University of Southampton

SCHOOL OF BIOLOGICAL SCIENCES
DIVISION OF BIOCHEMISTRY & MOLECULAR BIOLOGY
FACULTY OF SCIENCE

STUDIES ON THREE-WAY DNA JUNCTIONS RELATED TO THE DEVELOPMENT OF A NOVEL METHOD FOR THE DETECTION OF GENETIC POLYMORPHISMS.

René Assenberg

A thesis submitted for the degree of Doctor of Philosophy,
Faculty of Science, University of Southampton

January, 2001

University of Southampton

Abstract

FACULTY OF SCIENCE
SCHOOL OF BIOLOGICAL SCIENCES
DIVISION OF BIOCHEMISTRY & MOLECULAR BIOLOGY

Doctor of Philosophy

Studies on three-way DNA junctions related to the development of a novel method for the detection of genetic disease

René Assenberg

Mutations in the genetic material of living organisms can have major effects on the survival of the individual. Genetic variations influence both the virulence and treatment-susceptibility of micro-organisms and affect the host-susceptibility for pathogenic micro-organisms. In addition, many common genetic disorders arise from defects in multiple alleles. This has important implications for health diagnostics as most available methodologies for genetic screening are not suitable to detect these. We have therefore developed a novel nucleic acid screening method, named signal mediated amplification of RNA technology (SMART), for detection of genetic polymorphisms. The technology uses two nucleic acid probes that anneal to a specific target sequence and to each other, resulting in the formation of a three-way DNA junction (TWJ). One probe (extension probe) can be extended against the remaining probe (template probe) by a DNA polymerase. This activity yields a functional T7 RNA polymerase promoter, and addition of T7 RNA polymerase results in RNA production. The RNA produced functions as a signal for the presence of a specific target sequence. These studies have focussed on optimising the junction design in SMART to allow discrimination between targets containing only small genetic variations. The results indicate that a number of factors can enhance the specificity of SMART: 1) incorporation of PNA (peptide nucleic acids) in the target complementary regions of the two probes, 2) shortening the size of the target containing duplex arms, 3) shortening the complementary sequence shared by the two probes, 4) elevating the reaction temperature. A further DNA-analogue (Locked Nucleic Acids, LNA) was also assessed but found to be less sensitive for mutation detection. The optimised assay is able to discriminate between the wild-type and single- and three-base variants of the cystic fibrosis transmembrane conductance regulator and Factor V genes, using either synthetic or PCR amplified genomic material. Additional studies also revealed that the structure of the TWJ may affect the RNA yield. In the presence of divalent cations, three-way junctions adopt two different coaxially stacked conformations. The preferred conformer depends on the sequence around the branch point and this affects the efficiency of the assay. We have used a gel-electrophoresis method to probe the structures adapted by different junctions. The major factor influencing the conformer distribution is the arrangement of purines and pyrimidines around the junction, although some junctions appeared to adopt unusual conformations. Together these findings have improved the understanding of TWJ conformation and allow for more accurate structure prediction based only on the DNA sequence.

Finally, a novel cleavage activity of restriction endonucleases was discovered and characterised. The activity was observed only for DNA constructs containing a non-DNA linker (e.g. hexaethylene glycol, abasic site) and involved cleavage of the phosphate located directly 5' of the linker. The cleavage was further found to be independent of the presence of canonical substrate and appeared to be modulated by methyltransferase treatment.

PREFACE

This work was performed between September, 1997 and November, 2000 in the Division of Biochemistry & Molecular Biology, University of Southampton and at Cytocell Limited, Banbury.

Some of the results presented in chapter 5 have been accepted for publication in: *Electrophoresis*.

I would like to thank the following people for various reasons:

Cytocell Limited is greatly acknowledged for generously (with respect to my research budget) sponsoring my thesis. I would like to thank Dr. Anthony Weston and Dr. Don Cardy in particular for their support and suggestions. In addition, I would like to thank all the team-members at Cytocell, Peter Marsh, Susan Wharam, Graham Mock, John Lloyd, Julie McPhee, Sarah Cunningham, Trevor Ray, Philip Brown for making my time at Cytocell thoroughly enjoyable both scientifically and socially. Gordon Munro is acknowledged for introducing me to SMART. I also wish to thank Cytocell for creating and sustaining the Chromoprobe department, as their strict drinking discipline has been a continuous source of inspiration on many stress-releasing nights for me.

I would further like to gratefully thank Professor Keith R Fox for his support and many suggestions while I worked at the university. Professor Tom Brown is also acknowledged for his suggestions and discussions. I express my gratitude to Oswel for producing the numerous oligos used in this thesis.

Of course, my life in Southampton would not have been the same without the enlightening presence of some special characters by the names of Kris, Claire, James, Sarah, Paul, Manuel, Richard, Tracey and Bruce. I will remember all the "special times" we had together...

A special thanks to Kris for his motivation and suggestions during the writing of my mini and Sarah for discussions on the HEG research and various other bits.

My parents, realising that surviving on the salary scales PhD students have to put up with in this country is a simple impossibility, are gratefully acknowledged for their substantial financial (and emotional) support.

Finally, I am of course grateful to Magda, for her love and support.

Abbreviations

A	adenosine
AB	annealing buffer
APC	activated protein C
AP	alkaline phosphatase
ARMS	amplification refractory mutation system
ASO	allele-specific oligonucleotide
ATP	adenosine 5'-triphosphate
bp	base pair
BSA	bovine serum albumin
C	cytosine
CCD	charge-coupled device
CCR5	cysteine-cysteine linked chemokine receptor 5
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
Ci	curie
COP	competitive oligonucleotide primer
cps	counts per second
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel electrophoresis
DMS	dimethyl sulphate
DNA	deoxyribonucleic acid
DNAP	DNA polymerase
DNase	deoxyribonuclease
dNTP	deoxyribonucleic acid triphosphate
ds	double-stranded
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ELOSA	enzyme-linked oligonucleotide sorbency assay
FEN	flap endonuclease
FRET	fluorescence resonance energy transfer

FWJ	four-way DNA junction
G	guanosine
HA	heteroduplex analysis
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HPLC	high-performance liquid chromatography
IFN	interferon
IL	interleukin
LD	linkage-disequilibrium
LNA	locked nucleic acid
LSA	long-short arm assay
MALDI-TOF	matrix-assisted laser desorption ionisation-time-of-flight
MBL	mannose binding lectin
MHC	major histocompatibility complex
mRNA	messenger RNA
MS	mass spectroscopy
MT	methyl transferase
NASBA	nucleic acid sequence-based amplification
NC-IUBMB	nomenclature committee of the international union for biochemistry and molecular biology
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser electron spin
NRAMP1	natural resistance associated macrophage protein 1
ODU	optical density units
Oligo	oligonucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEDIAT	primer extension dependent isothermal amplification technology
PNA	peptide (or polyamide) nucleic acid
pNPP	4-nitrophenyl phosphate
R	purine
RCA	rolling circle amplification

R-ENase	restriction endonuclease
RNA	ribonucleic acid
RNAP	RNA polymerase
RNase	ribonuclease
R-ENase	restriction endonuclease
rNTP	ribonucleic acid triphosphate
rpm	revolutions per minute
RT	reverse transcriptase
SMART	signal mediated amplification of RNA technology
SNP	single nucleotide polymorphisms
ss	single-stranded
SSCP	single strand conformation polymorphisms
SSOP	sequence-specific oligonucleotide probe
T	thymidine
TBE	tris-boric acid-EDTA
TBM	tris-boric acid-magnesium
TEMED	N,N,N',N'-tetramethylethylenediamine
TMA	transcription-mediated amplification
Tris	2-amino-2-(hydroxmethyl)-1,3-propandiol
TWJ	three-way DNA junction
UV	ultraviolet
VDR	vitamin D receptor
Y	pyrimidine

Contents

Figure Index

Chapter One

Chapter Two

Chapter Three

1 Introduction

1.1 Genetic polymorphisms in infectious and heritable diseases and their treatment.

This thesis details the development of a novel diagnostic method for detecting genetic polymorphisms. There is considerable interest in this area as a result of the large-scale collaborative efforts such as the human genome project and more recently the single-nucleotide polymorphism consortium (discussed below). Both efforts are aimed at elucidating genotype complexity, and how this is expressed at the phenotype level. Although many genetic diseases have been described, these are mainly rare, Mendelian-type diseases, caused by a single dominant/recessive allele. Most common genetic diseases (e.g. cancer, diabetes) remain poorly characterised at the genotype level because most of them involve a number of alleles that are difficult to detect by traditional methods.

The effect of genetic variations is not restricted to genetic diseases. It also affects the action of drugs used for treating disease (pharmacogenomics), one's susceptibility for infectious diseases and the pathogenicity of micro-organisms and their susceptibility to treatment.

These developments have major implications for the methods used to diagnose genetic variations. As will be shown later in this chapter, most available methods are unsuitable for diagnostic purposes because the detection of single base changes is technically demanding (resulting in expensive and impractical methods). Another crucial problem is that most PCR-based methods can yield false positive/negative results, something that is unacceptable when the diagnosis of a patient is concerned, as the outcome could alter his/her entire life. Novel approaches are therefore required and this thesis deals with one such method that may circumvent some of these problems.

1.1.1 Genetic polymorphisms in pathogens involved in virulence

Genetic variation is the single factor that allows individuals to adapt to the ever-changing environment. From a human perspective the process of natural selection and consequent adaptation is not always a beneficial one. For one, evolution does not harbour much sympathy for maladaptations. This means that genetic variations that reduce one's reproductive success are usually disposed of. However, by no means every fitness-reducing variation is discarded. Situations may occur where, for instance, a variation enhances the fitness of one trait whilst reducing it for another trait (trade-off). Also, evolutionary conflicts may arise between genes that have different transmission patterns but reside within a single host. An example of the latter is the conflict that occurs between host genes and the genes of an invading pathogen. Both adapt to each other to minimise the destructive forces exerted by the other, in order to maintain reproductive ability. It is an intriguing battle of fitness as the strategies deployed by both differ radically. Micro-organisms, reproducing largely asexually (although this disregards horizontal gene transfer), compensate the consequently lower rate of genetic variability by reproducing rapidly and in large numbers (Stearns, 1999). In this manner rapid adaptation may be achieved to host immune attack through, for instance, variations in cell surface proteins that function as potential antigenic focus points. In fact, most pathogenic species appear to contain small but specific hypervariable DNA regions (referred to as contingency loci) often encoding part of the cell surface proteins responsible for adherence and/or antigenic variation (Deitsch *et al.*, 1997). Examples are *Neisseria gonorrhoeae* and *N. meningitidis*, which possess surface protein complexes (referred to as pili) that mediate their attachment to host epithelium cells. These pili contain repeating subunits (pilins) and the genes encoding them were found to contain three different regions with respect to their genetic variability referred to as constant, semivariable and hypervariable (Diaz *et al.*, 1984, Perry *et al.*, 1987). The hypervariable regions appeared to determine the antigenicity of the pathogen as well as its tissue specificity (tropism). As pointed out previously, this immune-evasion mechanism is found in many pathogens (refer to Cross, 1996, Deitsch *et al.*, 1997, for reviews), including viruses and eukaryotic parasites such as *Trypanosoma bruceii* (sleeping sickness) and *Plasmodium falciparum* (malaria). The presence of specific variable genetic regions is by no means restricted to cell surface proteins. In many pathogens the

virulence genes are located in so-called "pathogenicity islands" (Deitsch *et al.*, 1997, Hacker *et al.*, 1997), in which promoter elements or parts of genes themselves contain repeated DNA segments. The latter may result in strand slippage during DNA replication, causing the deletion or addition of these repeated segments. This may ultimately affect the promoter activity (and thus, the quantity of associated gene product) or induce translational frameshifts. Both evoke a similar effect in that protein activity is either maintained or eliminated (often referred to as an "on/off" switch or phase variation), although reduced levels of transcription may occur as well. Other mechanisms exist as well that can invoke an "on/off"-state of expression through site-specific DNA alterations, including variable DNA methylation, inversion of genetic elements, and intra- or inter genomic recombination (Baumberg, 1999).

These examples explain why such large heterogeneity has been observed even among members of the same pathogenic species with respect to their virulence and infectivity. The genetic variability at a restricted number of loci, due to selection pressure exerted by the host environment (and the genetic process involved in its accomplishment) also play a dominant role in the acquisition of antibiotic resistance. Over the past decade substantial evidence has been gained that the large-scale administration of antibiotics has increased the number of resistant strains (Levin & Anderson, 1999). This man-tailored selection mechanism is assumed responsible for the emergence of multiple resistant pathogens such as found for *Streptococcus pneumoniae*, *Mycobacterium tuberculosis* and *Staphylococcus aureus* and others (Claverys *et al.*, 2000, Klepser *et al.*, 1997, Maiden, 1998, McCormick, 1998, Tomasz, 1999). A number of mechanisms may underlie the acquisition of antibiotic resistance. Plasmid and transposon transfer at the inter- or intraspecies level is a common mechanism of genetic exchange in bacteria, and is considered as one of the major transmission routes of antibiotic resistance genes. A different genetic mechanism entails the mutation of the antibiotic target such that its affinity for the antibiotic is reduced like, for instance, the penicillin-binding protein encoded by the *mecA* gene in methicillin-resistant *S. aureus* (MRSA) (Neu, 1992). Some micro-organisms also are able to remove the antibiotic from the cell interior, prevent the antibiotic from reaching the cell interior or enzymatically modify the antibiotic to render it inactive (for reviews refer to Maiden, 1998, Neu, 1992, Nikaido, 1994). In addition, horizontal exchange of genetic material via transformation, often induces

recombination between related but distinct sequences (recA for instance can tolerate a 30% difference in sequence homology). The result is the generation of so-called mosaic genes, which can be regarded as novel alleles of the affected chromosomal target gene and may be retained in subsequent progeny. This intrageneic recombination process may also produce antibiotic resistant proteins (Maiden, 1998).

These defences are only a few of the many a pathogen can use in response to negative influences exerted by the host environment. Most others are a result of whole-gene acquisition and/or random mutation coupled to long-term evolution. Also house-keeping genes, generally regarded as less genetically flexible, are often involved in the induction of a pathogenic state. As pointed out previously, many bacteria contain pathogenicity islands and these may include genes whose products interfere with antibody or complement-mediated attack in a number of different ways that do not necessarily involve loci of increased genetic variability. Instead, a plethora of proteins have been characterised from a large range of pathogens that contribute to the evasion of host defence systems. These include proteins that degrade or release antigen-IgA complexes from the outer membrane, inhibit transportation of antigen-presenting complexes to the surface of the infected cell etc (Henderson *et al.*, 1998, Henderson *et al.*, 1999).

In summary, pathogens have evolved intricate genetic mechanisms to introduce genetic variations in DNA segments that are involved in interactions with the host. These localised variations may aid the pathogen in successful invasion and proliferation within the host by evading immune responses and antibiotic treatment and may be crucial in the diagnosis and treatment of infected individuals.

1.1.2 Genetic factors affecting human susceptibility to infectious agents and genetic disease.

The evasive strategies deployed by pathogens are no guarantee for survival as among humans substantial differences have been observed with respect to their susceptibility to infectious diseases. In other words, a certain pathogen may proliferate within one host, but be eliminated in a different, yet related host. Evidence exists that implicates polymorphisms in the genes encoding the interferon-

γ-receptor (IFN-γ), natural-resistance-associated macrophage protein 1 (NRAMP1), mannose-binding lectin (MBL), vitamin D receptor (VDR) and the HLA system as contributors to susceptibility to a variety of infectious diseases (reviewed in: Altare *et al.*, 1998, Bellamy & Hill, 1998, Gruenheid & Gros, 2000).

Diseases that enhance the invasive potential of microorganisms, including commensals, can further increase the host susceptibility. Cystic fibrosis (an inherited disease) for instance, arises due to mutations in a gene (cystic fibrosis transmembrane regulator gene, CFTR) encoding a plasma membrane protein involved in chloride ion secretion and regulation of other ion channels. Numerous disease-causing mutations have been characterised for this gene although the majority of the affected, Caucasian persons carry the ΔF508 deletion (a three-base deletion leading to loss of a phenylalanine). Ultimately these mutations lead to increasing loss-of-function of the lungs as a result of (predominantly) *Pseudomonas aeruginosa* infection (for a recent review on cystic fibrosis, refer to: Pier, 2000). Interestingly, *P. aeruginosa* is an opportunistic lung commensal and will become invasive only after a genetic conversion switches it to a mucoid state. For as yet unknown reasons, the lung environment of CF patients is thought to increase the selection on mucoid variants of *P. aeruginosa* (Boucher *et al.*, 1997, Pier, 2000), illustrating again the complex interactions between the host and pathogen genomes.

These examples suggest that optimal disease diagnosis involves not only the determination of which pathogen is infecting a patient, but also whether (and to what extent) his/her genetic make-up harbours susceptibility to this pathogen. Other factors play an important role as well, such as immunogenic life-history and life-style.

1.1.3 Genetic polymorphisms contributing to susceptibility for human diseases and their treatment

Susceptibility to infectious disease is only one genetic condition where genetic variation plays a prominent role. Awareness is growing that most common genetic diseases are multifactorial, controlled by more than two alleles and environmental factors. Although many "simple" Mendelian diseases have been mapped to their respective chromosome locations (refer for instance to the Online Mendelian Inheritance in Man Database (http://www3.ncbi.nlm.nih.gov/omim), at the time of

writing listing over 5,000 disorders linked to single-gene defects), these are, in most cases, rare. More common diseases including cancer, diabetes and coronary heart diseases appear to involve more than two alleles (dismissing the simple (co)dominant/recessive transmission pattern) and are often polygenic. As an example, studies on asthma and atopic allergies mediated by IgE expression, have revealed a large and complex network of interactions between different genes (Cookson, 1999). Candidate gene studies (or association studies) have revealed numerous polymorphisms that may predispose one to asthma and allergy, involving the human leukocyte antigen (HLA) system, major histocompatibility complex (MHC), FcεRI (or IgE receptor), IL-4 and IL-4 receptor. However, as with most complex genetic diseases analysed so far, such associations require more detailed analysis to verify the predisposing nature of the polymorphisms.

The complexity described here, exemplified by asthma, explains why certain predisposed individuals that appear to carry the same genotype show considerable differences in disease pattern with respect to the age of on-set or severity. Even for cystic fibrosis, a classic example of a monogenic trait where in 70% of the affected (Caucasian) cases one specific, three-base deletion is the causative factor (Kerem *et al.*, 1989), widespread differences have been observed in the age-of-onset. Only recently, a CF modifier locus (modifier loci do not affect the severity of the disease itself) was detected (Zielenski *et al.*, 1999), indicating that cystic fibrosis can be considered a polygenic trait. Similar arguments have been made for other, classically considered monogenic traits such as phenylketonuria (Scriver & Waters, 1999).

Another related area, that of pharmacogenomics, is perhaps the best example of the complex relationship between genotype and phenotype. Drugs for medical treatment are usually designed for the "average person" yet such a person is in fact non-existent due to large differences in many genes involved in the drug-metabolising pathways (e.g. those of the cytochrome P450). These polymorphisms affect the rate at which the drug is broken down, and whether or not (and how many) damaging reaction intermediates (radicals) are formed (Evans & Relling, 1999, Taningher *et al.*, 1999). This has implications for the effectiveness and toxicity of the drug in question. The affected drug-metabolism caused by the same polymorphisms may also be considered as a complex disease since the chances of cancer

development may be increased due to abnormal levels of radicals generated by a defective (or less/over) active enzyme (Taningher *et al.*, 1999).

These examples illustrate the basic problem underlying current genetical understanding: a shortage of sequence and association data. The human genome project is but one step on the long route towards filling this knowledge gap. This project will provide a reference sequence but will not reveal the extent of genetic polymorphism that resides within our genomes much less how the latter predisposes one to a certain disease. In order to achieve this, large populations need to be scanned and tested on associations between variations in the genome and a certain disease trait via comparative analysis of affected and unaffected individuals (association studies). With this objective in mind, a consortium of companies and institutes was recently formed (the single nucleotide polymorphism consortium) focussing exclusively on mapping SNPs in the human genome (Collins *et al.*, 1997). The reason why efforts are focussed on single-nucleotide variations is because they represent the most abundant form of genetic variation. In addition, they are also relatively easy to apply to high-throughput analysis. Linkage-disequilibrium (LD) mapping is the method of choice when SNPs are concerned (essentially LD is a method to assess associations between SNPs and diseases) (Cargill *et al.*, 1999, Kruglyak, 1999). Linkage disequilibrium occurs when the SNP resides so close to the actual predisposing gene that recombination becomes negligible. This means that even in unrelated persons a close linkage between the SNP and the gene remains due to the lack of recombination. This approach is in stark contrast to the more traditional linkage methods, which relies on pedigree analysis of trait inheritance and consequently relies on recombination to occur. The use of pedigrees has obvious problems (most notably the need for an accurate model of inheritance and informative pedigrees), although LD analysis holds its own problems as well (Kruglyak, 1999; Lander & Schork, 1994; Schafer & Hawkins, 1998; Terwilliger & Weiss, 1998).

One major drawback of the population genetical LD approach is that large numbers of affected and non-affected persons need to be screened to establish which SNPs are significantly associated with a disease. Thus, LD-analysis requires a dense map of SNPs to be established, which in turn requires enormous high-throughput ability of the techniques involved. Non-coding SNPs that reside within regulatory

elements (promoter, silencer, enhancer, splice-regions) present another challenge as less is known about these DNA elements and the way they control gene expression. Finally, major statistical hurdles are present in any association-based method, and is most notably affected by the choice of population used. Admixture and founder effects are notorious for their ability to generate statistically significant, yet anomalous associations between polymorphisms and traits (Terwilliger & Weiss, 1998; Wright *et al.*, 1999). Nevertheless, in terms of disease diagnosis these developments mean that in the future certain patterns of SNPs can be related to a specific disease predisposition, or susceptibility to a certain drug.

In summary, this section has provided an introductory glance at the current status of genetic polymorphisms related to pathogenesis. From a diagnostic perspective the trends observed here mean that disease diagnosis will become much more focussed on individual genetic variability since both susceptibility to and treatment of disease is highly dependent on it (Biesecker & Marteau, 1999). Once established it may allow for better prediction and treatment of a disease condition, but also establish the disease-cause more rapidly. The latter is important especially in microbiological pathogenesis, for instance for the diagnosis of tuberculosis, which can take up to two weeks from patient sampling to actual diagnosis using traditional culturing methods for *M. tuberculosis*. Novel nucleic acid screening methods may allow more rapid detection of disease cause, but also which specific variant one is infected with and whether or not the infecting agent is antibiotic-resistant. With respect to human genetic disease they may facilitate the design of individual-tailored drugs, or drug-treatment, and improve one's quality of life by establishing which disease one may be predisposed to before it actually emerges. The next section will therefore analyse in more detail the current technologies available for genetic screening and what their advantages and drawbacks are.

1.2 Methods for Mutation Detection

The previous section briefly outlined the immense complexity that exists at the genetic level of common infectious and heritable diseases. The genetic heterogeneity of pathogenic micro-organisms and humans that defines virulence and disease-

susceptibility remains, to a large extent, to be uncovered. Research at present is faced with a number of practical obstacles, in particular with respect to the diagnosis of and association with a certain disease. Increasing emphasis is put on the ability to conduct large-scale, high-throughput experiments in the search for clinically relevant SNPs. The development of genetic diagnostics (or genetic screening) is moving in the same direction, but with even greater additional emphasis on speed, low cost and ease-of-use of the methods involved. As will be discussed further below, current methodologies available for genetic screening are largely unsuitable for diagnostic applications because they do not fulfil these requirements. However, genetic screening technologies are subject to intensive development. Several novel approaches have already begun to emerge, addressing the issues of cost and practicality thus providing a hopeful perspective on the progress in medical diagnostics.

A number of different methods are available to analyse and screen DNA sequences for the presence of mutations (summarised in table 1.1). Most of these rely on PCR to provide the required amplification of the target sequence in order to produce sufficient quantities for further analysis. PCR coupled to automated dideoxy sequencing remains a popular research tool, but a number of complementary techniques are available to reduce the workload and cost such an approach produces. These technologies essentially screen the PCR products for the presence or absence of mutations, reducing the number of required sequencing runs to characterise a particular mutation further. Significant differences are not only apparent in the methodologies themselves but also in the results they produce. Some techniques merely detect whether a mutation is present or not but do not reveal the exact identity or location of the mutation. Other techniques provide the exact location and identity of a mutation but may be relatively expensive or difficult to apply for high-throughput screening. The latter is becoming more important due to the requirements of human-based linkage and association studies, used to establish whether a particular polymorphism has clinical relevance or not (discussed previously). Finally, there are methods that require the location and identity of the mutation to be established beforehand. These methods are not only useful for association studies but also for diagnostics.

With respect to diagnostic assays, increasing emphasis is being put on the elimination of PCR (Eng & Vijg, 1997, Whitcombe *et al.*, 1998) and the associated

Table 1.1: Advantages and disadvantages of several mutation detection methods.

Method	Type	Main Advantages	Main Disadvantages
Dideoxy Sequencing	PCR screening	• Highly accurate • Amenable to high-throughput analysis • High sensitivity	• Very expensive • Laborious
MALDI-TOF	PCR-screening	• Highly sensitive • High level of accuracy • No sequencing required	• Only short DNA fragments can be analysed • Very expensive • Miniaturisation problematic
SSCP/HA	PCR screening	• Simple to perform • Amenable to high-throughput analysis • Less expensive • Detects novel mutations	• Target-dependent optimisation often required • Fragments must be 200 bp or less • Requires post-sequencing step • Variable sensitivity
DGGE	PCR screening	• Can handle slightly longer fragments than SSCP • High sensitivity • Detects novel mutations	• Fragment must be less than 600 bp • Requires post-sequencing step • Variable sensitivity • Expensive equipment required for temperature gradient • Specialised gels required
Chemical / Enzymatic cleavage	PCR screening	• Large sequences (1-2 kb) can be analysed	• Chemical cleavage: hazardous chemicals involved • Requires post-sequencing step • More laborious than other screening methods
Mini-sequencing	PCR screening	• Highly sensitive • Amenable to high-throughput analyis • No (Dideoxy) sequencing required	• Laborious, expensive • Does not detect novel mutations
ARMS	Modified PCR	• Sequencing not required • Multiplexing possible • Amenable to high-throughput analysis	• Target-dependent optimisation required • Does not detect novel mutations
Invader/SDA/ RCA	PCR-independent	• Isothermal • Simple to perform • No sequencing required • Amenable to high-throughput analysis • Circumvents PCR & Taq polymerase patents	• Not extensively/independently tested • Does not detect novel mutations

requirements for thermal cycling equipment (expensive), technical know-how (expensive, impractical) and subsequent PCR based mutation screening methods (impractical and can be expensive). Instead, focus is shifting towards applications that are rapid, cost-effective detection methods with minimal practical requirements, high sensitivity, high-throughput and a high level of reproducibility.

1.2.1 Methods that screen standard PCR fragments for novel and/or existing mutations

Single Strand Conformation Polymorphism (SSCP) and Heteroduplex Analysis (HA)

SSCP relies on the observation that, under certain conditions, ssDNA molecules form sequence-dependent tertiary structures due to intramolecular base pairing interactions. This property is exploited by applying an amplified target sequence (using PCR) in a denatured state to a non-denaturing gel. Even a single base change may alter the resulting conformation of the ssDNA molecules, and this is reflected by a change in band mobility (Liu *et al.*, 2000, Orita *et al.*, 1989). The technique has the advantage that it is relatively simple to perform, although it should be noted that to achieve the highest sensitivity and reproducibility, rigorous temperature control during electrophoresis may be required. Temperature is important because it is known to affect the tertiary structures of ssDNA molecules. Other conditions such as buffer pH and the level of gel cross-linking may also require optimisation steps in order to achieve mutation detection. SSCP can only be used to directly detect the presence but not the identity nor the location of a mutation in the PCR fragment. Nevertheless, band-shifted fragments may be isolated, re-amplified and sequenced to establish this (Suzuki *et al.*, 1991). This makes the technique however more labour intensive, as well as expensive. Finally, reported sensitivities of the technique vary between 60%-90%, indicating that the method is not able to detect all mutations, presumably because not all mutations cause a significant change in DNA conformation.

A substantial number of studies have used SSCP to detect mutations in a variety of disease-related genes, including those involved in familial hypercholesterolemia (Humphries *et al.*, 1997), myopathy (Nowak *et al.*, 1999) and the CCR5 receptor involved in HIV1 mediated pathogenesis (Dean *et al.*, 1996). Often SSCP is used in

conjunction with a related technique referred to as heteroduplex analysis (HA, refer for instance to Dean *et al.*, 1996). This can be carried out on the same gel used for SSCP analysis, since a significant portion of the denatured PCR fragments will re-anneal to form duplexes. In heterozygous persons heteroduplex formation occurs when a wild-type strand re-anneals to a mutation-containing strand. Again the technique relies on differences in gel-mobility, as heteroduplexes tend to migrate slower through the gel than homoduplexes.

Denaturing Gradient Gel Electrophoresis (DGGE)

DGGE (Fischer & Lerman, 1979) uses gel-electrophoresis where a gradient of denaturant is applied to the gel. When an amplified PCR fragment is applied, it will experience an increase in temperature or concentration of denaturing solvent whilst migrating through the gel matrix. This ultimately results in a denaturation event, which halts further migration through the gel of the DNA. The gel-location where melting occurs is highly sensitive to the DNA sequence and thus forms the basis of discrimination between wild-type and mutant targets. As with SSCP/HA, fragments can be recovered, re-amplified and sequenced after electrophoresis, allowing the identity and location of a certain mutation to be established. The technique has been successfully applied in identifying a number of potential disease-related mutations including, for instance, CFTR, p53 and β-globin genes (Cremonesi *et al.*, 1999, Day *et al.*, 1999, Gelfi *et al.*, 1997).

The method involves relatively complicated procedures for optimisation and in the case of a temperature gradient additional equipment is required. Also, optimisation appears to be required for each individual fragment and the length of the fragments used is limited to 600 bp or less.

Chemical and Enzymatic cleavage of mutations

Several methods have been developed that exploit chemicals or enzymes that act specifically on mutations. Chemicals such as hydroxylamine and osmium tetroxide attack specific bases when these are exposed to solvent (i.e. not hydrogen bonded). This property has allowed them to be used for the detection of mutations. However,

the most significant drawback in this approach is the high level of toxicity of the chemicals involved and enzymatic cleavage of mutations is therefore the preferred method. Enzymes such as T4 endonuclease VII recognise minor disruptions in heteroduplex DNA and therefore require wild-type DNA to be mixed with the mutation-containing DNA before the assay to generate these heteroduplex molecules (Youil *et al.*, 1995). Once formed, incubation with T4 endonuclease VII leads to a double stranded cleavage event (if a mutation is present) and the resulting fragments can be analysed on gel where the cleaved fragments migrate faster than wild-type fragments. This technique has been used to detect a number of mutations in different genes, including BRCA1, β-globin (Youil *et al.*, 1996) and p53 (Giunta *et al.*, 1996).

Although much larger fragments can be analysed than with SSCP or DGGE and an approximate indication of the mutation location can be obtained, establishing the exact identity of the mutation requires additional sequencing and is due to the presence of the extra cleavage step more laborious.

1.2.2 PCR based methods that screen for known mutations

Amplification Refractory Mutation System (ARMS)

In ARMS (Newton *et al.*, 1989, and Fig. 1.1a) either a first wild-type specific or a mutant-specific primer is coupled to a second common PCR primer (i.e. the first primer binds to one target strand and the common primer to the complementary target strand). In the case of homozygous wild-type DNA exponential amplification is only achieved when the wild-type primer is included, whereas homozygous mutant DNA will only yield a PCR product when the mutant primer is included. Heterozygous material will yield a product in both cases. This technique is probably the most widely applied to screen PCR products for known mutations and is adaptable to multiplexing. Screens for multiple mutations involved in cystic fibrosis (Bradley *et al.*, 1998, Ferrie *et al.*, 1992) and colorectal carcinoma (Fox *et al.*, 1998) have been described, illustrating the power of the technique. High levels of sensitivity may be obtained, although careful, target dependent optimisation is usually required (as in all PCR based applications). A closely related technique involves the addition of a so-called competitive oligonucleotide primer (COP) to a

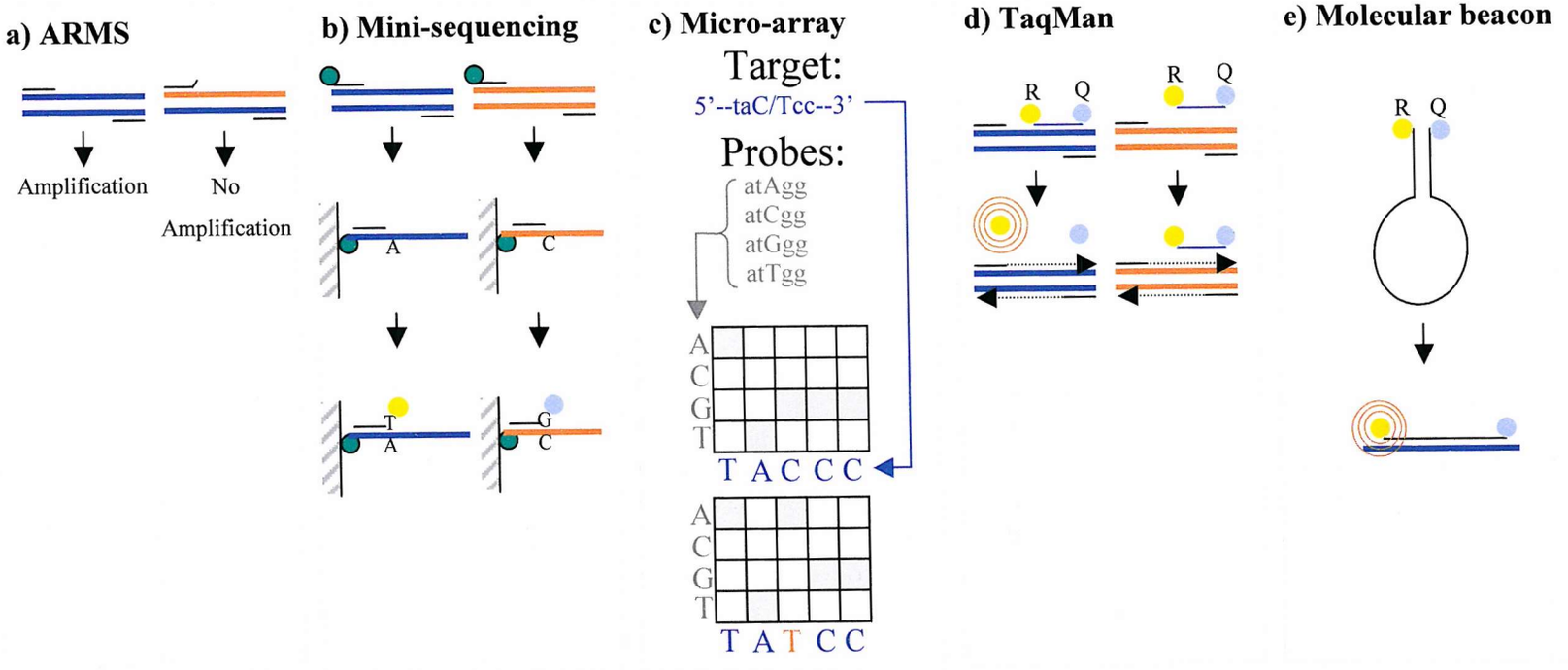

a) ARMS
b) Mini-sequencing
c) Micro-array
d) TaqMan
e) Molecular beacon
Amplification
No Amplification
A
C
T
A
G
C
Target:
5'--taC/Tcc--3'
Probes:
atAgg
atCgg
atGgg
atTgg
A
C
G
T
T A C C C
A
C
G
T
T A T C C
R Q
R Q
R Q

Fig. 1.1 Nucleic acid amplification technologies

a) ARMS. Allele-specific primers are designed such that the 3' end of the primers exactly match the sequence of the wild-type target, but produce a mismatch when hybridised to a mutant allele (or *vice versa*). The presence of this mutation will prevent efficient extension by the polymerase, inhibiting further amplification. b) Solid phase mini-sequencing. After PCR with biotinylated primers, the amplicon is captured on solid support and washed. A primer is then added with one labelled nucleotide and incorporation of this labelled nucleotide thus generates a signal only when a complementary base is present in the target. By using different labels one can detect whether wild-type or mutant target was present (and even which specific base was present). c) Microarray (adapted from Chee *et al.*, 1996). Each of the tile in the array contains thousands of copies of a probe specific for one of the four possible bases (A, C, G or T shown to the left) at each position of a target sequence. A single base change will cause a change in hybridisation signal as a different probe will now become complementary to the altered base as shown in the bottom array. Note that the probes shown at the top are only for querying the position where the mutation occurred in this example. The other bases in the target sequence (*i.e.* all other tiles in the array) were queried by probes specific for that location. d) TaqMan. A probe containing two fluorophores (a reporter (R) and quencher (Q)) specific for wild-type (or mutant) is added to a PCR reaction. The reported fluorophore remains quenched (through FRET) due to the close proximity of the quencher fluorophore unless it hybridises to a complementary target site. When this occurs, the Taq polymerase exonuclease activity will degrade the R/Q-oligonucleotide while it is extending the PCR primers, which releases the reporter fluorophore and it starts to fluoresce. e) Molecular beacon. A molecular beacon is a hairpin and like the TaqMan system, contains a reporter (R) and quencher (Q) fluorophore. The hairpin is complementary to a target sequence (wild-type or mutant specific) and will therefore only open-up in the presence of a complementary target strand. This produces a fluorescent signal.

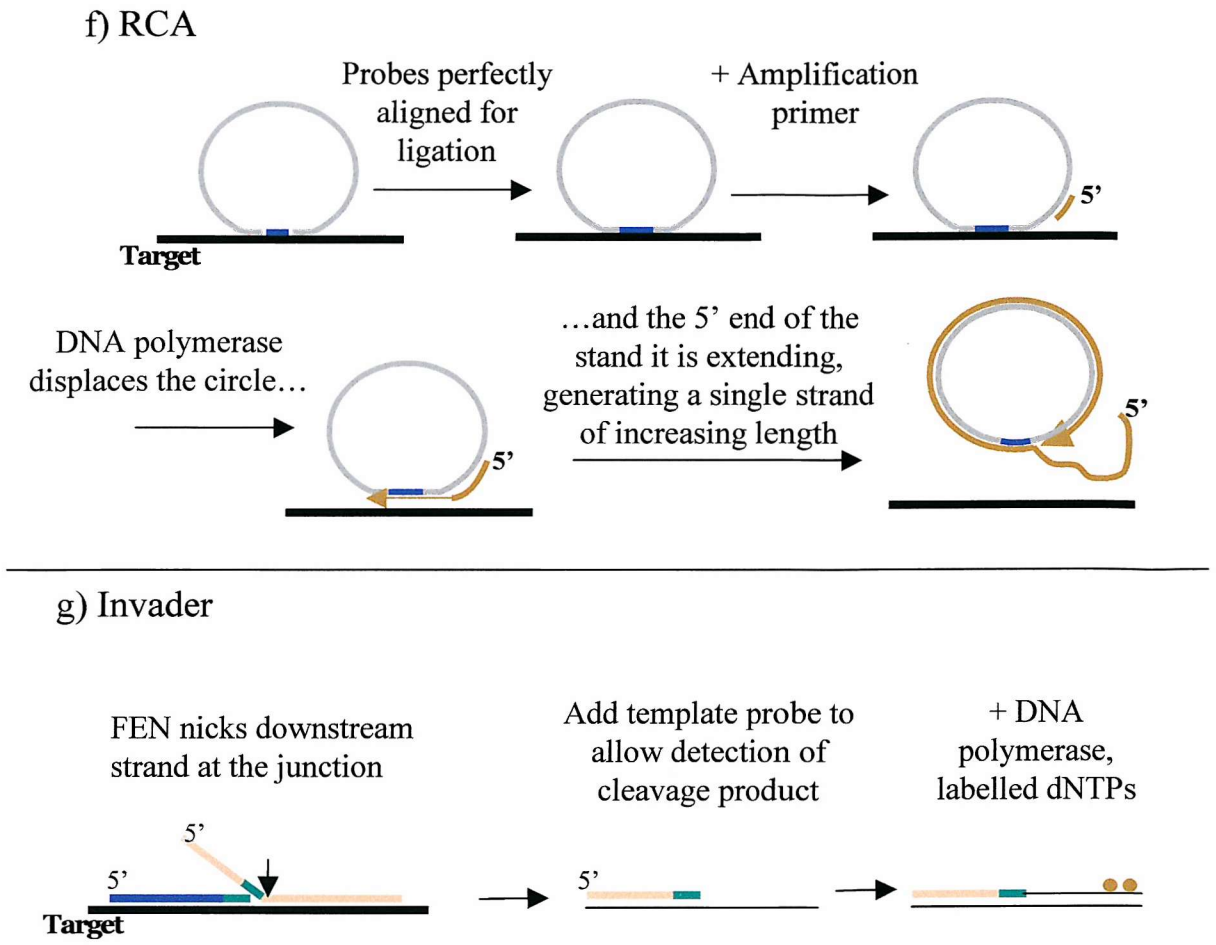


Fig. 1.1 (continued). Nucleic acid amplification technologies

f) RCA. Two probes are added to a target sequence. The blue coloured probe is a "gap-fill" oligonucleotide, which is specific for wild-type in the example above. The other, green coloured probe anneals to the target at its 3' and 5' ends, generating a looped molecule with a gap between its 3' and 5' ends where the gap-fill probe binds. Only if the sequence of the gap-fill oligo is perfectly matched with the target at its 3' and 5' end can ligation occur to create the Padlock probe. The third oligonucleotide then binds to the probe and through DNA polymerase-mediated strand-displacement of the target and itself a long single strand of DNA is generated by continuous and repeated replication of the circle. The reaction can only succeed if the blue and green probes align perfectly for ligation. Thus, if a mismatch is present at the junction between the green and blue probes, no ligation can occur, enabling discrimination between wild-type and specific mutatations. The signal is generated by incorporation of for instance, a fluorescent labelled dNTPs into the growing strand.

g) Invader. An upstream (blue/green) and downstream (orange/green) oligonucleotide are added to a target sequence. FEN cleaves the flapped structure only if the 3' overlapping bases are identical in both strands allowing further detection of the cleaved product. Repeated cleavage is achieved by performing the assay at a temperature close to the melting temperature of the downstream strand/target duplex.

PCR reaction where, as in ARMS, one of the primers and the COP cover the region where a certain mutation may be expected. The COP is selective for the mutation so that both primers are competing for binding. By differentially labelling the two competing primers (e.g. different fluorophores) one can discriminate between homozygous wild-type, heterozygous and homozygous mutant targets since each target will generate a different fluorescent colour. Thus, the application of COPs allows ARMS to be performed in a single tube. Using a PNA (Peptide Nucleic Acids, described below) oligomer as a COP, the deep venous thrombosis-causing SNP in the Factor V gene (Factor V Leiden) could be diagnosed in this manner (Behn & Schuermann, 1998b).

Solid Phase Mini-sequencing

In mini-sequencing (Syvanen *et al.*, 1990, Fig. 1.1b) one PCR primer is labelled with a biotin group to allow capture on a solid support coated with streptavidin. After PCR, the products are denatured, captured and the non-captured material washed away. A third oligonucleotide is then added, which binds to the trapped PCR strand such that its 3'-end is positioned next to the base in the PCR strand where a mutation can be expected. By adding either a labelled dNTP complementary to wild-type or to the mutation, the presence and zygosity of the mutation can be established. This method has been widely applied for the discovery of mutations, including ones involved in aspartylglucosaminuria (Hietala *et al.*, 1996), deep venous thrombosis (Pecheniuk *et al.*, 2000), mitochondrial disorders (Tully *et al.*, 1996) and is relatively easy to convert to chip-format (discussed below) (Syvanen, 1999).

DNA micro array technology: Mutation detection using sequence (or allele)-specific oligonucleotide probes (SSOPs or ASOs).

One of the oldest methods known to detect mutations is hybridisation with SSOPs (sometimes also referred to as allele-specific oligonucleotides, ASOs). It has made a remarkable comeback in the past few years due to advent of DNA micro-arrays. The latter are mostly based on binding of PCR products to SSOPs. Another well-known application is filter hybridisation (dot-blots), which is essentially the forerunner of the DNA chip. Filter hybridisation involves a short oligonucleotide, which under

appropriate conditions anneals only to its complementary sequence present in the PCR product but not to a mutation-carrying fragment. The target to be detected can be labelled prior to application (for instance with a fluorescent label). By additional labelling of the probe, for instance with biotin, and binding it to a streptavidin coated membrane, or immobilising it to a nylon membrane through oven baking or UV cross-linking, the probe is attached to a solid support. Labelled, denatured target is then added and under controlled conditions annealed to the probe. This approach has significant disadvantages however, mainly caused by the fact that each probe-target complex is likely to possess a different melting temperature. The latter stands in the way of generating a generalised protocol. Also gel-verification of successful PCR amplification is required prior to SSO treatment. As mentioned before however, micro-array technology depends on SSOs and major efforts have been and are being put into overcoming these practical hurdles. In conventional DNA chip designs, oligonucleotides are also attached to a solid support (usually silica-based) but at very high density (micro-arrays are often referred to as high-density DNA arrays). This allows for an enormous increase in parallel screening capacity as a large number of genes (or SNPs) can be scanned at once. However, it should be noted that a major bottleneck limiting the rapid high-throughput screening potential of this technology remains PCR. Screening of large numbers of genes means that these genes need to be amplified before micro-array screening. High-throughput, parallel performance of PCR remains highly problematic since multiplex PCR (i.e. amplification of different target sequences in a single tube) becomes less efficient and more prone to yield non-specific amplification when the extent of multiplexing increases. In addition, post-PCR labelling with fluorophores is required for most chip formats to allow detection.

Large-scale SNP mapping has been performed on the human genome using micro-array technology (Wang *et al.*, 1998), but the detection of SNPs required a substantially complicated protocol (involving post PCR labelling and extensive hybridisation (15h)) and took two days to perform. The chip format used involved a silicon based solid surface divided into separate sections. Each section contained a large number of the same oligonucleotide, containing either wild-type or a single point mutation at a defined position. Hybridisation success was translated into a measurable, visual signal, which was quantitated by computer analysis of the signal intensity of each section (Fig. 1.1c shows an example). Although far less

experimental handling was required compared to conventional PCR-based sequencing, for this approach to be of diagnostic use, major improvements are still required. Also, a significant error rate was apparent as the chip yielded 12% false-positives, requiring careful re-examination of results obtained afterwards. Some of these errors were random (eliminated by repeating the experiment) but some were recurring and apparently dependent on the methods involved. Other studies, using essentially the same chip design and protocols, confirmed that additional verification (through sequencing) of candidate SNPs was essential (Cargill *et al.*, 1999, Halushka *et al.*, 1999, Wang *et al.*, 1998). To reduce the number of experimental steps, substantial multiplexing of PCR was required (*i.e.* the performance of multiple PCR reactions in the same tube), which has technical problems of its own, usually resulting in a reduction of PCR efficiency as the extent of multiplexing increases. Finally, some methods for producing chips (*i.e.* those made by Affremetrix Inc.) are limited due to the fact that oligonucleotide synthesis on the chip is allowed in only one direction (3'→ 5'), and that only very short oligonucleotides can be used.

Thus far, the most reliable results for mutation screening have been obtained with chips designed for well-studied targets including BRCA1 (involved in breast cancer), cystic fibrosis and the human mitochondrial genome (Chee *et al.*, 1996, Cronin *et al.*, 1996, Hacia *et al.*, 1996). These chips were based on sequences already available, reducing the number of required control experiments.

In spite of current problems, the use of micro arrays offers an enormous improvement in assay throughput and different chip designs attempting to solve those problems are becoming available at rapid rates. (*e.g.* Gilles *et al.*, 1999, Steemers *et al.*, 2000).

Taqman 5' nuclease assay

A widely used approach for PCR mutation detection involves the binding of a DNA oligonucleotide conjugated with two different fluorophores to the region of the PCR amplicon (Fig. 1.1d). The oligonucleotide will bind only to the amplicon if its sequence fully matches that of the amplicon. When this occurs the 5' exonuclease activity of the Taq polymerase degrades the oligonucleotide during the successive rounds of amplification, releasing the two fluorophores. In the oligonucleotide-bound form the two fluorophores are quenched (fluorescence resonance energy transfer),

and degradation of the oligonucleotide therefore generates a fluorescent signal (Holland *et al.*, 1991, Fig. 1.1d). Combinations of spectrally different fluorophores can be used for oligonucleotides designed against either wild-type or mutant sequences, allowing for the discrimination between homozygous and heterozygous material in a single tube.

