antibiotic resistance making it difficult to treat. Early detection to allow treatment and isolation of the patient are vital in fighting this potential killer. Viral pathogens

such as bird flu (Horimoto and Kawaoka, 2001) and SARS (Groneberg *et al.*, 2005) have and continue to threaten global epidemics if left unchecked.

There is a need to be able to identify sequences of DNA rapidly, both the genes and alleles which relate to our health and treatment, and to identify pathogens to prevent the spread of disease. This has led to an increase in the development of tools for DNA diagnostics.

1.2.3 Clinical DNA diagnostic tools

The previous two sections have highlighted the need for DNA diagnostics, that is, the identification of specific sequences of DNA, be it pathogen or human mutation and variation. In either case, the most frequent problem is detecting down to the levels of DNA present in a cellular sample. Generally an amplification step is needed as well as an initial sensor. Therefore in any assay we should have three stages – sensor – amplification – detection. Whatever technology is chosen it must be “sensitive, specific, simple and quick” (Whitcombe *et al.*, 1998). There may be only small differences between sequences, little DNA with a large background, untrained staff and a need of a rapid response.

A number of techniques are available for specific sequence detection. These are such as padlock probes, amplification refractory mutation system (ARMS), single strand conformation, heteroduplex analysis, restriction analysis, invader, molecular beacons, taqman, rolling circle amplification and microarrays. There are new techniques being developed all the time, many of which are variations and improvements on the methods listed above. Some of the techniques most relevant to this thesis have been described below.

Sequencing

The most accurate technique is sequencing an area to understand the nature of the mutation. This is probably the most informative but least cost effective option. It cannot easily be performed on a large scale without complicated and expensive automated equipment. Sanger Dideoxy sequencing using chain terminating

nucleotides (Sanger *et al.*, 1977) is the most common method used currently. By labelling each of the four dideoxy nucleotides fluorescently, the sequences can be read by lasers. The fragments of different length produced can be separated by capillary gel electrophoresis (Karger., 1996), capillary array (Kheterpal *et al.*, 1996) or MALDI-TOF mass spectrometry (Edwards *et al.*, 2001). This method is generally saved for large scale genome sequencing.

Analysis by gel electrophoresis

The polymerase chain reaction (PCR) can be used to amplify regions of interest and the sequences analysed by a number of methods involving gel electrophoresis. Restriction fragment length polymorphism (RFLP) is a common method of analysing different disease alleles, where an allele contains a restriction site, disrupted in an alternative version. The products of restriction analysis will vary in length depending on the presence and position of the restriction site. This has been used to monitor the number of repeats in Huntington's disease (Moncke-Buchner *et al.*, 2002), where affected patients have CAG trinucleotide repeats lengths of 30-180 (unaffected patients have 6-37 repeats). RFLPs have also been used to detect female carriers of the X-linked severe combined immune deficiency gene, X-SCID (Puck *et al.*, 1987)

Heteroduplex analysis uses T4 endonuclease VII to identify single base mismatches between a probe and amplified DNA target (Youil *et al.*, 1995). Unlike RFLP analysis, there is no need for the presence of a specific restriction site. When the probe and target are annealed, any mismatches are cleaved by T4 endonuclease VII and the products can be seen on a gel. This technique has been used to detect point mutations in the Haemophilia B gene (Mondandon *et al.*, 1989). Other gel techniques include single strand conformation polymorphism, SSCP (Orita *et al.*, 1989). Single base changes can change the way a single stranded fragment of DNA runs through a gel. This is compared to a wild type version of the sequences and used to screen for allelic variants. Again, this technique is not dependent on presence of a restriction site so can potentially be used for more targets.

A more accurate method of gaining information about a potential mutation or allelic variant is to use a combination of techniques. Recently, multiple restriction fragment analysis has been combined with single strand conformation analysis to investigate mutations in the PKDI gene (Thongnoppakhun *et al.*, 2004).

Allele Refractory Mutation System (ARMS)

A number of techniques have been designed as variations or improvements on PCR to produce a direct result rather than going through later detection stages. ARMS is a variation of PCR with the probes designed to pick up single base differences between alleles. If there is a mismatch at the 3' end of the probe, extension will either not take place or will be poor. Two probes are designed which differ by one base at the 3' end, to complement the two different alleles. The results are then simply analysed on a gel (Newton *et al.*, 1989). Homozygous samples will show strong extension with one probe and poor with the other. Heterozygotes will show equal extension for both probes. A variation on this technique was recently developed to improve detection of sequences which did still extend despite the mismatch by analysing the kinetics of extension (Hultin *et al.*, 2005). Proteinase K was added to the reaction to digest the polymerase. Fully matched probes would be extended quickly and so a band would be seen. In the slower reactions, the enzyme would be degraded before extension was able to take place. This addition to the technique vastly improves the sensitivity. The reaction was also set up using differently labelled probes for multiplexing.

Taqman

Taqman is a PCR based system involving the binding of a probe with a quencher at one end and fluorophore at the other. The probe binds exactly to its target. During subsequent rounds of replication, Taqman degrades the loose ends of the oligos, releasing the fluorophore and quencher. When the fluorophore is no longer quenched, it fluoresces, indicating the presence of the target (Holland *et al.*, 1991). The assay can be multiplexed using different colour fluorophores for the wild-type and mutant sequences to more accurately show heterogeneity in the sample.

Molecular Beacons

Molecular beacons can be used for multiple allele discrimination (Tyagi *et al.*, 1998). They consist of a stem-loop structure with a quencher at one end and a fluorophore at the other. The target recognition region is in the loop section. When the target is present, the structure reorganises so that the centre is bound to the target and the ends are free. The fluorescence increases up to 100 times, indicating the presence of the target. The stem loop structure is sensitive enough to detect single base mismatches, over and above a linear probe. This is due to the lower stability from the free ends. Molecular beacons are often used as a detection step, with PCR or an alternative method being used to amplify the target(s). Recently, this technique has been used to test for resistance to Itraconazole in *Aspergillus fumigatus*, a life-threatening invasive fungal disease (Balashov *et al.*, 2005). The majority of mutations conferring resistance involve a substitution of Gly54 in the Cyp51A gene product. Molecular beacons were made for the wild-type and seven mutants.

Microarrays

DNA microarrays are essentially based on hybridisation of a target and probe. The development of DNA micro-arrays has enabled analysis of multiple alleles simultaneously. This has been of immense value for analysing the expression profiles of cancer cells (Garber *et al.*, 2001; Hendrix *et al.*, 2001; Takahashi *et al.*, 2001). The information derived from these profiles is complicated and generally on a whole genome scale. A large amount of analysis is needed afterwards to look at the profile as a whole (Rhodes and Chinnaiyan, 2005). They are less useful for looking at individual genes of interest. Micro-arrays can also be used for pathogen detection and have recently been used to characterise strains of Salmonella for drug resistance (van Hoek *et al.*, 2005).

Padlock probes

Padlock probes are double stranded probes of DNA which recognise a target and bind to it to creating a complete circle. DNA ligase then joins the probes and targets together creating a catenane with the duplex target. These products can be analysed

on a denaturing polyacrylamide gel. The reaction is highly specific to the target with little background (Nilsson *et al.*, 1994). This method of detection has been combined with microarray technology to create a multiplexed assay for plant pathogens (Szemes *et al.*, 2005).

The work in this thesis focuses on a novel DNA detection technique with no requirement for PCR, Signal Mediated Amplification of RNA technology (SMART). This method is described in the following few pages.

1.3 Signal Mediated Amplification of RNA technology (SMART)

The Signal Mediated Amplification of RNA Technology (SMART) assay system developed by Cytocell Ltd. is designed to detect the presence of a specific nucleic acid sequence in cell lysates (Wharam *et al.*, 2001). The target sequence can be DNA or RNA, making it ideal to detect infectious agents such as viruses, which may use RNA as their coding information. One target for the SMART system is Methicilin Resistant *Staphylococcus aureus* (MRSA), the hospital “superbug” (Levi *et al.*, 2003). The assay should enable rapid, cheap and safe detection of infected samples from patients, in hospital labs. It is also aimed at identifying common genetic disease alleles such as major deletion mutants in cystic fibrosis. The SMART assay has also been used to detect marine cyanophage DNA (Hall *et al.*, 2002). Eventually, the assay should be sensitive enough to detect Single Nucleotide Polymorphisms (SNPs) for disease analysis or potentially for high throughput pharmacogenetic analysis.

SMART is a nucleic acid detection system that uses DNA probes to test for the presence of a sequence. The two probes are designed to recognise a sequence of about fifty bases. Each recognises half the sequence and when both probes anneal to their target, they form a short overlap with each other, not related to the target sequence (Figure 1-11). The overlap is only about five bases long to prevent the two probes annealing to each other when the target is absent. It is the formation of the overlap that creates the correct conditions for the signal to be produced.

The signal used by SMART to indicate the presence of the target sequence is the production of an RNA transcript. The longer of the two probes acts as a template. Bst. DNA polymerase, which is added to the reaction, extends the shorter probe (the extension) to full length. When this occurs, a T7 RNA polymerase promoter site is completed (Figure 1-12). One of the properties of T7 RNA polymerase is that it cannot bind to a single stranded promoter, only to a double stranded. So, for the RNA polymerase to bind, the target has to be present, the two probes must anneal to it, the short overlap must be formed stably and DNA polymerase must extend the short probe.

Once the RNA polymerase has bound to its promoter, the sequence downstream can be transcribed. The whole of the overlap region has been optimised to ensure high-level production of the transcript. This region is always the same, no matter what the target is. In this way the target is not amplified, and the same RNA transcript can be employed for all the potential targets, both DNA and RNA. Many copies of the transcript are produced but it is not yet present in high enough concentration to detect. An amplification step is needed.

Amplification of the transcript occurs by it binding to another probe, this time linear. The RNA is then used as a primer and extended by DNA polymerase to the full length of the probe. This completes another RNA polymerase site and a new transcript is produced. Using this amplification method, between 10^3 and 10^4 fold amplification was obtained.

An Enzyme Linked Oligosorbant Assay (ELOSA) is used to detect the quantity of the second RNA. Samples from the polymerase reaction are added to a microtitre plate containing two more DNA probes which bind to each end. The first probe has a biotin molecule attached to the end. This is the capture probe and binds to streptavidin coating the wells, anchoring the RNA. The second probe is attached to alkaline phosphatase and enables detection of the RNA. The plate is washed out to remove anything that is not anchored, and a substrate is added. The substrate reacts with the alkaline phosphatase to generate a colour product, detectable at 405nm. A standard curve can be set up to detect the concentration of RNA present and therefore give an indication of the level of the original target. This is useful for detecting viral or bacterial pathogens because the scale of infection can be analysed. An overview of the entire assay can be seen in Figure 1-13. 10amol of RNA can be detected using the ELOSA reaction. The SMART assay can be used to detect down to 20-50fmol nucleic acid target.

The junctions used in SMART contain a bulge in the template strand at the branch point to allow them to fold and base pair fully. The composition of the bulge is a chemical linker, either an octanediol or a hexaethylene glycol. These are simply short linkers that act in the same way as the extra bases to extend the backbone and allow folding. The reason for using chemical linkers is two fold. They help reduce the

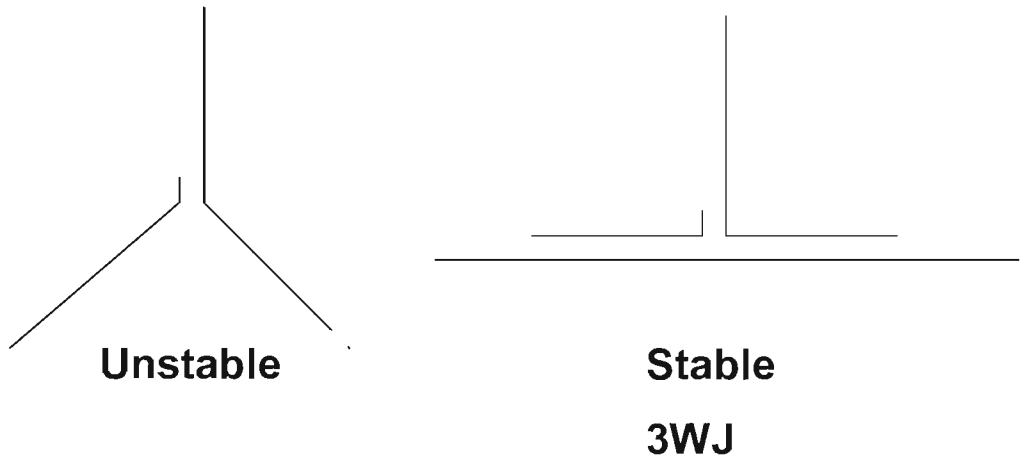


Figure 1-11: Binding of the probes in the presence and absence of the target. The probes do come into contact in solution and may bind transiently to each other. The overlap is kept short to discourage stable formation of this structure and possible extension by DNA polymerase. When the probes are annealed to the target, they have a stable base from which extension and transcription can take place.

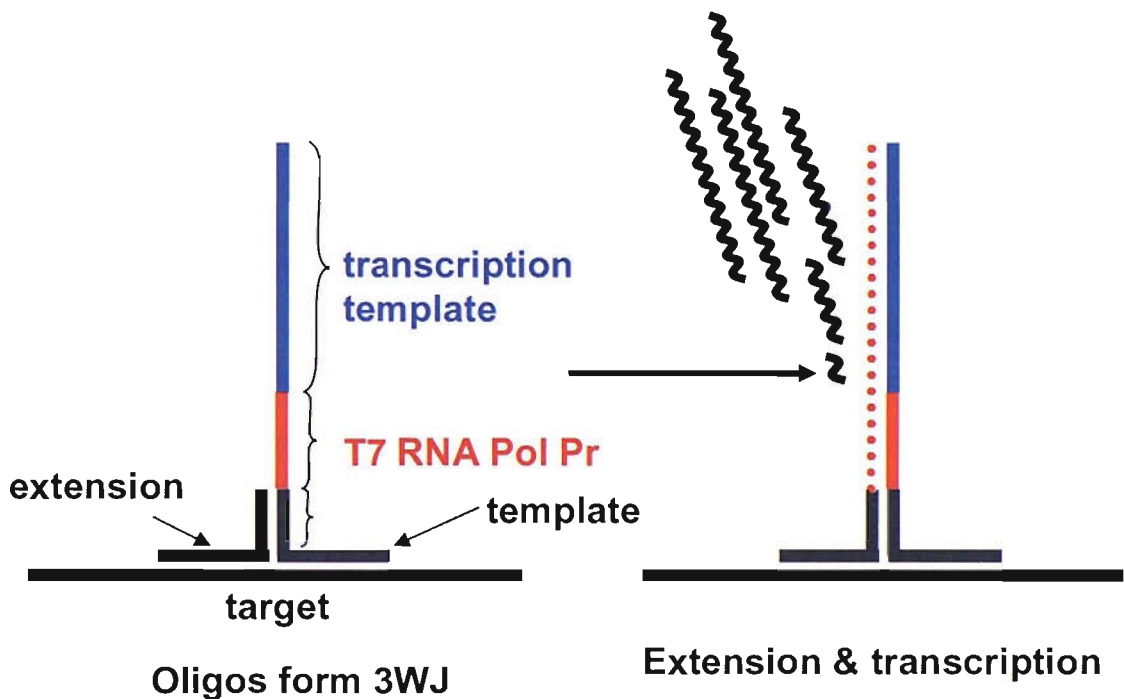


Figure 1-12: Extension and transcription of the 3WJ. The 3WJ is formed in the presence of the target and extended by DNA polymerase. The T7 RNA polymerase promoter is located just after the overlap region between the two probes. RNA polymerase can bind when the promoter is double stranded and reads along to a transcription start site. The RNA produced and all the sequences around the transcription site are optimised for a high yield.

background signal caused by random priming of *Bst* DNA polymerase and they also eliminate errors in base pairing across the junction branch point. Placing them in the sequence gives better target specificity by reducing inappropriate extension and transcription.

The target sequence used by SMART is adaptable to any genomic target. The sequence of the T7 region and beyond is the only fixed part. The system can recognise RNA targets as well as DNA without needing reverse transcriptase so is useful for targeting viral RNA. It is also quantitative and so can detect the extent of infection. It is the signal that is amplified, not the target, unlike PCR. This is advantageous because the sequence of RNA amplified (the signal) can be optimised to allow maximum transcription. Unlike PCR there is no need for a thermal cycler and highly trained operators. SMART can be isothermal, cheap and may be fairly easily automated for high throughput analysis.

1.4 Split Promoter Amplification Reaction (SPAR)

The Split Promoter Amplification Reaction (SPAR) is a second-generation technology of SMART. It is designed for the same purpose, to detect specific nucleic acid sequences, and uses the same principles to do so, using DNA probes and producing an RNA signal. The junction has been redesigned to remove the need for DNA polymerase. Using only one enzyme, RNA polymerase, simplifies the system and makes it cheaper. The junction used by SPAR is termed a two-and-a-half way junction (2.5WJ) as opposed to the 3WJ of SMART. It is still formed by two probes and a target sequence but the second probe (the complement), anneals only to the first probe, not to the target. In this way it forms an incomplete 3WJ (figure 1-14).

In the 3WJ, the DNA polymerase produced a double stranded T7 RNA Polymerase promoter when a stable 3WJ had been formed. It was this step that controlled when T7 RNA polymerase could bind to the junction and start transcription. In the 2.5WJ, the promoter is not complete unless the full junction is formed. The first three bases of the promoter are on the target itself. The polymerase needs the ATT on the target to complete the site and allow binding (figure 1-15). RNA is produced as before and then amplified and detected. This technology is still in development but will be used

Scheme for Isothermal Amplification using SMART

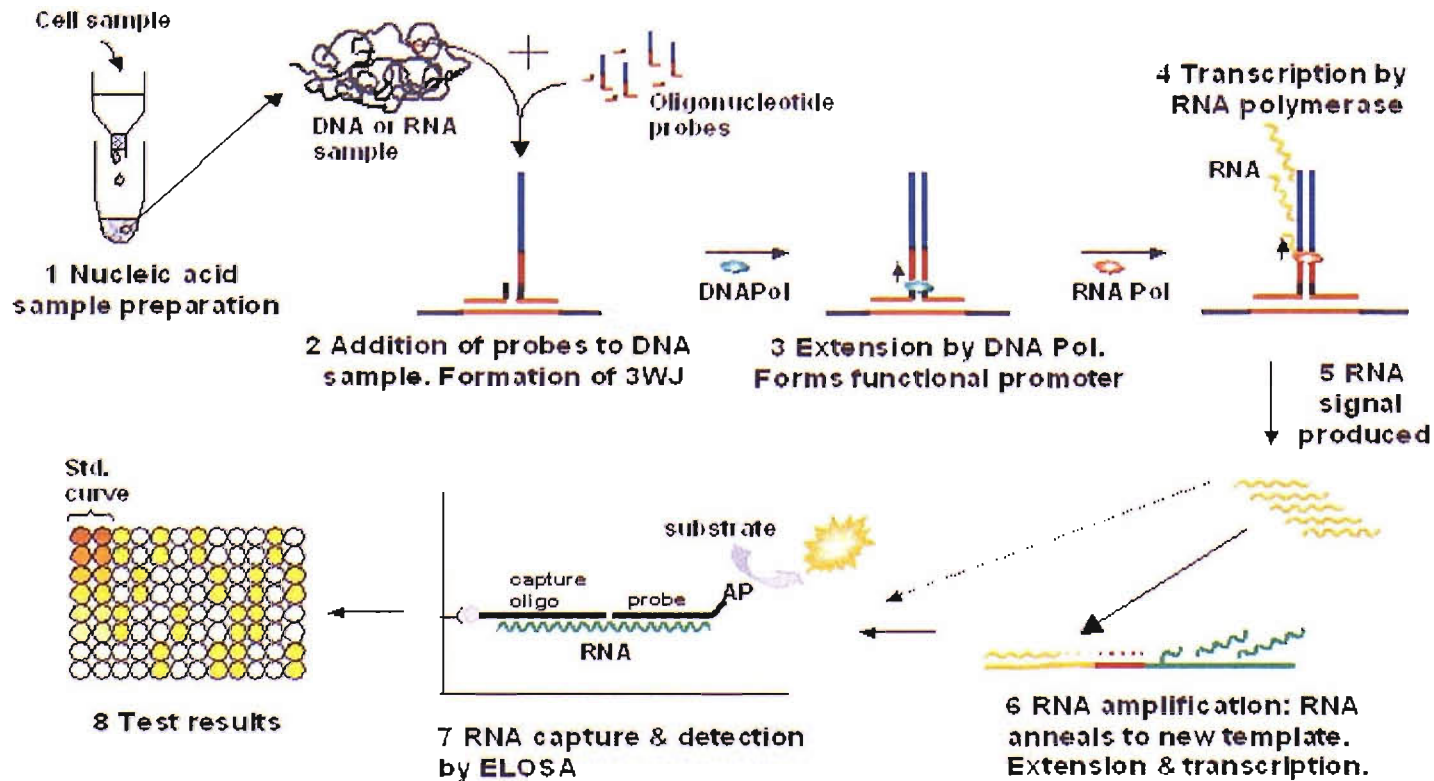


Figure 1-13: SMART schematic showing the assay, amplification and detection steps. A cell sample is lysed to release the nucleic acid. The probes are added, then the nucleic acid sample is denatured and reannealed. The probes will form a 3WJ with the target if it is present. DNA polymerase can then extend this structure completing the T7 RNA polymerase binding site. RNA 1 is then produced off the junction and binds to an amplification probe. The RNA is extended and RNA 2 is transcribed. The samples are transferred to streptavidin coated microtitre plates and two more probes are added. These probes secure the RNA to the plate and allow detection by an enzyme reaction. The amplification of RNA is in the order of 10^3 to 10^4 . Approximately 10 amols of RNA is detected in the ELOSA. This diagram is reproduced courtesy of Cytocell Ltd.

3WJ (SMART)

2.5WJ (SPAR)

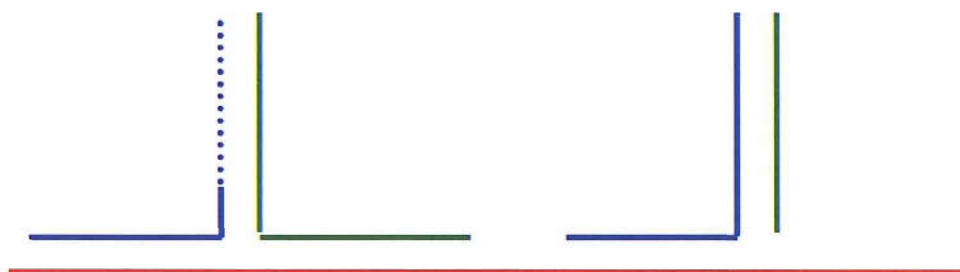


Figure 1-14: SMART 3WJ and SPAR 2.5WJ junctions. The genomic target nucleic acid is shown in red along the bottom. The blue probe (the extension) serves as the start site for T7 RNA polymerase. In SMART, DNA polymerase extends the probe once the 3WJ is formed (dotted line). The green probe is the template or complement. It contains the transcription region that will be transcribed by RNA polymerase as a signal of the presence of the target sequence. In the 2.5WJ, the complement probe binds only to the other probe and not to the target. As a result, the structure is slightly less stable than the 3WJ.

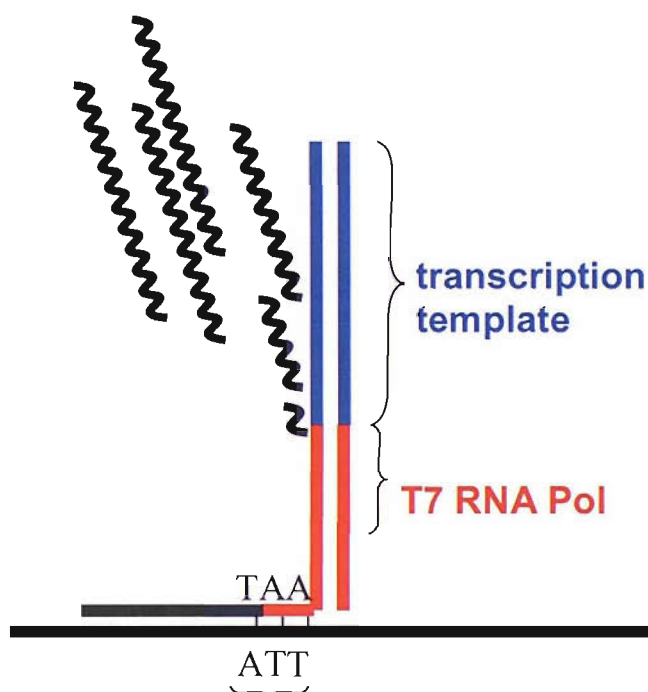


Figure 1-15: Transcription from the 2.5WJ. T7 RNA polymerase recognises the completed double stranded promoter, binds and transcribes RNA from the junction. As in the SMART assay, the production of this RNA is the signal for recognition of the target site and the completion of a 2.5WJ. The junction is orientated to resemble the 3WJ of SMART. In reality, the RNA polymerase would be unlikely to read round the corner of the junction around kinked DNA. What is more likely to happen is that the extension strand containing the full promoter lies flat while the target is kinked to facilitate polymerase binding.

in the discrimination of mismatches e.g. SNPs. SPAR does have a problem with high background, probably due to the probes binding to similar sequences containing ATT.

1.5 Aims of the project

The aim of this thesis is to elucidate the relationship between structure and sequence of three-way junctions. This information would then be used to choose sequences that would work well for SMART (different sequences have been shown to respond differently in the SMART assay). By following a simple set of rules, the SMART assay could be optimised.

The primary aim was to be achieved by using a variety of structural, chemical, enzymatic and melting studies of three-way junctions with different sequences. The structural studies consisted of identification of the dominant conformer by gel mobility studies. Chemical probing was used to identify base pair interactions and fluorescent melting studies to analyse stability of the junction.

A secondary aim was to optimise the SMART or SPAR assay to recognise single nucleotide polymorphisms (SNPs). This involved looking at the best position for a mismatch to be recognised and looking at increasing the stringency of recognition.

Chapter 2

MATERIALS AND METHODS

2 Materials and Methods

2.1 Reagents

All chemicals were purchased from Sigma unless otherwise stated. Redivue radioactive gamma and alpha ^{32}P dATP was purchased from Amersham Biosciences at 3000 Ci per mmol.

2.1.1 Buffers

5X TBE	54 g Tris, 27.5 g boric acid, 10 mM EDTA in 1 litre. pH 7.4
10X TBM	108 g Tris, 55 g boric acid, 10 mM MgCl_2 in 1 litre. pH 7.4
10X NEB Buffer 2	500 mM Tris-HCl pH 7.9, 500 mM NaCl, 100 mM MgCl_2 , 1 mM DTT.
Formamide stop solution	80% Formamide, 10 mM EDTA, 0.1% Bromophenol blue.
Non-denaturing loading dye (Ficoll)	20% ficoll, 0.1% bromophenol blue.
Tris EDTA (TE)	10mM Tris, 0.1mM EDTA. pH 7.4
DNaseI buffer	2mM MgCl_2 , 2mM MnCl_2 , 20mM NaCl.
10XTBE+urea	108 g Tris, 55 g boric acid, 9.4 g EDTA, 500 g urea and hot distilled water to 1 l
Diluent	500g urea and hot distilled water to 1 l.
5X lyophilisation concentrate (SMART)	25 g sucrose, 5 g Ficoll, 5 g polyvinylpyrrolidone and made up to 100 ml with molecular grade water
Buffer C (SMART) - the final volume of the buffer is given at 5 μl for one reaction. This should be scaled up for the number of reactions needed.	2.5X transcription buffer (Ambion) (160 mM Tris, 24 mM MgCl_2 , 8 mM spermidine, 40 mM DTT), 600mM NaCl, 1X Lyophilisation concentrate. Molecular grade water to give 5 μl per reaction.
Hybridisation buffer (SMART)	20 mM EDTA (Ambion), 1 M NaCl, 50 mM Tris.HCL to pH 8.0, 1% w/v BSA plus sterile water to give 1 l.

2.12 Oligonucleotides

All oligos were purchased from Oswell DNA services on the 40nmol or 0.2 μ mol scale and contained additions as specified. All oligos used for the light cycler were HPLC purified.

a) Oligos for SMART and extension assay.

Oligos manufactured for use in the SMART assay consisted of a synthetic target, long template probe and short extension probe (Figure 2-1). The template sequence contains an octanediol at the branch point to aid folding. To detect the presence of MRSA, probes were designed to detect two target sequences from different genes in the same genome: *mecA* and coagulase genes. The use of shorter target sequences allowed the behaviour of the probes to be more easily analysed. When using *Salmonella typhimurium* 16SrRNA as a target, two junctions were designed during the course of the research. JO and J2 differed by a few bases in the target sequence they recognised. J2 was adopted later due to better specificity for *Salmonella*. The sequences of the probes and the artificial targets are shown in Table 2-1.

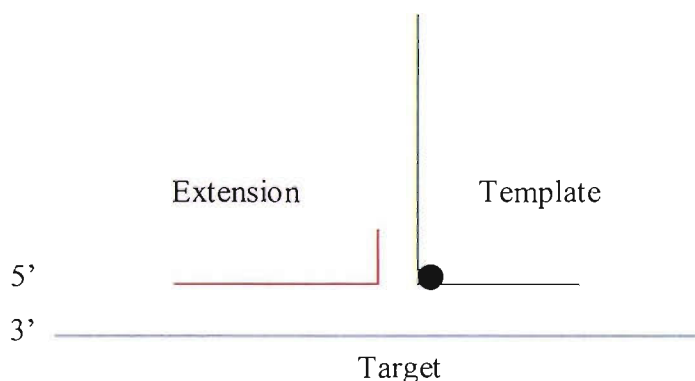


Figure 2-1: SMART junctions. A short, synthetic target was created for use in assays. This, together with the extension and template probe, forms the SMART 3WJ. These oligos were used to test the system and analyse the efficiency of the extension step of the assay. The *Salmonella* junctions used in Chapter 4.2 also include facilitator oligos placed either side of the extension and template oligos. These were designed to open up the DNA and prevent secondary structures from forming along the target region.

mecA target	AAA TGA AAC AAG GAG AAA GTG GCA GAC AAA TTG GGT GGT TTA TAT CAT ATG ATA AAG ATA ATC CAA ACA TGA TGA TGG CTA TTA ATG TTA AAG ATG TAC AAG ATA AAG GAA TGG CTA GCT ACA ATG CCA
mecA template	TCG TCT TCC GGT CTC TCC TCT CAA GCC TCA GCG CTC TCT CTC CCT ATA GTG AGT CGT ATT AAT TTC GAA-(OCTANEDIOL)-GGA TTA TCT TTA TCA TAT GAT ATA AAC CAC CCA ATT T
mecA extension	TTG TAC ATC TTT AAC ATT AAT AGC CAT CAT CAT GTT TTT CGA AAT
coag target	CAA AAC AAG GCA TCA GAA ACA AAC GCA TAT AAC GTA ACA ACA CAT GCA AAC GGC CAA GTA TCA TAC GGA GCT CGC CCA ACA CAA AAG AAG CCA AGC GAA ACA AAT GC
coag template	TCG TCT TCC GGT CTC TCC TCT CAA GCC TCA GCG CTC TCT CTC CCT ATA GTG AGT CGT ATT AAT TTC GAA-(OCTANEDIOL)-GGC CGT TTG CAT GTG TTG TTA CGT TAT AT
coag extension	GTG TTG GGC GAG CTC CGT ATG ATA CTT TTC GAA AT
<i>Salmonella</i> JO synthetic DNA target	GCC GCG TGT ATG AAG GAA GCC TTC GGG TTG TAA AGT ACT TTC AGC GGG GA GGA AGG TGT TGT GGT TAA TAA CCG CAG CAA TTG ACG TTA C
JO template	TCG TCT TTC CGG TCT CTC CTC TCA AGC CTC AGC GCT CTC TCT CTC CCT ATA GTG AGT CGT ATT AAT TTG AA – (OCTANEDIOL) – TCC CCG CTG AAA GTA CTT TAC AAC CCG AAG
JO extension	TAT TAA CCA CAA CAC CTT CCT TCG AAA T
JO extension facilitator	GTA ACG TCA ATT GCT GCG GT
JO template facilitator	GCC TTC TTC ATA CAC GCG GC
J2 template	TCG TCT TTC CGG TCT CTC CTC TCA AGC CTC AGC GCT CTC TCT CTC CCT ATA GTG AGT CGT ATT AAT TTG AA – (OCTANEDIOL) – CCT CCC CGC TGA AAG
J2 extension	TAA CCA CAA CAC CTT TTC GAA AT

J2 extension facilitator	ACG TCA ATT GCT GCG GTT AT
J2 template facilitator	TAC TTT ACA ACC CGA AGG CC

Table 2-1: Sequences of all the SMART junctions. The junctions MecA and Coag were used to investigate the presence of MRSA while JO and J2 were used to detect *Salmonella*. The sequences are shown 5' to 3'. An octanediol has been covalently linked into the template oligos and its location is indicated in the sequence.

A number of variations on the MecA junction were made to complement the above oligos. These new junctions contained additional octanediols to add extra flexibility and to raise the RNA polymerase recognition site away from the branch point. This allowed the enzyme better access and, in theory greater enzyme efficiency. The sequence of the junction remained unchanged but 2, 3 and 4 octanediols were added to the branch point of the template strand and 1,2,3 and 4 octanediols were added to the branch point of the extension strand.

Different versions of the *Salmonella* J2 extension probe were tested for reactivity with RNA polymerase. The modified versions are listed in Table 2-1 with their British Biocell International (BBI) designation.

Ampi 1881	TAA CCA CAA CAC CTT - (OCTANEDIOL)- TTC GAA AT
Ampi 1882	TAA CCA CAA CAC CTT CTT GAA AT
Ampi 1883	TAA CCA CAA CAC CTT TTC GCA GC
Ampi 1884	TAA CCA CAA CAC CTT CTT GCA GC

Table 2-2: Modified extension sequences for J2. The sequences used additional chemical linkers or alternative T7 RNA polymerase binding sites to prevent or reduce interaction with RNA polymerase.

b) Oligos for LSA and chemical probing.

Fully complementary oligos were made, modified from the original SMART junctions to include a unique restriction site in each arm; SspI, DraI, EcoRV (Figure

2-2). The original sequence was modified as little as possible. The restriction sites were chosen for their similarity to the sequence and so that they would leave blunt ends when they cut. This was important for the long-short arm assay because overhanging ends can affect the mobility of the junction in the gel (Assenberg and Fox, 2002). The sequences of these junctions are shown in Table 2-3.

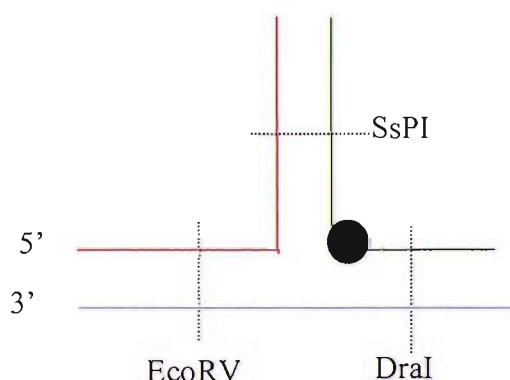


Figure 2-2: Full length LSA junction. The MecA and Coag junctions were modified to give full length arms and contain blunt cutting restriction sites for the long-short arm assay. The strands retained the name of the probe they took their sequence from. The red strand is the extension, green is the template and blue is the target, as before. Both junctions contained an octanediol in the template strand at the branch point (shown as a black circle).

LSA MecA ta	TTA TAT CAT ATG ATT TAA ATA ATC CAA ACA TGA TAT CGG CTA TTA ATG TT
LSA MecA te	TAT AGT GAG TCG TAA TAT TTT CGA A-(OCT)-GG ATT ATT TAA ATC ATA TGA TAT AA
LSA MecA ex	AAC ATT AAT AGC CGA TAT CAT GTT TTT CGA AAA TAT TAC GAC TCA CTA TA
LSA Coag ta	AAC GTA ACA ACA CTT TAA AAC GGC CAA GTA TGA TAT CGA GCT CGC CCA AC
LSA Coag te	TAT AGT GAG TCG TAA TAT TTT CGA A-(OCT)-GG CCG TTT TAA AGT GTT GTT ACG TT
LSA Coag ex	GTT GGG CGA GCT CGA TAT CAT ACT TTT CGA AAA TAT TAC GAC TCA CTA TA

Table 2-3: Sequence of the long-short arm assay junctions.

Extra oligos were made to complement the above. Mte 3-oct contained three octanediols at the branch point of the MecA template sequence. Mex 1-oct and 3-oct contained one and three octanediols at the branch point respectively.

c) Oligos for light cycler

All light cycler oligos were based on a shorter junction sequence (Figure 2-3). A number of shorter oligos were made which, together, completed the junction. These were made to assess the stability of the junction in more detail. The main junctions used were a complete 3WJ with a variety of linkers (CC, TT, Hexaethylene glycol (HEG) and 2 HEG), a 2.5WJ, a 2.5WJ with a helper oligo (split 3WJ) and a duplex (Figure 2-4). A fluorophore, (fluorescein) and a quencher, (methyl red), were present on opposite strands in each junction. The sequences of the oligos are shown in Table 2-4.

Target	GGG AAA GAG AAC TTA GAA ATG CTG T-FAM
Extension	MeRed-CAG CAT TTC TAA TAC GAC TCA CTA TA
Template. X can be CC, TT, HEG, 2HEG.	TAT AGT GAG TCG TA X GTT CTC TTT CCC
2.5WJ complement	TAT AGT GAG TCG TA
2.5WJ foot	GTT CTC TTT CCC
Duplex complement	TAT AGT GAG TCG TAT TAG AAA TGC TGT-FAM

Table 2-4: Sequences of the light cycler junctions. The presence of methyl red is indicated by MeRed while fluorescein is abbreviated to FAM.

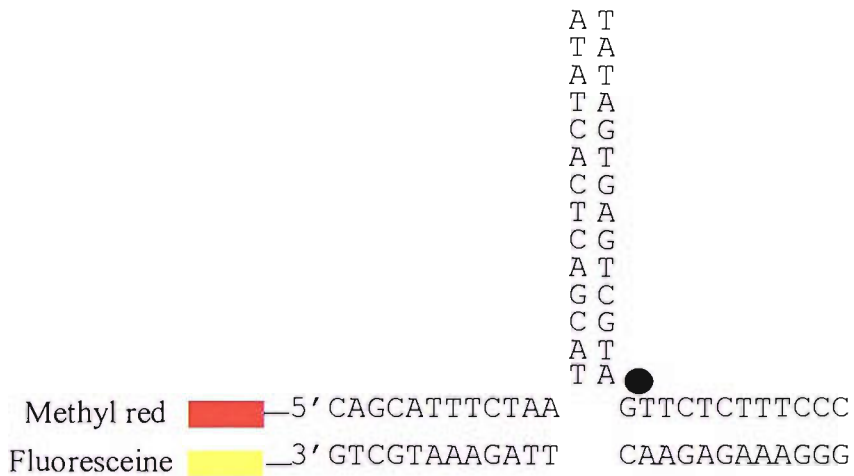


Figure 2-3: Outline of the light cycler junctions. A set of oligos were made to form various junction structures, all based on the above sequence. A fluorophore and quencher were attached to opposite strands in order to follow the melting of the strands.

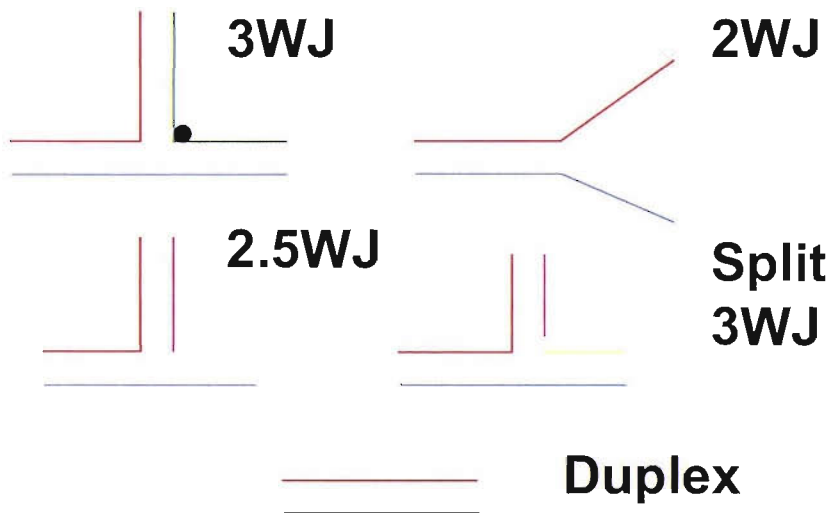


Figure 2-4: Light Cycler junctions. The junctions above were constructed from component oligos based on the sequence in Figure 2-3. They were used to compare the stability of a variety of junctions by following the melting and annealing of the strands under various buffer conditions.

To analyse the effect of mismatches on junction stability, a series of three target strands were created containing a single mismatch at different positions (Figure 2-5 and Table 2-5).

TA-MUT1	GGG AAA GAG AAC CTA GAA ATG CTG TT-FAM
TA-MUT2	GGG AAA GAG AAC TTA GAA ACG CTG TT-FAM
TA-MUT3	GGG AAA GAG AAC TTA GAA ATG CCG TT-FAM

Table 2-5: Sequences of the target strands in the Light Cycler junction. Mut 1-3 all contain a single base mismatch but at different positions. These targets were used with the template and extension oligos from Table 2-3.

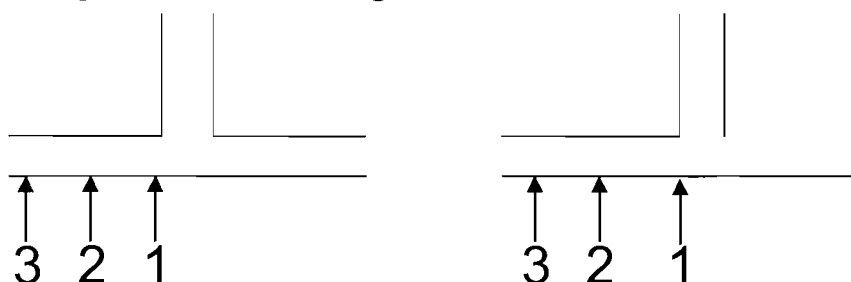


Figure 2-5: Position of mutations in the target strand. Three new, labelled target strands were created, each with a mutation in a different position. The diagram shows the approximate location of the mutation (T to C) in strands 1, 2 and 3. A 3WJ and 2.5WJ are shown.

d) FRET oligos

Junctions containing donor and acceptor fluorescent molecules were used for Fluorescence Resonance Energy Transfer type experiments. Two junctions of slightly different sequence were used. For each junction, there were three different combinations of donor and acceptor pairs, to cover each pair of arms. The positions of the fluorescent molecules and the sequences of these junctions are shown in Figures 2-6 and 2-7.

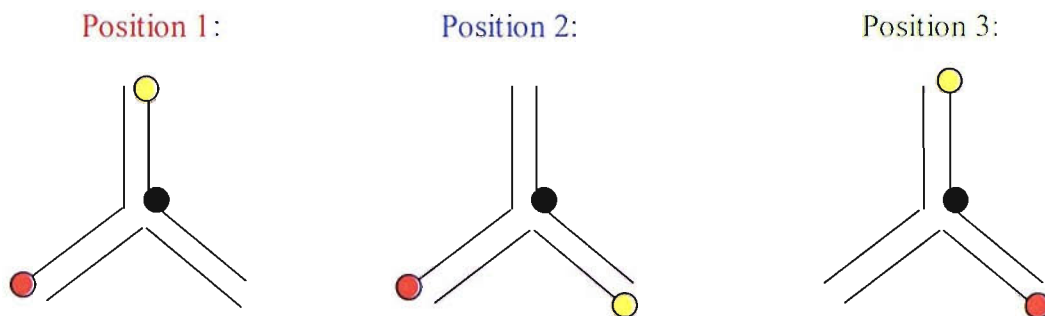


Figure 2-6: Positions of the donor and acceptor molecules on the three-way junctions. The three variations of the same junction are colour coded as positions 1-3 in red, blue and green, respectively. These correspond to the colours used on the graphs in Chapter 6. The donor and acceptor are shown as circles on the end of the relevant strand to represent the covalent attachment of the molecules. Fluorescein is shown in yellow while tamra is red. The position 1 junction contains the donor on arm B and the acceptor on arm A. Position 2 retains the acceptor on arm A but has the donor on arm C. Finally, position 3 uses the donor on arm B and the acceptor on arm C. In each case the distance between the donor and acceptor molecules, and hence the relevant arms, can be measured using FRET.

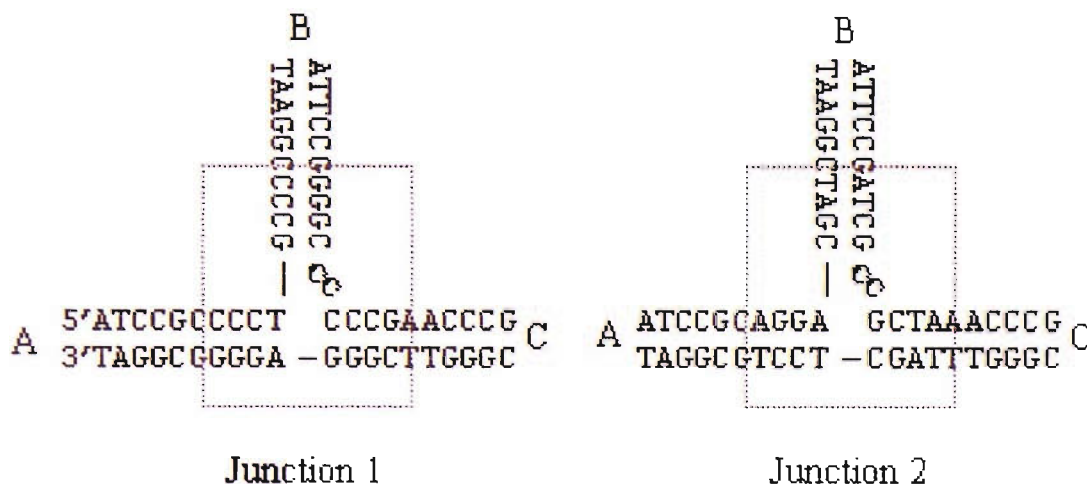


Figure 2-7: Sequences of the two FRET junctions. Junctions 1 and 2 are identical in sequence except for the central four base pairs on each arm (highlighted in the box). Both junctions use two unpaired cytosines as the linker. For each junction there is a set of oligos with donor and acceptor molecules attached to different strands. They are combined as above to allow the distance between each pair of arms to be measured.

2.2 5' end labelling of oligonucleotides

100pmol of the appropriate oligo (in a volume of less than 16µl of TBE) was added to 2µl polynucleotide kinase buffer, 1µl T4 polynucleotide kinase (NEB 10,000U/ml), 1µl ³²P-gamma ATP and xµl water to make a 20µl reaction volume. The labelling mix was left at 37°C for 1h and stopped with 20µl formamide stop solution. The solution was denatured by heating at 100°C for 3mins then crash cooled by placing on ice.

2.2.1 Gel electrophoresis

Denaturing polyacrylamide gels were made as a 50ml solution containing x ml Sequagel (30% A: B 39:1)(National Diagnostics Inc.) , 5ml 10X TBE+urea, 200µl 20% ammonium persulphate (AMPS), 40µl N,N,N',N'-tetramethylethelenediamine (TEMED) and diluent up to 50ml. 1ml Sequagel in 50ml gives a gel of 0.6% acrylamide. Non-denaturing gels were made using x ml Accugel (40% acrylamide: bis-acrylamide 19:1), 10ml 5XTBE or 5ml 10XTBM, 200µl AMPS, 40µl TEMED and water up to 50ml. 1ml Accugel in 50ml gives a gel of 0.8% acrylamide. In both cases, gels were left to set for 30 mins then fixed to a vertical gel tank. Denaturing gels were run at 1500V for 1-3h. Non-denaturing gels were run at 800V for 2-3h. Gels were removed from the tanks, the glass plates opened and the gel fixed in 10% glacial acetic acid (Fisher Inc) for 10mins. They were then transferred to 3mm Watman paper (Fisher Inc) and dried on a gel dryer for 1h.

2.2.2 Gel purification of oligonucleotides

Radiolabelled oligos were purified from the other components of the reaction by running on a 10% denaturing polyacrylamide gel. The labelled DNA was identified by autoradiography of the wet gel and the X-ray film used as a guide to excise the bands. Gel slices were incubated overnight in a P1000 pipette tip filled with glass wool and 300µl TE and sealed at both ends with parafilm.

2.2.3 Elution and precipitation of DNA from polyacrylamide gel

After incubation in TE overnight, the pipette tip was unsealed and spun down to release the TE containing the DNA. 1/9th total volume of 3M sodium acetate pH 7.8 was added then 2X total volume of 100% ethanol. The solution was placed on dry ice for 10mins then spun down in a microcentrifuge at 14K rpm for 10mins. The supernatant was carefully removed and discarded. The remaining pellet was washed twice with 200µL 70% ethanol and spun down each time at 14Krpm for 5mins. The remaining ethanol was evaporated off in a speedvac.

2.3 Annealing of oligos

2.3.1 Non-radiolabelled junctions

100pmol of each strand was added to a 50µl reaction volume with 5µl NEB buffer 2. The annealing mix was heated to 95°C then left to cool to room temperature then in a fridge to 4°C.

2.3.2 Radiolabelled junctions

Radiolabelled oligos were resuspended to approximately 100 counts/µl. To create junctions with only one strand labelled, excess of the unlabelled strands (5µl of each) were added to 10µl of the purified, labelled strand in a 50µl reaction volume with 5µl NEB buffer 2. The annealing mix was heated to 95°C then cooled to 4°C as before.

2.4 Long-short-arm assay

The fully complementary junctions described in 2.12a have a unique restriction site in each arm, SspI (New England Biolabs 5,000 U/ml), EcoRV (NEB 20,000 U/ml) and DraI (NEB 20,000 U/ml). This enables each arm to be cut independently of the

others. For each junction, four identical samples were prepared. One was left uncut, the other three were cut with one of the restriction enzymes as follows. 7.5 μ l (15pmol) of the junction solution was added to 2 μ l NEB buffer 2, 2 μ l R-ENase and 8.5 μ l water. The restriction mix was incubated at 37°C for 3h then 10 μ l non-denaturing loading dye was added. The junctions were run on a 12% minigel in a Novex gel cassette, containing 10mM magnesium. The gel was run at 100V for 3-4 h.

2.4.1 Staining with ethidium bromide

Minigels were stained in ethidium bromide (300 μ g/200ml) for 15 mins on a shaker then imaged using a UV transilluminator.

2.5 Chemical probing of three way junctions

2.5.1 Diethyl pyrocarbonate: 2 μ l 5' labelled junction DNA was incubated at 41°C (the SMART assay temperature) for 5 mins then 5 μ l diethyl pyrocarbonate (DEPC) was added. The reaction was left for 5mins then another 5 μ l DEPC was added and the reaction allowed to proceed. After a further 5mins, the reaction was stopped with 2 μ l β -mercaptoethanol. The reaction volume was made up to 20 μ l with water and the DNA was then ethanol precipitated (section 2.2.2). The precipitated DNA was then cleaved with piperidine (section 2.5.3) and run on a 15% denaturing polyacrylamide gel and the pattern of cleavage viewed using a phosphoimager screen (Amersham biosciences).

2.5.2 Osmium tetroxide: Osmium tetroxide was obtained as a solution dissolved in 2-methyl-2-propanol (Sigma #208863). It was mixed with pyridine in a 4:1 ratio to the volume required. 2 μ l junction DNA was reacted cooled on ice, then 5 μ l of the osmium-pyridine solution was added. The reaction was allowed to proceed for 15mins. The reagent was extracted twice with 300 μ l diethyl ether and the DNA precipitated. The samples were then treated with 10% piperidine as described in 2.5.3

2.5.3 Cleavage of DNA with piperidine

50µl 10% piperidine was added to the precipitated DNA and incubated at 100°C with the Eppendorf cap closed. After 30mins the piperidine was evaporated off in the speed vac and the pellet washed twice with 70% ethanol. The DNA was then resuspended in 20µl water which was then evaporated off to give a cleaner pellet. The DNA was resuspended in 8µl formamide stop solution as a loading dye prior to separation on a gel.

2.5.4 Preparation of a GA ladder

1µl 5'-labelled Junction DNA was added to 25µl water and 5µl formamide stop solution. The solution was incubated at 100°C with the cap of the microcentrifuge tube open for 60mins. The remaining solution was run along side the chemical probing samples as a sequence ladder. The G and A residues are visible as bands, having been cleaved by the formamide in the stop solution.

2.6 DNaseI protection.

A stock solution of DNaseI (7,200U/ml) was diluted to give 0.6U/ml in DNaseI buffer. 2µl of labelled junction DNA was incubated at 41°C for 10mins then 2µl of diluted DNase I was added. The reaction was allowed to proceed for 1min exactly then stopped by adding 4µl formamide stop solution. The products were run on a 16% denaturing gel.

2.7 SMART reaction

10X oligo mixes were prepared and aliquoted as shown in Table 2-6.

Oligo Type	Concentration in TE	Volume per 5µl of 10XA
1 st 3WJ template	100fmol/µl	0.5µl
1 st 3WJ extension	200fmol/µl	0.5µl
1 st 3WJ extension facilitator	20pmol/µl	0.5µl
1 st 3WJ template facilitator	20pmol/µl	0.5µl
2 nd 3WJ template	1.5pmol/µl	0.5µl
2 nd 3WJ extension	6pmol/µl	0.5µl
2 nd 3WJ template facilitator	30pmol/µl	1.0µl
TE	-	1.0µl

Table 2-6: Dilutions of the SMART oligos.

Cultures were prepared and grown 17hours in advance of the assay. Cells were diluted in peptone water to provide 1/10 or 1/100 dilutions as required for the experiment. Samples were assayed in duplicate. *E.coli* cells were monitored alongside *Salmonella* to give a negative control.

Mix A was prepared from the 10X A concentrate described in Table 2-6, plus NTPs. For each reaction, 0.5µl 10XA was added to 3.2µl 25mM rNTP mix (final concentration of each nucleotide is 16mM), 0.11µl 10mM dNTP mix (final nucleotide concentration of 0.22mM). 0.1µl 9pmol/µl 5'biotinylated probe, 1µl 5X lyophilisation concentrate and molecular grade water.

For each reaction, 5µl A was added to 5µl buffer C (section 2.1.1) and 10µl put into each reaction tube. 1µl 20KU/µl Lysozyme was added to 100µl of each broth dilution and the cells left to lyse for 10mins at 41°C. 5µl of each diluted culture was added to the appropriate reaction tubes. The tubes were then placed in a thermal cycler for 5mins at 95°C then cooled to 41°C at 0.1°C/sec. They were held at 41°C for 60mins to anneal.

An enzyme mix is made up to 30 μ l per reaction as follows: 4.5 μ l 10X transcription buffer (Ambion), 16.5 μ l molecular grade water, 6 μ l 5X lyophilisation concentrate, 2 μ l T7 RNA polymerase (200U/ μ l), 1 μ l Bst DNA polymerase (8U/ μ l). 30 μ l was added to each reaction while at 41°C. The reaction proceeds at 41°C for a further 90mins.

2.7.1 Enzyme Linked Oligosorbant assay (ELOSA)

The Horse Radish Peroxidase conjugated detection probe was added to the hybridisation buffer (section 2.1.1) to a concentration of 0.5pmol/reaction. 55 μ l of H+G was then placed in each well of a 96well micro titre plate. The approximately 50 μ l of the SMART reaction is transferred to the relevant well. An ELOSA positive well was created by the addition of an oligo complementary to the detection probes. The plates are incubated on a plate shaker for 30mins at room temperature.

A solution of TMB Blue is prepared for use in the ELOSA and kept in the dark until needed. PBST is used as a wash solution. 2X 200 μ l of the solution are used to wash out the wells, the contents being discarded each time. 100 μ l TMB Blue is added to each well and incubated at room temperature for 3 mins exactly. 100 μ l stop solution (1 part HCl, 11 parts sterile water), is added after the incubation period. The optical density of each well is read in a plate reader.

2.7.2 DNA Polymerase extension assays

The junctions used in this assay are the same as those used in the SMART assay. Only the extension oligo is 5' labelled. 1 μ l junction DNA is added to 1 μ l of a 10mM dNTP mix, 5 μ l of the polymerase 10X buffer and 43 μ l water. Microcentrifuge tubes were prepared, labelled 0,1,2,5,10 and 30mins. 4 μ l formamide stop solution was added to each. 8 μ l of the reaction solution was removed and added to tube 0 before the addition of the enzyme. The remaining reaction solution was incubated at the appropriate temperature (37-50°C) for 10mins to bring it up to temperature. 1 μ l of the polymerase was added and 8 μ l samples removed at each of the appropriate time

points. The samples were denatured by crash cooling and run on a 12% denaturing polyacrylamide gel.

DNA polymerases used for the extension assays were Bst DNA polymerase (New England Biolabs (NEB) 8U/ μ l), Klenow (NEB 5U/ μ l), KOD (Novagen 2.5U/ μ l), Fast start Taq (Roche 5U/ μ l) and Sequenase (USB 13U/ μ l).

2.7.3 Transcription assay

The SMART assay was set up using J2 oligos, all of which were unlabelled. The reaction mix contained both DNA and RNA polymerase. 1 μ l of 32 P labelled α -UTP was added to the reaction along with the standard concentration of NTPs. Conditions were kept as sterile as possible, having used RNase Zap (Sigma) to eliminate any RNase and DEPC water was used where needed. The reaction was allowed to proceed for the full 90mins and then stopped by putting on ice and precipitating. The products were resuspended in 10 μ l of RNase-free ficoll loading dye and loaded onto a 15% denaturing polyacrylamide gel.

2.8 RNA purification

Purification of total cell RNA was performed using RNeasy Mini kit from Quiagen. *E. coli* and *S. typhmuri* cells were cultured for 17 hours over night at 37°C. A 1 in 10 dilution of the cells was made to give approximately 10^9 bacteria per ml. 12 samples were set up for each of the cultures. Following purification, those of the same culture were pooled to give a greater amount of RNA.

The bacteria were harvested by centrifuging at 5000g for 5 mins at 4°C. The supernatant was decanted and the remaining media carefully aspirated. The pellet was loosened by flicking the tube, and was then resuspended in 100 μ l lysozyme containing TE. The suspension was then incubated at room temperature for a further 5 mins. 350 μ l of the supplied buffer RLT was added and mixed by vortexing. 250 μ l

96% ethanol was added and mixed by pipetting. The samples were applied to an RNeasy column in a 2ml collection tube. The tube and contents were centrifuged briefly for 15 seconds and the flow-through discarded. 700 μ l of buffer RW1 was applied to the column, centrifuged and the flow-through discarded. The column was transferred to a fresh collection tube, then 500 μ l of buffer RPE was added to the column and centrifuged. This step was repeated once more to wash the silica gel membrane.

The column was transferred to a 1.5ml tube for elution of the RNA. 40 μ l RNase-free water was pipetted onto the membrane and the tube spun gently to release the RNA. After being pooled, the RNA samples were aliquoted into 5 μ l amounts and stored at -80°C.

2.8.1 Quantification of the RNA yield.

The RiboGreen kit from Molecular probes (R-11490) was used to quantify the RNA. The RiboGreen reagent was thawed from -80°C and kept in the dark as much as possible to prevent photo-degradation. 20X stock TE was diluted to 1X with DEPC treated water (0.01%DEPC incubated overnight and autoclaved). A low range working concentration of the reagent was made using a 200 fold dilution.

To prepare a standard curve, 2 μ g/ml RNA stock was diluted 20 fold to make 100ng/ml. Dilutions were prepared as described in Table 2-7.

Volume TE (μ l)	Volume 100ng/ml RNA stock (μ l)	Volume 2000-fold diluted RiboGreen reagent (μ l)	Final RNA concentration in assay
0	100	100	50ng/ml
50	50	100	25ng/ml
90	10	100	5ng/ml
98	2	100	1ng/ml
100	0	100	blank

Table 2-7: dilution of the stock RNA to create a standard curve. The above dilutions are for the low-range RiboGreen assay.

1 in 10 and 1 in 100 dilutions of the purified *Salmonella* and *E.coli* RNA were prepared in duplicate and run alongside the standard curve. 100 μ l of each sample was added to 100 μ l RiboGreen as above. The samples were placed in a 96 well clear plate and incubated at room temperature for 5 mins. The plate was read in a Flurostar Optima BMG fluorescence plate reader and the concentration of RNA calculated. Since the majority of RNA in the cells would be from the 16S Ribosome, the concentration of target was taken as the total yield of RNA.

2.9 Redundant junctions

2.9.1 Annealing and purification

To determine whether junctions of different sequence could be separated out on a gel, a series of junctions were made containing redundant, N bases at four positions around the branch point (see Figure 2-8). The result is a set of sixteen possible junction combinations. The oligos were annealed in high concentration (800pmol) in a 60 μ l volume with 10 μ l NEB buffer 2. The oligos were annealed as slowly as possible in 1l water at 95°C, and left to cool to room temperature. 30 μ l Ficoll was added as a loading dye. After annealing, 10 μ l of the junction mix (80pmol) was run on a 20% non-denaturing gel at 100V for 8h. Bands were identified and cut out

using a UV transilluminator following staining with ethidium bromide. These bands were incubated in TE overnight as described in Section 2.2.3.

2.9.2 Restriction digest with BamHI

The DNA purified from the gel was precipitated and reannealed in 10µl water. 2µl Buffer E (Promega) was added along with 2µl BamHI (Promega 10,000U/ml) and 6µl water. The reaction was incubated at 37°C for 3h.

To inactivate the enzyme, the digestion mix was precipitated and washed with 70% ethanol. The DNA was resuspended in 20µl water.

2.9.3 Plasmid ligation

10µl of the digested DNA was added to 2µl 10X ligase buffer (Promega), 1µl pUC18 cut with BamHI, 5µl water and 2µl DNA ligase (Promega 3,000 U/ml). The reaction was left at room temperature for 3h.

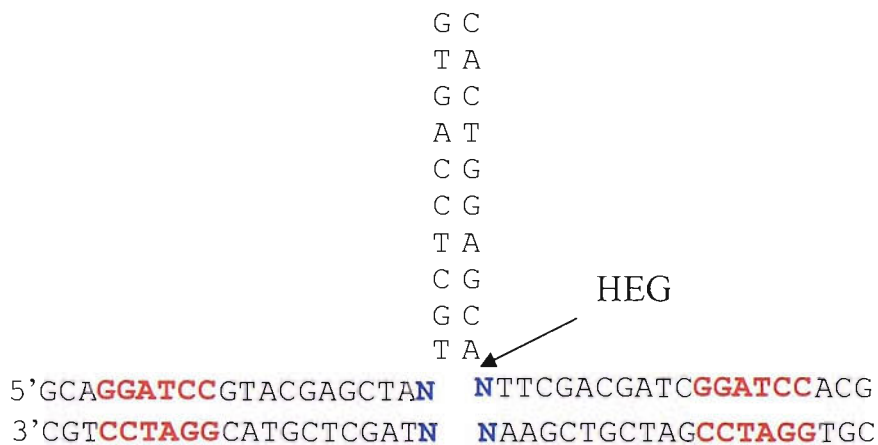


Figure 2-8: Redundant junction. The junction shown above contains one fixed base pair around the junction, in this case TA, and four N bases. This leads to sixteen possible combinations of junction. Three other sets of junctions were made, containing GC, AT and CG fixed base pairs. The arm shown as extended in the diagram is shorter than the other two. This means that the junction will run differently on the gel depending on which conformer is dominant in each junction. The junction contains two BamHI sites, marked in red. These allow ligation into a plasmid for sequencing.

2.9.4 Transformation in to E.coli (TG2)

10 μ l of the ligated DNA was added to 200 μ l of TG2 competent cells and left on ice for 30mins. The cells were incubated at 45°C for 1 min exactly and then returned to the ice to heat shock them. 20 μ l of the transformation mix was transferred to an agar plate containing 0.1mg/ml carbenicillin, 0.02% X-gal and 1mM IPTG. The remaining mix was transferred to a second plate and both were grown at 37°C overnight.

2.9.5 Selection of transformants

White colonies on the plate were selected and picked using an inoculating loop. The colony was then transferred to sterile 5ml culture containing 0.1mg/ml carbenicillin. The culture was grown in a shaking incubator at 37°C overnight. When the culture had grown, cells were transferred to a microcentrifuge tube and the remainder discarded. The tube was spun down for 5mins at 6000 rpm in a microcentrifuge and the supernatant discarded. A Quiagen miniprep kit was used to purify the DNA from the cells.

2.9.6 Sequencing of the plasmid and insert.

10 μ l of 2M NaOH was added to 40 μ l of the plasmid DNA and left at room temperature for 10mins. The DNA was precipitated by adding 15 μ l 3M sodium acetate, 35 μ l water and 300 μ l 100% ethanol. This was left on dry ice for 10mins then spun down in a microcentrifuge at 14K rpm for 10mins. The pellet was washed with 70% ethanol. The DNA was resuspended in 10 μ l DNA and 2 μ l annealing buffer and 2 μ l universal primer (USB T7 sequencing kit) were added. The reaction was left at 37°C for 20mins and at room temperature for a further 10mins. Meanwhile, for each sequence, four tubes were prepared labelled A,C,G and T. 2.5 μ l of the appropriate dideoxy mixes A,C,G,T-short were added to the tubes. The polymerase mix was prepared on ice in two separate stages, to one tube, 12 μ l of label mix A was added to 7 μ l of water and 1 μ l α -³²P ATP. In a separate tube, also on

ice, the enzyme was diluted by adding 6.5 μ l enzyme dilution buffer to 1.5 μ l T7 DNA polymerase. The diluted enzyme was added to the labelling mix when needed. These quantities were sufficient to sequence four sequences.

After the annealing time, 6 μ l of the polymerase mix was added to each of the sequences. These were left at room temperature for 5 mins while the dideoxy tubes were incubated at 37°C for 5 mins. Following this time, 4.5 μ l of the sequence was added to each of the four dideoxy tubes. The reaction was left for 5mins then stopped with 5 μ l stop solution. The samples were crash cooled and run on a 12.5% denaturing gel.

2.9.7 Analysis of sequencing results

The sequencing results were run for about 2 hours, until the first marker had reached the bottom of the gel. Once the gel was dry, it was exposed to a phosphor-imager screen for 20mins only and scanned. The sequence between the two BamHI sites was located and related to the sequence of the junction. Since only the longer arms of the junction were cut with BamHI and could anneal into the plasmid, the junction could clone in one of two ways. Whichever way the junction was incorporated, when the plasmid was taken up by the cells and replicated, only the continuous strand should remain. The other two forming the third arm should be discarded because the replication machinery should not be able to read around them. The opposite strand would then be copied using the continuous one as a template. The sequence between the BamHI sites was checked for mutations and the two N bases identified.

Chapter 3

ANALYSIS OF THE CONFORMATIONAL DIFFERENCES BETWEEN THREE- WAY JUNCTIONS

This chapter looks at junctions known to behave differently in the SMART assay (MecA and Coag) and uses a combination of gel mobility assays and chemical probing to analyse the conformational differences between them. The work then progresses to altering the sequence around the branch point to alter the overall conformation.

3 Analysis of the conformational differences between three-way junctions

3.1 The conformation of DNA three-way junctions

Three-way junctions are expected to exist in an equilibrium between two stacked conformations in solution. The difference in free energy between the two forms varies. In most cases, the difference is enough to favour one conformation over the other (the dominant conformation). This can be seen most clearly in global analysis over a long time scale. Analysis of the mobility of junctions under electrophoresis is a commonly used technique and very useful for analysing the relative size of the angles between the arms in the dominant conformation (long-short arm assay). Occasionally, due to the dynamic nature of the equilibrium, intermediate patterns are obtained. These may represent un-stacked junctions, those with no differences in free energy between the two forms and those with alternative conformations. These patterns will be discussed in greater detail later on in this chapter. Gel mobility has been confirmed as a viable technique by comparison to other methods such as AFM and FRET studies (Cooper and Hagerman, 1987).

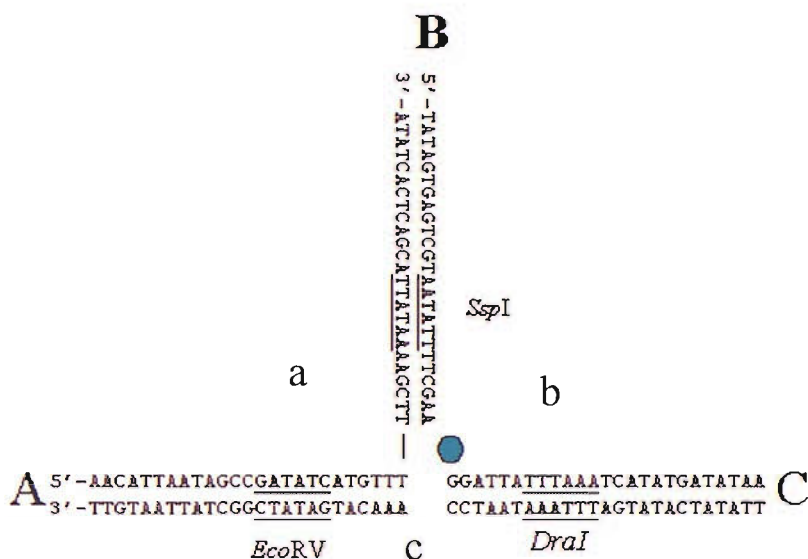
The sequence of the junction determines the bias of the equilibrium but the rules that govern preference for folding are unknown (Welch *et al.*, 1995, van Buuren., *et al* 2001, Assenberg *et al.*, 2001). The junctions used in SMART can be taken as basic bulged 3WJs. During its development, the SMART assay has been used to detect a variety of different target sequences in genes and bacterial sequences. A number of different sequence versions of the SMART junction exist. Two of these junctions are MecA and Coag, designed to detect the *mecA* and coagulase genes present specifically in MRSA. A positive result from the two tests performed on the same sample, confirms the identity of the infection as MRSA and not a similar strain. These two junctions are important to understand because they behave very differently in terms of signal strength in the SMART assay.

To understand the difference between the two junctions, full-length versions were made for easier use in the assays (section 2.12b). Synthetic oligos for the target, extension and template probes gave a junction with three equal arms of twenty-five base pairs. Unique restriction sites were incorporated into each arm a short distance from the branch point. These sites were changed as little as possible from the original sequence to avoid effects on junction conformation. All restriction enzymes used were blunt cutting to ensure that overhangs after cutting did not affect gel mobility (Assenberg *et al* 2002). The terminology used in this chapter is based on that set out by Altona (1996), where the arms are labelled A, B and C and the strands of DNA are labelled a, b and c in a clockwise direction starting from the bottom left.

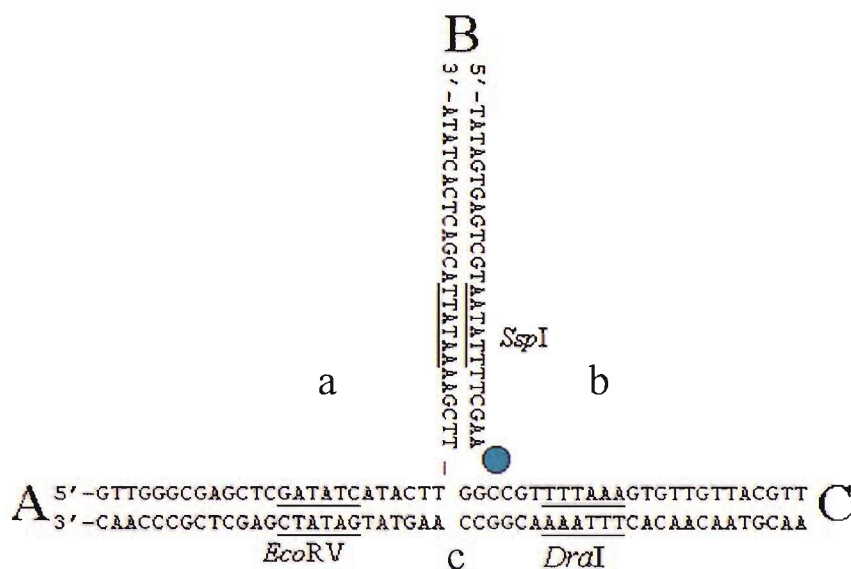
3.2 MecA and Coag

Junctions MecA and Coag give very different signal outputs from the SMART assay. One possible source of this difference was that the two adopted different dominant conformations. The sequences of the two junctions can be seen in figure 3-1. MecA is very AT rich along the target region, arm AC. This may cause it to differ substantially in conformation from Coag (the A-T base pair contains only two hydrogen bonds whereas the G-C contains three, making it much stronger). Since Coag is the more efficient of the junctions at producing a signal in the SMART assay, it is possible that this is because it is more stable.

Figure 3-2 shows the pattern of mobility of MecA and Coag when each arm is cut in turn in the long-short arm assay. The fastest migrating fragment contains the co-linear (stacked) arms. The identity of this fragment defines the dominant conformation. In the case of both MecA and Coag, the fastest running fragment in the presence of magnesium is AB. This means that arms A and B are stacked and the junctions exist predominantly in conformer AB. The distinction between the fragments on the gel is good and there is no visible difference between the junctions. Also shown in figure 3-2 is a simplified version of the result showing the interpretation of the pattern and the angles between the arms.



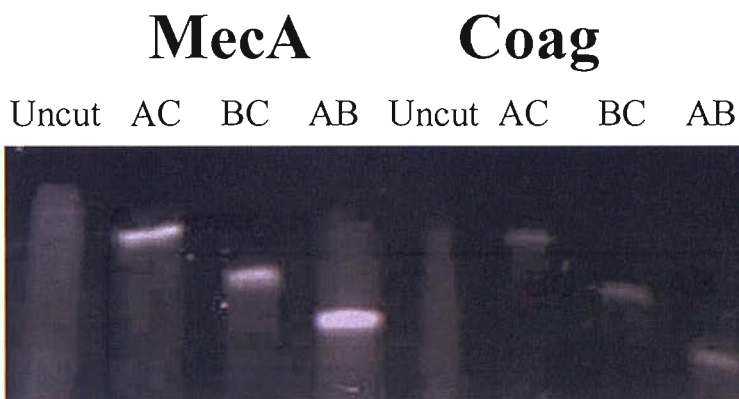
MecA junction



Coag junction

Figure 3-1: The MecA and Coag junctions displayed with the position of the restriction sites and octanediol (blue circle). The junctions above are displayed in the same format and orientation as those used in SMART. See junction list in chapter 2 for a comparison with other formats used in different papers. Strand a, which runs 5' to 3' through arms A and B, would be the extension probe in the SMART version with strands b and c (the template and target, respectively) clockwise around the junction.

a)



b)

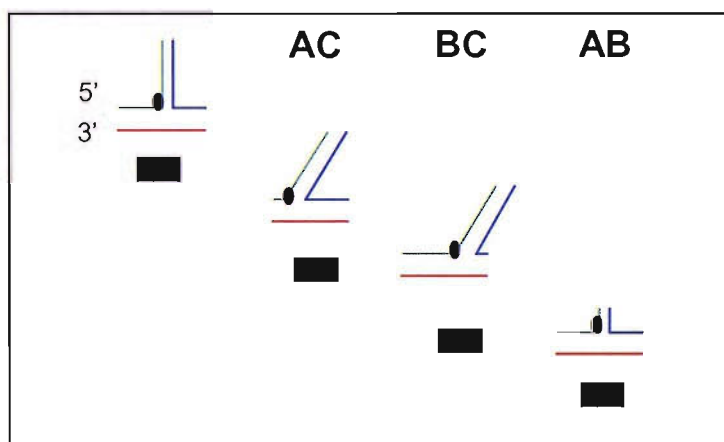
Conformer AB

Figure 3-2: The long-short arm assay of three-way junctions MecA and Coag with 1mM magnesium. 2a shows the mobility with 1mM magnesium in the buffer and polyacrylamide gel, allowing the junctions to maximise their stacking interactions. The two junctions are shown side by side for comparison. The labels uncut, AB, BC and AC refer to the state of the junction, either a full three-way junction or the pairs of arms which remain. 2b shows how the bands on the gel correspond to the angles between the arms for two different junctions. The bottom figure is a partial reproduction of figure 1-10 showing the different patterns produced by conformers AB and AC.

At first, it is surprising that both MecA and Coag show an identical pattern in the LSA when they show such different results in the SMART assay. When the core sequence of the two junctions are compared, the centre of the junctions, up to two base pairs along each arm, is identical. Whatever causes the difference between them in terms of performance, it is not the overall conformation. In this case the identical central sequence probably controls the conformation.

3.3 Probing the junctions for single stranded regions

In order to identify other potential causes of the differences between MecA and Coag, chemical probing was used to search for single stranded regions which might show one of the junctions not folding correctly. For this osmium tetroxide was used to react with single stranded thymines and diethyl pyrocarbonate (DEPC) to probe for single stranded adenines. Both methods have a much stronger reaction with single stranded than native DNA, due to the mode of attack by both chemicals, from above the plane of the base. The stacking of the bases in the double helix gives protection against this. The MecA junction, being very AT rich would be more likely to loop out due to the lower number of hydrogen bonds between the AT base pairs compared to the GC base pairs.

Figure 3-3 shows a control reaction with DEPC and osmium tetroxide. A single stranded oligo was exposed to each chemical and stopped at the indicated time points. The sample then underwent cleavage with piperidine, and was run on a polyacrylamide gel to view the length of the fragments produced. By 5 mins exposure to DEPC and 3 mins exposure to osmium, the single stranded DNA has been completely digested due to the intensity of the reaction. In the experiments with junction DNA, trial and error showed that a reaction time of 10mins gave the best results. The results on the single stranded DNA provide a measure of the reactivity of both compounds with any single stranded regions.

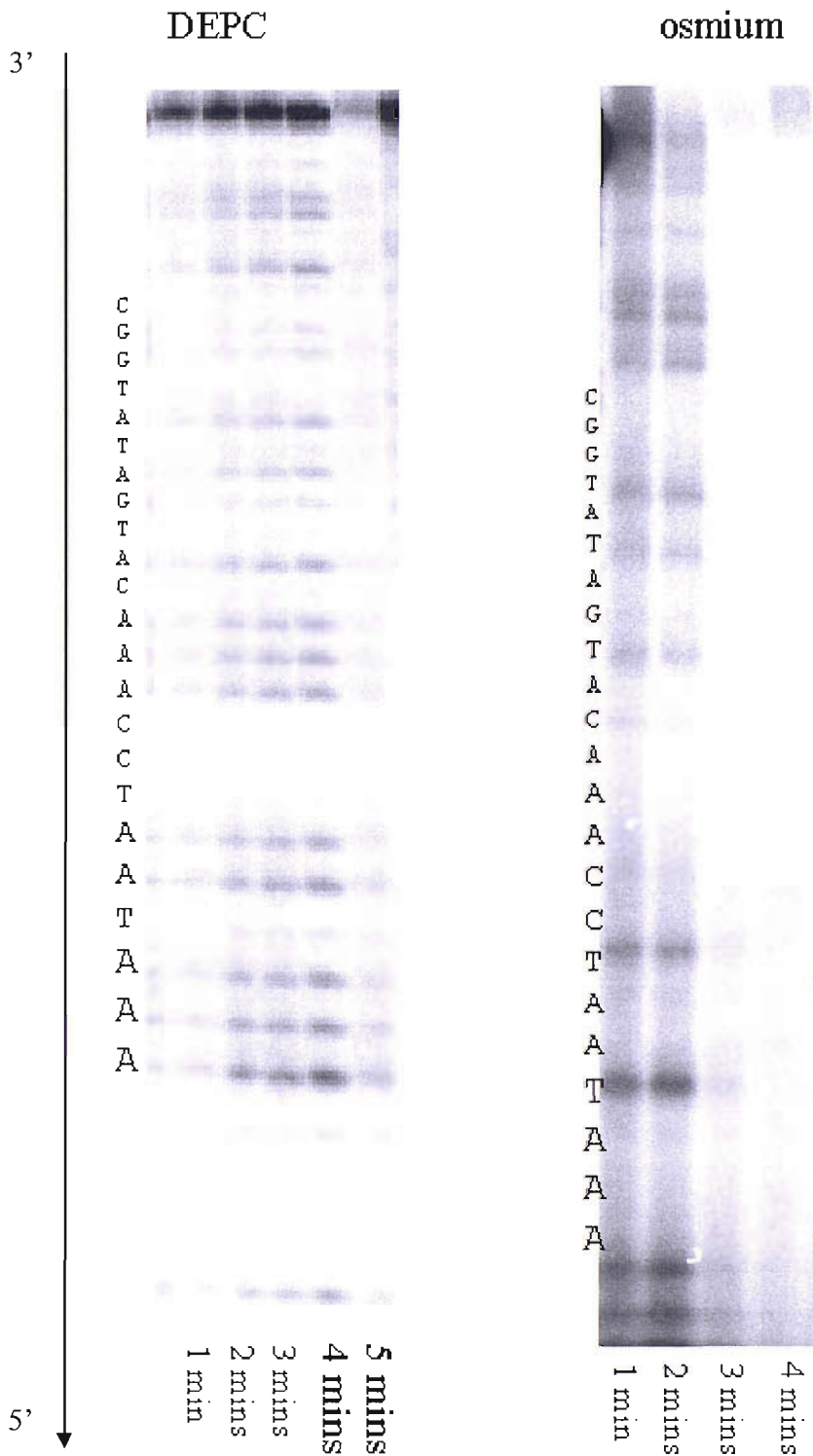


Figure 3-3: Control gels showing the effect of exposure of single stranded DNA to DEPC and osmium tetroxide. The sequence used for this control was Meca target (c) and is shown on the left hand side of the gels, matching up to the cleaved bases. The reaction was stopped after progressively longer periods up to 5 mins.

Figure 3-4 shows the results of a 10 minute exposure of junctions MecA and Coag to DEPC. The assay was performed at 41°C to make the conditions as similar as possible to the SMART reaction. There was magnesium in the junction buffer to allow folding and stacking of the bases. Neither MecA nor Coag shows a strong reaction with DEPC, not even around the branch point. There is a background level of banding due to some reaction with the double stranded DNA but there are no adenines that stand out. In contrast, Figure 3-5 shows a strong reaction of thymines around the branch point with the osmium tetroxide in both junctions. A box has been used to highlight the two bases either side of the branch point in the figure. The intensity of the bands should only be compared within the lane, not between lanes. Although the bands appear more intense in strand a of the MecA junction than in strand a of Coag, the background cleavage in the former is also more intense. This could be due to small differences in the time taken to stop the reaction, the concentration of the DNA and the amount loaded in the well.

The results from Figure 3-3 demonstrated that the reaction of single stranded DNA with osmium tetroxide is stronger than that with DEPC. Since the results from figure 3-5 show reproducibly that there are thymines around the branch point which can be single stranded, it could be that this looping out of bases only occurs in a minority of the junction population and that DEPC exposure is not sensitive enough to pick this up.

To demonstrate that a looped out base could be detected by DEPC within the junction context, two different b strands containing unpaired bases instead of the octanediol as the linker were made for MecA. One contained two unpaired adenines, which should show a strong reaction to DEPC, the other contained two thymines to be shown up in the osmium tetroxide reaction. The results are shown in Figure 3-6. In both cases, the unpaired bases do show up well against the background. The results from the osmium tetroxide exposure are slightly over digested but do show the bases around the branch point in strand a and the unpaired bases in strand b are very reactive.

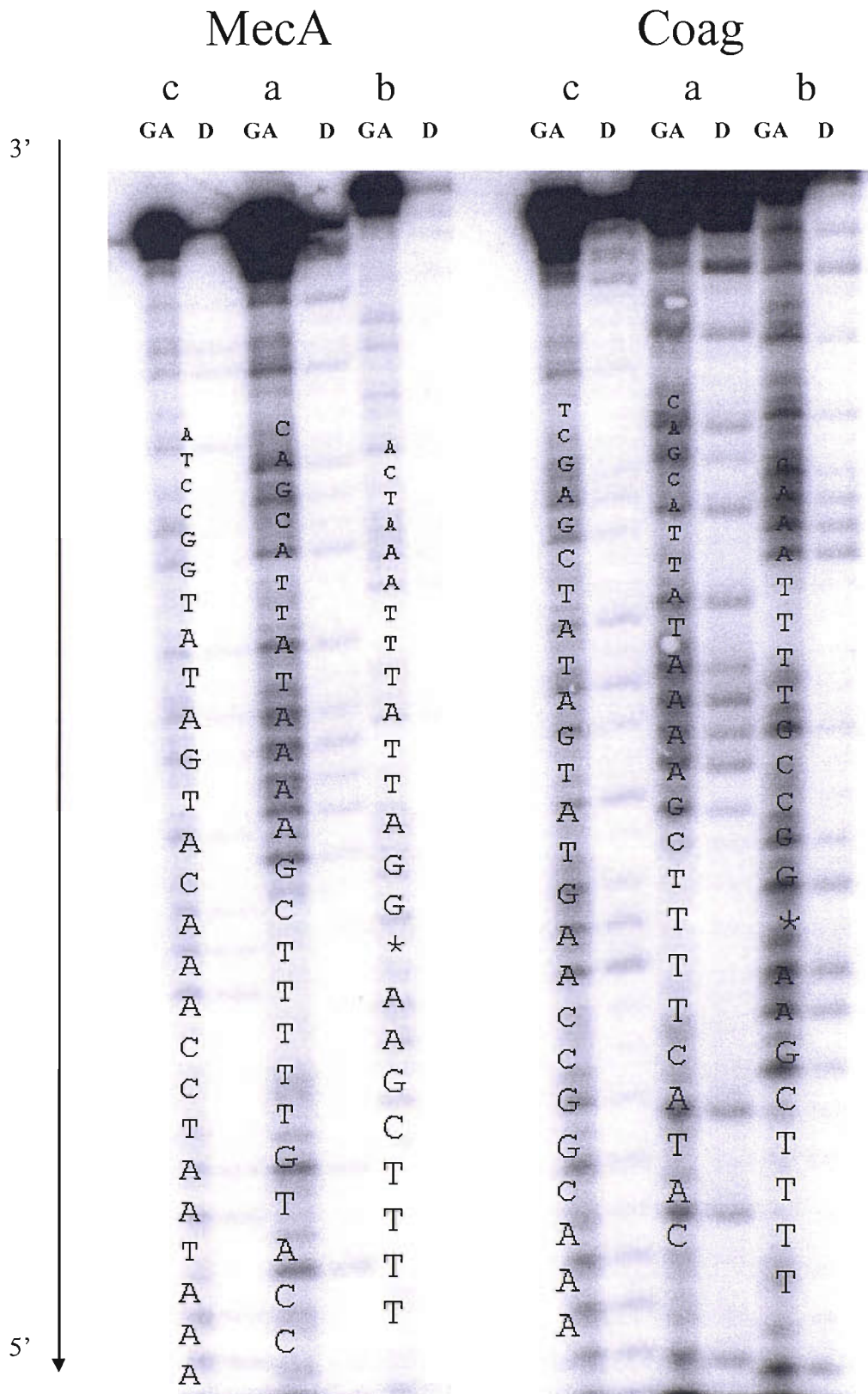


Figure 3-4: 10 minute exposure of junctions MecA and Coag to DEPC. The labels a, b and c refer to the identity of the labelled strand. In each case it is the 5' end of the DNA that is labelled. The same junction is probed three times, each time with a different strand labelled to give the results from a different point of view. A GA sequence lane is shown prior to each result for comparison, indicated by GA above the lane. The label D above the lane refers to the results of the DEPC exposure. The * symbol in the sequence lane indicated the position of the chemical linker.

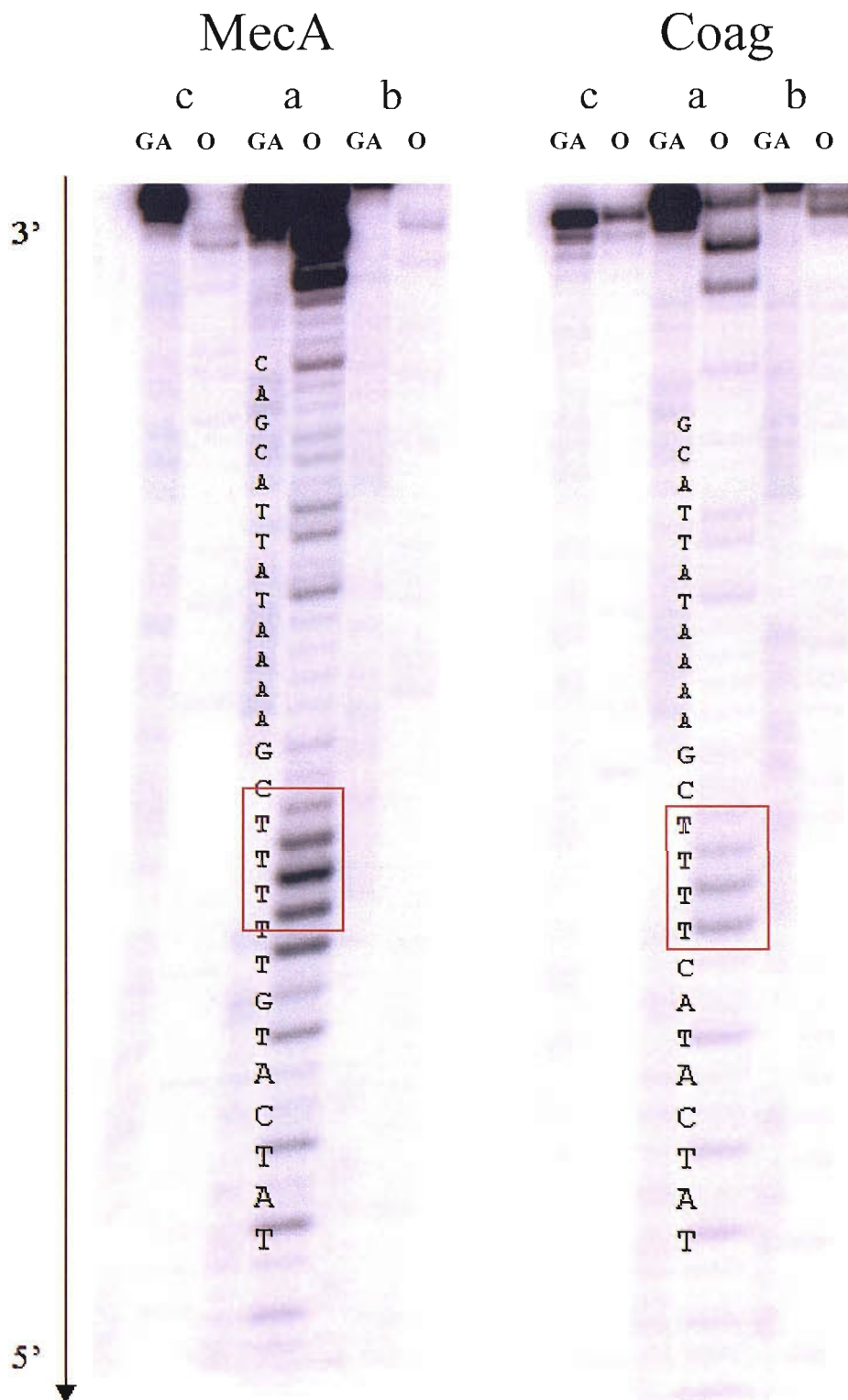


Figure 3-5: 10 minute exposure of junctions MecA and Coag to osmium tetroxide. The figure above takes the same format as the DEPC probing shown in Figure 3-3. Only the sequence of lane a is shown in each junction to highlight a region of thymines around the branch point. The location of the branch point is shown by a red box surrounding the two bases on either side of it. The GA designates a GA sequence lane while O indicates the reaction with osmium tetroxide.

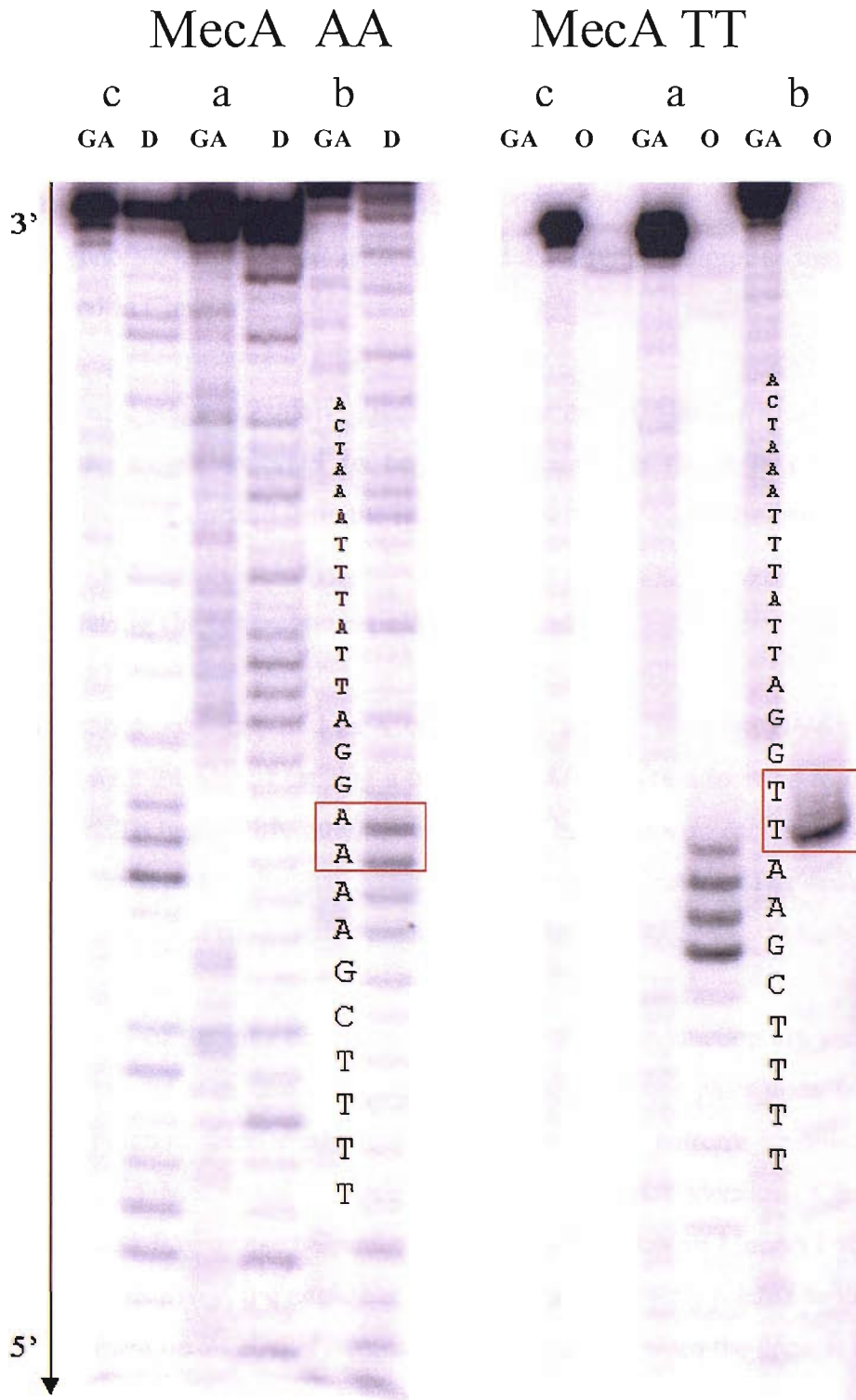


Figure 3-6: Control gels showing the effect of using unpaired bases at the branch point rather than a chemical linker. Two different b (template) strands were made to complement the MecA junction. One used two unpaired adenines at the branch point and the other, two thymines (indicated by a red box). This allows the effect of chemical probing to be seen in a junction context. Only the sequence of strand b is indicated on the gel, with a red box surrounding the two unpaired bases acting as the linker. The left hand gel with MecA AA was exposed to DEPC only while the right hand gel, MecA TT was exposed to osmium tetroxide only.

Both MecA and Coag show the same pattern of cleavage and the same conformation in the long-short arm assay. It appears, so far at least, that the two have very similar folding patterns. It may be that, while the versions of MecA and Coag adapted for these assays do not show any differences, the original SMART junctions (consisting of a genomic target, short extension and long template probe) may. This will be explored in Chapter 4.

As the next stage in exploring junction conformation, MecA was taken as a standard junction, which appeared to be predominantly in the A/B conformation, and compared to a junction from the literature reported to adopt conformer A/C.

3.4 Altering the central base pairs of the junction

The central sequence of MecA was altered to that used by J1V6 (Welch *et al.*, 1995), with the aim of causing a conformer A/B junction to swap to that of an A/C junction. To begin with, only the central two base pairs on each side of the junction branch point were changed, generating junction MecAV1. The sequences of the central regions of the junction are shown in Figure 3-7.

The results of the long-short arm assay with this new junction are shown in Figure 3-8. It can be seen that changing the central two base pairs does not change the MecA junction from conformer A/B to A/C. The patterns produced from the cleavage of MecA and MecAV1 are, however, not identical. Close inspection reveals that fragments BC and AB are closer together in MecAV1 than in MecA. Since in this assay, the angle between the arms is directly related to the mobility of the fragment on a gel, this suggests that the angle between the arms is smaller in the new junction.

Since changing the central two base pairs was not sufficient to alter the conformation to an A/C conformer, similar experiments were performed with sequences in which the central six base pairs were changed. This led to the construction of junction MecAV2, containing the central six base pairs of J1V6. In all other ways MecAV2 was identical to MecA and MecAV1. The gel in Figure 3-9

shows the long-short arm assay of MecAV1 and V2. This time the patterns from the two junctions differ greatly, but V2 still does not show the pattern of a conformer AC. If this pattern represents a single dominant conformation, then the greatest angle must be between arms B and C. The linker is placed between these two arms. Arm A would be extended with equal angles between arms AB and AC. This would represent an unstacked conformation. Since the junction appears to be predominantly unstacked, this may reflect that it is in a rapid dynamic exchange between conformers and the unstacked form is only the averaged conformation.

Where junctions are suspected to exist in a more open conformation, chemical probing can provide some much more specific information about which residues are exposed more often than others. The chemical probing of junction MecA V1 and V2 with DEPC is shown in Figure 3-10. This reveals a strong reaction with an adenine on the 3' side of the octanediol in strand b, which appears to be equally reactive in both junctions.

Similar experiments examining the exposure of thymines to reaction with osmium tetroxide are shown in Figure 3-11. It can be seen that the thymine opposite the exposed adenine in strand b is also hyper-reactive to osmium tetroxide. This reaction is so strong that even when the exposure time is dropped to 7 minutes, the DNA in strand c is still over digested. The over digestion of the DNA in these lanes means that any exposed bases 3' to the reactive thymine are not visible. There is another thymine on the 3' side of the branch point on strand c. It is possible that this one is also reactive, but if this was the case, the adenine opposite, on strand a, should also be much more intense than the background bands. The band is visible in Figure 3-10 but does not show a strong reaction with DEPC. This suggests that it is only the thymine on the 5' side of the branch point in strand c that is single stranded and that the thymine on the 3' side is correctly base paired.