The method has been successfully applied for the detection of mutations in the NAD(P)H quinone oxidoreductase (involved in drug metabolism), Factor V and CCR5 genes (Shi *et al.*, 1999, Happich *et al.*, 1999, Yuan *et al.*, 2000), and for genotyping of the influenza virus and chlamydiaceae (Schweiger *et al.*, 2000, Reiss *et al.*, 1998). Also, Perkin-Elmer currently offers a TaqMan kit for high-throughput genotyping (Livak, 1999).

A drawback in the technique is the high cost of the reagents and equipment. Only a limited number of fluorophore combinations are available, limiting the extent of multiplexing that can be performed. Also, contamination has been reported as a potential source of error but even more, methods that reduce the chances of contamination affected the sensitivity of the assay (Corless *et al.*, 2000). Tapp *et al.* further reported that molecular beacons (discussed in the next section) appeared more sensitive for mismatches than the TaqMan assay when applied to the gene encoding the estrogen receptor (Tapp *et al.*, 2000).

Other developments

A substantial focus of research attention remains the further optimisation and simplification of the PCR itself. This includes for instance the ability to monitor the formation of PCR amplicons in real-time, using molecules such as a molecular beacon. The latter is a DNA hairpin with its loop complementary to the amplicon. The ends of the hairpin contain two different fluorophores, which through resonance energy transfer remain quenched until the hairpin binds to a complementary strand and opens up (Fig. 1.1e). The sensitivity of current molecular beacon designs are however low, and they usually do not detect the formed amplicon until 15-20 cycles or higher. Nevertheless, they have been successfully applied to detect the presence of mutations (and their zygosity) in PCR amplicons (Marras *et al.*, 1999).

Also, efforts are being put in replacing the sequencing step many screening methods require by other methods. Matrix-assisted laser desorption-ionisation-time-

of-flight mass spectroscopy (MALDI-TOF MS) is increasing in popularity for the analysis of PCR products, although severe limitations are still attached to it (Griffin & Smith, 2000; Laken *et al.*, 1998). Its main advantage is its accuracy and it can be used for instance for mini-sequencing, and through modification of the PCR reaction design as a replacement for sequencing or gel-electrophoretic analysis. However, at present it is hindered by substantial costs, expertise requirements, the sheer size of the equipment involved (hindering miniaturisation), inability to perform high-throughput analyses and the fact that only very short DNA fragments can be analysed. Although often heralded as a technology of the future, unless these major limitations are overcome large-scale implementation of this technique remains unlikely in a diagnostic environment. Alternative methods that can be used as replacement for the sequencing and/or electrophoresis step are capillary electrophoresis and HPLC, but both have similar limitations as MS.

A number of studies have further shown that the PCR reaction time can be significantly reduced. Roche Diagnostics currently sells a glass-capillary based thermal cycler (LightCycler^{TM}), which can achieve rapid cooling rates (>10°C/sec) and reducing the time needed for a 30 cycle reaction to less than an hour. Even more impressive was a report by Kopp *et al.* (1998), who, using a continuous flowing PCR reaction through a temperature gradient on a glass chip, were able to reduce the time required for a 20-cycle PCR reaction to less than 90 seconds (Kopp *et al.*, 1998).

An alternative improvement in the detection of mutations lies in the modification of the oligonucleotides used in genetic screening methods. A significant development in this area has been the advent of peptide (or polyamine) nucleic acids (PNA). PNA consist of a pseudo peptide backbone (N-(2-aminoethyl)glycine units) coupled to standard Watson & Crick bases (Fig. 1.2b) and has proven a remarkable molecule in that it retains normal Watson-Crick base pairing rules, and assumes a similar double helical structure as DNA when bound to a complementary PNA, DNA or RNA strand (Brown *et al.*, 1994, Eriksson & Nielsen, 1996, Nielsen *et al.*, 1991, Rasmussen *et al.*, 1997). One of the most important differences between PNA and DNA is the absence of a negatively charged backbone and the absence of a sugar ring. The lack of electrostatic repulsion is thought to be the main reason for the observed higher stability of PNA-containing duplexes when compared to their DNA-

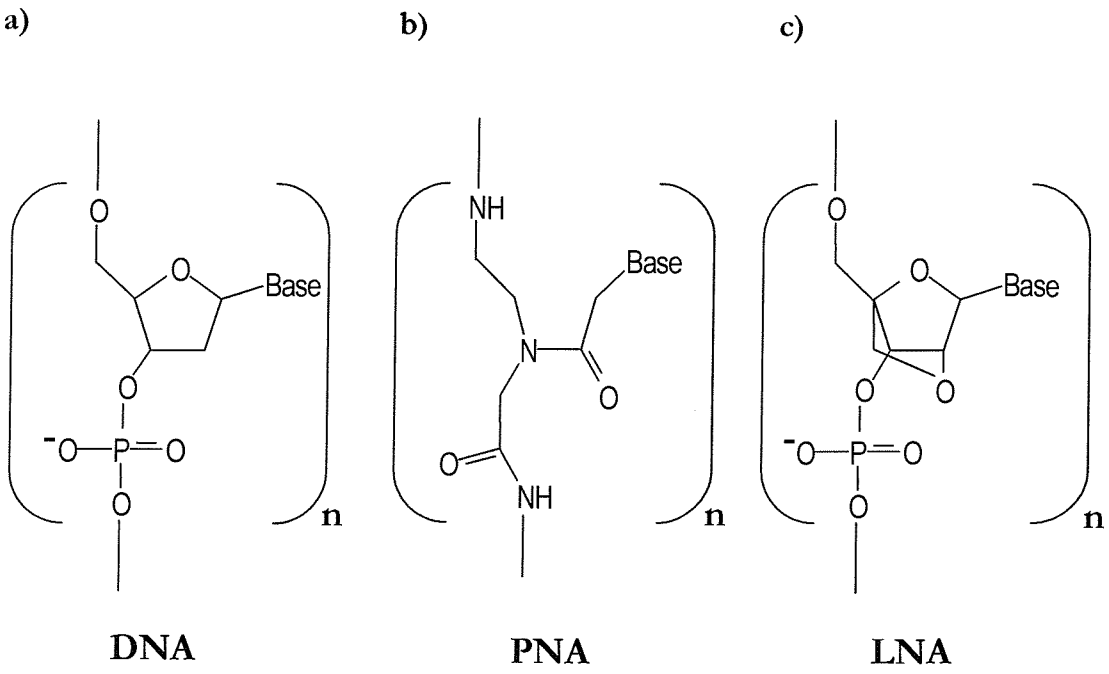


Fig. 1.2 Differences in chemical structure between DNA, Peptide Nucleic Acids (PNA) and locked nucleic acids (LNA).

equivalents. Another important feature of PNA-DNA duplexes is their increased sensitivity for mismatches and other mutations (see for instance: Eriksson & Nielsen, 1996, Igloi, 1998) in that the melting temperature drops substantially more than is the case for DNA-DNA duplexes when confronted with a mutation. This property has already been exploited in PCR based assays, both in SSCP and ARMS where a PNA COP was used to detect the Factor V Leiden SNP (Behn & Schuermann, 1998a, Behn & Schuermann, 1998b). Although less well studied than PNA, a RNA analogue named Locked Nucleic Acid (LNA, Fig. 1.2c) also has been reported as more sensitive for mutations and has been applied for Factor V Leiden detection (Nielsen *et al.*, 1999, Orum *et al.*, 1999). LNA contains a methylene linker connecting the 2' oxygen with the 4' carbon in the furanose ring, which presumably reduces the conformational freedom of the sugar ring (Koshkin *et al.*, 1998, Nielsen *et al.*, 2000).

Finally, improvements in the sensitivity of the end-detection system can have important value as it would reduce the level of required target amplification. Molecular beacons (described above) constitute one example of an improved, more sensitive end-detection system compared to conventional electrophoretic analysis. Another notable development in this area was described recently by Jenison *et al.* (2001). They developed an optically active solid surface to which a target-capture oligonucleotide was attached. After annealing of the target to this capture probe, a biotin-labelled detection oligonucleotide was added that bound to another region of the target. Addition of horseradish peroxidase-labelled anti-biotin, washing and subsequent addition of enzyme substrate resulted in the formation of a precipitate that formed a film on the surface. This caused an alteration in the reflectance pattern of light from the solid surface, which was detected by using a CCD (charged-coupled device) camera. Using this technique, as few as ~60,000 copies of target could be detected (Jenison *et al.*, 2001).

In summary, PCR derived applications for mutation screening are available and used widely in research laboratories and major efforts are ongoing improving PCR-based mutation detection/screening methods. Nevertheless, for diagnostic applications more robustness is required than PCR can currently offer in particular with respect to the high incidence of false positive and negative results that are often observed and their requirement for technical know-how and specialised laboratories

(Bustin & Dorudi, 1998; Whitcombe *et al.*, 1998). The following section will therefore outline some of the alternative technologies, which may have potential as PCR replacements for direct mutation screening. It should be noted that most recorded successes of the technologies described below originate from the labs they were created, and should therefore be regarded as potential mutation detection methods until independent studies have verified their potential.

1.2.3 Isothermal Mutation Detection methods

Rolling Circle Amplification (RCA)

Rolling circle amplification (Fig. 1.1f, Lizardi *et al.*, 1998) is a strand-displacement driven method, which means that amplification success depends on the strand displacing activity certain DNA polymerases possess. Whilst DNA synthesis is proceeding, these polymerases are capable of melting duplex regions that they encounter ahead of them. The initial stage of the reaction involves the formation of a so-called Padlock probe, which is a circularised single stranded DNA molecule annealed to a specific target sequence. The formation of this probe is mediated by ligating a short DNA oligonucleotide to both ends of a longer DNA oligonucleotide. As shown in Fig. 1.1f, the latter oligo binds to a target sequence at its 3' and 5' end, so that the intervening sequence forms a large loop. A small gap exists between the 3' end 5' ends of this probe and this is where the second probe binds to the target. Subsequent ligation is then dependent on proper base pairing at the ends of this oligonucleotide. If a mismatch occurs where the two probes meet, ligation (and subsequent Padlock probe formation) is inhibited, allowing for discrimination between wild-type and mutant sequences. Amplification is achieved by adding a third DNA oligonucleotide, which binds to another part of the Padlock probe. DNA polymerase extension then ensues displacing the target and, once the polymerase has moved round the circle, displacing the replication primer itself. This process generates a continuously growing DNA chain. By incorporating labelled dNTPs both sensitive and accurate detection of point mutations in the cystic fibrosis gene was achieved in less than two hours (Lizardi *et al.*, 1998).

Invader

The invader assay (Fig. 1.1g) is based on an endonuclease (Flap endonuclease, FEN) that recognises and cleaves flapped DNA structures. Two oligonucleotides (upstream and downstream strands) anneal to a target sequence flush against each other but with one containing a 5' single-stranded flap at the point where the two probes meet. However, cleavage activity of the FEN is dependent on the presence of sequence-overlap between the two probes at the junction. Experiments indicated that FEN cleaved the flapped probe at the position where the overlap in sequence between the two probes ended, and a single base change at that position was sufficient to eliminate cleavage activity. The prevention of cleavage was further independent of the type of mismatch (Lyamichev *et al.*, 1999). Once the flap sequence is cleaved, it can bind to and extended along (by a DNA polymerase) an additional oligonucleotide, incorporating labelled dNTPs to allow further detection. Repeated cleavage of the flapped strand can be achieved by increasing the assay temperature close to that of the melting temperature (Tm) of the downstream strand/target duplex, which stimulates rapid dissociation and annealing of the downstream strand. Used in this format, the Factor V Leiden SNP and the ΔF508 CFTR deletion described earlier were detected in blood samples of affected patients (Lyamichev *et al.*, 1999).

1.3 Signal Mediated Amplification of RNA Technology (SMART)

The previous sections showed that there is a need for the development of novel diagnostic detection methods. As pointed out, reliability, cost, user-friendliness and sensitivity remain issues that continue to hamper large-scale implementation of nucleic acid detection methods in medical environments. Cytocell Limited therefore initiated the development of a novel, isothermal nucleic acid amplification technology: the signal mediated amplification of RNA technology or SMART, invented by Cytocell Limited in 1991 (Cardy & Delnatte, 1991). As will be further outlined below, SMART is unique in that amplification depends on the formation of a three-way DNA junction. This unstable structure was considered to be more sensitive to the introduction of mismatches than a standard DNA duplex. This potentially makes it more useful for mutation detection. In addition, it is an isothermal technology, and is not reliant on PCR to achieve the required sensitivity levels. Moreover, it does not involve amplification of the target sequence itself but a sequence unrelated to the target, which can be optimised to yield the highest and most consistent levels of amplification.

1.3.1 Description of the technology

1) The first amplification stage.

One of the characteristic features of SMART is the target-dependent formation of a three-way DNA junction (TWJ) as outlined in Fig. 1.3. This is achieved by adding two DNA probes, called the extension probe and the template probe, to a specific target sequence (steps 1 and 2, Fig. 1.3). These probes contain so-called foot-regions, which are complementary to adjacent regions of the target. In addition, both probes contain a short complementary sequence, which in the presence of target results in the formation of an overlap duplex (step 2). The important feature of the technology is that this overlap duplex is not stable in the absence of target, and is therefore necessarily short.

The template probe contains two additional sequence features, namely a 17 base T7 RNA polymerase (RNAP) promoter sequence and a transcription template. SMART exploits the fact that T7 RNA polymerase will only bind to a double-stranded T7 RNAP promoter. Thus, this enzyme is not active on the template probe strand alone. A DNA polymerase is then used to extend the 3' end of the extension probe (step 3) thereby producing a double stranded promoter, hence activating multiple rounds of RNA transcription by the T7 RNAP (steps 4 and 5). Thus, a first level of amplification is achieved as multiple copies (typically 100) of RNA are produced from one junction molecule. However, 100-fold amplification is not sufficient in those cases where low (i.e. sub atomolar) concentrations of target are concerned. To increase the sensitivity of the technology an additional amplification step(s) is required, which is outlined in Fig. 1.4 and discussed below.

2) The second amplification stage

To amplify the RNA produced during the first stage, two additional probes are added, called amplification probe 1 and amplification probe 2. Both probes contain three distinct sequence regions: a RNA binding, a T7 RNAP promoter and a transcription template sequence. Amplification probe 1 binds to the RNA produced from the junction such that the 3'-end of the RNA is extended by the DNA polymerase, yielding a second active transcription unit. This unit produces a second, unique RNA sequence, which binds to the second amplification probe followed by DNA polymerase extension and T7 RNAP transcription. A critical feature of the

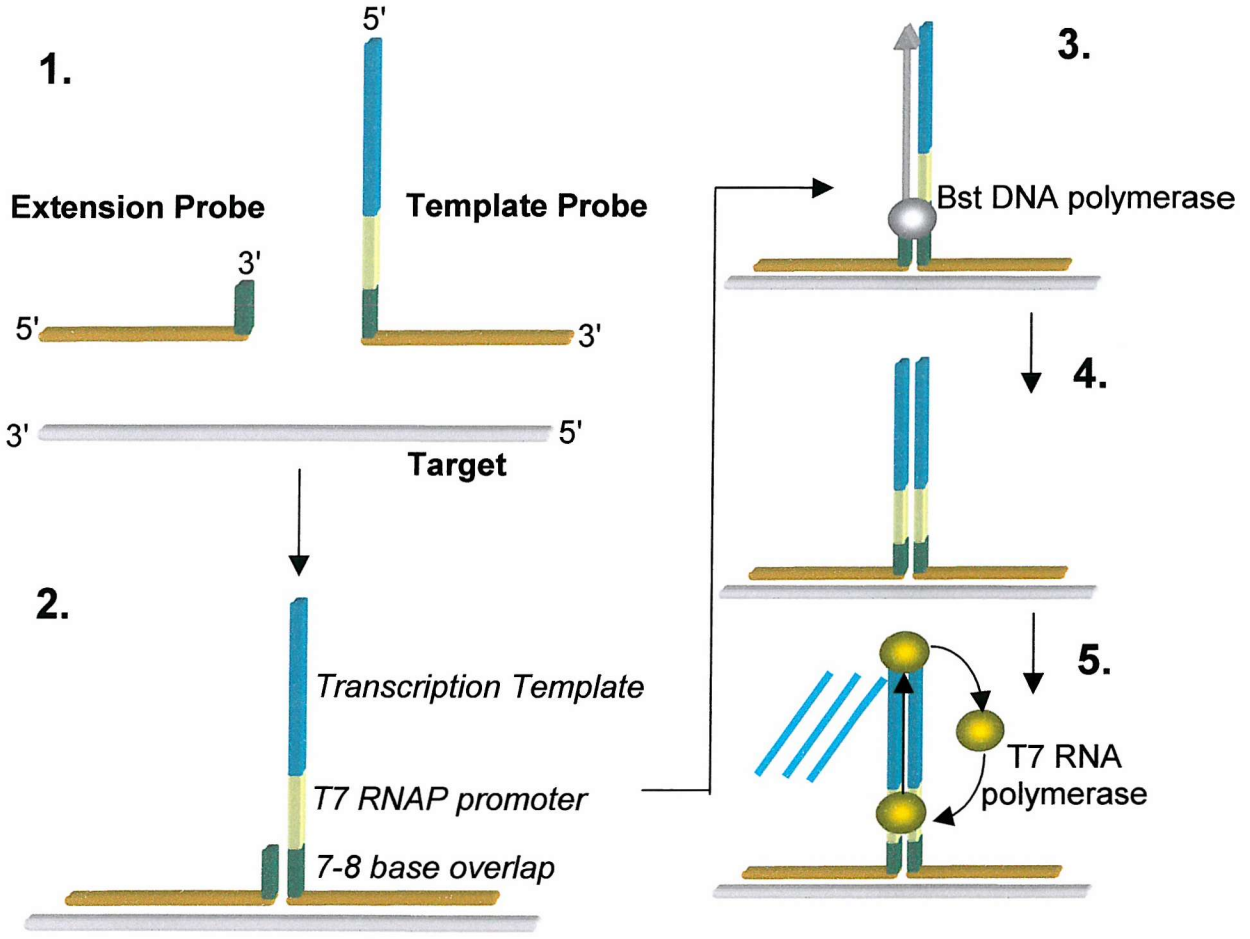


Fig. 1.3 Schematic overview of the first amplification stage in SMART. Binding of two probes to a target sequence facilitates the formation of the overlap region, which is unstable in the absence of target. Subsequent extension of the extension probe by a DNA polymerase yields a double stranded T7 RNA polymerase promoter. Only when this promoter is double stranded does the T7 RNA polymerase initiate processive transcription. Since one junction molecule yields multiple copies of RNA, the function of transcription is to yield a detectable signal (RNA) indicative of the presence of a specific target.

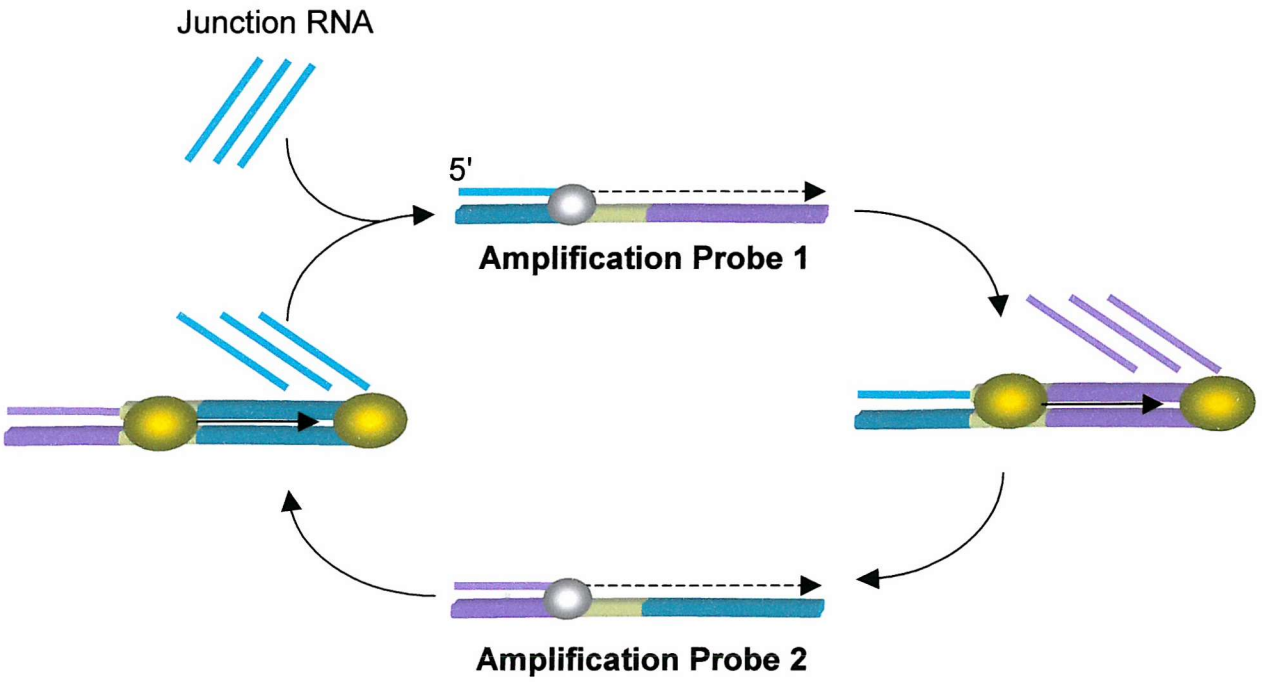


Fig. 1.4 Schematic overview of the second amplification stage in SMART.
The same symbols and colours are used as in Fig. 5a (e.g. the grey shaded circle represents the DNA polymerase). Both amplification probes contain 3 distinct regions: a RNA binding, a T7 RNAP promoter and a transcription sequence. Probe 2 yields the same RNA sequence as the RNA that is transcribed off the junction whereas probe 1 produces a different RNA sequence. Refer to text for further details.

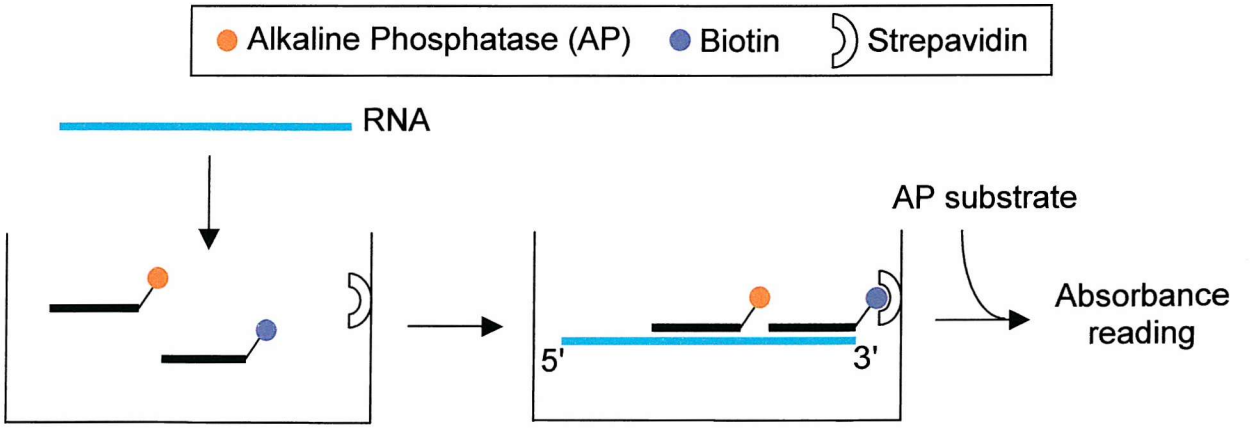


Fig 1.5 Simplified scheme for the detection of RNA. The SMART reaction mixture is added to a streptavidin-coated well, containing two labelled probes that are complementary to different parts of the RNA. One probe is labelled with biotin to allow the RNA to be trapped in the well. After trapping, excess of probes can be removed by a wash step. Chromogenic substrate is then added to quantify the amount of AP (and thus RNA) present.

second amplification probe is that its transcription template matches that of the template probe, thus yielding the same RNA sequence. Consequently, the RNA produced by the second transcription unit can bind again to amplification probe 1, generating an exponential RNA amplification cycle.

3) Detection of the RNA signal

Fig. 1.5 outlines the detection method for the RNA produced by SMART. Two probes, one labelled with biotin and one conjugated to an alkaline phosphatase (AP), are added together with the reaction mixture to a microtiter plate well, coated with streptavidin. The two probes bind to (different parts of) the RNA. The biotinylated probe acts to trap the RNA in the well through the strong binding of biotin to streptavidin so that unbound probes and remaining reaction constituents can be washed away from the well. Adding an appropriate substrate (e.g. 4-nitrophenyl phosphate, pNPP) for the AP yields a colour reaction that can be followed spectrophotometrically. The signal produced by the latter is directly proportional to the amount of AP, and therefore RNA, present in the well. Using a standard curve of synthetically prepared RNA it is possible to quantify the absolute amount of RNA.

1.3.2 Important features of SMART

The previous section detailed the processes involved in SMART. Several issues however, are apparent that may affect the technology. One critical point is the potential of background signal (noise) generation in the absence of target. Also, since the RNA produced acts as the signal for the presence of a specific nucleic acid sequence, its yield determines the sensitivity of the technology. Since a number of factors involved in SMART may affect both the signal to noise ratio and RNA yield, this section will analyse these factors in more detail.

1) The overlap region

One potential source of non-specific signal is the overlap region. The assay is therefore designed so that this mini-duplex does not form at the assay temperature. Any target-independent formation of this duplex in the absence of target, will give rise to a signal. The base composition of the overlap is therefore likely to play a crucial factor in controlling the background as it afffects its stability. The presence of

the DNA polymerase also act as a supplementary stabilising factor since the addition of only one or a few bases to the extension probe will increase the lifetime of these non-specific interactions and stimulate further, non-specific polymerisation. Ultimately this could result in the formation of a functional T7 RNAP promoter and the generation of (detectable) RNA in the absence of target. As indicated in Fig. 1.3, reducing the size of the overlap will prevent spurious extension events. However, it is here that the junction itself may prove to be a hindrance as a shorter overlap means that the DNA polymerase has to bind closer to the branch point. The presence of the two target-containing arms could potentially interfere with binding of the *Bacillus stearothermophilus* (Bst) DNA polymerase to the overlap region, reducing the efficiency or the sensitivity of SMART. It is not known whether DNA polymerase binding is inhibited by the structure of the TWJ, although previous results obtained at Cytocell Limited have indicated that the overlap may not be shorter than 5 bp. Usually the best results (i.e. good yield with good signal to noise ratios) were obtained when the overlap comprised 7 to 8 bp. Co-crystal structures of Bst DNA polymerase with DNA bound to its active site (Fig. 1.6) have further revealed that a large portion (containing the active site) of the enzyme essentially wraps around its substrate (Kiefer *et al.*, 1998). The protein covered a region of approximately 10 bp, further substantiating the hypothesis that the short size of the overlap region, coupled to the close proximity of two additional DNA arms, may pose a steric barrier for the enzyme.

Nevertheless, at present a clear understanding of the relationship between TWJ structure and the RNA produced in SMART is largely lacking and requires further exploration as will be further outlined below. The situation is likely to be complex as multiple factors may be interacting at once, affecting both the overlap stability and DNA polymerase accessibility.

2) T7 RNA polymerase

The T7 RNA polymerase is an enzyme with wide-ranging applications in the life sciences. Processive, commercially available, relatively inexpensive and a single-subunit enzyme are the main properties that make this enzyme an attractive one for commercial exploitation. In addition, several nucleic acid amplification technologies (e.g. TMA (transcription mediated amplification), NASBA (nucleic acid sequence based amplification) (Compton, 1997, Kwoh *et al.*, 1989)), as well as SMART,

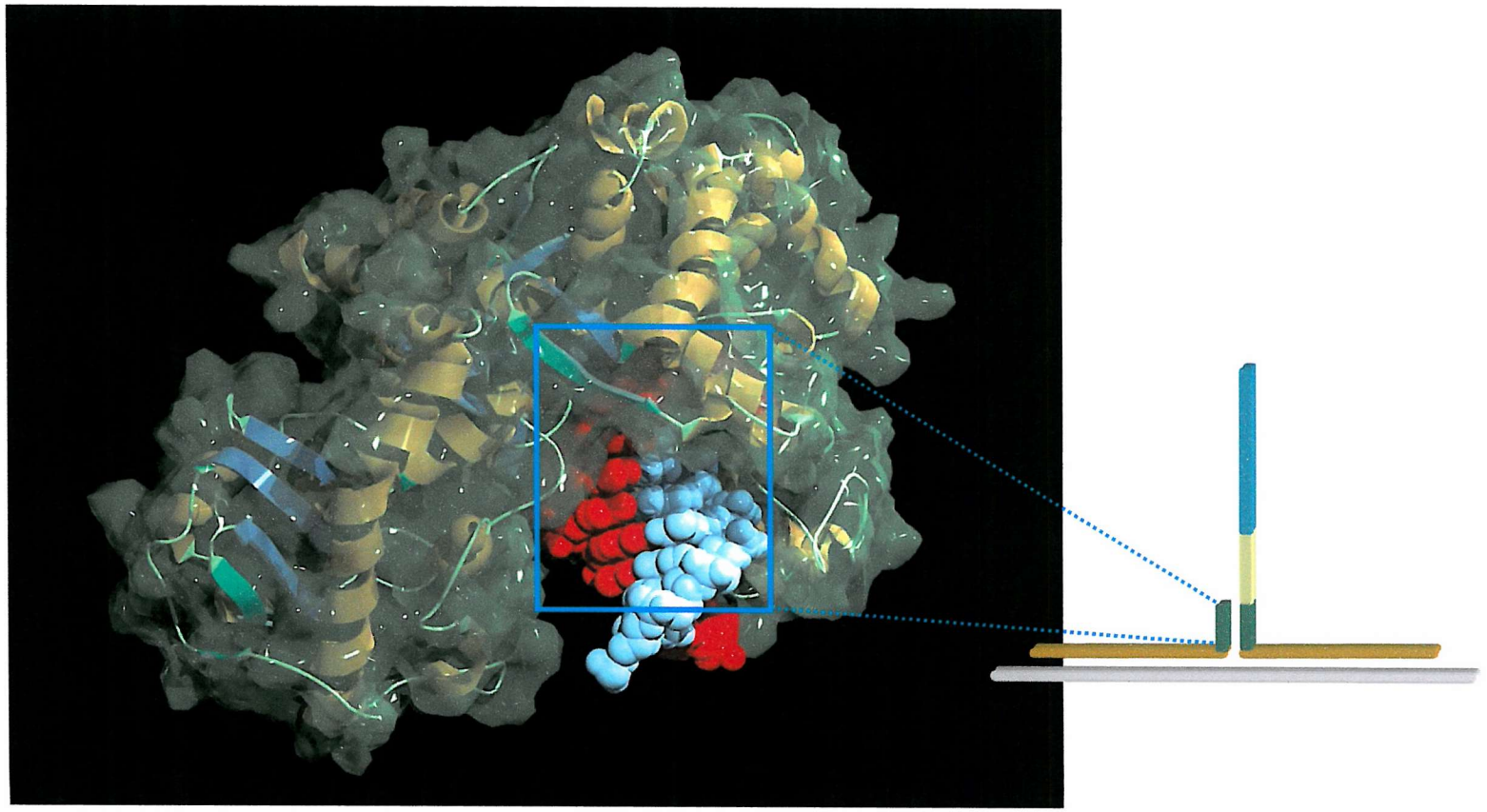


Fig. 1.6. Co-crystal structure of the Bst DNA polymerase bound to a 9 bp DNA duplex (Kiefer *et al.*, 1998, pdb entry 2bdp). The beta sheets, alpha helices and loops of the protein are shown with the protein molecular surface superimposed in green. The DNA is shown in a VDW representation with one strand in light blue and the other strand in red. The DNA polymerase wraps around most of the DNA except for the part shown in the figure. The narrow space available for DNA binding could therefore mean that the conformation of the three-way junction formed in SMART (shown to the right) affects the binding of the polymerase due to the close proximity of two additional DNA arms. The co-crystal structure was rendered with Swiss-pdb viewer and Pov-Ray for Windows.

exploit the fact that T7 RNA polymerase acts processively only in the presence a double stranded promoter sequence, for which it is also highly specific. Studies have however indicated that T7 RNAP is able to initiate non-specific transcription using single stranded DNA molecules albeit with low processivity. Moreover, a nuclease activity has been observed for T7 RNAP in the presence of ssDNA but only when the ssDNA : dsDNA ratio was very high (Sastry & Ross, 1997). Also, Biebricher & Luce (1996) observed auto-catalytic replication of single-stranded RNA molecules at high RNA to DNA ratio's. These properties constitute potential sources of background signal and may cause a reduction in assay efficiency and yield. Even if the amount produced in the first amplification stage is below the detection limit, a small number of RNA molecules may be sufficient to trigger the second amplification stage to yield detectable amounts of RNA background, urging for careful design. Also, a high ssDNA:dsDNA ratio is present in the initial stages of amplification and a reduction in this ratio due to nuclease activity may reduce the sensitivity of the assay because less probes are available to form three-way DNA junctions.

1.3.3 Potential advantages of SMART

The reaction schemes shown in Fig. 1.3 and 1.4 in conjunction with the previous discussion reveal several noteworthy properties of SMART, which distinguishes it from other nucleic acid amplification based detection technologies. Importantly and in contrast to PCR, the target sequence itself is not amplified. Instead, the target merely induces the amplification of a chosen RNA species that is unrelated to it. It should be stressed that SMART acts only as a detection method for known sequences and is incapable of detecting novel sequences. Thus, amplification, in stark contrast to PCR, functions only to produce a detectable signal. This property has a substantial advantage as the sequence of the transcription template, present in the template probe, can be altered as to optimise the RNA yield. Even more, the optimised transcription template sequence can be transferred to any junction, regardless of which target sequence one wants to detect. This could improve the reproducibility of the assay across different target sequences, even those that are GC-rich (known to be detrimental to PCR efficiency).

Further, RNA targets are expected to act as target as efficiently as DNA, thus eliminating the requirement for an additional reverse transcriptase step and allowing

the direct monitoring of gene expression. The latter has important clinical applications in for instance cancer diagnosis. Also, RNA provides a more accurate indication of the number of viable cells (cancerous or pathogenic) present in a human body. Detection of RNA transcripts should, theoretically, be easier to achieve as they are often present in multiple copies. In addition, as the amount of RNA produced by SMART is directly proportional to the initial number of target molecules present, quantitative gene expression analysis should be a feasible option. RNA detection is however, less useful in the assessment of genetic variations since clinically important variations may reside in the coding regions as well as in the non-coding (e.g. promoter-, enhancer-, silencer-) regions of a certain gene. RNA may also be more difficult to obtain, as gene-expression obviously is dependent on the cell-type. For example, extracting RNA from a liver cell is more demanding than from blood cells.

1.3.4 SMART as a detection assay for small genetic variations: The TWJ conundrum

The detection of single or larger nucleotide variations ultimately involves a comparison between wild-type and mutant targets. For SMART this means that this discrimination must be achieved during the first step of the technology, by controlling the formation of the three-way junction. An attractive method involves the control of duplex stability through temperature alterations. The TWJ involved in SMART contains two duplex regions involving the target. One region is created by the annealing of part of the extension probe to target, and the other by the annealing of part of the template probe to target. The important question that needs to be answered is: which part of the duplex formed on the target shows the highest destabilisation when presented with a mutation? To answer such a question one could divide the target-spanning sequence of the TWJ into three regions (Fig. 1.7): The end-regions of the two target duplexes (comprising the terminal base pairs), the centre-regions of the two duplexes and the branch point region. It is well established that in DNA duplexes, the termini show the least sensitivity towards mutations. Thus, based on these observations these regions appear to be the least sensible position for mutation detection. In fact, most studies have observed that it is the centre of DNA duplexes that is the most sensitive for mismatches (Guo *et al.*, 1997). However, a complication arises when the TWJ is considered as a whole. Are the two target-

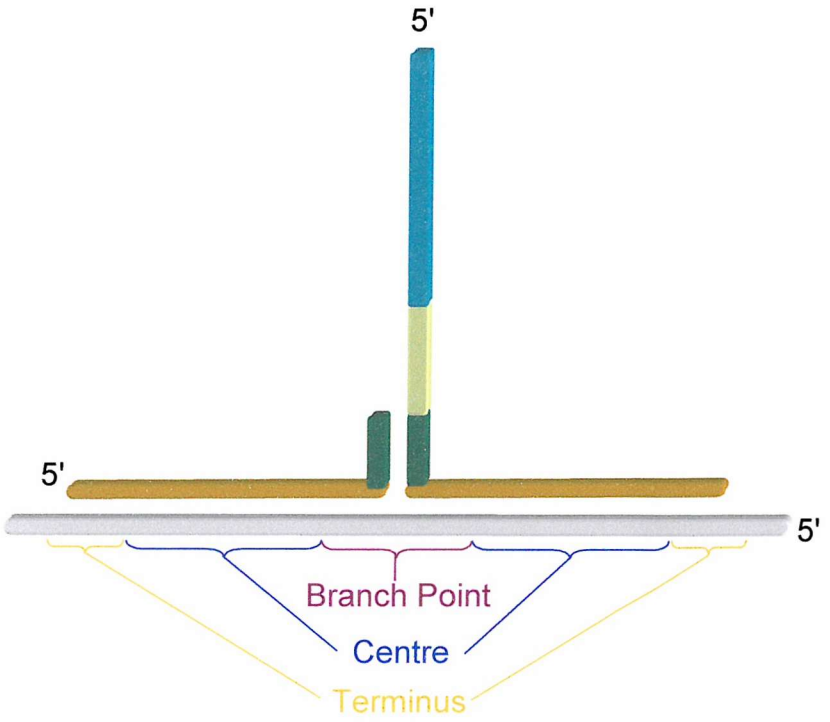


Fig. 1.7 The TWJ and mutation sensitivity. As a first approximation, the target-spanning region in the TWJ can be regarded as three duplex regions, each representing distinct areas with respect to mutation sensitivity.

containing duplexes truly separated in a thermodynamic sense, or do they interact? If so, it is less clear as to what should be considered as the centre. To answer the questions posed here, the following section will detail the current state of knowledge regarding TWJ structure and stability, followed by a discussion on how these properties may affect mutation sensitivity of the junction.

1.4 Three-Way Junction Structure

1.4.1 The complexity of junction folding: the importance of base stacking, DNA backbone and DNA-solvent interactions.

Three-way DNA junctions belong to a class of DNA molecules sometimes referred to as branched DNA. Other molecules belonging to this class are for instance the Holliday or four-way DNA junctions and bulged DNA duplexes. Generally, branched DNA molecules have been proposed to play an important role in various DNA rearrangements processes. These include homologous recombination, strand slippage and mutation (Lilley, 1993). Recent years have seen a tremendous increase in the analysis of the conformation of branched DNA, and its effect on DNA arrangements. Importantly, three- and four-way DNA junctions appear to share many conformational properties, although the former has been less well studied (for reviews refer to Lilley, 1993, Seeman & Kallenbach, 1993, Altona *et al.*, 1996). Nevertheless, one of the main characteristics of DNA junctions is their propensity to engage in coaxial stacking interactions. This important property of DNA junctions is further explained in Fig.1.8a, showing the three possible conformations of a four-way DNA junction. The centre structure is what is referred to as the open-, unstacked- or unfolded form. Essentially, this conformation consists of four standard B-DNA helices connected via their sugar-phosphate backbone chains at the branch point. The reason why this is called the open form is that a considerable gap exists between the four base pairs located at the branch point. To appreciate the significance of this gap it is useful to consider a standard DNA double helix. In a DNA duplex, neighbouring base pairs are separated by an average distance of 3.4Å (the helical rise). This strict separation distance is maintained because of so-called stacking interactions between neighbouring base pairs. These are a complex interplay of different forces including Van der Waals (VDW) and electrostatic forces. In fact, the value of 3.4Å is close to the optimal VDW distance and therefore, these are the

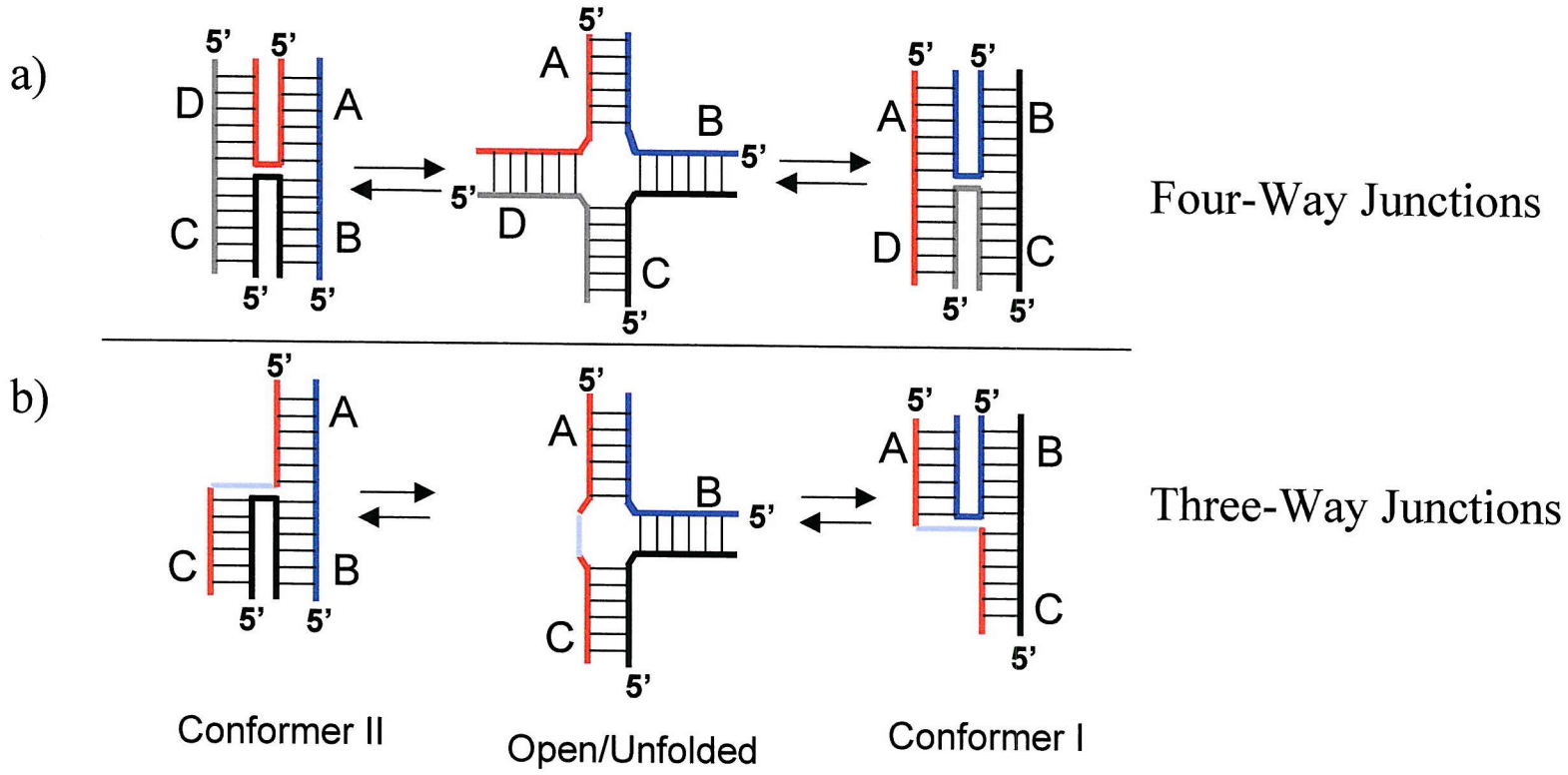


Fig. 1.8. The conformations of four-way and three-way DNA junctions. Unpaired bases are indicated in light blue. Refer to the text for further details.

forces contributing most to the helical rise (Lu & Hunter, 1997). Nevertheless, atomic partial charges, repulsive interactions between the π electrons of adjacent bases and attractive interactions between negatively charged π-electrons of one base and the positive atomic nuclei of a neighbouring base probably also contribute to the overall stacking arrangement of base pairs (Hunter & Sander, 1990; Hunter, 1993). The latter for instance explains why bases are not ideally stacked but show what is referred to as off-set stacking, which maximises attractive interactions between π-electrons and nuclei (Hunter & Sander, 1990; Hunter, 1993; Hunter & Lu, 1997). To illustrate this, Fig. 1.9 shows the arrangement of two neighbouring base pairs (two GC pairs) taken from a crystal structure (two base pairs from the non-junctional region of the FWJ studied by Eichman *et al.*, 2000). Fig. 1.9b shows the VDW surface of the two pairs, which are viewed from the minor groove (same orientation as shown in Fig. 1.9a). Fig. 1.9b clearly shows that the two base pairs are in close VDW contact. Fig. 1.9c displays the two pairs from above, down the helical axis and shows that the hydrophobic ring systems of the two pairs overlap only partially (off-set stacking). Fig. 1.9c also indicates that guanosine possesses a relatively large negative partial charge, contributed by the N7 and O6 groups (orange oval), which overlaps with the relatively large positive partial charge present on the C5 in the neighbouring cytosine (blue circle) (Packer *et al.*, 2000a). This explains for instance, why GC steps have a natural tendency for sliding motions, as these allow the favourable overlap illustrated in Fig. 1.9c, highlighting the importance of electrostatic interactions. In addition, a well-documented property of many base pair steps is a propensity for propeller twisting (especially ApA steps), which is the anti-rotation of the two bases in a pair relative to each other. The most likely explanation for this behaviour is that this rotation avoids a steric clash between the thymine methyl group and the furanose ring of the neighbouring base in the major groove, although other explanations have been offered (El Hassan & Calladine, 1996). Finally, the sugar-phosphate backbone is often considered semi-flexible and to act as a constraining factor on extreme deviations in base stacking arrangements that would otherwise occur in its absence (Packer & Hunter, 1998). For instance, Packer & Hunter (1998) noted that certain movements of base pairs relative to adjacent base pairs (slide and shift) are affected by, and can themselves affect the geometry of these adjacent base pairs through changes in the backbone torsion angles.

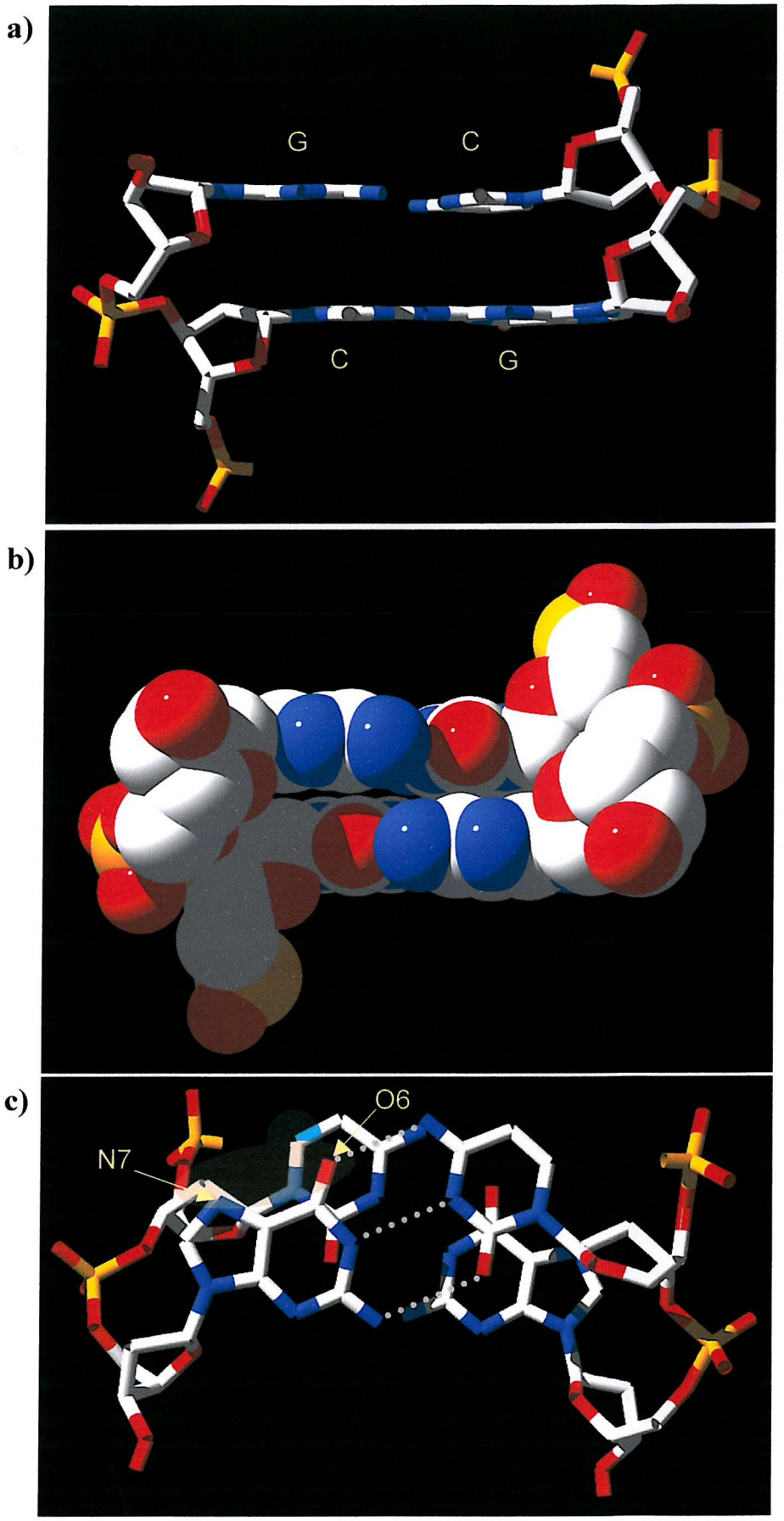


Fig. 1.9. Stacking interactions in DNA. The two base pairs were taken from a DNA crystal structure (pdb entry: 1dcw). Blue: nitrogens, red: oxygens, yellow: phosphates, white: carbons. Hydrogen bonds are indicated only for the top GC pair in c) by the broken grey lines. The three figures were rendered using Swiss-PDB viewer and Pov-Ray. Refer to text for further details.

In short, DNA stacking interactions are highly favourable and sequence-specific interactions, determined by both VDW and electrostatic interactions.

With respect to the previous discussion, it is now clear that the gap present at the branch point in Fig. 1.8 represents a lack of stacking interactions across the branch point in FWJs. The base pairs flanking the branch point are not in close VDW contact with each other. At the same time it appears that the FWJ does harbour the potential to further maximise stacking interactions. For instance, by rotating the arms, arm A can be connected (stacked) to arm B and arm C to arm D (Fig. 1.8a). Alternatively, A could be connected to D and B to C. The remaining possibility (A with C and D with B) is not feasible. Formally, for the fully stacked conformations the arms are said to be coaxially stacked, forming two quasi-continuous double helixes. Thus, in addition to the unfolded/open form, two different, fully stacked conformations are possible. Indeed, studies have verified the existence of each of these three conformations by comparative electrophoresis, FRET (Miick *et al.*, 1997), NMR (Altona *et al.*, 1996) and X-Ray crystallography (Eichman *et al.*, 2000). Whether a junction adopts a coaxially stacked or open form is determined primarily by the presence and concentration of cations. Even from the schematic illustrations in Fig. 1.8a it is clear that coaxial stacking involves a close approach of negatively charged DNA strands. Cations are therefore assumed to be required to screen these negative charges.

The folding of three-way DNA junctions shows many similarities with that of FWJs as judged from comparative gel-electrophoresis, chemical probing, NMR and FRET experiments (Duckett & Lilley, 1990, Guo *et al.*, 1990, Leontis *et al.*, 1991, Lu *et al.*, 1991, Stuhmeier *et al.*, 1997a, Stuhmeier *et al.*, 1997b, Welch *et al.*, 1993, Welch *et al.*, 1995, Yang & Millar, 1996, Zhong *et al.*, 1993, Zhong *et al.*, 1994). As for FWJs, an open form and two coaxially stacked forms are possible as shown in Fig. 1.8b, with the occurrence of coaxial stacking again depending on the presence of cations. The two conformers are usually referred to as conformer I and II and are discriminated easiest by following the polarity of the bulge-containing strand since they are different in both conformers.

TWJs can, in many respects, be regarded as FWJs lacking one arm. However, an important difference between TWJs and FWJs is that TWJs require the presence of

two or more unpaired bases at the branch point to allow coaxial stacking to occur. The simplest explanation for this is outlined in Fig. 1.10 (adapted from Altona *et al.*, 1996), showing that in the absence of unpaired bases insufficient space is present to allow two arms to engage in coaxial stacking interactions. Therefore, in the absence of unpaired bases, the TWJ will always adapt a Y- or pyramidal shaped conformation, lacking stacking interactions across the branch point.

NMR studies (Leontis *et al.*, 1994, Rosen & Patel, 1993a, Rosen & Patel, 1993b) have offered further insight into the characteristics of unpaired bases when present in a TWJ. These experiments revealed that the identity of the unpaired bases does not affect the choice of stacking partners. One NMR study however showed small, localised differences when a bulge of two adenosines (both located within the junction) was replaced with two cytosines. One of the cytosines was located outside the TWJ interior (Rosen & Patel, 1993b). A thermal melting study on a junction containing the same 4bp around the branch point, and which had shown to fold in the same manner as the junctions studied by Rosen and Patel (1993a, 1993b), revealed differences in stability between the two differently bulged junctions (Welch *et al.*, 1995). Thus, although the composition of the bulge does not appear to influence the choice of stacking partners, it may affect the stability of the junction.

Finally, the effect of increasing the bulge size was studied by Welch *et al.* (1993), who found a closer approach of the unstacked arm towards the stack when more unpaired bases were added to the bulge.

The similarities between TWJs and FWJs are further illustrated by the two high-resolution structures shown in Fig. 1.11 and 1.12. The figures compare a FWJ crystal structure (Eichman *et al.*, 2000) with a conformer II TWJ (Rosen & Patel, 1993b). Apart from highlighting the strong resemblance between the quasi-continuous helixes and B-DNA, Fig. 1.11 and 1.12 also illustrate that the FWJ contains two different faces: a major groove face (which is the viewpoint of the figures) and a minor groove face. A similar situation is present in the conformer II TWJ, where the major groove faces of the branch point base pairs in the quasi-continuous helix face the minor groove face of the unstacked helix. Due to the differences in stereochemistry between conformer I and II TWJs, the latter situation is reversed in conformer I junctions (i.e. the minor groove of the coaxially stacked helixes faces the major groove of the unstacked arm). It should be noted however, that the position of

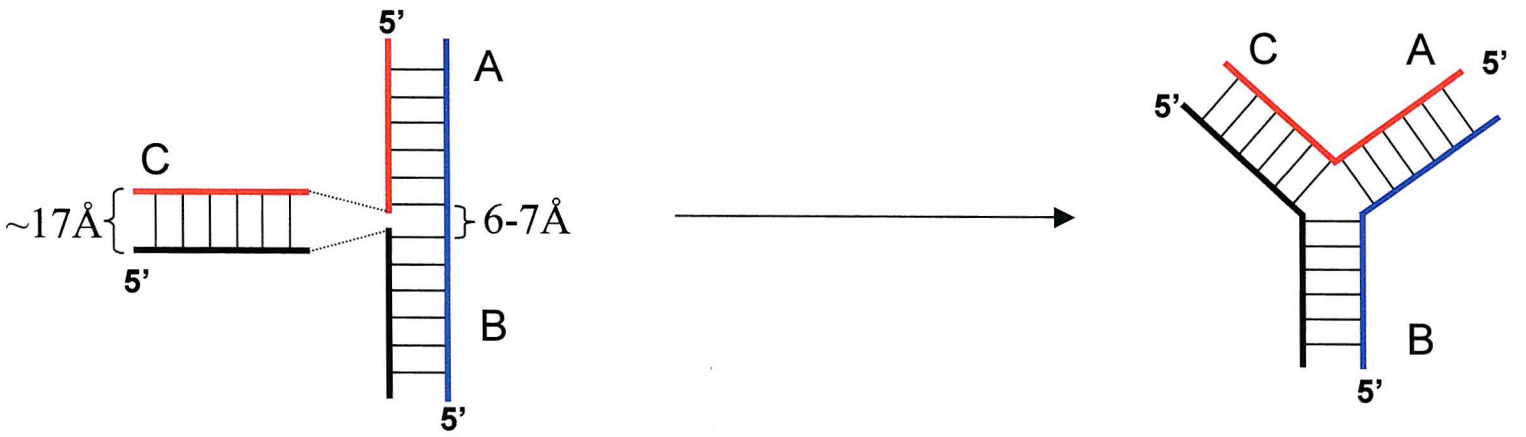


Fig 1.10. The effect of unpaired bases on TWJ conformation. The construction of a TWJ lacking unpaired bases can be viewed as an attempt to couple a DNA helix to another DNA helix in the centre. The diameter is much larger than the distance separating the two phosphate groups of the two bases where the second helix is to be coupled to. Thus, stacking interactions between arm A and arm B cannot be maintained in the absence of unpaired bases and, in stead, the junction assumes a Y- or pyramidal shaped structure.

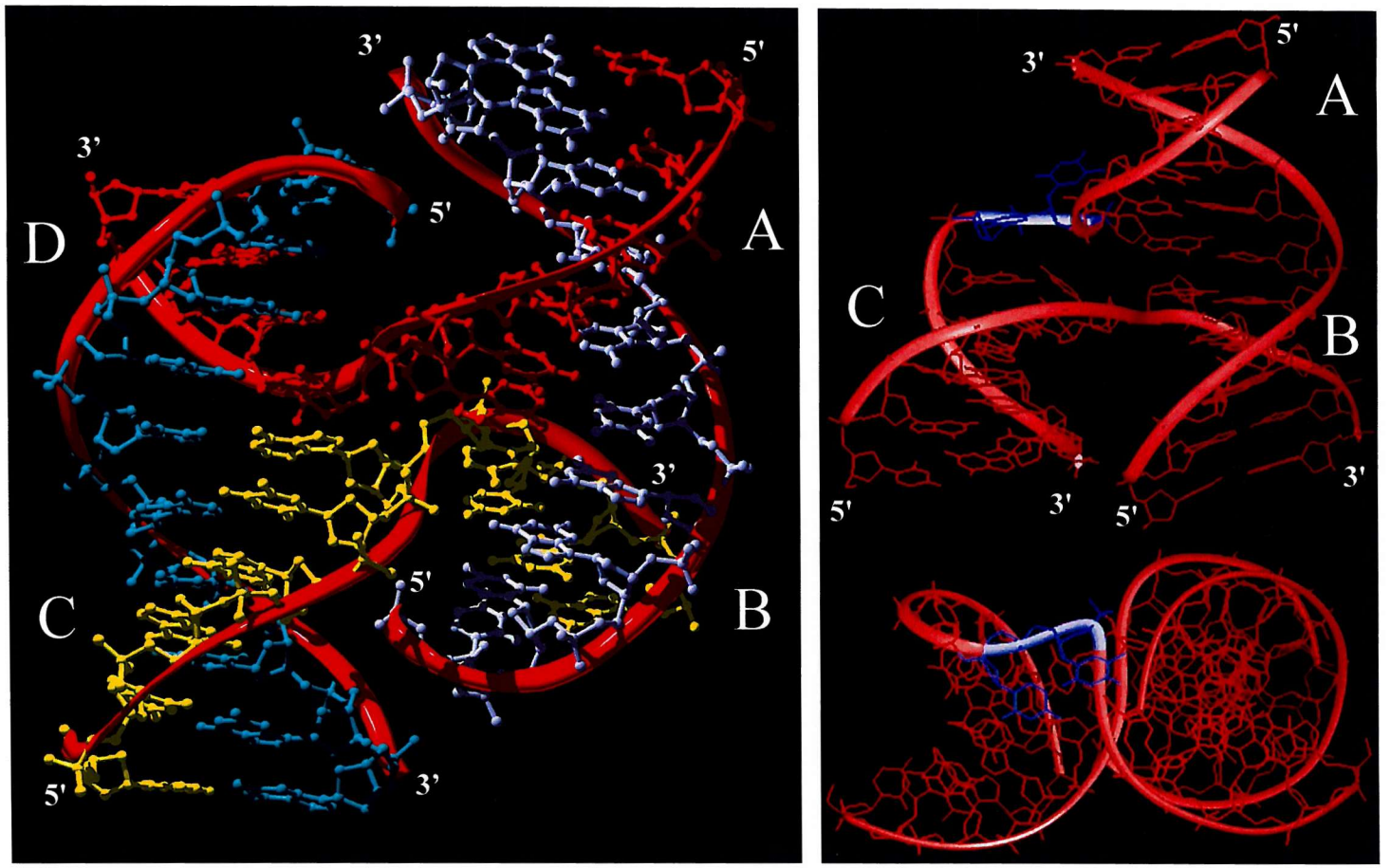


Fig. 1.11. Crystal and NMR structures of a FWJ (left) and TWJ(right). The strand colours in the FWJ are same as in Fig. 1.8a, except that the black strand is shown in yellow here. The FWJ structure was determined by Eichman *et al.* (2000), PDB-entry 1dcw. The TWJ was assessed by Rosen & Patel (1993b) and the figure was taken from Neidle, S (1999). The TWJ is viewed from the major groove of the coaxial stack (top) and down the helical axis of the coaxial stack (bottom). Unpaired bases are indicated in blue.

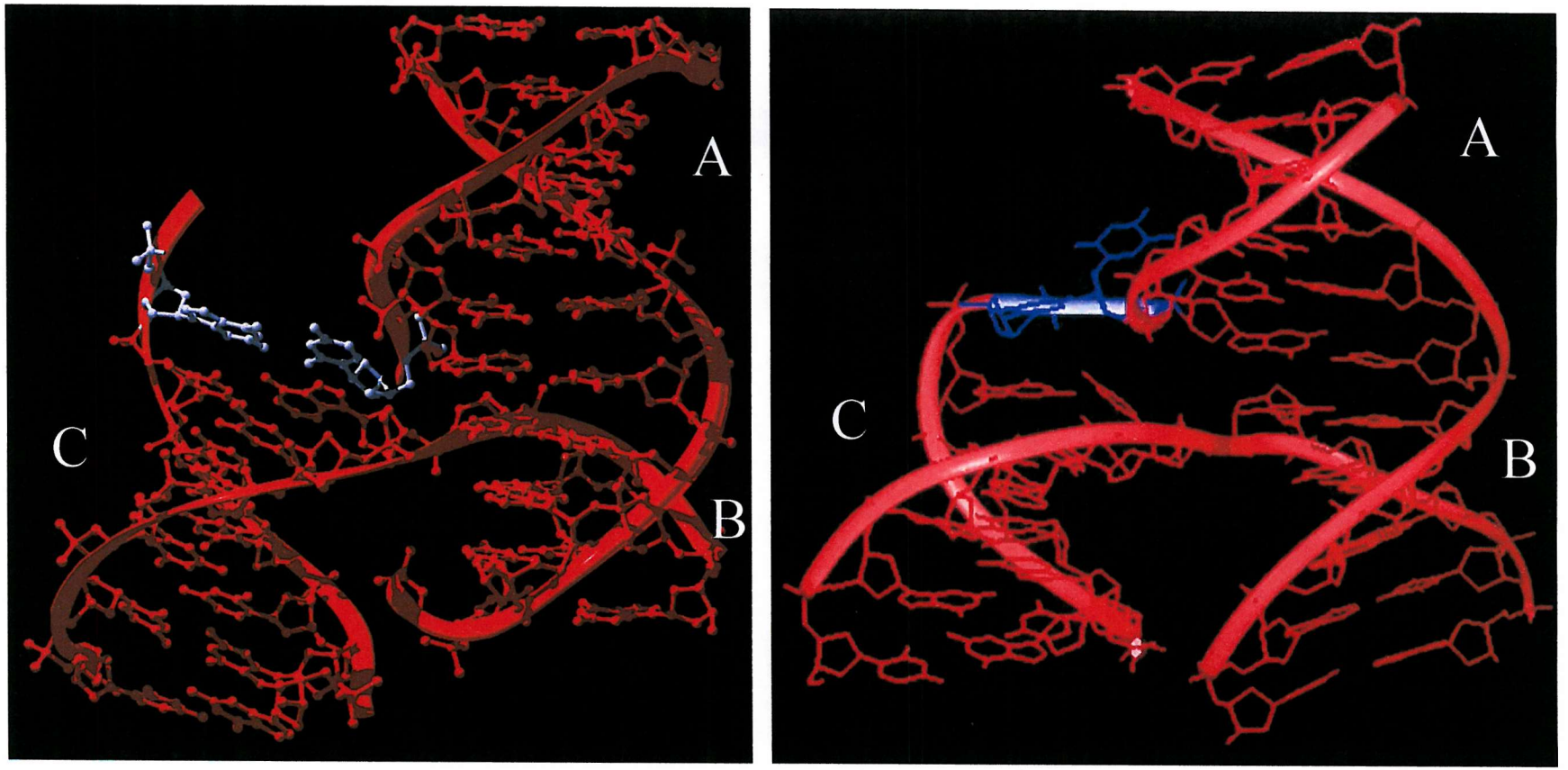


Fig. 1.12. Similarities between FWJs (left) and TWJs (right). The TWJ can be regarded as a FWJ lacking one arm as shown on the left. The figures further show that the most significant difference is found in the backbone conformation of the strand containing the unpaired bases in the TWJ. To overcome the gap present in the FWJ structure, the backbone must be rotated to accommodate the unpaired bases whilst maintaining coaxial stacking interactions.

the unstacked arm relative to the coaxially stacked helixes in a TWJ may show variation due to increased flexibility [Rosen & Patel, 1993b, Leontis *et al.*, 1999 (see also pdb entry: 1EKW)]. Finally, Fig. 1.12 highlights that the most significant difference between TWJs and FWJs is in the conformation of the backbone containing the unpaired bases in the TWJ.

One important question that remains is which process(es) determines which coaxially stacked conformation a FWJ or TWJ assumes upon junction formation (providing cations are present)?. The answer to this question probably involves forces acting within the helix (i.e. stacking interactions) and those acting on the helix (i.e. forces exerted by the solvent). Although the negative charge present on the DNA backbone explains why coaxial stacking occurs only in the presence of cations, cation-phosphate oxygen contacts do not explain why certain junctions display conformer I stacking and why others adopt the alternative coaxially stacked conformation. Studies have shown that both stacked configurations are feasible and, unsurprisingly, the choice determining which arm stacks with which arm is determined by sequence. In fact, only sequences located close to the branch point appear to influence the conformation of DNA junctions, and even a single base pair change can result in a complete rearrangement of coaxial stacking interactions (Welch *et al.*, 1995). It seems clear that the explanation for this must involve differences in base pair stacking interactions. In other words, the coaxial stacking sequences formed in the adopted conformer must be the energetically most favourable. However, the two coaxially stacked conformations, apart from showing differences in base stacking interactions, are likely to show differences in DNA-solvent interactions as well. This can be understood when considering the fact that the phosphate oxygens are not the only atoms capable of interacting with cations and water. The minor and major groove faces of a given base pair also contain (sequence-specific) partially charged amine and carbonyl groups that can interact with water and cations. Since the electrostatic surface of DNA is therefore also sequence dependent, it is possible that junction conformation is the result of a trade-off between optimal stacking and optimal solvent-DNA interactions. Although the conformation of standard B-DNA helixes can be reasonably well explained on the basis of stacking interactions (Packer *et al.*, 2000a, 2000b, El Hassan & Calladine 1996), clear evidence has been obtained suggesting that cations such as Mg^{2+} can

affect stacking interactions in duplex DNA as observed for instance by Dickerson *et al.* (2000). Moreover, it is important to realise that in DNA junctions the branch point constitutes a highly unusual environment because of the high charge density, which complicates prediction of the influence of solvent on TWJ conformation. Only a small number of studies have attempted to assess the importance of solvent-DNA interactions in junction folding by, for instance, substituting phosphate groups located at the branch point with neutral analogues (e.g. methyl phosphonates, Pikkemaat *et al.*, 1995). In one case, substitution of a single phosphate group resulted in a complete stacking rearrangement (i.e. conformer conversion) but this only highlights the importance of the backbone charges, and reveals little about the influence of sequence on the solvent arrangement. The latter is experimentally difficult to test since substitutions of polarised atoms in the bases affect the electronic structure of these bases, which complicates discrimination between stacking effects and solvent effects. In addition, different sequences may affect the local positioning of phosphate groups (for instance by causing differences in base pair twisting), which again complicates accurate analysis of cation-TWJ interactions. Thus, the observed sequence-dependency of junction folding largely remains an unsolved and highly complex issue.

In summary, the global conformation of DNA junctions is affected by three factors: the DNA sequence close to the branch point, the presence (and concentration) of cations and, in the case of TWJs, the presence of unpaired bases. Of these factors, the sequence-dependency of junction folding remains the least understood.

1.4.2 Dynamics of junction folding

One aspect that has not been addressed so far is the dynamics of junction conformation. FRET and NMR studies have revealed compelling evidence for the existence of a dynamic conformational equilibrium between the two possible stacked conformations of FWJs (Miick *et al.*, 1997, Overmars & Altona, 1997). However, in most cases it appears that a clear conformer preference exists. This is reflected in the relative distribution of the two stacked conformers at any given point in time, where the relative amount of the dominant conformer usually ranges between 60% and

90%. The conformational residence times have been estimated to be in the millisecond range (Overmars & Altona, 1997), and this explains why comparative gel-electrophoresis experiments never revealed the existence of a conformational equilibrium (further discussed below). This is a crucial revelation since the majority of structural data regarding the structures of FWJs and TWJs has been obtained using a technique developed by Cooper & Hagerman (1987) sometimes referred to as the long-short arm assay, outlined in Fig. 1.13. The assay incorporates three different restriction endonuclease (R-ENase) recognition sites, each positioned in a different arm equally spaced relative to the branch point (Duckett *et al.*, 1990, Welch *et al.*, 1995). By separate incubation with each of the appropriate R-ENase, followed by analysis on a non-denaturing polyacrylamide gel, information is obtained on the curvature of each of the three fragments. Differences in migration through the gel can result only because of differences in curvature since all three fragments contain exactly the same number of base pairs. Since DNA bending has been shown to cause retardation in polyacrylamide gels relative to linear duplex DNA (Hagerman, 1990), the fastest species is assumed to constitute the quasi-continuous helical fragment. The gel migration of the other two TWJ fragments thus provides information on the position of the unstacked arm relative to the stacked arms. Since the duration of native gel electrophoresis is hours rather than milliseconds (the time-scale involved in FWJ conformer conversion), the structural data obtained is averaged, presumably reflecting the dominant conformer. This is probably the reason why the LSA has never exposed the existence of two conformationally different versions of the same junction in a single solution.

Regarding their similarities with FWJs it appears likely that TWJs also reside in a dynamic conformational equilibrium. However, it should be stressed that so far experimental studies (NMR and FRET, Overmars *et al.* 1995, Stuhmeier *et al.* 1997a & 1997b) have not been able to verify the existence of such an equilibrium for TWJs. These observations do not exclude the existence of an equilibrium between structural states for TWJs but call for more studies on junctions for which studies indicate multiple conformations may exist. An example is a junction named YAAAA (Zhong *et al.*, 1994), which showed the appearance of two bands rather than one in a single lane in the long-short arm assay. The study did not include control experiments to determine whether the observed bands were experimental artefacts (e.g. DNA

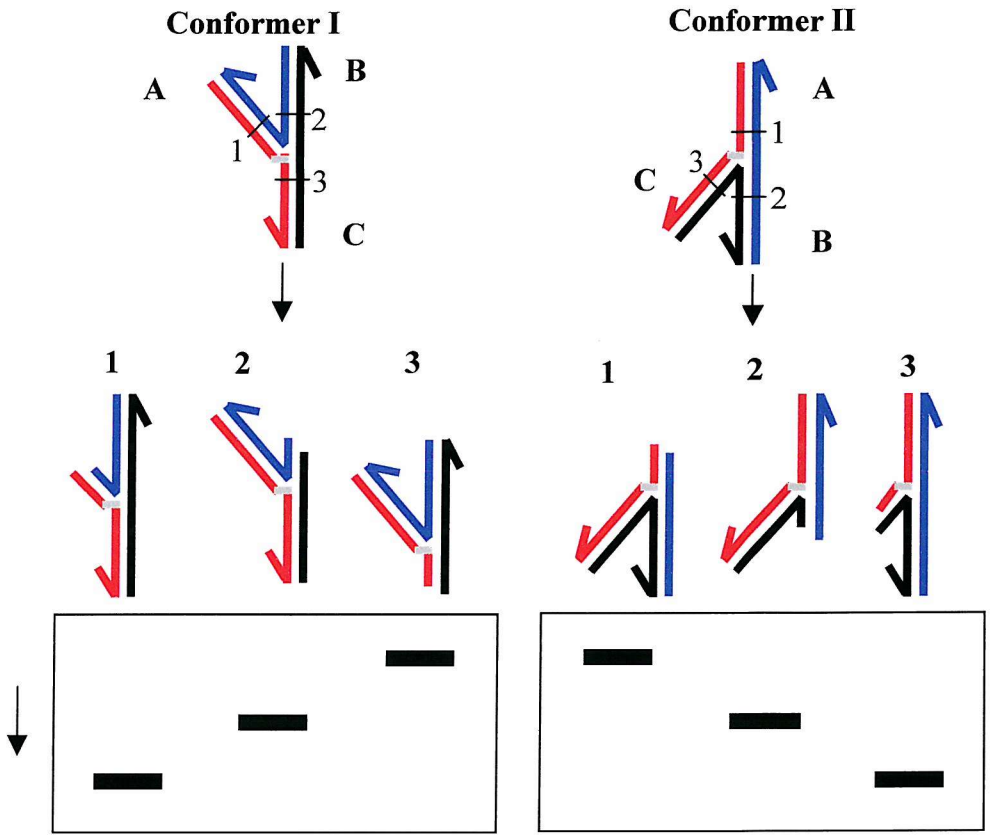


Fig. 1.13 Principles of the LSAa. A TWJ with equal length arms and three unique R-ENase cleavage sites (1-3), one in each arm (A, B and C as in Fig. 1.8) located at equal distances from the branch point, is designed. When this junction is subjected to cleavage by each R-ENase, three fragments are obtained, each with a unique combination of two intact arms possessing different angles. Because all fragments are of equal molecular weight, their gel mobility is directly dependent on this angle (the smaller the angle the greater the retraction). Thus, the gel provides information on the folded state of the junction. The schematic gels further indicate that each conformer yields a unique mobility pattern for the same order of restriction fragments.

contamination, incomplete digestion products or band-shift) or not and therefore require more rigorous analysis. It has been suggested that junction J1V9 studied by Welch *et al.* (1995), exhibited an equilibrium between conformer I and conformer II. This junction showed equal mobilities for all restriction R-ENase fragments when subjected to the long-short arm assay. Such a pattern can be interpreted as an average between the two conformers if rapid changes in conformation occur during migration through the gel. A pyramidal-shaped junction lacking stacking interactions between arms would also explain the data however, again urging further experiments to resolve the issue.

1.4.3 Mutations and three-way junction structure

Little attention so far has been paid to the effect of mismatched bases on the three-way junction. To our knowledge only two studies have been conducted on the effect of mismatches in TWJs (Zhong *et al.*, 1997, Zhong *et al.*, 1993). The study concerned a junction with T-T pairs at the branch point and the results showed a pronounced effect compared with a junction containing a 'perfect' T-A pair, as studied by the LSA. A clear dependency on the position at the branch (i.e. in which arm it was located) was observed with respect to its structure. The outcome was quite dramatic as changing the location of the T-T mismatch around the branch point, altered the mobility pattern of the R-ENase-fragments from conformer I to conformer II. Calorimetric studies indicated only small differences in the melting temperatures of the mismatched junctions relative to the perfect junction (Zhong *et al.*, 1997), suggesting that the mismatch might have acted in a similar way as a bulge by providing folding flexibility.

Most studies have focussed on the effect of introducing two or more unpaired bases at the branch point as mentioned earlier. These junctions could be considered as junctions resulting from annealing of the two SMART probes to a target containing a deletion or insertion. The inclusion of unpaired bases usually adds stability to the junction (Leontis *et al.*, 1991, Zhong *et al.*, 1994), probably because it allows stacking to occur between two of the three arms, but (as pointed out before) can have a substantial impact on the conformation as well.

1.4.4 Significance of the thermodynamic and structural properties of TWJs for SMART

The previous section showed that the conformation of TWJs shows a clear dependency on the sequences flanking the branch point. Provided that sufficient cations and flexibility are present (e.g. unpaired bases or a chemical linker) stacking between two helixes may occur. It should be noted that SMART junctions usually incorporate a chemical, rather than a nucleotide linker in the template probe at the branch point as experiments have shown that this reduces background signals. Two conformations are therefore possible, and these are reproduced in Fig. 1.14 but this time using a SMART junction as an example. Fig. 1.14 demonstrates that the position of the overlap containing arm relative to the remaining two arms differs substantially between the two conformers. In conformer I the two target spanning arms are localised closer to the overlap region than in a conformer II. It appears more likely that conformer I provides less steric accessibility for DNA polymerase than conformer II, as in the latter the two target-containing arms are moved away from the overlap. We therefore propose that the conformation of the three-way junction formed in SMART will affect the amount of RNA produced and that conformer II will yield more RNA than a conformer I. In the case of a structural equilibrium we should rephrase the latter as a junction whose dominant conformation is conformer II will yield more RNA than a junction whose dominant conformation is conformer I. If this hypothesis is correct, it follows that junctions based on different target sequences (either different genes or mutant variants of the same gene) may differ in conformation and so will produce different amounts of RNA. In terms of mutation detection two possibilities can be envisaged that achieve discrimination between wild-type and mutant targets (Fig. 1.15). One method involves a conformational change of the TWJ induced by the presence of a mutation but this requires intimate knowledge of the conformational properties of the TWJ. It is possible to determine consensus sequences for both conformers given the observation that TWJ structure is regulated by only a small sequence region flanking the branch point (2 bp in each arm but possibly more). Determining these would allow one to design conformer II junctions by positioning the branch point at the conformer II-dictating sequence in a given target. The branch point should be positioned at the location where a mutation is expected. The introduction of a mutation effectively alters the sequence at the branch point, yielding the alternative and less accessible conformer I. Thus, in

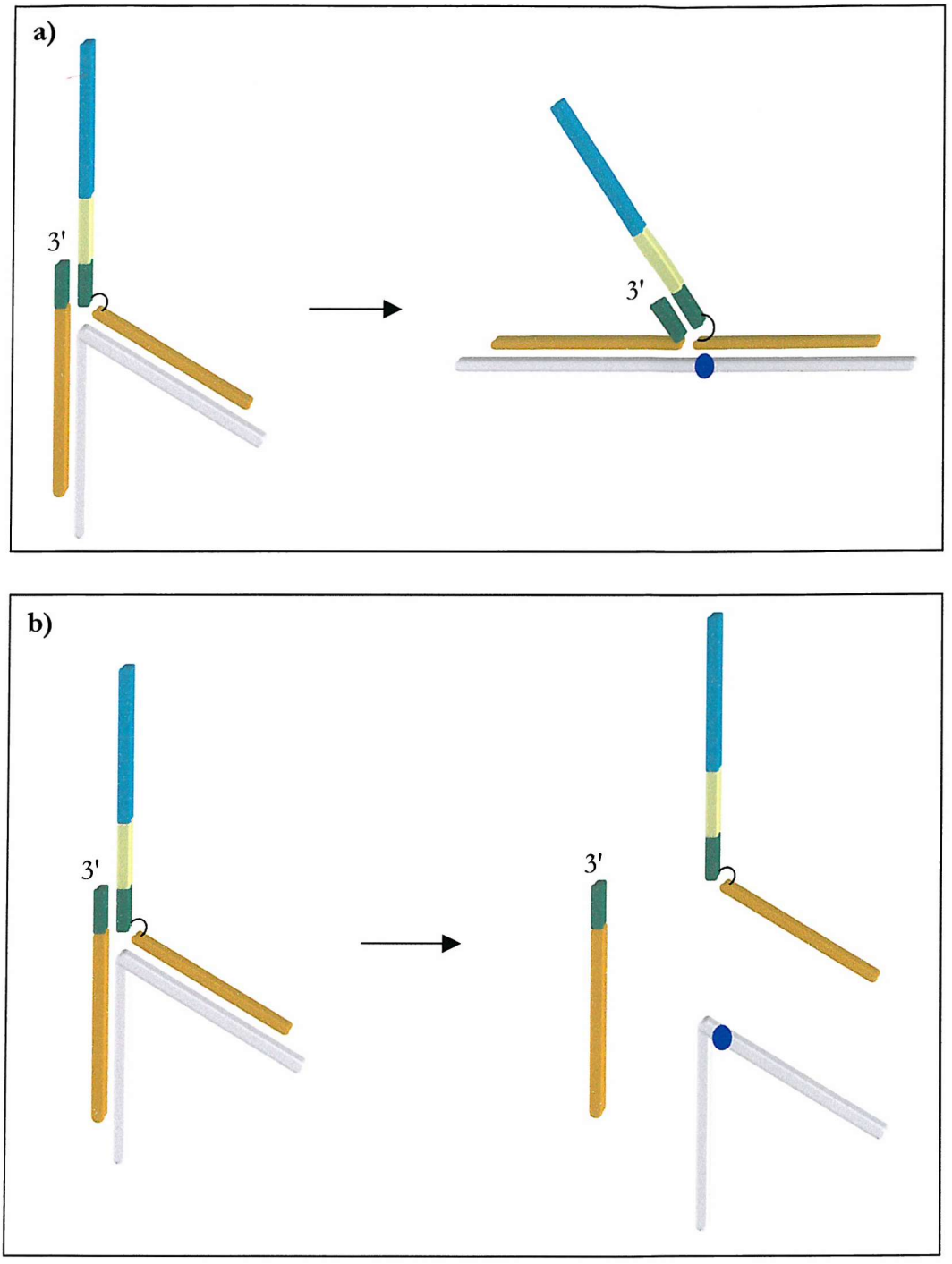


Fig. 1.15. Possible approaches to achieve discrimination between wild-type and mutant targets using SMART, either through a TWJ conformational change (top) or through reduction of junction stability (bottom). The blue circle indicates a mutation, but note that in the second example the optimal location for the mutation remains to be determined.

addition to structural consensus sequences, understanding the effect of mutations on TWJ conformation is required.

An alternative method is based purely on DNA hybridisation theory and the observation that at least in standard DNA duplexes, mutations usually reduce the Tm. In theory it should therefore be possible to design the junction in such manner that in the presence of the wild-type target the junction is formed, whereas the presence of a mutation destabilises the junction. This approach is less dependent on TWJ conformation (although a favourable conformation may improve discrimination), but it does require the assessment of the most mutation-sensitive area in the junction. Empirically, the latter approach is the most attractive as one can simply introduce mismatches at various locations around the branch point to assess their effect on the RNA yield by comparing with wild-type target. Nevertheless, the structure-yield relationship does require attention for it forms a potential source of inconsistency in SMART. Therefore, as outlined in section 1.5, both approaches will be assessed in this thesis.

1.5 Purpose of this work

The objectives of this thesis are:

1) To investigate and improve the ability of SMART to discriminate between wild-type targets and targets containing small sequence differences. These studies will use synthetic oligonucleotides as wells as PCR-amplified fragments obtained from affected and unaffected persons.
2) To assess the effect of mutations on the conformation of the TWJ formed in SMART.
3) To test whether TWJ conformation affects RNA yield in SMART.
4) To develop an empirical model that allows accurate predictions to be made of TWJ conformation based solely on the target sequence.

2 Materials & Methods

2.1 Reagents

2.1.1 Chemicals and enzymes

Chemicals were purchased from Sigma-Aldrich Inc. unless stated otherwise. All enzymes were purchased from New England Biolabs Inc. except T7 RNA polymerase (5,000 U/ml, Promega Inc) and DNaseI (2,000 U/ml, Ambion Inc.).

2.1.2 Buffers

5x TBE	0.45 M tris, 0.45 M of boric acid, 10 mM EDTA (pH 8.0) in 1 litre deionised water
10x TBM	0.89 M tris, 0.89 M boric acid, 50 mM $MgCl_2 \bullet 6H_2O$, in 1l deionised water, pH 7.5
10x urea gel buffer	7 M urea, 1 M tris-borate, 20 mM EDTA (pH 8.0)
Diluent	8 M urea
1x Denaturing loading dye	80% (v/v) formamide, 10 mM EDTA (pH 8.0), 2 mM NaOH and 0.1% bromophenol blue (w/v)
1x Non-denaturing loading buffer	1x TBM, 0.1% (w/v) bromophenol blue, 20% (w/v) ficoll
10x kinase buffer	700 mM tris-HCl (pH 7.6), 100 mM $MgCl_2$, 50 mM DTT, supplied by Promega
10x NEB buffer 2	500 mM tris-HCl pH 7.9, 500 mM NaCl, 100 mM $MgCl_2$, 1 mM DTT
Stop solution	1.5 M NaOAc pH 7.0, 1 M β-mercaptoethanol
5x TBS (Tris Buffered Saline)	30.25g tris pH 8.0, 40.3g NaCl, 1g KCl, 5 ml Tween 20 in 1l deionised, RNase free water.
Hybridisation Solution	58.4g NaCl, 40 ml 0.5 M EDTA pH 8.0, 6.1g tris, 1g Bovine Serum Albumin in 1l RNase-free deionised water

2.1.3 Nucleic acid sequences

DNA oligonucleotides were synthesised by Oswel DNA service (Southampton, UK), on a 40 nM scale with HPLC purification. PNA-DNA chimeras were synthesis by PNA Diagnostics A/S (Denmark) or Oswel DNA service. LNA-DNA chimeras were synthesised at Exiqon A/S (Denmark). The sequences of the chimerical probes are shown in the text.

a) DNA sequences for TWJ conformational studies

Junction designs used for TWJ structural analysis are indicated in the text, including the sequences of the oligonucleotides used.

b) DNA/RNA sequences used for SMART assays

Target sequences

A large range of different but related target oligonucleotides were available, representing the wild-type or mutant CFTR gene. These are shown in Fig. 2.1. CFTR targets were obtained from Genzyme Inc. (USA). Information regarding the reaction conditions used to generate these were not provided, only the resulting target yields. The presence or absence and zygosity of the ΔF508 and ΔI507 deletions were confirmed (by Genzyme) by dot-blotting with appropriate allele-specific oligonucleotides.

Additionally, single-stranded synthetic targets representing either the wild-type or mutant human Factor V gene were prepared:

Wild-type:

5'GGACTACTTCTAATCTGTAAGAGCAGATCCCTGGACAGGC**G**AGGAATA
CAGAGGGCAGCAGACATCGAACAGCAGGCTGTGTTTGCTGTG3'

Factor V Leiden (bold-underlined indicates the SNP):

5'GGACTACTTCTAATCTGTAAGAGCAGATCCCTGGACAGGC**A**AGGAATA
CAGAGGGCAGCAGACATCGAACAGCAGGCTGTGTTTGCTGTG3'

Extension probe and template probe sequences

For SMART TWJs, the arms interacting with the target are variable in sequence, as is the size of the overlap region. The designs of the individual junctions are shown in the various chapters of this thesis. The promoter/transcription template sequences

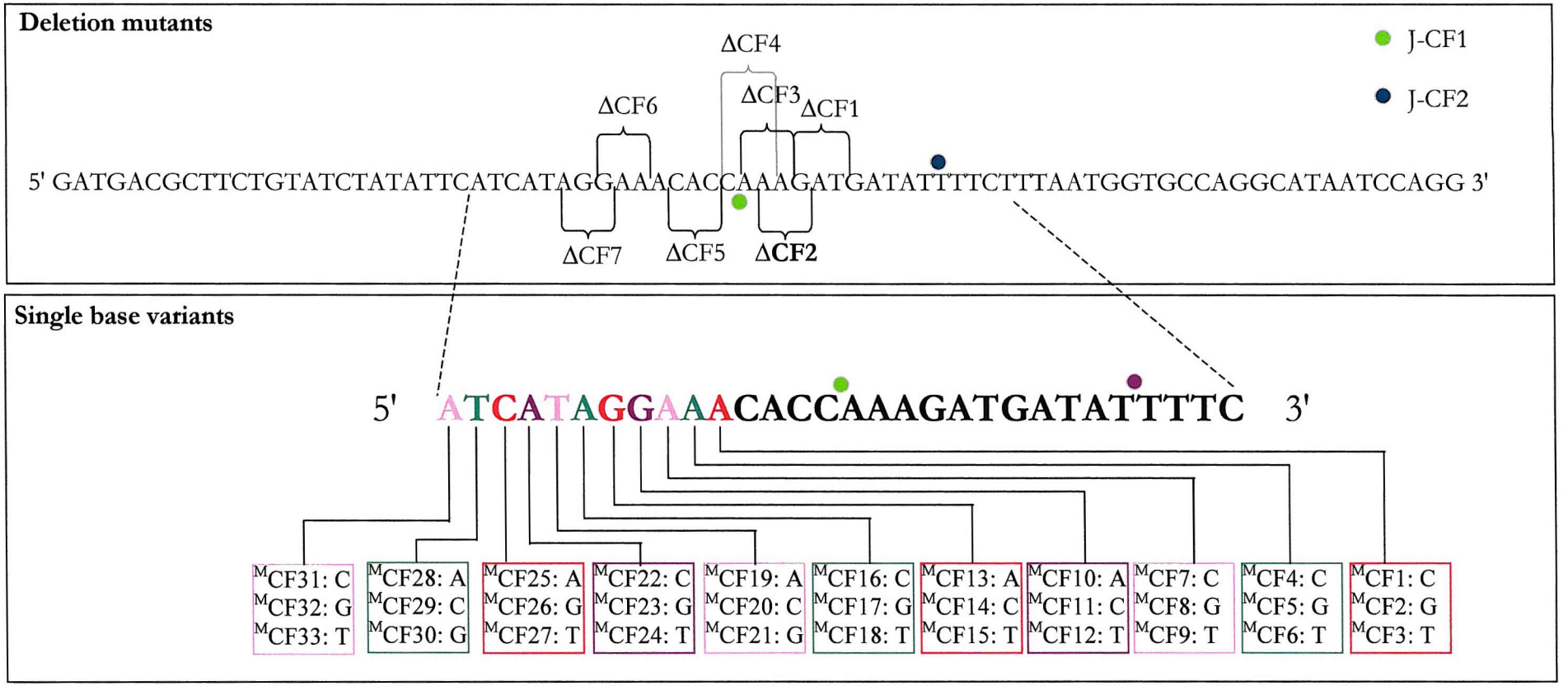
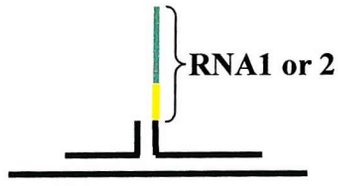


Fig. 2.1 The CFTR target and derivatives containing deletions (top) or single base substitutions (bottom). The full wild-type sequence of CFTR is the top sequence. Deletions are three-base deletions, e.g. ΔCF1 has 5'GAT3' deleted from CFTR-1. The names of the oligos are indicated in the figures with ΔCF for deletion mutations and ^{m}CF for CFTR derived targets containing a single base change. The filled circles indicate the branch points of the different junctions tested (described in chapters 3). Disease-predisposing deletions: ΔF508 = ΔCF2 and ΔI507 = ΔCF1.

are common to many junctions and are described in this section. Only two different sequences covering the overlap and transcription template in the template probe are used, producing the RNA species named RNA1 and RNA2:



The RNA1 and 2 producing versions of the template probe are simply referred to as A or B respectively in naming of junctions (refer to 2.9 for further details). The sequences are as follows (yellow indicates the T7 RNAP promoter sequence):

RNA1 (A)

5'TGCCTCCTTGTCTCCGTTCTGGATATCACCCGATGTGTCTCCCTATAGTG AGTCGTATTA3'

RNA2 (B)

5'TCGTCTTCCGGTCTCTCCTCTCAAGCCTCAGCGCTCTCTCTCCCTATAGT GAGTGAGTCGTATTA3'

The overlap and target complementary sequences of each of the extension and template probes are indicated in the text.

Oligonucleotides used for the ELOSA (described in section 2.8)

Oligonucleotides for the detection of RNA1 and RNA2:

Synthetic DNA copy of the RNA produced in the SMART assay for standard curve:

5'-GGGAGACACATCGGGTGATATCCAGAACGGAGACAAGGAGGCA-3'

Capture Probe:

5'-biotin-TCCGCTGCCTCCTTGTCTCCGTTCT-3'

Alkaline Phosphatase (AP) probe:

5'-GGATATCACCCG-Alkaline Phosphatase-3'

2.2 5'-End labelling of oligonucleotides (Sambrook *et al.*, 1989)

Oligonucleotides were labelled in assay mixtures containing 1µl of fresh γ-^{32}P-ATP (Amersham, ~3000 Ci/mmol), 1µl of T4-polynucleotide kinase, 10µl (~100 pmol) of oligonucleotide, 2µl of 10x kinase buffer and 6µl of deionised water. Reactions were incubated for 1-2 hr at 37°C. 10 µl of denaturing gel loading buffer was added. Labelled oligonucleotide was purified on a 12.5% denaturing polyacrylamide gel, run at 1500V, 42W for about 2 hr (25ml of sequagel solution (National Diagnostics Inc.), 5ml of 10x urea gel buffer, 15ml of diluent, 40µl TEMED, 200µl 20% (w/v) ammoniumpersulphate solution) with 1x TBE as running buffer. After excision, the position of the appropriate band was located by autoradiography and excised. The gel slices were transferred to a 1ml Gilson pipette tip, which had a glass wool plug at the bottom, and eluted into 300µl of deionised water. The tips were placed in a 1.5ml Eppendorf tube and incubated over night with shaking at 37°C. The DNA-containing solution was transferred to an Eppendorf tube and purified by ethanol precipitation using 3 volumes of ethanol and 1/9 volume of 3 M NaOAc, followed by a 10 minute incubation on dry-ice, centrifugation at 14,000 rpm for 10 min at 4°C and lyophilisation in a speed-vac. The oligonucleotides were further subjected to two washing steps involving re-dissolving the DNA in 70% ethanol and lyophilisation. Finally, the purified, labelled oligomers were dissolved in deionised water to give a concentration of around ~10 counts cpm per µl (as measured with a Geiger counter).

2.3 Annealing of radiolabelled junctions

The three-way DNA junctions were annealed so that the resulting TWJ contained one labelled and two unlabelled strands. The annealing solution contained 10µl of the labelled, 2x 5µl of unlabelled oligonucleotide present in vast excess, 5µl of 10x NEB buffer 2 and 25µl of deionised water. Tubes containing these solutions were transferred to a beaker with 200ml of water of ~90-95°C and left to cool slowly to room temperature.

2.4 TWJ structure analysis: the 'long-short arm' assay.

For restriction enzyme digestion the following solution was prepared: 5µl of labelled junction, 2µl of 10x NEB buffer 2, 1µl of R-ENase (10-20 units), and 11µl of deionised water. The solutions were incubated for 3-4 hr at 37°C after which 10µl of TBM loading buffer was added. The reactions were run on a 40 cm x 20 cm x 0.4 mm non-denaturing 12%, (w/v) polyacrylamide gel (15 ml Accugel (National Diagnsotics Inc), 5 ml 10x TBM, 40µl TEMED, 200µl 20% (w/v) ammoniumpersulphate solution and 30 ml deionised water). Electrophoresis was performed at room temperature for ~4hr at 800V, 20W. Under these conditions the gel warmed to ~35-45°C during running. After the run, the gel was fixed in 10% (v/v) acetic acid and transferred to Whatman 3MM paper, dried under vacuum at 80°C for 1 hr, and finally exposed at –80°C to a Kodak autoradiographic film using an intensifying screen.

2.5 Non-radioactive mini-gel format of the long-short arm gel electrophoresis assay

An alternative, non-isotopic version of the long-short arm assay described in sections 2.2-2.4 was developed, which differed from the original assay in several ways:

The junctions were annealed as follows: 10 pmol of each oligonucleotide was added to 5µl of 10x NEB buffer 2 and water was added to make the total volume 50µl in 200µl PCR tubes. The resulting mixture was placed in a PTC-200 thermal cycler (MJ Research Inc.) and incubated for 25 minutes at 95°C. The temperature was dropped to 20°C at a rate of -0.1°C /sec. The mixtures were either subjected to further digestion reactions or stored at –20°C.

7.5µl of the TWJ solutions was added to 2µl of 10x NEB buffer 2, 2µl of restriction R-ENase (10-20 units) and 8.5µl of water. The mixtures were made up in 200µl PCR tubes and incubated for 3 hr at 37°C in a PTC-200 thermal cycler. 10µl of TBM loading buffer was added to each mixture, followed by brief vortexing. 3.5µl of the reaction solutions was transferred to a 10 cm x 10 cm x 0.1 mm Novex mini-gel cassette, containing a non-denaturing 12%, (w/v) polyacrylamide gel (7.5 ml

Accugel (National Diagnsotics Inc), 2.5 ml 10x TBM, 20µl TEMED, 100µl 20% (w/v) ammonium persulphate solution and 15 ml deionised, molecular biology grade water). Note that for the LSA experiments described in chapters 5 and 6 the TBM electrophoresis buffer was replaced by a tricine/triethanolamine buffer (25 mM each, pH 7.5). This buffer was previously shown to be more resistant to heating and to have higher voltage and current stability during electrophoresis compared to tris-based buffers without loss of resolution (Liu *et al.*, 1999). The gel was run for 3-4hr at room temperature at 90V (voltage was held constant), with the cassette being fully immersed in running buffer on both sides. A DNA ladder ranging from 10-330 bp with 10 bp increments (Gibco/Life Technologies, UK) was run as well to allow cross gel reference. After running, the gel was stained with ethidium bromide/water solution (300µg/200 ml) for 15 minutes. The gel was placed on a transilluminator and photographed using a Kodak Polaroid DS-34 camera with an ethidium bromide specific filter (Elchrom Scientific).