It is interesting to note that MecA V1 and V2, which have the same sequence right at the branch point, have very different folding patterns yet show a similar pattern of chemical cleavage. This suggests that both MecA V1 and V2 spend more time in between conformers and may have hyper-exposed residues for this reason.

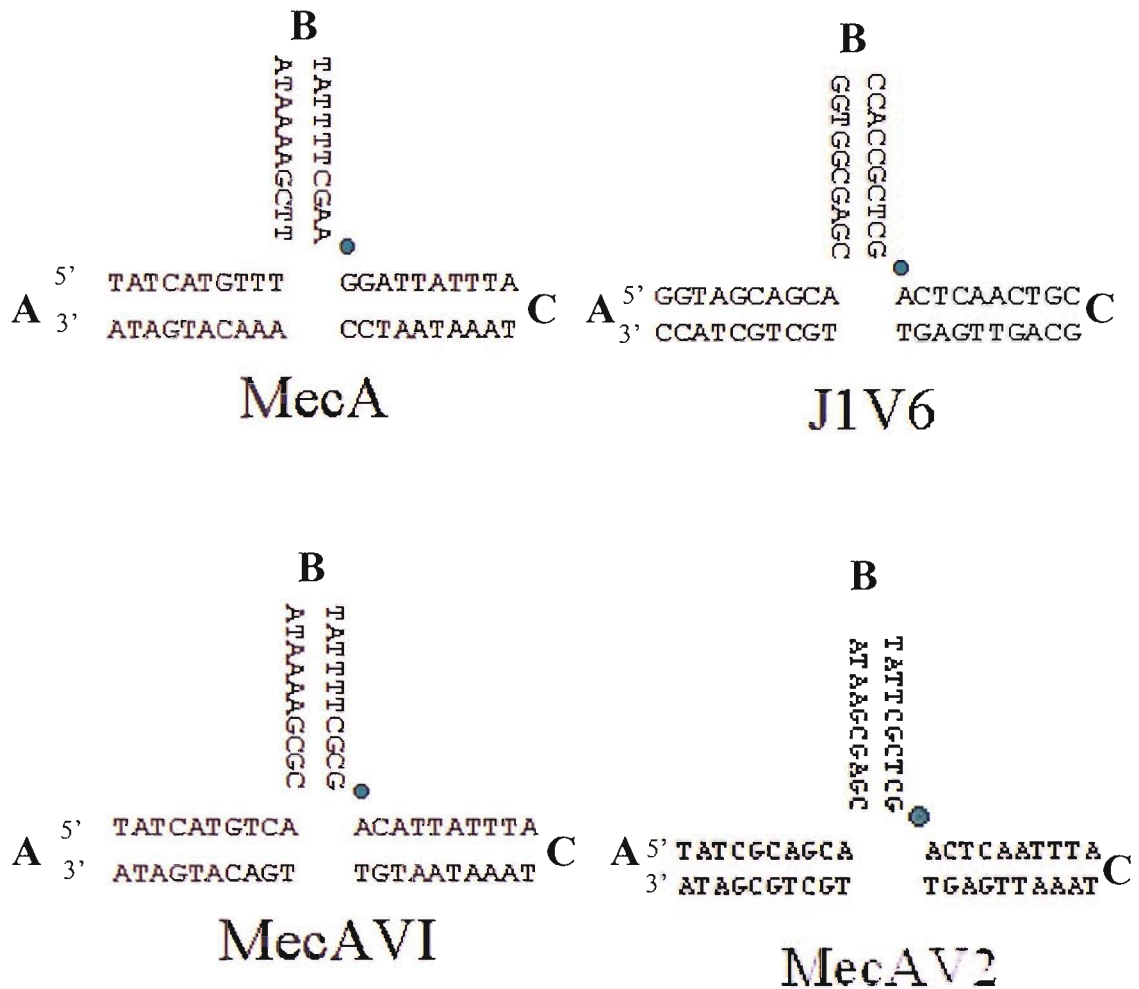


Figure 3-7: Comparison of the central ten base pairs of MecA, J1V6, MecAV1 and MecAV2. In MecAV1, the central two base pairs on each arm of the MecA junction, just around the branch point, have been swapped for those of J1V6. The aim of this being to cause MecAV1 to adopt a conformer A/C. The blue circle shows the position of the octanediol in MecA, MecAV1 and MecAV2. J1V6 uses two unpaired adenines as its linker (Welch *et al.*, 1995). Only the central base pairs are shown. The remainder of the sequence is identical to that shown for MecA in Figure 3.1.



Figure 3-8: Long-short arm assay using MecA and MecAV1. As in Figure 3-2, the two junctions have been run on the same polyacrylamide gel for comparison in the presence of 1mM magnesium chloride in the buffer and gel. An uncut junction can be seen in lane 1. The remaining lanes all show junctions with one arm shortened. The identity of the remaining full length arms is indicated above the lane. The gel was run for 8hours at low voltage (100V) to obtain good separation between the bands.

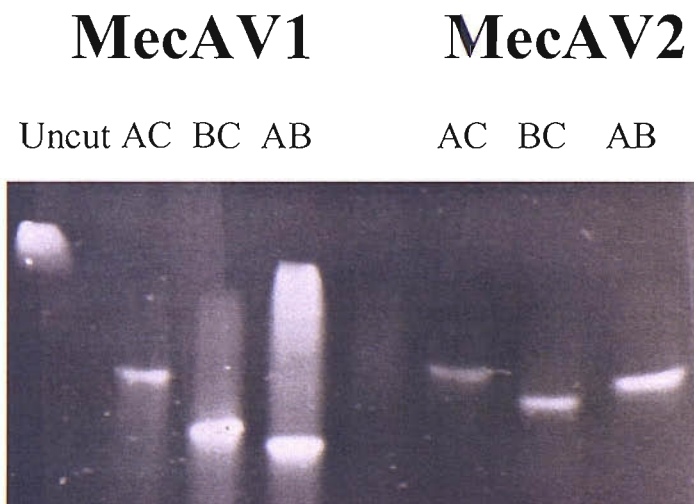


Figure 3-9: Long-short arm assay using MecAV1 and MecAV2. Junction MecAV2 includes the central six base pairs from the J1V6 junction and, as such, should be expected to show a conformer AC pattern. The conditions are identical to those described above in Figure 3-8.

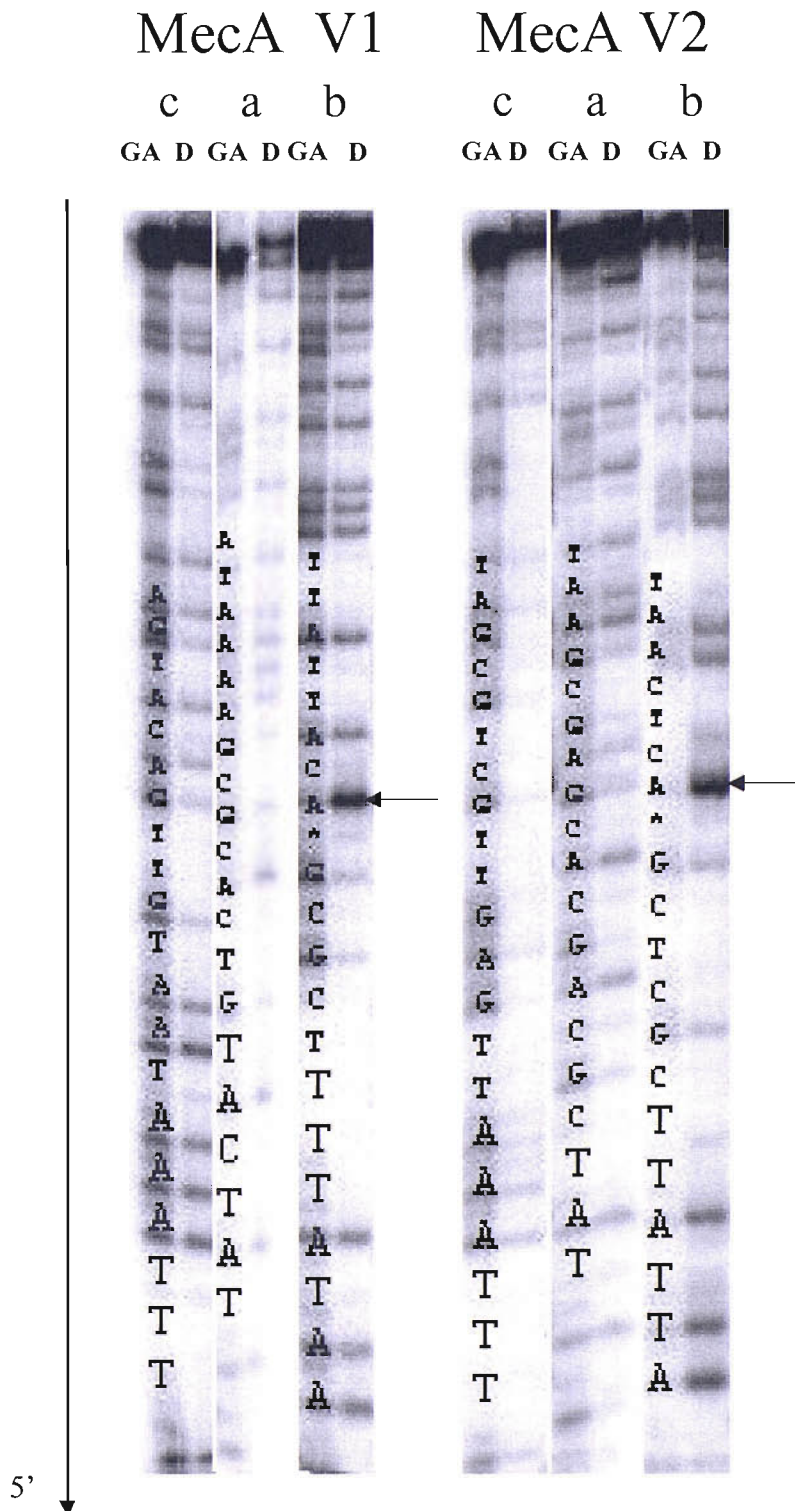


Figure 3-10: DEPC probing of MecAV1 and V2. The assay conditions used to produce this gel are identical to those used for MecA. The reaction with DEPC is stopped after 10mins and the products are treated with piperidine. The sequence of the labelled strand is overlaid on the GA marker lane previous to each result. The DEPC treated DNA is indicated with a D above the relevant lane. An arrow has been used to pick out a particularly reactive adenine in strand b of both junctions.

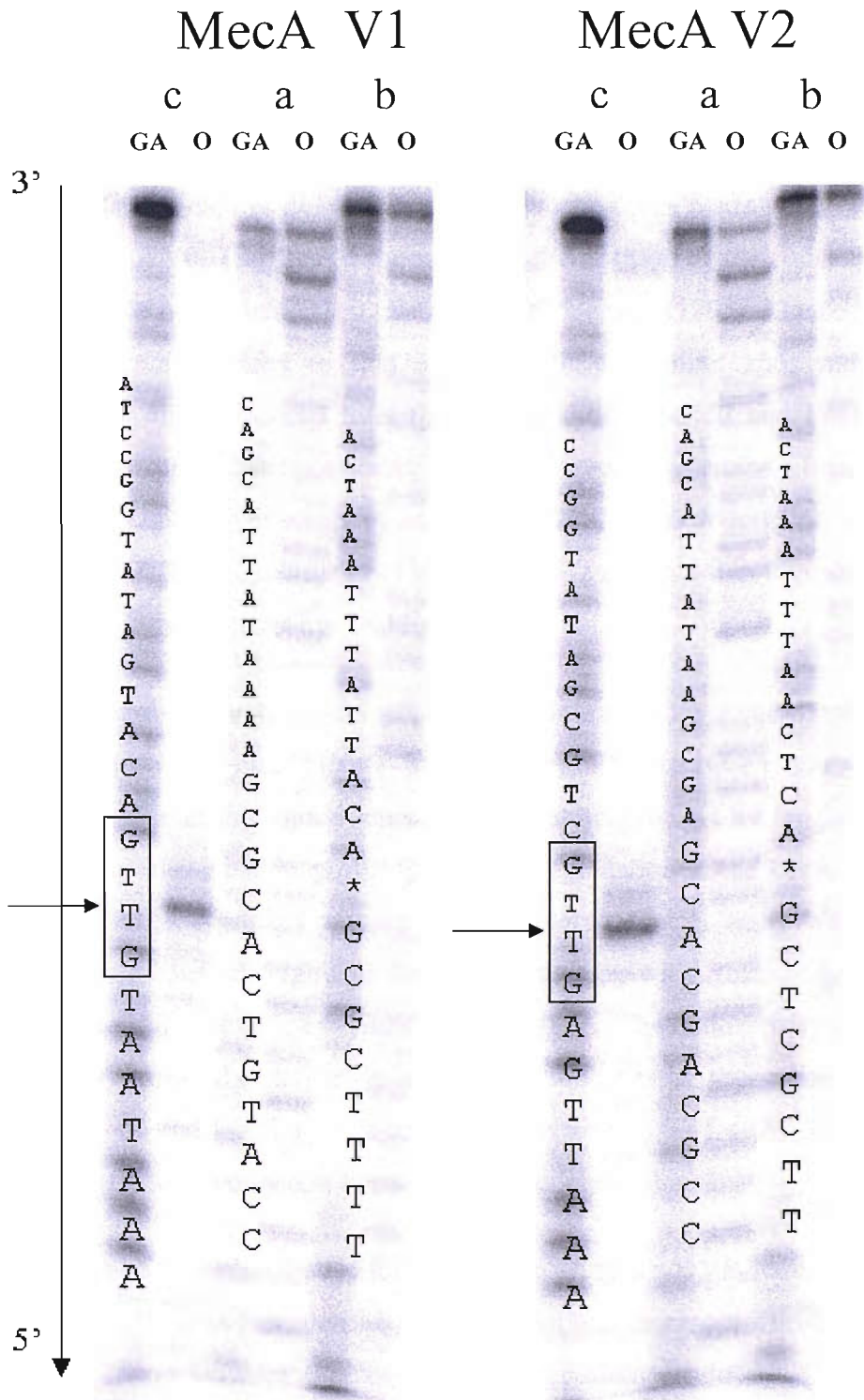


Figure 3-11: Exposure of MecAV1 and MecA V2 to osmium tetroxide. The reaction time was dropped from 10mins to 7 mins to give the result above. The DNA in lane 2 in both junctions is over digested even at the lower reaction time. This means that the bands 3' of the branch point in strand b are not visible on this gel. The reaction with strands a and c are normal. The branch point on strand b has been marked with a box over the two bases on either side. The reactive thymine on the 5' side of the branch point has been marked with an arrow.

3.5 Exploring the effect of sequence further from the branch point.

To confirm that the original junction J1V6 does adopt conformer A/C as reported in the literature (Welch *et al.*, 1995), another junction was made which was closer to the original J1V6. The size of the junction, the position of the restriction sites and the sequence were identical to that from Welch *et al.*, (1995). The restriction sites, however, were changed to that of MecA to give blunt ends rather than the overhanging ends produced from the original XbaI, HindIII and EcoRI. This was necessary to rule out the possibility that the 3' overhangs would affect the mobility of the cut fragments (Assenberg and Fox, 2001). The new junction was given the name J1V6-2. The sequences of J1V6 and J1V6-2 are shown in Figure 3-12 with the differences in restriction sites highlighted.

Under the non-radioactive mini-gel assay conditions, J1V6-2 could not be separated enough to provide a clear picture. Running the gel for the twelve hours at 100V was not sufficient to achieve good separation and running the gel for longer meant losing the intensity of the bands, probably through degradation of the DNA. To overcome this problem, a modified version of the long-short arm assay was used, using radioactively labelled fragments rather than staining with ethidium bromide. These were run on full size polyacrylamide gels at 100V. The radioactivity gave increased sensitivity while the larger gels allowed for greater separation between the fragments. 5' end labelling was used in the same way as for the chemical probing experiments, where only one strand in the junction is labelled.

The results of the radiolabelled long short arm assay for J1V6-2 are shown in Figure 3-13. It can be seen that, despite the larger gel and increased sensitivity, there was little separation between the fragments. The pattern produced in the absence of magnesium and presence of EDTA is also shown in Figure 3-13. It shows that in the absence of magnesium, the shortened junctions can be separated in the expected manner. This would suggest that the junction is annealed correctly and there are no other problems. Compared to the MecA and Coag junctions used so far, the arms of J1V6-2 are long and the restriction sites quite far from the branch point. In the J1V6 junction, restriction sites are positioned 12 base pairs away from the branch point,

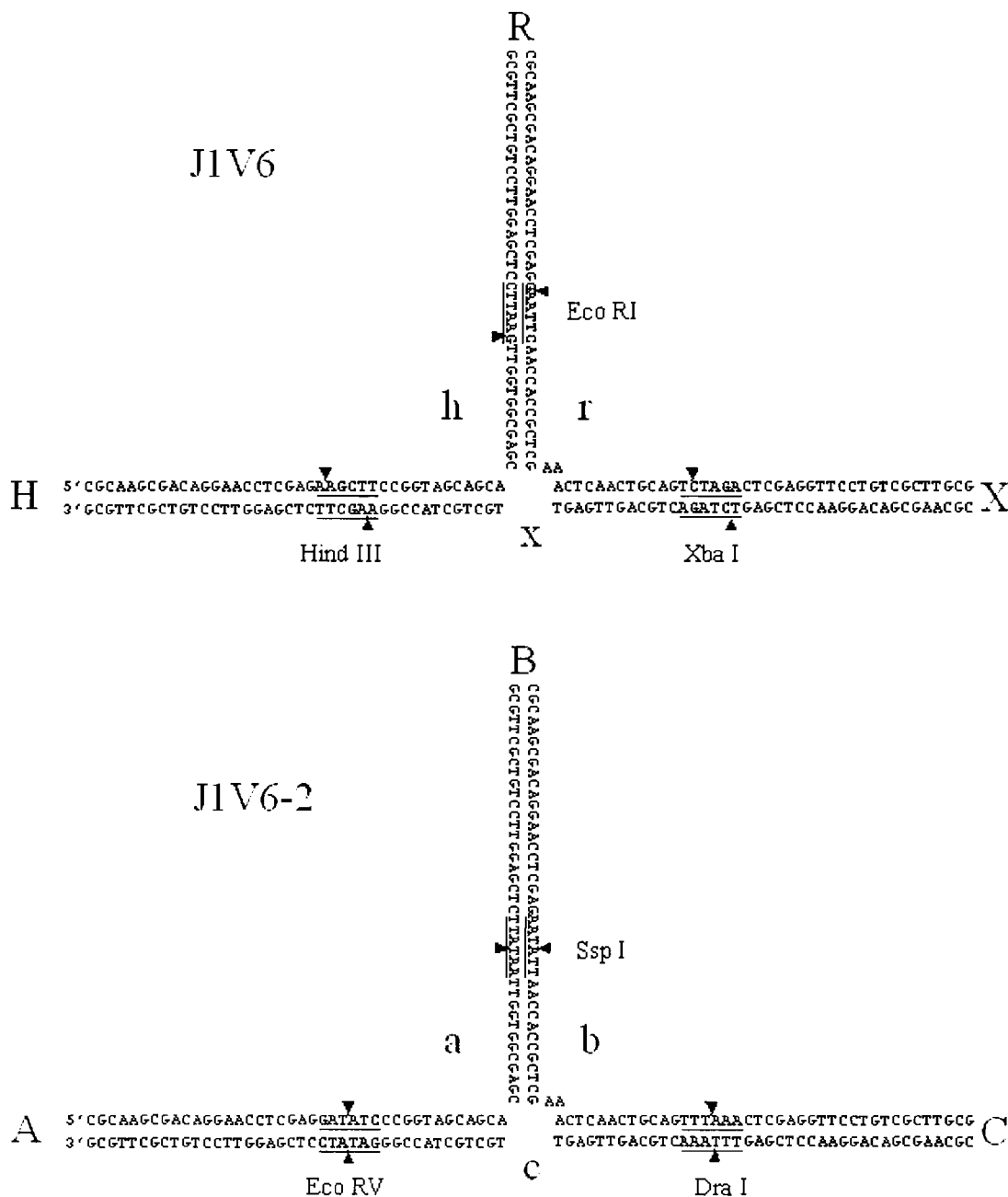


Figure 3-12: The full sequences of junctions J1V6 and J1V6-2. J1V6 was designed and used by Welch *et al.*, (1995). J1V6-2 has been modified from J1V6 for use in this study. Both junctions have arms of 40 base pairs with the restriction sites 12 bases away from the branch point on each side. The restriction sites are shown underlined and with the positions at which the enzymes cut indicated by an arrow head. The three sites from the J1V6 junction yield overhangs which can affect the mobility of the junction through a polyacrylamide gel (Assenberg and Fox, 2001). The restriction sites used in J1V6-2 are identical to those used for MecA and Coag (refer back to Figure 3-1).

compared to the 6 used in MecA and its related junctions. The arms of J1V6 and MecA also differ greatly, the former having arms of 40 bases and the later just 25. This would mean that the junctions would move more slowly through the gel and may be hindered by the length of the shortened arm remaining. In the Welch *et al* paper, the gel used was an 8% polyacrylamide gel with 1mM magnesium run for 20hours. Even after that length of time, the separation between the shortened junctions is not great but does show the mobility pattern of a conformer A/C. In Figure 3-13, the junction with arms AC remaining shows the slowest mobility. It seems unlikely that even with a much longer running time, the J1V6-2 junction would show the expected conformer A/C pattern.

Another version of the J1V6 junction was made with shorter arms and restriction sites at the same position as MecA. The sequence of this new junction, J1V6-3, is shown in Figure 3-14. The results of the long-short arm assay with this junction under non-radioactive mini-gel conditions are shown in Figure 3-15. It gave slightly better results in the long-short arm assay than J1V6-2, in terms of the separation between the shortened junctions, but ultimately did not show the fast-intermediate-slow pattern of a conformer A/C. Instead the pattern is closer to the slow-intermediate-fast of conformer A/B. The angles between the arms must be very similar to give shortened junctions of similar mobility. The differences are slight, suggesting that there is no dominant conformation A/C or A/B. There may be an alternative folding arrangement that it adopts.

Figure 3-16 shows the effects of probing J1V6-3 with DEPC and osmium tetroxide. In the DEPC reaction, four prominent bands can be seen. In strand b, there are three adenines in a row reacting strongly. These bands correspond to the two unpaired adenines at the branch point, replacing the octanediol as the linker, and the base to the 3' side. This is equivalent to the band highlighted in MecA V1 and V2 (Figure 3-10). There is another base in the DEPC reaction which stands out and is not present when MecA V1 and V2 are analysed. It is on the 5' side of the branch point in strand a. These two adenines around the branch point are matched by reactive thymines in the reaction with osmium tetroxide. The result marks two unpaired base pairs which are highly reactive. It also suggests that the junction is quite open at the branch point, more so than any of the other junctions studied so far.

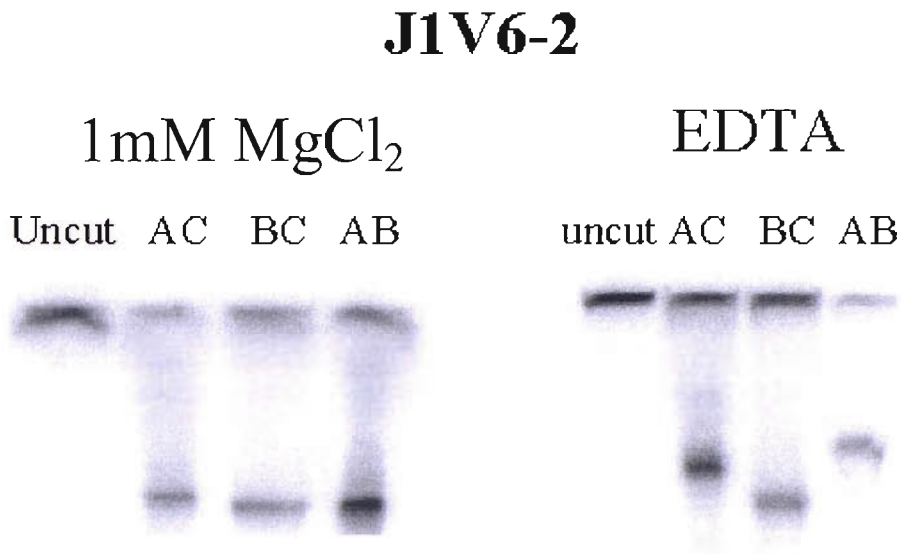


Figure 3-13: Radioactive version of the LSA assay for J1V6-2. One of the strands that would remain intact after the restriction digest was labelled at the 5' end with ³²P-γ-ATP rather than staining with ethidium bromide. The result on the left was obtained with a 12% non-denaturing polyacrylamide gel, with 1mM magnesium chloride in the buffer and gel. The result on the right used a 12% denaturing polyacrylamide gel containing EDTA. They were run for 12 and 8 hours respectively at 100V.

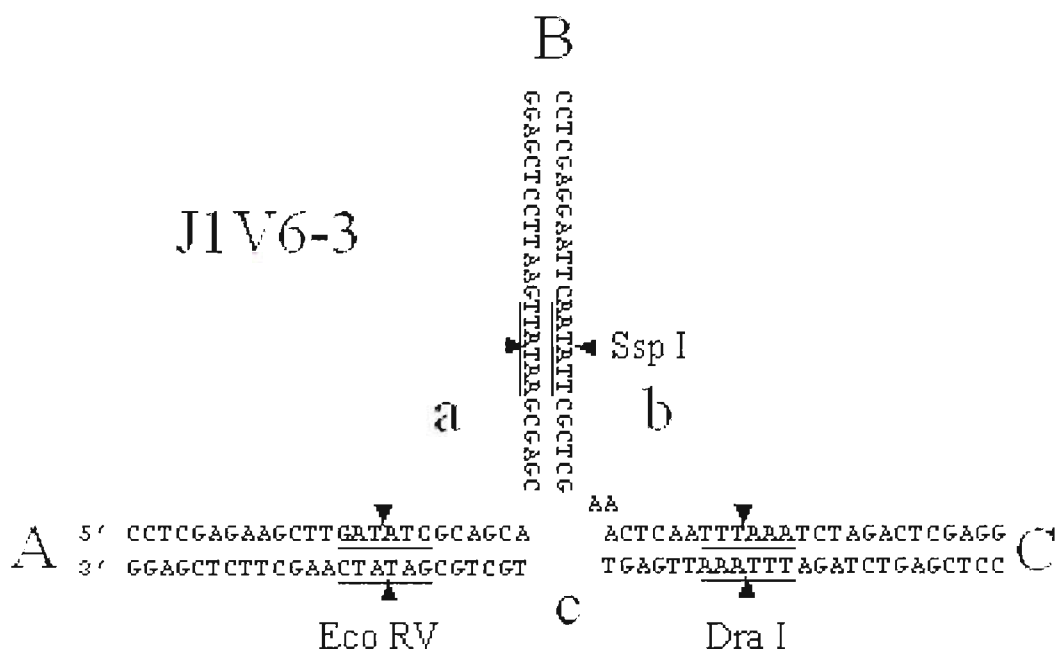


Figure 3-14: The full sequence of J1V6-3. The J1V6 junction was altered to give a junction the same size as MecA with the restriction sites blunt cutting and the same distance away from the branch point as in MecA. Compared to the original junction in Figure 3-12, the arms have been shortened from 40 to 25 base pairs and the restriction sites have been placed 6 base pairs away from the branch point rather than 12. The position of the cut made by the restriction enzyme is indicated by an arrow head.

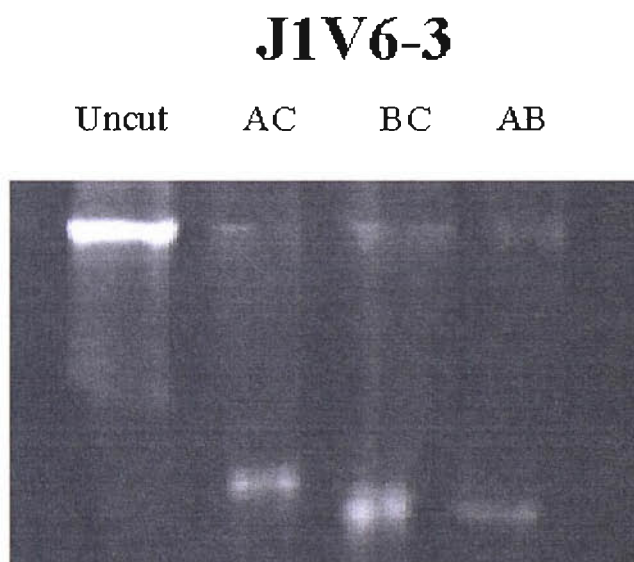


Figure 3-15: Non-radioactive LSA for J1V6 3. This new junction was designed with shorter arms to match MecA and the same restriction sites. While separation was hard to achieve, it does appear to have a slow-fast-fast pattern of mobility. It is clear that the junction does not adopt a dominant conformation AC as predicted. This may be due to the change in restriction sites.

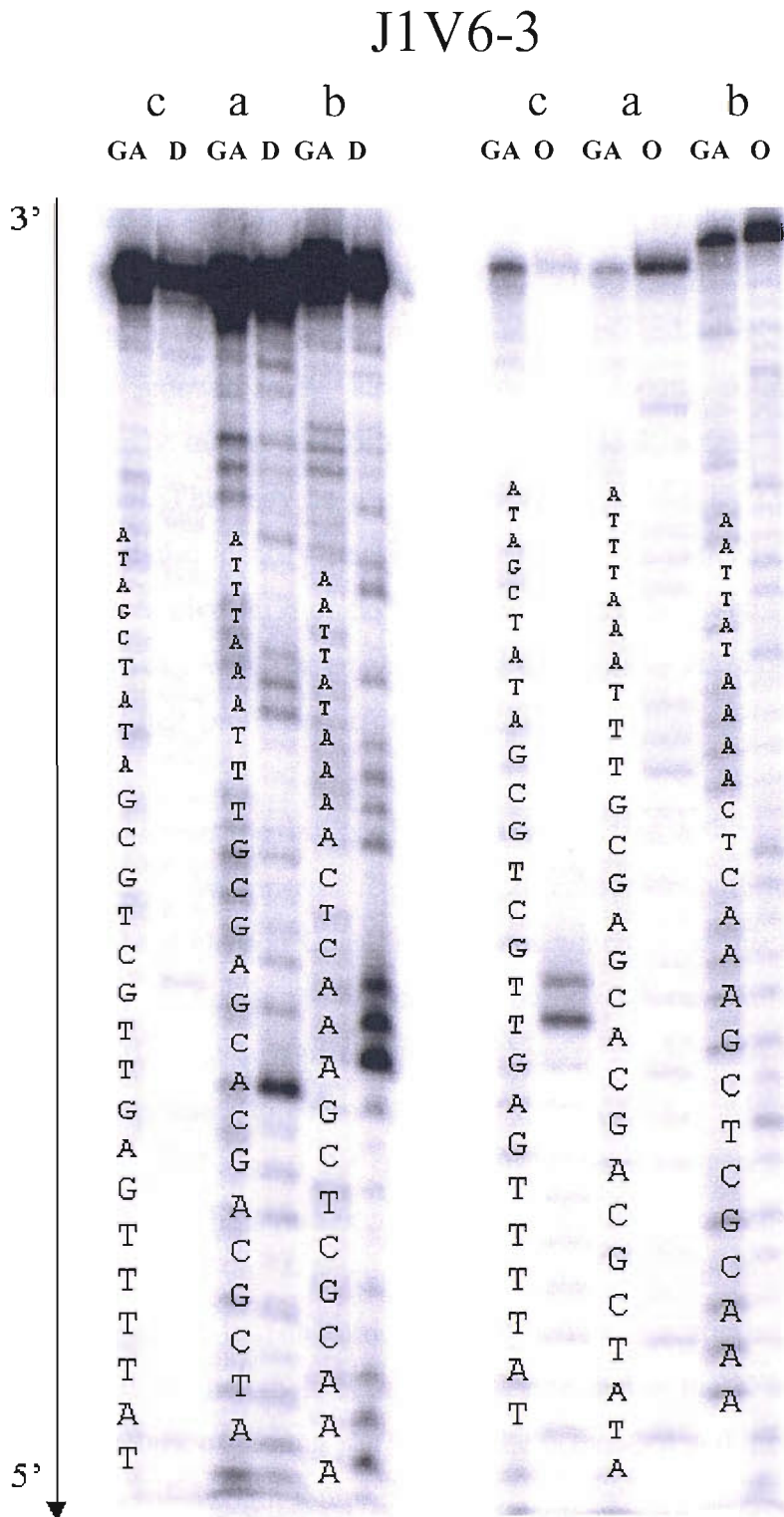


Figure 3-16: Chemical probing of the J1V6-3 junction with DEPC and osmium tetroxide. The sequences of all the strands of DNA are shown alongside the lanes. The gel on the left shows the results of chemical probing of J1V6-3 with DEPC while the gel on the right shows the results of the reaction of the same junction with osmium tetroxide.

3.6 Protection from cleavage around the branch point

DNaseI is a large enzyme which cleaves double stranded DNA non-specifically and can be used to identify protected regions of DNA by the footprint they leave on a cleavage gel (Galas and Schmitz, 1978). This is a technique normally used for identifying the regions where proteins or drugs bind to DNA. The folding of the three-way junction may also provide protection. A junction which folds predominantly in one conformation should block access to DNaseI around the branch point. This may be a useful technique to identify the conformation of the junction and also identify junctions which spend the majority of their time unstacked and therefore give little or no protection. Figure 3-17 shows the principle of this protection assay with the enzyme being too large to cut close to the branch point. There is no such protection in unstacked junctions.

Figure 3-18 shows the results of the DNaseI reaction with MecA and Coag. It can be seen that there are apparent regions of protection for both junctions and that these are particularly clear in Coag. The regions which show this apparent protection do not fall directly across the branch point but in each case fall slightly to the 3' side. To ensure that the lack of cleavage at these points was not sequence dependent, a duplex complement to each of the MecA strands was made and tested in the same way. Figure 3-19 shows the comparison between each strand in a junction and a duplex context. These results confirm that the protection seen is real and that it is the junction structure which is responsible for this. The one exception is the region on the 5' side of the branch point in strand b. The area appears to show quite a wide footprint across the whole branch point, extended to the 3' side. The duplex strand mirrors the lack of cutting just around the octanediol and the guanine to the 3' side but there is still a slight footprint beyond.

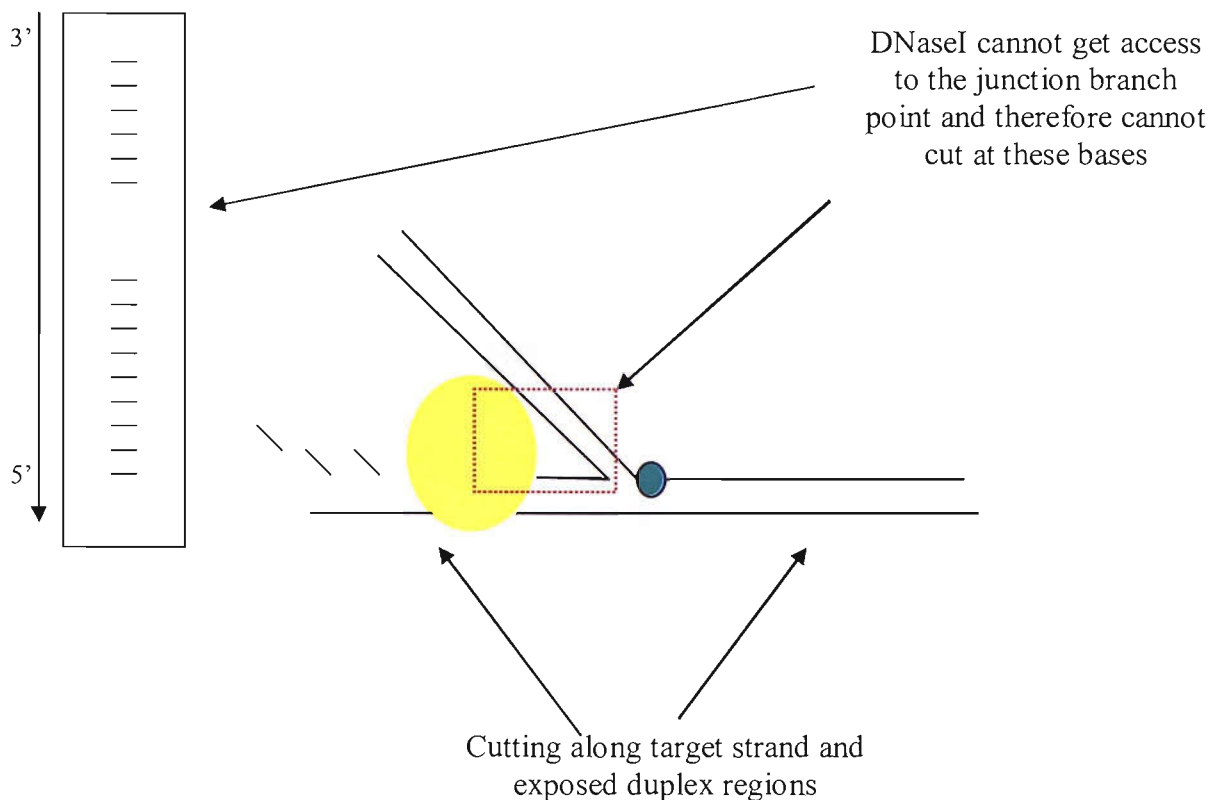


Figure 3-16: The cutting of double stranded DNA by DNaseI. The DNaseI enzyme is shown as a yellow circle digesting the double stranded DNA along one arm of the junction. The enzyme does not cut in a sequence specific manner and so will cut at any base, unless prevented from doing so. The reaction length is timed so that each junction in the population is cut once only, but at different locations. The folding of the 3WJ causes regions around the branch point to be protected from the enzyme. This is represented by a small footprint on a denaturing polyacrylamide gel similar to that shown on the left of the picture.

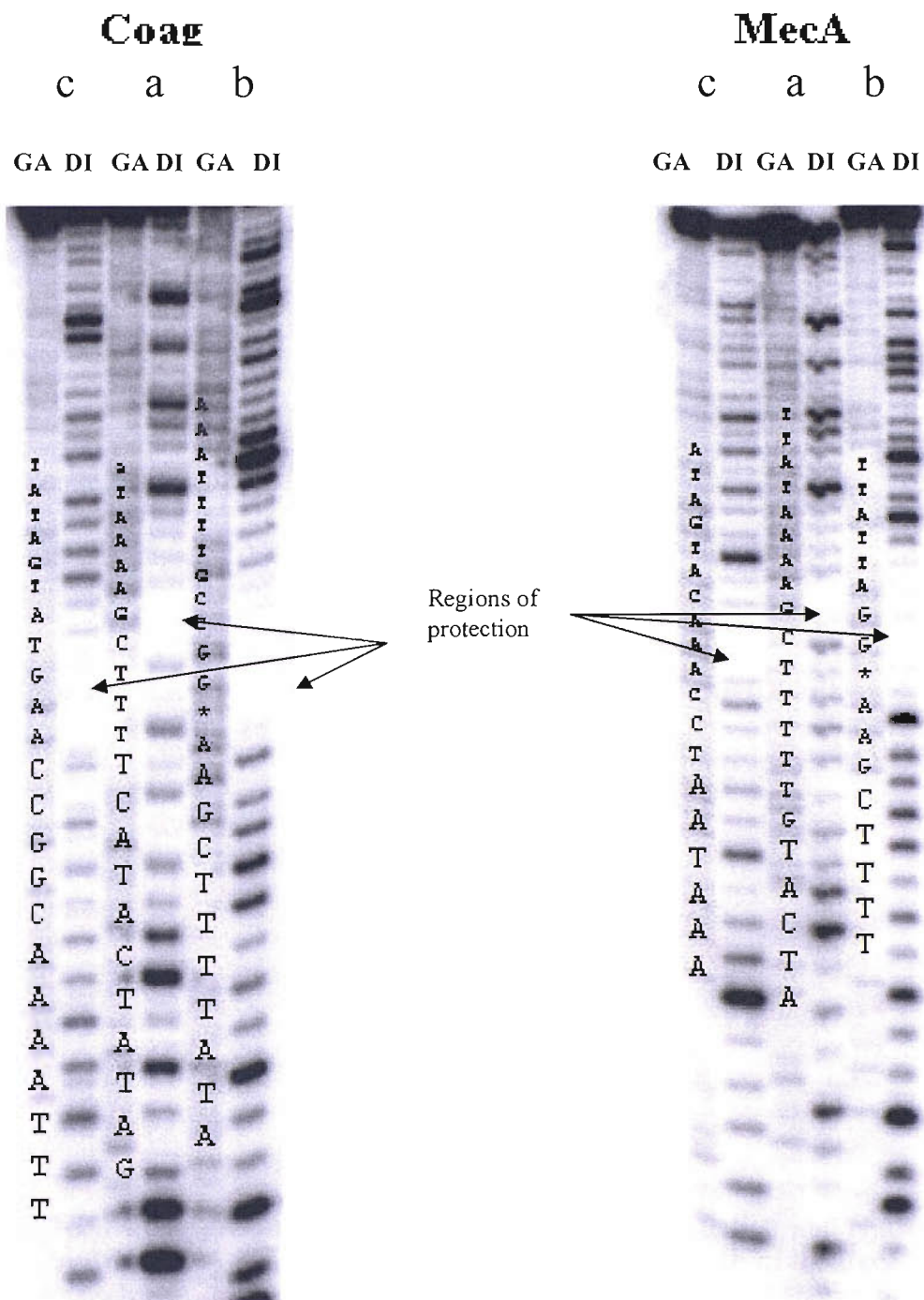


Figure 3-18: DNaseI digestion of junctions MecA and Coag. Since DNaseI cuts along double stranded DNA in a non-sequence dependant manner, bands should be seen for every band that the enzyme has access to. The results above are from a 15% denaturing polyacrylamide gel, identical to those used for the chemical probing experiments. As before, a GA marker lane has been run alongside each result and the sequence of the appropriate strand overlaid. DI has been used to indicate the results of DNaseI reactions. Areas of the gel which are totally or relatively free from cutting have been indicated by arrows.

MecA junction and duplex

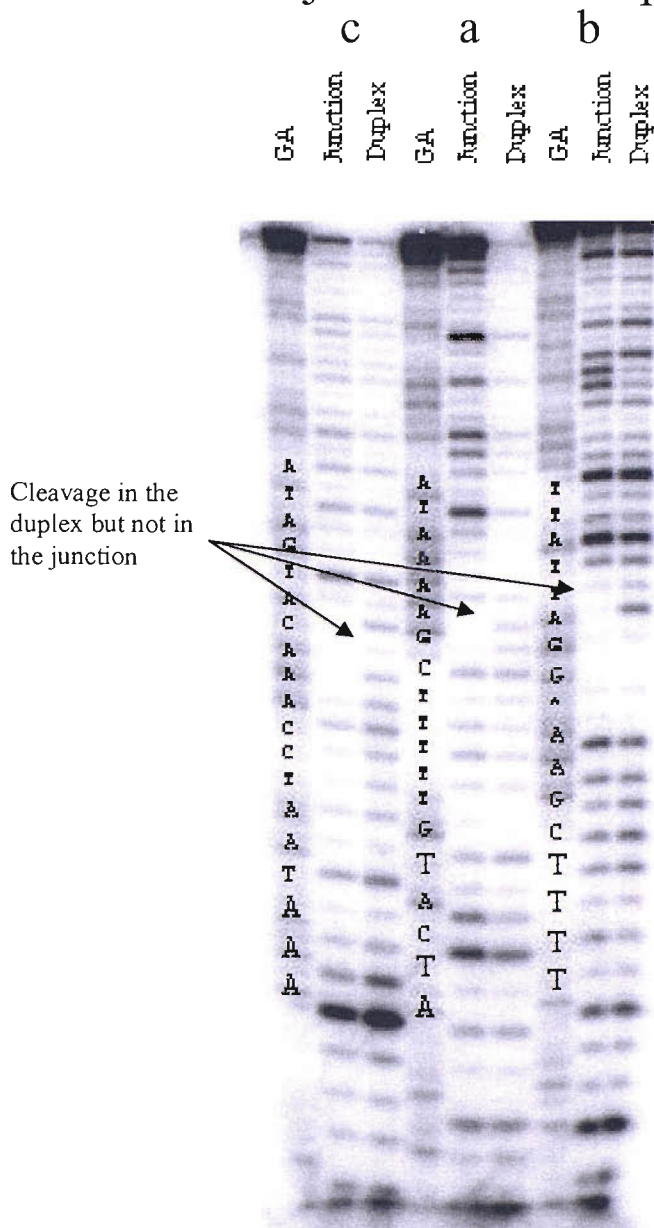


Figure 3-19: Comparison of the cleavage of the MecA junction and equivalent duplex by DNaseI. A complementary strand was made for each of the three oligos that make up MecA, to form a set of three duplexes. The above gel shows the effect of DNaseI on each of the strands when in a junction or duplex. Where there is cleavage in the duplex but not in the junction, it has been highlighted by an arrow.

Junctions MecA V1 and V2, whose sequence around the branch point was changed in section 3.4, were exposed to DNaseI. The results can be seen in Figure 3-20. Looking around the branch point in strands a and c for MecAV1, there does appear to be a region free from bands. The bands on strand b appear continuous with no breaks. MecAV2 shows very little protection from DNaseI. There is a region on strand c where the cutting appears less.

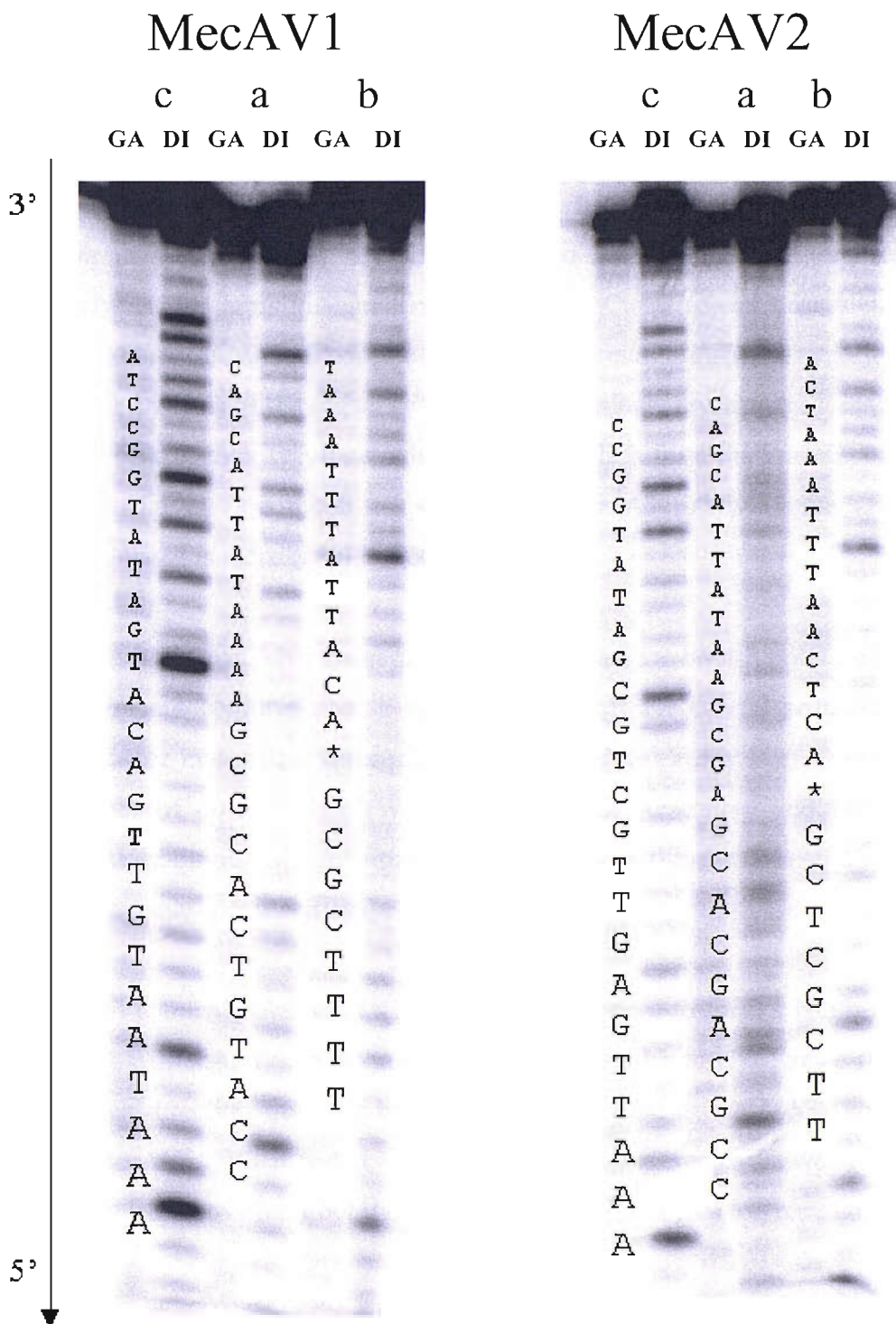


Figure 3-20: Exposure of junctions MecA V1 and MecAV2 to DNaseI. The data shown above is for the junctions MecA V1 and V2 only. The sequence of the appropriate strands has been overlaid on the GA lanes.

3.7 Discussion

MecA and Coag are junctions that are known to behave differently in the SMART assay. Modified versions of the three-way junctions with full length arms and restriction sites incorporated, have been made to investigate the differences. The angles between the arms in junctions MecA and Coag, appear near identical as judged by gel mobility. Both adopt a standard conformation A/B the majority of the time. Closer analysis of the sequence of the two junctions, reveals that it is the same around the central core, two base pairs in each direction from the branch point. The sequence further away is much less similar. MecA is AT rich while the balance of AT and GC bases is more equal in Coag.