2.6 Chemical probing with diethyl pyrocarbonate (DEPC) or potassium manganate ($KMnO_4$).

2.6.1 DEPC probing (Bailly & Waring, 1997)

2µl of DEPC was added to 5µl of a three-way junction solution, prepared as described in 2.1-2.2, and incubated at room temperature for 2, 10, 15 and 30min. Subsequently 2µl of β-mercaptoethanol, 2µl of 3 M sodium acetate (NaOAc) and 60µl of ethanol were added to the assay mixture to stop the reaction. The precipitated DNA was then washed 2x with 70% ethanol and dried using the speed-vac. Finally, the mixture was subjected to piperidine cleavage, which is described below.

2.6.2 $KMnO_4$ probing (Sambrook *et al.*, 1989)

2µl of a 100 mM $KMnO_4$ solution was added to 5µl of three-way junction solution and incubated for 1-2min at room temperature and the reaction was terminated by addition of 2µl of β-mercaptoethanol, 2µl of 3 M sodium acetate (NaOAc) and 60µl of ethanol. The precipitated DNA was then washed 2x with 70% ethanol and dried using the speed-vac. Finally, the mixture was subjected to piperidine cleavage, which is described below.

2.6.3 Preparation of the G- ladder (Sambrook *et al.*, 1989)

For determining the location of chemical attack by $KMnO_4$ or DEPC, a G sequence ladder of the same 5' labelled oligonucleotide was used. The ladder was prepared by adding 1µl of 10% dimethylsulphate (DMS) to 5µl of labelled oligonucleotide (as prepared in 2.1). The mix was incubated for 2 min at room temperature, after which 50µl of stop solution and 750µl of ethanol was added to precipitate the DNA. Piperidine cleavage as described below was carried out on the samples.

2.6.4 Piperidine cleavage (Sambrook *et al.*, 1989)

Reaction mixtures were placed on dry ice for 10 min, transferred to a pre-chilled (4°C) centrifuge and spun down at 14,000 rpm for 10 min. The supernatant of the mixtures was removed and the pellets washed twice with 70% (v/v) ethanol with 2 min spins at 14,000 rpm in an Eppendorf centrifuge after each washing step. After the second washing step the reaction were tubes placed in a speed-vac and dried for ~5 min. 50µl of 10% piperidine was added to the pellets and the resulting mixtures were incubated for 30 min at 100°C. After the incubation the reaction tubes were placed back into the speed-vac and left until all piperidine had evaporated. The remaining pellets were washed twice with 70% (v/v) of ethanol and twice with water (each washing step was followed by lyophilisation). Finally, 6µl of gel-loading buffer (refer to section 2.1) was added to the pellets and the resulting solutions were subjected to gel electrophoresis as described for the labelling of oligonucleotides in section 2.1.

2.7 The SMART assay

Reactions were carried out using either 50 fmol or 5 fmol of target unless stated otherwise. A typical reaction contains 250 fmol extension probe, 50 fmol template probe and 50 fmol of target, or 25 fmol extension probe, 5 fmol template probe and 5 fmol of target. Reactions were prepared in 20 µl volumes containing, in addition to DNA, 1x T7 transcription buffer (Promega, contains: 40 mM Tris pH 7.9, 6 mM $MgCl_2$, 50 mM NaCl, 8 mM Spermidine) and deionised, double autoclaved water, made up in 200 µl PCR tubes.

The reaction mixture was subjected to a denaturing step at 90°C for 3 min in a PTC-200 thermal cycler, after which the temperature was ramped down to 47°C at –0.1°C/sec. Holding the tubes at 47°C, 1μl of *BstI* DNA polymerase (8,000 U/ml) and 1μl of the dNTP mix (100 μM each, diluted from a 100 mM stock supplied by Amersham-Pharmacia Inc) was added, followed by a 30 min incubation at 47°C. The temperature was then decreased to 37°C (at maximum cooling rate) and 2μl of T7 RNA polymerase and 2μl of rNTP mix (stock solution containing 20 mM of each NTP) was added, bringing the total reaction volume to 20μl. An incubation of 3 hr at 37°C followed, after which 1.6μl of RNase-free DNaseI was added and left for 15 min at 37°C. The reactions with 50 fmol target were subsequently diluted 1:5 by the addition of 79.4μl of deionised, double autoclaved water, to avoid overloading of the well (which can only hold a maximum of 900 fmol of RNA as described below). 5μl of the reaction was subjected to the ELOSA protocol, described next, for the determination of the amount of RNA produced.

2.8 Enzyme-Linked Oligonucleotide Sorbency Assay (ELOSA)

A mixture containing the 'capture probe' and 'alkaline phosphatase probe' (sequences are shown in section 2.1.3) were added to hybridisation solution such that 200μl of solution contained 900 fmol of capture probe and 6 pmol of alkaline phosphatase probe. 200μl of the solution was added to each well in the streptavidin-coated microtiter plate (with the exact number of wells depending on the number of reactions that needed to be assayed) together with 5μl of the reactions described in paragraph 6. The plate was incubated at room temperature for 1 hr with continuous shaking at 300 rpm. The wells were then washed four times with 200μl of 1x TBS and once with 200μl of 1x substrate buffer (diluted from a 5x stock supplied by Boehringer Mannheim). Following these washing steps, 200μl of substrate solution (5 mg/ml p-nitrophenyl phosphate (pNPP), supplied by Boehringer Mannheim, dissolved in deionised water) was added to each well. The plate was transferred to an EMS Lab Systems plate reader, pre-heated at 37°C. The optical density units (ODU) at 405 nm in each well was measured every 30 sec for 30 min with continuous shaking at 300 rpm between readings. The plate reader was connected to a standard PC and the data of the readings recorded by a program (Genesis) provided by the

manufacturer of the plate reader. After the assay, the data was transferred to an Excel 8.0 (Microsoft) workbook and the slope of the line representing the increase of OD units in time (ODU rate) was calculated using the 'SLOPE' function provided with the software package. Data for the first 10 min were discarded since previous measurements indicated that less than 10 min were required for the plate to heat-up to 37°C. Also, all data above ODU=3.5 were discarded since the manufacturer claimed accuracy of reading up to ODU=4. Finally, if the correlation coefficient (R^2) of the linear regression analysis for the ODU increase in time was less than 0.90 the data was discarded.

To determine the concentration of RNA produced in the reactions described in section 2.7, a serial dilution series of a synthetic oligonucleotide with the same sequence as the RNA was made (see 2.1.3). It should be noted that the oligonucleotide was made of DNA rather than RNA and resulting concentrations should therefore be interpreted as approximate. The known amounts of RNA were used to generate a trend line (using the 'TREND' function of the Excel 8.0 spreadsheet program), correlating RNA concentration to the corresponding rate of the ODU increase in time. The ODU rates calculated for the reactions with unknown amounts of RNA were then correlated to this trend line, giving approximate values for the amount of RNA produced. Again the data were discarded if the trend line had a correlation coefficient of less than 0.90.

Finally, for each experiment or set of experiments two types of calculations were carried out. The procedure described above yields absolute values for the amount of RNA produced. An alternative calculation method however, produces a clearer distinction between the RNA produced in the presence of wild-type targets and in the presence of mutant target. The absolute amount of RNA produced in the presence of mutant target is simply expressed as a percentage of the absolute amount of RNA produced in the presence of wild-type. In case of multiple repeat experiments values were averaged but it is important to realise that differences exists in the calculations for the average absolute values and the average relative values. Averages for absolute values were calculated as averages over all experiments involved:

Average absolute amount of RNA ($\bar{A}$) over N experiments:

$$\bar{A} = \frac{\sum_{i=1}^{N} A_i}{N} \quad (1)$$

The average relative amount of RNA ($\bar{A}_{rel}$) over N experiments was calculated as:

$$\bar{A}_{rel} = \frac{\sum_{i=1}^{N} \left(A_{i,mut} / A_{i,wt} \right) \bullet 100}{N} \quad (2)$$

$A_{i,mut}$ and $A_{i,wt}$ are the absolute yields obtained in the presence of mutant and wild-type target respectively in experiment i. Note that equation (2) does not calculate A_{rel} using the average absolute amounts of RNA produced. This means that graphs for absolute values and relative values will not necessarily show the same patterns.

2.9 Naming conventions for different junctions

2.9.1 Naming of junctions used for structure assessment

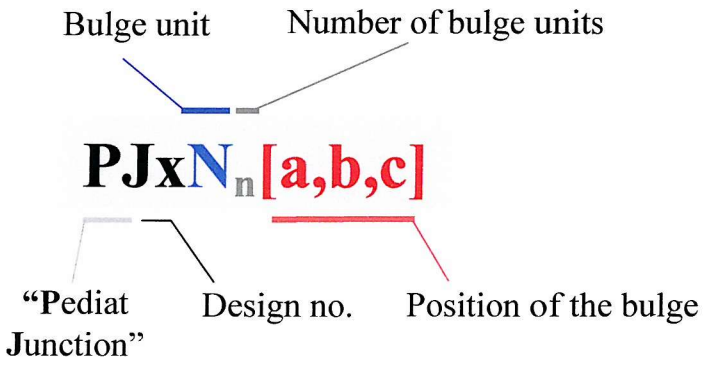
Although the nomenclature committee of the international union for biochemistry and molecular biology (NC-IUBMB) recently presented several guidelines for the naming of branched DNA species (Lilley *et al.*, 1995), the present study uses a slightly different method for naming the junctions. The main reason for this is that the official guidelines do not include conventions for junctions containing multiple bulges in different strands, or conventions for indicating the position of a bulge within the junction. For the present study, it is important to indicate whether a bulge is positioned within the extension, template or target DNA. Therefore, the following rules were applied:

1) All junctions start with PJ for PEDIAT Junction (PEDIAT stands for primer extension dependent isothermal amplification technology, which during the course of this thesis was changed into SMART).
2) The number following PJ indicates the chronological order of design. Thus, PJ3 would be the third junction designed after the start of the project.
3) Then follows the type, number and position of the bulge(s). The composition of the bulge is indicated by one-letter abbreviations (e.g. C for cytosine) followed

by a subscript number indicating the number of bulge units present. Then follows a lower case letter indicating the position in the junction at the branch point; a = template probe, b = extension probe, c = target. Thus, PJ3C_1a indicates a PJ3 junction with one unpaired cytosine present in the template probe strand. Note that the subscript number refers to bulge *units*, which means that PJ3CT_1a is a PJ3 junction with one 5'CT3' bulge present in the template probe strand (all sequences are printed 5'→3') and PJ3CT_2a contains a 5'CTCT3' bulge.

4) Multiple bulges in different strands are indicated in a consecutive manner, following the strands in alphabetical order (i.e. from a to c). For instance, PJ3C_2aT_3c is a PJ3 junction with two unpaired cytosines in the template probe strand and three unpaired thymidines in the target strand. If identical bulges are present at multiple sites at the branch, the naming can be abbreviated as follows: PJ3C_2abc.

This can be summarised as follows:



A number of examples are shown in Fig. 2.2.

2.9.2 Naming of SMART junctions

The naming of the junctions used in SMART is severely complicated by the asymmetric structures of the TWJs used. The three different arms may have different numbers of base pairs, for instance the overlap region is much shorter than the remaining arms. Also, different transcription template sequences, different chemical linkers present at the branch point, DNA and DNA analogs etc are used, which complicate a coherent naming convention. Therefore, a simplified naming

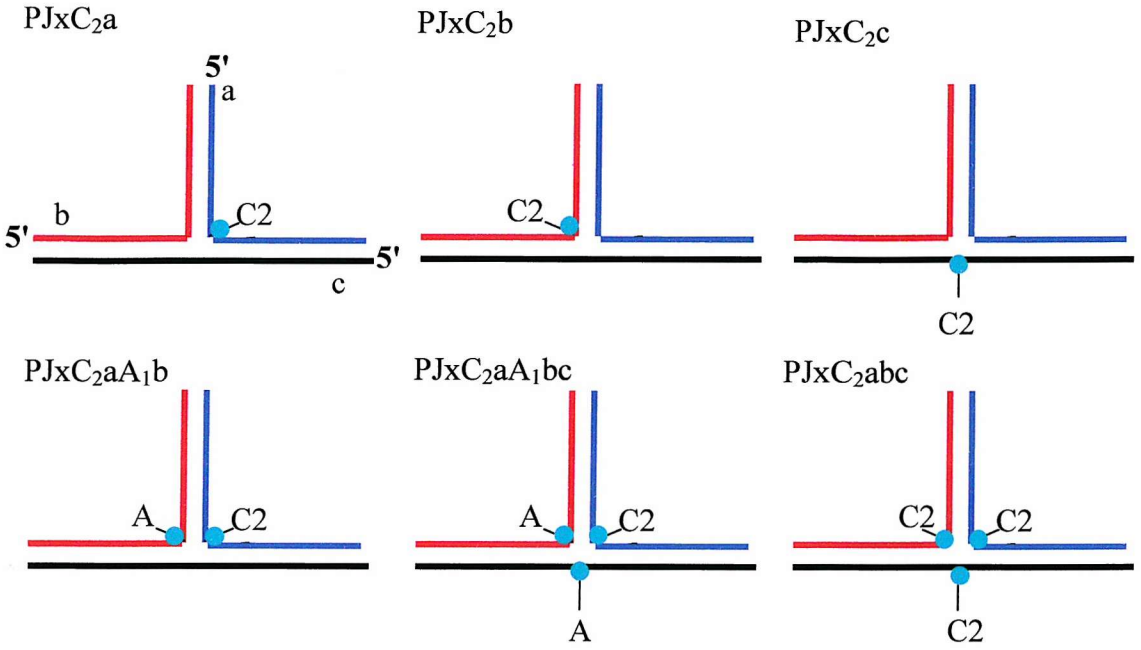


Fig. 2.2. Naming convention for junctions used for structural analysis. Each strand in a junctions is designated a (blue), b (red) or c (black) which correspond to the SMART template probe, extension probe and target strands respectively. Bulges are indicated by the green circles and the identity of the unpaired base(s) is indicated.

convention was used, which describes the most important characteristic design features of a certain TWJ. A typical example is as follows:

J-CF1A-(20^{PNA}-7-16^{PNA})

J-CF indicates that the (J) junction involved part of the CFTR gene, with number 1 symbolising the location of the branch point, which is illustrated elsewhere in the text. Then follows the letter A or B, where A indicates that the TWJ produced RNA1 and B indicates RNA3 production (refer to 2.1.9). Between brackets the size of the different arms are indicated in the order: extension probe/target arm - overlap arm - template probe/target arm. Finally, in superscript the type of nucleic acid involved is indicated. Note that this refers only to the sequences present in the extension probe and template probe since both the target and the overlap/transcription arm are always DNA. For clarity, the inclusion of chemical linkers or any other modification (e.g. changes in the overlap sequence) is omitted and will be indicated in the text.

3 Development of SMART for the detection of small genetic polymorphisms

3.1 Introduction

The general principle of the SMART assay is that even a single base change in the target-spanning region should prevent the formation of the three-way junction. As outlined in chapter one, several options are available in the context of SMART that may enhance its sensitivity towards mismatches. This chapter deals with one of these, namely the higher sensitivity some DNA analogues have shown towards mutations when compared to DNA duplexes. Since initial experiments were not able to discriminate between targets containing 1-3 base differences, further optimisation experiments were performed to establish the optimal reaction conditions that would allow such discrimination.

3.2 Assay optimisation and assessment of PNA and DNA for the detection of small genetic variations

3.2.1 Design of junctions J-CF1A-(15^{PNA}-7-16^{PNA}) and J-CF1A-(20^{DNA}-7-16^{DNA})

The effect of PNA on mutation sensitivity was assessed using the PNA-DNA junction J-CF1A-(15^{PNA}-7-16^{PNA}) shown in Fig. 3.1. This was compared with the DNA-only junction, J-CF1A-(20^{DNA}-7-16^{DNA}). In J-CF1A-(15^{PNA}-7-16^{PNA}) the target complementary sequences in the extension and template probes were replaced by PNA with lengths of 15 (extension) or 16 (template) bases. The PNA sequence was coupled to the remaining DNA sequences via a synthetic linker [(penta- (extension probe) and hexa- (template probe) methylene]. The overlap region was not converted into PNA for two reasons. Firstly, DNA polymerases are not able to initiate polymerisation from a PNA residue. Secondly, PNA increases the stability of PNA/DNA and PNA/PNA duplexes substantially when compared to standard DNA duplexes. The latter could, potentially, enhance the production of non-specific (i.e. target-independent) signals due to probe-probe interactions.

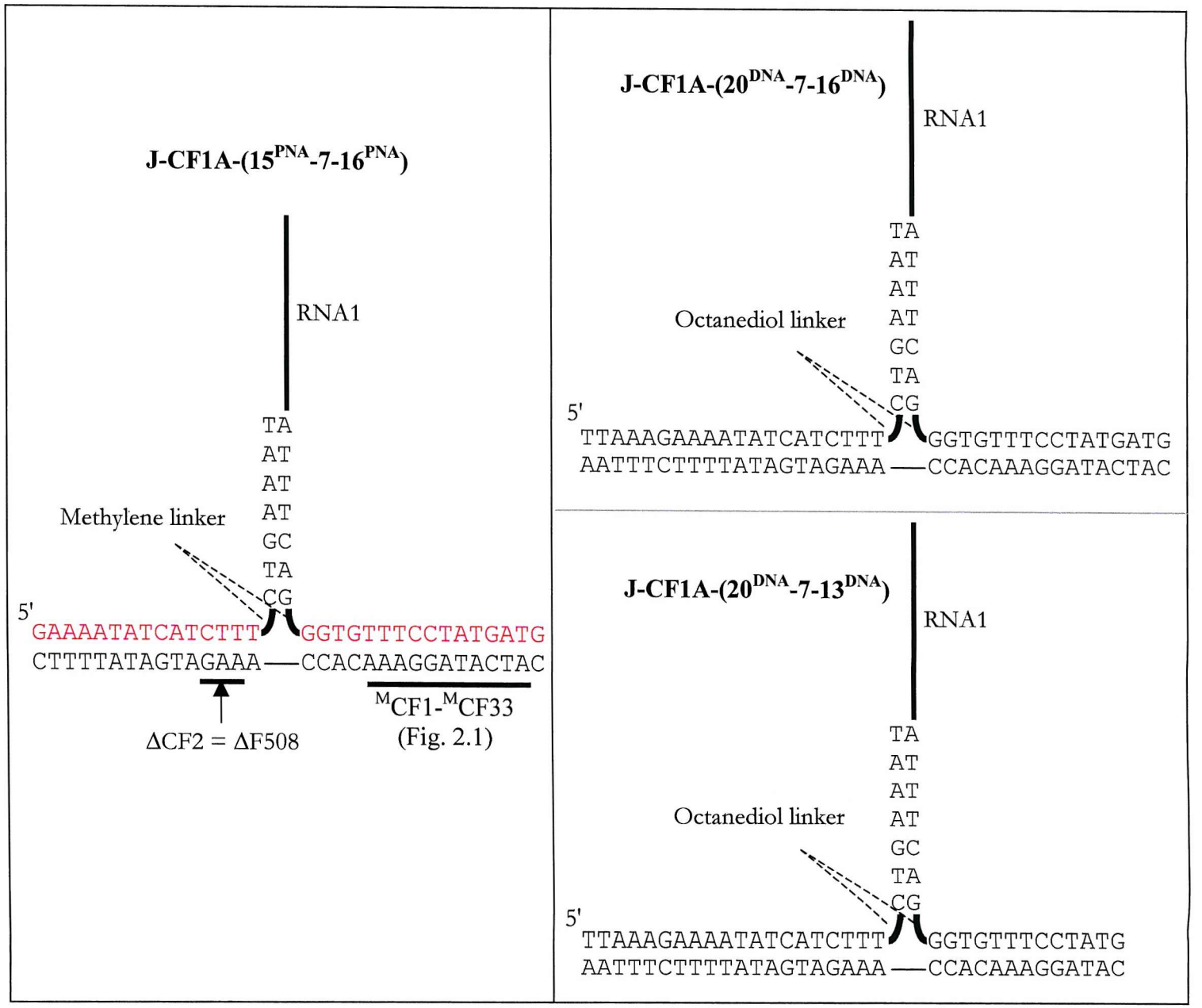


Fig. 3.1. Junction designs based on the CFTR gene. Nucleotides drawn in red are PNA. The locations of the mutations studied in this chapter are indicated in the first junction (see also Fig. 2.1). ΔCF2 is a three base deletion, whereas ^{M}CF1-^{M}CF33 represent each possible single base change in the underlined region.

J-CF1A-(15^{PNA}-7-16^{PNA}) and J-CF1A-(20^{DNA}-7-16^{DNA}) were designed incorporating part of the cystic fibrosis transmembrane regulator (CFTR) gene. The most abundant mutation occurring in the human, Caucasian population is usually referred to as the ΔF508 deletion. This results in the deletion of three base pairs corresponding to a phenylalanine amino acid in the sodium ion channel, causing the disease cystic fibrosis (refer to section 1.1.2). Another naturally occurring deletion is ΔI507. The two regions where the deletions occur are located in the foot region of the extension probe of the two junctions. The ΔF508 deletion (ΔCF2 in Fig. 2.1) was located close to the branch point whereas ΔI507 was located closer to the centre of the extension probe/target arm. Finally, a range of mismatched target oligonucleotides were synthesised as shown in Fig. 2.1. All these target oligonucleotides were based on the wild-type CFTR gene, but each containing a single mismatch at different positions in the template probe/target arm and extension probe/target arm. These oligonucleotides were used to test the sensitivity of the PNA chimeras for mismatches.

There are also differences between J-CF1A-(15^{PNA}-7-16^{PNA}) and J-CF1A-(20^{DNA}-7-16^{DNA}). These are outlined in Fig. 3.1, which shows that the DNA junction contained a longer foot region in the extension probe. The DNA junction was designed prior to the PNA-DNA junction and standard SMART junction designs usually incorporate longer feet regions. However, due to solubility problems the maximum size of PNA molecules that could be synthesised was around 16 bases, which is why the extension probe in J-CF1A-(15^{PNA}-7-16^{PNA}) was shorter. The DNA junction also contained octanediol linkers rather than hexa- and pentamethylene linkers between the target complementary sequences and the overlap sequences, again due to differences in synthesis requirements. A direct comparison between DNA and PNA is complicated due to their different hybridisation kinetics and stabilities. Ideally, one would design the DNA junction so that the melting temperature of both target-containing arms is equal and that both junctions assume the same conformation. The effect of PNA on TWJ formation has so far not been studied and only a small number of studies have attempted to predict melting temperatures for DNA/PNA duplexes. Thus, a comparison between the two different junctions can only be approximate.

3.2.2 Analysis of the optimal DNA polymerase incubation temperature for mismatch discrimination with J-CF1A-(15^{PNA}-7-16^{PNA})

As outlined in Fig. 1.3 and section 2.7, the first part of the SMART consists of two major steps: a DNA polymerase incubation step and a DNAP plus T7 RNAP incubation step. Since the DNA polymerase used in SMART was isolated from a thermophilic *Bacillus stearothermophilus* strain, and possessed an optimal activity temperature of 65°C, the DNAP plus T7 RNAP incubation step constituted the step with the most flexibility for manipulating the formation of the TWJ by temperature. By testing a range of temperatures it should be possible to find optimal conditions where wild-type target induces formation of the TWJ whereas mutant target will not. It was necessary to test the effect of temperature as no accurate predictions could be made on the effect of mismatches on the melting temperature of the TWJ. The reasons for this include the fact that TWJ melting behaviour remains poorly characterised and that significant levels of monovalent and polyvalent cations ($MgCl_2$ and spermine) were present in the reaction buffer. The latter can have a substantial impact on the melting. Experimentally tested nearest-neighbour values (used to predict melting temperature values) are currently only available for B-DNA-duplexes in the presence of 1M NaCl. Thus having the ability to vary the assay temperature in at least one of the two steps could potentially serve as an indirect indication for the melting temperature of junctions in the presence/absence of mutations. More signal is expected at temperatures below the Tm of the junction and less signal above its Tm.

Fig. 3.2 shows the effect of different DNA polymerase incubation step temperatures, using three mismatches (^{M}CF40, 41 and 45) located in the centre of the template probe/target arm in J-CF1A-(16^{PNA}-7-15^{PNA}). The upper plots show the total RNA yield under the various conditions, while the lower plots show the yield as a percentage of that with wild-type target. It can be seen that performing the DNA polymerase incubation step step at 47°C appeared optimal (>80% signal difference between wild-type and mutant targets) since above this temperature there was a significant drop in RNA yield and no improvement in mismatch sensitivity. At 37°C, the discrimination between wild-type and mutant targets ^{M}CF40 and ^{M}CF41 was

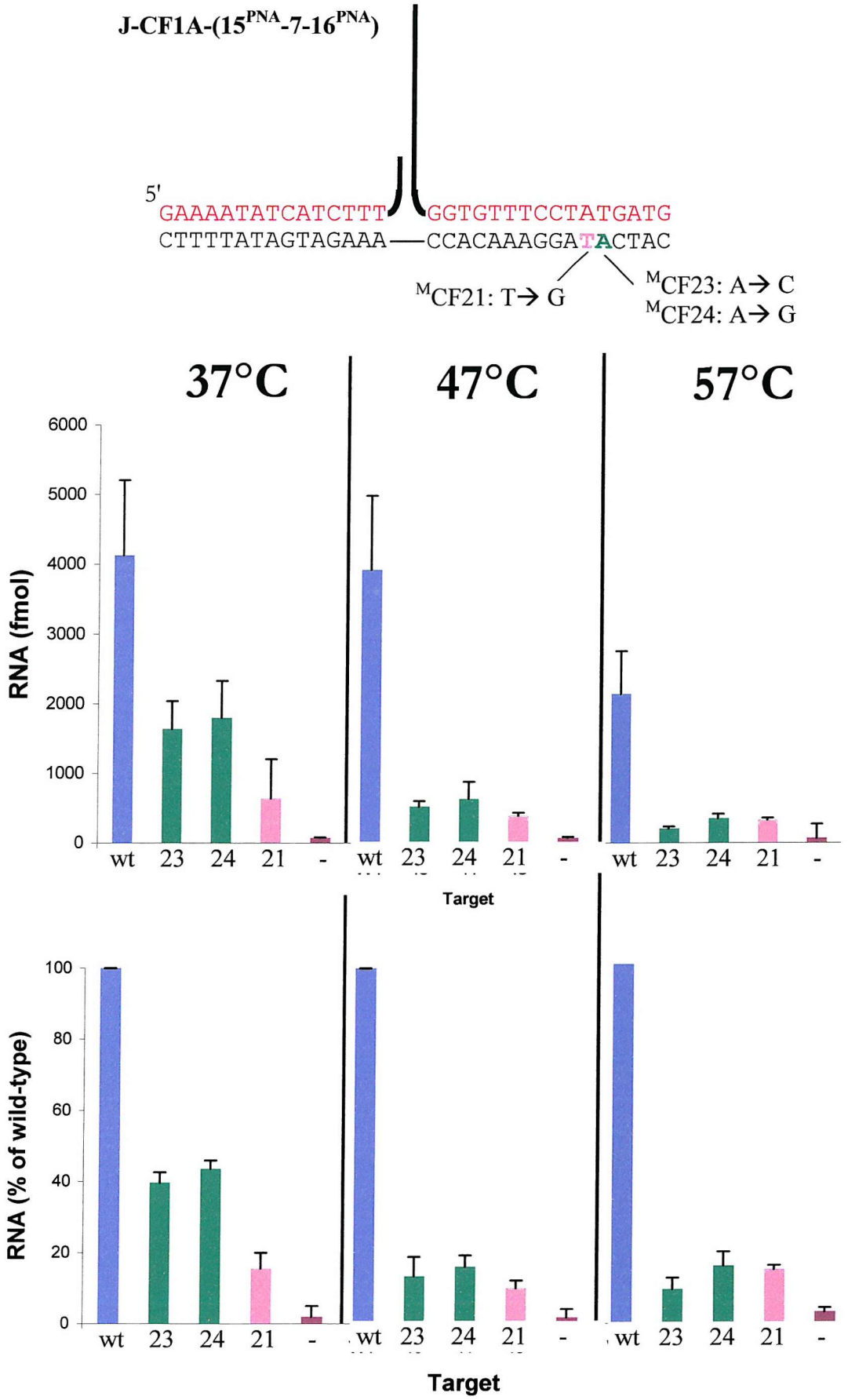


Fig. 3.2. Effect of DS incubation temperature on the mismatch sensitivity of J-CF1A-(15^{PNA}-7-16^{PNA}). The numbers correspond to ^{M}CF21, ^{M}CF23 and ^{M}CF24 shown in Fig. 2.1 and above. The experiment was repeated five times using 50 fmol of target, 250 fmol extension probe and 50 fmol template probe.

reduced compared to 47°C. Based on these results, further reactions were carried out at 47°C during the DNA polymerase incubation step.

3.2.3 Assessment of the optimal location for mutations using J-CF1A-(16^{PNA}-7-15^{PNA}) and comparison with J-CF1A-(20^{DNA}-7-16^{DNA})

Having established that mismatch discrimination (using synthetic ssDNA target) was possible using PNA/DNA chimeras, experiments were initiated to map the mutation sensitivity of the TWJ. Deletions and mismatches were assessed at varying positions from the branch point and the results are presented in Fig. 3.3 and Fig. 3.4. Fig. 3.3 shows the effect of 3-base deletions around the branch point. Experiments were repeated five times, except for deletion ΔCF1, which was repeated only once. Clearly, discrimination between wild-type and mutant sequences was obtained as all deletion targets showed reductions in signal. This was corroborated by the statistical analysis, shown in table 3.1. All deletions except ΔCF1 show significant p-values whether based on relative or absolute values. The exception, ΔCF1, was maybe due to the very small data set. Also, notable differences were observed between those deletions located at the branch point and those located further into the template probe/target arm, with the ΔCF4 deletion showing the least discrimination. The results imply that the region around the branch point is the least sensitive area, while it appears that deletions located closer to the centre of the template probe/target arm are more destabilising giving greater discrimination. The extension probe/target arm was not investigated so no conclusions can be drawn on the sensitivity of this arm. Differences are also apparent between the two graphs as the patterns for the absolute amount of RNA produced do not match those found for the relative amount. This can be explained by referring to Fig. 3.4, where the standard deviations calculated for the average absolute amounts of RNA and relative amounts of RNA are compared. Clearly, standard deviations are lower for the majority of targets when signals were calculated as a percentage of the wild-type signal. This suggests that although the absolute RNA yields varied from experiment to experiment, internally experiments were more consistent. In other words, the variation between experiments is much greater than the errors within each experiment. Fig. 3.4 indicates that calculating the amount of RNA produced as a percentage of wild-type signal corrects for part of the

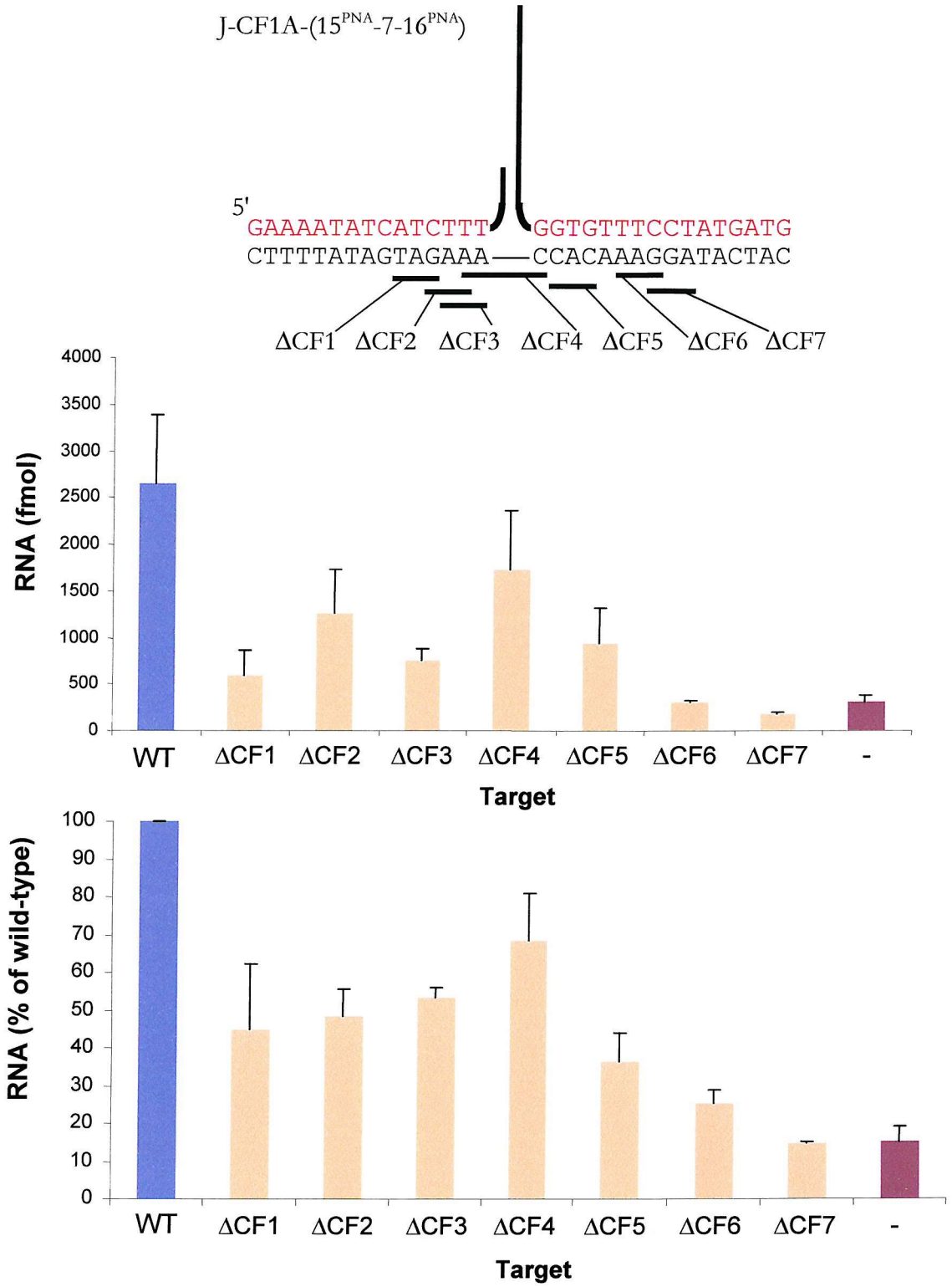


Fig. 3.3 Effect of deletions on SMART signals using J-CF1A-(15^{PNA}-7-16^{PNA}). Experiments were conducted with 250 fmol extension probe, 50 fmol template probe and 50 fmol single-stranded target. The deletions are indicated by the lines in the top figure (see also Fig. 2.1). The ΔF508 deletion is ΔCF2 and the ΔI507 deletion is ΔCF1. Averages were calculated over at least three repeated experiments. The top graph shows the absolute RNA yields in fmol/reaction. The bottom graph shows the RNA yields calculated as a percentage of the absolute signal with wild-type target. Refer to section 2.8 for information on the calculations performed. WT= wild-type, - = no target.

Table 3.1. Statistical analysis of the effect of deletions on the RNA produced with J-CF1(15^{PNA}-7-16^{PNA}). RNA yields were compared with yields in the presence of wild-type using a one-tailed, paired t-test (α=0.05, N=5, except for ΔCF1 which was repeated only once). Bold are values less than 0.05.

Target	p-value using absolute RNA yields	p-value using relative RNA yields
ΔCF1	0.0995	0.1326
ΔCF2	**0.0093**	**0.0023**
ΔCF3	**0.0029**	**0.0022**
ΔCF4	**0.0046**	**0.0248**
ΔCF5	**0.0191**	**6.64E-05**
ΔCF6	**0.0165**	**0.0197**
ΔCF7	**0.0133**	**0.0020**
No target	**7.32E-07**	**2.22E-16**

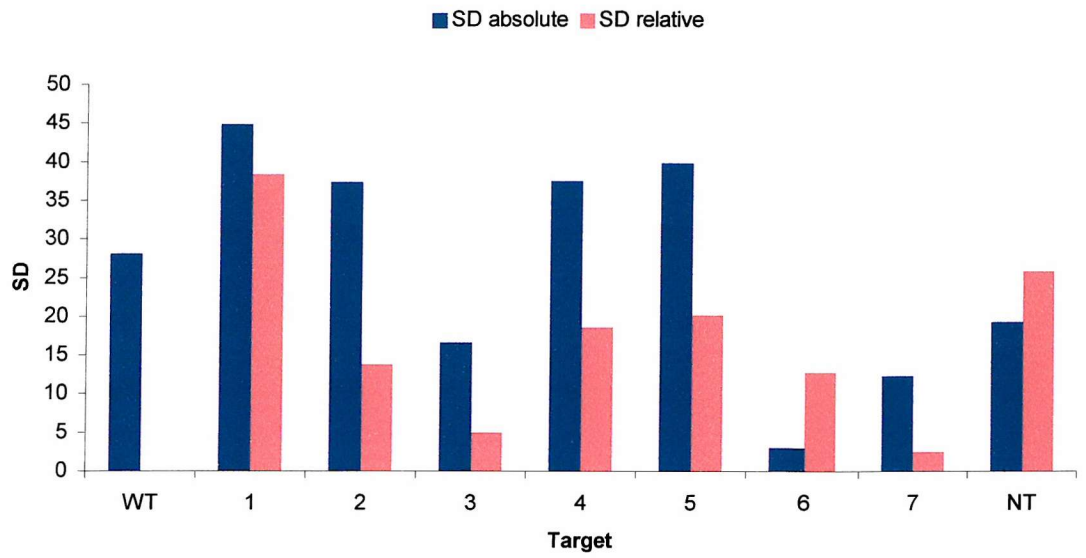


Fig. 3.4. Difference in standard deviation (SD) calculated for the average absolute amount of RNA and relative amount of RNA. To allow comparison, the SD was calculated as a percentage of the average calculated signal (*i.e.* normalised). Thus, SD absolute was calculated as the percentage of the average absolute amount of RNA and SD relative was calculated as a percentage of the average relative amount of RNA.

observed experimental variation. Finally, a one-tailed, paired t-test comparing the normalised standard deviations of the absolute and relative values suggested they were significantly different ($p=6.41*10^{-5}$).

We next examined how single base changes at different positions in J-CF1A-(16^{PNA}-7-15^{PNA}) affected the signal from the SMART assay. The results are presented in Fig. 3.5 and reveal that single base changes in J-CF1A-(16^{PNA}-7-15^{PNA}) produced a relatively constant decrease in yield at all positions (base pairs 5-15 relative to the branch point). Most mismatches decreased the RNA yield by between 60-85% and all displayed significant differences with wild-type yields ($p<0.05$, paired, one-tailed t-test). One exception was the A→C mismatch at position 15, yielding only ~25% reduction in signal compared to wild-type ($p=0.068$). The reason for this remains elusive, although the affected base was located close to the terminus of the template probe/target arm. However, the A→G and A→ T mismatches at position 15 showed a much larger reduction in signal despite being located at the same position as the A→C mismatch. Other studies have shown that the level of destabilisation caused by a mismatch is dependent on the surrounding sequence (see for instance Allawi & SantaLucia, 1998), which could explain why some mismatches yielded a higher reduction in signal than others.

Fig. 3.6 shows the results of the same experiment but using DNA instead of PNA probes and reveals that substantially less discrimination was detected between wild-type and mutated targets. This supports other studies that have reported a higher sensitivity of PNA to mismatches. However, as noted before, there are other differences between the PNA and DNA probes which could potentially affect the mutation sensitivity and TWJ conformation.

Finally, as noted for the deletion experiments, there is a substantial difference in standard deviations calculated for the absolute and relative yields with J-CF1A(15^{PNA}-7-16^{PNA}) ($p<<0.05$). Comparison between the yields obtained over different experiments for wild-type and a A→C-mismatched (at position 10 in Fig. 3.5) target shows a strong correlation ($n=10$, $R^2 = 0.90$, Fig. 3.7). The correlation coefficients for other mismatches compared to wild-type yields were also significant, ranging between 0.63 and 0.98. This indicates that absolute yields may vary substantially from experiment to experiment. Thus, as noted before, absolute yields

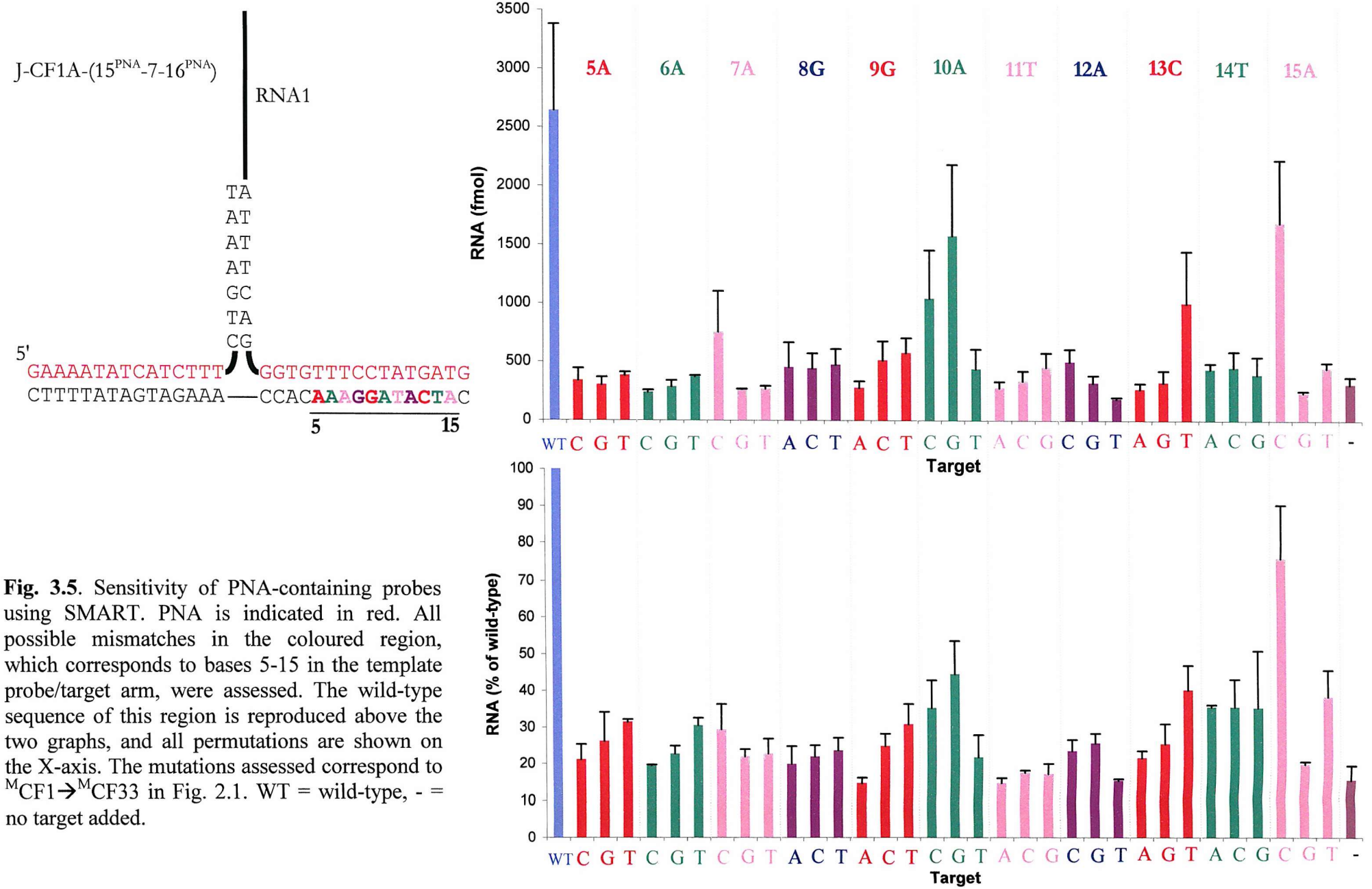
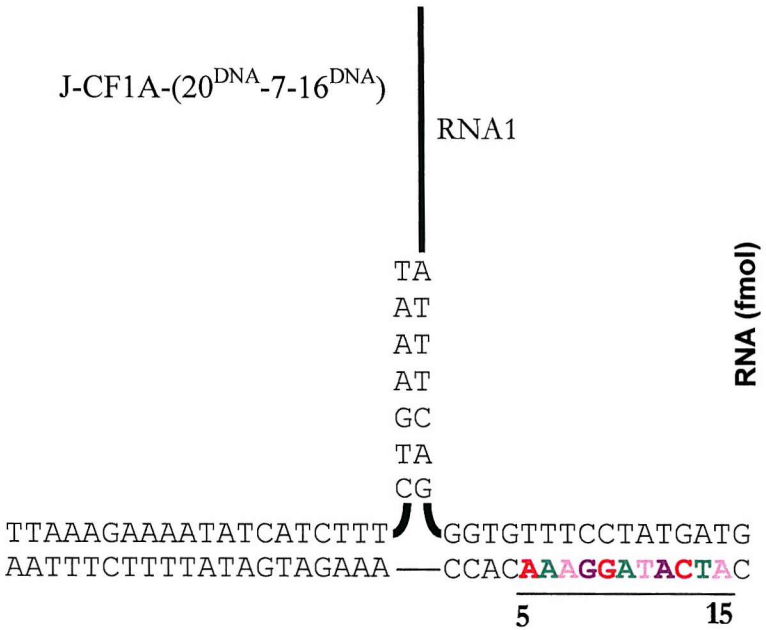


Fig. 3.5. Sensitivity of PNA-containing probes using SMART. PNA is indicated in red. All possible mismatches in the coloured region, which corresponds to bases 5-15 in the template probe/target arm, were assessed. The wild-type sequence of this region is reproduced above the two graphs, and all permutations are shown on the X-axis. The mutations assessed correspond to ^{M}CF1→^{M}CF33 in Fig. 2.1. WT = wild-type, - = no target added.



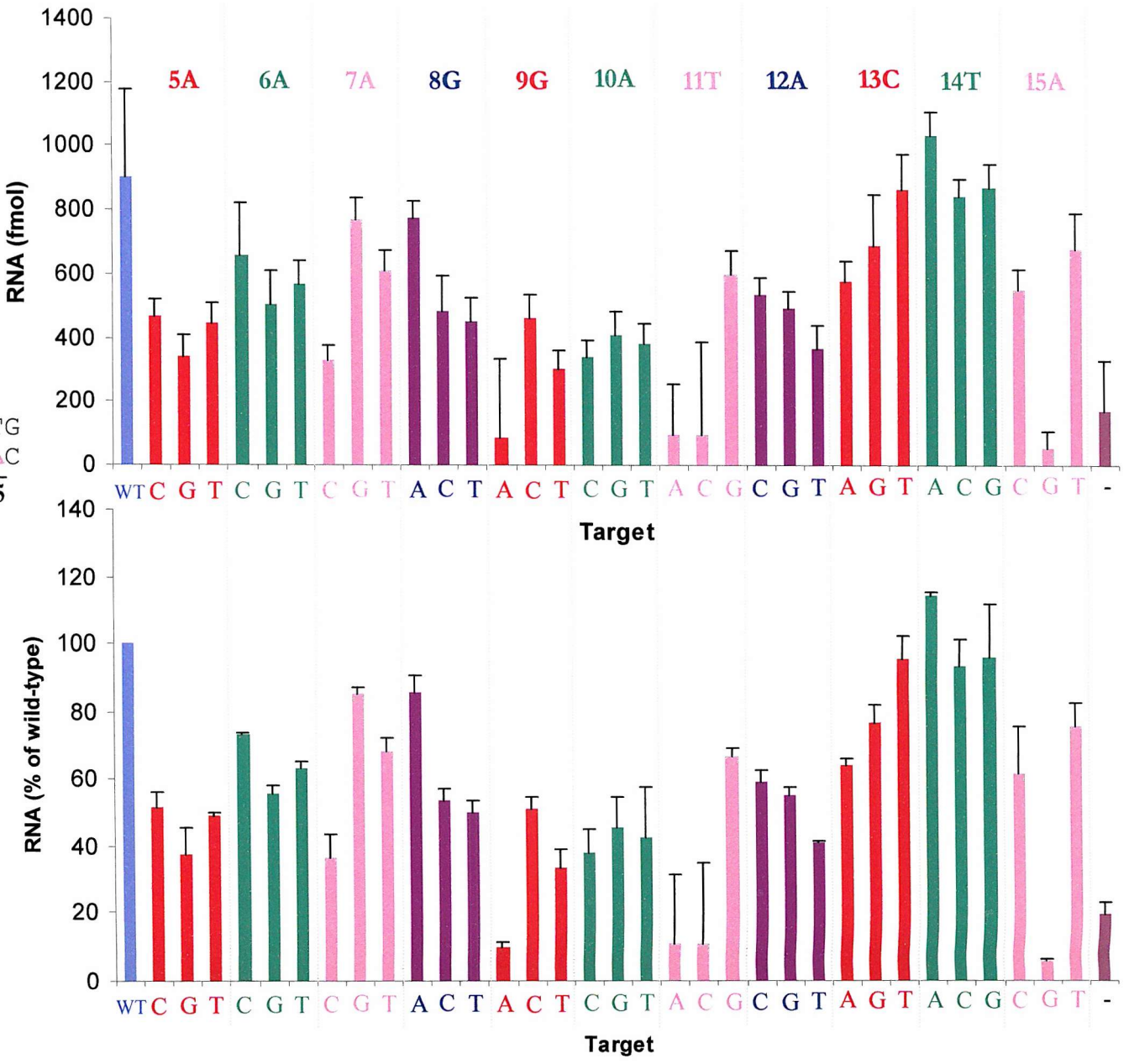


Fig. 3.6. Sensitivity of DNA-containing probes using SMART. All possible mismatches in the coloured region, which corresponds to bases 5-15 in the template probe/target arm, were assessed. The wild-type sequence of this region is reproduced above the two graphs, and all permutations are shown on the X-axis. The mutations assessed correspond to ^{M}CF1→^{M}CF33 in Fig. 2.1. WT = wild-type, - = no target added.

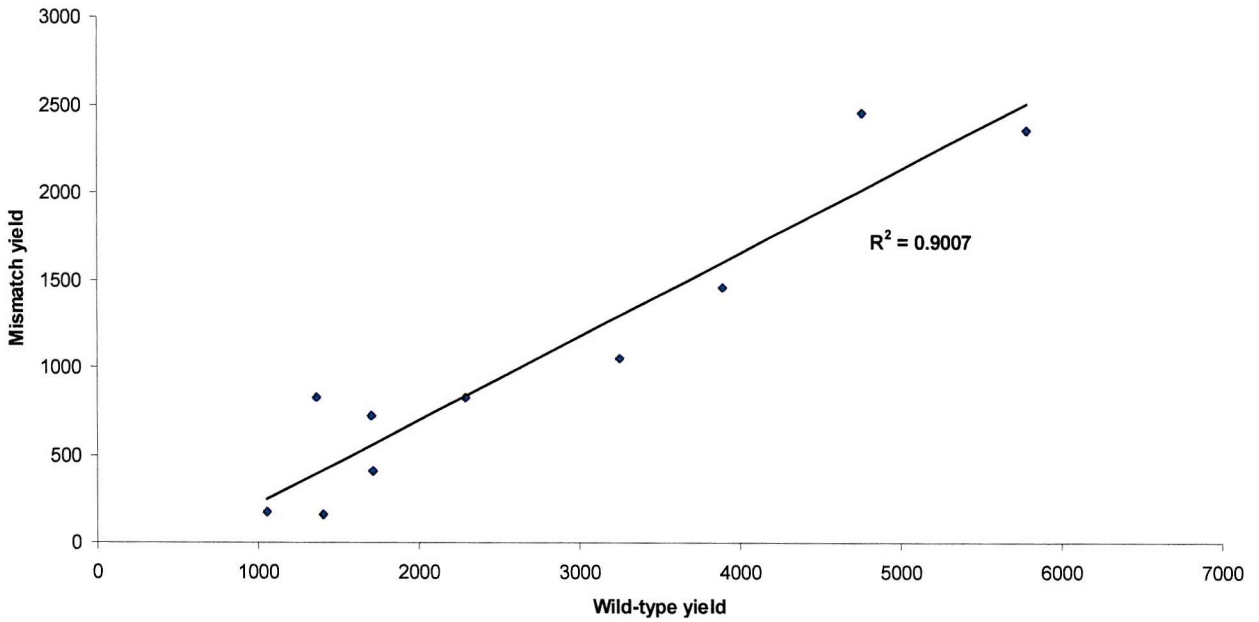


Fig. 3.7. Correlation in wild-type and CA mismatch (position 10 in Fig. 3.5) yields over 10 experiments. The strong correlation clearly shows that although absolute yields vary substantially from experiment to experiment, the relative difference between wild-type and mutant target yields remains largely unaffected.

vary from experiment to experiment but the difference in yield between wild-type and mutant targets remained relatively constant. Therefore, the relative yields provide a more accurate reflection of the effect of mismatches since it removes effects caused by internal experimental errors. The source of the experimental variation remains unknown but it should be noted that the assay involves a great number of steps that each can contribute to the observed error. Extensive multivariate analysis could provide a more accurate indication of where the experimental errors originated from by varying different reaction conditions and assessing the effect on RNA yield.

3.2.4 Influence of template probe/target arm size on mismatch sensitivity using DNA probes

The length of the template probe target-complementary sequence was reduced to 13 bases (J-CF1A-(20^{DNA}-7-13^{DNA}), Fig. 3.1) to assess whether lowering the Tm of the template probe/target arm would affect the mismatch sensitivity. The results are shown in Fig. 3.8 and were obtained using 50 fmol of target with mismatches located in the template probe/target arm. Compared to a 16 base template probe foot, a substantial increase in sensitivity was detected for most mutations when the size of the foot region was reduced to 13 (compare Fig. 3.6 with Fig. 3.8). Less sensitivity was observed for mismatches located close to the terminus of the template probe/target arm (mismatches at positions 12 and 13), presumably due to a higher degree of fraying at the termini of DNA duplexes. Also, the A→G mismatch at position 7 showed a comparatively high yield (reduced by only ~30% compared to wild-type), whereas other mutations at the same location caused a much greater reduction in signal (~75%). All mismatches showed significant differences from wild-type yields ($0.012 < p < 0.042$, one-tailed, paired t-test).

Finally, the absolute yields showed a strong correlation between wild-type and mutant targets (R^2 between 0.92 and 0.99), supporting earlier conclusions stating that the relative yields more accurately reflect the effect of mismatches.

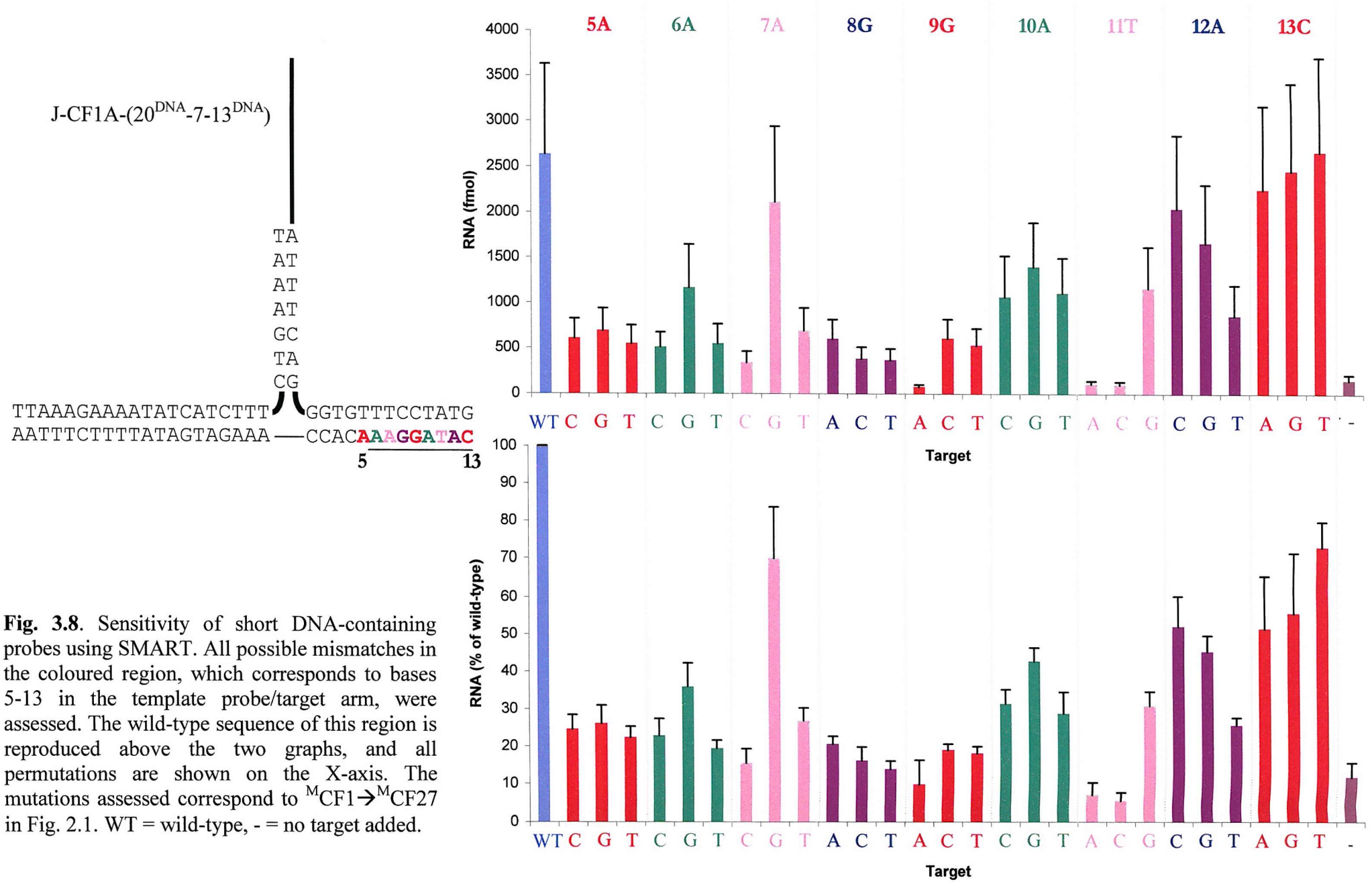


Fig. 3.8. Sensitivity of short DNA-containing probes using SMART. All possible mismatches in the coloured region, which corresponds to bases 5-13 in the template probe/target arm, were assessed. The wild-type sequence of this region is reproduced above the two graphs, and all permutations are shown on the X-axis. The mutations assessed correspond to ^{M}CF1→^{M}CF27 in Fig. 2.1. WT = wild-type, - = no target added.

3.3 Assessment of the mutation sensitivity of LNA in SMART

3.3.1 Mutation sensitivity of J-CF1A(9^{LNA}-7-9^{LNA})

The preceding section highlighted the potential of a DNA-analogue, PNA, for the detection of point mutations using SMART. However, as outlined in section 1.2.2 alternatives to PNA are available, such as locked nucleic acids (LNA). We were therefore interested to assess whether LNA could yield a similar (or better) improvement in mutation sensitivity as observed with PNA and altered J-CF1A so that both target-complementary sequences in the two probes now contained LNA (Fig. 3.9, J-CF1A-(9^{LNA}-7-9^{LNA})). These LNA-DNA chimeras were obtained from Exiqon A/S, Denmark and contained only 9 target complementary bases in each probe because of the (allegedly) much higher stability of LNA when compared to DNA (Koshkin *et al.*, 1998).

Fig. 3.10 shows the results of an experiment comparing the RNA yields in the presence and absence of target (carried out using 50 fmol of each probe and target). Clearly, unexpected high background signals were obtained, and the control experiment shown in Fig. 3.11 indicates that this background originated from target-independent interactions between the two probes. However, where the extension probe was replaced with DNA (the same extension probe used in J-CF1A-(20^{DNA}-7-14^{DNA}) while retaining the LNA template probe, clear discrimination was achieved as shown in Fig. 3.12. However, this figure also indicates a substantial reduction in RNA yield (700 fmol) compared to J-CF1A-(20^{DNA}-7-14^{DNA}) (1200 fmol) and J-CF1A-(15^{PNA}-7-16^{PNA}) (2500 fmol).

3.3.2 Mutation sensitivity of junctions containing mixed DNA/LNA target complementary sequences

In the light of the previous results a modified, LNA-containing junction was designed (J-CF1A($15^{DNA/LNA}$-7-$16^{DNA/LNA}$), Fig. 3.9), in which the extension probe/target and template probe/target arms contained only three LNA residues at the centre of each arm. It was hoped that reducing the number of LNA residues would reduce the large background signal, and improve the sequence specificity of the junction (since J-CF1A-(9^{LNA}-7-9^{LNA}) contained only 9 target complementary bases

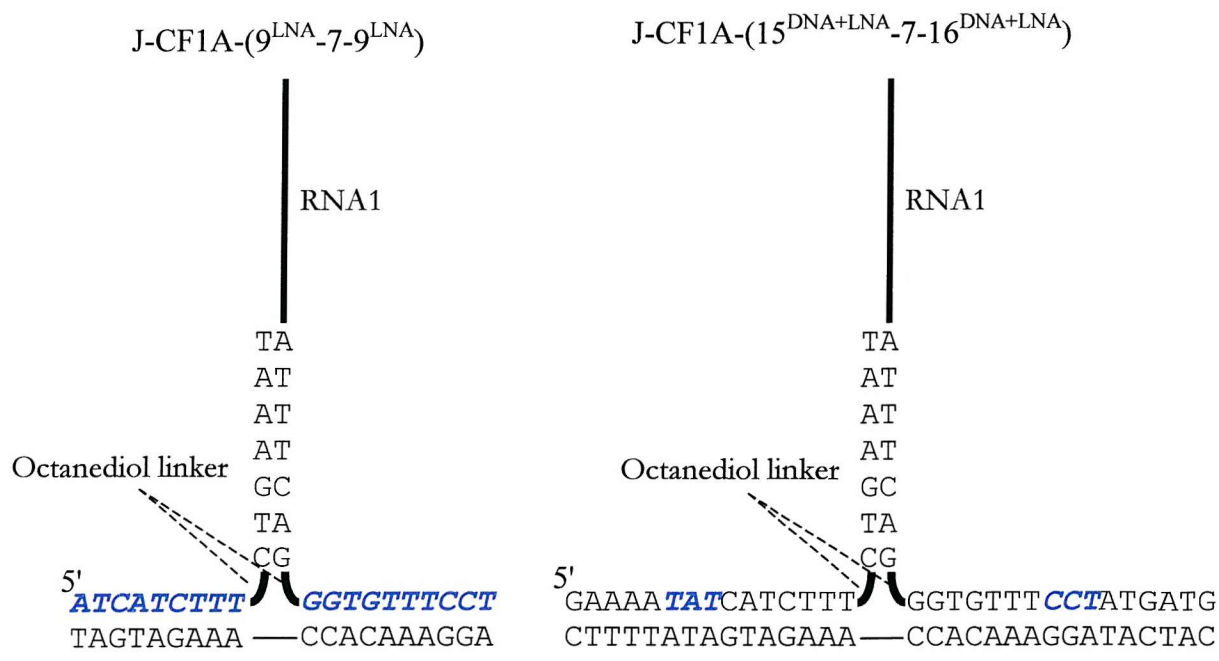


Fig. 3.9. Design of LNA-containing J-CF1A junctions. Blue italics indicates LNA.

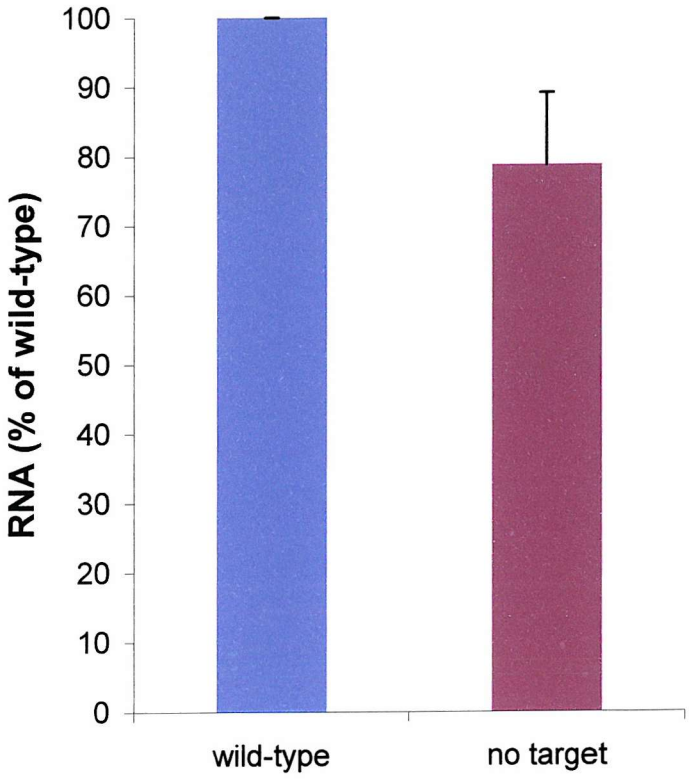


Fig. 3.10. SMART results for J-CF1A-(9^{LNA}-7-9^{LNA}). N=3.

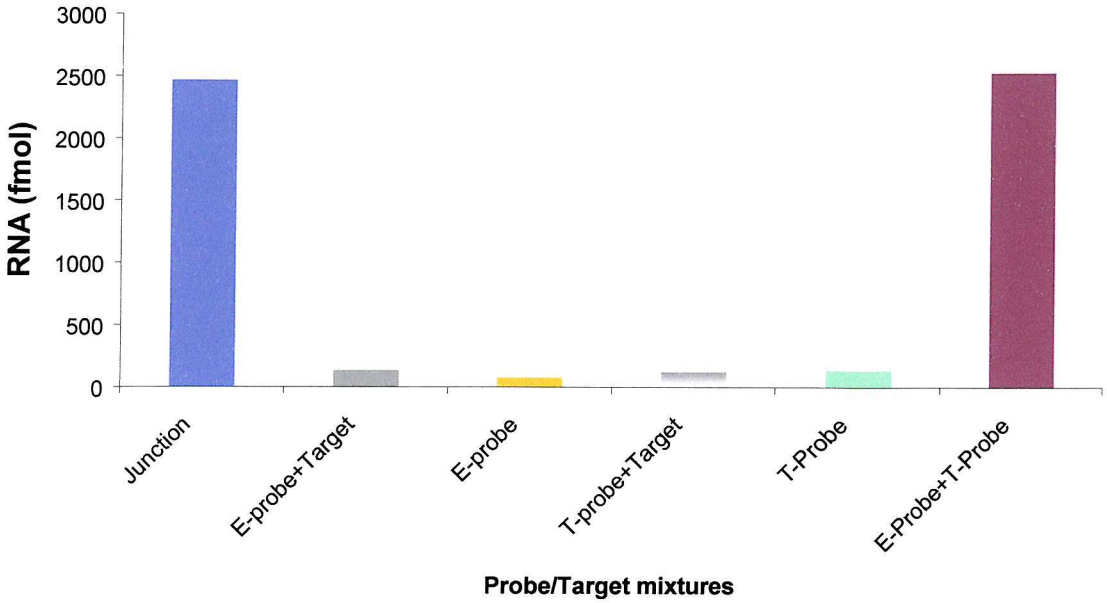


Fig. 3.11. Analysis of background signals generated by J-CF1A-(9^{LNA}-7-9^{LNA}). All possible probe/target combinations were assessed, using 250 fmol extension probe (E-probe), 50 fmol template probe (T-probe) and/or 50 fmol target. The graph clearly shows that the high background was caused by some interaction between the extension probe and template probe.

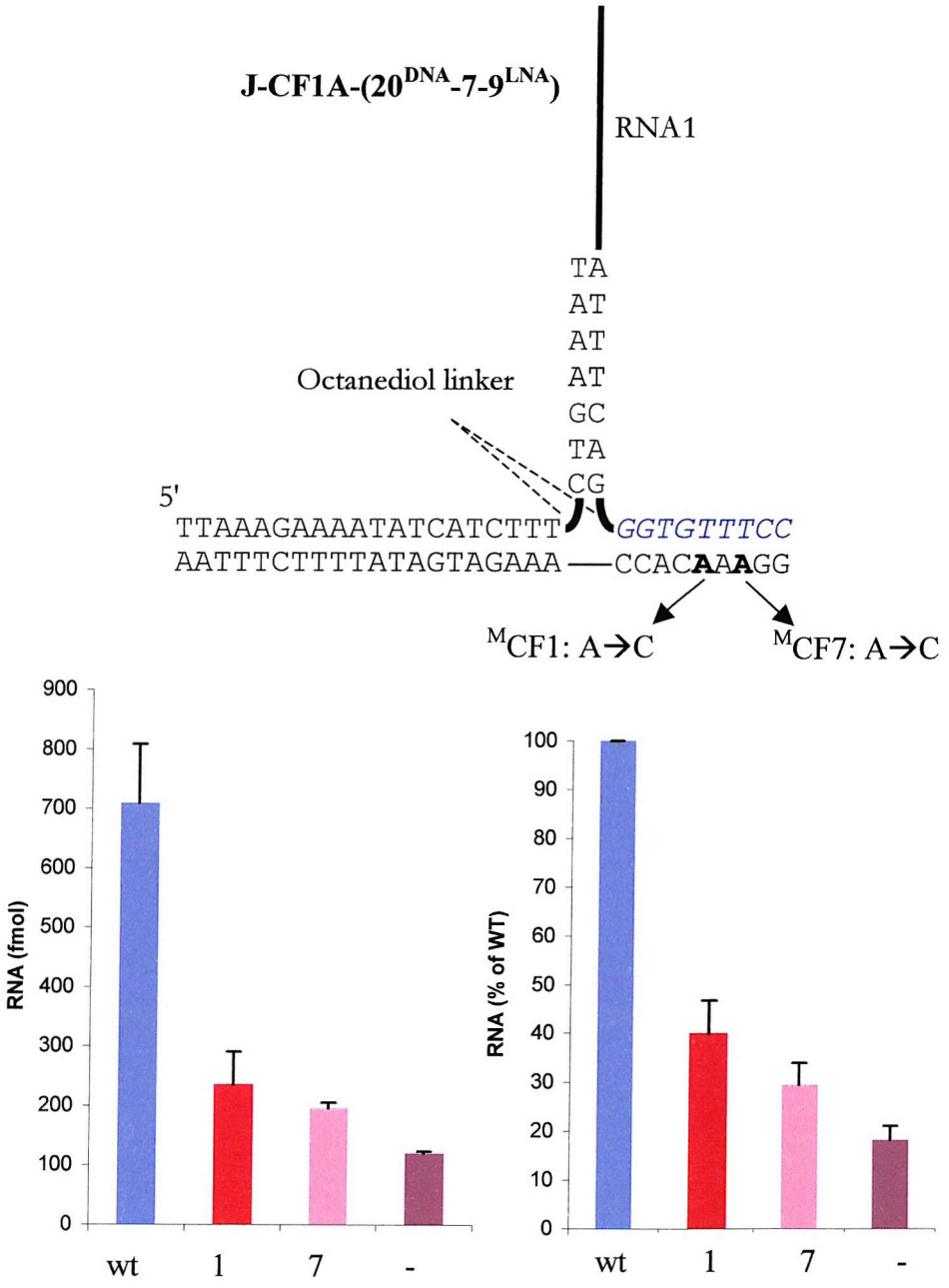


Fig. 3.12 Analysis of mutation sensitivity of J-CF1A-(20^{DNA}-7-9^{LNA}). 1 and 7 correspond to ^{M}CF1 and ^{M}CF7 in Fig. 2.1. The blue bases represent LNA. The experiment was repeated 5 times.

in each probe). The results, presented in Fig. 3.13, confirm the lower background. However, this figure reveals that the introduction of LNA did not improve the sensitivity for single base changes even when compared to a DNA 16 base template probe foot (Fig. 3.6). Interestingly, three mismatches (T→A and T→C at position 11 and A→G at position 15) showed a marked reduction in signal, close to background level. However, inspection of other graphs (Fig. 3.5, 3.6 and 3.8) reveal that these targets always generate large reductions in signals, even with junctions such as J-CF1A-(20^{DNA}-7-16^{DNA}), which showed a low sensitivity for mismatches at most other positions. This could therefore indicate that the oligonucleotides contained impurities, or that these mismatches had an unusually high propensity for junction destabilisation. The latter possibility appears less likely since the signals for all junctions in the presence of these targets dropped to background level and the fact that the other two mismatches at position 15 yielded much lower reductions in signal than the A→G mismatch.

3.4 Using SMART for the detection of the ΔF508 CFTR mutation in PCR samples

3.4.1 Effect of moving the branch point of J-CF1 on the detection of the ΔF508 CFTR mutation

Section 3.2.3 revealed that deletions located close to the branch point led to a smaller reduction in signal than those located further into the template probe/target arm. Since the disease-relevant deletion, ΔGAA (ΔF508, or ΔCF2 in Fig. 2.1), was located at a non-optimal location with respect to mutation sensitivity (i.e. close to the branch point) a new CFTR gene based junction was designed involving probes that contained only DNA (J-CF2B(20^{DNA}-8-14^{DNA})) (Fig. 3.14). A shorter target-complementary sequence was included in the template probe as this was previously shown to improve the mismatch sensitivity of J-CF1A. Other studies had shown that RNA2 (see chapter 2) coupled to an 8 bp overlap yielded substantially higher RNA yields and these modifications were also included in the design of the two junctions. Finally, the branch was moved such that now the ΔF508-deletion was located in the centre of the template probe/target arm.

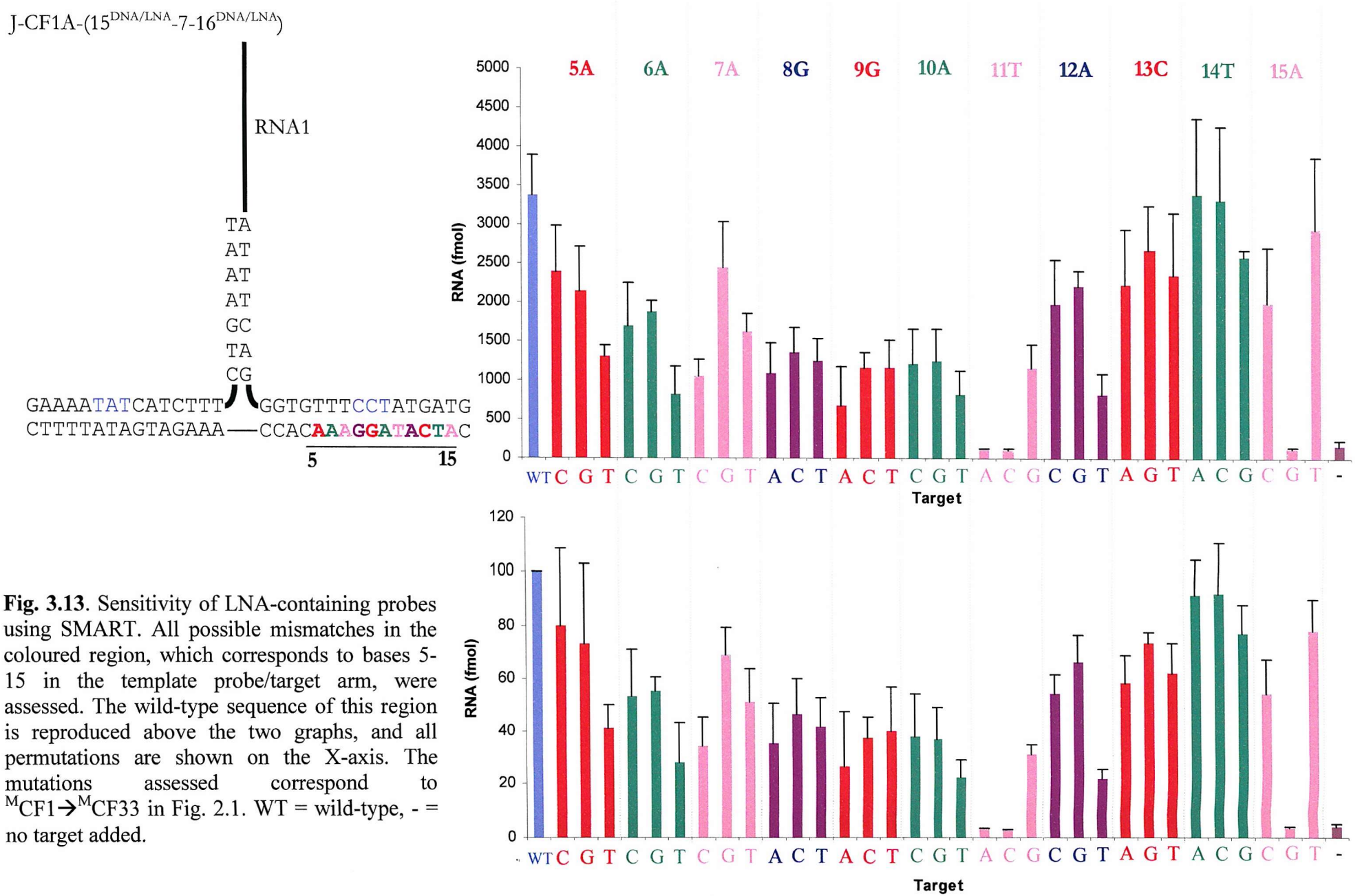


Fig. 3.13. Sensitivity of LNA-containing probes using SMART. All possible mismatches in the coloured region, which corresponds to bases 5-15 in the template probe/target arm, were assessed. The wild-type sequence of this region is reproduced above the two graphs, and all permutations are shown on the X-axis. The mutations assessed correspond to ^{M}CF1→^{M}CF33 in Fig. 2.1. WT = wild-type, - = no target added.

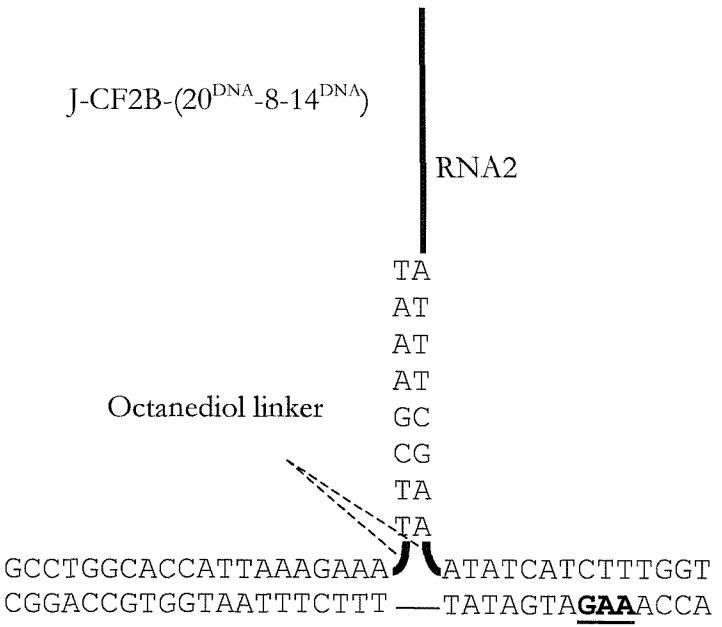


Fig. 3.14. New junction design based on the CFTR gene, using DNA probes for detection of the ΔF508 deletion. The bases printed in bold and underlined indicate the location of the ΔF508 deletion (=ΔCF2 in Fig. 2.1).

Fig. 3.15 shows the results of an experiment with this new TWJ using 25 fmol of extension probe, 5 fmol of template probe and 5 fmol of target (note that previous J-CF1A experiments were conducted with 10x more target present) using the optimised reaction conditions described previously. The sensitivity of the junction was tested by adding a synthetic target containing either the ΔF508 or the ΔI507 deletion (the latter is also a disease-predisposing deletion but occurs less frequently in the population and corresponds to ΔCF1, Fig. 2.1) and comparing the signal to that obtained in the presence of wild-type DNA. Fig. 3.15 reveals that clear discrimination was obtained between wild-type and deletion-containing targets, probably due to the combination of using a shorter template probe/target arm and placing the deletion close to the centre of this arm.

Fig. 3.16 shows that J-CF2B is also sensitive to single base changes, as judged from the reduction of signal compared to that obtained with wild-type DNA. This indicates that significant mutation sensitivity is not limited to a single junction (i.e. J-CF1A), but is a general property of SMART. In contrast to J-CF1A, PNA did not appear to enhance mismatch sensitivity as the level of discrimination for the two junctions was approximately the same, and the yield for J-CF2B-(20^{DNA}-8-14^{DNA}) was even higher than for the PNA-containing junction (compare the left upper and lower graphs in Fig. 3.16).

3.4.2 Discrimination between wild-type and Hetero- and Homozygous ΔF508-containing CFTR PCR fragments using J-CF2B-(20^{DNA}-7-14^{DNA}).

Progress so far has revealed that junction design is a critical issue in obtaining mutation sensitivity. This can be achieved either by using DNA-analogues such as PNA or LNA, or by using DNA-template probes with a reduced target complementary sequence. However, all the experiments described so far were carried out using synthetically prepared single stranded target sequences, which is clearly different from distinguishing a single base change within the background of, for instance, human genomic material. One step towards the detection of mutations in such natural genomic samples is to detect mutations present in (double stranded) PCR amplified sequences. Such experiments would indicate whether the double stranded character of DNA interferes with probe binding and/or mutation sensitivity.

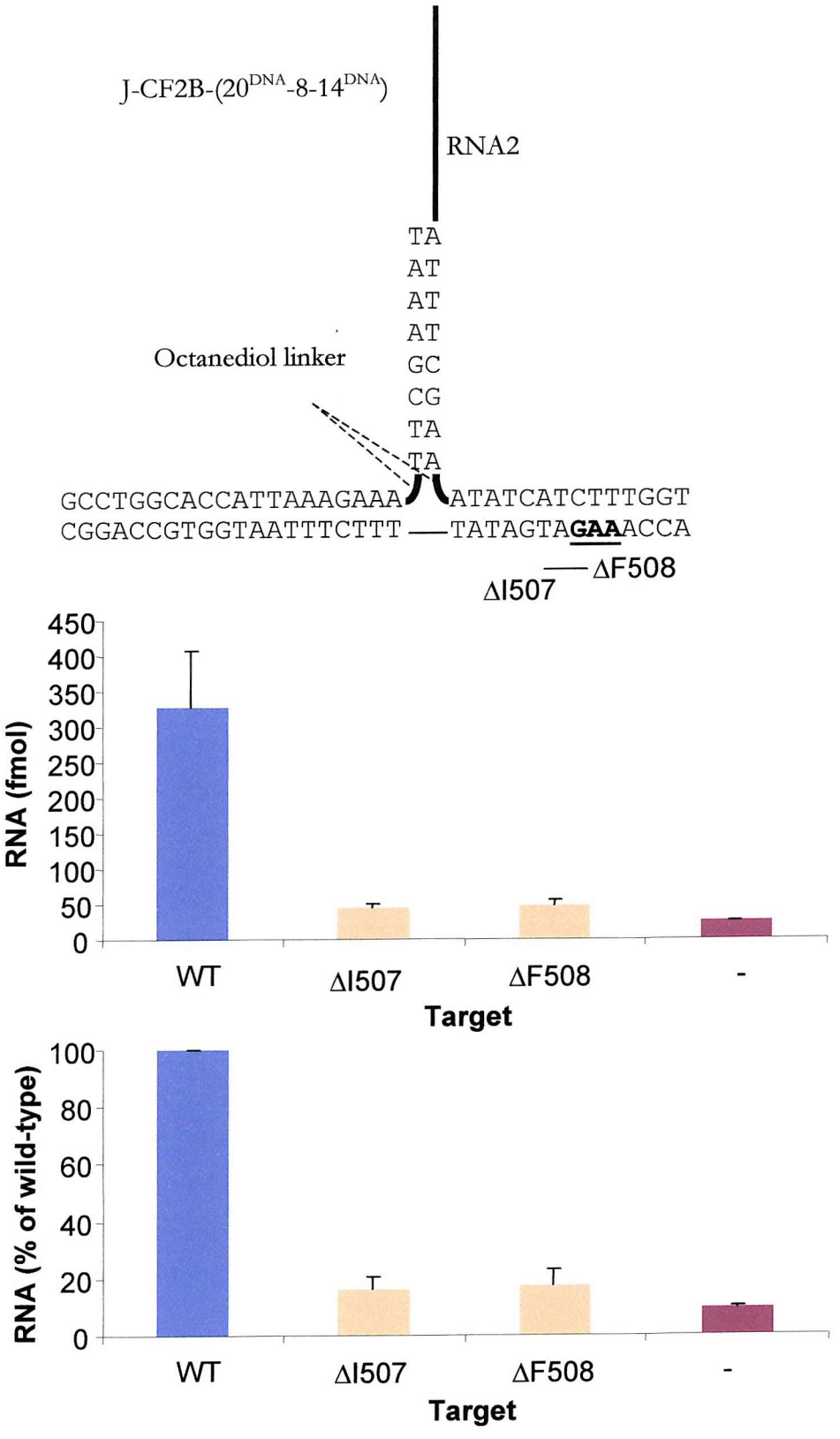


Fig. 3.15 Sensitivity of J-CF2B-(20^{DNA}-8-14^{DNA}) for CFTR gene deletions, using 5 fmol of target. ΔI507 = ΔCF1 and ΔF508 = ΔCF2 in Fig. 2.1.

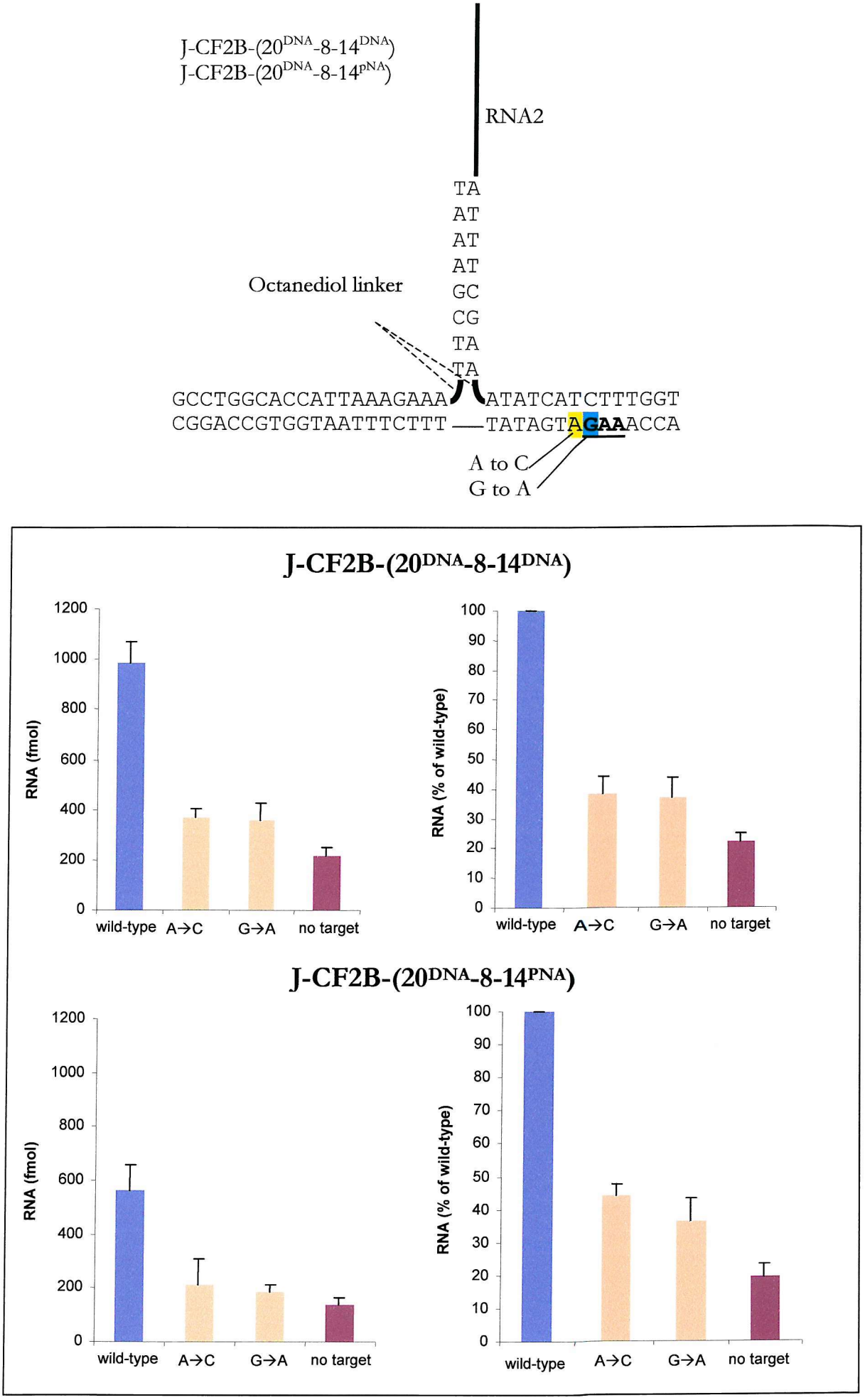


Fig. 3.16. Comparison between J-CF2B-(20^{DNA}-8-14^{DNA}) and J-CF2B-(20^{DNA}-8-14^{PNA}).

To this end, wild-type, ΔF508 and ΔI507-containing CFTR PCR products were obtained from Genzyme Inc., USA. Unfortunately, primer sequence data was unavailable and only the total 220 bp amplicon sequence was provided along with their concentrations. Control dot-blot experiments were carried out by Genzyme to confirm the ΔF508 and ΔI507 zygosity of the different PCR samples provided.

Experiments were carried out using the extension probe and template probe for J-CF2B (20^{DNA}-8-14^{DNA}) with 50 fmol of PCR target, 80 fmol of extension probe, 50 fmol of template probe and the same reaction conditions as described previously. Fig. 3.17 indicates that not only were the probes able to discriminate between wild-type and homozygous ΔF508 PCR targets, but a reproducible difference between heterozygous and homozygous material was observed as well. Statistical analysis performed on these data (shown in table 3.2) further corroborate these conclusions. The table shows that homozygous ΔF508 material shows no significant differences with exon 20 (non-specific) DNA (p=0.1677), yet significant difference with wild-type target (p=0.0004). This indicates that accurate discrimination is possible using double-stranded DNA of much longer length (~200 bp) than the synthetic target oligonucleotides used thus far. Even more, the statistical analysis shows that "quantitative" discrimination is feasible as heterozygous and homozygous material were clearly discriminated (p=0.0076), yet no significant difference was apparent between two different heterozygous targets (p=0.4833, *i.e.* they both yielded a similar reduction in signal compared to wild-type).

3.5 Using SMART for the detection of a SNP causing deep venous thrombosis: Factor V Leiden

This chapter showed that by using DNA or DNA/PNA chimerical probes, detection of single base pair changes can be achieved by SMART. However, the single base substitutions assessed so far were synthetic, i.e. not known to cause a human hereditary disease. Therefore, a novel target was chosen for which a well-characterised SNP was known to cause a disease. This target was the platelet Factor V gene, and its disease predisposing variant is usually called Factor V Leiden. A single base mutation (G to A in codon 506, nucleotide 1691), leading to a Arg → Glu substitution, occurs in possibly as many as 40% of all patients with venous

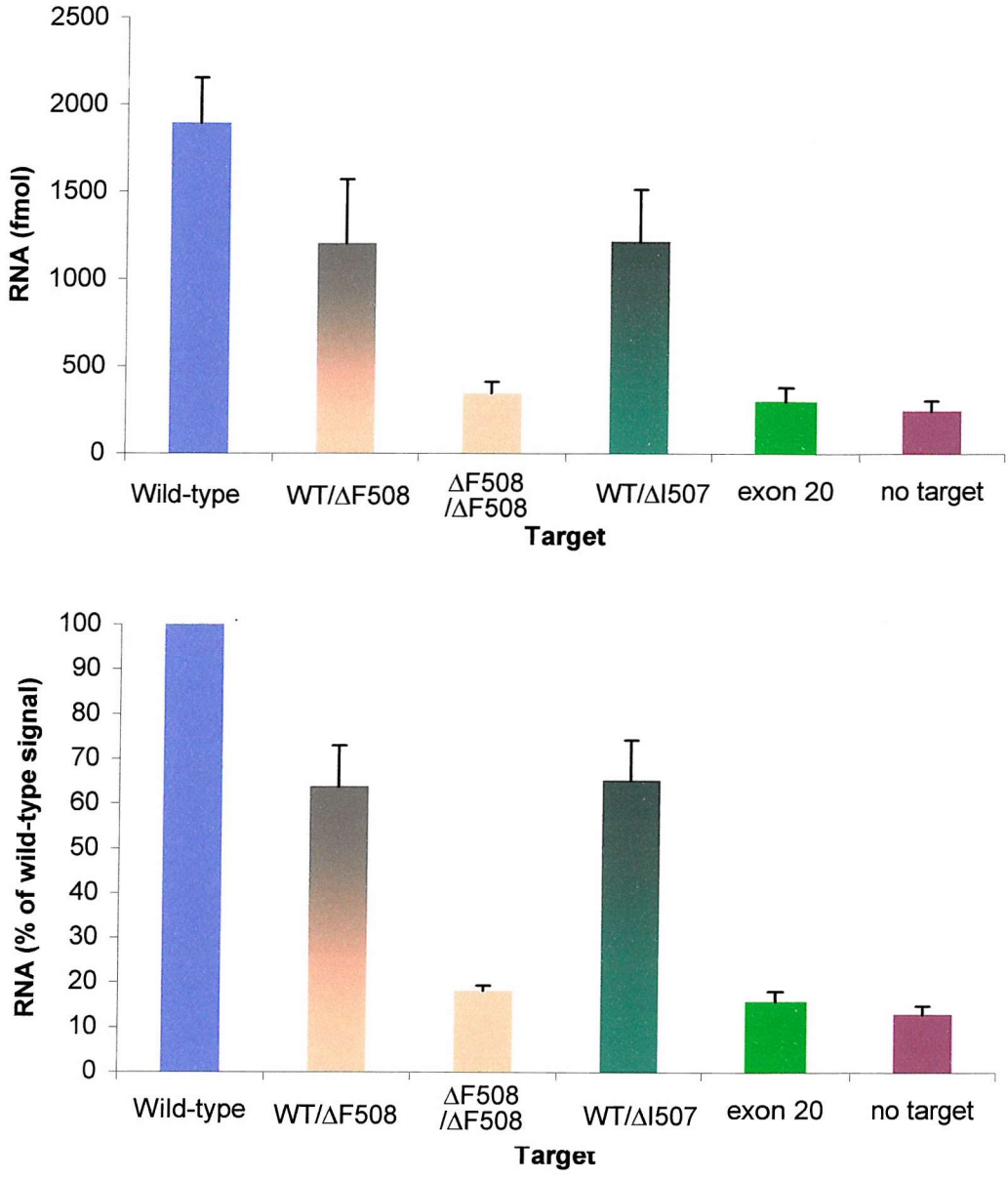


Fig. 3.17 Sensitivity of J-CF2B(20^{DNA}-8-14^{DNA}) for wild-type and disease-predisposing deletions in the CFTR gene, using PCR-amplified targets. Exon 20 stands for a PCR-amplified segment of the CFTR exon 20, which was used as a non-specific target control (the ΔF508 and ΔI507 deletions are located in exon 10). Genzyme did not provide the homozygous ΔI507 fragment. Experiments were repeated 5 times and the average values obtained are shown here together with the respective standard deviations. WT = wild-type.

Table 3.2 Statistical analysis of wild-type, homozygous and heterozygous ΔF508 and ΔI507 CFTR deletion detection with SMART. A one-tailed, paired t-test was applied to the data presented in Fig. 3.18, comparing the absolute (A) or relative (B) yields obtained in the presence of different targets. Data points with $p<0.05$ are printed bold.

a) Absolute yields

	wild-type	WT/ΔF508	ΔF508/ΔF508	WT/ΔI507	exon 20	no target
wild-type		**0.0154**	**0.0004**	**0.0260**	**0.0007**	**0.0006**
WT/ΔF508	**0.0154**		**0.0076**	0.4833	**0.0116**	**0.0073**
ΔF508/ΔF508	**0.0004**	**0.0076**		**0.0040**	0.1677	**0.0231**
WT/ΔI507	**0.0260**	0.4833	**0.0040**		**0.0022**	**0.0019**
exon 20	**0.0007**	**0.0116**	0.1677	**0.0022**		0.0693
no target	**0.0006**	**0.0073**	**0.0231**	**0.0019**	0.0693	

b) Relative yields

	wild-type	WT/ΔF508	ΔF508/ΔF508	WT/ΔI507	exon 20	no target
wild-type		**0.0143**	**0.0000**	**0.0153**	**0.0000**	**0.0000**
WT/ΔF508	**0.0143**		**0.0073**	0.4583	**0.0101**	**0.0059**
ΔF508/ΔF508	**0.0000**	**0.0073**		**0.0053**	0.1660	**0.0109**
WT/ΔI507	**0.0153**	0.4583	**0.0053**		**0.0031**	**0.0028**
exon 20	**0.0000**	**0.0101**	0.1660	**0.0031**		0.0707
no target	**0.0000**	**0.0059**	**0.0109**	**0.0028**	0.0707	

thrombosis although there is evidence for the involvement of additional (unidentified) genes. Factor V is a cofactor for a protease named activated protein C (APC), and the arginine present in the wild-type protein functions as cleavage substrate for protein C, which inactivates Factor V. The Arg → Glu deficiency inhibits Factor V inactivation and leads to accumulation of thrombin, resulting in thrombosis (Bertina *et al.*, 1994; Gandrille *et al.*, 1995; Svensson & Dahlback, 1994).