Since MecA and Coag have the same sequence around the branch point and share a similar dominant conformation, it would appear that the central sequence is the dominant factor in deciding the conformation. This has already been demonstrated by a number of other studies (Welch *et al.*, 1995, Assenberg and Fox, 2001). Experiments modifying MecA further, however, suggest that the core sequence is by no means the only consideration.

Junction J1V6 from Welch *et al.*,(1995), has been shown to form a conformer AC, opposite from MecA. Taken as an opposing junction, the core sequence (two base pairs in each direction from the branch point), was used in the MecA junction to produce MecAV1. When tested, it was shown that the angles between the arms changed compared to MecA but the overall dominant conformation remained AB.

A further experiment using the core six base pairs on each arm (MecAV2), gave an intermediate conformation typical of a junction with a preference for neither conformer. Both junctions MecA V1 and V2 showed hyper-reactive bands around the branch point indicating a more open structure.

After full-length versions of the original J1V6 junction failed to separate at all on a gel, a shorter version was made. Once the restriction sites had been brought inwards, the new J1V6-3 junction became almost identical to MecAV2 for the core twelve base pairs on each arm. The difference between them was the choice of linker, an

octanediol in the MecA junction and two unpaired adenines in the J1V6 junction. The sequence further than the restriction sites also differed between the two junctions. The result for the new J1V6-3 junction suggested more or less equal angles between all arms.

Figure 3-21 shows the models of the dominant conformation pieced together from gel mobility and chemical probing experiments. MecA shows the clear AB pattern reflected in the very different angles between the arms. In MecAV1, the angles between BC and AB are more equal but the pattern still suggests that arms A and B are co-linear. MecAV2 shows a classic unstacked conformation demonstrating a preference for neither conformer. Both of these suggest that sequence further from the branch point than the central core can have a substantial effect on overall conformation. This may be a knock on effect which influences the interactions of the base pairs at the branch point.

The pattern in J1V6-3 is the most difficult to explain. It does not adopt the classic unstacked/ intermediate conformation pattern yet in the absence of magnesium it does. The pattern could be interpreted as slow-intermediate-fast pattern of conformer AB but the difference in the angles between the arms seem so slight it is unlikely that any pair of arms are stacked. The structure must therefore be open. Normally the arms containing the linker are forced outwards compared to the others. In this case it may be that the unpaired adenines slip and cause the next adenine to be slipped out instead. This may explain the difference in pattern. It is presumed that J1V6-3 is an intermediate and that there is no preference for either stacked conformer.

The final aspect of this work to consider is that the shorter version of J1V6 did not show the same pattern as that given in the literature, although the techniques and buffer conditions were made as close to those used in the paper as possible. It may be that the difference in restriction sites had a sequence dependent effect in itself or that the change to blunt cutting enzymes from those that leave 3' overhangs altered the pattern of mobility.

The DNaseI studies of protection in the 3WJ did show some protection in the MecA and Coag junctions. This was proved to be due to the junction structure when the results were compared to those obtained for a duplex of similar sequence. One feature of the protection noted, was that it always occurred on the 3' side of the branch point. This was also noted by Murchie *et al* 1990, who did a similar study of 4WJs. Their explanation was that this was the region of backbone located in the major groove of the opposing helix (Figure 3-22).

DNaseI may yield some additional information on a three-way junction which can be used in connection with other data to establish a portfolio of the junction folding. Based on the data presented here, it is unlikely to be able to give useful information as to the dominant conformation alone. One of the problems with this techniques is that the junctions are not likely to be static but in a dynamic equilibrium. When a junction is in the unfolded state, it is vulnerable to attack by DNaseI at any position. If it does fold into the opposite folded conformation for any length of time, it will be vulnerable to attack at different points and protected at others. It is perhaps most useful for looking at junctions where the equilibrium is highly biased towards one conformer or has very little bias and spends most of its time unstacked. These would provide the clearest patterns.

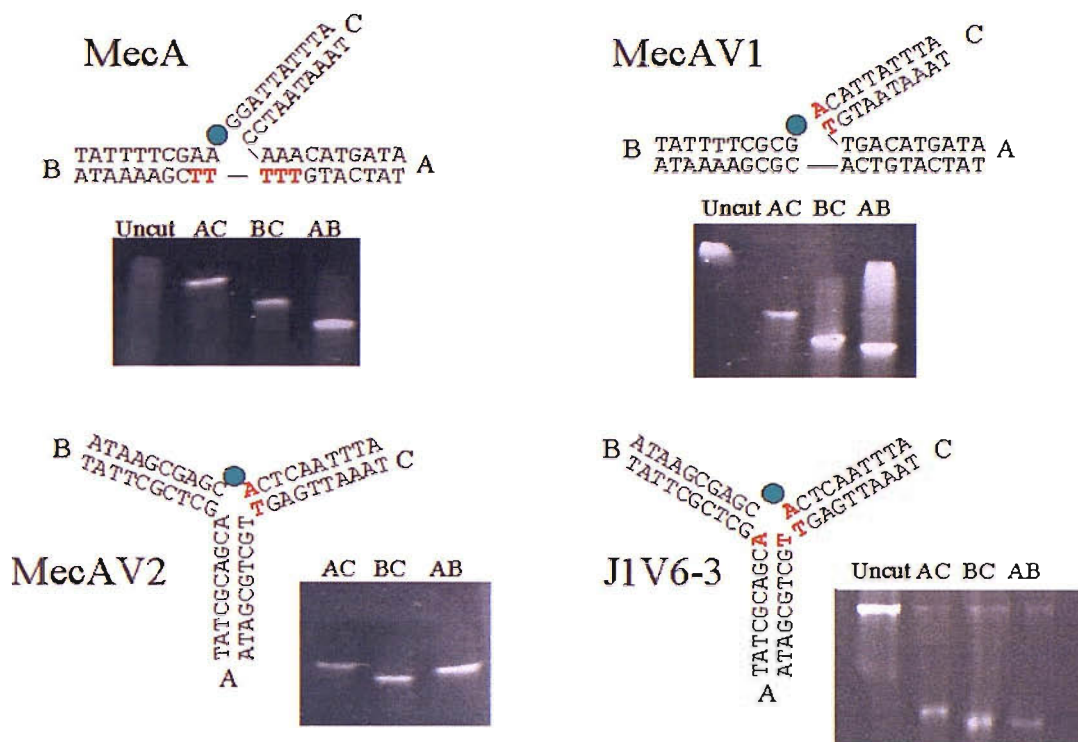


Figure 3-21: Models of the three-way junctions suggested by conformation and chemical probing studies. The junctions are shown in the dominant state suggested by the gel mobility assays. Since these molecules are planar, a fair guess can be made about the actual angles between the arms based on their relative mobility. The bases highlighted in red are those picked out well by chemical probing. It should be noted that the state of the GC bases cannot be accurately determined with the chemicals used.

The green circle is used to show the position of the linker, as in previous diagrams. In the case of the MecA junctions and variants, the linker is an octanediol. In keeping with the J1V6 junction from the literature, J1V6-3 uses two unpaired adenines as its linker. This may explain the slight differences in conformation between junctions MecAV2 and J1V6-3 despite the fact that the sequence around the branch point is identical up to 12 bases away along each arm.



Figure 3-22: Model of the protection from DNase I in a folded 4WJ. The diagram is taken directly from Murchie *et al.*, 1990. It shows the regions of protection in yellow. The crossing of the strands and in particular the way sequence 3' of the branch point sits in the major groove of the opposing helix, afford protection from DNase I but not the smaller dimethyl sulphate.

Chapter 4

ACCESSIBILITY OF THE THREE-WAY JUNCTION TO ENZYMES

Part 1: The extension step

In this section, the SMART assay is broken down into stages to follow the reaction more closely. The extension of the DNA primer by *Bst* DNA polymerase is investigated under simplified reaction conditions. Improvements to the MecA SMART junction are tested for the efficiency of primer extension, and for conformation using the long-short arm assay and chemical probing.

Part 2: The full SMART assay

This section details later work which involved investigating extension under full SMART assay conditions. The involvement of other molecules in the reaction are analysed, including the interaction of RNA polymerase with the 3WJ. *Salmonella* rRNA is used as the target.

4 Accessibility of the three-way junction to enzymes

4.1 Review of the SMART reaction

During the SMART reaction, the probes anneal to their target and form a stable 3WJ structure. At this stage the shorter probe overlaps with the longer probe by only eight bases. DNA polymerase present in the reaction uses the short probe as a primer and adds bases on to it complementary to those on the longer strand. A long duplex arm of about one hundred bases is produced (Figure 4-1). This is an important step in the reaction because it generates a double stranded T7 RNA polymerase promoter. Once this has occurred, the RNA polymerase can bind and begin transcription of the downstream region. The extension of DNA by DNA polymerase is slow due to the need for accuracy, whereas RNA polymerase works by making lots of copies, not all of which are necessarily accurate. DNA extension is the slowest step in the reaction and as such is the rate determining step.

To investigate the efficiency and rate of this step, a simple extension assay was designed using a 5' radiolabelled extension probe. A time course reaction was used to monitor the extension of the short probe. The samples taken were run on a denaturing gel to follow the progress of the probe population from all unextended to all full length or as far as the reaction would go (Figure 4-2). This gave a measure of efficiency of the DNA polymerase at extending the 3WJ.

4.2 Extension of MecA and Coag junctions

In order to increase the efficiency of the MecA junction at giving a signal in the SMART reaction, a series of junctions were created with extra octanediols at the branch

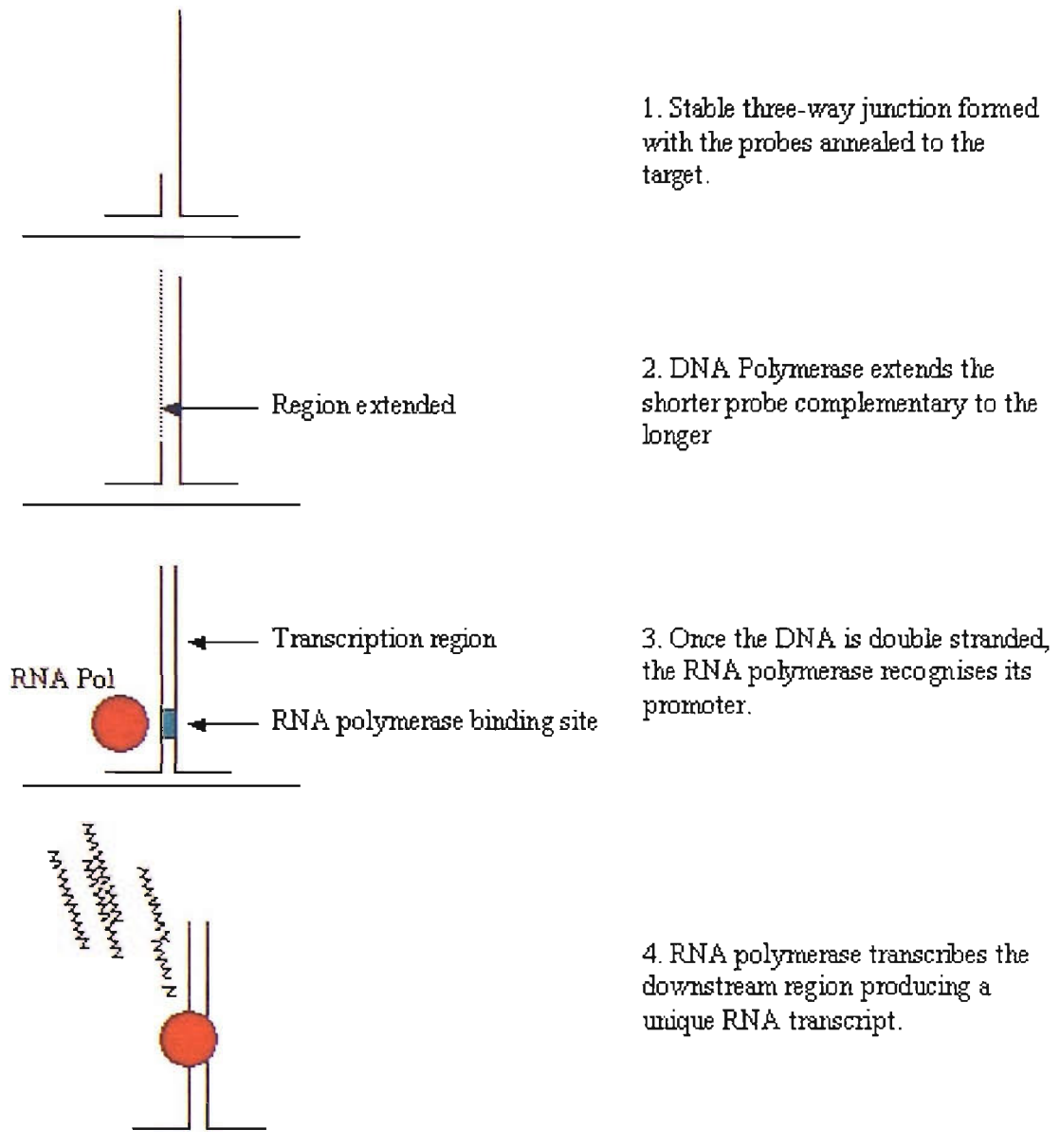


Figure 4-1: Review of the extension and transcription steps of SMART. The four sequential steps outlined above show the reaction from the formation of the junction to production of the RNA transcript as a signal.

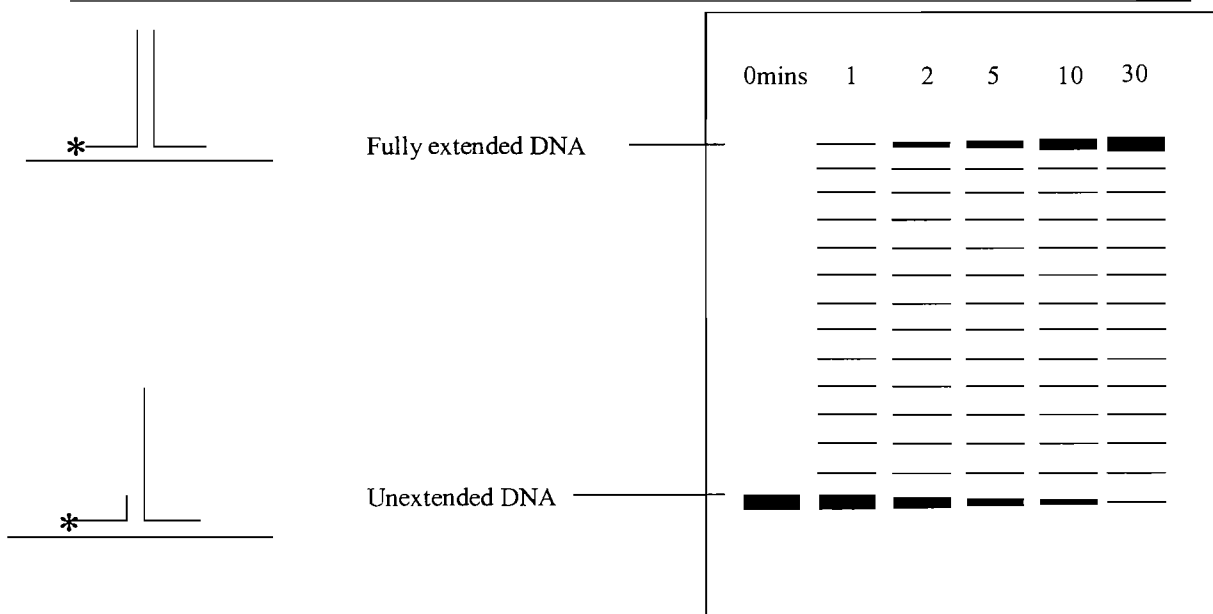


Figure 4-2: Schematic representation of the DNA extension assay. A 5' radiolabelled primer was used to chart the progress of extension. Samples were removed from the reaction at certain time points, 0, 1, 2, 5, 10 and 30mins. The samples were stopped by placing on ice so that no further extension could take place, and then run on a 15% denaturing polyacrylamide gel. The 0 min sample was removed before the addition of the enzyme and so contains only unextended DNA as a control. The remaining samples showed the gradual extension of the population of DNA molecules by the lessening in intensity of the band containing the unextended DNA and an increase in intensity of the fully extended band. In between the two was a ladder of semi-extended DNA. At these points, the enzyme reaction was either in progress or the enzyme had dissociated from the DNA leaving the reaction uncompleted.

point. The insertion of the extra linkers was designed to lift the T7 RNA binding region away from the branch point to improve enzyme accessibility. These new junctions contained extra octanediols in both template and extension strands at the branch point. The standard junctions contain a single octanediol in the template strand (Figures 4-3 and 4-4). The four new junctions used one octanediol in each strand (1:1) then two (2:2), three (3:3) and four (4:4) (Figure 4-5). These new junctions had performed well in the SMART reaction, 3:3 giving the best signal and 4:4 the worst. The original MecA and Coag junctions were tested alongside the modified versions in the extension reaction. (Figure 4-6).

It can be seen from the gels in Figure 4-6 that the labelled DNA was extended to a product of expected length. Although samples were taken at regular time points, the reaction seems to have reached completion within the first few minutes and no further extension is observed at longer times. In each case there was a large amount of DNA remaining un-extended despite the excess of the other strands present allowing ample opportunity for junction formation.

Table 4-1 contains a more detailed breakdown of the results. An analysis of the intensity of regions of the gel leads to an estimation of the percentage of DNA present in different states within any one lane. The states identified were totally un-extended, partially extended (any DNA present in the DNA ladder) and full length (Figure 4-7). The first table shows only the percentages of un-extended and full length to display the pattern more clearly.

The extra octanediols did not appear to make a great difference to the amount of full length product formation. However, while the amount of full length oligo only increases a few percent by adding octanediols, the percentage remaining un-extended after 30mins goes down from 41% in MecA to 13% in the 3:3 junction. The remainder is found in the ladder of partially extended bands. A large amount of the DNA in the ladder will still be usable in SMART because it will have the completed T7 RNA polymerase promoter, albeit the transcript template region may remain single stranded

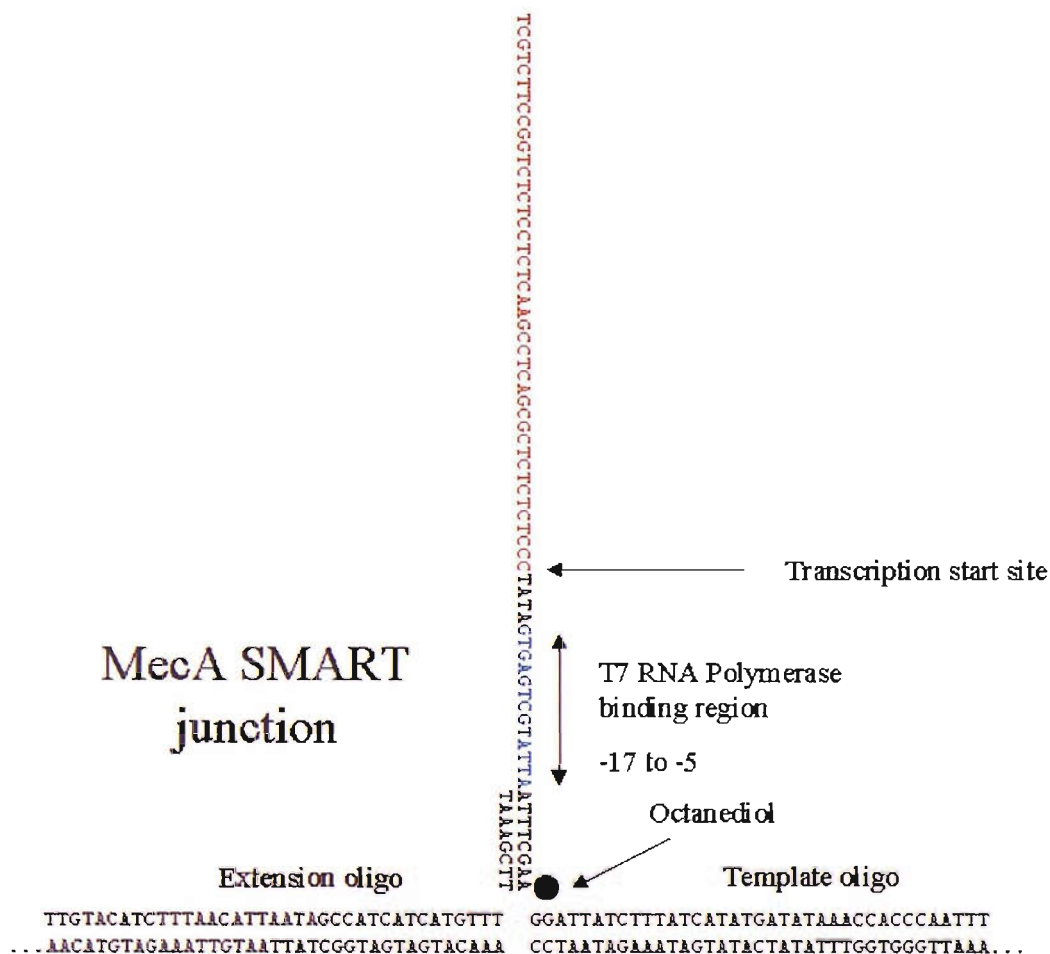


Figure 4-3: Sequence of the MecA SMART 3WJ. The diagram shows the two probes, the extension and template, annealed to the target sequence. The two probes are positioned on the target so that they anneal with an 8 base pair overlap only. The template oligo contains a number of important sequences. Shown in blue is the T7 RNA polymerase binding region. The RNA polymerase only recognises this site when it becomes double stranded after extension of the short probe. The transcription start site follows, and RNA complementary to the sequence shown in red will be produced. It does not matter if the transcription region is double stranded (Kukarin *et al.*, 2003).

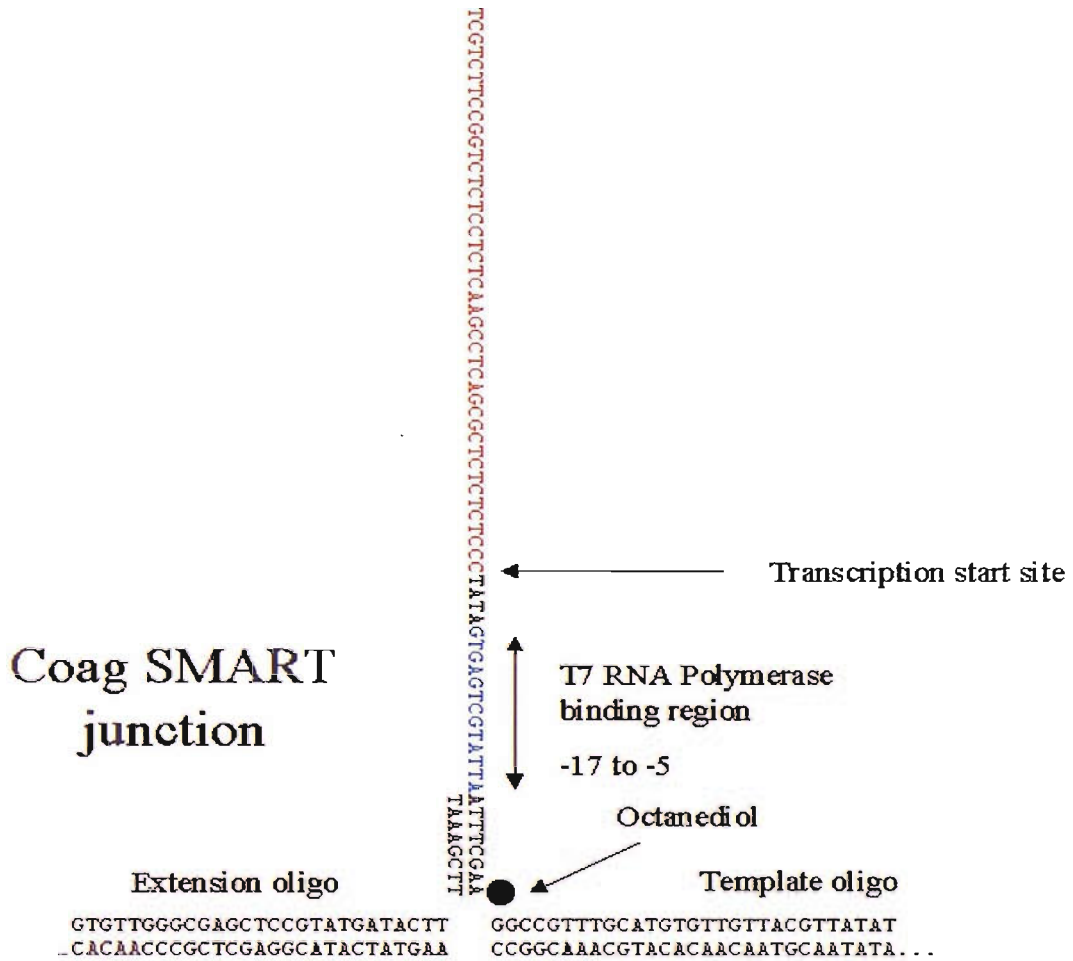


Figure 4-3: Sequence of the Coag SMART junction. The extended arms of the MecA and Coag junctions are identical. The sequence along this arm is optimised to produce the best yield of RNA. The sequence recognised as the target is part of the coagulase gene in the genome of MRSA. Sequences from both the mecA and coagulase genes are used to ensure specificity.

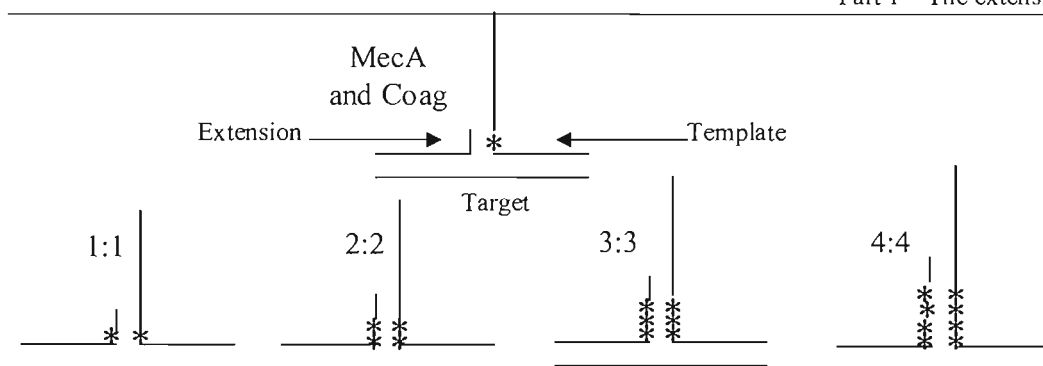


Figure 4-5: Position of the octanediols in the modified MecA junctions. These junctions contain the same sequence as MecA but include extra octanediols (*) at the branch point on the template and extension probe. The standard MecA and Coag junctions have a single octanediol in the template strand.

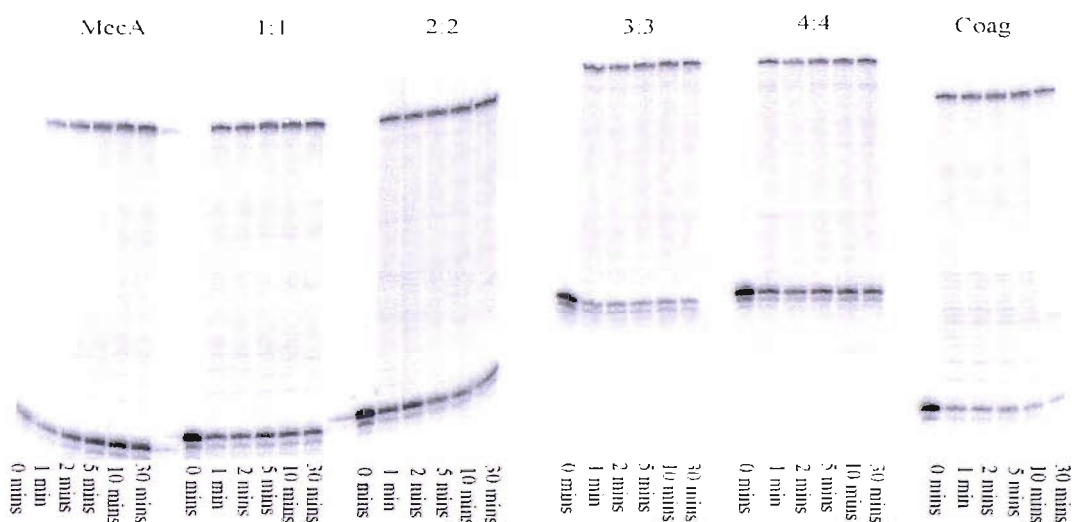


Figure 4-6: Extension of MecA, Coag and the modified MecA junctions. The samples above were run on a 15% denaturing polyacrylamide gel containing TBE. In each case a time course reaction was performed, with one sample (0mins) being removed before the addition of the enzyme. Samples were taken at each of the indicated time points and the reaction stopped. Unextended DNA can be seen as a band at the bottom of the gel (the 0mins sample is 100% unextended), fully extended DNA at the top and the DNA ladder in between.

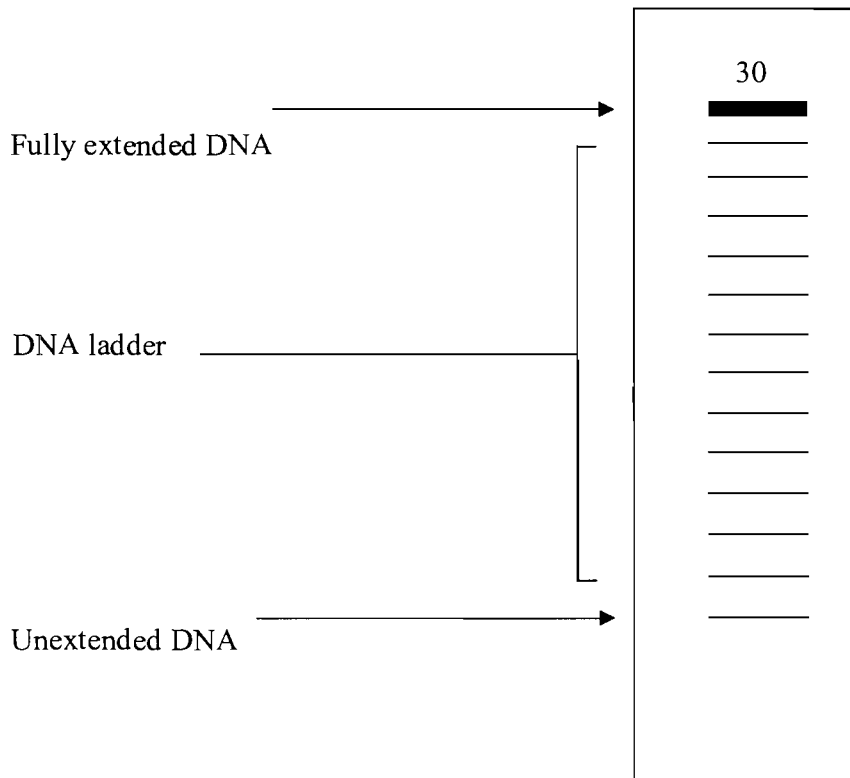


Figure 4-7: Diagram of the regions of DNA into which each lane is divided. The intensity of each region is measured using ImageQuant and expressed as a percentage of the total DNA in the lane.

	MecA	1:1	2:2	3:3	4:4	Coag
Percentage unextended	41	31	27	13	31	9
Percentage full length	27	31	29	29	21	38

Table 4-1: The percentage of DNA extended after 30mins using *Bst* DNA polymerase at 41°C. The percentage of DNA extended fully and unextended is shown for each junction. The remaining DNA unaccounted for is in the DNA ladder. The percentages above were averages obtained over 3 gels.

(Kukarin *et al.*, 2003). The 4:4 junction is not extended well. The standard Coag junction is substantially more efficient than MecA with 38% full extension after 30mins and only 9% remaining unextended.

In summary, the addition of up to three octanediols on each side of the branch point increases the frequency of initiation of extension by *Bst* DNA polymerase. This is probably due to the lifting of the branch point facilitating interaction with the enzyme. This increased initiation does not seem to result in increased full length product formation. This suggests that the enzyme may dissociate further along the extension although there is no single point at which this occurs.

4.21 Effect of inserting extra linkers on the conformation of the 3WJ

To investigate whether the added linkers affect the conformation of the three-way junction, two of the modified junctions (1:1 and 3:3) were synthesised as full length products, with restriction sites incorporated into them for use in the long-short arm assay. The sites were identical to those used in the full length MecA junction (Figure 4-8). Junction 1:1 contained one octanediol in the extension and one in the template strand (arms AB and BC, respectively) adjacent to the branch point. Junction 3:3 contained three octanediols in the extension and template strands. These junctions were made to test the conformational preference of these new junctions with a view to better explaining their success in the extension assay.

The results of the long-short arm assay for junctions MecA, 1:1 and 3:3 are shown in Figure 4-9. It can be seen that introducing extra linkers adjacent to the branch point changes the pattern of bands. This suggests that the extra linkers increase the flexibility of the junction and cause a change in the angles between the arms. By adding one extra octanediol into MecA in the 1:1 junction, the pattern is shifted so that the angles between the three arms are closer together. While the pattern is still that of an AB junction, fragment AB runs slower than its equivalent in MecA. At the same time, fragment BC is running faster, showing an increased angle between arms B and C. These results combined suggest that the octanediol in AB introduces enough flexibility

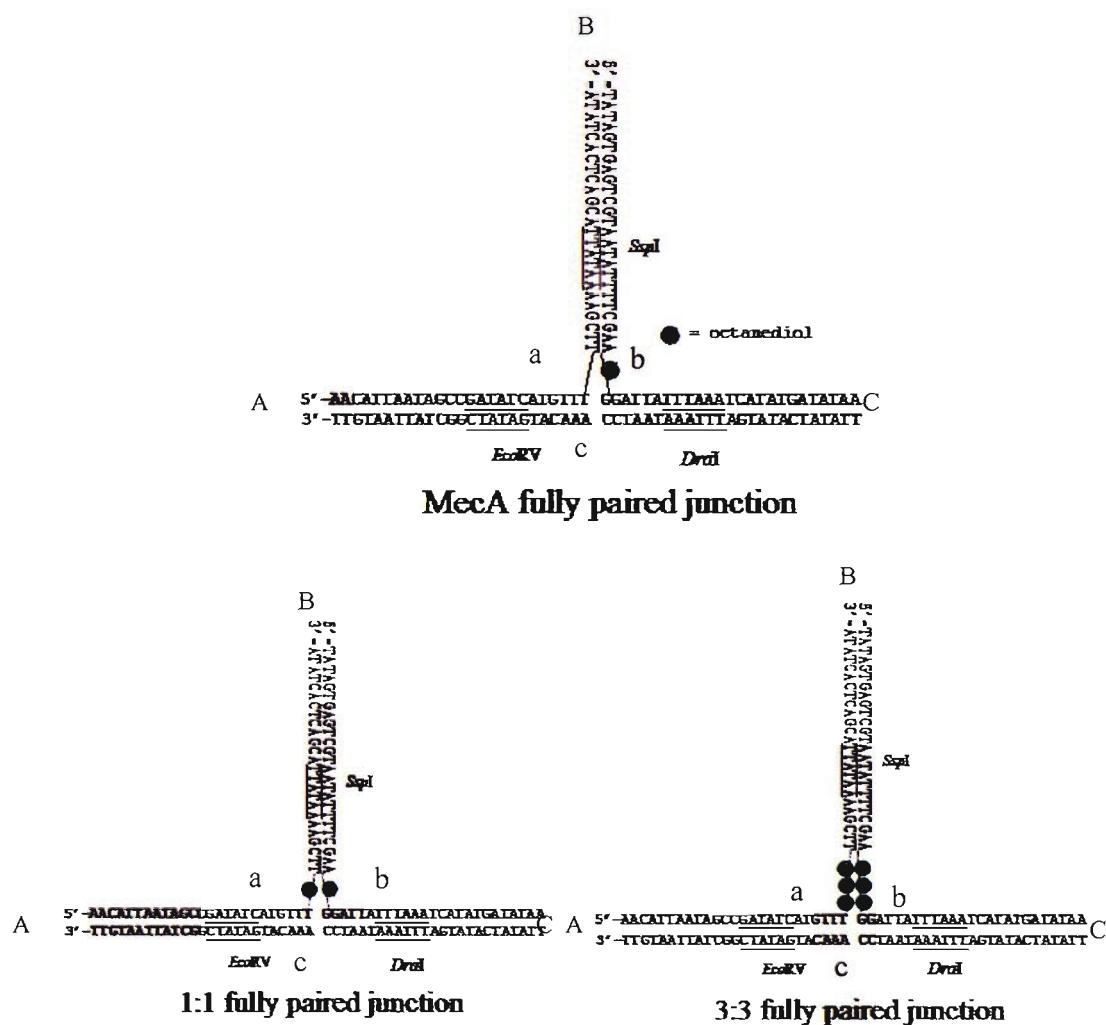


Figure 4-8: Fully complementary versions of junctions MecA, 1:1 and 3:3. The sequence of the SMART junctions were modified to incorporate unique restriction sites in each arm of the junction. All arms are 25 base pairs long. The 1:1 and 3:3 junctions are identical to the modified version of MecA but contain extra octanediols (blue circle) at the branch point. The arms of the junction are labelled A, B and C while the strands are a, b and c.

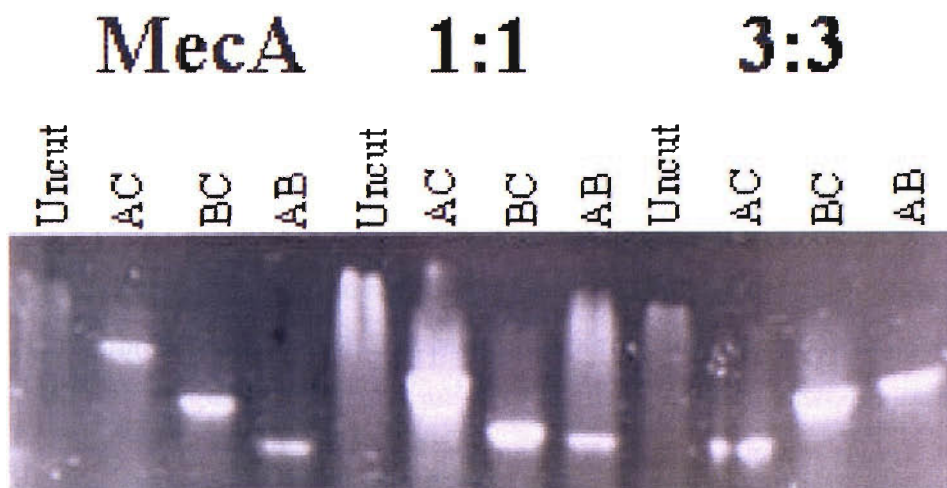


Figure 4-9: Long-short arm assay using MecA and junctions 1:1 and 3:3 containing extra octanediols at the branch point. The 1:1 junction contains a single octanediol in the equivalent of the extension and template strands, while 3:3 contains 3 extra octanediols in between arms AB and AC. The pattern made by MecA is shown for comparison. The gel is a 20% non-denaturing polyacrylamide gel. It was run for 12 hours at 100V in the presence of 1mM magnesium chloride, then stained with ethidium bromide. The fastest running sample contains the junction with the greatest angle between its two remaining arms.

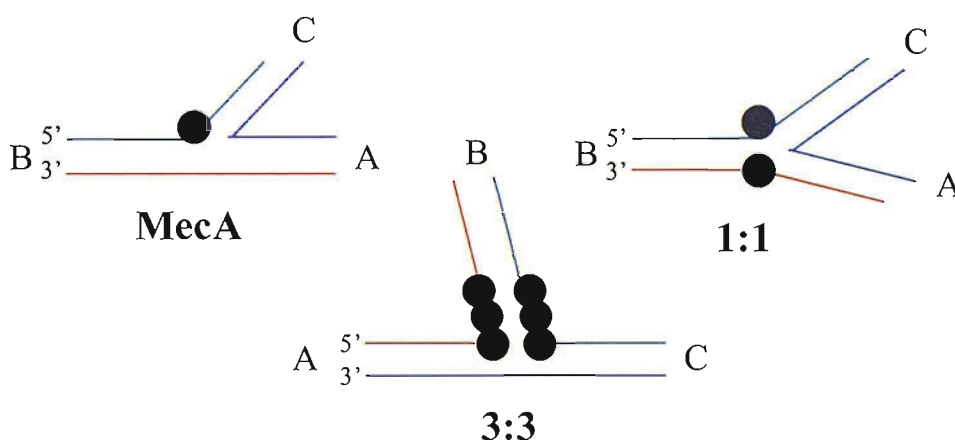


Figure 4-10: Model of the averaged conformations of junctions MecA, 1:1 and 3:3. The extra flexibility provided by the octanediols appears to cause a change in the angle of the arms in the junctions. In the 1:1 junction, arm AB may kink from the insertion of the octanediol in its length. The angle AB then decreases, dropping arm B and lengthening angle AC as a consequence. Introducing 3 octanediols into each arm causes arm AC to become co-axial regardless of the sequence around the branch point. The angles between arms AB and AC are similar but AB does appear to be smaller.

to generate a kink in the co-axial arms. This lengthens angle AC and makes angles AB and BC much more similar .

By adding three octanediols on two sides of the branch point, the junction adopts a pattern resembling that of a conformer AC. This represents a complete shift in conformation with arm AC becoming co-axial. Fragment AC in the 3:3 junction is runs at a similar mobility to fragment AB in MecA, supporting the suggestion that the former is now co-axial. Angles AB and AC are very similar giving an extended arm that is almost at 90° to the co-axial arm. Fragment BC has a slightly higher mobility suggesting that the arm is leaning slightly with angle AB being smaller than BC (Figure 4-10). Both octanediol junctions show quite a smeared pattern, possible due to their flexibility. The long-short arm assay gives an averaged conformation over a period of a few hours. If the conformers are flipping in the equilibrium or if the extended arm is very flexible, a slightly blurred pattern would be expected.

4.22 Chemical probing of junctions 1:1 and 3:3

The conformation around these junctions was investigated further by chemical probing with DEPC and osmium tetroxide. The results for these reactions are shown in Figure 4-11 and 4-12 respectively. Chemical probing reveals a few exposed bases around the branch point, the adenines and opposing thymines immediately adjacent to the branch point. The reactions with the exposed bases are not particularly strong in comparison to those in MecA V1 and V2 in chapter 3, and the presence of the octanediols does not appear to encourage looped out regions around the branch point. If anything they probably aid folding by increasing the backbone length and releasing tension around the branch point. There is no evidence that the junctions spend more time in the unstacked form than in the original MecA junction. Both junctions appear stable.

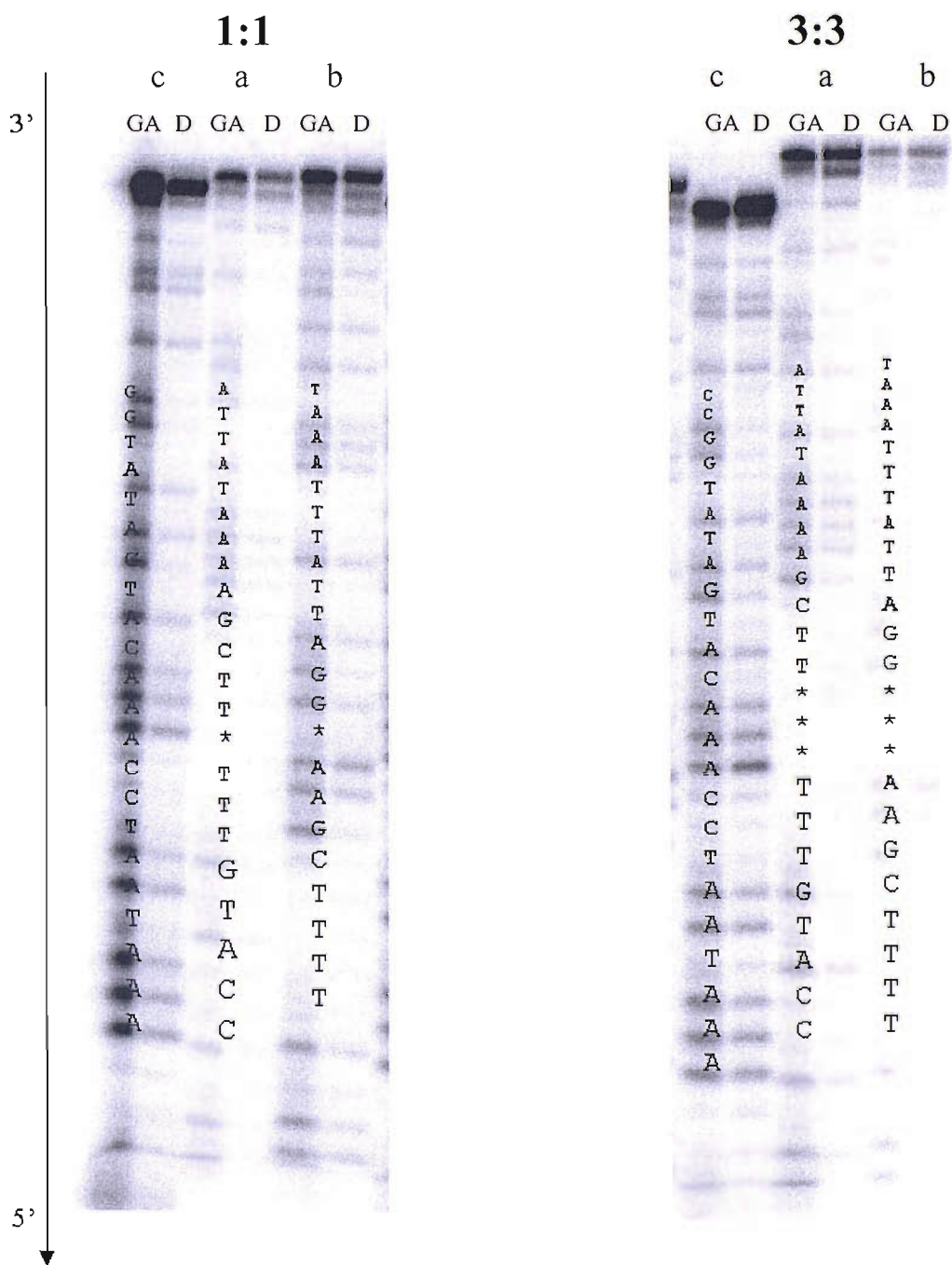


Figure 4-11: Chemical probing of junctions 1:1 and 3:3 with DEPC. The gel shows the exposure of the fully complementary junctions 1:1 and 3:3 to diethyl pyrocarbonate (DEPC) as described in section 2.51. The products of cleavage were run on a 15% denaturing gel. The identity of the labelled strand in each case is shown as c, a or b. The labels GA and D refer to either a sequence lane highlighting G's and A's or the results of the DEPC cleavage respectively. The sequence of each strand is overlaid on the sequence lane with the position of the octanedioles shown by *.

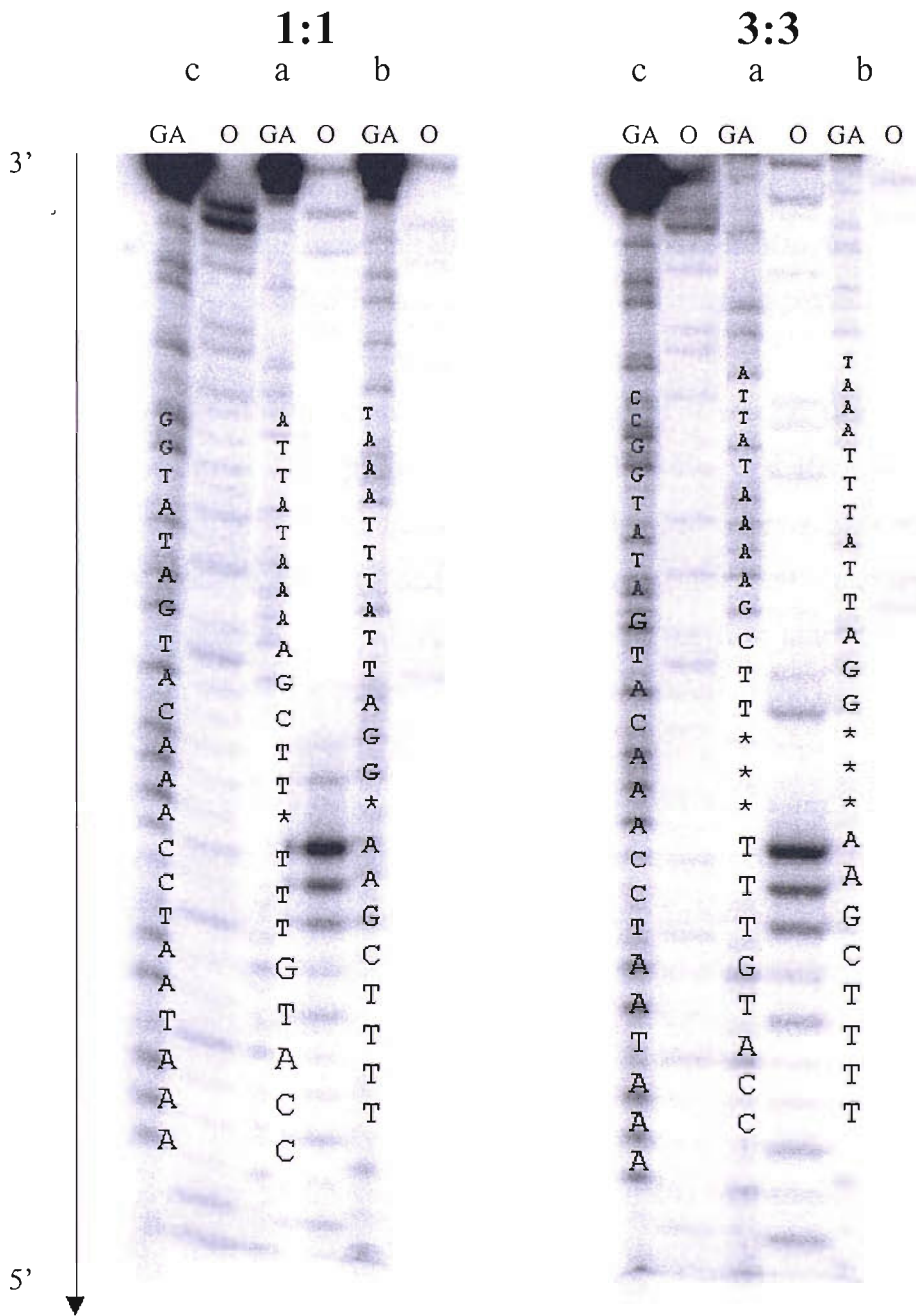


Figure 4-12: Chemical probing of junctions 1:1 and 3:3 with osmium tetroxide. The gel above uses an identical layout to figure 4-11 but shows the products of cleavage with osmium tetroxide. The individual lanes are labelled either GA or O for the GA sequence lane or results of the osmium probing.