3.5.1 Assessment of effect of overlap length and type of nucleic acids used in junctions involving part of the Factor V gene.

The previous chapter indicated that a short target complementary sequence in the template probe yielded the highest sensitivity for mutations in the CFTR gene. To verify this observation, analogous junctions were designed for the Factor V target, one containing a 14 base target complementary region in the template probe and one with a template probe containing a 14 base PNA foot (J-LV1B-(16^{DNA}-8 or 7 or 6-14^{DNA}) and J-LV1B-(16^{PNA}-8 or 7 or 6-14^{PNA}), Fig. 3.18). Further, as for the CFTR junction J-CF2B-(20^{DNA}-7-13^{DNA}), octanediol linkers were incorporated in both probes at the branch point as well as an optimised transcription sequence (RNA3, refer to materials & methods). All experiments described in this section used 50 fmol extension probe, 25 fmol template probe and 5 fmol target. The length of the overlap region contained 8, 7 or 6 bp (shortening the extension probe 3' end). This was done because an 8 bp overlap did not show a detectable difference between wild-type and mismatched targets as shown in Fig. 3.19a, b. This figure also shows that when the overlap region was reduced to 7 or 6 bp a clear enhancement in mismatch sensitivity (80% signal reduction when compared to wild-type) occurred, while not substantially decreasing the RNA yield. This observation underlines the complex nature of the TWJ as modifications in at two arms can have a significant effect on mutation sensitivity, even if the modified arm does not contain the mutation. These results therefore suggest that the overlap-containing arm was thermodynamically linked to the template probe/target arm.

Fig. 3.19c and 3.19d further show that when PNA was incorporated in the template probe (but not in the extension probe), an 8 bp overlap yields a substantially higher sensitivity for the Factor V Leiden target compared to the corresponding DNA

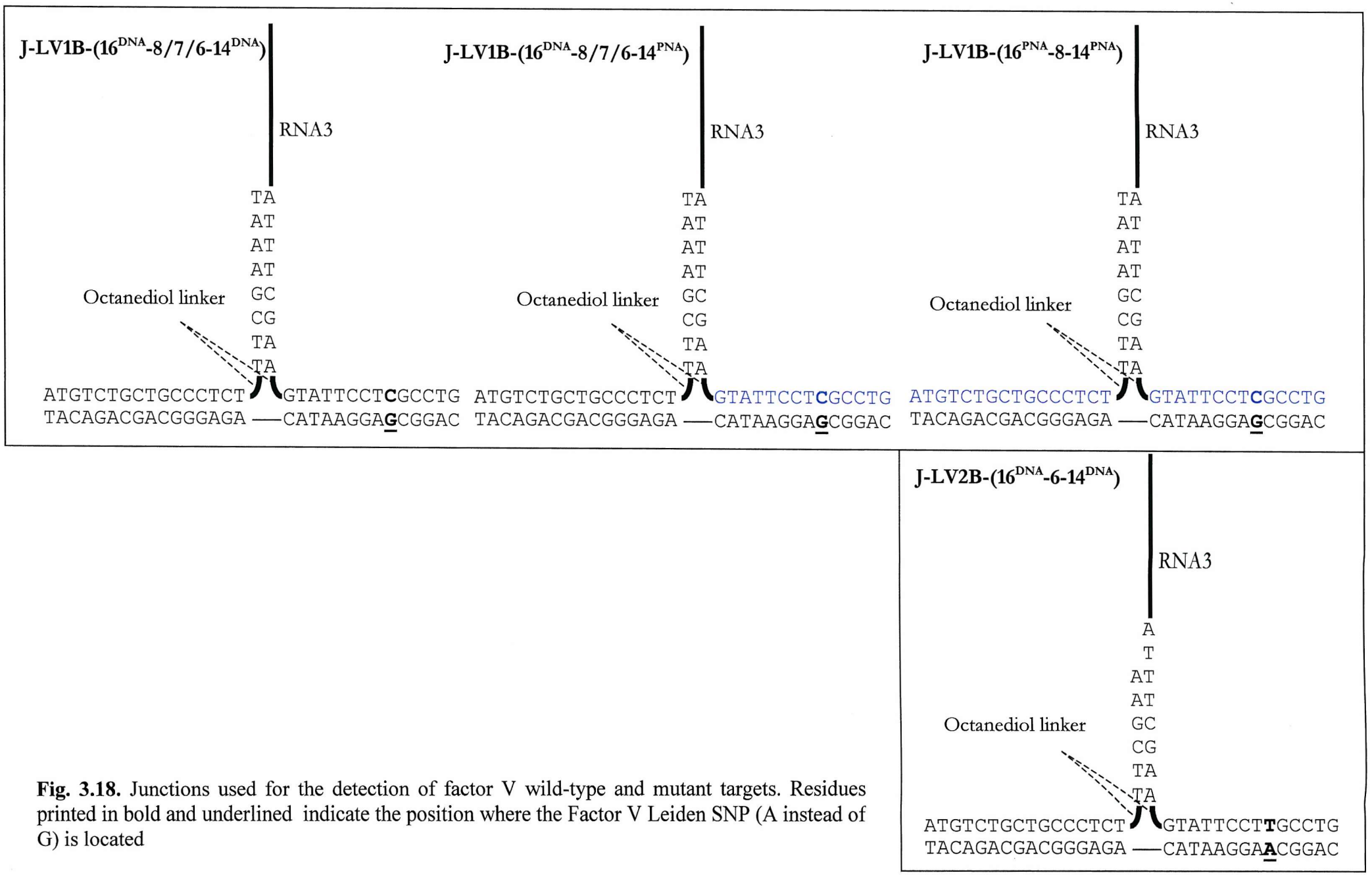


Fig. 3.18. Junctions used for the detection of factor V wild-type and mutant targets. Residues printed in bold and underlined indicate the position where the Factor V Leiden SNP (A instead of G) is located

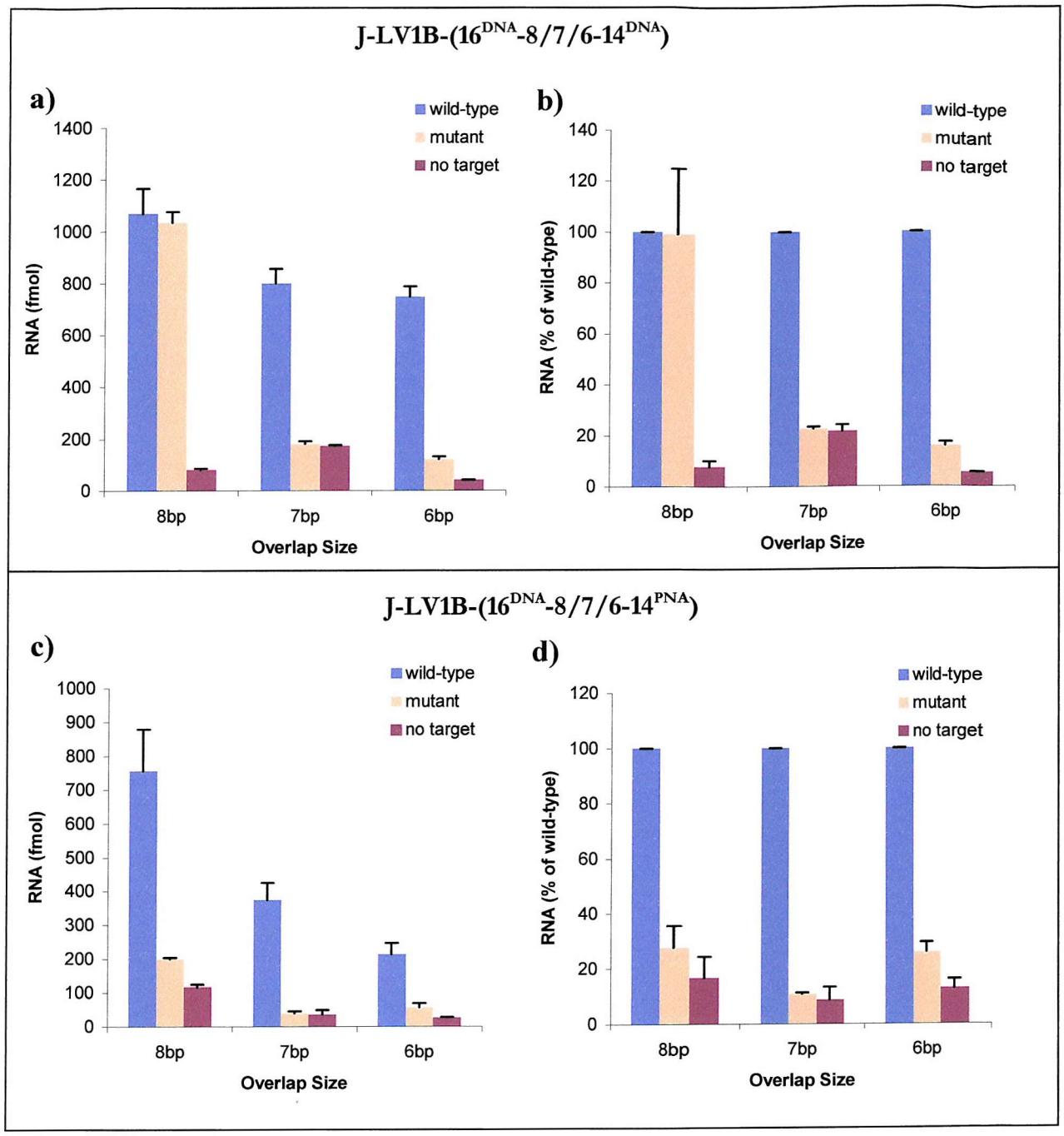


Fig. 3.19 Assessment of the effect of shortening the overlap and using PNA or DNA in the template probe on mismatch sensitivity. The mutation used was the G→A factor V Leiden SNP. The experiment was repeated twice.

template probe (~70% versus ~0%). The reason for the unusual high variability of the signal detected for J-LV1C-(16^{DNA}-8-14^{PNA}) could be due to experimental errors as the reaction was repeated only once. Fig. 3.19 also shows that reducing the overlap with a PNA-containing template probe resulted in a more pronounced reduction in overall signal than was the case for the DNA-only junction. Since the DNA and PNA versions of J-LV1C possessed the same sequence and linkers, these results provide a more accurate comparison between the two types of nucleic acids than those obtained with the CFTR junctions and show that inclusion of PNA benefited this particular junction design. These results support those in chapter three, where a positive effect of PNA was also observed for J-CF1. However, Fig. 3.20 shows that when PNA was incorporated into the extension probe, a large decrease in RNA yield was observed regardless of whether DNA or PNA was present in the template probe (a reduction from 1000 fmol to 150 fmol, Fig. 3.20 and Fig. 3.19). The reason for this is unclear, although the HPLC trace of the PNA/DNA extension probe indicated that the quality of the PNA oligomer was good (Oswel DNA service, UK, personal communication). As noted before, the effect of PNA on TWJ conformation has not been studied to date but is likely to have a substantial effect due to the reduction in negative charge. Whether a structural effect was the cause for the reduction in signal observed for J-LV1C-(16^{PNA}-8-16^{PNA}) remains to be determined.

3.5.2 Specific detection of Factor V Leiden

So far, junctions were designed to detect wild-type sequences whereas for proper disease diagnosis it is desirable to detect the presence of a specific disease predisposing allele. The Factor V Leiden target served as a convenient target as it contained only a single base change, as opposed to the ΔF508 CFTR target, allowing for better assessment of the robustness of SMART in detecting such genetic deficiencies. The template probe used in J-LV1C-(16^{DNA}-7-14^{DNA}) was redesigned such that it now matched the Factor V Leiden target rather than the wild-type gene, subjected to SMART and compared with J-LV1B-(16^{DNA}-7-14^{DNA}). Fig. 3.21 shows that J-LV2B-(16^{DNA}-7-14^{DNA}) accurately detected the SNP-containing target (the same conditions were applied as described in section 4.2.1). Although substantial variation in the absolute amount of RNA produced was apparent for J-LV2B-(16^{DNA}-

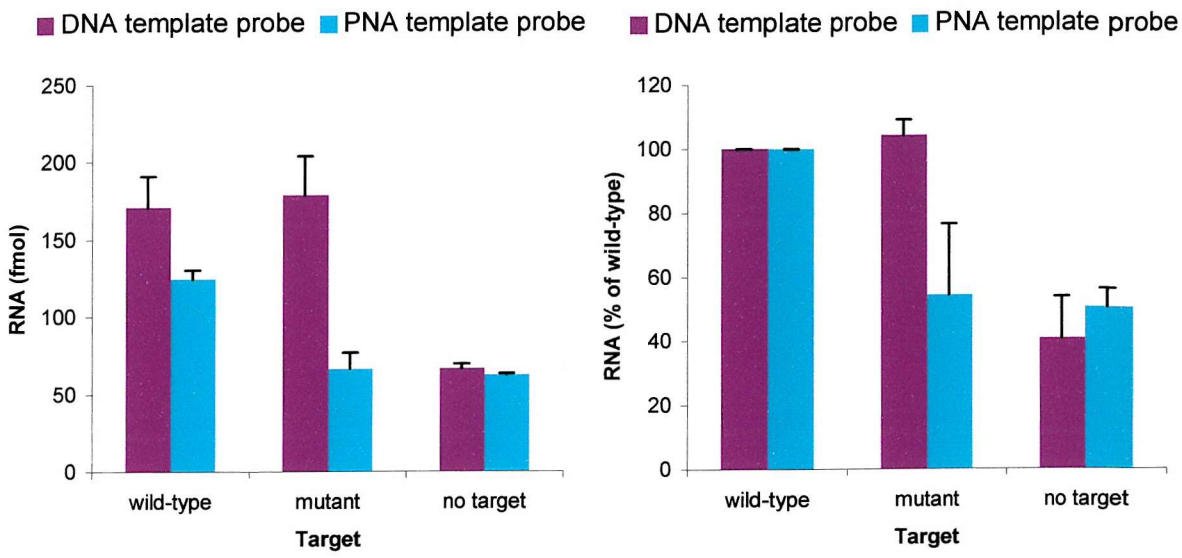


Fig. 3.20 Results for J-LV1B using a PNA-containing extension probe coupled to either a DNA or PNA template probe. The mutant target is factor V Leiden.

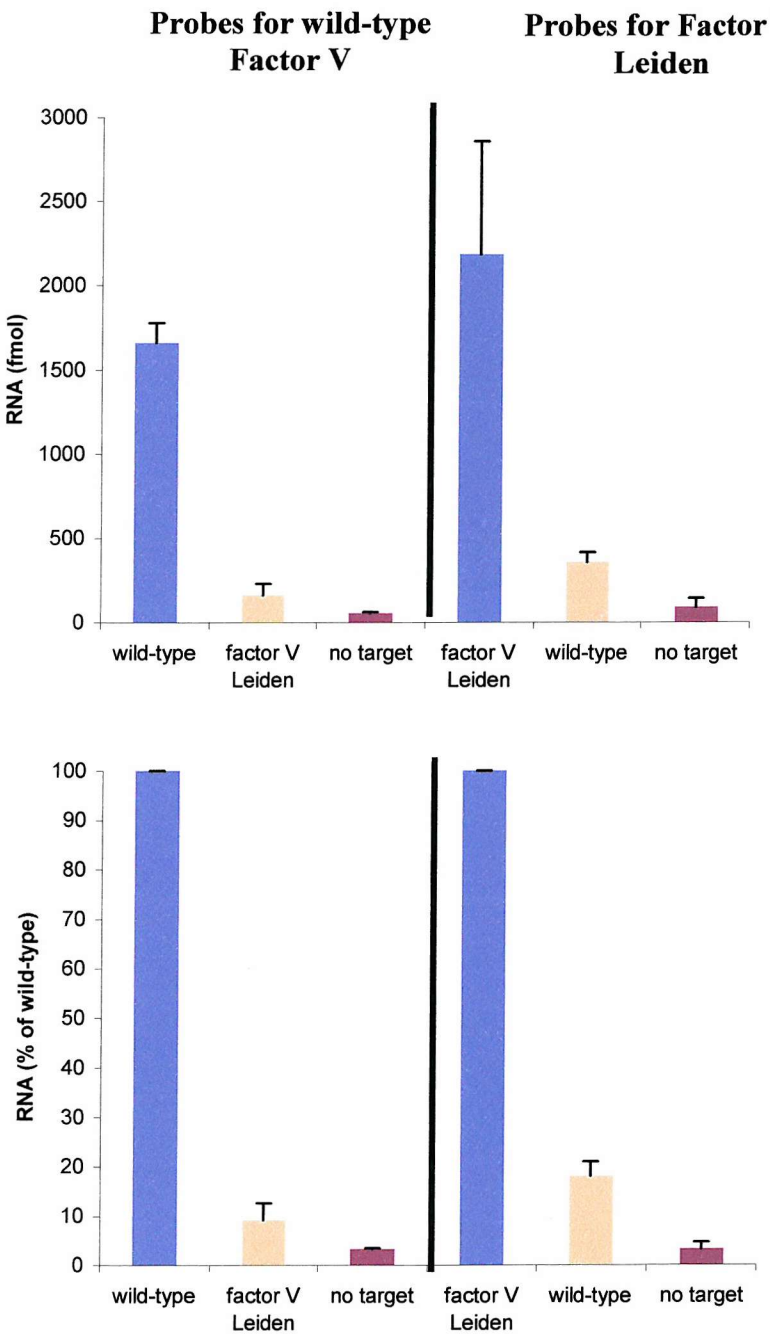


Fig. 3.21 Factor V wild-type and SNP detection, using J-LV1B-(16^{DNA}-7-14^{DNA}) and J-LV2B-(16^{DNA}-7-14^{DNA}) respectively.

7-14^{DNA}), the relative difference between the SNP-containing and wild-type target remained the same. Similar results were already obtained for J-CF1A junctions, supporting the conclusion that discrimination is not affected by small variations in the total RNA yield.

3.6 Discussion

This chapter has indicated that sensitive discrimination between wild-type target and single-base variants thereof, can be achieved with SMART using either synthetic or PCR amplified target DNA. Evidence was presented, suggesting that PNA is more sensitive for mutations than DNA. Two very similar junctions, one containing PNA and the other DNA as the target complementary sequences in the two probes showed substantial differences in discrimination (section 3.2.3). However, it was also found that reduction of the size of the template probe-foot region markedly improved the sensitivity of DNA probes (J-CF1A-(20^{DNA}-7-13^{DNA}) and J-LV1B-(16^{DNA}-7-14^{DNA})). The most likely explanation for the latter is that the melting temperature was reduced by the reduction in foot length, which would improve its mutation sensitivity.

With respect to LNA, the very high background signals observed for J-CF1A-(9^{LNA}-7-9^{LNA}) were surprising. The reasons for the high background remain elusive at present, but are surprising because of the previously reported high sequence-specificity of LNA (Koshkin *et al.*, 1998, Nielsen *et al.*, 2000). It should be noted that the higher specificity of LNA is backed only by a handful of studies carried at Exiqon A/S, and further studies are therefore required to verify the universality of the reported specificities. The high background observed for J-CF1A-(9^{LNA}-7-9^{LNA}) is particularly disturbing since it suggests that interactions between the target complementary sequences in the two probes were actually promoted by LNA. It seems unlikely that for instance trace contaminants that were not removed during LNA synthesis caused the high background since J-CF1A-(15$^{DNA/LNA}$-7-16$^{DNA/LNA}$) showed very low background signals. The oligos used for this junction were synthesised by the same company, and in addition, the HPLC traces of all probes indicated that no contamination was present. The high background problem was eliminated by replacing the LNA-extension probe with a DNA extension probe. However, substantial reductions in yield and no improvement in sensitivity was detected. Other probe constructs containing a reduced number of LNA residues also

did not show a detectable difference in mutation sensitivity when compared to DNA probes. Together these results cast serious doubts as to whether LNA is in fact more sensitive for mutations than DNA.

An additional problem that remains is predicting the effect of incorporating PNA or LNA in the TWJ on its conformation and stability. The absence of negative charge and/or the hydrophobic character of PNA probably affect TWJ folding and stability as might the constrained nature of the LNA backbone. J-LV1B-(16^{PNA}-8-14^{PNA}) for instance, showed reduced RNA yields compared to J-LV1B-(16^{DNA}-8-14^{PNA}) and J-LV1B-(16^{DNA}-8-14^{DNA}). This shows that PNA may have a negative effect on RNA yields. Interestingly, the yield reduction was observed only when PNA was incorporated into the extension probe, not when PNA was present only in the template probe. However, whether this was due to differences in sequence length, a structural effect or some other factor remains to be verified.

Together, these results show that, although PNA may improve the mutation sensitivity of TWJs in some cases, using shorter DNA probes offers better reproducibility and prediction of its influence on TWJs since DNA is much better characterised than PNA. Using a shorter (13 or 14 bases instead of 16 bases) template probe foot improved the sensitivity greatly without compromising the RNA yield for junctions based on the CFTR and Factor V targets. One could argue that reducing the length of this target complementary sequence could potentially reduce the specificity of the template probe since the shorter this sequence becomes the more likely it becomes that the same sequence is present elsewhere in the human genome. This is supported by the observation that a Blast search (http://www.ncbi.nlm.nih.gov/blast/, advanced blast using the Blastn database with "expected" set to 1000 and using the PALM30 -9-1-0.89 search matrix, carried out on 03-08-2000) using the 13 base J-CF1A(20^{DNA}-7-13^{DNA}) template probe foot sequence, yielded hits for nearly all human chromosomes, often with multiple hits per chromosome. However, the length of the template probe is not the only factor that determines the specificity of SMART as the total target-spanning region in the junction determines this. Performing a blast search on the entire target spanning sequences of the two probes in J-CF1A-(20^{DNA}-7-13^{DNA}) yielded only hits for the CFTR gene. Therefore, increasing the length of the extension probe/target arm could well compensate for any loss in specificity caused by reducing the template probe/target arm. One could still argue that a small template probe could be titrated

away by other complementary sites present in the genome, but addition of more template probe to compensate for its loss elsewhere may solve this. Whether increasing the length of the extension probe or adding more template probe affect the sensitivity for mutations remains to be tested.

Finally, it is interesting to note that the standard deviations calculated for the relative amounts of RNA in most cases were substantially reduced when compared to the standard deviations calculated for the absolute values. Regarding the significant correlations between yields for wild-type and mismatched targets for repeated experiments, it appears that although the absolute RNA yields can vary substantially, this variation does not affect the relative yield differences.

3.7 Concluding remarks

The results shown here illustrate that SMART can be used for the detection of disease predisposing SNPs as accurate detection of both wild-type and disease predisposing mutations was achieved using synthetic and PCR-amplified material. Junction design is a critical factor in achieving the required sensitivity to allow discrimination between targets that differ by one or more bases. This chapter showed that the length of the template probe foot and the overlap region are important factors contributing to mutation sensitivity.

The next major goal is to increase the sensitivity of SMART to allow the detection of polymorphisms in patient material (*e.g.* blood samples). Experiments at Cytocell Limited have already indicated that the scheme presented in Fig. 1.3 is not sufficient to detect sequences present at levels below 1 fmol and that further amplification (Fig. 1.4) is required. Future work will therefore focus on coupling the optimised junction designs established here to generate further exponential amplification of the RNA produced.

4 Critical analysis of the long-short arm assay used for the elucidation of junction conformation.

4.1 Introduction

Chapter one showed that TWJ structure is affected by sequences which flank the branch point as well as the presence of cations. In the SMART assay, sufficient cations (NaCl, $MgCl_2$ and spermidine) are present to allow co-axial stacking of two helixes to occur so that the conformation will be dictated by the sequences at the junction. As postulated in chapter one, the conformation of the TWJs involved in SMART may affect the RNA yields. To investigate this we assessed the structures of a number of junctions studied previously, including those based on the CFTR and Factor V genes. Comparison between the conformations observed for these junctions and their yields (noting that a substantial difference in yield for J-CF1A and J-LV1A for instance was observed in chapters 3 and 4) could provide insight into the influence of TWJ conformation on RNA yield.

4.2 Structural analysis of literature-derived and J-CF1A based junctions

4.2.1 Design of junctions

The initial approach comprised the investigation of structural aspects of three-way junctions. The long-short arm assay, first applied by Cooper & Hagerman (1987) and later adapted by Welch *et al.* (1993), served as a convenient and rapid method for comparing the structures of several different three-way junctions. Additionally, several junctions were probed with $KMnO_4$ and DEPC, which are specific for unpaired thymine and purine residues respectively to assess the stacking arrangement of the junction. These chemical probes generate alkali labile adducts that can be cleaved by piperidine (Bailly, 1997). The cleavage products are then analysed on denaturing polyacrylamide gels to assess the sites of cleavage.

The basic design of the junctions is outlined in Fig. 4.1. Since the idea was to establish a link between TWJ structure and DNA/RNA polymerase activity, the

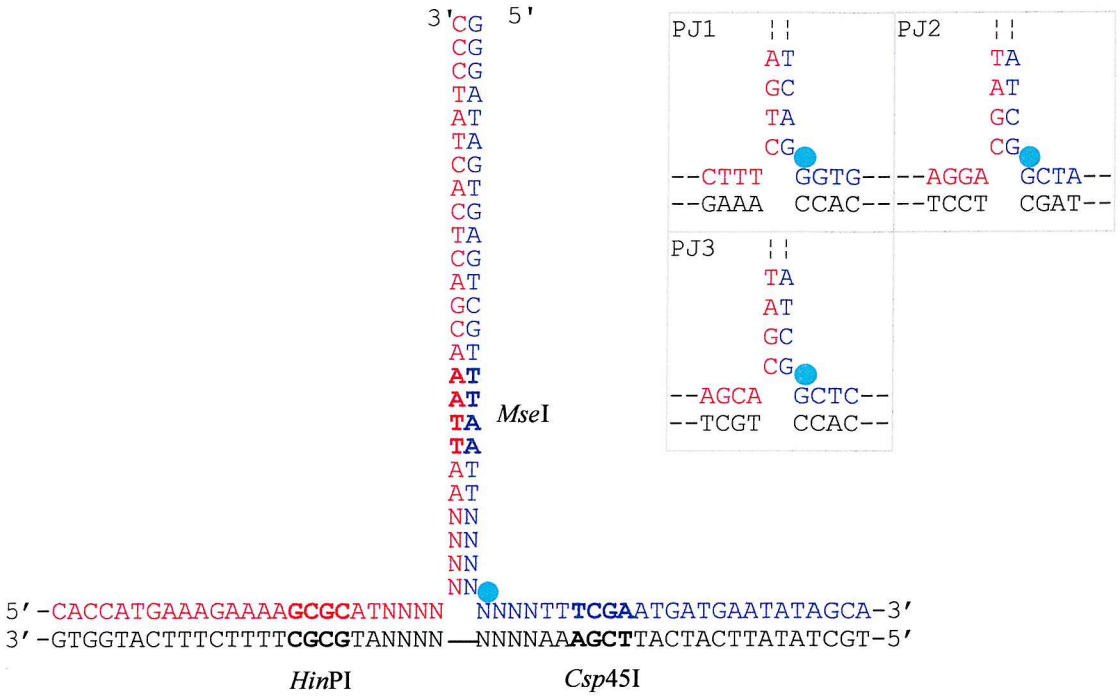


Fig 4.1 Designs of PJ1, PJ2 and PJ3. Red indicates the extension primer sequence, blue the template primer sequence and black the target sequence. 'N' indicates the area where PJ1-3 differ in sequence as outlined in the upper right hand boxes. Restriction endonuclease recognition sequences are printed in bold. The green circle indicates the position of the linker (unpaired bases).

design incorporated the sequences used in SMART assays for the detection of the CFTR gene (J-CF1, chapter 3). In the figure the extension primer sequence is indicated in red, the template primer sequence in blue and the target in black. Each arm contained 25 bp and a unique R-ENase recognition site starting 7 bp away from the branch point. Note that these junctions are necessarily different from those described in chapters 3 & 4 since all the arms must be of equal length and restriction endonuclease cleavage sites needed to be introduced.

Adapting the approach followed by Welch *et al.* (1995), sequence differences were only introduced in the first four base pairs of each arm relative to the branch point, which in the figure are indicated in the upper right hand boxes. PJ1 was based on J-CF1 (chapter 3). PJ2 incorporated the same sequence at the branch point as the conformer II junction studied first by Rosen & Patel (1993b) by NMR and later by Welch *et al.* (1995) by electrophoresis. Finally, PJ3 possessed the same DNA sequence as a junction studied by Welch *et al.* (1995) called J1V9. J1V9 showed equal mobilities for the three restriction digests when analysed on polyacrylamide gel applying a similar protocol as described in sections 2.1-2.3, which implied that the junction folded into an unstacked pyramidal-shape, or that the junction existed in two different conformations in rapid equilibrium with each other.

The three junction sequences were chosen for two reasons. First, for comparative analysis with the CF junction and second, to confirm we could reproduce the results of two earlier, independent, studies. It should be noted that the sequences beyond the first four base pairs around the branch point in each arm are different from those used by Rosen & Patel (1993b) and Welch *et al.* (1995).

4.2.2 Long-short arm assay results for PJ1C$_2$a, PJ2C$_2$a and PJ3C$_2$a

When PJ2C$_2$a and PJ3C$_2$a were subjected to the long-short arm assay as described in 2.1-2.3, the electrophoresis patterns shown in Fig. 4.2 were observed. For PJ2C$_2$a the *Mse*I fragment showed the highest mobility, indicating that this was probably the linear or coaxially stacked fragment. The *Csp*45I fragment displayed the highest mobility, which implies that this fragment contained the lowest degree of bending, while the *Mse*I fragment has the lowest mobility. Thus, the structure corresponded to a conformer II junction. In contrast, for PJ3C$_2$a all three fragments have very similar mobility suggesting that it adopted an unstacked pyramidal shape.

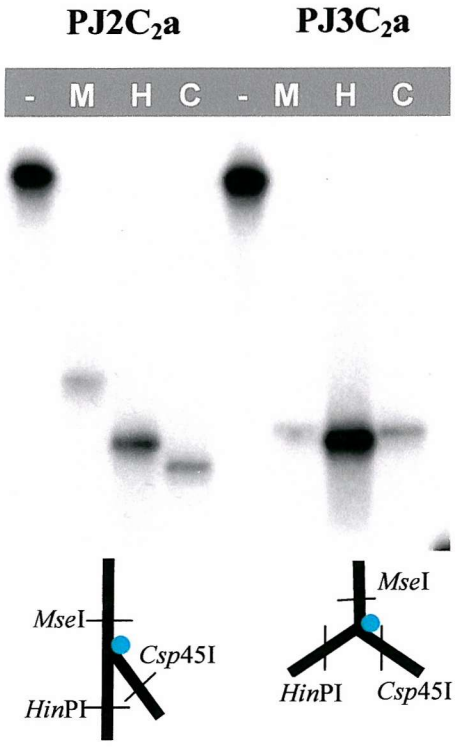


Fig. 4.2 Gel mobility patterns for PJ2C$_2$a, PJ3C$_2$a and extrapolated structures.

- = uncleaved junction, M = *Mse*I, H = *Hin*PI, C = *Csp*45I. Green circle: unpaired bases.

Alternatively, the mobility pattern could reflect a rapid equilibrium between conformer I and II (Welch *et al.*, 1995).

The results obtained for PJ2C$_2$a and PJ3C$_2$a correspond well to earlier studies. They further confirm the notion that the identities of the base pairs around the branch point are the important determinants for the resulting TWJ structure since the remainder of the sequence was identical for both junctions.

4.2.3 Effect of increasing the bulge size in PJ1

As the junctions used in SMART usually contain linkers at the branch point, we attempted to establish the effect of increasing the number of unpaired bases in PJ1 to assess the influence of the size of the linker on TWJ conformation.

Fig. 4.3 shows the R-ENase digestion patterns when the cytosine bulge in PJ1 was increased from 0, 1, 2, 3 to 10 unpaired cytosine residues. Both the isotopic and non-isotopic long-short arm assays, described in section 2.4 were used. The gels in Fig. 4.3b show that the mini-gel format provided a useful alternative for the assay as the observed mobility patterns corresponded well to the ones where radiolabelled probes were used (Fig. 4.3a). It is interesting to note that PJ1, lacking a bulge at the branch point, still showed differences in the mobilities of the three fragments, suggesting that it folds as a conformer I junction. This is somewhat surprising as junctions lacking unpaired bases usually adapt an unstacked conformation. More surprising is the observation that when one unpaired cytosine residue was inserted at the branch point, the mobility of the *Csp*45I fragment changed from slowest (PJ1) to fastest (PJ1C$_1$a). Thus, by inserting a single cytosine the conformation appeared to change to a conformer II-type of TWJ. However, judging from the ethidium bromide-stained gel, mobility differences between the different fragments of PJ1 and between those of PJ1C$_1$a were relatively small, especially when compared to other junctions discussed below and PJ2C$_2$a (Fig. 4.2). Therefore, it is possible that stacking did not occur in PJ1 or any of the bulge-containing PJ1-derivatives but rather, that they assumed a structure lacking coaxial stacking (Fig. 4.3c).

When the number of cytosines was further increased to 2 and 3 the conformation altered again. For PJ1C$_2$a the differences in mobility between the three fragments became even smaller, suggesting the junction now folded clearly into a Y-shape much like PJ3C$_2$a (Fig. 4.2). The PJ1C$_2$a *Mse*I and *HinP*I fragments showed the

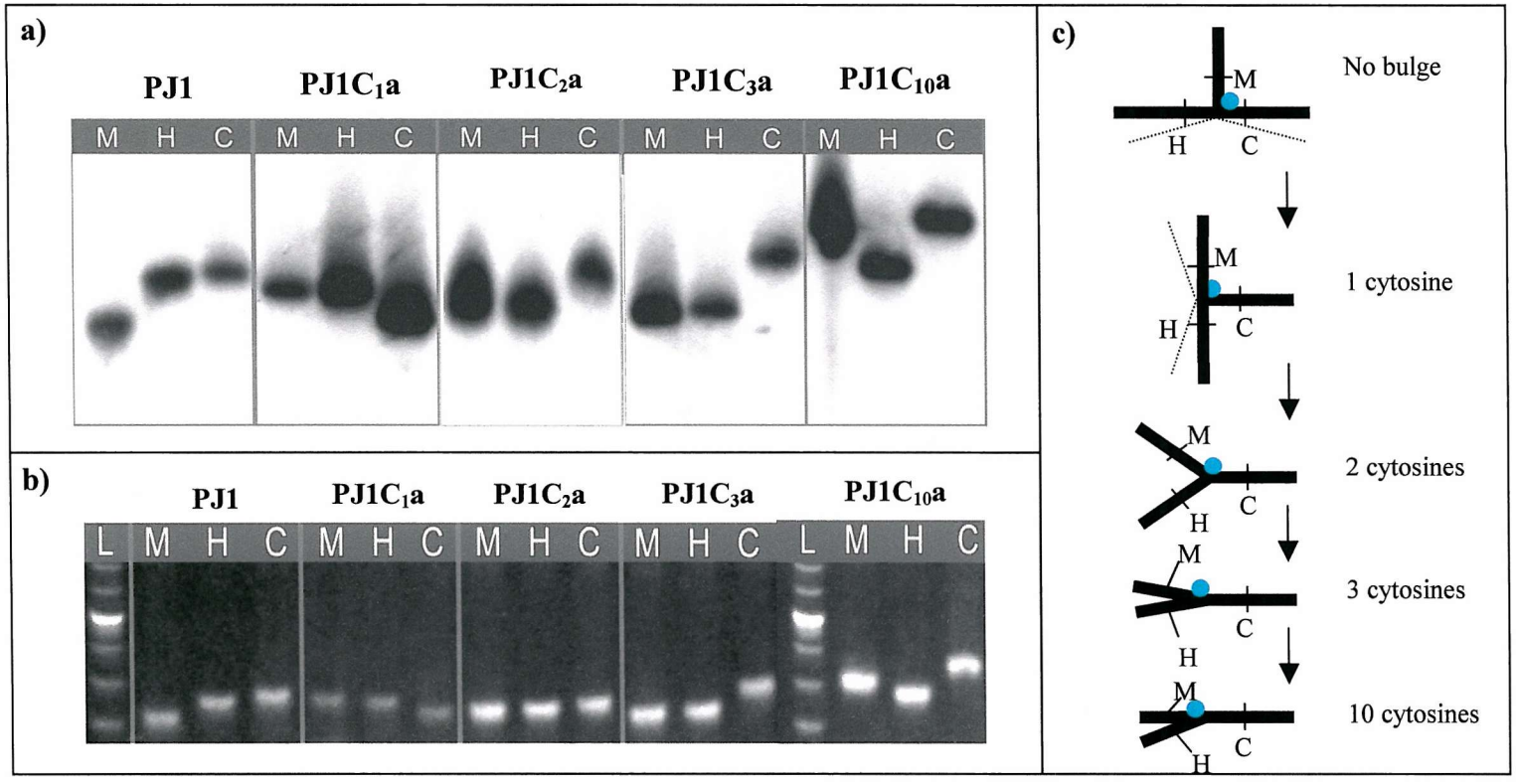


Fig. 4.3. Effect of increasing the number of unpaired bases at the branch on PJ1 and extrapolated structures. The two gels (a and b) show the results of two repeat experiments, one with radiolabelled probes (a) and one with non-radiolabelled probes (b, PJ1C$_{10}$a was run on a separate gel). Restriction endonuclease names are abbreviated to their first letter; M =*Mse*I, H = *Hin*PI, C = *Csp*45I. L indicates a 10bp DNA ladder (Gibco-Life Technologies, UK) with the 100 bp band appearing brighter than the others. For *Mse*I digestions junctions with labelled target were used, and for the *Hin*PI and *Csp*45I digestions junctions with labelled template probe. The structures in c) shows schematically the effect of increasing the number of unpaired cystosines. The structures are drawn as in Fig. 4.2 with the green circle indicating the position of the cytosine bulge. Some uncertainty in the interpretation of the structures in PJ1 and PJ1C$_1$a exists, since although the slow-to-fast fragment orders correspond to a conformer I and II respectively, mobility differences are small. The latter could mean that these junctions resided in an unstacked pyramidal shape (broken lines) or a dynamic conformational equilibrium where one coaxially stacked configuration dominated only marginally (solid lines). The patterns for PJ1 with 2 or more unpaired cytosines indicate an unstacked conformation.

same mobility, while the *Csp45*I ran slightly slower. The retardation of the *Csp45*I band relative to the other two bands increased when a third cytosine residue was added (PJ1C_3a), while the other two fragments migrated more rapidly. The mobility differences between the *Mse*I and *HinP*I bands remained however, negligible. When the size of the bulge was increased to 10 cytosine residues a difference in mobility between the *Mse*I and *HinP*I became apparent. In this case the slow-to-fast order of fragments became *Csp45*I – *Mse*I – *HinP*I. This order corresponds to an extended structure with the extension probe/target arm positioned closely to the extension probe/template probe arm. Thus, as outlined in Fig. 4.3c the effect of increasing the bulge size in PJ1 is a decreasing angle between the extension probe/target and extension probe/template probe arms. Similar results were obtained in previous studies by Welch *et al.* (1993) and Zhong *et al.* (1994), who also observed that increasing the size of the bulge decreased the angle between two arms in the TWJs.

Fig. 4.4 shows the results of DEPC and $KMnO_4$ probing of PJ1 (gels 1 & 2) and PJ1 containing varying bulge sizes (gel 3). Gels 1 shows the effect of increasing amounts of $KMnO_4$ on PJ1 lacking a bulge using labelled extension probe, while gel 2 shows the effect of increasing incubation times with DEPC with labelled target strand. In each case it evident that certain bases close to the branch point are much more susceptible to modification than in the remainder of fragment (Fig. 4.4b). Clearly, the three AT pairs in the extension probe/target arm flanking the branch point are prone to attack by both chemical probes. In each case the AT pair located at the branch point shows the highest degree of reactivity, which is most noticeable with DEPC probing for which only the A directly flanking the branch point shows a high degree of reactivity. These results suggest that these bases were either unpaired or resided in a more transient hydrogen bonded state.

The presence of (perhaps transient) unpaired bases at the junction could explain the unusual folding of PJ1 as it clearly does not assume a perfect pyramidal shape. Since unpaired AT pairs can form pseudo-bulges, these might act as flexible linkers, allowing co-axial stacking to occur. This possibility was recently suggested by Altona *et al.* (1996) from analysis of a number junctions that lacked unpaired bases, yet displayed clear folding (Guo *et al.*, 1991, Lu *et al.*, 1991). Altona *et al.* (1996) further postulated that AT pairs located close to the branch point should be more

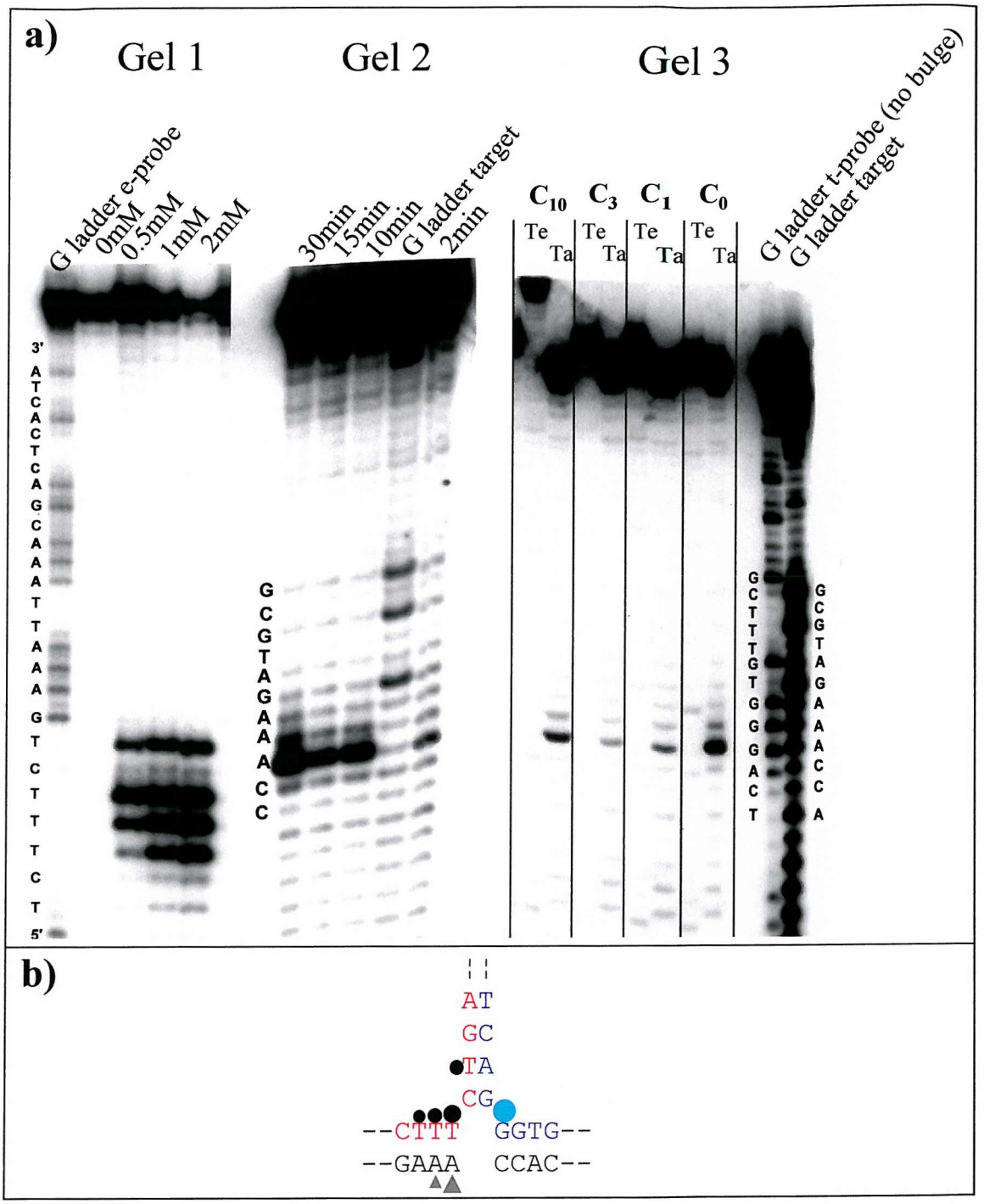


Fig. 4.4 Chemical probing of PJ1 (gel 1 & 2) and PJ1 with different sized bulges (gel 3). PJ1 was probed at room temperature with both $KMnO_4$ (gel 1, using a 5' radiolabelled extension probe (e-probe) and different amounts of $KMnO_4$) and DEPC (using 5'-radiolabelled target and various DEPC incubation times, 1µl of a 100% DEPC solution was applied in all cases). Bulged PJ1 junctions (gel 3) were probed only with DEPC, using either radiolabelled target (Ta) or template probe (Te). Fig. 5.4b shows the bases reactive to $KMnO_4$ (black circles) and/or DEPC (triangles) as deduced from gels 1-3. The size of the symbols provide an approximate indication of the reactivity of the bases. The green circle shows the location of the unpaired bases.

prone to fraying than GC pairs, simply because the former contains one less hydrogen bond.

The observed dissociation of the AT pairs in PJ1 does not exclude the possibility that the extension probe/target arm was not stacked on the template probe/target arm. Naively one might expect that those base pairs flanking the branch point and part of the stack would be more stable than those in the unstacked arm. Other studies have shown that base pairs that are part of the stacked arms can show substantially disrupted hydrogen bonding, even in the presence of unpaired bases (Zhong *et al.*, 1994). It is also worth noting that chemical probing itself may affect the conformation of the junction. For chemical probing to produce reliable results, probing should be performed under "single-hit" conditions, meaning that a single junction molecule is cleaved no more than once by a chemical probe. It is evident that cleavage of a transient base pair can facilitate the cleavage of an adjacent or nearby located pair. The cleavage of the first pair could reduce the stability of its neighbour(s). This could explain the higher reactivity of $KMnO_4$, since gel 1 in Fig. 4.4a shows that much more junction was cleaved than for DEPC cleavage (gel 2), as judged from the intensity of the uncleaved junction bands. In other words, it appears unlikely that $KMnO_4$ probing was performed in the single-hit regime, whereas DEPC probing was.

Finally, gel 3 in Fig. 4.4 reveals that increasing the bulge size does not appear to affect the reactivity of the branch point A-T pair for DEPC (30 minute reaction at room temperature). In other words, the changes in the angles between the different arms upon increasing the bulge size (Fig. 4.2c) have no effect on DEPC reactivity. Previous studies indicate that if different co-axial stacked conformations are assumed, usually the sensitivity for chemical probes also alter because of the different stacking arrangements. Although this is not always the case, this could mean that the PJ1-junctions with different sized bulges all resided in a similar conformation. This supports the suggestion in Fig. 4.2c, that PJ1 (lacking unpaired cystosines) and PJ1C_1a were unstacked rather than co-axially stacked.

4.2.4 Control experiments for the LSA assay: effect of changing the R-ENase sites in PJ1- and PJ2-based junctions

The sequence of PJ1 was based on the original SMART J-CF1 in order to retain as much sequence similarity as possible and therefore contained sites for R-ENases that have not been characterised as extensively as the more commonly used *Bam*HI or *Eco*RI. Therefore, PJ1 and PJ2 were redesigned such that they contained a number of different R-ENase cleavage sites as outlined in Fig. 4.5. PJ11, 14 and 18 were based on PJ1, while PJ12, 13 and 15 were based on PJ2. In contrast to the previous junctions, the restriction enzymes used all produce cleavage products with 4-base overhangs. These junctions at the same time allowed testing of the assumption that the composition of the restriction enzyme sites (and their overhanging staggered products) does not affect the mobility of the resulting fragments.