4.3 Re-designing the extension assay to SMART conditions

Work with British Biocell progressed to using a junction designed to recognise *Salmonella* 16S ribosomal RNA (rRNA). At this point it was deemed necessary to bring the extension conditions in line with those in the SMART assay to compare data more directly. The SMART assay works by detecting very low levels of target DNA and uses pmol concentrations of the probes. It also uses a more complex buffer and includes other factors such as RNA polymerase present in solution.

The assay is further complicated by the use of a second three-way junction to amplify the initial RNA produced (Figure 4-13). The first RNA is used as a target with two new probes binding to it. A second round of extension and transcription ensues to produce RNA of different sequence to the first. It is the second RNA which is eventually detected. All the oligos for the first junction, second junction and detection stages are present in the reaction, as well as the target rRNA.

All the buffers, solutions, oligos, and enzymes were made from the stocks at BBI to avoid any differences in batches. Initially a short, synthetic DNA target was used to test the junction but RNA was also isolated from *Salmonella* and *E.coli* cultures. All oligos and RNA were aliquoted and frozen to avoid contamination and damage from freeze-thawing. Chapter 2 contains the details of the buffers, oligos and RNA isolation.

During the course of this section of work, a number of different junction sequences were used. Junction J0 and J2 were the first junctions used for the detection of *Salmonella* RNA. The latter was taken as the preferred junction for its better specificity for *Salmonella* over other similar species. The sequence recognised by J2 is 2 base pairs removed from that of J0, hence the name. Only the foot sequences were different, the optimised sequence along the transcription arm is identical in both (Figures 4-14 and 4-15). Both of these junctions use short oligos called facilitators to bind to the target on either side of the junction and help open up the DNA, or RNA in this case. They are designed to prevent secondary structures encroaching on the 3WJ. Examples of work with both junctions are shown through this chapter.

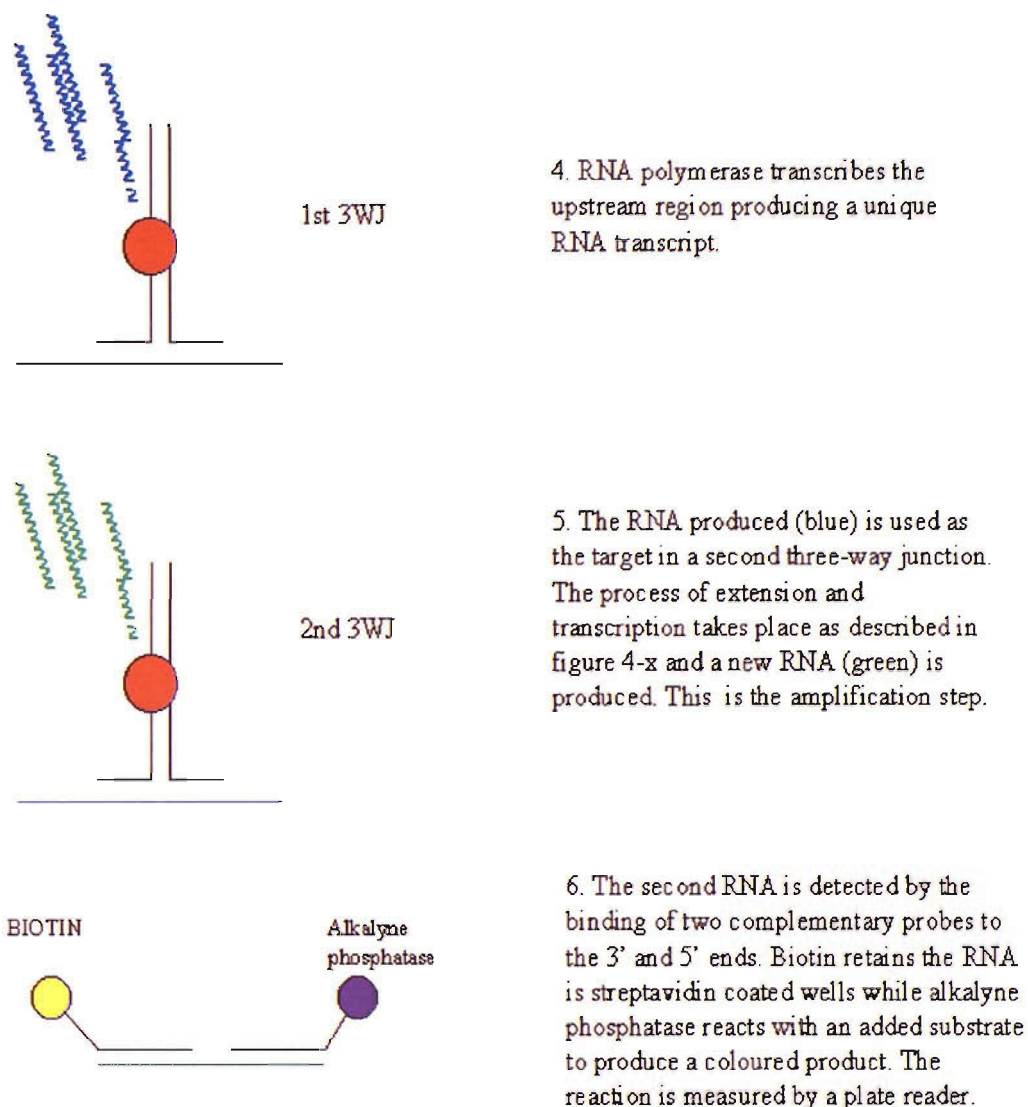


Figure 4-13: Outline of the SMART reaction from transcription of the first RNA to detection. The diagram follows on from figure 4-1. In the first step, the initial RNA is produced from the optimised sequence at the end of the junction. The RNA produced is then used as a target for a second set of probes. This is called the second junction. The next round of transcription produces RNA of different sequence, shown in green, which is then used as a target itself, this time for two linear probes. The probes are covalently attached to biotin or alkaline phosphatase to allow capture and detection. Biotin binds to streptavidin, which coats the walls of wells in a microtitre plate. This ensures the retention of the RNA in the wells. Alkaline phosphatase then undergoes a chemical reaction with its substrate to produce a colour reaction.

Original Junction (JO)

-20bp feet

Oligos:

Synthetic target 1733

Extension 1734

Template 1735

Extension Facillitator 1736

Transcription Facillitator 1760

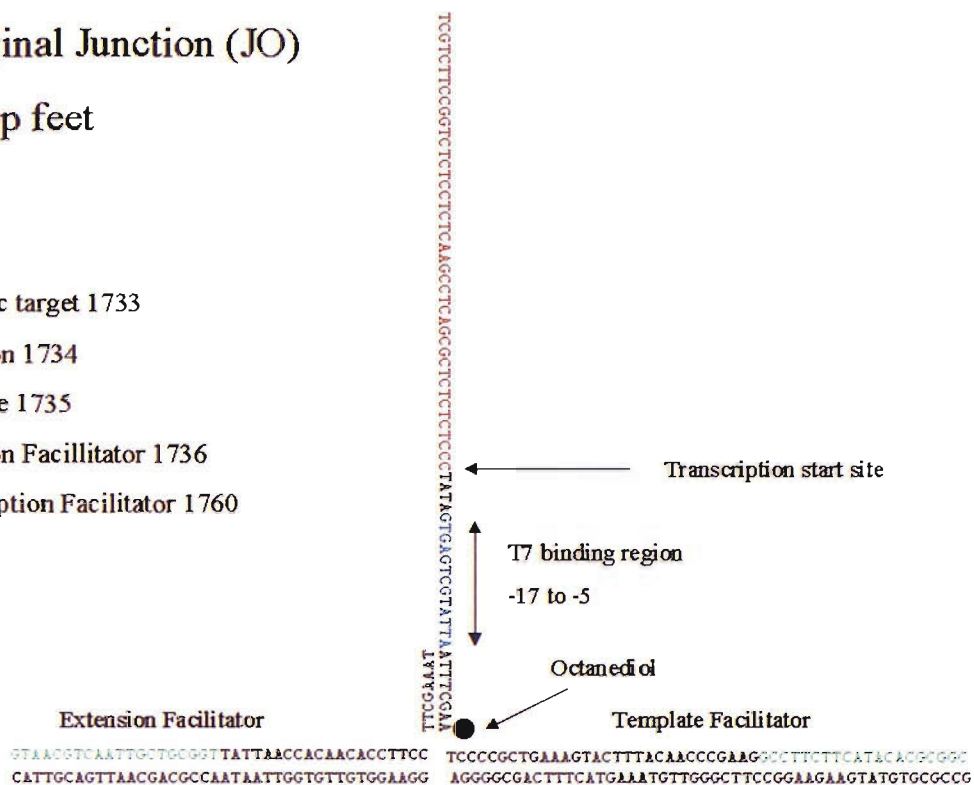


Figure 4-14: Sequence of junction JO plus facilitators. The target sequence is shown across the bottom running from 5' on the right to 3' on the left. Shown bound to it are the extension and template probe plus the two facilitators (shown in green). As in the MecA and Coag junctions, the overlap between the extension and template is only 8 base pairs. The T7 promoter is shown in blue. This is the region which must be double stranded for the RNA polymerase to bind. The numbers given to the oligos have been indicated at the side.

Junction position 2

-15bp feet

Oligos:

Synthetic target 1733

Extension 1848

Template 1849

Extension Facillitator 1850

Transcription Facillitator 1851

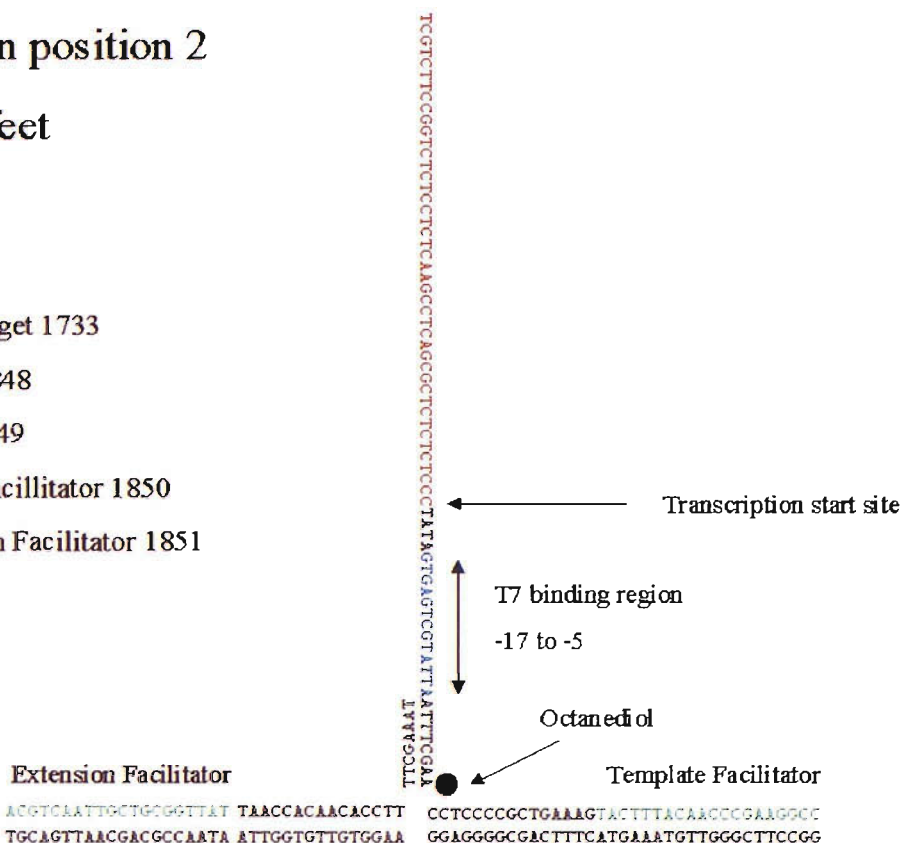


Figure 4-15: Sequence of Junction J2. This junction uses the same target region as the J0 junction but the sequence recognised has been moved by a couple of bases to give better discrimination between similar micro-organisms. The one big difference between J0 and J2 is the size of the feet (the region complementary to the target on the extension and template). J0 uses 20bp feet whereas J2 uses 15bp. The size of the feet can make a large difference to the specificity and target signal in the assay. Long feet have more chance of binding to an unrelated sequence and producing a signal, however they will also be more stable on the true target. There is a greater signal strength but also a greater background signal. Shorter feet provide a smaller but more specific signal necessary to prevent false positives.

4.31 Extension under SMART assay conditions

The result of the extension using *Salmonella* rRNA as a target is shown in Figure 3-16. The probe from JO starts off un-extended at 28 bases (the length of the foot region plus the overlap with the template). When fully extended it should reach 88 bases. No full length product was produced in any of these reactions. Instead, there was a short product about 45 bases long. This would mean that the extension probe had been extended just 17 bases and no further. By 90mins this distinct band has dissipated and there is a band another 4 bases further up, suggesting that there was some extra addition of nucleotides late in the reaction. A similar pattern is obtained with junction J2. A band is produced at just under 40 bases, indicating that approximately 13 nucleotides have been added.

The first assumption was that foreign DNA present in one of the components was causing the short probe to be extended to a different length. Alternatively, one of the junction oligos might have been degraded, leading to only a partial junction being formed. All the batches of the buffers and enzymes were changed, new tubes of oligos were used. The formation of the junction was confirmed by running each oligo and combination thereof on a gel. No source of the contamination could be found.

The next stage was to break down the assay and remove each extra component in turn to see which parts were responsible for the anomalous result. The results of this analysis are shown in Figure 4-17. It can be seen from the results that removing the template, target and facilitator oligos did not prevent extension. The short extension product is independent of the three-way junction and must be formed in preference to extension from the junction. This implies that something is interfering with the normal extension process when all the components are present. Normal extension is prevented but maybe some other secondary structure is being extended.

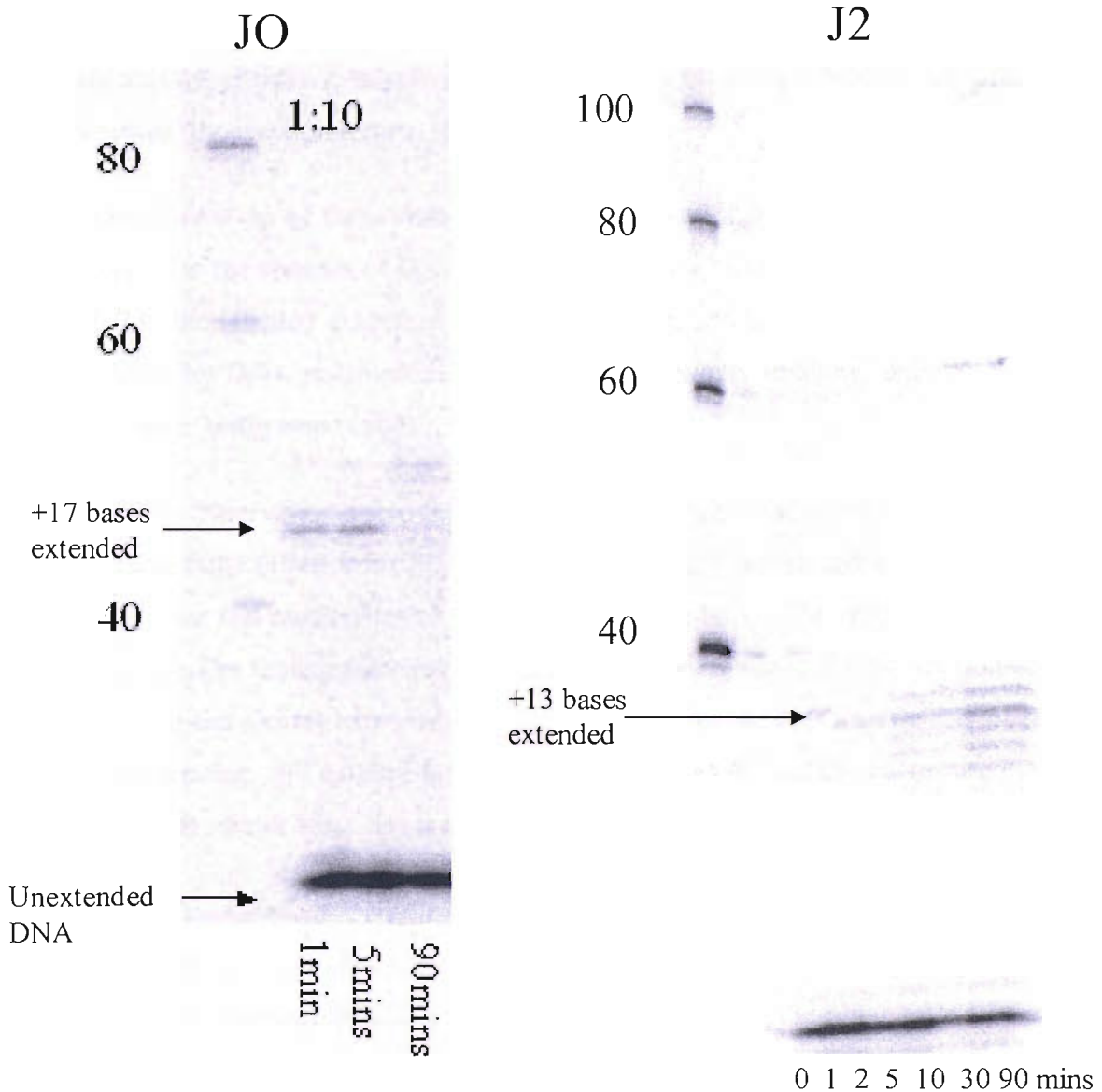


Figure 4-16: Extension of Junction JO and J2 under full SMART conditions. The reaction was performed using purified Salmonella RNA (1:10 dilution) as the target but the same result has also been shown for a synthetic DNA target. The radio-labelled probe is added into the reaction at the same concentration as in the SMART assay protocol (see materials and methods). Single stranded oligos of length 40, 60, 80 and 100 nucleotides have been used as markers. These have also been labelled and run in a single lane. The length of the anomalous product is indicated in both cases.

The results also show that a product of expected length occurs in the absence of either RNA polymerase or rNTPs. One conclusion from this gel would be that the transcription process is interfering with the DNA extension. When transcription is prevented, the reaction returns to normal.

Further follow-up of these results, however, showed that the anomalous bands are generated in the absence of DNA polymerase (Figure 4-18). Nucleotides were being added to the labelled extension probe by some process other than by normal DNA extension by DNA polymerase. The evidence, however unlikely, pointed to RNA polymerase being responsible.

Taking the observation a step further, a series of reactions were set up using only the extension oligo (1848 from J2), the standard SMART buffer and a combination of polymerases and nucleotides to generate every possibility. The results are shown in Table 4-2. The table quite clearly shows that the minimum components needed for generating the shorter extension are RNA polymerase and rNTPs. If either of those two are missing, full extension is seen. The shorter extension always occurs in their presence no matter what else is included.

Promoter-independent DNA-led primer extension has been reported for wild-type T7 RNA polymerase (Krup, 1989; Sousa and Padilla, 1995). Krup (1989) reported that extension occurred with short oligos (25mers) when the oligo folded back on itself to create a hair-pin loop (Figure 4-19). The RNA extension was complementary to the remainder of the oligo, which served as both primer and template. It is possible that something similar would be seen with the *Salmonella* extension oligos. On examining the sequence, it can be seen that a homo dimer could be formed between two oligos. The same is true of JO and J2 extension oligos. This is demonstrated in Figure 4-20.

To test this theory, a random 35mer oligo was tested in the system. The results of this extension and the exact sequence used are shown in Figure 4-21. Oligo 1 was similarly extended in the presence of RNA polymerase and rNTPs, but for at least 25 bases. This is much further than the extension product from JO and J2. It is likely that under such forced conditions as the SMART reaction, T7 RNA pol can take advantage of any structures formed by short oligos.

Components test – Original junction

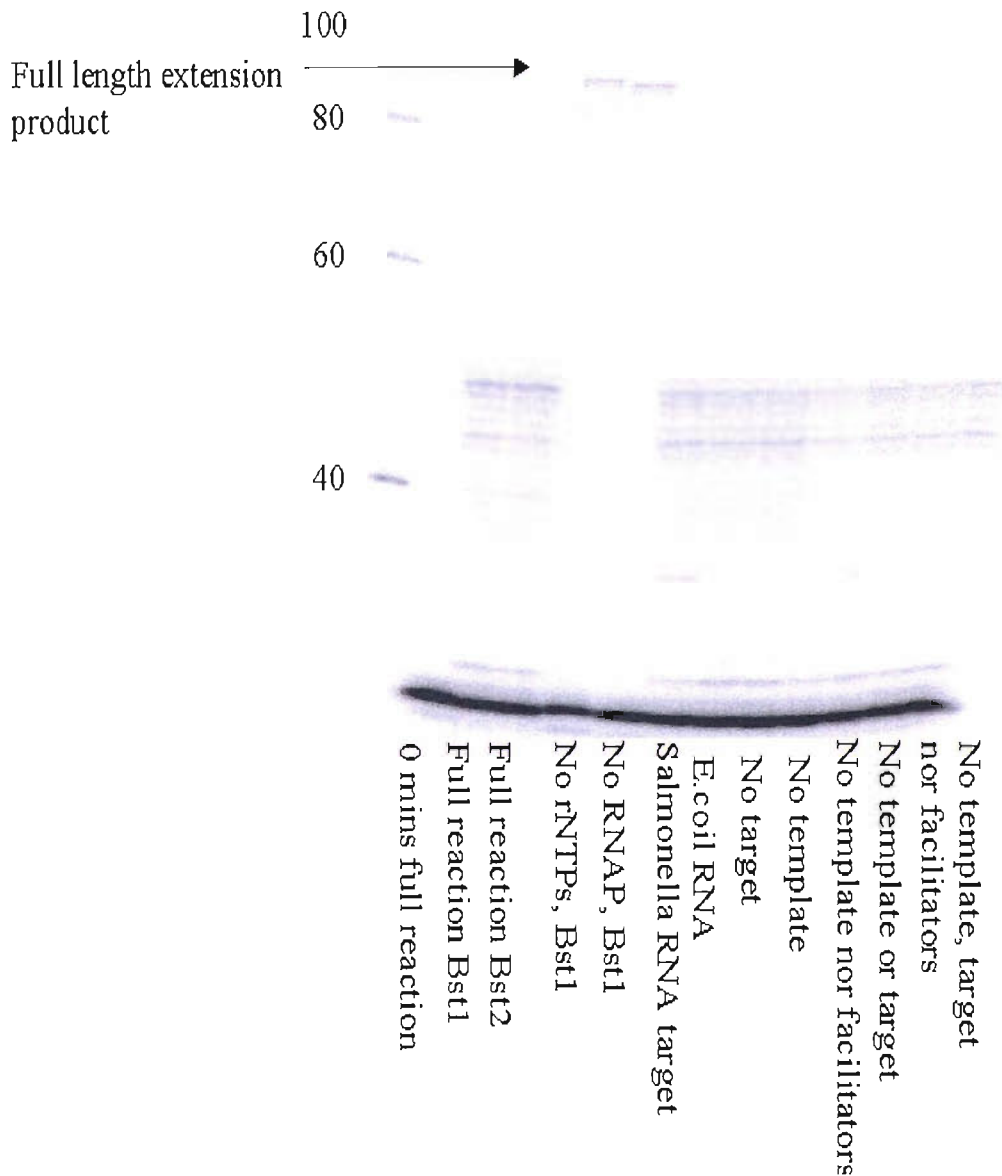


Figure 4-17: Extension of the Salmonella junction under varying assay conditions. The conditions used are stated below each lane. Single-stranded oligos of known length have been labelled and used as markers. These are shown in the first lane. The reactions cover just about every variable. The first lane shows a 0min reaction, that is without the addition of *Bst* DNA polymerase or RNA polymerase (the two are added together after annealing and equilibration of the samples at 41°C). The reactions with *Bst1* and *Bst2* refer to 2 separate batches of the DNA polymerase. All other reactants are identical to the full SMART reaction. The reaction then eliminates components from the full reaction. First rNTPs are left out, then RNA polymerase. *Salmonella* rRNA has been used as a target to eliminate the synthetic target as a problem. *E.coli* is also used as a target as a background control. Finally different combinations of oligos are used.

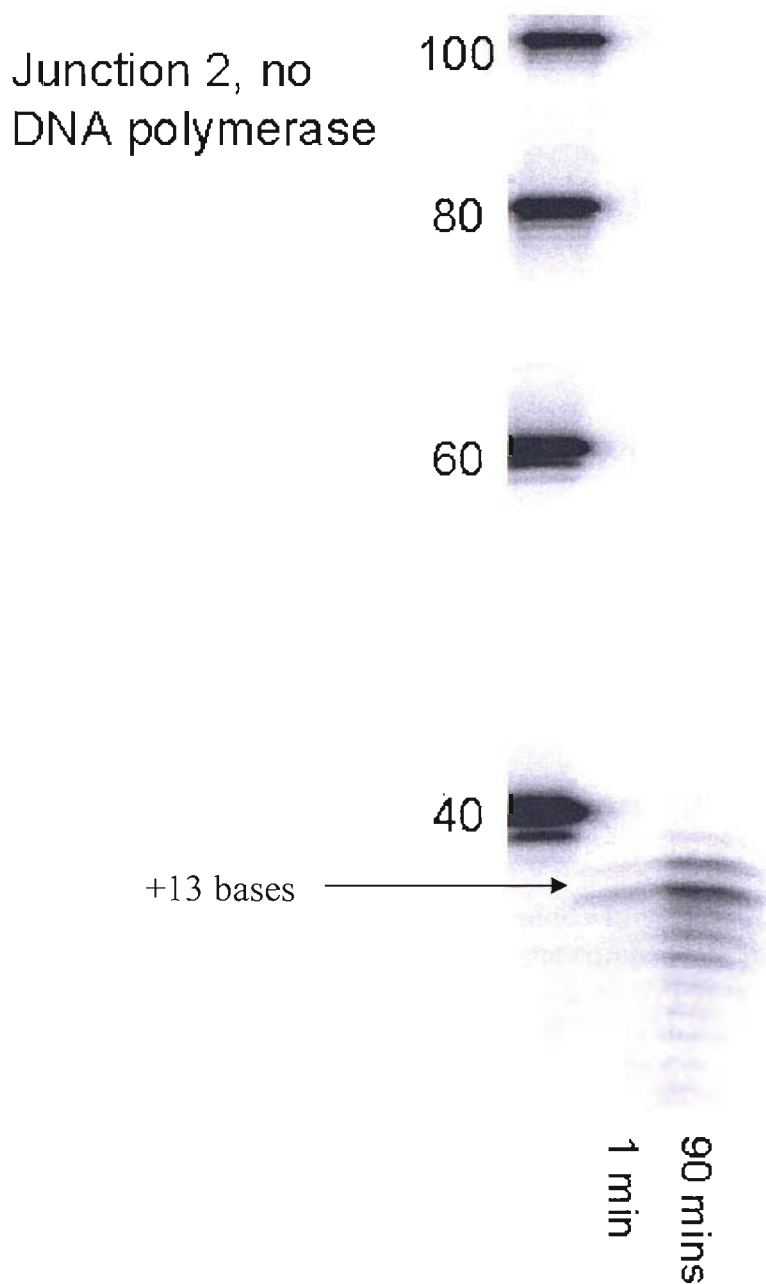


Figure 4-18: Extension of oligo 1848 (J2 extension) by RNA polymerase using rNTPs. No DNA polymerase was present in this reaction. Only the labelled extension oligo is present, it is not part of a junction. Only the 1 and 90 min reactions are shown for comparison.

Conditions under which banding pattern is seen

	DNA pol	RNA pol	dNTPs	rNTPs	Shorter banding pattern seen
1	Y	Y	Y	Y	YES
2	Y	N	Y	Y	NO
3	Y	Y	Y	N	NO
4	N	Y	Y	Y	YES
5	N	Y	N	Y	YES
6	N	Y	Y	N	NO

Table 4-2: Analysis of the conditions under which the shorter extension product is seen. Reactions 1-6 are detailed with the components they contain. Each of the polymerases and nucleotides are ticked with a yes (Y) or no (N) for their presence in the reaction. The column at the end notes whether the shorter extension product was present or not. The row highlighted in red indicates the minimum components necessary for the banding without full length extension.

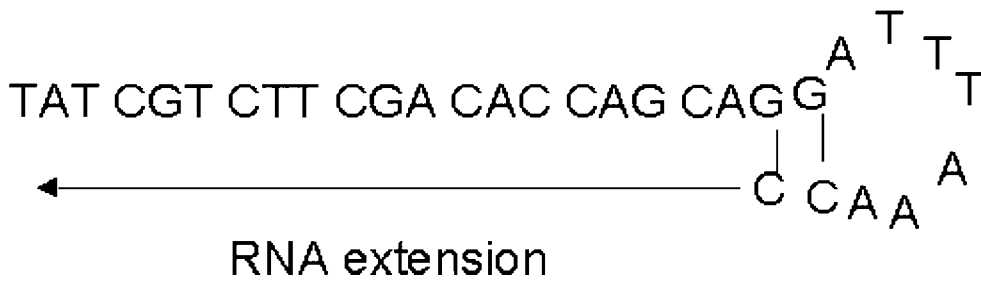
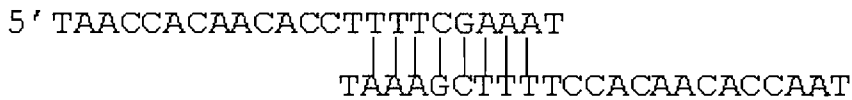


Figure 4-19: Hairpin loop structure formed by a short DNA template (Krup 1989). The structure folds back on itself generating a very short overlap of 2 base pairs. These interactions are enough to stabilise the structure so that T7 RNA polymerase can use them as a template to add nucleotides. The sequence of the RNA produced was confirmed by sequencing to be complementary to the region shown above. This activity is independent of the T7 RNA polymerase promoter.

Ampi 1848



Ampi 1734



4-20 Possible dimers formed in oligos 1848 and 1734. Both ampi 1734 and 1848 show the potential to concatemerise. This is due to a palindromic sequence between the foot and the overlap. Ampi 1848 shows a greater potential to bind to itself but 1734 also binds along the same region.

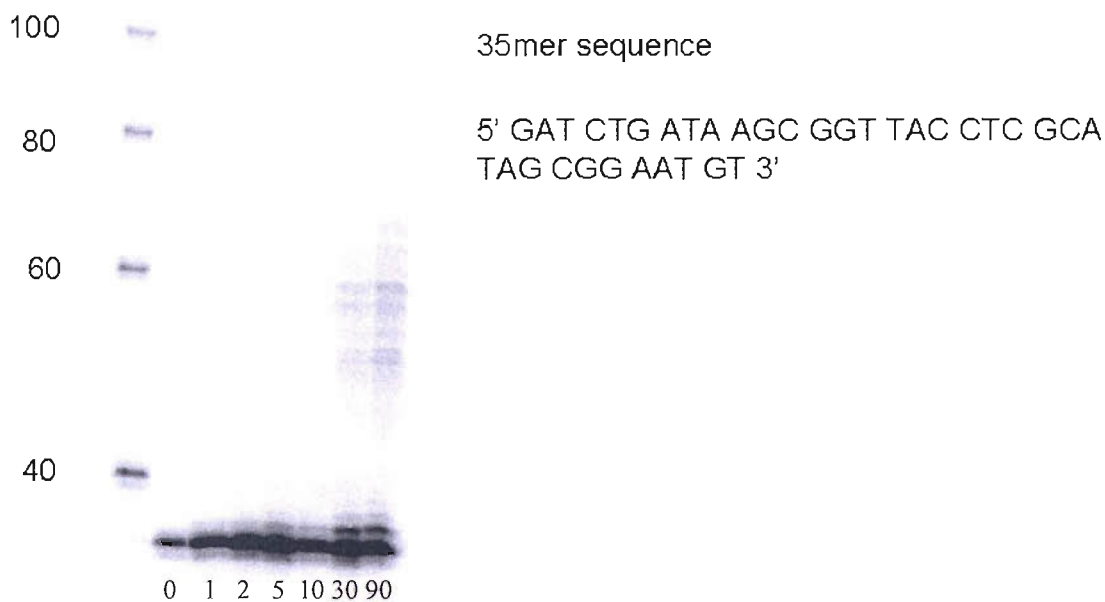


Figure 4-21: Extension of an oligo unrelated to the SMART sequences, by T7 RNA polymerase. The 35mer has been 5' labelled and placed under the SMART buffer conditions. There was no complementary oligo, DNA or dNTPs present in the reaction, only RNA polymerase and NTPs. The time in minutes at which each reaction was stopped is shown across the bottom of the gel. The time course covers 0-90mins. Markers to show the length of the oligo at the start and after extension are shown on the left.

4.3.2 Adding octanediols to prevent non-desired extension

To find out if an octanediol at the branch point would make any difference to non-desired DNA primed RNA extension, ampi 1253 (MecA extension) and 1695 (1:1 extension) were compared in the SMART reaction. The components of the reaction were identical to standard CytAMP conditions but lacked dNTPs and DNA polymerase. Ampi 1695 used the same sequence as 1253 but contained a single octanediol at the branch point.

It can be seen from the results in Table 4-4 that Ampi 1253 did show extension of the DNA by RNA polymerase. Approximately 30 bases were added. The pattern of extension using ampi 1695 is identical. Analysis of the results showed that 90% of the ampi 1253 DNA is extended in this way after 90mins. Ampi 1695 shows no lessening of this at all giving 94% extension. Ampi 1701 (2:2 extension) and 1706 (3:3 extension) were also tested in the RNA only system. These two oligos contained 2 and 3 octanediols at the branch point respectively. Both showed the same pattern of extension as before, giving 88 and 90% extension respectively.

The remaining two oligos connected to the Salmonella SMART reaction were ampi 1734 (J0) and 1805 (2nd junction). These were extended by the RNA polymerase but not to the same extent as MecA. Ampi 1734 showed 32% extension after 90mins while 1805 showed 30%.

4.3.3 Using alternative RNA polymerases

Ampi 1848 was used with a variety of alternative polymerases (T7 plus, T3, SP6 and Epicenter T7). The results of these experiments are also shown in Table 4-3. Analysis of the results showed that standard Ambion T7 RNA polymerase extended 33% of the DNA (table 1). Of the other polymerases tested, only SP6 showed very little interaction with ampi 1848 (6%).

4.3.4 Changing the overlap region to prevent concatemerisation

Since the addition of the octanediol to the extension oligos made no difference to the extension, a further set of oligos, based on the junction 2 extension oligo, were synthesised. These contained alternative T7 RNA polymerase binding sites. These would not have been as efficient at recruiting the polymerase to the SMART junction but did lessen the ability of the oligos to dimerise or fold back on themselves. The results are shown at the bottom of Table 3-4. Amp 1881, containing an octanediol at the branch point did reduce the extension by RNA polymerase slightly (24% compared to 40% in the reaction with 1848). Amp 1882 which contained the alternative –8 to –6 region showed 30% extension, while amp 1883 containing the alternative –1 to –5 showed just 9% extension. Amp 1884 which contained both alternative regions showed just 7% extension (barely visible on the gel image).

Oligo	Description	% extension by RNA polymerase after 90mins
Adding octanediols		
Ampi 1253	MecA extension oligo	90%
Ampi 1695	1:1 extension	94%
Ampi 1701	2:2 extension	88%
Ampi 1706	3:3 extension	90%
Other Salmonella oligos		
Ampi 1734	Salmonella J0	32%
Ampi 1805	Salmonella 2 nd junction	30%
Alternative RNA polymerases		
Ampi 1848 T7 RNA pol	JP2 extension	33%
Ampi 1848 T7 Plus RNA pol		44%
Ampi 1848 T3 RNA pol		36%
Ampi 1848 SP6 RNA pol		6%
Ampi 1848 Epicenter T7 RNA pol		29%
Alternative overlap regions		
Ampi 1848	JP2 extension	33%
Ampi 1881	As 1848 but with octanediol	24%
Ampi 1882	Alternative –8 to -6	30%
Ampi 1883	Alternative –5 to -1	9%
Ampi 1884	Both alternative regions	7%

Table 4-3: Analysis of the interaction of various oligos with T7 and other RNA polymerases. The table is divided into four sections to cover the different methods tried to eliminate the extension of DNA by RNA polymerase. A description of each oligo or experiment is given and the percentage extension by RNA polymerase. The extension was measured by comparing the intensity of the unextended band on the gel with that in the rest of the lane. This, therefore, represents any extension, unlike the DNA polymerase assays which measured the amount in the unextended band, the ladder and the full length band.

4.4 Transcription reactions

In the SMART reaction, after the extension of the overlap by DNA polymerase, RNA polymerase transcribes the downstream region, producing large quantities of RNA and amplifying the signal. Despite the tendency of the extension probe to dimerise, results are seen in the SMART reaction following the amplification steps and detection in the ELOSA. There must be DNA extension taking place and transcription to allow the final detection.

A transcription experiment was carried out using the same conditions as the full SMART reaction. ^{32}P α -UTP was used in the reaction to be incorporated into the transcribed RNA. The experiment was carried out under RNase free conditions to protect the RNA as much as possible. Despite this precaution, there were no products visible on the gel, even after long exposure. The quantity of rNTPs used was 16mM of each nucleotide in each reaction. This is extremely high compared to the 1/6th of the 1mM radioactive UTP in the extension reaction. The volume of labelled UTP added was increased to 5mM (5 μ l) but a high number of counts appeared in the supernatant when the DNA/RNA was precipitated, suggesting it had not been incorporated. Putting any more in the reaction was becoming very expensive and difficult to work with. The only solution was to buy separate nucleotides and reduce the amount of standard UTP added to the reaction, in the hope that the labelled UTP would be better incorporated. This would have proved expensive for a trial run and, because the concentration of UTP would have to be dropped substantially, would probably mean less RNA would be produced. This would, therefore, not be representative of the true reaction.

4.5 Discussion

The primer extension assay used in the first section to analyse the SMART junctions is a simplified version of the extension step in the SMART assay. Although it does not use the same buffers and uses only minimal components, it has been useful for analysing potential problems in the larger reaction and has highlighted a fundamental difference between junctions MecA and Coag. This difference lies in the efficiency of the extension of the primer by DNA polymerase. The simplified assay eliminates all other factors in the reaction and limits it down to the interaction between the junction and polymerase.

Looking at the basic MecA and Coag junctions themselves, there are two major differences in their extension. The first difference is the proportion of the primer reaching full length. This is 38% in Coag and 27% in MecA. The values obtained have been averaged over three sets of gels. None of the band intensities measured varied any more than within 10% of the mean. The difference between the extension of the primer to full length is, therefore, significant. The second difference between the junctions is the percentage of primer remaining un-extended. 41% of the primer remains un-extended in the MecA junction and yet only 9% of the primer in the Coag junction remains un-extended. The concentration of primer and complementary strands in both junctions is identical, and the two junctions were annealed under identical conditions. This means that any differences in the annealing of the junctions are not due to the conditions or concentrations used.

It is possible that MecA, being particularly AT rich would not anneal and hold together as well as Coag. It would take less energy to be able to pull the junction apart. However, this seems unlikely to account for the large difference in the initiation of extension. The chemical probing of MecA and Coag in chapter 3 demonstrated that the two appeared tightly base paired. It is also a possibility that the angles formed between the arms are obstructing the access of the enzyme to the branch point. However, the long-short arm assay demonstrated that the two adopt a similar dominant conformation.

The addition of extra linkers to the branch point improved the signal output in the SMART assay up to the addition of three linkers on each side of the branch point. In the scaled down extension assay, the junctions with the extra linkers did show some improvement in the proportion of primer extended to full length, but more strikingly, the proportion remaining un-extended dropped dramatically from 41% in the standard MecA junction to just 13% in the 3:3 junction. A high percentage of the DNA in the ladder would extend past the T7 RNA polymerase promoter site and would, therefore, be of use in the later parts of the SMART reaction. This would confirm the increased signal output shown by the 3:3 junction in full SMART.

The 3:3 junction was reported to be much improved in signal output in the SMART reaction compared to the standard MecA. The results of the long-short arm assay for the modified 3:3 appear to suggest the reason why. The conformation is completely altered to a conformer A/C. Since the sequence has not been changed from that of MecA, this must be a result of the extra octanediols. This is a better position for DNA polymerase to get access to the overlap region at the branch point. The 1:1 junction does not switch conformation but does show a difference in the angles formed between the arms. The angles of the arms are more equal in 1:1 than in MecA, probably also aiding interaction with the overlap region. Chemical probing shows identical results for the 1:1 and 3:3 junctions. This is unexpected if they do form different dominant conformations. There were a few bases which were particularly reactive around the branch point, suggesting that these two may spend more time unstacked than MecA. It is not clear how these results related to the detection of a genomic target. The DNA/RNA targeted would be much longer and may show less of a tendency to bend than a short length of DNA used here. If this is the case, then a junction which is dominantly in conformer A/C may have a more energetically favourable situation than a dominant A/B junction which is forced into A/C due to the inflexibility of the target.

It is interesting to note that although *Bst* DNA polymerase is reported to be a highly processive polymerase and does not have 3'-5' proofreading exonuclease activity (Aliotta *et al.*, 1996), it does appear to dissociate from the primer and template before it has reached the end. It is unclear why this should be. There are no specific sites at which it dissociates, instead it produces extension products stopped at every

base. The junctions containing the extra octanediols were intended to improve extension by lifting the overlap region away from the branch point and facilitating interaction by DNA polymerase. Although this improves the signal in the SMART assay by increasing the amount of DNA which clears the promoter region, it does not result in significant changes in full length extension.

In the second section of work, there is no doubt that the RNA extension reduces the efficiency of the DNA extension. This places it beyond the limits of detection by autoradiography. This is also the case with the RNA as a target even in the absence of RNA polymerase. Reactions have been tried using neat *Salmonella* RNA to no avail. The RNA has been quantified to 90ng/ml using the RiboGreen quantification kit. This is enough to give a positive result in the SMART reaction but not enough to detect on a gel. It may be that even if the junction is improved substantially, it may not be possible to have a particularly detailed view of the reaction under current CytAMP conditions.

The promoter-independent extension of a DNA or RNA primer is not unheard of. As already mentioned, Krup, (1989) saw extension of a DNA hairpin by T7 RNA polymerase. Cazenave and Uhlenbeck, (1994) also reported similar behaviour when a transcript produced from DNA-dependent promoter driven extension, folded back on itself to give a primer and template. The T7 RNA polymerase extended the primer in a promoter independent fashion. They tried this with a number of different transcript and found that the same activity occurred with transcripts that formed relatively unstable hairpins. They suspected that the interaction with the RNA polymerase then stabilised these structures. Figure 4-20 showed complementarity between the short extension probes being used. In both of these it appears to be dimerisation of the sequences rather than hairpin structures responsible but the effect is the same. A DNA primer is generated with a short template. It would be expected that, with such a structure in the reaction, DNA polymerase would fill in the rest of the template. It does not appear to. This could be a function of the quantities of dNTPs and rNTPs in the reaction. Per reaction in the SMART assay, there are 16mM rNTPS compared to 0.2mM dNTPs. The difference exists because there should be a much higher amount of RNA being transcribed than DNA being extended. This is part of the amplification process. The large amount of rNTPs and

RNA polymerase probably encourages the DNA primer led extension by T7 RNA polymerase.

Measures put in place to attempt to limit the extension by RNA polymerase showed some success but were limited in their practicality. The addition of octanediols to the extension probe did reduce the extension of oligo ampi 1848 (J2 extension probe) from 33% to 24% but, to have greatest effect, the linker would probably have to be at the end of the overlap rather than at the branch point. It is unclear what effect this would have on the 3WJ formation needed for the SMART reaction. Changing the sequence to reduce complementarity was the most effective measure, but these oligos were not as successful at recruiting the RNA polymerase to the promoter and therefore did not give a good signal output in the full SMART. The first few bases on the overlap between the extension probe and template are designed as part of the promoter site. These first few bases are important in the efficiency of the promoter (Imburgio *et al.*, 2000). A quick review of the literature reveals that the best advice given for preventing RNA primed RNA extension is to alter the sequence (Nacheva and Berzal-Herranz, 2003). It appears that this extension is a feature of the SMART reaction which may be minimised but probably not eliminated.

Chapter 5

SEPARATING CONFORMERS IN DNA THREE-WAY JUNCTIONS

A novel technique is presented in this section, based on the long-short arm assay. The aim is to clarify the relationship between sequence and conformation by analysing a large number of sequences at once.

5 Separating conformers in DNA three-way junctions

Three-way junctions fold so as to maximise the stacking interactions between the arms and adopt two possible conformers. The dominant conformer is determined by the sequence around the branch point, in a way that is not fully understood. To understand the rules that determine folding and the equilibrium between conformers, a series of junctions with different sequences would need to be constructed and analysed by gel electrophoresis. While some data for different sequences has been collated and analysed, many more combinations are needed.

As an alternative approach, this chapter describes a novel method whereby pools of junctions with redundant (random) bases around the branch point have been prepared. In this way every possible combination of sequences at this position is generated. Junctions which are strongly biased towards one or the other conformer were separated by their mobility on polyacrylamide gels. These were then isolated, cloned and sequenced to reveal the identity of the sequences contained within each conformer.

5.1 Designing the redundant junctions

The method used in this set of experiments was designed to be similar to the gel mobility assay used in chapter 3. The original assay used three-way junctions with arms of equal length, all 25 base pairs in the case of MecA and Coag. The arms incorporated unique restriction sites in each arm so that each could be shortened in turn. The restricted products were run on a polyacrylamide gel and the pattern of mobility was interpreted to give the identity of the arms with the greatest and smallest angles between them.

In the assay to separate redundant junctions, a simpler method was used, with smaller junctions. Instead of using unique restriction sites in each arm, one arm of the junction is pre-shortened (Figure 5-1). With one arm shorter than the other two, the two possible conformers should look very different from each other and therefore move through the gel differently (Figure 5-2).

Around the branch point, one of the three base pairs was fixed while the other two were redundant, allowing sixteen possible combinations. The pool of assorted strands of the junction were mixed, heated to 95°C and allowed to cool slowly to room temperature. The slow cooling allowed accurate base pairing to occur between the strands when they found the correct partners. It was a concern that mis-pairing would occur where there was only one base pair difference. In these cases there would have been looped out bases at the branch point and it was suspected these looser junctions would run differently to the fully base paired junctions. However, as described below, control experiments suggest that these redundant three-way junctions adopt the correct base pairing

The mixture of folded junctions, now fully annealed, was run on a 20% polyacrylamide gel at low voltage (100V) over a twelve-hour period. The buffer and gel contained 1mM magnesium to aid the folding of the junctions into the appropriate conformers. The gels used were identical to those used for gel mobility assays in chapter 3.

Four different pools of junctions were created with different fixed base pairs in the short arm B: TA, GC, AT and CG (Figure 5.1 illustrates the case for TA) and a hexaethylene glycol spacer in strand (ii). The results of all four gels are shown in Figure 5-3. On close inspection it was possible to separate a faster moving band containing the majority of the DNA and a slower moving more diffuse band. These are marked as bands L and H for low and high mobility respectively. It was assumed that these correspond to the presence of different conformers. The two bands were excised from the gel and eluted into TE buffer, then ethanol precipitated.

The purified DNA was cut with BamHI (sites shown in red in Figure 5-1) to generate sticky ends for ligation into the plasmid pUC18. The junction could ligate into the plasmid one of two ways (Figure 5-4). The ligated plasmids were transformed into *E.coli* by heat shock and grown in media containing ampicillin to select for bacteria containing the plasmid. When the plasmid is replicated inside the cells, replication of the upper strand (containing the branch point) is not possible as it is not continuous and in addition contains the non-nucleosidic hexaethylene glycol linker. In contrast the lower strand is contiguous and so is faithfully copied. The strand containing the short arm is therefore discarded and only the complement to the bottom strand is produced. This situation was observed for all the clones that were examined (over 100), that are described below.

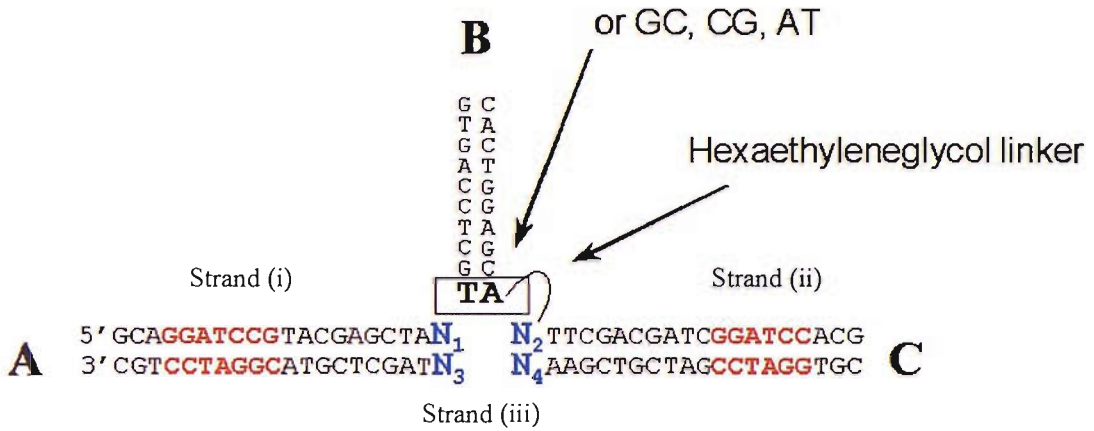


Figure 5-1: Design of junctions containing redundant bases. The junction contains two long arms, A and C, each 20 base pairs long, and one short arm (B), containing 10 base pairs. Around the branch point there is one fixed base pair (TA in the diagram above, but can be GC, AT, or CG) and four redundant bases designated N₁ to N₄. In this way 16 sequence properly paired possibilities are available for each junction set. A hexaethyleneglycol linker was incorporated into strand (ii) to allow the junction to fold (shown at the branch point). The sequences shown in red are Bam HI sites used to clone the junction into a plasmid for sequencing.

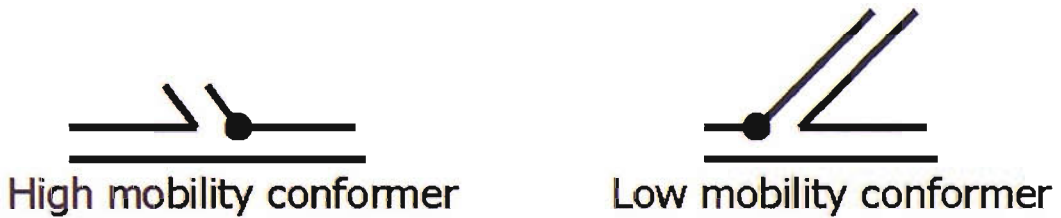


Figure 5-2: The two possible conformers resulting from the folding of the junction in Figure 5-1. The folding of the junction depends on the sequence, particularly around the branch point. Sequences which are biased towards either of the conformers above will run differently on a polyacrylamide gel, either as a high or low mobility conformer. The high mobility conformer shown has the long arms stacked continuously, allowing it to run easily through the gel with few restrictions. The low mobility conformer has one of the long arms protruding from the stacked duplex, hindering its movement through the gel.

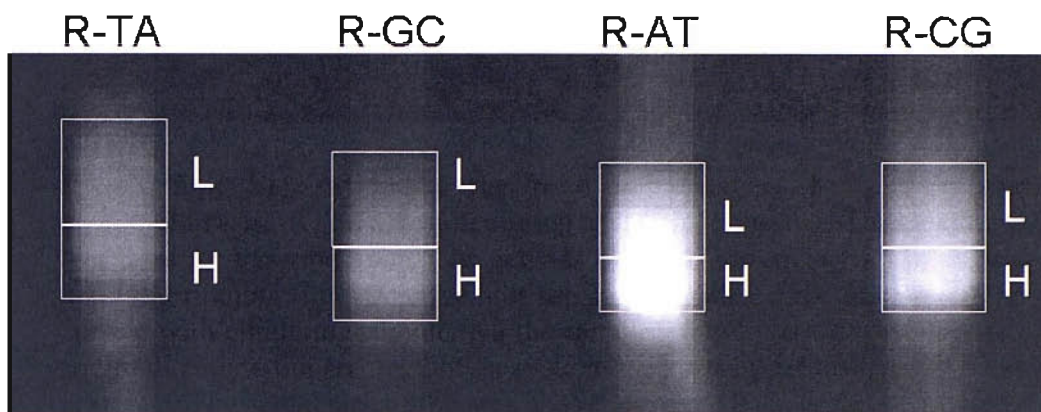


Figure 5-3: Separation of the pool of junctions by polyacrylamide gel electrophoresis. A 20% non-denaturing gel containing 1mM magnesium chloride was used for the separation. The gel was run for 12hours at 100V. The low voltage and long running time gave the best separation between the junctions. The gel was stained with ethidium bromide. All the samples above were run on the same gel and so are directly comparable. Only a section of the main gel (total size, 10cm x10cm) is shown. The actual size of the area above is about 10 cm wide x 3cm high.

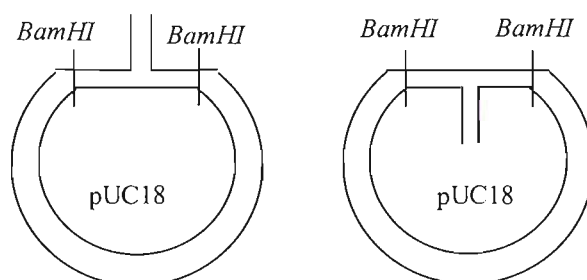


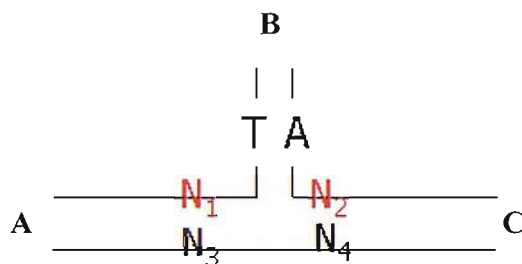
Figure 5-4: Cloning of the purified three-way junction into the pUC18 plasmid. The 3WJ is cut at a BamHI site in each of the two long arms and ligated into the BamHI site of pUC18. There is no selection for the orientation of the junction so two resulting products are possible. Both are viable but it is presumed that only the long, continuous strand (the bottom in the first picture and the top in the second) is retained during replication. It is unlikely that DNA polymerase will be able to read across the gap in the junction and past the non-nucleosidic linker. Instead the continuous strand is used as a template to generate a complete plasmid.

The selected colonies were sequenced and the identity of redundant bases N_1 and N_2 recorded. Since the three-way junction could anneal in either orientation, two possible sequences can be generated that could be read in the sequenced strand between the two BamHI sites, the bottom, continuous junction strand or its complement. The sequences therefore gave the identity of bases N_1 and N_2 directly or N_3 and N_4 from which N_1 and N_2 could be worked out. With the large number of sequences gathered it was necessary to have a simple comparison of the junctions. Table 5-1 shows the results obtained from sequencing the redundant junction with the TA fixed base pair (R-TA).

Also included with the data from the two purified bands is a column called ‘un-purified’. With each pool of junctions a control was used to ensure that the annealing, cloning and sequencing process did not favour certain sequences over others. The un-purified junctions were cloned and sequenced without having been run on the gel. The solution contains all the possible combinations of junctions, whichever conformer they might adopt. The data show a good representation of the sixteen sequences expected. There are examples of each of the four bases occurring in position N_1 and N_2 (except G in N_2). No one base dominates in either position and no one sequence is over-represented. This is a good indication that no sequences are favoured by the process and therefore the results are not biased.

The data for junction R-TA, show a clear pattern of bases in the different arms. There is a clear preference for cytosine in the first (N_1) position in the low mobility band. In contrast there does not appear to be a preference for the identity of the base at position N_2 . However, examination of the step N_1N_2 reveal that most are YY (9/13) with a few YR (4/13) and no examples of RY or RR steps. The clones from the high mobility band show a much greater diversity of sequences, as expected since this is the predominant band in the gel electrophoresis. However, there are a high number of purines in the first position, while N_2 can be any base. There is no overlap found between the sequences in the two bands and no dinucleotide occurs in both the LM and HM columns..

R-TA (N_1N_2)		
U	LM	HM
CA	CA	GA
AT	CA	GT
GC	CT	GT
AC	CT	GC
GA	CT	GC
CC	CT	GG
CT	CT	GG
TT	CT	GG
GC	CG	GG
	CG	AT
	CC	AT
	CC	TA
	TC	AA
		AC



A/B A/C

Table 5-1: Identity of bases N_1 and N_2 obtained from sequencing clones from the R-TA junction pool. Three columns are shown, un-purified (U), low mobility (LM) and high mobility (HM). The low and high mobility data are from the bands extracted from the gel. The un-purified data are controls obtained from sequencing the junction after annealing rather without gel separation. The control was necessary to show that the procedure of cloning does not favour the incorporation of certain sequences over others. The diagram beside the table illustrates the positions of bases N_1 and N_2 .

5.3 Confirming the method

To confirm that sequences cloned from the low mobility band correspond to those that predominantly adopt the A/B conformer, and therefore run slower than those from high mobility band, two junctions were constructed. In these the fixed base pair was again TA, while N_1 and N_2 were either CT or GG (corresponding to the most common sequences in the low and high mobility bands respectively). The new junctions, termed TA-CT and TA-GG, were run both separately and together on a polyacrylamide gel alongside the R-TA junction pool. The results shown in Figure 5-5, show a clear difference in mobility of the two junctions. The predicted low mobility junction, TA-CT, runs more slowly than the predicted high mobility TA-GG. This therefore confirms that the two sequences, which should be fully base paired, adopt the predicted different conformations, which can be separated by their gel mobility.

The oligonucleotides that were used to form these two specific junctions were used to further test the stringency of the annealing process. We were concerned that, for the pool of junctions of similar sequence, there might be some mis-pairing, forming either larger loops or mismatched base pairs. To minimise the problem, the oligos were annealed over as long a time as possible, when preparing the mixtures for separation. This was typically achieved by slowly cooling from 95°C to room temperature over a period of about 90 minutes.

To test the efficiency and specificity of this slow cooling method reactions were set up annealing the six oligo forming junctions TA-CT and TA-GG together in one tube. This provided potential mis-pairing opportunities. The annealing was performed in two ways. In one case the annealing was done over 90 minutes while the other was faster, taking only 10 minutes to cool from 95 °C to 4 °C. This faster cooling was expected to facilitate mis-annealing. In a further reaction the samples were deliberately mis-annealed by mixing the wrong combination of oligonucleotides, mixing $N_1N_2 = CT$ with $N_3N_4 = CC$, generating CC and CT mismatches.

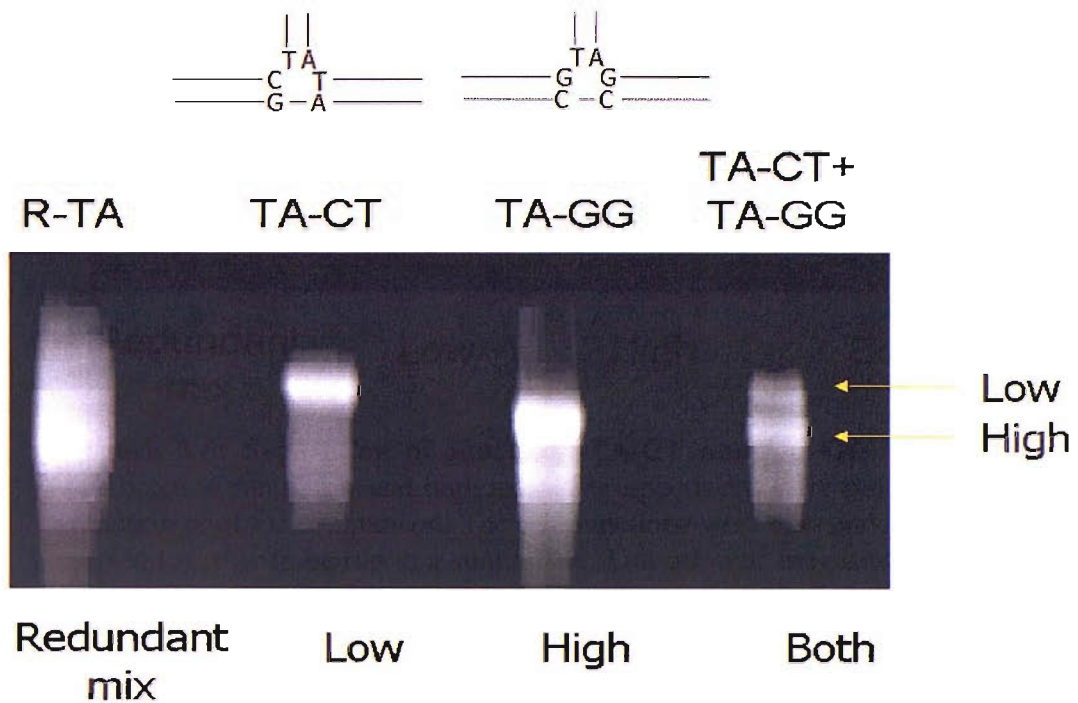


Figure 5-5: Separation of junctions TA-CT and TA-GG. The two junctions, predicted to run at low and high mobility respectively, were run alongside the R-TA junction pool for comparison. The two junctions were also annealed separately then mixed together to be run in a single lane. This allowed the exact separation between the two bands to be seen. The sequence around the branch point for each junction can be seen above the appropriate lanes of the gel.

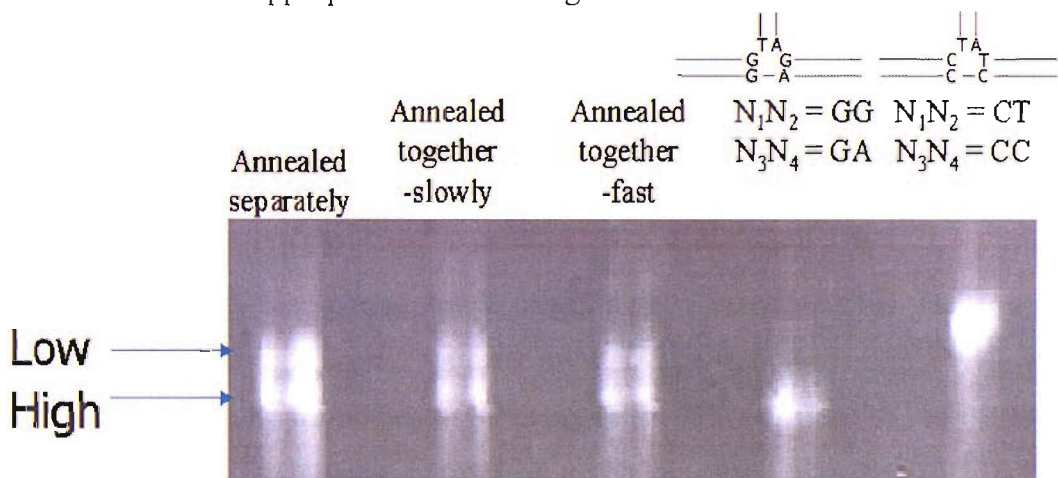


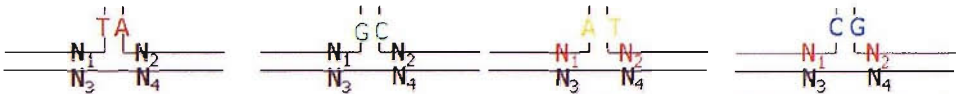
Figure 5-6: Testing the ability of complementary junctions to base pair accurately when other similar partners are available. The gel shows junctions TA-CT and TA-GG under different annealing conditions. The first lane shows the two annealed separately and then mixed together. The oligonucleotides for the junctions in the second and third lanes were mixed together first and then annealed, allowing the strands to pair correctly or with a 2 base pair mismatch. In the second lane the annealing took place slowly over about 90 minutes while the mixture was cooled rapidly to 4°C in about 10 mins in the third lane. The samples in the final two lanes have been forced to form a mismatched junction, by combining strand C from the one junction with strands A and B from the other.

Figure 5-6 shows that annealing the junctions together has little effect on the pattern of mobility and it appears that the correctly junctions are formed. Two separate bands of low and high mobility can be seen in conditions of both rapid and slow annealing, suggesting that there is a strong preference for the junction to form correctly. There is no sign of any other products being formed. More importantly, when the junction is forced to mis-pair (final lane), the resulting junctions run at a different mobility to the fully paired versions.

5.4 Other junctions

Table 5-2 shows the sequencing data obtained by sequencing the low and high mobility bands from other the junctions with different fixed base pairs (GC, AT and CG). The four different fixed base pairs combined with the redundant bases allow for every combination of the bases directly around the branch point. As already described the pattern for the R-TA junction is quite striking, with a clear preference for a C as the base in the first position in the low mobility band, generating an A/B conformer. In contrast the junction with the same sequence except for a purine (G or A) at this position, folds as a high mobility (A/C) conformer (Figure 5-7).

Examination of Table 5-2 reveals that the patterns for the remaining junctions are less clear, and in contrast to the result with R-TA there are several instances where a nucleotide step is found in both the high and low mobility bands. This could arise because there is no clear preference in the folding pattern or from inadequate separation of the different species. No clear rules emerge from these data. However, there is a tendency for pyrimidines to be found at the first position (N_1) in the low mobility conformer. The YR step is common in the low mobility bands (12/50), but is underrepresented in the high mobility bands 3/56). It is also noticeable that the RR step is completely absent from the low mobility bands, and the RY step accounts for only 7/50 sequences in contrast to contrast to 19/56 for the high mobility bands.



R-TA			R-GC			R-AT			R-CG		
U	LM	HM	U	LM	HM	U	LM	HM	U	LM	HM
CA	CA	GA	CT	TT	TC	TC	TT	GC	TC	TC	TC
AT	CA	GT	CT	TT	TT	TT	TT	GC	TG	TC	TC
GC	CT	GT	CT	TC	TT	CA	TT	GG	TT	TC	TC
AC	CT	GC	CC	TG	CG	CC	TC	GA	GC	TG	TT
GA	CT	GC	TA	TA	CC	GC	CT	GA	GT	AT	TT
CC	CT	GG	TG	TA	CC	GA	CT	GA	CG	AC	TT
CT	CT	GG	TG	CT	GC	GT	TA	GA	CA	GC	CT
TT	CT	GG	GA	CT	GC	AT	TA	GT	CC	CT	CT
GC	CG	GG	GT	CG	GC	AG	CC	TC	CC	CC	CA
	CG	AT	GG	GC	GT		CC	TC	CC	CC	CC
	CC	AT	AC	AC	GA		CC	CC	CT	CC	CC
	CC	TA	AT		AC		CA	CC			GT
	TC	AA			AT		GG	CC			
		AC			AT		GC	CT			
					AT		AC				
					AG						

Table 5-2: Sequencing results (N_1 and N_2) for the pools of junctions with the four different fixed base pairs. The junctions are listed in the table as R (redundant) followed by the identity of the fixed base pair (for example R-TA). The columns are coloured to help separate the results. For each junction there are three columns, U (un-purified), LM (low mobility) and HM (high mobility). These are from the control junction mix, and the two bands purified from the gel. In each column, the two bases found at positions N_1 and N_2 in each of the clones sequenced are listed.

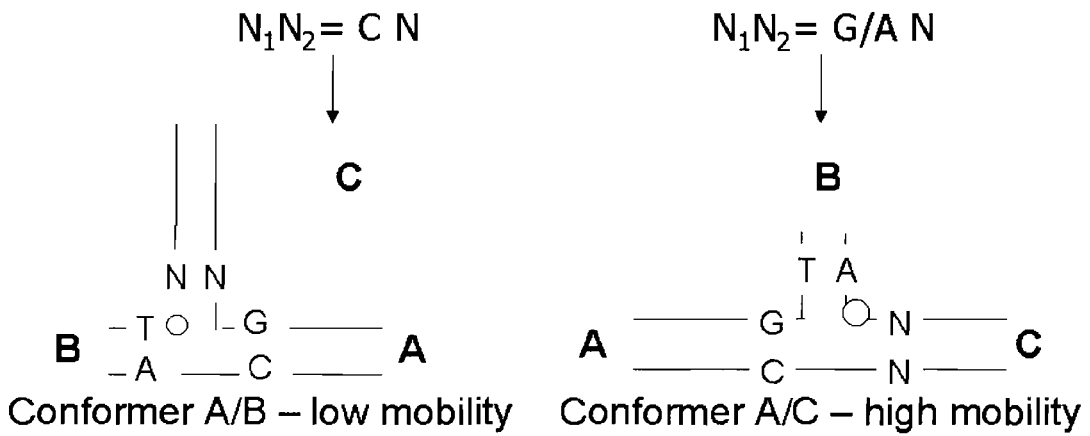


Figure 5-7: Conformers formed by placing different bases at position N_1 . A single base difference in the junction can mean a large change in the average conformation. The model above shows the predicted conformers when a C or a purine is in N_1 . Those bases still labelled N can be any base.

5.5 Discussion

A novel technique has been presented based on the long-short arm assay, to clarify the relationship between sequence and conformation. The junctions used have one pre-shortened arm of 10 base pairs while the remaining arms are 20 base pairs long. One of the base pairs around the branch point is fixed while the others can be any base. This creates a pool of 16 sequence combinations. A set of four different experiments varying the fixed base pair has led to every combination of bases immediately around the branch point being covered.

The technique involves separating the pool of junctions on a polyacrylamide gel according to the average conformation adopted, extracting the DNA from different bands and sequencing. This reveals the identity of junctions which adopt each conformer. There is clear evidence that junctions with common sequence features will migrate on the gel at the same rate and fall into the same band. These features are largely based on having a purine or pyrimidine at a certain location.

The separation of redundant junctions described here has been proved to have some accuracy, particularly when the fixed base pair is TA. The major limitation of the technique lies in the dynamic equilibrium of the three-way junctions. As with the long-short arm assay, the gels are run over a number of hours (12 hours in this case). During that time the junctions will most probably be rapidly exchanging between conformers, even in the temperature controlled mini-gels. McKinney *et al.*, (2003) showed that four-way junctions were in a constant state of flux, switching between conformer at the rate of between 6.1 to 12 times per second. Despite this, the normal long-short arm assay does normally show good, clean separation between the bands. There was very little separation between the bands in this case. Rather one observed a distinct band at the front end followed by the majority of the DNA. With such a large combination of junctions there must be a variety of angles adopted by the arms of the junctions providing the blur. There was evidence for changing angles in chapter 3 with the MecAV1 and 2 junctions and the 1:1 and 3:3 junctions in chapter 4.

There is a possibility that, with all the redundant oligos annealed together, there will be mismatches formed. To counter this problem, oligos were annealed as slowly as possible to allow accurate base pairing. The junctions synthesised later on to test the predictions made about folding, differed by two bases at the branch point. When the oligos for these junctions were annealed rapidly together and then analysed, the patterns formed were identical to those for the junctions annealed separately. There was no evidence of mismatches at all. When the junctions were forced to mispair by the use of the strand c from the opposing junction, the pattern of mobility was altered. This suggests that, should any mismatches occur, they will show a very different pattern of mobility to the fully annealed junctions. Of course the forced mismatching in this case involved two bases at the branch point. It is possible that when only a single mismatch occurs, the mobility of the junction does not alter substantially. However, the ability to accurately predict the mobility of the two sequences tested, does speak well of the technique.

It is likely that with a larger volume of data, the patterns relating sequence to mobility would become clearer. In these experiments where only the bases immediately around the branch point are changed, it should be possible to lay down some definite rules like those for R-TA. These rules may, however, change substantially when the next base pair away from the junction is changed. Sequence further away may affect the conformation of the junction, as has been demonstrated in this thesis. With the current data available, this technique cannot necessarily be used to predict the conformation of unrelated sequences. The next stage would be to analyse a larger volume of data for the bases adjacent to the branch point, and then move one base further out.

Chapter 6

MELTING STUDIES OF DNA JUNCTIONS

Part 1: Local melting analysis using a fluorophore and quencher

The melting of a set of DNA junctions is analysed by fluorescence. A fluorophore and quencher on one arm of the junction reveal the temperature at which the strands on that arm dissociate and give an indication of the stability of the junction

Part 2: Global melting analysis using SYBR Green

SYBR Green, a sequence independent DNA binding dye, binds to double stranded DNA and fluoresces when bound. It is tested here as an indicator of junction stability in melting studies.

Part 3: Measuring the angles between the arms in the junction

Fluorescence Resonance Energy Transfer occurs between a donor and acceptor molecule on different arms of the junction. The study shown here is qualitative rather than quantitative, analysing the relative positions of the arms and, therefore, the conformation.

6 Melting Studies of DNA junctions

6.1 Measuring the melting temperature of DNA junctions

The stability of any DNA structure is dependent upon the bonds that hold it together. These are usually the standard Watson-Crick base pairing which involves hydrogen bonding between complementary bases. A short duplex sequence can vary greatly in stability depending on its sequence. This is because a GC pair shares three hydrogen bonds whereas a TA pair shares only two. This difference in the stability can be most easily measured by melting the strands of DNA apart and monitoring the temperature at which this happens.

The Roche LightCycler is a device normally used for Real-time PCR. The production of the amplicons is monitored by fluorescence (either by a generic DNA binding dye eg SYBR Green or specific hybridisation probes). In both cases there is an increase in fluorescence which accompanies the increase in the amount of amplicon produced. This avoids the need to take aliquots to run on an agarose gel to check the amount of product.

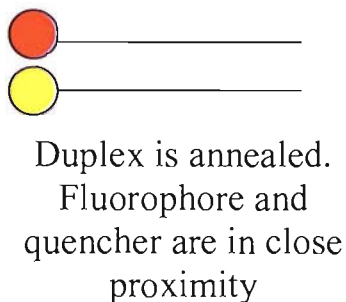
The fluorescence monitoring and programmable cycles of the LightCycler also lend themselves to examining the melting of DNA, providing a profile from which thermodynamic information can be derived. In this chapter, a number of fluorescence methods are used to analyse the kinetics of three-way junctions and related DNA structures. The first uses an attached fluorophore and quencher to measure the point at which the junction melts. The second uses SYBR green, a dye which fluoresces upon binding to double stranded DNA. Finally, FRET (fluorescence resonance energy transfer) is used to analyse the relative distance between pairs of arms in the junction.

6.2 Measuring the melting of individual arms in the junction – fluorophore and quencher.

In this method, the melting of the junction was measured using a quencher (methyl red) and fluorophore (fluorescein) attached to opposite strands of the same arm of the junction. When the junction is folded, the quencher and fluorophore lie in close proximity. The fluorescence emitted by fluorescein is quenched by methyl red when they are in close contact. As the temperature is raised, the strands of the junction melt apart. The distance between the fluorophore and quencher increases and as it does so, the fluorescence increases (Figure 6-1). The temperature at the mid-point of this transition is the melting temperature (T_m). This value can be used as a measure of stability under different experimental conditions. This technique has been used successfully for the study of duplexes, quadruplexes and triplexes (Darby *et al.*, 2002).

A short junction with an approximately equal distribution of GC and AT bases pairs was used for this set of experiments. The sequence of the junction is shown in Figure 6-2, but within this frame several variations were also introduced. This set of experiments was designed to compare the melting of parts of the junction, eg duplex compared to a junction or a 2.5WJ, partially complementary duplex and so on. The range of junctions used is shown in Figure 6-3.

Fluorescence quenched



Increase in temperature

Fluorescence high

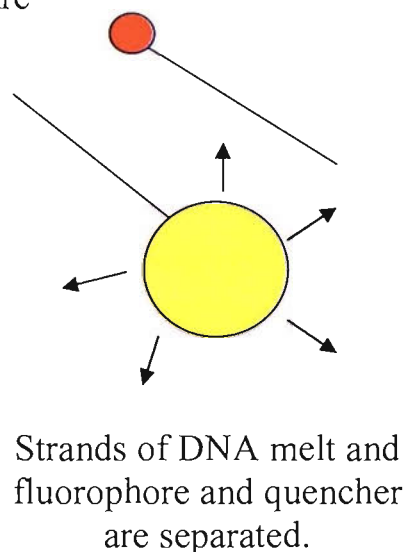


Figure 6-1: Illustration of how the melting of the DNA is measured by fluorescence. In the first part of the diagram, the DNA is shown annealed and stable. The fluorophore and quencher are adjacent to each other and the overall fluorescence is quenched. As the temperature increases, the strands of DNA will gradually melt apart and the fluorophore and quencher will be increasingly separated. During this stage, the fluorescence will increase, gradually at first and then rapidly as the two molecules are separated in solution.

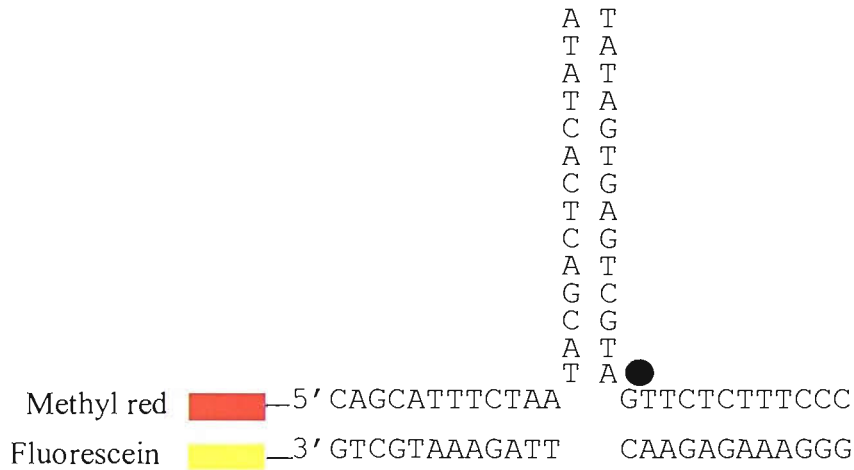


Figure 6-2: Outline of the light cycler junctions. A set of oligos were made to form various junction structures, all based on the above sequence. A fluorophore and quencher were attached to opposite strands in order to follow the melting of the strands. As in previous chapters, the filled circles indicates the position of the flexible linker (an octanediol in this case).

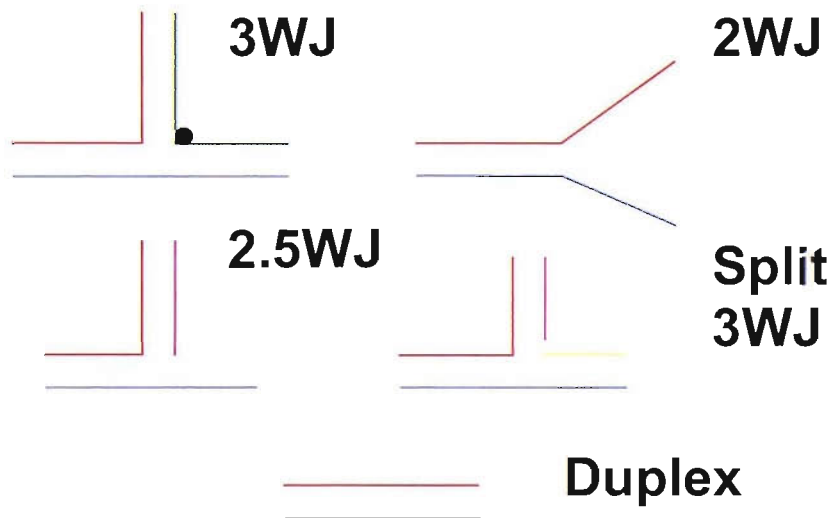


Figure 6-3: Light cycler junctions. The junctions above were constructed from component oligos based on the sequence in Figure 6-2. They were used to compare the stability of a variety of junctions by following the melting and annealing of the strands under various buffer conditions. In each case the fluorophore and quencher are located at the left hand side on the lower and upper strands, respectively.

6.2.1 Melting profile of a three way junction

The melting profile of a three-way junction and a duplex, in 50mM sodium phosphate and 2mM magnesium, are shown in Figure 6-4. The graph shows the change in fluorescence measured as the temperature of the reaction is increased. At the start of the experiment, both samples are fully annealed with the fluorophore and quencher in close proximity. The fluorescence emitted is effectively quenched.

As the temperature increases, the strands melt apart and the fraction of annealed molecules decreases. Overall the quenching diminishes and fluorescence increases. Once all the strands have melted, the full, un-quenched fluorescence is seen. The T_m corresponding to the mid-point of the transition, is estimated from the first differential of this graph (Figure 6-5).

The two structures shown in Figure 6-4, the three-way junction and the duplex, melt very differently. The duplex is considerably more stable than the 3WJ as judged by melting temperature. The duplex melts at 61°C compared to 48°C for the 3WJ (a significant difference of 13°C). This difference is expected due to the added complexity of the 3WJ. The base pairing in the duplex is continuous and stable. The pairing in the 3WJ is disrupted by the presence of the third arm. Any torsional stress would destabilise the entire structure.

Not only is there a difference in stability, there is a difference in the way the structures melt, as judged by the melting profile. The three-way junction has a slightly steeper melting profile than the duplex. The transition is much more sudden. This may be due to a larger number of hydrogen bonds in the duplex. The ΔH would affect the steepness of the curve and the height as well.

Following the melting transition, the fluorescence measured for both structures then decreases slightly. This depression is due to temperature dependence of the fluorescence and is not a function of either structure.

Melting profile of a three-way junction and a duplex

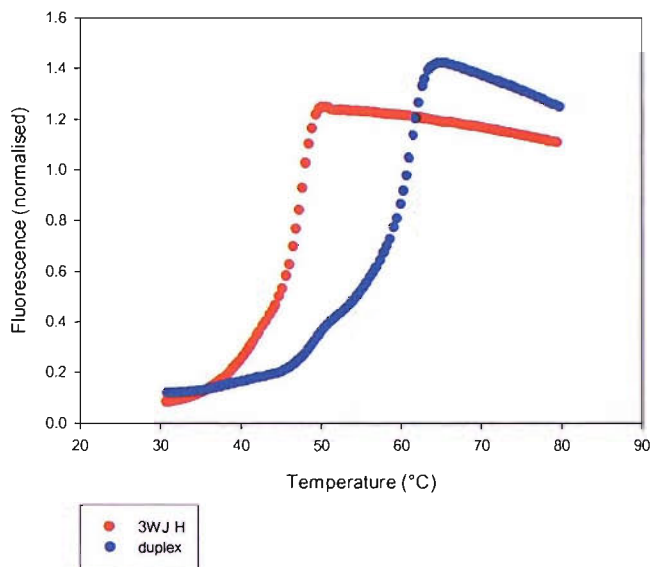


Figure 6-4: melting profile of a three-way junction and duplex of equivalent sequence. The fluorescence values are recorded at every degree change in temperature from 30 to 95°C. Only the values up to 80°C are displayed here to give a better view of the important aspects of the data. The fluorescence values have been normalised.

Rate of change in fluorescence with temperature

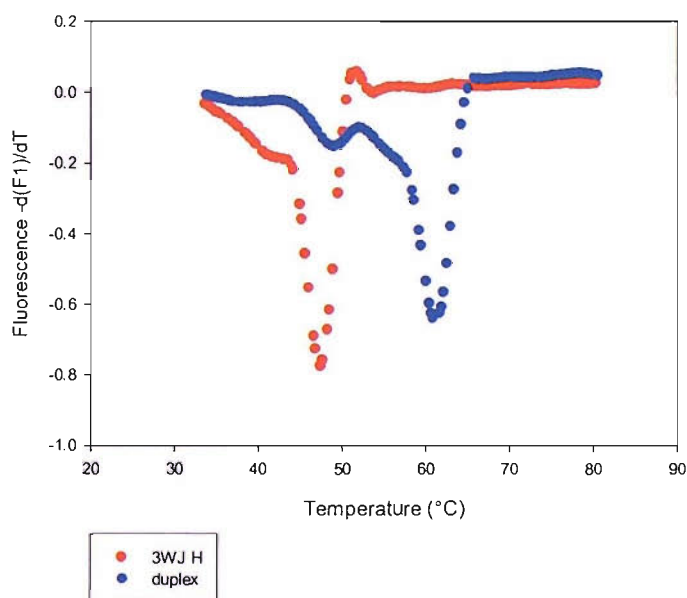


Figure 6-5: The rate of change in fluorescence over increasing temperatures for a 3WJ and duplex. The graph is obtained by differentiation of Figure 3-4. The troughs represent the maximum rate of change in fluorescence and therefore the melting temperature.

6.2.2 Effect of the identity of the linker on junction stability

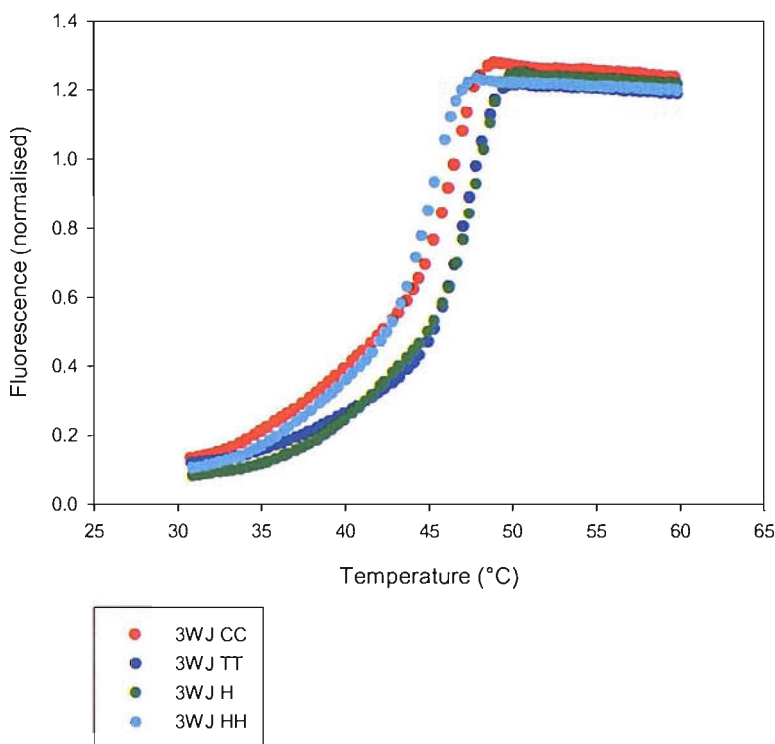
The three-way junctions used in this thesis all contain a linker of some form to give the junction the torsional freedom to stack. Previous work (Welch *et al.*, 1995) has shown that the identity of the linker makes little difference to the junction conformation, although Rosen and Patel, (1993a) showed that there may be structural differences in the way unpaired bases sit at the branch point (Section 1.1.5). These studies looked at using unpaired bases as the linker. Occasionally this can cause complications when similar sequences occur around the branch point, resulting in slippage and the ambiguities in looped and paired bases. To resolve this, a number of the SMART junctions used chemical linkers to artificially extend the backbone of the DNA. Initially hexaethylene glycol was used and later octanediois.

A set of four junctions of identical sequence was tested in the LightCycler to look for differences in the melting profile of junctions that contained different linkers. Those chosen were 2 cytosines (CC), 2 thymines (TT), a single hexaethylene glycol (H) and a double hexaethylene glycol (HH). Each base adds eight extra bonds onto the backbone, giving 16 for the double bases used here. A hexaethylene glycol contains eighteen extra bonds, giving thirty-six for the double.

As can be seen from the data in Figure 6-6, there are few differences between the melting profiles or temperatures of the four junctions. All show the same three-way junction melt and the T_m 's are within two degrees of each other. A difference of up to one degree is probably not significant and can be accounted for by experimental error. The difference of two degrees between the single and double hexaethylene glycol, although small, may suggest that the double linker introduces too much freedom into the junction, reducing stability slightly. The single hexaethylene glycol is the most stable of the linkers, though there is little difference between this, the TT and CC linkers.

Melting profiles of three-way junctions containing different linkers

A



B

Junction	T _m (°C)
3WJ CC	46.5
3WJ TT	47.1
3WJ H	47.7
3WJ HH	45.1

Figure 6-6: Comparison of the melting profiles and melting temperatures of three-way junctions with different linkers. Diagram A shows the melting profiles of the four 3WJs. The sequence of the junctions is identical, only the identity of the linker changes (2 cytosines – red, two thymines – blue, single hexaethylene glycol – green, double hexaethylene glycol – cyan). The table underneath shows the melting temperatures. Repeat experiments showed that the T_ms are reproducible to within 0.5° C

6.2.3 Stability of partial junctions

These results show that there are differences between the melting of a three-way junction and a duplex of equivalent sequence. This is presumably a result of the disruption of the helix at the branch point leading to fewer hydrogen bonds in the three-way junction. The next set of experiments examines which aspects of the structure act to stabilise or destabilise it, by comparing the stability of a series of related partial junctions.

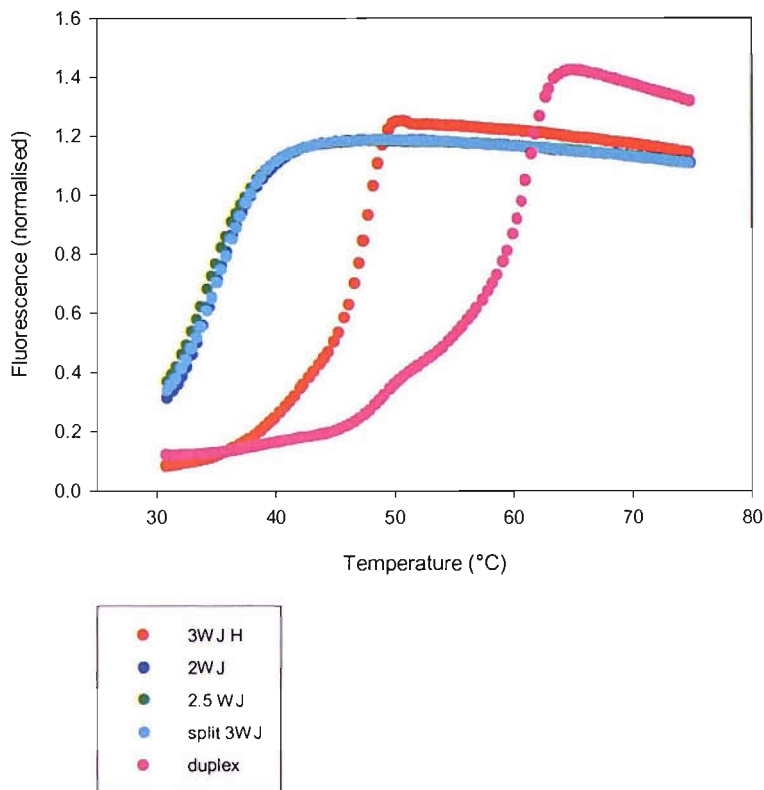
The sequences and structure of the partial junctions are shown in Figure 6-2 and 6-3. These oligos combine to give partial junctions. The two-way junction consists of a duplex with one half fully base paired leaving single-stranded ends (ie omitting the third strand of the 3WJ). The two-and-a-half-way junction (2.5WJ) has been mentioned in the introduction as the structure used for the SPAR reactions. It is intermediate between the 2 and 3WJ's, having an accessory oligo attached to arm B, but not to arm C. Finally, the split three-way junction uses the accessory oligo of the 2.5WJ with another attached to arm C. This is, in effect the same as the 3WJ but there is a nick in one of the strands.

The results for the melting of these junctions are shown in figure 6-8. It can be seen that the partial junctions all melt together while the three-way junction and duplex are entirely separate. The three partial junctions are all considerably (13°C) less stable than the three-way junction. As with the duplex, the number of hydrogen bonds in the partial junctions, and therefore the ΔH will be less than in the 3WJ and so will affect the shape of the curve to an extent.

It is clear that the accessory strands do not aid the folding of the junction as effectively as the full third strand containing the linker. More than that, since the 2.5WJ and split 3WJ show no difference to the 2WJ which has no helper strands, it is clear that they do not aid folding at all, or not in a way that can be measured using this method (the melting is only measured from the arm at the left of the junction).

A

Melting profiles of a three-way junction, partial junctions and a duplex



B

Junction	T_m (°C)
3WJ H	47.7
2WJ	34.7
2.5WJ	34.3
Split 3WJ	35.3
Duplex	61.2

Figure 6-8: Comparison of the melting profile and melting temperatures of a three-way junction, partial junctions and duplex. Panel A shows the melting profiles of the junctions in a buffer containing 50mM sodium phosphate and 2mM magnesium chloride. The data values up to 75°C are shown, after which all junctions have melted. The melting temperatures are shown in the table in panel B. The junctions used are shown in Figure 6-3.

6.2.4 The effect of placing the fluorophore and quencher on a different arm of the junction.

The placing of the fluorophore and quencher on one arm of the junction means that the melting temperature recorded may be very dependent on the sequence of the arm in question. If the sequence along this arm is very G-C rich, then it may hold together after the remainder of the junction has dissociated. It should be remembered that this technique only provides information on the proximity of the fluorophore and quenchers, which may not necessarily reflect the stability of the entire structure.

To determine the effect the local sequence had on the melting temperature, a new set of oligos was made with the fluorophore and quencher attached to arm B (figure 6-9). These oligos were combined to construct three-way, two-way and two-and-a-half-way junctions.

The T_m values for these junctions are shown in Table 6-1, from which it can be seen that the melting temperature of the three-way junction measured on arm A is similar to that measured from arm B, with a difference of about 2°C. The differently labelled partial junctions, however, show much greater differences, about 9°C. For the two-way junctions, different sequences are used to form each junction, oligos a and c to make a fully annealed arm A and oligos a and b to make a fully annealed arm B. Since the melting will be dependant on sequence, this may explain the 9° C difference in T_m between labelled arms A and B, since B is longer and has a higher percentage of GC residues.

The melting temperatures of the two two-and-a-half-way junctions are even more different than those of the two-way junctions, showing a difference of 12°C. If, as was discussed in section 6.24, the third strand does not help the stability of the junction, it would be expected that the 2.5WJ measured from arm b would be equivalent to showing arm a annealed to a short version of oligo b. This structure should melt relatively easily and certainly much lower than its equivalent three-way junction.

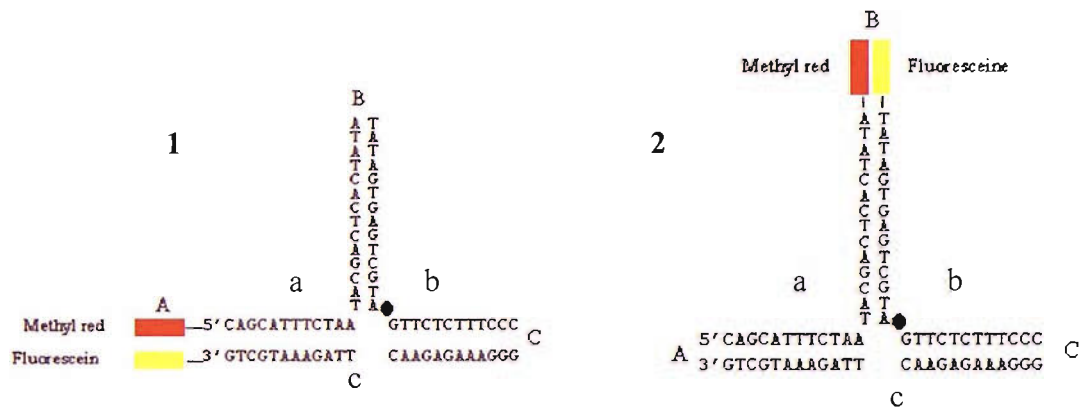


Figure 6-9: Positions of the fluorophore and quencher at position 1 on arm A and position 2 on arm B. The sequence of both junctions is identical but the positions of the labels are different. The fluorophore and quencher are attached to the end of arm A or B so that they are in close proximity when the junction is folded. The full 3WJs are as shown above, the 2WJ is missing the unlabelled strand in each case. The 2.5WJs both use a short version of oligo b and so are directly comparable.

	T_m arm A	T_m arm B
3WJ H	47.7	45.5
2WJ	34.8	43.7
2.5WJ	34.3	46.2

Table 6-1: Melting temperatures of junctions measured from arms A and B. The three-way junctions used are shown in figure 6-9. The fluorophore and quencher are either attached to the end of arm A or arm B. This means that the melting temperature recorded may be very dependent on the melting temperature of the arm in question. The two way junction consists of the strands containing the molecular beacons, either a and c in the case of the beacons at position 1 or a and b for arm B. The third, unlabelled strand is omitted. The 2.5WJ consists of a shorter length oligo b in each case.