Fig. 4.6 and 4.7 show the LSA assay results for PJ11-15 and PJ18, using the non-isotopic format of the LSA assay. Since the R-ENase site should not affect the conformer equilibrium, it was expected that junctions sharing the same branch point sequences would yield the same mobility patterns. Fig. 4.6 & 4.7 clearly show that this was not the case, indicating that the choice of R-ENase affects the mobility pattern. Comparing PJ12C_2a , PJ13C_2a and PJ15C_2a a substantial difference can be seen in the relative mobility of the different fragments (Fig. 4.6a). Interpretation of the mobility pattern for PJ15C_2a alone would suggest that this junction folds in an extended conformation since the *Bam*HI and *Ava*I cleavage products have identical mobilities and are faster than the *Eco*RI fragment. In contrast, PJ12C_2a and PJ13C_2a gave the pattern expected for a conformer II junction. Although one could argue that the altered R-ENase cleavage sequences affected the conformation of the junction, a more likely explanation is that the different staggered ends of the fragments were responsible instead. To resolve the issue synthetic complexes were prepared, which were identical to the cleaved products of PJ15C_2a. The results for these complexes are shown in Fig. 4.6b. It can be seen that junctions PJ24C_2a and PJ25C_2a (which are identical to the *Bam*HI and *Ava*I fragments of PJ15C_2a respectively) have identical mobilities, both of which are faster than PJ23C_2a (identical to the *Eco*RI fragment of PJ15C_2a). To assess whether the altered mobility of the PJ15C_2a *Ava*I fragment was dependent on the position of the *Ava*I site, an additional junction (PJ26C_2a) was designed, which contained the same sequences but in which the restriction sites were

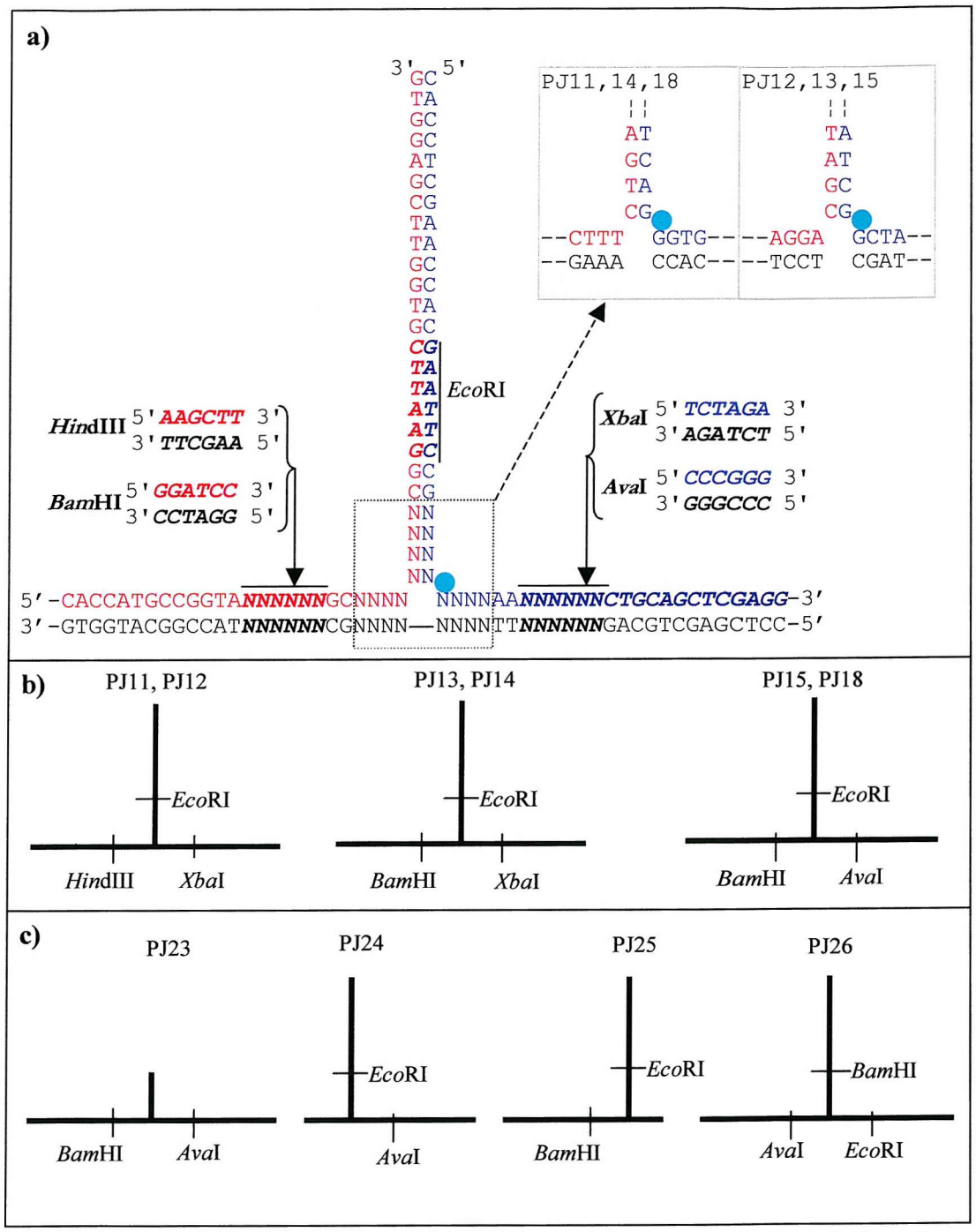


Fig 4.5 Designs of junctions used for assessing the effect of single stranded overhangs on fragment mobility. A) Red indicates the extension primer sequence, blue the template primer sequence and black the target sequence. 'N' indicates the area where the junctions may differ in sequence as outlined in the upper right hand boxes. R-ENase recognition sequences are printed in bold and italics. The junctions contained different R-ENase sites in the extension probe/target and template probe/target arm, which is indicated in b). Synthetic restriction fragments of *Eco*RI, *Bam*HI and *Ava*I-cleaved PJ15 were made as well (c) and a junction where the restriction sites were swapped (PJ26) compared to PJ15.

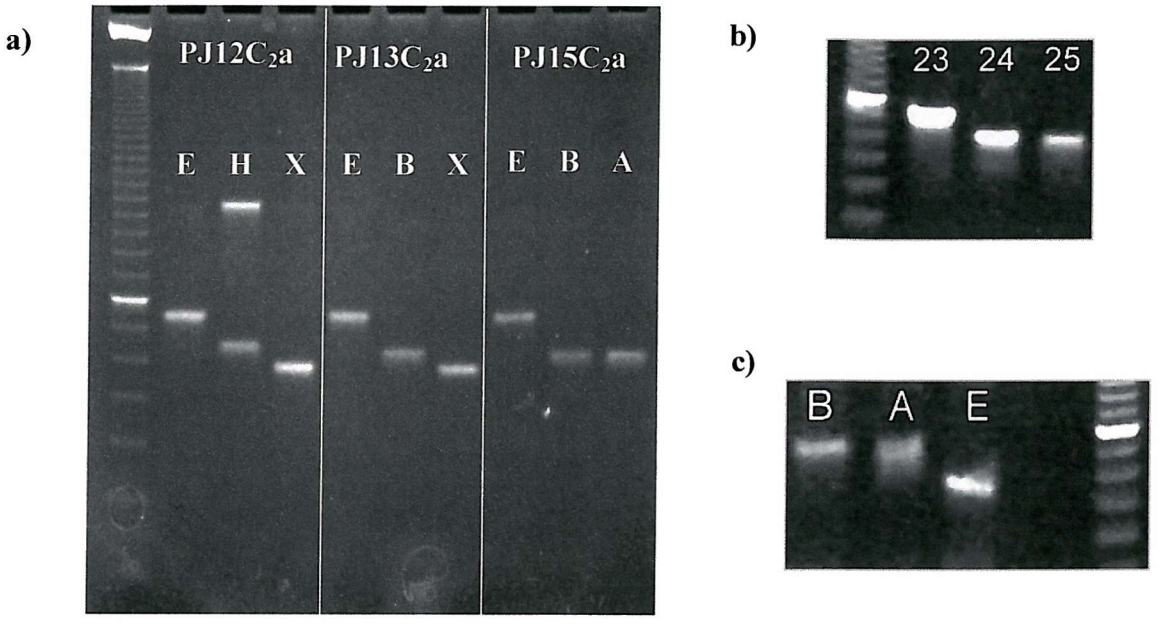
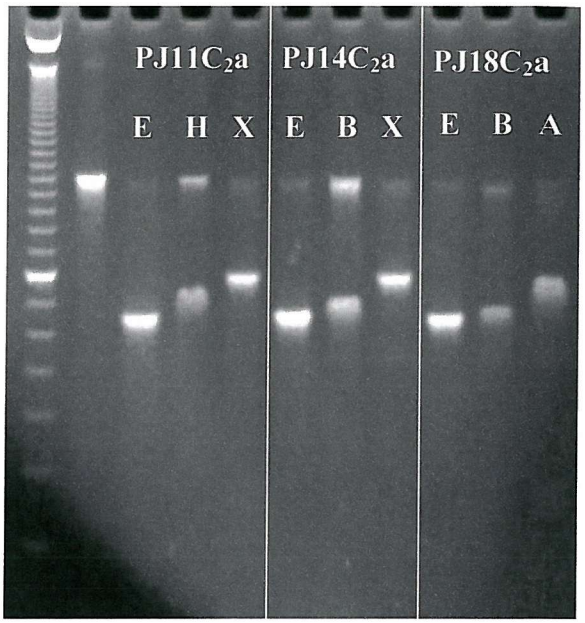


Fig 4.6. Effect of R-ENase cleavage site sequence on the LSA for PJ12C_2a, 13C_2a and 15C_2a. Fig. 4.6b shows synthetic prepared *Eco*RI (PJ23C_2a), *Bam*HI (PJ24C_2a) and *Ava*I (PJ25C_2a) fragments of PJ15C_2a. Fig. 4.6c shows the mobility pattern of PJ26C_2a. For the sequences of the junctions refer to Fig. 4.5. R-ENases are abbreviated to their first letter: E = *Eco*RI, H = *Hin*dIII, X = *Xba*I, B = *Bam*HI and A = *Ava*I.

a)



b)

Fig. 4.7. Effect of R-ENase cleavage site sequence on the LSA for PJ11C_2a, 14C_2a and 18C_2a. The top gel shows the LSA assay results for PJ11C_2a, PJ14C_2a and PJ18C_2a. The bottom gel shows the restriction fragments of PJ11C_2a, 14C_2a and 18C_2a sorted on R-ENase. The order is always: PJ11, PJ14, PJ18. E = *Eco*RI, H = *Hin*dIII, X = *Xba*I, B = *Bam*HI and A = *Ava*I.

swapped so that *Bam*HI, *Ava*I and *Eco*RI sites replaced the *Eco*RI, *Bam*HI and *Ava*I sites in PJ15C_2a respectively (Fig. 4.5c). The results of this experiment are shown in Fig. 4.6c and show that the *Bam*HI and *Ava*I fragments have identical mobilities, which are slower than the *Eco*RI fragment. In each of these cases it appears that the *Ava*I fragments have unusual mobilities when compared to the corresponding *Hin*dIII and *Bam*HI fragments of PJ12C_2a, -13C_2a and -15C_2a.

Similar gel-mobility differences were also apparent for PJ11C_2a, -14C_2a and -18C_2a as illustrated in Fig. 4.7a and b. In Fig. 4.7a it can be seen that the relative mobilities of the three restriction products are not the same in each case. In particular, the *Bam*HI fragments of PJ14C_2a and PJ18C_2a have similar mobilities but migrate faster than the *Hin*dIII fragment of PJ11C_2a. In Fig. 4.7b the fragments were sorted by the different R-ENases shared by each arm in the junctions rather than the junctions themselves, allowing for more accurate comparison between the mobility of the different fragments. From this it is clear that the *Ava*I (PJ18C_2a) fragment has an anomalously fast mobility compared with the *Xba*I cleavage of PJ11C_2a and PJ14C_2a, and that *Hin*dIII cleavage of PJ11C_2a gives a slower migrating species than *Bam*HI cleavage of PJ14C_2a and PJ18C_2a at the same position.

Additional experiments were carried out to further examine the unusual mobility patterns observed for PJ15C_2a. The junction was incubated with the single-strand specific Mung bean nuclease after treatment with each of the enzymes to remove the overhanging ends. Fig. 4.8 shows that this treatment increases the relative mobility of the *Ava*I fragment significantly, so that the pattern now clearly resembles that of a conformer II. The mobility of this *Ava*I product now resembles that of the corresponding *Xba*I fragments of PJ12C_2a and PJ13C_2a. The pattern seen with *Xba*I cleavage of PJ12C_2a or after Mung bean nuclease digestion of PJ15C_2a is presumed to represent the correct conformational distribution since PJ15C_2a was based on the conformer II junction studied by NMR by Rosen & Patel (1993b). A further junction (PJ32C_2a) was designed, containing the same branch-point sequences as PJ15C_2a but with cleavage sites for the blunt cleaving R-ENases *Eco*RV, *Dra*I and *Pvu*II replacing those for *Eco*RI, *Bam*HI and *Ava*I respectively. The results are presented in Fig. 4.9, and shows the pattern expected for a conformer II. It therefore appears that the overhanging ends, in particular that of *Ava*I (5'-CCGG-3'), affect the mobility patterns in the long-short arm assay.

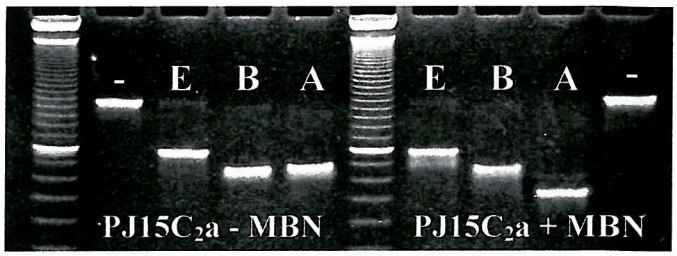


Fig. 4.8. Effect of mung-bean nuclease (MBN) treatment on the mobility of the restriction fragments of junction PJ15C_2a. E = *Eco*RI, B = *Bam*HI, A = *Ava*I, - = untreated junction.

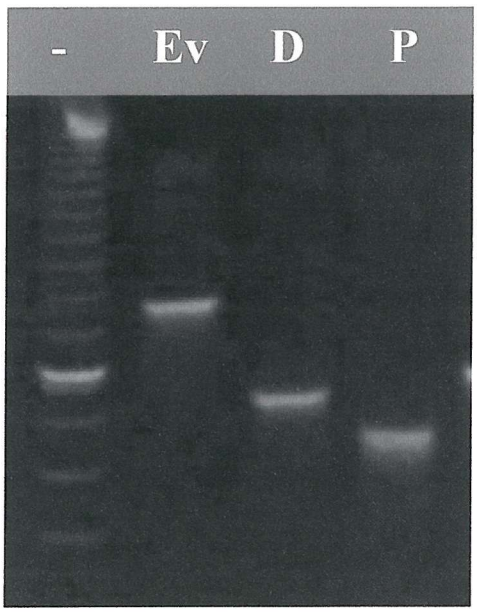


Fig. 4.9 LSA results for a PJ15C_2a-based junction with blunt cleaving R-ENase sites (PJ32C_2a). Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II

A number of PJ15C_2a-like junctions (based on the *Ava*I-cleavage product) were prepared to assess the effect of changing the sequence of the overhanging *Ava*I end on the gel mobility. Seven different sequence constructs were prepared and their sequences are shown in Fig. 4.10a. Electrophoresis was carried out at two different temperatures: and 40°C to test whether any observed mobility differences could be eliminated at elevated temperatures. Fig. 4.10 reveals that at 20°C the different junctions did not all have the same mobility, indicating that the sequence of the overhang affects the apparent conformer. It can be seen that junctions 1, 3 and 7 with overhangs GGCC, CCGG and GCGC respectively have slower mobilities than junctions 2, 4, 5 and 6. In addition, substantial smearing of the bands is apparent at 20°C. These differences were completely eliminated by increasing the electrophoresis temperature to 40°C. From these results it appears that all constructs that contained GC–overhangs produce anomalous mobilities.

4.3 Discussion

The results presented in this chapter provide a first insight into the conformational properties of TWJs. As observed previously, inclusion of a bulge and the size of this bulge can have a substantial impact on junction folding, but more importantly: so does altering the sequence close to the branch point. Unlike previous observations however, the inclusion of a bulge does not necessarily lead to a clearly coaxially stacked conformation, as evidenced by experiments on PJ1. However, it is important to note that the long-short arm assay, used here for structure determination, does not necessarily provide accurate information on the stacked conformation. First, it remains unclear if and to what extend TWJs exist in a dynamic conformational equilibrium like that observed for FWJs. Thus, it is impossible at this stage to conclude that no coaxial stacking interactions occurred in PJ1, only that on the time-scale involved in electrophoresis (i.e. ~4 hours) the average conformation reflects an unstacked conformation.

The remainder of this chapter revealed further, that the long-short arm assay can yield anomalous results when restriction sites are used that leave four-base overhanging ends after cleavage. The exact nature of the effect exerted by the overhanging ends on gel migration remains obscured, but it is clear that it can alter

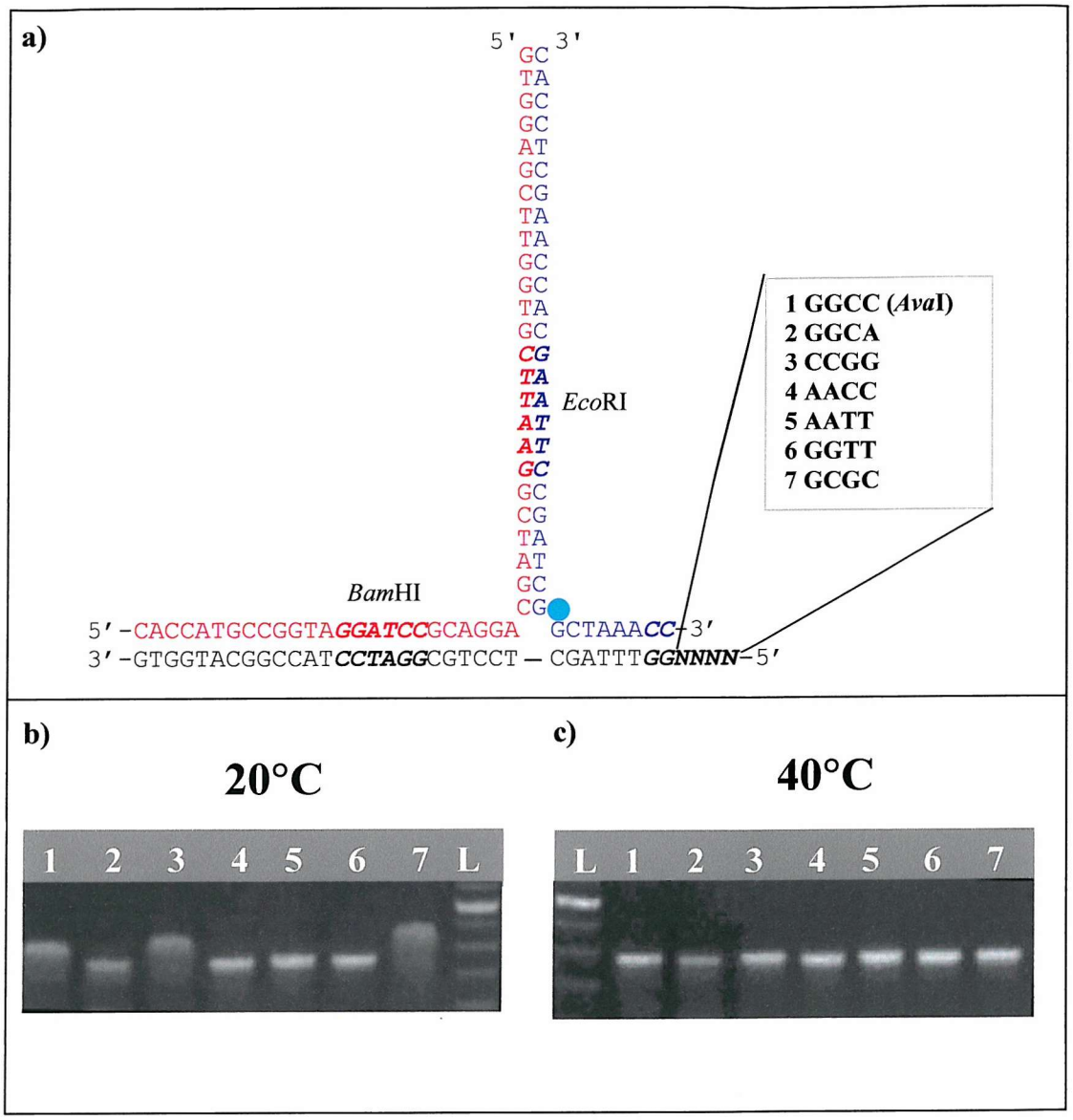


Fig 4.10 Effect of temperature on the mobility of differently staggered PJ15C$_2$a fragments. A) Design of the different junctions. B) & c) electrophoresis results at different electrophoresis temperatures. L = 10 bp ladder (Gibco Life Sciences Ltd., UK).

the mobility pattern independent of junction structure. It seems unlikely that the two terminal bases in the four base overhang could fold back to form Watson-Crick like interactions with the first two bases, as this would be a highly unusual structure. The temperature sensitivity suggests that the effect on gel mobility is not strong. Another possibility could involve transient base pairing between two junction molecules at their staggered ends during electrophoresis. However, the coupling of two junctions would be expected to give rise to much greater retraction. It is worth noting that similar results as presented here, have been reported by Muzniek & Doerfler (1998) and Woodcock *et al.* (1991). They observed that for duplexes ranging between 150-800 bp anomalous migration occurred when these duplexes contained GC-rich overhanging ends of at least four bases. In addition, Muizniek & Doerfler (1998) showed that *Hin*dIII fragments also migrated somewhat slower than expected on the basis of the molecular weight of the respective duplexes. These results therefore also support the observation that the anomalous gel migration is not a function of three-way junction structure but results from the overhanging ends.

Further, Muizniek & Doerfler (1998) noted that the anomalies depended on the presence of cations. Muizniek & Doerfler (1998) suggested that an unusual structure was probably the cause of the observed anomalies, proposing that electrostatic interactions between G and C residues mediated its formation through sliding motions. However, an explanation based solely on an unusual conformation is difficult to reconcile with the observed length-independency since fragments ranging from 57 (the junction fragments after cleavage) to 800 bp (Woodcock *et al.*, 1991) were similarly affected by a four-base overhang. This suggests that the effect exerted by the overhang dominated the migration through the gel rather than the molecular weight of the DNA polymer. From this perspective, it is interesting to note that Muiznik and Doerfler showed for a number of different overhanging sequences that the level of retardation is *inversely* proportional to polyacrylamide concentration. In other words, the overhanging sequences effecting anomalous gel migration do so more efficiently when the mean pore size is decreased. This strongly suggests that the effect of the overhang is not the same as for curved DNA for which the opposite is always found. Instead, it appears that a critical factor controlling the appearance of the anomalies observed here, is the solvation of the overhang. This follows from the observations that the anomaly can be affected by phosphatase treatment, cation concentration, cytosine methylation and the DNA sequence itself (Muiznieks and

Doerfler, 1998, Woodcock *et al.*, 1991). These four factors have in common a direct effect on the interactions between the overhanging sequences and cations and/or water. Thus, unusual interactions between the overhang and the electric field, the gel matrix, the cations present in the electrophoresis buffer or a combination of these as the most likely explanation for the effects observed. The exact nature of these interactions however, remains obscure since theories attempting to describe DNA motion through hydropolymer gels are presently incomplete (for a thorough discussion refer to Zimm & Levene, 1992). Further analysis of the overhanging sequences studied here and elsewhere, as well as their effects on DNA electrophoresis is therefore required to solve this interesting problem.

Not all restriction endonucleases tested, yielded fragments with anomalous migration. G/C rich fragments such as the one resulting from *Ava*I cleavage showed anomalous gel migration but also *Hin*dIII fragments (with a more AT-rich overhang).

Together these results illustrate that the choice of R-ENases in the LSA assay may affect the resulting mobility pattern and thus, the extrapolated structure. It should be noted that most LSA studies carried out to date (by others) did not include *Ava*I in the junction design. Instead, commonly used R-ENases are *Eco*RI and *Xba*I but also *Hin*dIII (for instance: Welch *et al.*, 1995). The results presented in this section suggest that the cleavage product for *Hin*dIII can also have an altered mobility (although less than *Ava*I) and should be interpreted with care. This effect might also be expected to occur with four-way DNA junctions. However, this study shows that the complications can easily be avoided either by post- R-ENase treatment with mung bean nuclease, elevating the electrophoresis temperature or (most attractively) by using blunt cleaving R-ENases.

5 Characterisation of the sequence-dependency of TWJ folding and its effect on RNA yield in SMART.

5.1 Introduction

5.1.1 Theoretical analysis of (potential) factors involved in TWJ folding

There is still lack of understanding on the factors that dictate the conformation of TWJs (and also FWJs). Although the influence of the sequence around the branch point is well documented as outlined in the introduction, it is still not clear why certain sequences engage in coaxial stacking and others do not. This information is crucial, not only for understanding the conformational properties of branched DNA molecules in general, but also to allow prediction of the TWJ structure based solely on its sequence. The latter would greatly facilitate junction design in SMART as target sites could be identified that are more likely to deliver high RNA yields.

Probably the most informative approach entails high-resolution structural analysis of TWJs differing in only a few bases around the branch point. However, NMR and crystallographic studies do not lend themselves (for practical reasons) for the relative high-throughput such an approach requires. Much has already been revealed about the sequence dependent conformation of B-type DNA from numerous crystallographic and NMR based studies. In this respect, it is fortunate that the helix formed by the two coaxially stacked helixes in a TWJ closely resembles B-type DNA. However, even for B-DNA many questions remain unanswered and theoretical models attempting to predict its conformation on the basis of sequence are still hampered by unknown factors. A full review on B-DNA conformation is beyond the scope of this thesis (the reader is referred to: Calladine & Drew, 1997; Dickerson, 1999; Hartmann & Lavery, 1996; Hunter, 1996; Saenger, 1984, for excellent overviews and introductions) but most models aimed at predicting B-DNA conformation from sequence are based on conformational properties of dinucleotide steps. Packer *et al.* (2000a, 2000b) recently showed that, although certain dinucleotide steps are relatively rigid and independent of base pairs located further away, the influence of neighbouring base pairs can be significant. They (and others) noted that DNA conformation can be regarded as a

result of a structural domino effect, *i.e.* the conformation of one base pair affects the conformation of a neighbouring pair, which affects its neighbouring pair *etcetera*. Another complicating issue concerns the influence of solvent on DNA conformation. Cations and water are known to interact with polarised amine and carbonyl groups in the nucleobases in addition to the charged phosphate oxygens. However, attempts to accurately model these interactions have been confounded mainly by the lack of a theory that couples both base-stacking and long-range electrostatic effects exerted by the solvent, usually forcing substantial simplifications to be introduced in the model.

In spite of these problems, there is an overall consensus that in many respects purines and pyrimidines can be regarded as separate groups regarding their influence on DNA conformation (Packer *et al.*, 2000a; Subirana & Faria, 1997). Fewer differences are usually observed between adenosines and guanosines, and between cytosines and thymidines than between purines and pyrimidines. Based on this we therefore suggest that the sequence-dependency of TWJ conformation is dominated by base-type (*i.e.* purine or pyrimidine) rather than specific nucleobase identity ("R/Y stacking hypothesis").

Further, because base-stacking interactions are considered as one of the major forces controlling DNA conformation (Hunter, 1993, Hunter & Lu, 1997), we hypothesised that conformer I and conformer II junctions share the same stacking-sequence preferences. Fig. 5.1 illustrates this idea further by introducing the concept of conformer conversion. If for a given conformer II junction the unstacked base pairs are swapped with two base pairs in one of the stacked arms such that the four-base pair coaxial stacking sequence is now positioned in a conformer I fashion, a conformer I junction may be expected. Although exactly the same sequences are present (retaining the same strand polarities), there are substantial differences with respect to the orientation of the base pairs relative to the branch point. In a conformer I junction, the major groove faces of the base pairs at the branch point in the unstacked arm are orientated towards the minor groove edges of the co-axially stacked base pairs. This situation is reversed in a conformer II junction, where the minor groove of the unstacked arm faces the major groove of the co-axially stacked arm. This can have important implication for the conformation of the TWJ as the TWJ-solvent interactions are likely to differ because of these differences.

Another difference between conformer I and II is the conformation of the backbones of the different strands. In Fig. 5.1, the orange strand of the left conformer

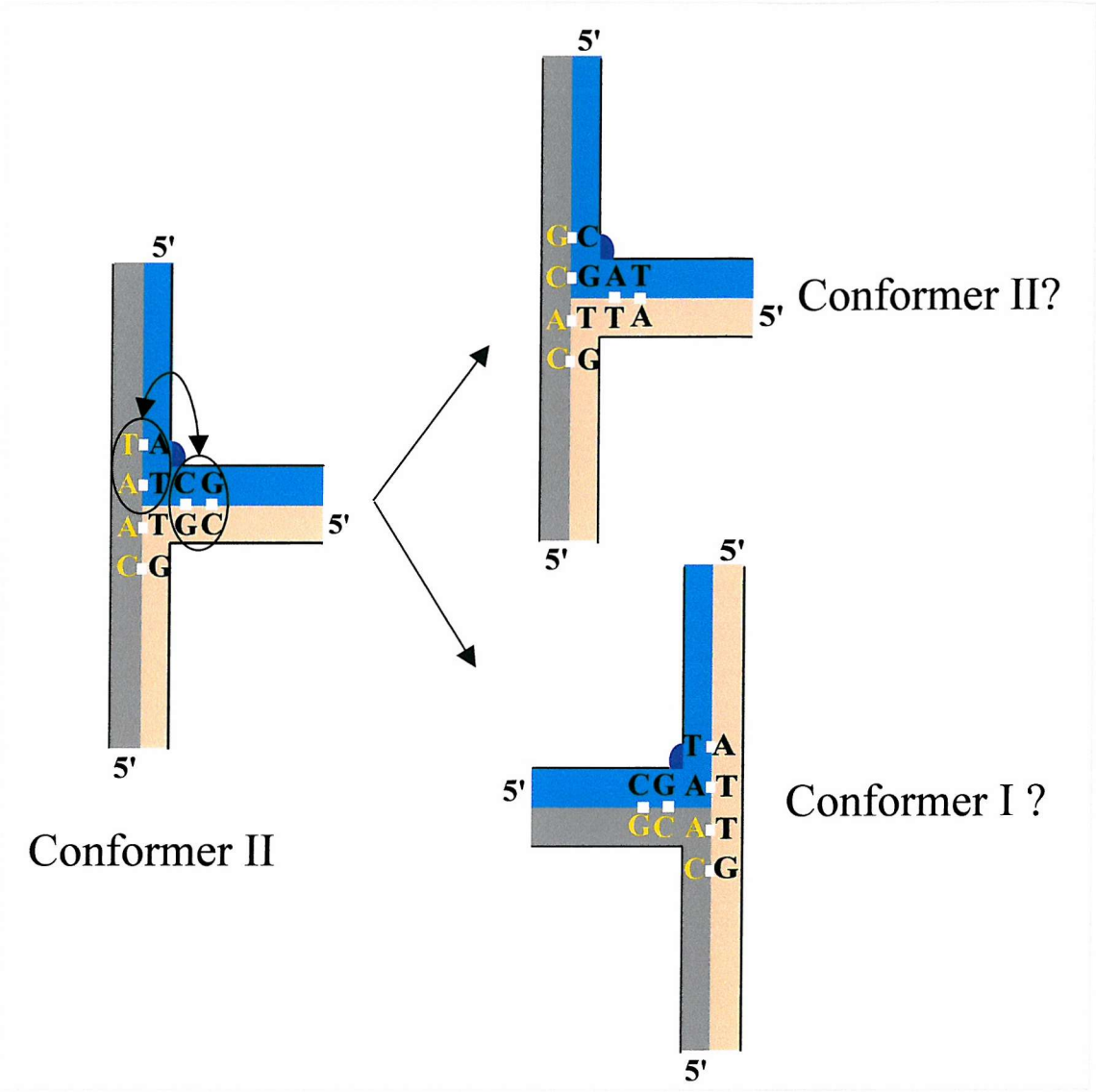


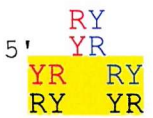
Fig. 5.1 Schematic representation of the conformer conversion approach to assess the importance of base-stacking interactions in TWJs. The three strands are indicated by the different colours and the unpaired bases by the dark blue half-circle. By swapping the two base pairs in the grey/blue arm with the two pairs in the blue/orange pair in the conformer II junction, the third junction from the left results. However, since the coaxial stack in the conformer II is CAAT/ATTG, and the sequences present in the third junction are exactly the same, it is anticipated that this junction will assume a conformer I junction to retain CAAT/ATTG stacking resulting in the junction shown to the far right. The reason why the two conformers are presented differently than before is to illustrate that although the stacked and unstacked sequences remain the same, their orientations relative to the branch point have altered substantially. This could influence for instance the local water and cation distribution. Also the backbone continuities have altered dramatically as in the conformer II CAAT runs continuous from one arm into the other, whereas in the conformer I the unstacked helix interrupts this sequence in the centre. Experiments such as these therefore allow an assessment of whether stacking interactions dominate over forces exerted by the backbone and solvent.

II junction is the exchanging strand, while the grey strand is the continuous strand. Although the blue strand is an exchanging strand as well, it is likely to be conformationally more flexible due to the larger steric space provided by the unpaired bases. In contrast to this conformer II junction, the conformer I junction in Fig. 5.1 has the grey strand as exchanging strand and the blue strand has now become the continuous strand.

The exchanging strand in a TWJ is conformationally rigid and its structure is virtually identical to the exchanging strands in a FWJ. Fig. 5.2 compares the exchanging strand of a FWJ (Eichmann *et al.*, 2000) and TWJ (Leontis *et al.*, 1994) with the backbone conformation of a standard B-DNA helix (Dickerson *et al.*, 1982). Several torsion angles in the FWJ and TWJ, especially χ, show large deviations from B-DNA values (see for instance: Saenger, 1984, Berman, 1997). Thus, these torsion angles probably establish the transition from the coaxial stack to the unstacked arm. Studies have determined that the geometry of base pairs can affect the conformation of the backbone (Packer & Hunter, 1998). Thus, the ultimate configuration of TWJs may involve a choice for base pairs that facilitate the formation of the required backbone conformation.

In summary, the two conformers are not equivalent and may show differences in stacking preferences due to dissimilarities in solvation and/or backbone conformation. The presence of stereochemical differences has been postulated to cause thermodynamic differences with respect to junction stability (Welch *et al.*, 1995). Together with the arguments discussed earlier, this could imply that even if the same stacked and unstacked sequences are present in both junctions, both may still show different stacking arrangements because of these differences. The conformer conversion method therefore allows one to obtain an indication of the importance of base-stacking interactions compared to other forces that act upon the junction such as those exerted by the solvent and the DNA backbones. To explore whether the conversion shown in Fig. 5.1 is feasible we first compared the stacked and unstacked sequences in junctions studied by others (table 5.1). The distinction between high and low energy stacking sequences was made, because in any given junction that undergoes coaxial stacking only two possible stacking sequences exist. For junctions where a clear conformer preference exists, one sequence is therefore more favourable (*i.e.* of lower free energy) in terms of coaxial stacking than the other

Table 5.1. Literature and experiment derived conformer I and II junctions are listed with their stacked and unstacked sequences. Colours are used to discriminate between the same and different R/Y sequences. Only junctions where a linker was incorporated at the branch point were included. Examples where no clear conformer I or II structure was apparent (e.g. Y-shaped) are omitted as well. For instance, the sequence of TWJ-TC (Leontis *et al*, 1994) was:



where yellow indicates the observed low energy stacking sequence. All junctions contained a two base bulge in the blue strand (corresponds to the template probe strand in the junctions described in this thesis). Refer to the appendix of this chapter for a full illustration of the conformation of the junctions listed here.

Conformer I junctions				Conformer II Junctions			
Low Energy	***High Energy***	***Junction***	***Reference***	***Low Energy***	***High Energy***	***Junction***	***Reference***
YRRY/RYYR	YRYR/YRYR	TWJ-TC	Leontis *et al*, 1994	RYYR/YRRY	RYRY/RYRY	TWJ2	Overmars *et al*, 1996
YRRY/RYYR	YRYR/YRYR	J1V6	Welch *et al*, 1995	RRYR/YRYY	RRRY/RYYY	TWJ1	Overmars *et al*, 1996
YRYY/RRYR	YRRR/YYYR	J1	Welch *et al*, 1995	RRYR/YRYY	RRRY/RYYY	J3CC	Rosen & Patel, 1993
YRYY/RRYR	YRYR/YRYR	J1V1	Welch *et al*, 1995	YYYR/YRRR	YYRY/RYRR	YTT	Zhong *et al*, 1994
YRYY/RRYR	YRYR/YRYR	J1V2	Welch *et al*, 1995				
YYRY/RYRR	YYRR/YYRR	J1V4	Welch *et al*, 1995				
YYYY/RRRR	YYRR/YYRR	J1V3	Welch *et al*, 1995				
YYYY/RRRR	YYRR/YYRR	J1V5	Welch *et al*, 1995				

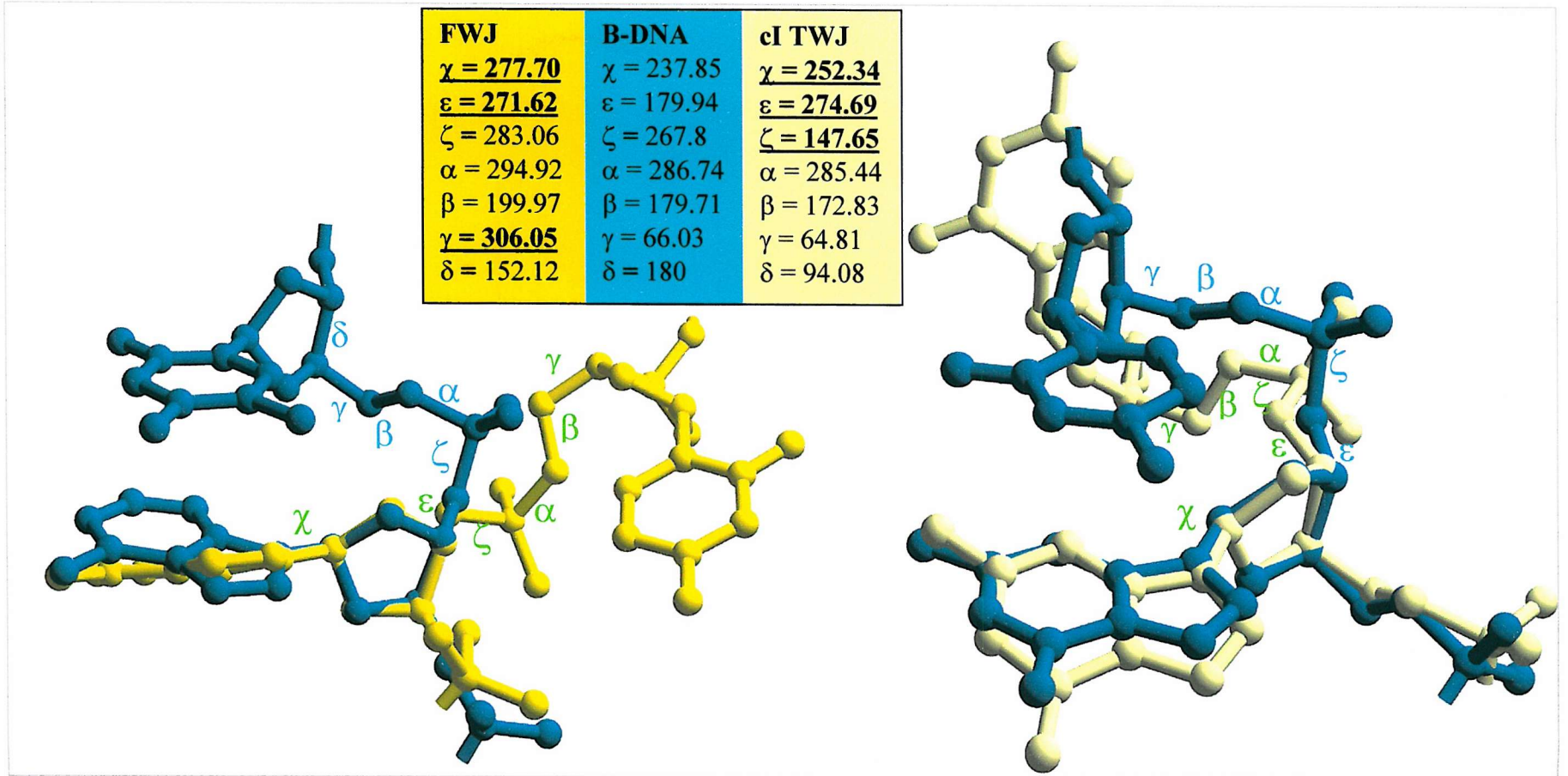


Fig. 5.2. Comparison of the backbone torsion angles of the exchanging strands in a FWJ (yellow, left), a conformer I TWJ (light yellow, right) and B-DNA (green). The figure highlights that FWJs and TWJs both differ in χ and ε torsion angles with respect to B-DNA and that the pseudorotation, α and β angles are close to B-DNA values (the pseudorotation angle describes the sugar pucker, which is C2'-endo). From this comparison it appears that the torsions χ, ε and γ or ζ are involved in establishing crossing over by rotating and extending the backbone away from the coaxial stack. The differences in the other angles between FWJs and TWJs could be caused by the observation that the conformer I TWJ showed a relatively large angle between the unstacked and coaxially stacked helixes (~90°). Please note that Eichman *et al.* considered the β torsion rather than the ε torsion significantly different (in addition to χ). However, they compared this angle with the average β torsions from the B-type nucleotides in their structure (146°), and not with the general range in which it occurs in B-DNA, which is the trans range (Saenger, 1984, Berman, 1997). Thus, it is the average they used that caused this misinterpretation. The angles were calculated using swiss-pdb viewer. Manual fitting of the furanose rings of the FWJ, TWJ and B-DNA was carried out with Swiss-pdb viewer. Swiss-pdb viewer and Pov-Ray for Windows were used to render the image.

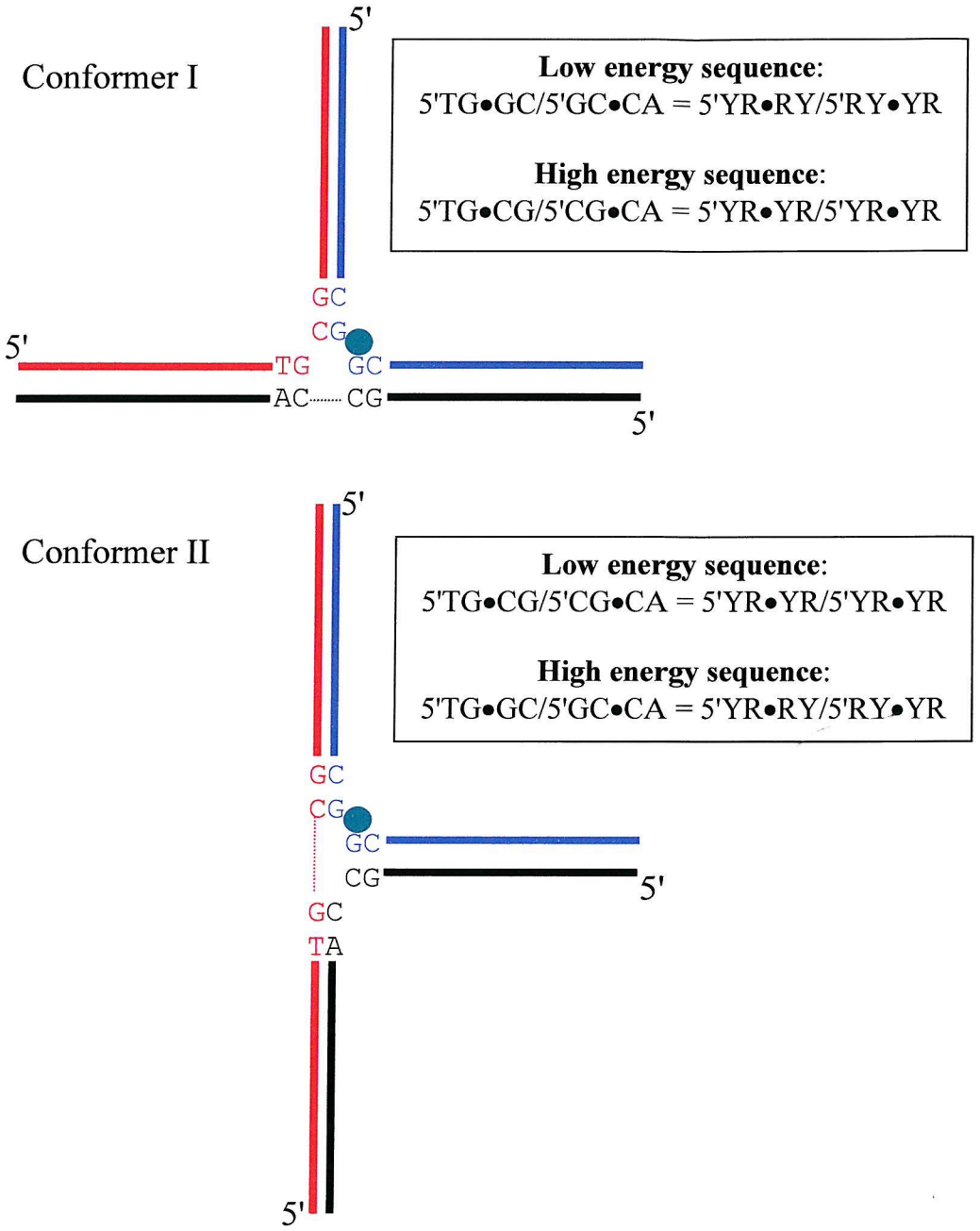


Fig. 5.3. Discrimination between high and low energy sequences. For any junction, two conformers are possible, and the choice between the two is governed by the sequence-dependent coaxial stacking interactions. Of the two possible, coaxially stacked structures, the one with the lowest energy will be favoured (providing a clear conformer preference exists). The junctions drawn here are both of exactly the same sequence, but one is drawn as a conformer I (top) and one as a conformer II (bottom), with the filled circle representing the unpaired bases acting as a bulge. The broken lines are shown to indicate there is no gap separating the bases it connects.

possibility. This is further explained in Fig. 5.3. Unfortunately, table 5.1 does not provide an example of a conformer I junction and a conformer II junction both having the same R/Y sequences around the branch point. Thus, table 5.1 cannot be used to draw conclusions about the differences or similarities in stacking preferences between the two conformers. However, table 5.1 does support the R/Y stacking hypothesis. Although the data set is very small, no high-energy sequence occurs in a low energy sequence column when the same R/Y sequences are present around the branch point. For instance, TWJ-TC and J1V6 both have different Watson-Crick bases around the branch point but the same R/Y sequence. Both show a preference for 5'YR•RY3'/5'RY•YR3' stacking (• = branch point) over 5'YR•YR3'/5'YR•YR3'. Unfortunately, the table does not contain an example of a conformer I and a conformer II junction both having the same R/Y sequences. When the actual nucleobase sequences were assessed, no obvious similarities between the stacking sequences of the two conformers were detected.

To test our hypotheses further, junctions were designed in order to

1. Investigate the effect of maintaining the same RY configuration around the branch point while varying the nucleobase identity
2. Assess whether the conformer conversion approach illustrated in Fig. 5.1 actually results in a conformer conversion by attempting to change a conformer I junction derived from table 5.1 to a conformer II and *vice versa*.