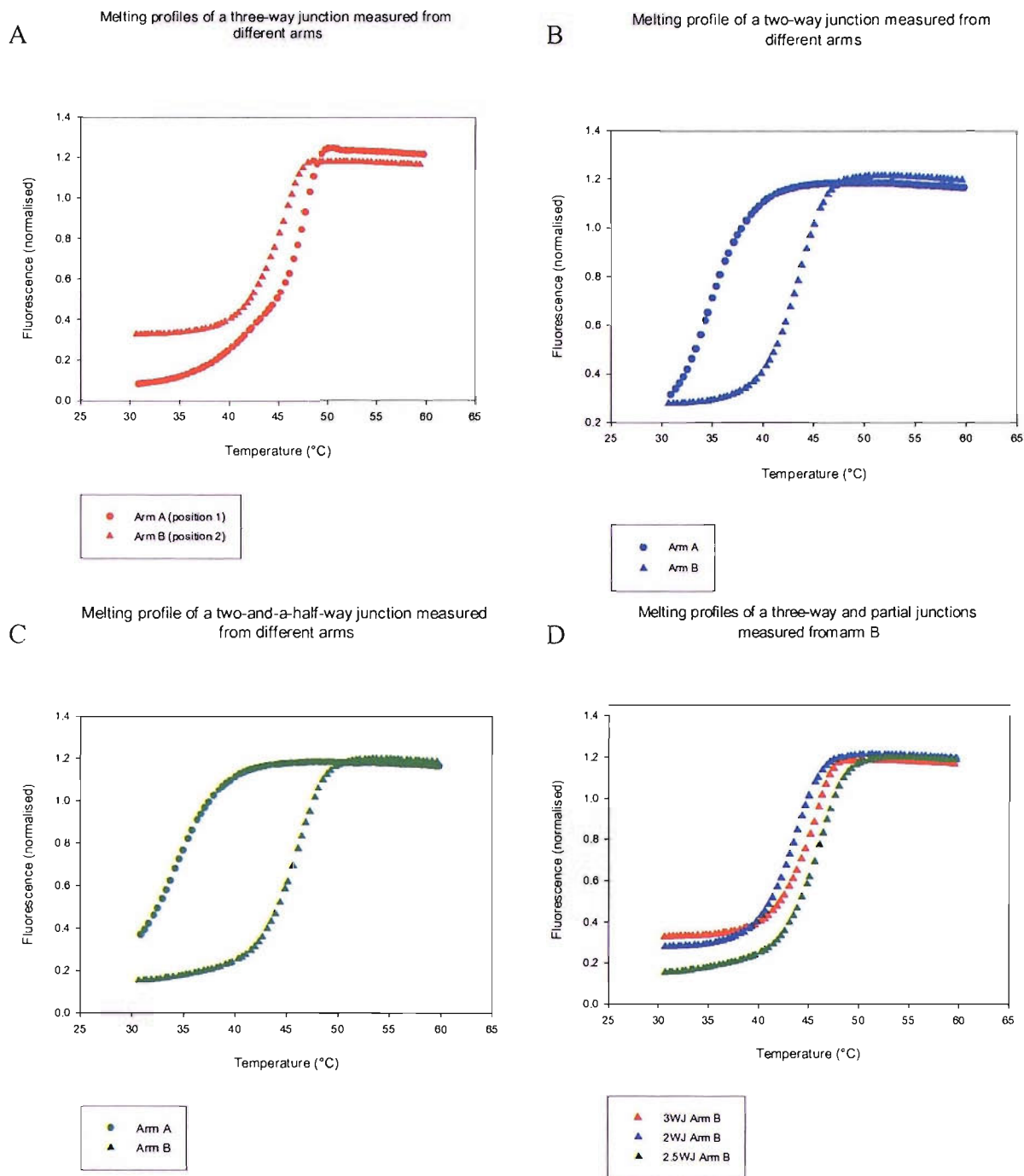


Figure 6-10: Melting profiles of junctions measured from arm A and arm B. Panel A shows a three-way junction melt measured from arm A (red circles) and arm B (red triangles). Panel B shows the same for the two-way junction melts and C shows a two-and-a-half-way junction melts. The final panel shows the three-, two- and two-and-a-half-way junctions measured from arm B only.

Figure 6-10 illustrates the differences in melting much more clearly. The three equivalent structures with the beacons on different arms are shown together in panels A-C. Panel D shows the melts of the three structures all measured from arm B. There is little to separate the three structures when the stability is measured from arm B.

6.2.5 The behaviour of a three-way junction under varying salt conditions

Previous work with both three- and four-way junctions has shown that magnesium or other divalent cations are needed to allow co-axial stacking. This is where the strands of DNA maximise their interactions, stacking right up to the branch point. The divalent cations screen the negative charges of the phosphates on the backbone of the DNA, which are concentrated at the branch point. A four-way junction will then form two co-axial helices which cross at the branch point, resembling a flattened X shape. The three-way junction containing a linker for flexibility will form an equivalent structure, which is a continuous helix with one arm extended from the branch point at an angle.

Divalent cations, such as magnesium are therefore much more effective than monovalent cations such as sodium. We compare the efficacy of sodium and magnesium at stabilising the three-way junctions at different ionic strength.

The ionic strength (I) is defined as half the sum of the concentration (c in moles) multiplied by the square of the valency (Z) for all ions. The calculations for a 1M solution of sodium chloride and 1M solution of magnesium chloride are shown on the next page.

$$I = 0.5 * \Sigma (c_i * Z_i^2)$$

Where c is the concentration of the ion in solution and Z is the valency.

So for a 1M solution of NaCl,

$$I = 0.5 * \Sigma (1*1^2) + (1*1^2) = 1$$

For a 1M solution of MgCl₂,

$$I = 0.5 * \Sigma (1*2^2) + (2*1^2) = 3$$

So this means that three times more sodium chloride is needed to give the same ionic strength as magnesium chloride. 3M NaCl is equal in ionic strength to 1M MgCl₂. The melting temperatures of 3WJ H, determined in a range of different ionic conditions are shown in table 6-2.

Ionic strength (M)	Tm		
	Magnesium chloride	Sodium chloride	Magnesium and sodium (2X ionic strength)
0	36.6	38.3	35.5
0.003	50.0	39.0	49.4
0.006	52.1	39.6	52.2
0.009	57.5	38.8	59.1
0.012	58.6	39.7	59.4
0.015	58.7	40.0	59.8
0.018	59.7	40.7	59.4
0.021	60.1	41.1	60.6
0.024	60.9	42.1	61.6
0.027	61.3	42.5	62.2
0.030	62.0	42.9	62.7

Table 6-2: Melting temperatures recorded for 3WJ H over varying magnesium and sodium chloride concentrations. In these experiments, the concentration of sodium phosphate used to buffer the solution was decreased from 50mM used in the previous reactions to 10mM. In the first column, magnesium concentration has been increased 1mM at a time leading to a final concentration of 10mM (I=0.03). Since 1mM magnesium chloride is equivalent in ionic strength to 3mM sodium chloride, the amount of sodium chloride in the reaction is increased by 3mM each time. This leads to a final concentration of 30mM. The final column shows the effect of combining the sodium and magnesium to see if the sodium has an additional effect above and beyond what the magnesium can accomplish alone. The ionic strength is therefore doubled by the combined magnesium and sodium.

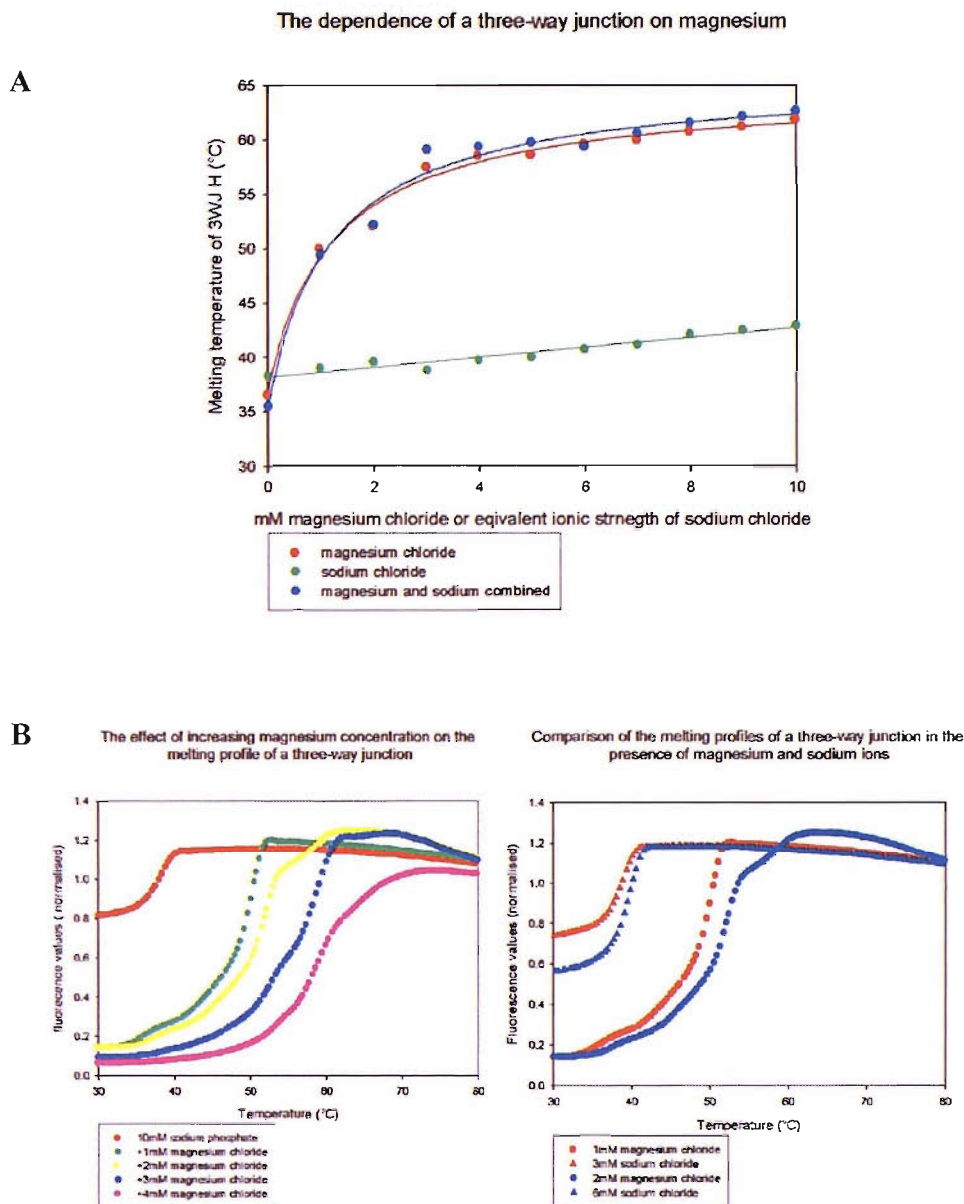


Figure 6-11: The effect of varying the concentration of sodium and magnesium on the melting temperature and melting profile of a three-way junction. Panel A shows the effect on the stability of 3WJ H of adding sodium chloride and/or magnesium chloride to 10mM sodium phosphate buffer. Approximate curves/lines as appropriate have been added to emphasise the trend. The scale along the x-axis shows the concentration of magnesium present or the equivalent ionic strength of sodium chloride (3X that of magnesium chloride).

Panel B shows how the melting profile of 3WJ H changes when the concentration of magnesium chloride is increased. Panel C highlights the difference in the melting profiles in equivalent ionic strengths of magnesium and sodium chloride. The circle points show the profiles in 1 and 2mM magnesium chloride while the triangle points show the same ionic strength (3 and 6mM) sodium chloride.

Figure 6-11 demonstrates more clearly the results from table 6-2. The first panel shows the effect on the melting temperature of increasing magnesium and/or sodium chloride concentrations.

It can be seen from Figure 6-11 that sodium chloride is not as effective as magnesium chloride at increasing the stability of the 3WJ at equal ionic strength. This suggests that there is a specific interaction with magnesium, probably at the branch point, rather than a general effect of charge screening. Adding sodium as well as magnesium has little further effect on the stability.

Panel B of figure 6-11 shows the change in melting profile of the three-way junction over increasing magnesium chloride concentration. The magnesium does have a quenching effect on the fluorophore, accounting for the differently shaped melting profiles after normalising. This does not, however, affect the melting temperature.

Panel C is included to demonstrate the difference between the equivalent ionic strengths of magnesium and sodium chloride. The profiles for 1 and 2mM magnesium chloride and 3 and 6mM sodium chloride are shown. Although 1mM magnesium and 3mM sodium are equivalent ionic strength, they have very different effects on the melting temperature and profile.

6.2.6 The effect of pH on the melting profile of the three-way junction.

A buffer containing 10mM sodium phosphate and 2mM magnesium chloride was adjusted to cover the range of pH's between 5 and 8.5. The results in table 6-3 show that the pH has little effect on the melting temperature of the three-way junction. The T_m is constant from pH5.0 – 7.0, though there is a slight decrease between pH7.5 and 8.5.

PH	T_m (°C) 3WJ H
5.0	52.9
5.5	53.0
6.0	52.6
6.5	52.8
7.0	53.8
7.5	50.9
8.0	48.4
8.5	46.9

Table 6-3: The effect of pH on the melting temperature of 3WJ H. The melting temperature recorded for 3WJ H at each of the pH values is shown above. The buffer contained 10mM sodium phosphate with different ratios of weak acid and salt to bring the solution to the appropriate pH. 2mM magnesium chloride was also present.

6.2.7 The ability of the three and two-and-a-half-way junctions to recognise mismatches

Although the SMART and SPAR junctions were designed to recognise the presence of relatively large sequences in nucleic acids (~30bases), one aim was to use them for detection of single nucleotide polymorphisms (SNPs). As mentioned in section 1.4, SNPs are largely responsible for genetic diversity and genotyping will become an important part of the diagnostics industry.

To test the ability of the 3 and 2.5WJ, representative of the SMART and SPAR junctions respectively, to recognise SNPs, mismatches were made at three locations on the target strand. Mismatch 1 was closest to the branch point with mismatches 2 and 3 increasingly further out as shown in Figures 6-12 and 6-13.

The results for melting of the 3WJ, 2.5WJ and 2WJ containing the single mismatches at different positions are shown in figure 6-14. It can be seen that the mismatches have a relatively small effect on the stability of the full 3WJ, producing a decrease in T_m of between 2-4° C. The effect is much greater for the 2.5WJ and 2WJ, which show very similar effects with T_m values of between 3-14° C. For all three types of junction, Mut2 has a greater effect than Mut3, while Mut1 causes the smallest change. This therefore suggests that a mismatch in the centre of the target foot will have a greater effect than one which is closer to either end.

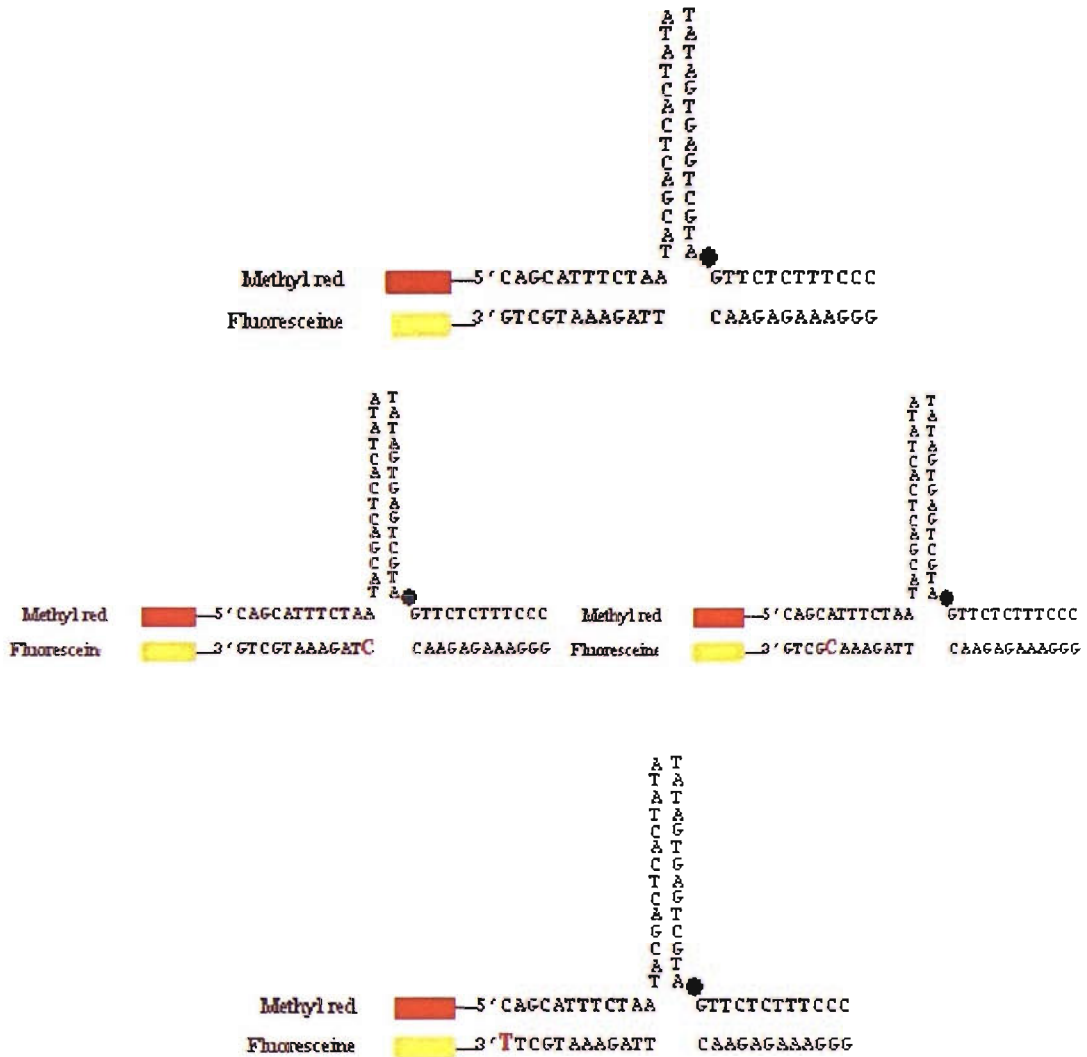


Figure 6-12: Position of the mismatches in target strands Mut (mutation) 1,2 and 3. The original, fully complementary junction is shown at the top with the variations on the target strand producing mismatches in the three panels underneath. The base that has been changed is shown highlighted in red.

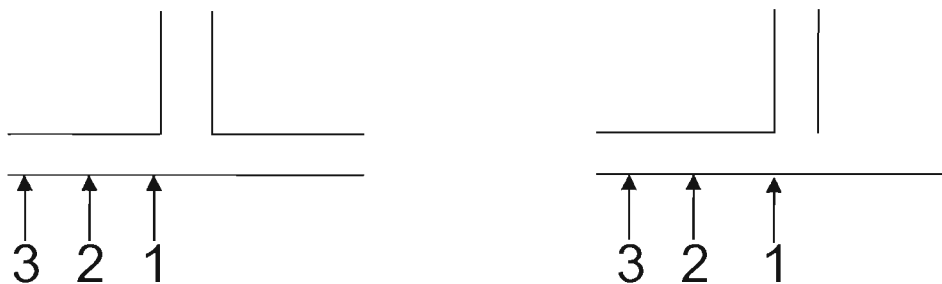


Figure 6-13: Position of the mismatches in the target strand. The simplified diagram above shows the position of the mismatches in the target strand in relation to the 3 and 2.5WJ. There is only ever a single mismatch on each strand.

A

Junction	Melting temp.			
	Normal	Mut1	Mut2	Mut3
3WJ H	59.8	57.5	55.2	56.6
2.5WJ	51.9	48.3	37.8	41
2WJ	52.3	49.1	38.8	41.7

B

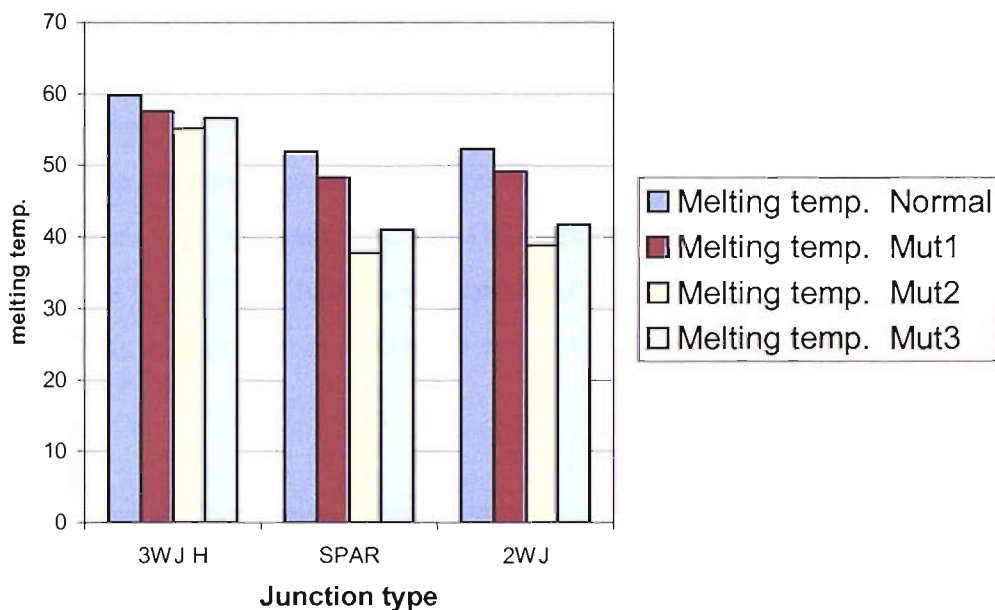


Figure 6-14: Melting temperatures of the 3WJ containing the hexaethylene glycol, the 2.5 and 2WJ. The mismatches represent potential SNPs in a target sequence. Mutations 1-3 are in different positions along the target as shown in figure 6-12- and 6-13. The melting temperature (°C) was recorded for each junction with each different target strand, giving the data in the table and in the graph.

6.3 SYBR Green binding to DNA

SYBR Green (Molecular probes) is a DNA binding dye which fluoresces upon binding to double stranded DNA. It has been used as a lower toxicity alternative to ethidium bromide staining of gels. It can be also used in real-time PCR to measure the quantity of amplicon produced throughout the reaction.

The mechanism of binding is thought to be two-stage. At low ratios of dye to base pairs, the dye intercalates with the DNA. At high ratios of dye to base pairs, SYBR Green also shows minor groove binding (Zipper *et al.*, 2004). SYBR Green also shows some sequence specificity characteristic of minor groove binding ligands. It binds better to AT sequences than to GC (Zipper *et al.*, 2004; Giglio *et al.*, 2003). This means that SYBR Green is not completely sequence independent, occasionally making this a problem in multiplex reactions but not for looking at individual samples.

The purpose of this section of work was to investigate whether SYBR Green is a useful tool for looking at the melting profiles of three-way junctions. Attaching fluorescent probes to the end of every junction to be studied is costly and so cannot be used to analyse high numbers of junctions. SYBR Green as a global DNA binding dye could be used to analyse any junction produced, on a high throughput scale.

A selection of non-labelled junctions (all previously analysed in this thesis), were melted and annealed along with SYBR Green. The profiles of the junctions and melting temperatures were compared. Since each of the junctions have been studied using other methods, the SYBR Green melts were intended as a companion to this information, to give a fuller picture.

6.3.1 Melting profile of a three-way junction

Junction MecA has been the base of much of the work in this thesis. It has been compared to Coag, a similar but much more stable junction than MecA, the latter containing a larger proportion of AT bases. In chapter 3, to determine whether the results of enzymatic cleavage using DNaseI were due to the three-way junction structure specifically, a complement was made for each of the three strands in the junction. This allowed three new duplexes to be made, each containing one of the original strands of the junction. Each of the duplexes was as close as possible in sequence to the original three-way junction.

Figure 6-14 compares the melting of three-way junction MecA with the duplex containing the target strand (Mta duplex), and also with three-way junction Coag. Shown in the diagram are the melting profiles, melting peaks and derived melting temperatures.

There are small differences between the 3WJ and duplex melting profiles. There is a suggestion of partial duplex structures being formed in the 3WJ. This is manifest in the slight second hump visible in the melting peak data, suggesting the melting temperature of a different structure. The most likely cause for this is a slight excess of some of the strands of DNA, leading to an incomplete 3WJ. In SYBR Green melting, all double stranded structures are seen. In the previous section, the fluorescence came only from the fluorophore labelled strand. By having an excess of the quencher and non-labelled oligos, virtually all the fluorophore-labelled strand was bound in junction form. An alternative explanation is that the 3WJ undergoes a two-stage melting process. This could occur if one arm was much more stable than the other two, maintaining a partial duplex after the third strand had melted. This appeared to be the case with the specific LightCycler junction, shown when the fluorophore and quencher were attached to arm B.

The actual melting temperatures recorded for MecA and the duplex are not as different as expected, although there is still a significant gap between them. The 3WJ (MecA) melted at 63.5°C while the duplex melted at 69°C. The difference is

much smaller than that between the 3WJ and duplex measured using the molecular beacons (the duplex melted at 61°C while the equivalent 3WJ melted at 48°C). The Mta duplex used here is longer than the specific LightCycler version (50 base pairs compared to 26). This accounts for its higher melting temperature under the same buffer conditions but not for the small difference between the 3WJ and duplex.

Figure 6-14 also shows the melting profiles of MecA compared to Coag. While overall the shapes of the profiles are similar, the starting fluorescence differs by just over three points. In the introduction to this section, it was noted that SYBR Green has a preference for AT rich sequences. MecA fulfils this criterion, and hence shows a high fluorescence due to more SYBR Green bound. This should not, as far as is known, affect the melting temperature recorded. As the junctions melt apart and the SYBR Green is dispersed, there is a sharp drop in fluorescence followed by a plateau, showing no more SYBR Green bound. Melting at 68°C, the Coag junction appears almost as stable as the MTA duplex. It is more stable than MecA by 4.5°C.

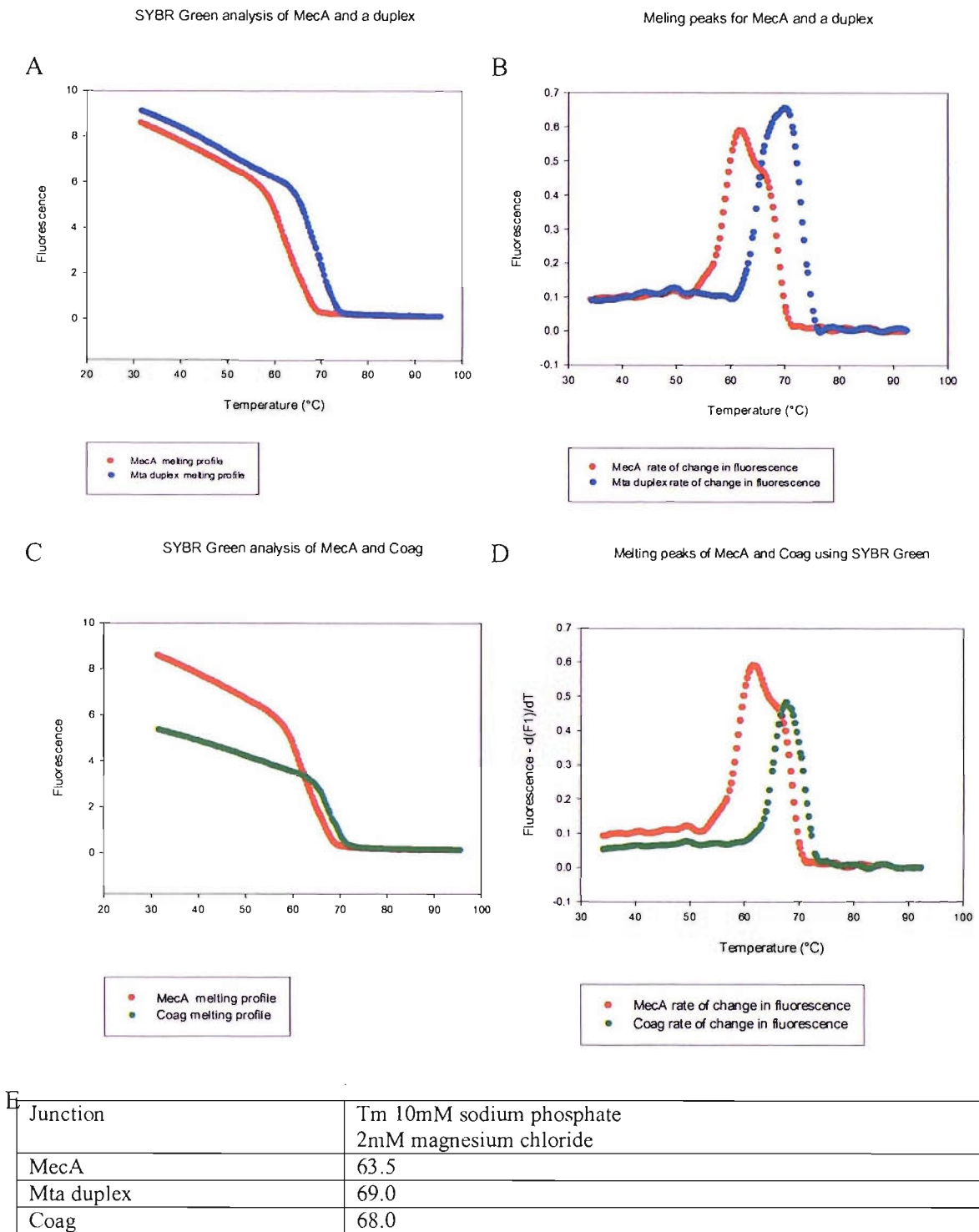


Figure 6-14: Melting profiles, peaks and melting temperatures of MecA, a duplex and Coag measured using SYBR Green. Panels A and B show the melting of a 3WJ and a duplex of as close a sequence as possible. The graphs have not been normalised to demonstrate any differences in signal strength due to preferential binding by SYBR Green. Panels C and D show the melting profile and peaks for 3WJs MecA and Coag. The melting temperatures, calculated from the peak data are shown in panel E underneath. The buffer used in this case was 50mM sodium phosphate and 2mM magnesium.

6.3.2 Comparison of the melting of different junctions

In chapter 4, junctions were created with extra linkers around the branch point to move the template arm away from the branch point. This allowed better extension and transcription. Of these junctions two, 1:1 and 3:3 were created as full-length versions to use for gel mobility and chemical probing studies. These junctions contain either one or three octanediols respectively on either side of arm B. Although chemical probing showed that they were fully base paired up to the branch point, they did show a high level of flexibility.

The melting profiles of junctions 1:1 and 3:3 are almost identical. Figure 6-16 panel A shows them alongside MecA. It can be seen that whilst MecA is similar to the other two in profile, the actual melting temperatures in Figure 6-17, show that MecA is more stable by a couple of degrees.

Panel B in Figure 6-16 shows MecA compared to two junctions of similar sequence. The sequence at the branch point has been changed in each case. These junctions were used to try to change the preferred conformer of MecA in chapter 3. There is some difference in signal strength, MecAV2 showing preferential binding of SYBR Green. After the initial difference in signal strength, V1 overlaps the profile for V2 and they melt at the same temperature, 3°C higher than MecA and almost on a par with the duplex.

The melting profiles in panel C show the effect of putting two mismatches at the branch point. Junctions TA-GG and TA-CT are almost identical except for two base pairs adjacent to the branch point (Chapter 5). The strands of DNA used for each junction can be switched to generate two mismatched junctions. Although the profiles appear slightly different due to the initial fluorescence, the actual melting temperatures are very similar, all within a couple of degrees. Junction TA-GG containing the mismatches melts 2 degrees lower than the fully matched junction, while junction TA-CT seems unaffected by the presence of a mismatch. The average T_m of the collection of redundant junctions (R-TA) was 69°C.

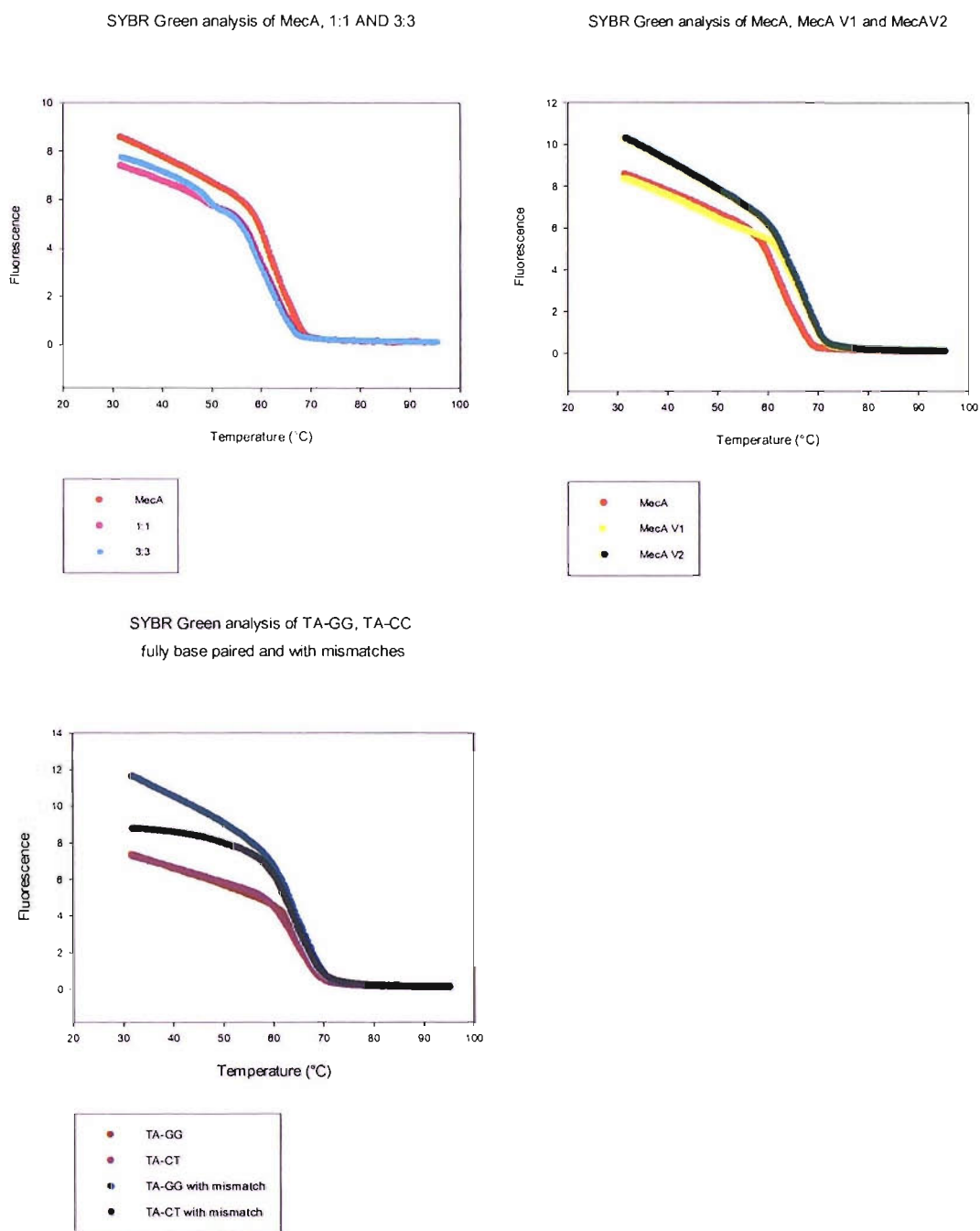


Figure 6-16: Melting profiles of related sets of junctions using SYBR Green. Panel A shows junction MecA compared to 1:1 and 3:3. Panel B shows MecA again but this time compared to MecA V1 and MecA V2. Panel C looks at a different set of junctions which came out of the folding prediction work in chapter 5. Junctions TA-GG and TA-CT differ only by 2 bases at the branch point on arms A and C. Strand c of each of them is switched to create junctions with mismatches around the branch point. All junctions are compared in 50mM sodium phosphate plus 2mM magnesium. The melting temperatures for these junctions are given in Table 6-5.

Junction	50mM sodium phosphate plus 2mM Magnesium chloride	50mM sodium phosphate only
MecA	63.5	49.7
Coag	68.0	56.7
1:1	61.5	48.5
3:3	61.4	48.0
Mta duplex	69.0	55.8
Mte duplex	66.7	58.1
Mex duplex	63.6	67.4
MecAV1	66.7	59.6
MecAV2	66.3	54.0
TA-GG	66.8	53.1
TA-CT	64.4	54.5
GG with mismatch	65.2	52.2
CT with mismatch	64.5	53.7
RTA	69.0	-

Melting temperatures of various DNA structures measured using SYBR Green

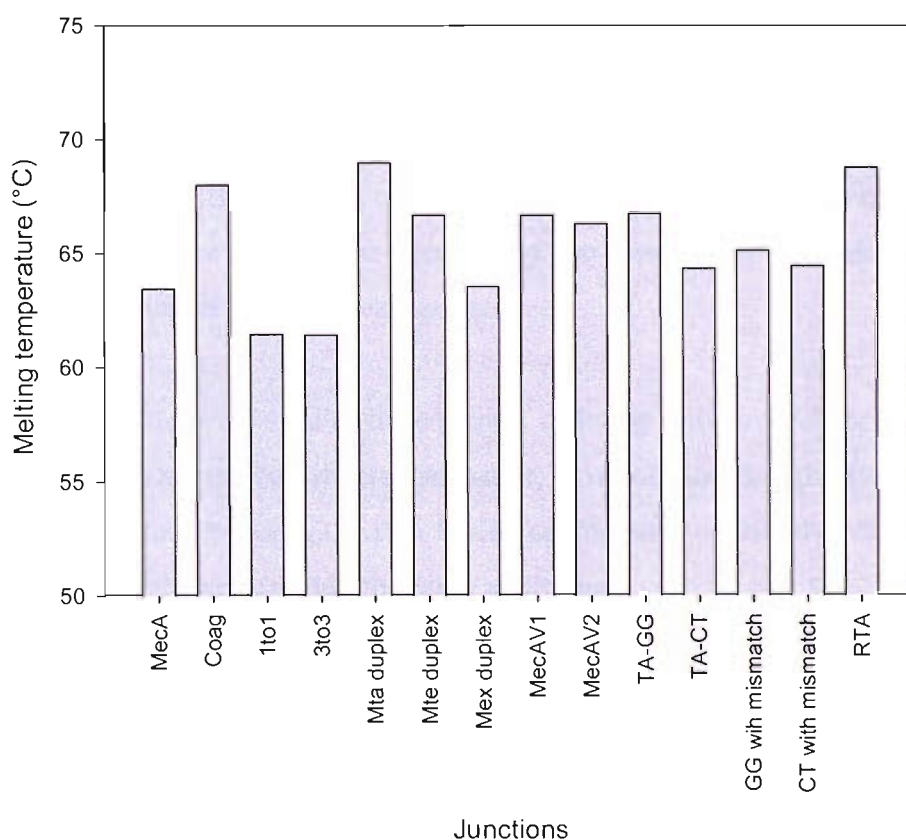


Figure 6-17: Melting temperatures of junctions in 50mM sodium phosphate plus and minus magnesium. The melting temperatures (°C) are displayed against each junction for a buffer containing 50mM sodium phosphate plus and minus 2mM magnesium chloride. The data recorded from the buffer containing magnesium is shown graphically in the bar chart in B. The melting temperature is displayed on the Y axis against each individual junction on the X.

6.4 Fluorescence Resonance Energy Transfer

Fluorescence Resonance Energy Transfer (FRET) involves the use the transfer of energy between donor and acceptor pairs to measure the distance between molecules or parts of the same molecule. The emission spectrum of the donor overlaps with the absorption spectrum of the acceptor. A donor in the excited state releases its energy to be absorbed by the acceptor. The energy is finally released as fluorescence. The efficiency of the energy transfer is inversely proportional to the 6th power of the donor-acceptor distance. The fact that each small change in distance is magnified by the change in the efficiency of transfer makes FRET quite a sensitive method. It is best used when the distance in question is between 1-10nm (for a recent review of FRET see Jares-Erijman and Jovin, 2003).

Sets of oligos comprising two three-way junctions with FRET donor and acceptor pairs on the arms had been made for a previous set of experiments. The oligos could be combined to give three variations on each junction, such that a donor and acceptor were on different arms of the junction. The three variations covered every combination (Figure 6-18). The identity of the donor was fluorescein (used in section 6.1) while the acceptor was tamra.

The two junctions were similar in sequence, differing only around the branch point (Figure 6-19). Due to the different sequence, it is possible that the two form quite different equilibria between the two folded conformers, or that they form different angles between the arms within the same conformer.

The LightCycler was used to give an indication but not an accurate measure of the angles between the arms of the junction. These experiments were used as a test to determine whether the LightCycler could provide useful information about the folding of the junction from the point of view of the separation of the arms. It must be noted that the junctions are in equilibrium. If each junction was measured at the same moment, we would see a range of values, probably leading to a biased distribution. The signal strength recorded from the sum of those junctions would give an average value. Within this, we may only be able to tell if the arms are close together or far apart. This should be enough to gauge the average conformation.

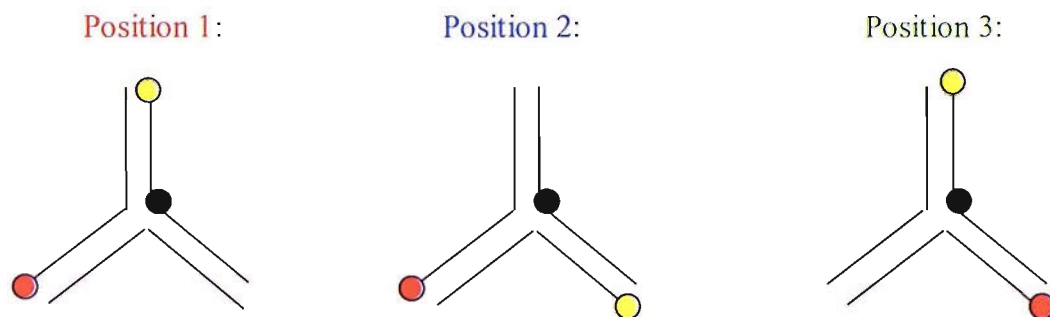


Figure 6-18: Positions of the donor and acceptor molecules on the three-way junctions. The three variations of the same junction are colour coded as positions 1-3 in red, blue and green respectively. These correspond to the colours used on the graphs later in the chapter. The donor and acceptor are shown as circles on the end of the relevant strand to represent the covalent attachment of the molecules. Fluoresceine is shown in yellow while tamra is red. The position 1 junction contains the donor on arm B and the acceptor on arm A. Position 2 retains the acceptor on arm A but has the donor on arm C. Finally, position 3 uses the donor on arm B and the acceptor on arm C. In each case the distance between the donor and acceptor molecules, and hence the relevant arms, can be measured using FRET.

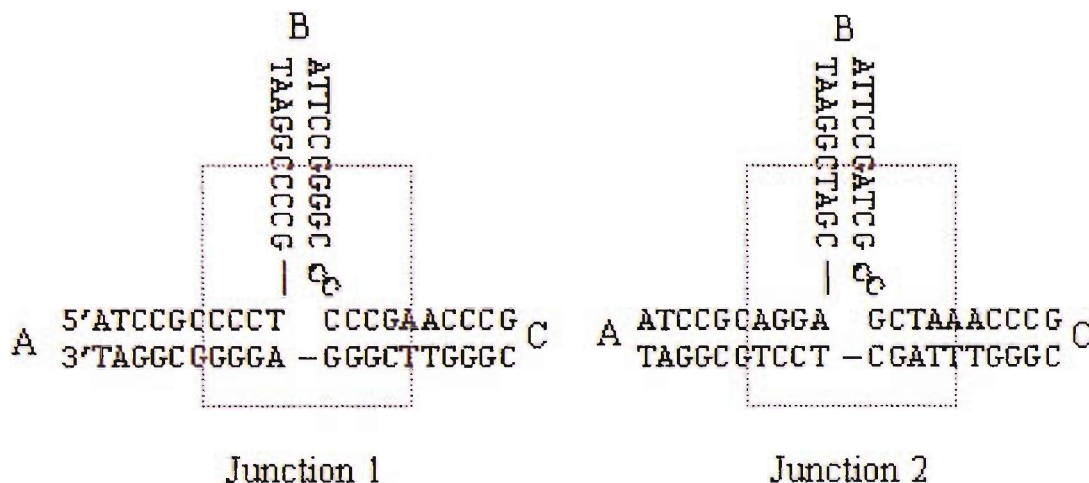


Figure 6-19: Sequences of the two FRET junctions. Junctions 1 and 2 are identical in sequence except for the central four base pairs on each arm (highlighted in the box). Both junctions use two unpaired cytosines as the linker. For each junction there is a set of oligos with donor and acceptor molecules attached to different strands. They are combined as above to allow the distance between each pair of arms to be measured.

6.4.1 Analysing the melting curves for the positions of the arms relative to each other.

The emission spectra of the donor and acceptor are measured on separate channels. Fluorescein is excited at 494nm and emits at 520nm (detected on channel 1). Tamra absorbs optimally at 544nm but overlaps with fluorescein's 520nm emission enough to be excited by energy transfer. It emits at 576nm and this is recorded on detection channel 2.

The two junctions have been tested using a buffer with and without magnesium to identify changes in the relative positions of the arms under different salt conditions. The strategy used previously was to use as low a concentration salt buffer as possible to allow better comparison between junctions. There is a clearer difference between junctions of different stability in a lower salt buffer than in a higher salt buffer. In this case, 50mM sodium phosphate was not sufficient to stabilise the junctions enough for the full melting profile to be seen. Using 100mM sodium phosphate just brought the profile back within the temperature range used. The melting profile for junction 1 and junction 2 are shown on channels 1 and 2, plus and minus magnesium. Each graph shows positions 1-3 together colour coded to match Figure 6-18.

The advantage of being able to measure the emission from both fluorescein as the donor, and tamra as the acceptor, is that the transfer process can be seen from two different angles. When the donor and acceptor molecules are close to each other, the fluorescence will be low on channel 1 and high on channel 2, because the resonance energy transfer will be good. When they are far apart, the signal will be high on channel 1 and low on channel 2. During the melting, if the two are close, there should be a transition of the fluorescence signal from channel 2 to channel 1 as the transfer efficiency lessens. Where the distance between the two was great to start with there should be a flat line on both graphs.

Figure 6-20 shows the results for junction 1 on channel 1 and 2, plus and minus magnesium. By comparing the transition of energy between channels 1 and 2 during melting, it can be seen that, when no magnesium is present, there is a clear lessening

of the energy transfer between fluorescein and tamra at position 1 as the junction melts. When the FRET pair is placed at position 2, there is a weak transfer of energy as the junction melts. Position 3 shows what appears to be a melt on channel 1 but nothing on channel 2. This suggests the change in signal on channel 1 may be due to self-quenching. Overall, the strongest FRET signal is between the donor and acceptor at position 1. This may indicate an open structure, with the bulging of the linker causing angle AB to be slightly smaller than the other two.

When magnesium is present in the buffer and the junction can fold, the strength of signal at position 1 increases across both channels. There is some signal from position 2 but no true FRET signal at position 3 (the melt is only seen on channel 2). This would point to angle AB being the smallest again. If AB is the exchanging strand then AC must be the continuous strand, giving a predominant conformer A/C.

The results of the fluorescence melting of junction 2 are shown in Figure 6-21. It can be seen that the overall pattern of melting appears very different from junction 1. The lessening of the transfer of energy between fluorescein and tamra shows a much sharper transition as the junction melts. Position 1 shows a signal on both channels but the transfer of energy is not very strong. Position 2 on the other hand, shows a very strong FRET signal. Position 3 again shows a melt on channel 1 but not on channel 2, suggesting that it is not a true FRET signal. If the donor and acceptor are closest in position 2, then arms A and B would be continuous and the junction would predominantly adopt conformer A/B.

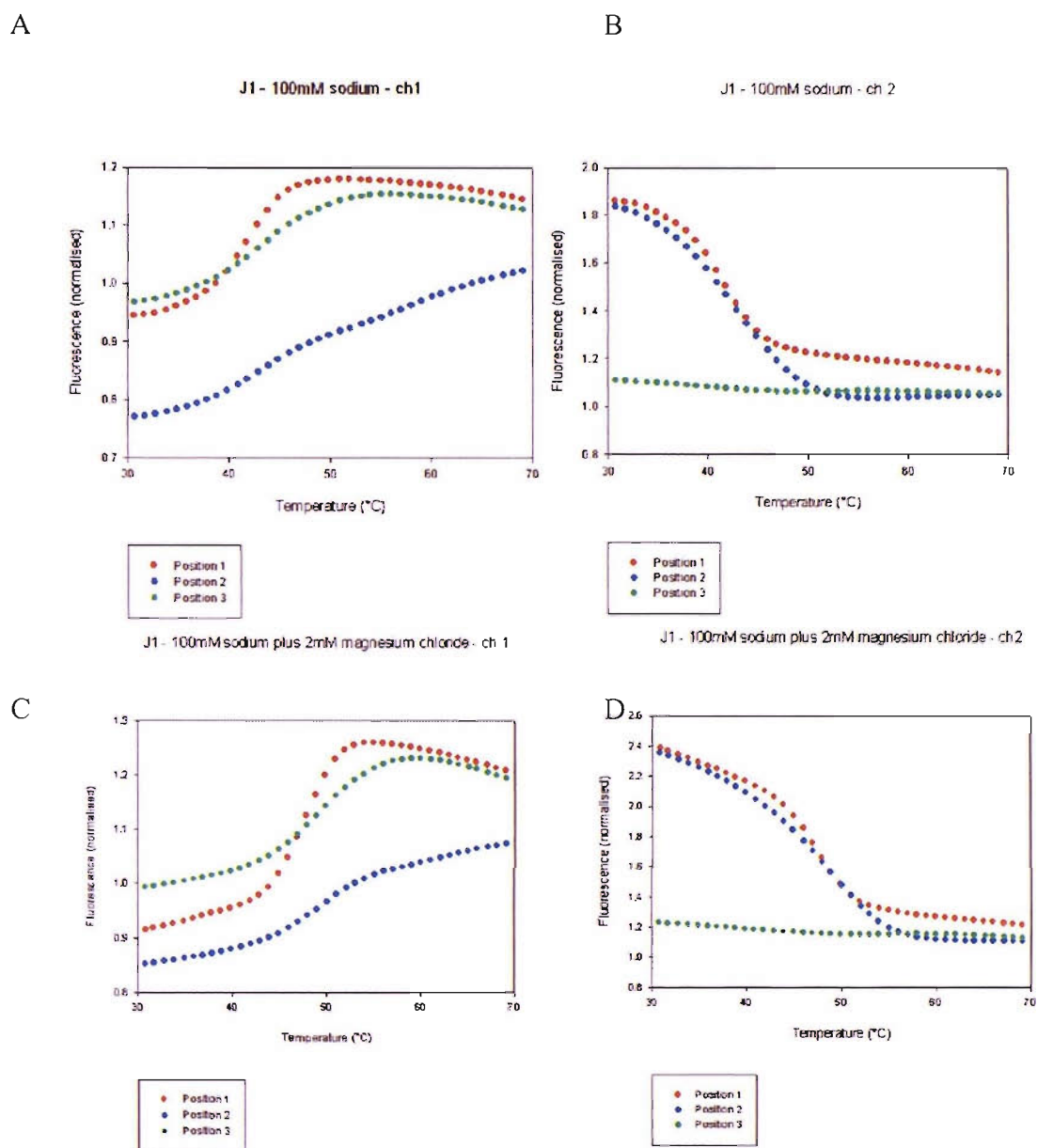


Figure 6-20: Melting profiles of junction 1 plus and minus magnesium. Panels A and B show the data obtained from junction 1 (J1) in a 100mM sodium phosphate buffer. A is measured on channel 1, whereas B is the same reaction measured on channel 2. The channels pick up the emission from the donor and acceptor respectively. Panels C and D show the data obtained from J1 in a 50mM sodium phosphate buffer containing 2mM magnesium to aid folding. In each graph the positions 1-3 refer to the location of the donor and acceptor laid out in Figure 6-18. The colours used for the different positions match those in Figure 6-18.

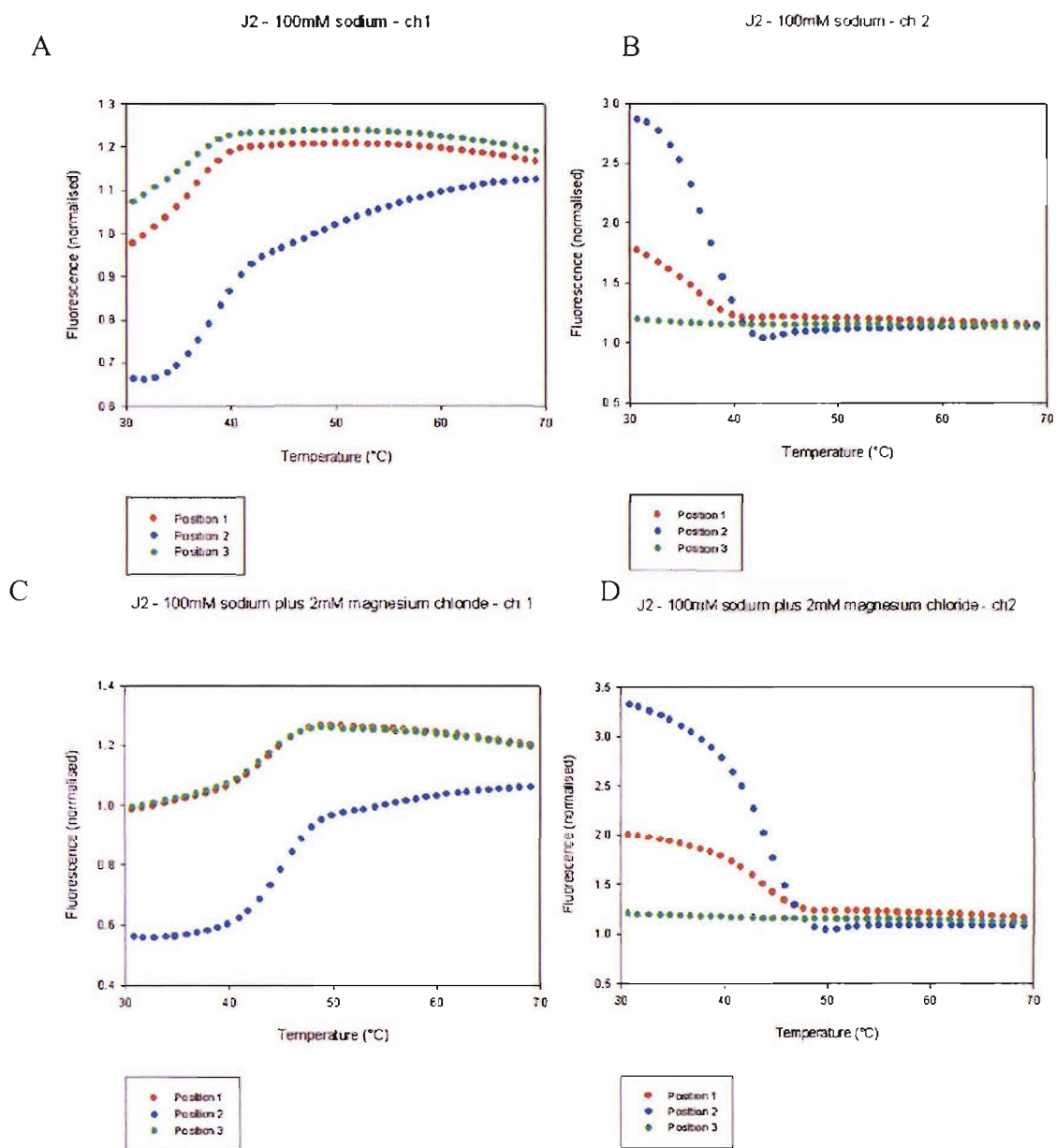


Figure 6-21: Melting profiles of junction 2 plus and minus magnesium. Panels A and C show junction 1 in 100mM sodium phosphate while panels C and D show junction 2 in 100mM sodium phosphate plus 2mM magnesium. The positions 1-3 refer to exactly the same arms as for junction 1. The sequence of the two junctions is different for the 4 bases on each side of the branch point (Figure 6-19).

Table 6-1 shows the actual melting temperatures recorded for the junctions plus and minus magnesium. The temperatures have been recorded at each position and on each channel to give an idea of the consistency of the data. An average temperature is given at the end of each line for better comparison. Overall, the melting temperatures are remarkably consistent across the different channels and positions. For junction 1 in 100mM sodium phosphate, there is less than a degree difference between the average temperature on channel 1 and channel 2. This confirms the accuracy of signal transfer data obtained.

	Position 1	Position 2	Position 3	Average T _m
J1 (100mM sodium) ch 1	42.1	43.2	43.7	43.0
J1 (100mM sodium) ch 2	41.5	42.7		42.1
J1 (100mM sodium plus 2mM magnesium) ch 1	48.0	49.8	49.1	49.0
J1 (100mM sodium plus 2mM magnesium) ch 2	47.6	49.3	-	48.5
J2 (100mM sodium) ch 1	37.0	38.5	36.5	37.3
J2 (100mM sodium) ch 2	36.9	38.0		37.5
J2 (100mM sodium plus 2mM magnesium) ch 1	44.2	45.0	43.6	44.3
J2 (100mM sodium plus 2mM magnesium) ch 2	43.9	44.6	-	44.3

Table 6-1: Comparison of melting temperatures for J1 and J2 plus and minus magnesium. T_m has been measured from each position on channel 1 and channel 2. An average temperature is included at the end of each set of conditions to compare data between the different channels. Where there is a blank space left in the chart it was because no signal was recorded to give a melting temperature.

6.5 Discussion

In section 1, the melting of the DNA junctions was followed by the change in fluorescence when a quencher and fluorophore on opposite strands of the same arm dissociated. The temperature at which this happened was taken as a measure of the stability of the junction as a whole. When the fluorophore and quencher were placed on arm A, this method showed a large difference between the stability of a 3WJ and duplex of similar sequence (T_m of 48°C compared to 61°C respectively). When partial junctions were used, it registered that the loss of the third, unlabelled strand (giving a 2WJ), lessened the stability substantially. The addition of accessory strands did not then improve the stability any as measured from arm A. This was probably because the third full strand held the other two together for longer. It was not necessarily a factor of how many base pairs there were in the junction.

When the fluorophore and quencher were moved to arm B, few differences were seen between the 3WJ and the partial junctions. There were differences in T_m of just a few degrees between the full 3WJ, the 2WJ and 2.5WJ. This suggested that the increase in the length of arm B and increase in GC content compared to arm A, gave the arm an increased stability. It may well have been the last part of the junction to melt. As such the loss of the third unlabelled strands and the addition of accessory strands made no difference to its melting. This affirms the result obtained using arm A, suggesting that they did show junction stability as a whole.

The mismatch work was very valuable in demonstrating that a 2.5WJ recognises mismatches well. Not only that, it demonstrated that, in every case, the best place for mismatch recognition was towards the middle edge of the target region, that is half way along one arm (Mut 2). The point where the mismatch was recognised least was in the centre of the target, underneath the branch point. As has been seen using unpaired bases or chemical linkers at the branch point, the loop created from a mismatch introduces flexibility and can be beneficial to folding and the stability of the junction, even if these appear in more than one strand (1:1 and 3:3 junctions in chapter 4).

In section 2 of this chapter, SYBR Green I was tested for its ability to follow the melting of 3WJs. The gaps between the melting of the junction and duplex were much smaller than they had been in section 1. It should be noted that the sequences used here were the longer junctions designed for the long-short arm assay, rather than the short junctions used specifically for the LightCycler work in part 1. The arms of MecA were 25 base pairs each compared to 12 base pairs in the LC junction. Despite the extra stability conferred by a doubling of the number of base pairs, there should still have been a large difference in the stability of MecA and a duplex of similar sequence. The duplex would be expected to be a lot more stable than the 3WJ, or put the other way round, the 3WJ should be much less stable than it is shown to be by the melting temperature. It is possible that when SYBR Green binds to the DNA, it stabilises the double stranded form. If so, this would have a large effect on the 3WJ but perhaps less so on the already stable duplex.

Junctions 1:1 and 3:3, containing the extra octanediols at the branch point, melt 2°C lower than MecA. As a more flexible junction with a large bulge at the branch point, 1:1 and 3:3 may be expected to be slightly less stable than their more rigid parent junction. The difference may be just a few degrees, but since there is only 4°C difference between MecA and the duplex, this may be more significant than it initially appears.

The redundant junction work in chapter 5 generated two junctions whose sequence varied only by two bases at the branch point. In one junction, TA-GG, the mismatches were reflected in the stability of the resulting junction, whereas in the CT junction, the mismatched junction showed no difference in stability to its parent. The work on mismatch detection from part 1 did show that single mismatches were least likely to be recognised at the branch point. This is, perhaps, a note of caution for the redundant junction assay.

Overall, the junctions and duplexes analysed by SYBR Green showed a high melting temperature and varied across a range of 10°C (between 60-70°C). When comparing these results to those obtained using the molecular beacons in part 1, it seems unlikely that these junctions would melt at the temperatures recorded without

some external stabilising factor. Since the buffer in the two experiments was identical, it suggests that it may be the binding of SYBR Green stabilising the DNA. This would not cause a problem in the normal application of SYBR Green. In a single PCR reaction, the production of a single amplicon is followed. The more of it is produced, the more fluorescence is seen in the reaction. In multiplex reactions, as long as the two amplicons differ in melting temperature, the production of both can usually be followed. Based on the research presented here, this technique is of limited use for following and comparing melting profiles of DNA structures. Although the technique has generated useful information about the relative stability of the 3WJ's here, it cannot be certain that, due to its slight preference for AT rich sequences, it will not bind to and stabilise certain sequences over others. This would distort the data.

The data obtained in part 3 from FRET donor and acceptor pairs did show definite differences between the two junctions studied. They did suggest that the junctions adopted different dominant conformations. Even without being able to measure the precise angles between the arms, this technique did have some usefulness. The next stage would be to test these junctions in the long-short arm assay to confirm the conformer preference. It would be ideal to use FRET to measure the angles between the arms in a number of junctions. This may confirm the range of angles adopted by the junctions within a stacked conformer.

It seems odd that there is never a signal obtained on channel 2 for position 3. This occurs in both junctions so would appear not to be a function of an individual oligo. The fact that there was no clear signal from this position may suggest that the largest angle was between arms B and C. This would mean that the linker was in the centre of the continuous strand, a structure which has not been reported so far. It seems unlikely that this would form unless there was some slippage of the sequence causing the unpaired bases to loop out elsewhere in the junction. Looking at the sequence, this is possible in junction 1 but not in junction 2.

There was also a difference in melting temperature between the two junctions, despite only a small change in sequence. This may be a factor of the stability of the junction due to different folding arrangements.

Chapter 7

DISCUSSION

7 Discussion

The aim of this thesis was to explore the folding of three-way DNA junctions in relation to sequence and to a novel DNA detection assay (SMART). A number of different techniques have been used to do this. Chapter 3 covered the use of gel mobility assays i.e. the long-short arm assay, to identify the dominant conformer adopted by a stacked junction. The research presented there gave a measure of the complication provided by sequence further away from the branch point. On one hand, the MecA and Coag junctions (which shared sequence up to 2 base pairs away from the branch point in each direction), adopted the same dominant conformer. On the other hand, changing the central sequence of MecA with that of a junction reported to adopt a conformer A/C, was not sufficient to cause MecAV1 to alter conformation. Chemical probing of the MecA V1 and V2 junctions did reveal bases which were exposed around the branch point. Since the folded dominant conformer did not appear to account for these bases being looped out, it is possible that they were a result of the junction spending a significant portion of its time unstacked.

Chapter 4 looked in more detail at the SMART assay and the variations in the efficiency of the extension step. The preferred conformer of a sequence probably has quite a large effect on the performance in the SMART assay. When binding to genomic DNA or ribosomal RNA, the probes and target may be forced to adopt an A/C conformation regardless of the conformer preference. This would place the target at the bottom and arm B, containing the extension probe, at an acute angle with the target. This would make access by DNA and RNA polymerase difficult. If the longer target sequence could be bent enough to accommodate the A/B conformer, the extension probe would lie at the bottom left and would be much more accessible.

Conformer preference is dictated by high and low energy situations. The three-way junction is able to adopt both conformers and move between the two through the unstacked state. It will spend most of its time in the lowest energy state available. If the difference between the states is small, the junction may be seen to be in equilibrium when assayed over a long time period. Within this, there was evidence in this thesis of varying arm angles (MecAV1 and V2 in chapter 3). Two junctions of different sequence may adopt the same dominant conformer but have slightly different angles between the arms. Watson *et al.*, (2004) have defined a number of parameters to better characterise the variation between junctions and compare junctions over high and low resolution techniques. The existing term, Inter-duplex angle (IDA) is defined as the positive (right handed) rotation of the back duplex relative to a static duplex held in front. The IDA has been estimated at 60° using low resolution techniques such as gel mobility, AFM and FRET. This angle can be divided into the factors of J slide, J roll and J twist.

J slide: “Translation of the duplexes along their respective helix axes either toward or away from the junction crossover”.

J roll: “...rotation of the two stacked duplexes about their respective helix axes, leading to a change in accessibility”.

J twist: “...the angle of the stacked duplex arms in the stacked X-form of the junction”. (Watson *et al.*, 2004)

Interduplex angles in stacked X junctions range from –80 to 55 degrees in crystal structures (Ho and Eichman, 2001). The sequence throughout a junction, not just at the branch point will control these parameters and will define the conformation adopted. Depending on the difference between the high and low energy states of the junction, different rates of flipping are seen (McKinney *et al.*, 2003).

The presence of the linker is of great importance to 3WJs. Leontis *et al.*, (1991) first demonstrated that adding between two to five unpaired bases into one strand at the branch point, stabilised the 3WJ and allowed it to fold in the presence of divalent metal ions, mimicking the 4WJ. To achieve enough flexibility in the backbone, there must be at least two unpaired bases. The stabilising effect provided by these bases is not universal. Kadramas *et al* (1995) showed that while unpaired bases may stabilise

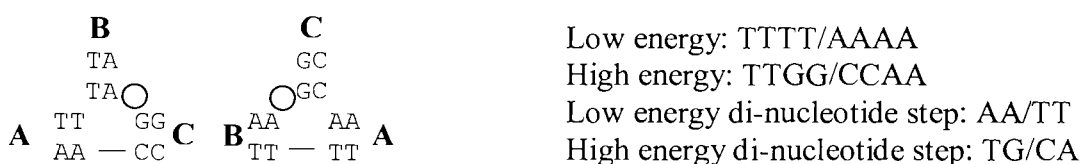
a 3 and 5WJ, they destabilise a 4WJ. It appears that if there are an odd number of helices, the lengthening of the backbone allows pair wise stacking. Adding bases to junctions with an even number of helices gives too much flexibility and disrupts pair wise stacking. A number of experiments presented in this thesis have used chemical linkers in the place of unpaired bases. The advantage of using these has been that there is no danger of mispairing occurring at the branch point between similar sequences. The melting studies using the quencher and fluorophore in chapter 6 part 1, demonstrated that a single hexaethylene glycol actually gave slightly more stability to a junction of the same sequence than two unpaired bases could. The addition of a second hexaethylene glycol, however, destabilised the junction. The difference between the stabilities of the junctions was only a few degrees difference in melting temperature, but was significant none the less. An unpaired base adds one nucleotide unit to the backbone and contributes six extra bonds in length. Two extra bases will contribute a total of twelve extra bonds. The hexaethylene glycol, $(C_3H_6)_6$, adds eighteen extra bond lengths to the backbone. The addition of a second hexaethylene glycol would give thirty-six extra bond length. It seems this flexibility may be slightly too much for the junction to accommodate stably.

The SMART junctions all used an octanediol, $O-C_8H_{16}-O$, as their chemical linker. This has a similar effect to the hexaethylene glycol and unpaired bases in stabilising the junction. It contains nine extra bonds to integrate into the backbone. From the long-short arm assay results of MecA, Coag, MecAV1 and V2, it would appear that the octanediol is sufficient to allow folding. Adding extra octanediols to two sides of the branch point produced some interesting results. The 1:1, 2:2, 3:3 and 4:4 junctions were synthesised with equal numbers of octanediols in strands a and b at the branch point. These were full SMART junctions, and the results from the assay showed that the 3:3 junction gave the best signal strength. 1:1 and 2:2 were an improvement on MecA but the 4:4 was not and gave the worst signal strength. When the 1:1 and 3:3 junctions were synthesised as long-short arm assay versions (arms of 25 base pairs with restriction sites incorporated into them), the folding of the junctions was shown to be quite different from MecA. The 1:1 junction, while fully base paired, did not appear to have the co-axial folding expected but had more equal angles between the arms. These results suggested that there was a kink in the continuous strand, formed by the octanediol on that side of the junction. Overall,

when analysed, it still adopted the A/B conformer of its parent junction. The 3:3 on the other had, swapped conformer to a dominant A/C. This was despite no change in sequence. It would appear that the bulge introduced by the linkers can have an effect on the angles in the arms and ultimately change the high and low energy conformations dictated by the sequence.

The ability of the 3WJ to recognise mismatches was tested by looking for changes in stability in the melting studies (Chapter 6 part 1). These experiments demonstrated that the 3WJ is not very good at recognising single base mismatches and so, unfortunately, would not be very good for SNP detection. The 2.5WJ, originally designed for SPAR, is much more sensitive to mismatches. The best place along the target for recognition was the centre of the arm of the junction, rather than at the branch point. While the SPAR assay and the 2.5WJ have fallen by the wayside in favour of SMART, it may be that this will prove useful in SNP detection.

To date, a number of different junctions have been studied. The sequences of all the 3WJs and their conformational preference have recently been studied by Assenberg *et al.*, (2002). This work looked at the sequence around the branch point and the relationship between what was a high or low energy sequence and was there a consistent choice of one sequence over another. The table from Assenberg, (2001) has been modified in Table 7-1 to include the sequences of junctions used in this thesis. Only junctions whose conformations are known and where there is a definite conformer preference are shown. Figure 7-1 explains how the sequences in the table relate to junction conformation.



Conformer A/C Conformer A/B

Figure 7-1: The high and low energy conformations of MecA. The junction is shown as both a conformer A/C and A/B. In the long-short arm assay it adopts a dominant conformer A/B. This is the low energy arrangement. The arms are labelled as A, B and C and the position of the octanediol is shown by a circle. Only the central sequence of the junction is shown.

<i>Low Energy</i>	<i>High Energy</i>	<i>Centre LE-DS</i>	<i>Centre HE-DS</i>	<i>Centre LE-DS</i>	<i>Centre HE-DS</i>	<i>Junction</i>	<i>References</i>
Conformer A/C							
TGGC/GCCA	TGCG/CGCA	GG/CC	GC/GC	RR/YY	RY/RY	TWJ-TC	Leontis <i>et al.</i> (1994)
CAAC/GTTG	CACG/CGTG	AA/TT	AC/GT	RR/YY	RY/RY	J1V6	Welch <i>et al.</i> (1995)
CGTC/GACG	CGAG/CTCG	GT/AC	GA/TC	RY/RY	RR/YY	J1	Welch <i>et al.</i> (1995)
CGTC/GACG	CGTG/CACG	GT/AC	GT/AC	RY/RY	RY/RY	J1V1	Welch <i>et al.</i> (1995)
CGTC/GACG	CGCG/CGCG	GT/AC	GC/GC	RY/RY	RY/RY	J1V2	Welch <i>et al.</i> (1995)
CCAC/GTGG	CCAG/CTGG	CA/TG	CA/TG	YR/YR	YR/YR	J1V4	Welch <i>et al.</i> (1995)
TCAC/GTGA	TCCC/GGGA	CA/TG	CC/GG	YR/YR	RR/RR	PJ46T _{2a}	Assenberg <i>et al.</i> (2002)
CCTC/GAGG	CCAG/CTGG	CT/AG	CA/TG	RR/YY	YR/YR	J1V3	Welch <i>et al.</i> (1995)
CCCC/GGGG	CCAG/CTGG	CC/GG	CA/TG	RR/YY	YR/YR	J1V5	Welch <i>et al.</i> (1995)
CTCC/GGAG	CTGC/GCAG	TC/GA	TG/CA	RR/YY	YR/YR	PJ16C _{2a}	Assenberg <i>et al.</i> (2002)
CCTG/CAGG	CCGT/ACGG	CT/AG	CG/CG	RR/YY	YR/YR	PJ42	Assenberg <i>et al.</i> (2002)
TGAG/CTCA	TGCC/GGCA	GA/TC	GC/GC	RR/YY	RY/RY	PJ45T _{2a}	Assenberg <i>et al.</i> (2002)
AGGT/ACCT	AGTG/CACT	GG/CC	GT/AC	RR/YY	RY/RY	TA-GG	This thesis
CTCC/GGAG	CTGC/GCAG	TC/GA	TG/CA	YY/RR	YR/YR	FRET	This thesis
Junction1							
Conformer A/B							
GCCG/CGGC	GCGT/ACGC	CC/GG	CG/CG	RR/YY	YR/YR	TWJ2	Overmars <i>et al.</i> (1995)
GTTA/TAAC	GTGA/TCAC	TT/AA	TG/CA	RR/YY	YR/YR	PJ55C _{2a}	Assenberg <i>et al.</i> (2002)
GGCG/CGCC	GGGT/ACCC	GC/GC	GG/CC	RY/RY	RR/YY	TWJ1	Overmars <i>et al.</i> (1995)
GACG/CGTC	GAGC/GCTC	AC/GT	AG/CT	RY/RY	RR/YY	J3CC	Rosen & Patel (1993b)
CGTT/AACG	CGGA/TCCG	GT/AC	GG/CC	RY/RY	RR/YY	PJ40C _{2a}	Assenberg <i>et al.</i> (2002)
TACC/GGTA	TAAG/CTTA	AC/GT	AA/TT	RY/RY	RR/YY	PJ44C _{2a}	Assenberg <i>et al.</i> (2002)
CTTT/AAAG	CTGT/ACAG	TT/AA	TG/CA	RR/YY	YR/YR	PJ41C _{2a}	Assenberg <i>et al.</i> (2002)
CTTC/GAAG	CTGC/GCAG	TT/AA	TG/CA	RR/YY	YR/YR	PJ51C _{2a}	Assenberg <i>et al.</i> (2002)
TTTC/GAAA	TTGT/ACAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ52C _{2a}	Assenberg <i>et al.</i> (2002)
TTTT/AAAA	TTGC/GCAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ53C _{2a}	Assenberg <i>et al.</i> (2002)
TTTC/GAAA	TTGC/GCAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ54C _{2a}	Assenberg <i>et al.</i> (2002)
TCTG/CAGA	TCAC/GTGA	CT/AG	CA/TG	RR/YY	YR/YR	YTT	Zhong <i>et al.</i> (1994)
GCTT/AAGC	GCAA/TTGC	CT/AG	CA/TG	RR/YY	YR/YR	PJ62	Assenberg <i>et al.</i> (2002)
GTTT/AAAC	GTGG/CCAC	TT/AA	TG/CA	RR/YY	YR/YR	PJ61	Assenberg <i>et al.</i> (2002)
ATCC/GGAT	ATGG/CCAT	TC/GA	TG/CA	RR/YY	YR/YR	PJ63	Assenberg <i>et al.</i> (2002)
TTTT/AAAA	TTGG/CCAA	TT/AA	TG/CA	YY/RR	YR/YR	MecA	This thesis
TTTT/AAAA	TTGG/CCAA	TT/AA	TG/CA	YY/RR	YR/YR	Coag	This thesis
CACG/CGTG	CAAC/GTTG	AC/GT	AA/TT	RY/RY	RR/YY	MecAV1	This thesis
ACTG/CAGT	ACTT/AAGT	CT/AG	CT/AG	YY/RR	YY/RR	TA-CT	This thesis
GAGC/GCTC	GAGC/GCTC	AC/GT	AG/CT	RY/RY	RR/YY	FRET	This thesis
Junction2							

Table 7-1: High and low energy di-nucleotide steps. This table has been adapted from Assenberg, (2001). Each junction is capable of folding into conformer A/B and A/C. It will spend more time in whichever conformation has the lowest energy. Column 1 shows the low energy, preferred sequence around the branch point. It shows the four bases across the continuous strand 5' to 3' from arm A and then the four bases which pair with them, also read 5' to 3'. The bases only represent the continuous arms, not the extended arm. The third and fourth columns show the di-nucleotide steps for the high and low energy situations. Some of the sequences have been highlighted and are discussed in the text.

Assenberg *et al.*, (2002) postulated that most of the sequences could be divided into two groups: Group 1 were the di-nucleotide steps AA/TT, AC/GT, AG/CT, GA/TC. Group 2 were GG/CC, GC/GC, CG/CG, CA/TG. It was found that, in general, a group 1 sequence was preferred over a group 2 sequence. It is certainly true to say that, looking at the purine/ pyrimidine di-nucleotide steps, RR/YY is commonly favoured over YR/YR (sequences highlighted in red). When conformer A/C is dominant, a few sequences showed preference for RR/YY over RY/RY (shown in blue), however, J1 (Welch *et al.*, 1995) shows a preference for RY/RY over RR/YY. There is still a limited amount of data from which to draw patterns. There will also be a large amount of interference from sequence much further from the branch point. What is needed is a singular study varying sequence close to the branch point and the further out. It is difficult to draw patterns from the data as it stands.

The work in this thesis has explored the relationship between conformation and sequence. The techniques used have analysed the dominant conformer, averaged over a long time period. With the advance of single molecule fluorescence currently being used to analyse 4WJs (McKinney *et al.*, 2004), it would be interesting to see how the rate of flipping compares between a 3WJ and 4WJ. FRET could also be employed to investigate the range of extended arm angles possible and how these vary with sequence. There has been an indication of this in this thesis but only of how the arms sit relative to each other. The Roche LightCycler has proved an invaluable tool for analysing the melting of the junctions by fluorescence. The next stage would be to analyse the precise kinetics of the melting of a three stranded or any other branched DNA structure. The technique with the most potential to come out of this work is the redundant junction assay. This variation of the long-short arm assay using pools of junctions with varying sequence showed some promising results. The next stage would be to do the same on a larger scale. That would involve using the same sequences initially but analysing a larger number to search for patterns. If this proved successful, then sequences further away from the branch point could be analysed. Eventually, this should provide some standard rules which could be used by secondary structure prediction programmes.

The properties of DNA junctions are becoming increasingly important for DNA nanotechnology as well as designing probes for the SMART reaction. DNA

nanotechnology uses the self-assembly principles of DNA i.e. specific base pairing, to create networks. Multi-helical DNA junctions are of particular use because of the arms can form multiple connection points (Du *et al.*, 1992). Figure 7-2 shows some examples of the basic structures used as building blocks. There are now many examples in the literature of 3WJs being used in nanocircuits. Charges can be transferred along a DNA junction by using donor and acceptor molecules or the junction is connected up to electrodes and appears to act as a nanowire (Cramer *et al.*, 2004). Understanding the behaviour of junctions and the variations produced by different sequences will allow the advancement of secondary structure prediction programmes. This will enable prediction of the folding of a random sequence or DNA, can be used to study RNA architecture (particularly for the development of Riboswitches to control gene expression – Tucker and Breaker, 2005), DNA nanocircuits and branched junctions *in vivo*.

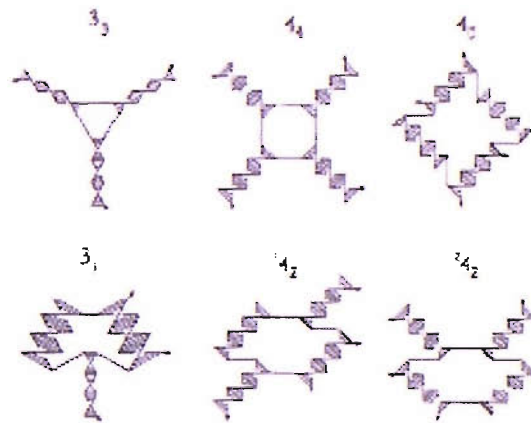


Figure 7-2: Structures created from branched DNA junctions. The diagram, taken from Du *et al.*, 1992, shows the arrangement of helices in these structures. These structures can be combined to give self-assembling networks to be used as scaffolding or nano-circuits.

References

- Aliotta, J.M., Pelletier, J.J., Ware, J.L., Moran, L.S., Benner, J.S., Kong, H. (1996). Thermostable *Bst* DNA polymerase I lacks a 3'-5' proofreading exonuclease activity. Genetic analysis: *Biomolecular engineering*. **12**: 185-195.
- Altona, C. (1996). Classification of nucleic acid junctions. *J. Mol. Biol.* **263**, 568-581.
- Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., Lander, E.S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407**: 513-516.
- Assenberg, R. and Fox, K.R. (2001). The electrophoretic mobility of DNA three-way junctions is affected by the sequence of overhanging single-stranded ends. *Electrophoresis* **22(3)**: 413-417
- Assenberg, R. (2001). Studies on three-way DNA junctions related to the development of a novel method for the detection of genetic polymorphisms. School of Biological Sciences, University of Southampton.
- Assenberg, R, Weston, A., Cardy, D.L.N., Fox, K.R. (2002). Sequence-dependent folding of DNA three-way junctions. *Nuc. Acids Res.* **30** (23), 5142-5150.
- Avery, O. T., MacLeod, C. M., and McCarty, M. (1979, reprint of 1944 paper). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Deoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III. *J. Exp. Med.* 149: 297-326.
- Azaro, M.A. and Landy, A. (1997). The isomeric preference of Holliday junctions influences resolution bias by λ integrase. *EMBO J* **16** (12), 3744-3755.

- Balashov, S.V., Gardiner, R., Park, S., Perlin, D.S. (2005). Rapid, high-throughput, multiplex, real-time PCR for identification of mutations in the *cyp51A* gene of *Aspergillus fumigatus* that confer resistance to Itraconazole. *J. Clin. Microbiol.* **43(1)**: 214-222.
- Bertram, L., Tanzi, R.E. (2005). The genetic epidemiology of neurodegenerative diseases. *J. Clin. Invest.* **115(6)**: 1449-1457.
- Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q., Lander, E.S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics.* **22**, 231-238.
- Cazanave, C and Uhlenbeck, O. (1994). RNA template directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA.* **91**: 6972-6976.
- Collins, F.S., Guyer, M.S., Chakravarti, A. (1997). Variations on a theme: Cataloging Human DNA sequence variation. *Science* **278 (5343)**: 1580-1581
- Cooper, J.P. and Hagerman, P.J. (1987). Gel electrophoretic analysis of the geometry of a DNA four way junction. *J. Mol. Biol.* **198**, 711-719.
- Cooper, J.P. and Hagerman, P.J. (1989). Geometry of a branched DNA structure in solution. *Proc. Natl. Acad. Sci. USA* **86**, 7336-7340.
- Cramer, T., Volta, A., Blumen, A., Koslowski, T. (2004) Theory and simulation of DNA charge transfer: From junctions to networks. *J. Phys. Chem. B* **108 (42)**: 16586-16592
- Cummings, C.T. and Zoghbi, H.Y. (2000). Fourteen and counting: unravelling trinucleotide repeat diseases. *Hum. Mol. Genet.* **9**:909-916

- Darby, R.A.J, Sollogoub, M., McKeen, C., Brown, L., Ristano, A., Brown, N., Barton, C., Brown, T., Fox, K.R. (2002). High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nuc. Acids. Res.* **30(9)**: e39.
- Du, S.M., Zhang, S.W., Seeman, N.C. (1992). DNA junctions, antijunctions and mesojunctions. *Biochemistry* **31 (45)**: 10955-10963
- Duckett, D.R. and Lilley, D.M. (1990). The three way DNA junction is a Y shaped molecule in which there is no helix-helix stacking. *EMBO J.* **9**, 1659-1664.
- Edwards, J.R., Itagaki, Y., Ju, J. (2001). DNA sequencing using biotinylated dideoxynucleotides and mass spectroscopy. *Nuc. Acids Res.* **29(21)** e104.
- Eichman, B.F., Vargason, J.M., Mooers, B.H.M, Ho, P.S. (2000). The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions. *Proc. Nat. Acad. Sci.* **97 (8)**, 3971-3976.
- Evans, W.E. and Relling, M.V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286 (5439)**: 487-491.
- Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., Enright, M.C., Berendt, T., Peacock, S.J., Maynard-Smith, J., Murphy, M., Spratt, B.G., Moore, C.E., Day, N.P.J. (2003). How clonal is *Staphylococcus aureus*? *J. Bacteriol.* **185(11)**: 3307-3316.
- Galas, D.J. and Schmitz, A. (1978). DNase Footprinting: a simple method for the detection of protein-DNA binding specificity. *Nuc. Acids Res.* **5(9)** 3157-3170.
- Garber, M.E., Troyanskaya, O.G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelback, M., van de Rijn, M., Rosen, G.D., Perou, C.M., Whyte, R.I., Altman, R.B., Brown, P.O., Botstein, D., Peterson, I. (2001). Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* **98(24)**: 13784-13789.

- Ghosh, A. and Bansal, M. (2003) A glossary of DNA structures from A-Z. *Acta Cryst.* **D59**: 620-626.
- Giglio, S., Monis, P.T., Saint, C.P.(2003). Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR. *Nuc. Acids. Res.* **31(22)** e136.
- Grainger, R.J., Murchie, A.I.H, Lilley, D.M.J. (1998). Exchange between stacking conformers in a four-way junction. *Biochemistry* **37**:23-32.
- Groneberg, D.A., Hilgenfeld, R., Zabel, P. (2005). Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir. Res.* **6(1)**: 8
- Hall, M.J., Wharam, S.D., Weston, A., Cardy, D.L.N., Wilson, W.H. (2002). Use of signal mediated amplification of RNA technology (SMART) to detect marine cyanophage DNA. *Biotechniques* **32(3)**: 604
- Hagerman, P.J. (1988) Flexibility of DNA. *Ann. Rev. Biophys. Chem.* **17**: 265-86
- Hargreaves, D., Rice, D.W., Sedilnikova, S.E. (1998). Crystal structure of *E.Coli* RuvA with bound DNA Holliday junction at 6 angstrom resolution. *Nature Struc. Biol.* **5(6)** 441-446
- Hendrix, M.J.C., Seftor, E.A., Meltzer, P.S., Gardner, L.M.G., Hess, A.R., Kirschmann, D.A., Schatteman, G.C., Seftor, R.E.B. (2001). Expression and functional significance of VE – Cadherin in aggressive human melanoma cells: role in vasculogenic mimicry. *Proc. Natl. Acad. Sci. USA* **98 (14)**: 8018-8023.
- Ho, P.S., Eichman, B.F. (2001). The crystal structures of DNA Holliday junctions. *Curr. Op. Struc. Biol.* **11 (3)**: 302-308
- Holland, P.M., Abramson, R.D., Watson, R. & Gelfand, D.H. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease

activity of the *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* **88**, 7276-7280.

Holliday, R. (1964) A mechanism for gene conversion in fungi. *Genet. Res.* **5**, 282-304

Horimoto, T., Kowaoka, Y. (2001). Pandemic threat posed by avian influenza A viruses. *Clin. Microbiol. Rev.* **14(1)**: 129-149.

Hultin, E., Kaller, M., Ahmadian, A., Lundeberg, J. 2005. Competitive enzymatic reaction to control allele-specific extensions. *Nuc. Acids Res.* **33(5)**: e48.

Imburgio, D., Rong, M., Ma, K., McAllister, W.T. (2000). Studies of promoter recognition and start selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**: 10419-10430

Jares-Erijman, E.A. and Jovin, T.M. (2003). FRET imaging. *Nature Biotech* **21(11)**: 1387-1395.

Jensch, F. and Kemper, B. (1986). Endonuclease VII resolves Y-junctions in branched DNA in vitro. *EMBO J* **5(1)**: 181-189.

Jensch, F., Kosak, H., Seeman, N.C., Kemper, B. (1989) Cruciform cutting endonucleases from *Saccharomyces cerevisiae* and phage T4 show conserved reactions with branched DNAs. *EMBO J.* **8(13)**: 4325-4334.

Kadmas, J.L., Ravin, A.J., Leontis, N.B. (1995). Relative stabilities of DNA three-way, four-way and five-way junctions (multi-helix loop junctions): unpaired nucleotides can be stabilising or destabilising. *Nuc. Acids. Res.* **23(12)**: 2212-2222.

Karger, A.E. (1996). Separation of DNA sequencing fragments using an automated capillary electrophoresis instrument. *Electrophoresis* **17(1)**: 144-151.

- Kerem, B.S., Rommens, J.M, Buchannen, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buckald, M., Tusi, L.C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* **254**: 1073-1080.
- Kheterpal, I., Schener, J., Clark, S.M., Radhakrishnan, A., Ju, J., Ginther, C.L., Sensabough, G.F., Matties, R.A. (1996). DNA sequencing using a four-colour confocal fluorescence capillary array scanner. *Electrophoresis* **17**: 1852-1859.
- Krup G (1989). Unusual promotor-independent transcription reactions with bacteriophage RNA polymerases. *Nuc. Acids Res* **17 (8)**: 3023-3036
- Lafontaine, D.A., Norman, D.G. and Lilley, D.M.J. (2001). Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *EMBO J.* **20(6)**: 1415-1424.
- Lee, S., Cavallo, L., Griffith, J. (1997). Human p53 binds Holliday junctions strongly and facilitates their cleavage. *J. Biol. Chem.* **272** (11), 7532-7539.
- Leontis, NB., Kwok, W., Newman, JS. (1991). Stability and structure of three way DNA junctions containing unpaired nucleotides. *Nuc. Acids Res.* **19**, 759-766.
- Leontis, N.B., Hills, M.T., Piotto, M., Ouporov, I.V., Malhotra, A. and Gorenstein, D.G. (1994). Helical stacking in DNA three way junctions containing two unpaired pyrimidines: proton NMR studies. *Biophys. J.* **68**, 251-265.
- Levi, K., Bailey, C., Bennett, A., Marsh, p., Cardy, D.L.N., Towner, K.J. (2003) Evaluation of an isothermal signal amplification method for rapid detection of methicillin-resistant *Staphylococcus aureus* from patient-screening swabs. *J. clin. microbiol.* **41(7)**: 3187-3191
- Livingstone, J., Axton, R.A. Gilfillan, A., Mennie, M., Ccompton, M., Liston, W.A., Calder, A.A., Gordon, A.J., Brock, D.J.H. (1994). Antenatal screening for cystic fibrosis – a trial of the couple model. *BMJ* **308** (6942): 1459-1462.

- McKinney, S.A., Déclais, A.C., Lilley, D.M.J., Ha, T. (2003). Structural dynamics of individual Holliday junctions. *Nature Struc. Biol.* **10**, 93-97
- Miick, S.M., Fee, R.S., Millar, D.P. and Chazin, W.J. (1997). Crossover isomer bias is the primary sequence-dependent property of immobilized Holliday junctions. *Proc. Nat. Acad. Sci. USA* **94**, 9080-9084.
- Minagawa, T., Murakami, M., Ryo, Y., Yamagishi, H (1983). Structural features of very fast sedimenting DNA formed by gene 49 defective T4. *Virology* **126**: 18-93
- Moncke-Buchner, E., Reich, S., Mucke, M., Reuter, M., Messer, W., Wanker, E.E., Kruger, D.H. (2002). Counting CAG repeats in the Huntington's disease gene by restriction endonuclease EcoP151 cleavage. *Nuc. Acids. Res.* **30(16)**: e83
- Montandon, A.J., Green, P.M., Giannelli, F., Bentley, D.R. (1989). Direct detection of point mutations by mismatch analysis: application to haemophilia B. *Nuc. Acid. Res.* **17(9)**: 3347-3358.
- Murchie, A.I.H., Carter, W.A., Portugal, J., Lilley, D.M.J. (1990). The tertiary structure of the four-way junction affords protection against DNaseI cleavage. *Nuc. Acids. Res.* **18(9)**: 2599-2606.
- Nacheva, G.A., Berzal-Herranz, A. (2003). Preventing undesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur. J. Biochem.* **270**: 1458-1465.
- Nadel, Y., Weisman-Shomer, P., Fry, M. (1995). The Fragile X syndrome single strand d(CGG)_n Nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* **270(48)**: 28970-28977.
- Newton, C.R., Graham, A., Hepinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., Markham, A.F. 1989 Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nuc. Acids Res.* **17 (7)**: 2503-2516.

Neidle, S (1999). Oxford handbook of Nucleic Acid Structure. *Oxford University Press*.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P., Landegren, U. (1994) Padlock probes: circularising oligonucleotides for localized DNA detection. *Science* **265**: 2085-2088.

Nowakowski, J., Shim, P.J., Prasad, G.S., Stout, C.D., Joyce, G.F. (1999). Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10-23 DNA enzyme. *Nature Struct. Biol.* **6(2)**: 151-156

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-stand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**: 2766-2770.

Oussatcheva, E.A., Shlyakhtenko, L.S., Glass, R. and Sinden, R.R. (1999). Structure of branched DNA molecules: gel retardation and atomic force microscopy. *J. Mol. Biol.* **292**, 75-86.

Oussatcheva, E.A., Hashem, V.I., Zou, Y., Sinden, R.R., Potaman, V.N. (2001). *J. Biol. Chem.* **276** (33), 30878-30884.

Overmars, F.J.J., Pikkemaat, J.A., van den Elst, H., van Boom, J.H., Altona, C. (1996). NMR studies of DNA three-way junctions containing two unpaired thymidine bases: the influence of the sequence at the junction on the stability of the stacking conformers. *J. Mol. Biol.* **255(5)**: 702-713.

Overmars, F.J.J. and Altona, C. (1997). NMR study of the exchange rate between two stacked conformers of a model Holliday junction. *Biochemistry* **273**:519-524.

Parkinson, M.J. and Lilley, D.M.J. (1997). The junction-resolving enzyme T7 endonuclease I: quaternary structure and interaction with DNA. *J. Mol. Biol.* **270**, 169-178.

- Pearson, C.E. and Sinden, R.R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35**, 5041-5053.
- Pearson, C.E., Wang, Y.H., Griffith, J.D., Sinden, R.R. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n(GAG)_n repeats from the myotonic dystrophy locus. *Nuc. Acids Res.* **26** (3): 816-823.
- Pearson, C.E., Tam, M., Wang, Y.H., Montgomery, S.E., Dar, A.C., Cleary, J.D., Nichol, K. (2002). Slipped-strand DNAs formed by long (CAG)_n(CTG)_n repeats: slipped-out repeats and slip-out junctions. *Nuc. Acids, Res.* **30** (20), 4534-4547.
- Perez-Roth, E., Lorenzo-Diaz, F., Batista, N., Moreno, A., Mendez-Alvarez, S. (2004) Tracking methicillin resistant *Staphylococcus aureus* clones during a 5-year period (1998-2002) in a Spanish hospital. *J. Clin. Microbiol.* **42(10)**:4649-4656.
- Puck, J.M, Nussbaum, R.L., Conley, M.E. (1987). Carrier detection in X-linked severe combined immunodeficiency based on patterns of X chromosome inactivation. *J. Clin. Invest.* **79**: 1395-1400.
- Rettberg, C.C., Prere, M.F., Gesteland, R.F., Atkins, J.F., Fayet, O. (1999). A three-way junction and constituent stem-loops as the stimulator for programmed -1 frameshifting in bacterial insertion sequence IS911. *J. Mol. Biol.* **286 (5)**: 1365-1378
- Rhodes, D.R., Chinnaiyan, A.M. (2005) Integrative analysis of the cancer transcriptome. *Nature Genetics supp.* **37**: S31-S37.
- Riordan, J.R., Rommens, J.M., Kerem, B.S., Alon, N., Rozmahel, R., Grzelczak, Z., Lok, S., Plavsic, N., Chou, J.L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., Tsui, L.C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **254**: 1066-1073.

- Rosen, M.A. and Patel, D.J. (1993a). Conformational differences between bulged pyrimidines (C-C) and purines (A-A, I-I) at the branch point of three-stranded DNA junctions. *Biochemistry* **32**, 6563-6575.
- Rosen, M.A. and Patel, D.J. (1993b). Structural features of a three-stranded DNA junction containing a C-C junctional bulge. *Biochemistry* **32**, 6576-6587.
- Sanger, F., Nicklen, S., Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. **74**: 5463-5467
- Shlyakhtenko, L.S., Potaman, V.N., Sinden, R.R., Gall, A.A. and Lyubchenko, Y.L. (2000). Structure and dynamics of three way DNA junctions: atomic force microscopy studies. *Nuc. Acids Res.* **28**, 3472-3477.
- Sousa, R. and Padilla, R. (1995). A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* **14 (18)**: 4609-4621
- Steitz, J.A (1992). Splicing takes a Holliday. *Science* **257**: 888-889
- Stühmeier, F., Welch, J.B., Murchie, A.I.H., Lilley, D.M.J. and Clegg, R.M. (1997). Global structure of three way DNA junctions with and without additional unpaired bases: a Fluorescence Resonance Energy Transfer analysis. *Biochemistry* **36**, 13530-13538.
- Szemes, M., Bonants, P., de Weerd, M., Baner, J., Landegren, U., Schoen, C.D. (2005). Diagnostic application of padlock probes – multiplex detection of plant pathogens using universal microarrays. *Nuc. Acids. Res.* **33(8)**: e70.
- Takahashi, M., Rhodes, D.R., Furge, K.A., Kanayama, H.O., Kagawa, S., Haab, B.B., The, B.T. (2001). Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc. Natl. Acad. Sci. USA* **98(17)**: 9754-9759.

- Thomas, P.D., Kejariwal, A.(2004). Coding single-nucleotide polymorphisms associated with complex vs Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc. Natl. Acad. Sci. USA* **101(43)**: 15398-15403.
- Tucker, B.J. and Breaker, R.R. (2005) Riboswitches as versatile gene control elements. *Curr. Op. Struc. Biol.* **15 (3)**: 342-348
- Tyagi, S, Bratu, D.P, Kramer, F.R (1998) Multicolour molecular beacons for allele discrimination. *Nature biotech.* **16**: 49-53.
- Van Buuren, B.N.M., Overmars, F.J.J., Ippel, J.H., Altona, C. and Wijmenga, S.S. (2000). Solution structure of a DNA three way junction containing two unpaired thymidine bases. Identification of sequence features that decide conformer selection. *J.Mol.Biol.* **304**, 371-383
- van Hoek, A.H.A.M., Scholtens, I.M.J., Cloeckert, A., Aarts, H.J.M. (2005) Detection of antibiotic resistance genes in different Salmonella serovas by oligonucleotide microarray analysis). *J. of microbial. Methods* **62**: 13-23.
- Wang, Y., Mueller, J.E., Kemper, B., Seeman, N.C. (1991). Assembly and characterisation of five-arm and six-arm DNA branched junctions. *Biochemistry* **30**, 5667-5674.
- Wannet, W.J.B., Spalburg, E., Heck, M.E.O.C., Pluisher, G.N., Willems, R.J.L., Neeling, A.J. (2004). Widespread dissemination in the Netherlands of the epidemic Berlin methicillin resistant *Staphylococcus aureus* clone with low level resistance to oxacillin. *J. Clin. Microbiol.* **42(7)**: 3077-3082.
- Watson, J., Hays, F.A., Ho, P.S. (2004). Definitions and analysis of DNA Holliday junction geometry. *Nuc. Acids. Res.* **32 (10)**: 3017-3027
- Welch, J.B., Duckett, D.R. and Lilley, D.M. (1993). Structures of bulged three way DNA junctions. *Nuc. Acids Res.* **21**, 4548-4555.

- Welch, J.B., Walter, F. and Lilley, D.M.J. (1995). Two inequivalent folding isomers of the three way DNA junction with unpaired bases: the sequence-dependence of the folded conformation. *J. Mol. Biol.* **251**, 507-519.
- Wharam, S.D., Marsh, P., Lloyd, J.S., Ray, T.D., Mock, G.A., Assenberg, R., McPhee, J.E., Brown, P., Weston, A. and Cardy, D.L.N. (2001). Specific detection of DNA and RNA targets using a novel isothermal nucleic acid amplification assay based on the formation of a three way junction structure. *Nuc. Acids. Res.* **29**, e54.
- White, M.F., Giraud-Panis, M.J.E., Pöhler, J.R.G., Lilley, D.M.J. (1997). Recognition and manipulation of branched DNA structure by junction-resolving enzymes. *J. Mol. Biol.* **269**, 647-664.
- Wu, B., Girard, F., van Buuren, B., Schleucher, J., Tessari, M., Wijmenga, S. (2004). Global structure of a DNA three-way junction by solution NMR: towards prediction of 3H fold. *Nuc. Acids. Res.* **32 (10)**: 3228-3239.
- Yang, M. and Millar, D.P. (1996). Conformational flexibility of three way DNA junctions containing unpaired nucleotides. *Biochemistry* **35**, 7959-7967.
- Youil, R., Kemper, B.W., Cotton, R.G.H. (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci. USA* **92**: 87-91.
- Zipper, H., Brunner, H., Bernhagen, J., Vitzhum, F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nuc. Acids. Res.* **32(12)**: e103.