5.2 A conformer conversion approach to assess the influence of TWJ structure on SMART.

PJ40C_2a (Fig. 5.4) was derived from a conformer I junction studied by Welch *et al.* (J1, refer to table 5.1 and Welch *et al.*, 1993 & 1995). This junction was used to assess the R/Y stacking hypothesis and whether swapping the base pairs in two arms would yield a conformer II rather than a conformer I as illustrated in Fig. 5.1. The base sequence flanking the branch point was therefore changed while maintaining the same R/Y stacking configuration, but repositioned such that a conformer II structure was expected to form (Fig. 5.4a). The same was done for a literature-derived conformer II junction, resulting in PJ42A_2a (based on junction YTT, see table 5.1 and Zhong *et al.*, 1994), which was therefore expected to become a conformer I

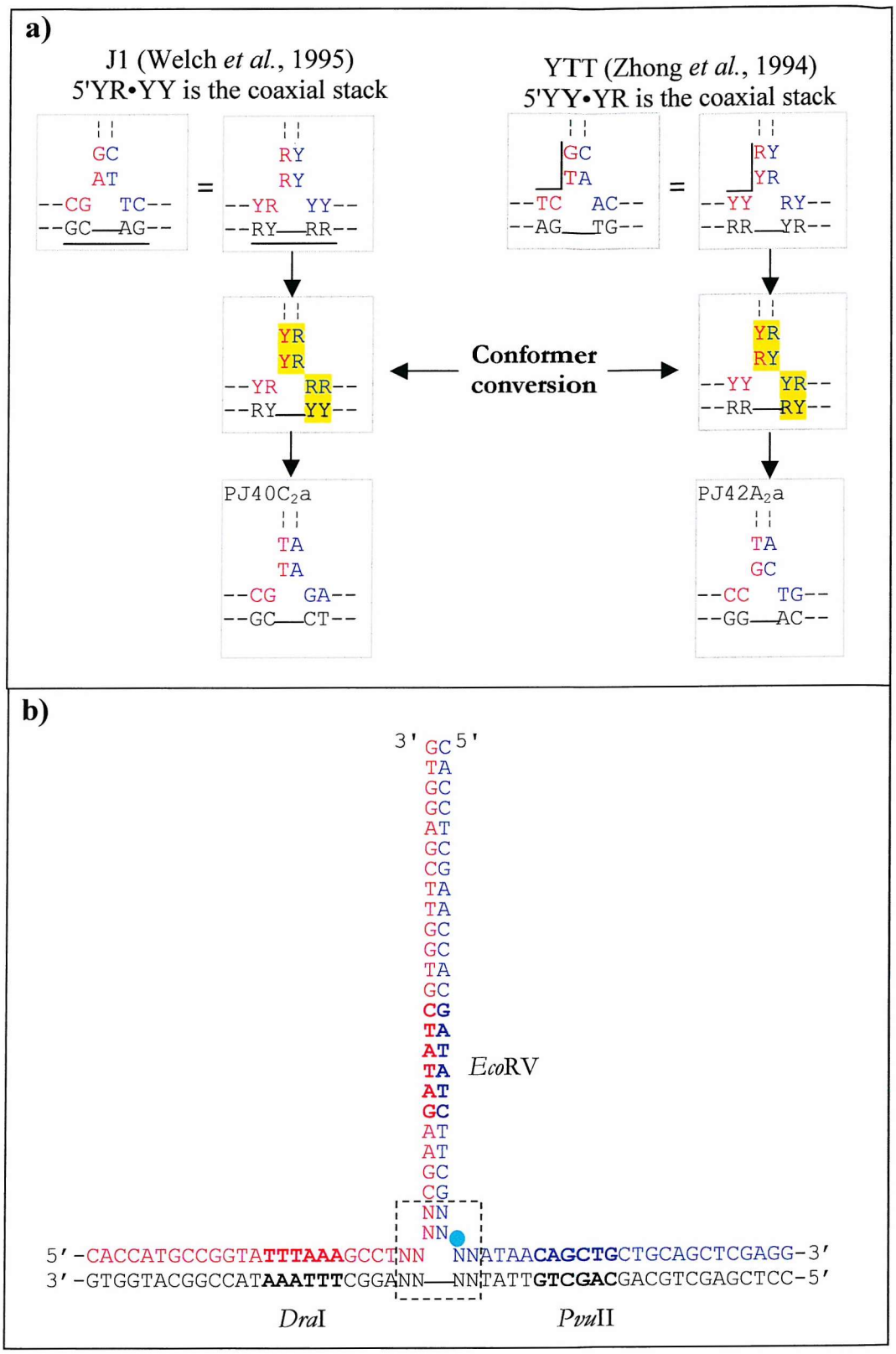


Fig. 5.4 Design of PJ40C_2a and PJ42A_2a. The rationale behind the junction designs is illustrated for both junctions at the top. The junctions are based on YTT and J1 (a conformer II and conformer I respectively, see table 5.1). Two base pairs were swapped between two arms according to Fig. 5.1 (conformer conversion) and indicated here in yellow, so that now the alternative conformer is expected to form. The coaxial stacking sequence is highlighted by a line.
Finally, retaining this R/Y configuration, the nucleobase identity was changed compared to the original junction. Note for instance that for YTT the 5' YY•YR coaxial stacking sequence was TCTG whereas this was changed to CCTG in PJ42A_2a. The junction sequence is indicated in b). The branch point sequence of PJ40C_2a and PJ42A_2a shown in a) are located in the boxed area in b). The green circle indicates the location of the unpaired bases.

junction. The reason why two unpaired adenines instead of cytosines were included is discussed below.

The LSA results for the two junctions are shown in Fig. 5.5. It can be seen that PJ40C_2a shows a clear conformer II mobility pattern. The mobility differences between the three fragments was less pronounced for PJ42A_2a but the pattern indicates that the *Pvu*II fragment is the slowest migrating band, consistent with a conformer I. However, the small differences may imply that the junction folded like a Y-shape rather than a true conformer I. In general, the results correspond to the predicted conformers. Two adenines were used as unpaired bases in PJ42 because PJ42C_2a yielded an unusual LSA mobility pattern that appeared to correspond to an unstacked structure (Fig. 5.6). Analysis of the sequence around the branch point suggested that branch migration was feasible since the two CG pairs located in the *Dra*I arm, could potentially pair with the two cytosines in the template probe strand, generating a bulge in the extension probe strand instead (Fig. 5.7). The fact that replacing the two cytosines with adenosines yielded a different conformation indicates that branch migration is indeed the most likely explanation since other studies found that altering the identity of the unpaired bases did not alter the conformation of the junction (see chapter 1).

To assess whether the TWJ sequences of PJ40C_2a and PJ42A_2a affect the RNA yield in SMART differently, their junction designs were altered. The target spanning arms remained as they were but the extension probe/template probe was truncated to 7bp. The RNA2 promoter/transcription template sequence was added to the 5' end of the template probe. Finally, the two junctions were subjected to a standard SMART reaction, containing 5 fmol of the target strand, 25 fmol of the extension probe and 5 fmol of the template probe. The results, shown in Fig. 5.8, clearly show a greater RNA yield for the conformer II junction ($p<0.05$, unpaired, one-tailed t-test), supporting the hypothesis that junction structure affects the RNA yield. The conformer II junction yielded ~50% more RNA than the conformer I junction. However, it is possible that other factors such as changes in probe secondary structure or differences in junction stability could also have an effect. Therefore six additional junctions were designed based on the conformer conversion approach to yield three conformer I and three related conformer II junctions. The design of the junctions is outlined in Fig. 5.9, showing that three pairs of junctions were designed based on junctions of known conformation. The alternative conformers of these

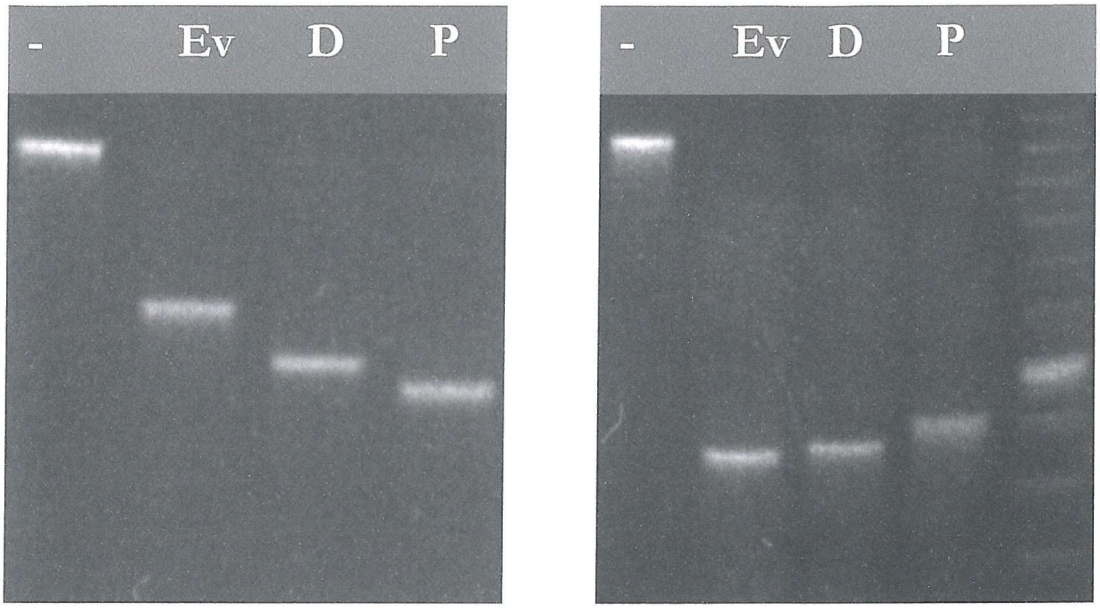


Fig. 5.5 LSA results for PJ40C_2a (left) and PJ42A_2a (right). Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II.

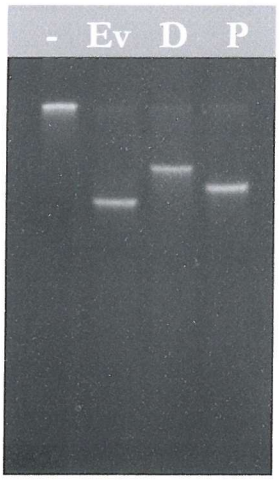


Fig. 5.6 LSA results for PJ42C_2a. Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II.

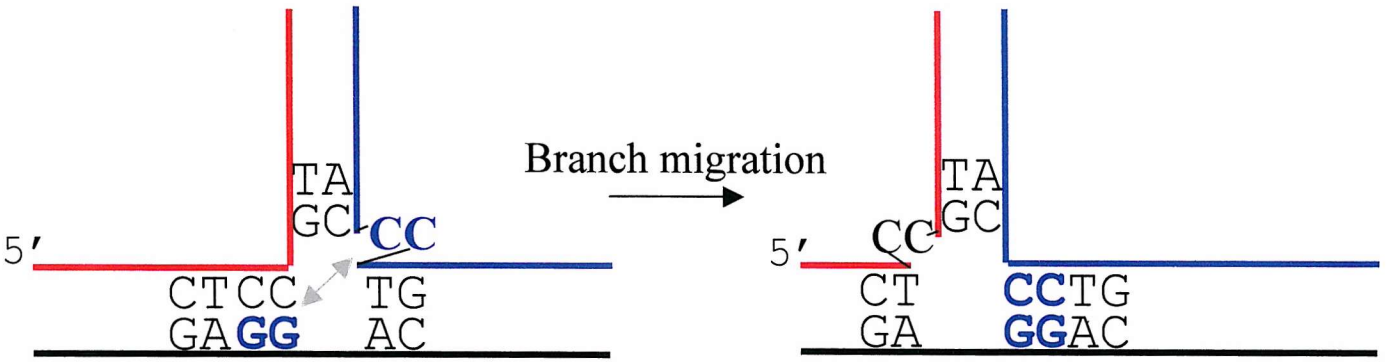


Fig. 5.7. Possible explanation for the LSA results of PJ42C$_2$a. Although the junction was designed to form as shown on the left, it is clear that the two unpaired cytosines (blue) have the potential to hydrogen bond to the two guanosines drawn in blue. This would yield the junction shown to the right and explains why replacing the two unpaired cytosines with adenosines altered the conformation. Clearly, the branch point sequence in the right junction has been altered. Also, the lengths of the different arms are not equal anymore, which invalidates the LSA.

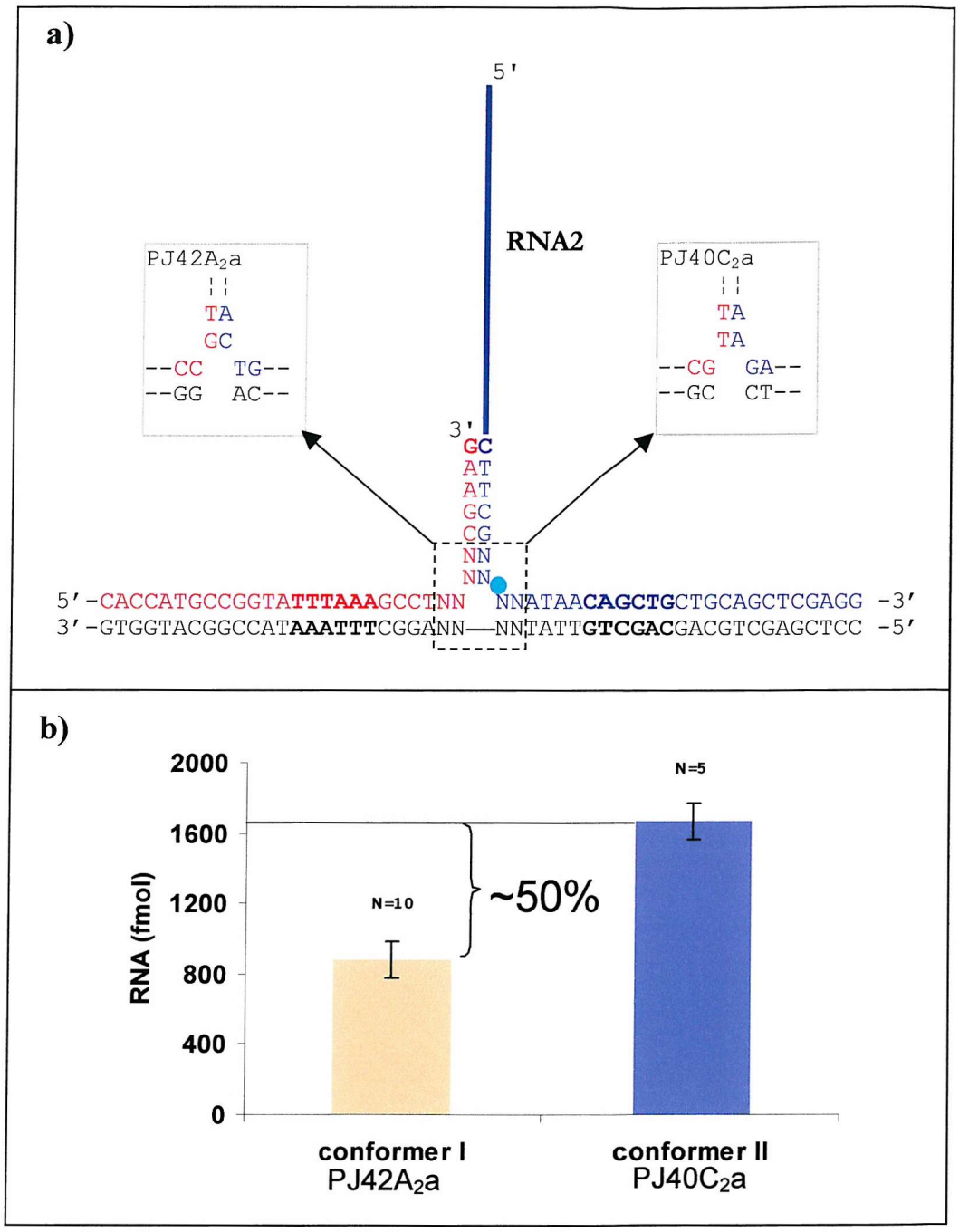


Fig. 5.8 Effect of TWJ structure on SMART RNA yield. Conformer I: $PJ42A_2a$-derived, conformer II: $PJ40C_2a$-derived. The circle indicates the location of the unpaired bases.

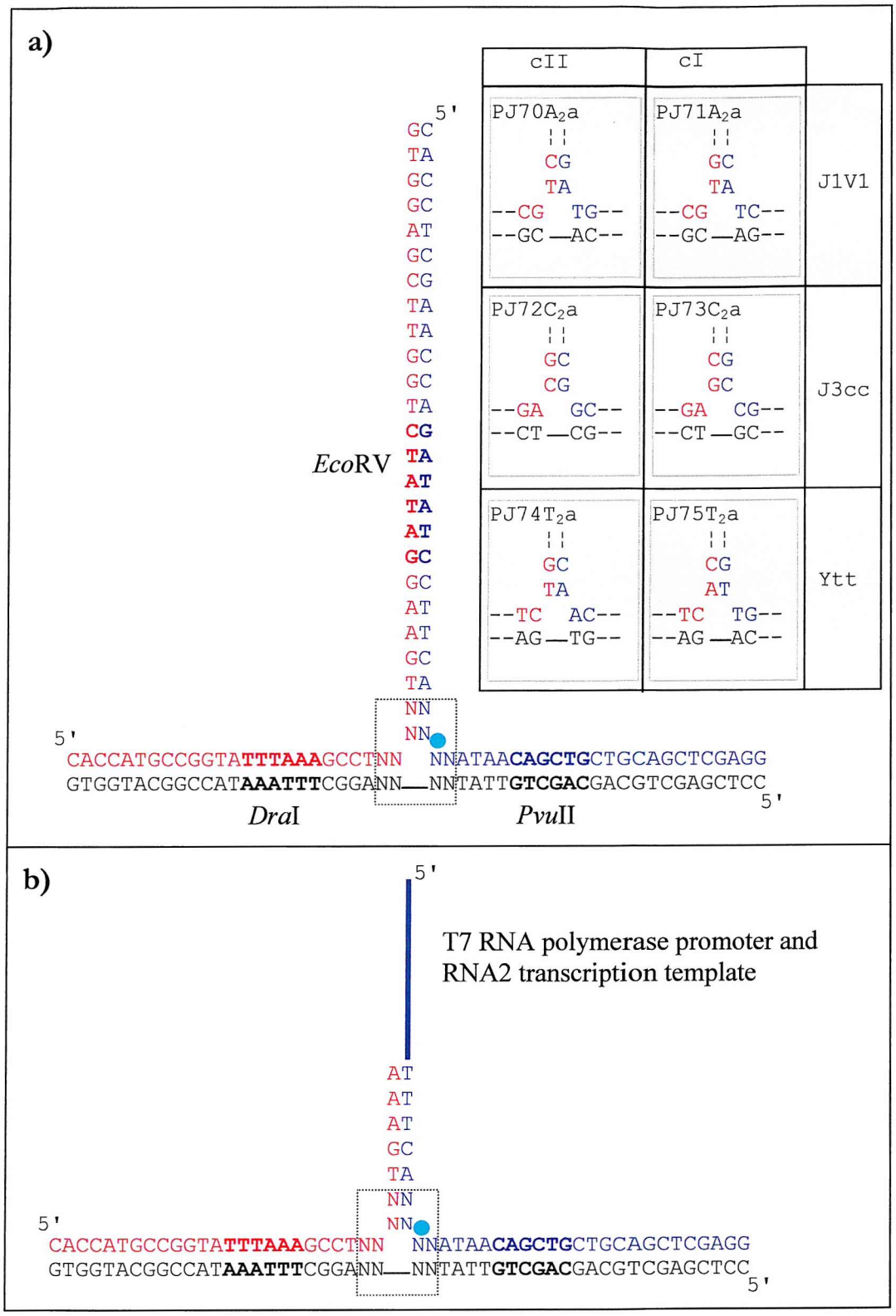


Fig. 5.9. Design of junctions PJ70-75. The branch point sequences of the different junctions are indicated above. Each pair was derived from a junction of known structure and their names are indicated to the right (refer to table 6.1). Using these junctions a conformer II and I version was designed by conformer conversion. The junctions were chosen such that both the original and conformer converted version contained the same number of AT and GC pairs in the overlap. This reduces differences caused by differences in overlap stability. B) Series of related junctions suitable for SMART analysis. The circle indicates the location of the unpaired bases.

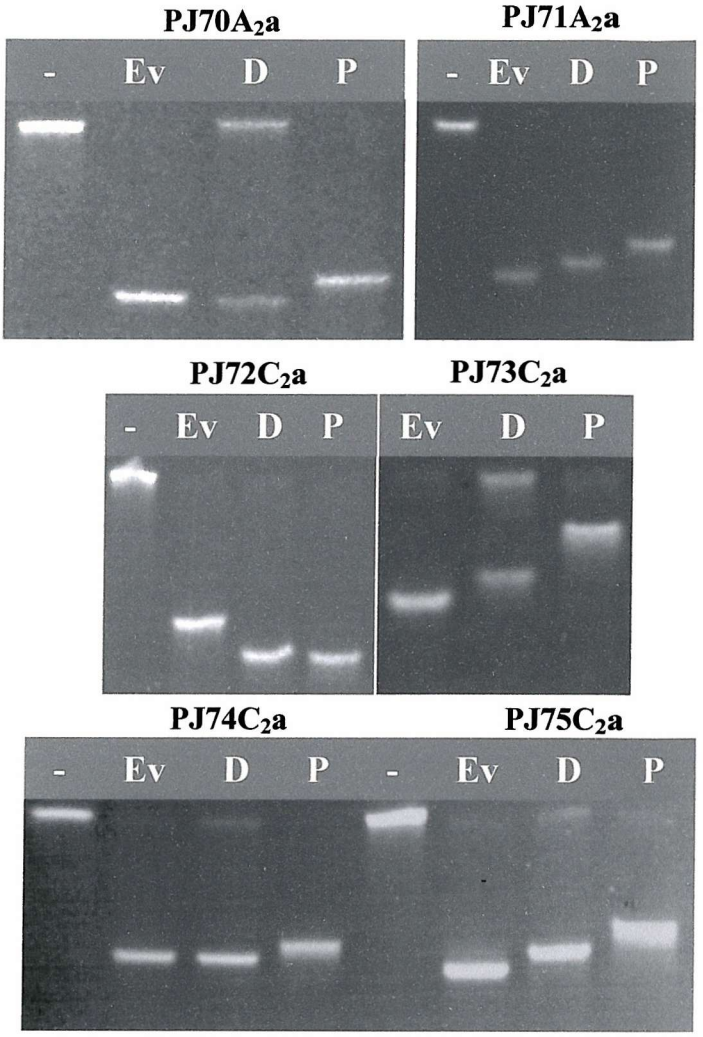
known junctions were designed using the conformer conversion approach. PJ71A_2a is equivalent to J1V1 (conformer I, Welch *et al.*, 1995) with respect to the branch point sequence. By swapping the two base pairs in the extension probe/template probe (red/blue) arm with the two pairs in the template probe/target (blue/black) arm, PJ70A_2a was obtained, which was therefore expected to fold as a conformer II. The other two pairs of junctions were designed in a similar manner. The unpaired bases were chosen so that the chances on alternative base pairing interactions were minimised. Finally, for each of these junctions, a similar junction was designed for determining the RNA yield in the SMART assay. The general design of these junctions is shown in Fig. 5.9b.

The conformer distribution of each of these junctions was determined by the long-short arm assay as shown in Fig. 5.10a. Fig 5.10b shows the RNA yields of these junctions. Interestingly, two of the proposed conformer II junctions (PJ70A_2a and PJ74C_2a) show a pattern resembling the unfolded form (the centre structure in Fig. 1.9b) since the *Dra*I fragments migrated slowest. Since PJ74T_2a was derived from a junction (YTT) for which conformer II conformation had been established, it is likely that sequences beyond the two base pairs from the branch point influenced the conformer equilibrium. Thus, it is not clear whether conformer conversion occurred in all junctions as the results preclude an accurate structural interpretation. Nevertheless, the activity data shows that all proposed conformer II junctions show substantially higher RNA yields than the proposed conformer I junctions.

5.3 Further assessment of the R/Y stacking hypothesis

To assess whether altering the base sequence but not its type (R or Y) affects the conformer equilibrium, PJ44C_2a and PJ45T_2a were designed and their structures analysed by the LSA (Fig. 5.11). These two junctions contained the same R/Y sequence as PJ40C_2a (Fig. 5.4a and section 5.2) but differed in the actual bases around the branch point. Two thymidines were included as unpaired bases in PJ45 rather than two cytosines to avoid base pairing interactions between the sequences flanking the branch point and the bulge as was observed for PJ42C_2a. The LSA results are presented in Fig. 5.11b. It can be seen that, instead of adopting the expected conformer II structure, both PJ44C_2a and PJ45T_2a show mobility patterns corresponding to an unfolded form. Changing one AT pair at the branch point of

a)



b)

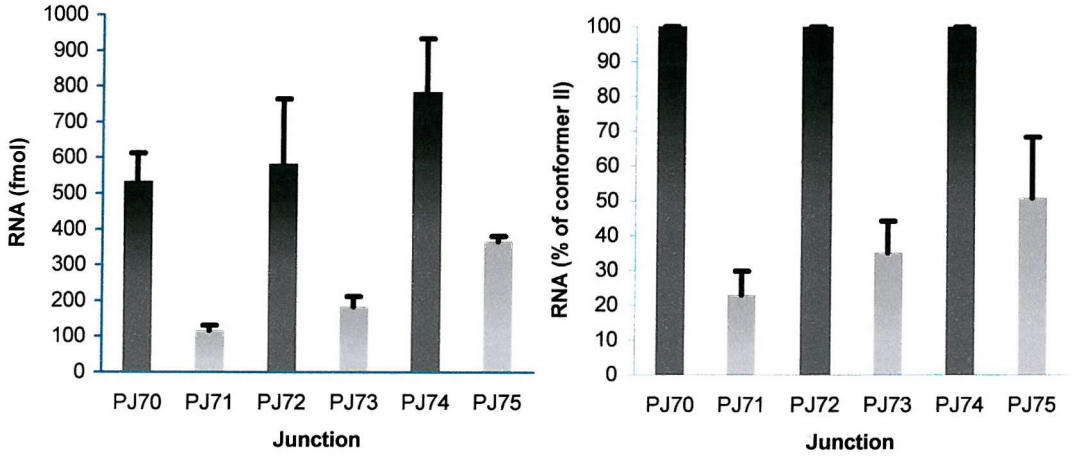


Fig. 5.10. Structure of PJ70-PJ74 (a) and RNA yield of junctions containing the same branch point sequences as PJ70-74 (b). The left graph shows the absolute RNA yield, whereas the signal is expressed as a percentage of the related (assumed) conformer II in the right graph. For instance, PJ70 was postulated to be a conformer II junction whereas the related junction PJ71 was postulated to be a conformer I. Therefore, in the right graph the RNA yield of PJ71 is expressed as a percentage of PJ70. The same was done for the PJ73-74 and PJ75-76 pairs.

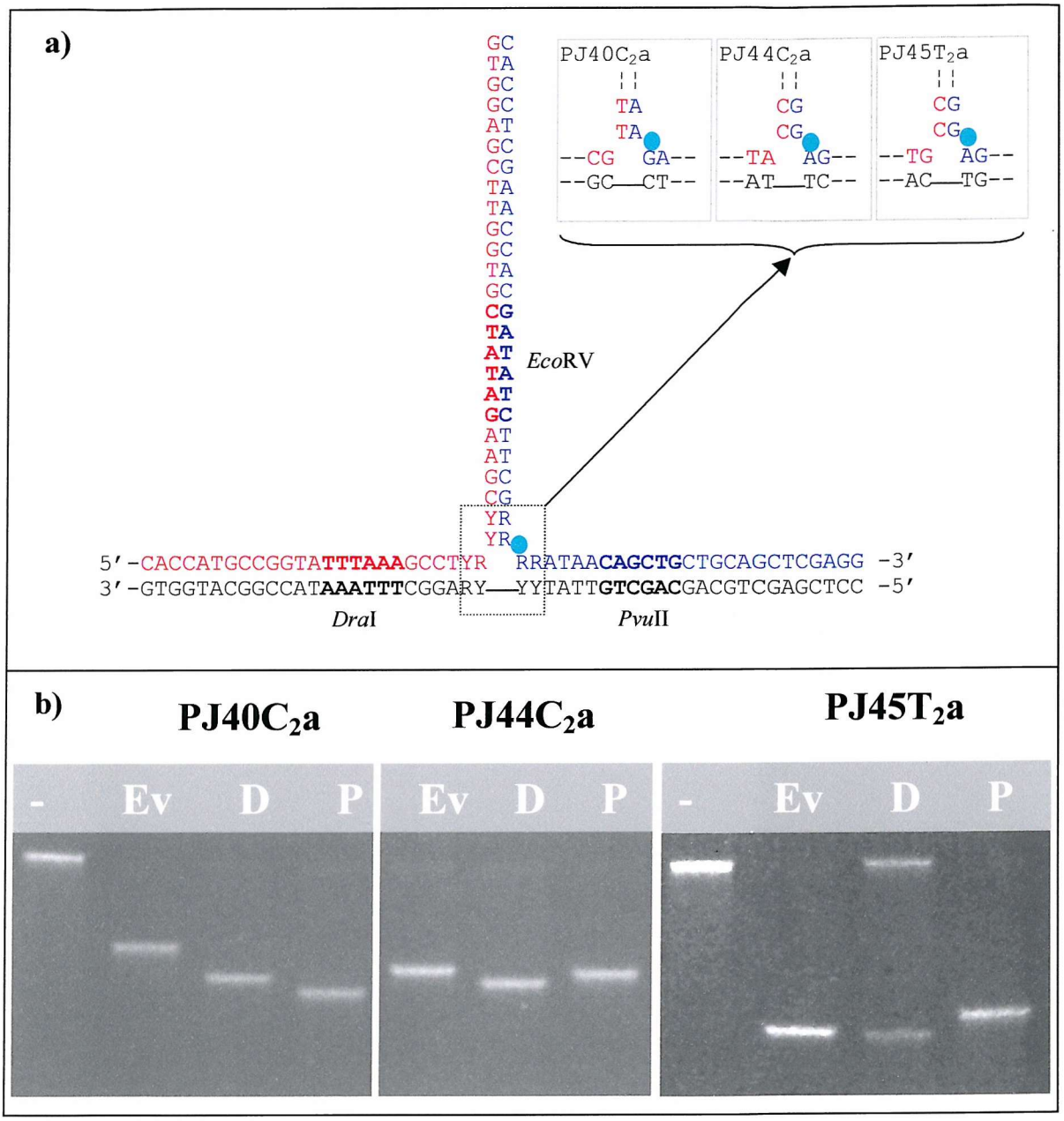


Fig. 5.11 Design and Long-short arm assay results for PJ40C_2a, 44C_2a, 45T_2a. Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II. The circle indicates the location of the unpaired bases.

PJ44C_2a to a GC pair (PJ45T_2a) to prevent any possible misalignment of the two adjacent AT pairs in PJ44C_2a, did not alter the overall pattern although the *Eco*RV now migrated faster than the *Pvu*II fragment. One possibility is that the two AT pairs at the branch point in PJ44C_2a have mispaired, producing a structure in which one AT pair is formed and the remaining A and T nucleotides are unpaired and form new bulges. However, this seems an unfavourable situation since it would involve a loss of hydrogen bonding and stacking interactions. In addition, mispairing appears even less likely in PJ45T_2a. A more likely explanation is that the junctions resided in a conformational equilibrium, where judging from the lower mobility of the *Eco*RV fragment, the conformer II state dominated slightly over the conformer I state. Reasoning along similar lines, PJ45T_2a could also have resided in a conformational equilibrium where the conformer I state dominated over the conformer II state (judged from the lower *Pvu*II fragment mobility).

Similar studies were performed on PJ41C_2a and a PJ41C_2a-derived junction, PJ46T_2a (based on the junction used for Factor V detection in chapter four) shown in Fig. 5.12. It can be seen that changing the bases at the junction while leaving the R/Y sequence unaltered, induced a conformational change, evident from the different relative mobilities. In contrast to the situation for PJ44C_2a and PJ45T_2a, a clear conformer I pattern was observed for PJ46T_2a (Fig. 5.12b). At first sight, these results suggest that the R/Y hypothesis is not correct, *i.e.* the specific base identity does matter. Further experiments were performed using PJ47-49C_2a and PJ51-54C_2a, which were designed based on PJ41C_2a, but each having one or more base pairs changed at the second position while retaining the same R/Y order as PJ41C_2a (Fig. 5.13). The results are shown in Fig. 5.14 and clearly show that these changes did not affect the structure, resulting in conformer II junctions like PJ41C_2a.

Finally, the LSA was applied to three additional junctions (PJ61-63C_2a, Fig. 5.15a), which were based on junctions used for detection of the 23S ribosomal RNA gene in *E. coli* by others at Cytocell Limited, and which showed a substantial degree of difference in RNA yield in the SMART assay. These junctions were derived from one of the junctions shown in table 5.1 (YTT, Zhong *et al.*, 1994), but with the sequences changed following the R/Y stacking hypothesis. The two base pairs in each arm around the branch point of each junction were included in junctions suitable for LSA analysis. The mobility patterns obtained (Fig. 5.15b) clearly show that each junction assumed the same conformation (conformer II). Although this

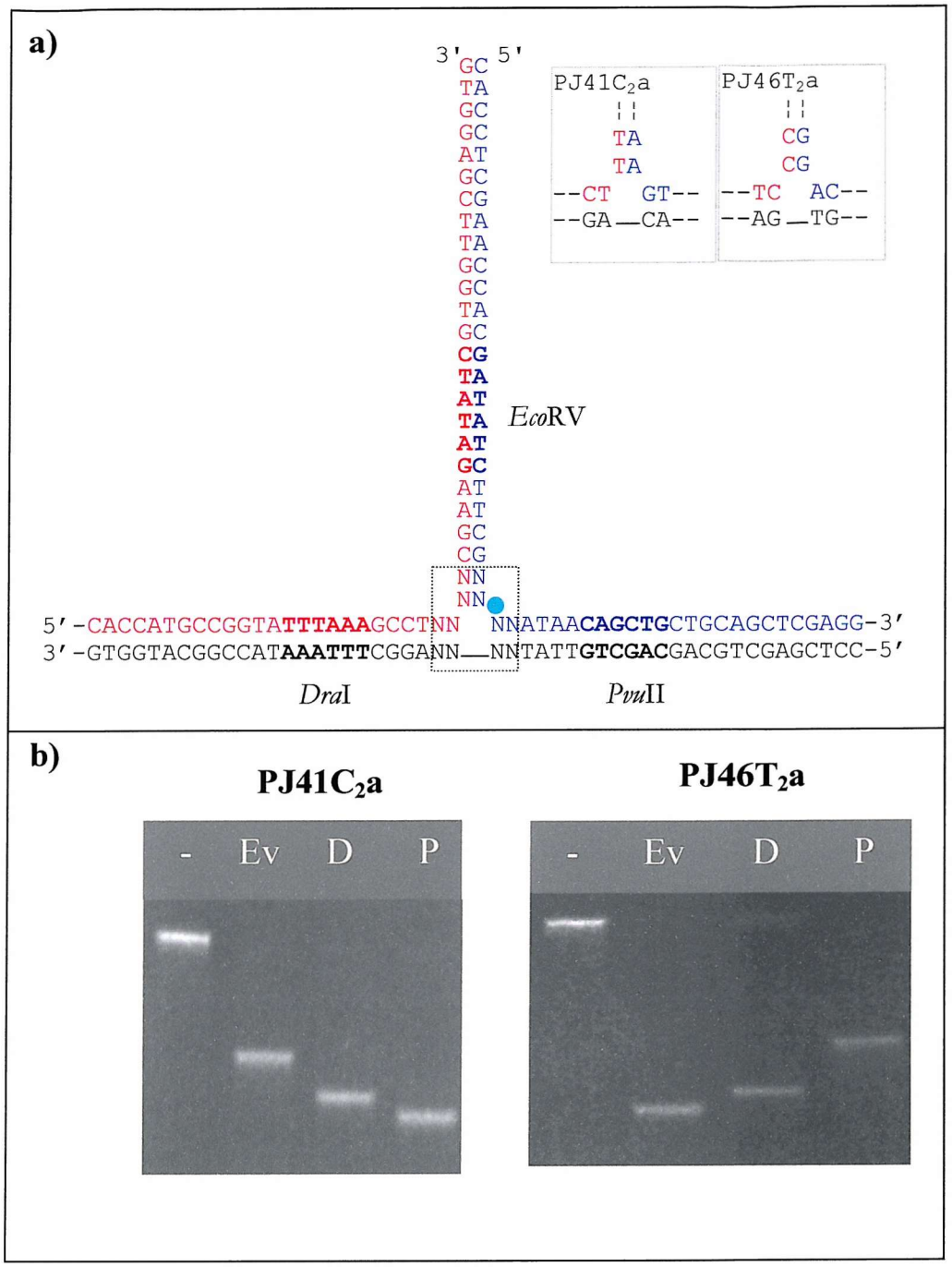


Fig. 5.12. Design and structure analysis of J-LV1C-derived junctions (PJ41C_2a and PJ46T_2a). Ev = *Eco*Rv, D = *Dra*I, P = *Pvu*II. The circle indicates the position of the unpaired bases.

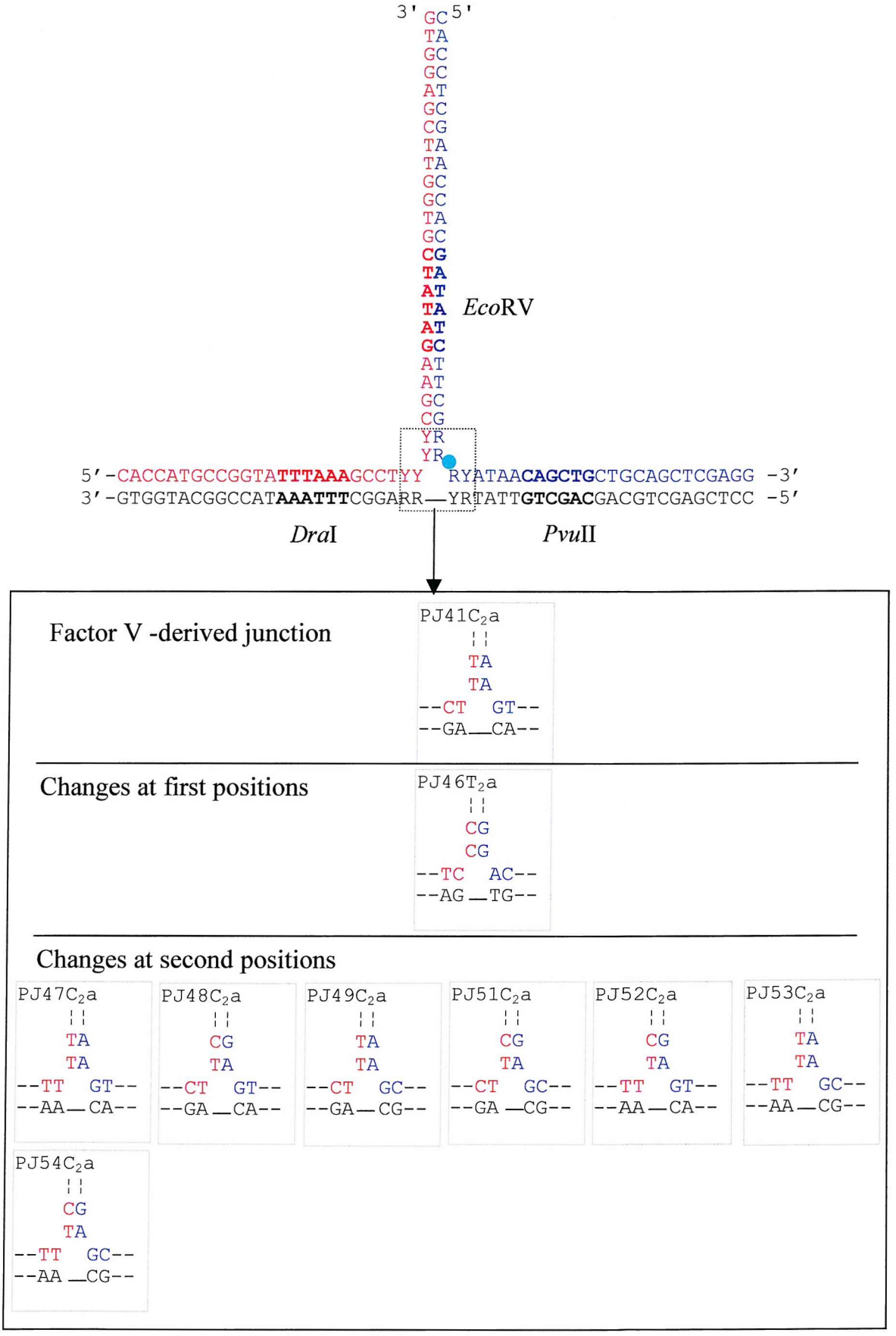


Fig. 5.13 Designs of junctions PJ41, PJ46-49 and PJ51-54 for testing the difference between specific nucleobase and purine/pyrimidine identity of sequences controlling TWJ structure. The circle indicates the location of the unpaired bases.

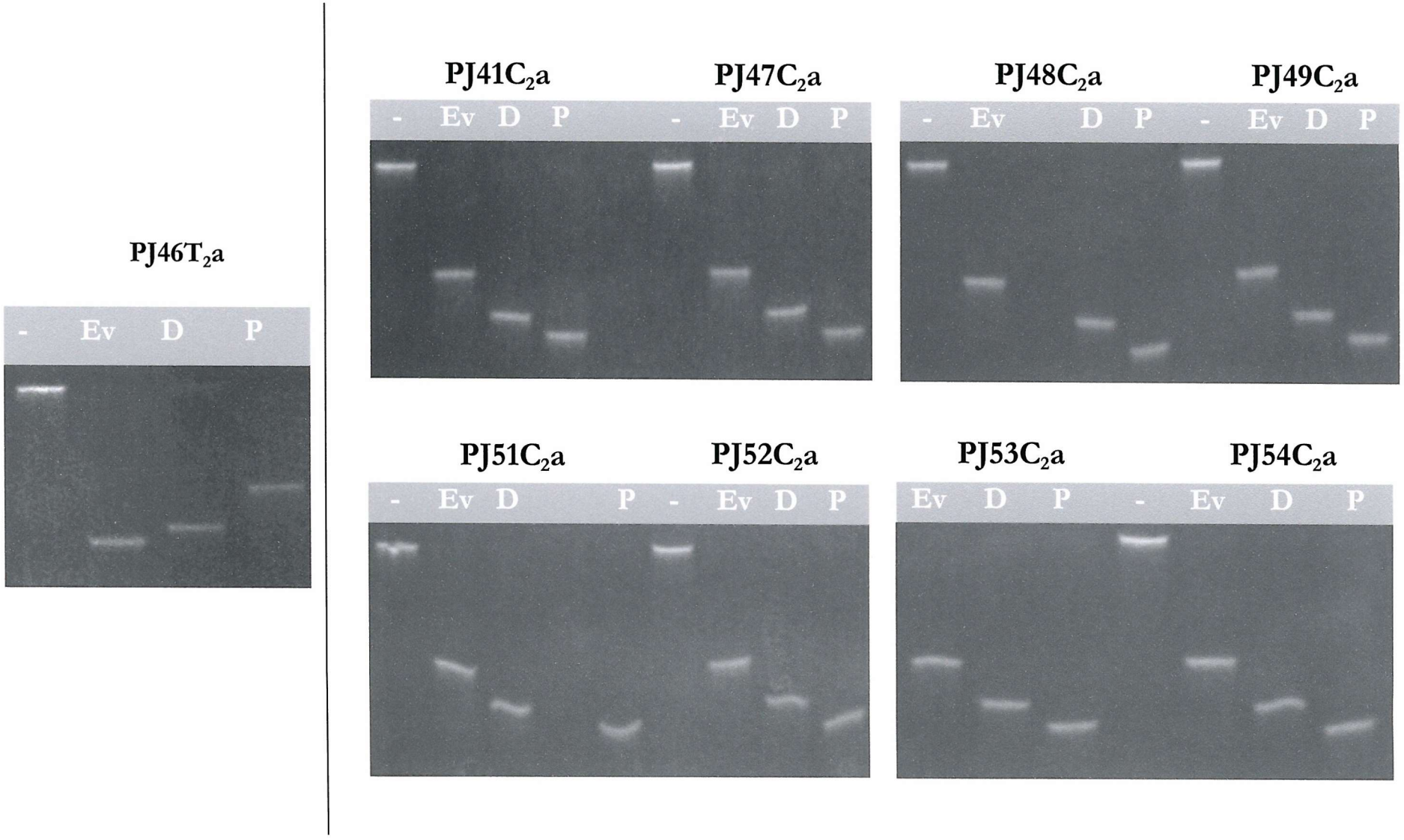


Fig. 5.14 LSA results for PJ41C_2a-derived junctions where base pairs were changed with respect to their identity but not type (i.e. R or Y) at the first and second (PJ46T_2a) or only at the second position (remaining junctions) relative to the branch point in different arms. The somewhat unusual appearance of some of the fragments was probably caused by gel impurities (LSA for PJ47-49C_2a was repeated and showed normal bands, data not shown).

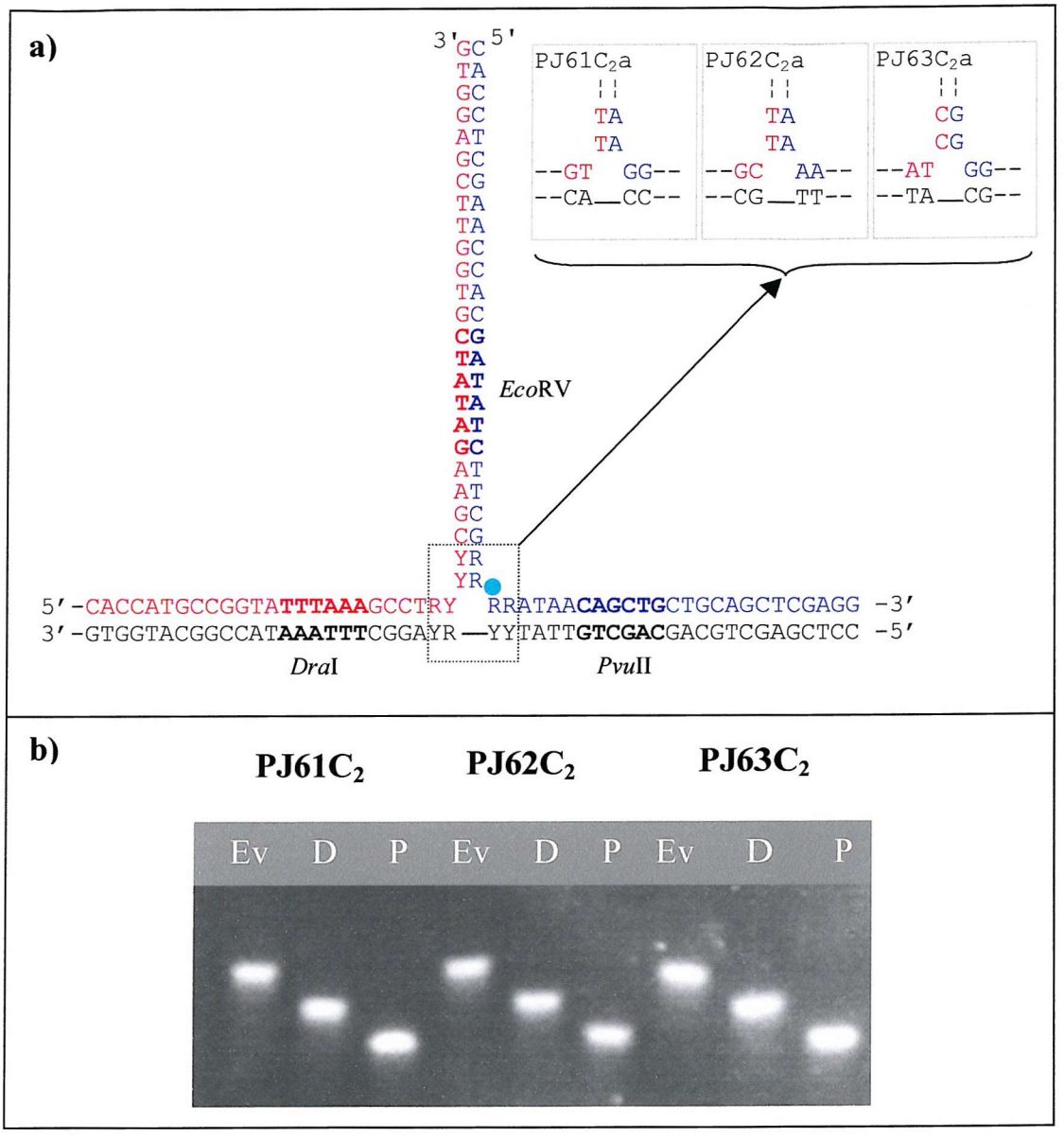


Fig. 5.15. Design and LSA results for junctions PJ61-63C_2a.

supports the R/Y stacking hypothesis, it does not support the hypothesis that junction structure controls RNA yield in SMART. However, it should be noted that the PJ63C_2a-equivalent SMART junction shows relatively high RNA yields, which may suggest that other factors in the less efficient junctions counteracted the effect of a beneficial conformation. Since these three junctions involve different 23S gene regions, differences in secondary structure of the target and/or the target-complementary sequences of the two probes could have affected the efficiency of junction formation.

5.4 Effect of slide flexibility of DNA conformation

5.4.1 The effect of slide on DNA conformation and theoretical assessment of the potential influence on TWJ conformation

The previous sections in this chapter have examined the R/Y-stacking hypothesis. However, this does not reveal much about the actual forces controlling coaxial stacking. The conformer conversion experiments on the other hand have shown that internal stacking interactions are probably not the only forces contributing to coaxial stacking as some junctions did not yield the expected conformer (PJ70A_2a, PJ74C_2a, PJ44C_2a, PJ45T_2a). This probably means that solvent and/or backbone dynamics must be taken into account, or sequences beyond the first two base pairs in each arm affect the equilibrium. Solvent-DNA interactions are notoriously difficult to assess both by experiment and by theoretical models. On the other hand, backbone dynamics are also controlled by stacking interactions. Packer & Hunter (1998) for instance, conducted an exhaustive theoretical analysis of the relation between stacking and backbone conformation using a large database of B-DNA crystal structures. They noted that some parameters (*e.g.* slide, twist) are strongly linked to the backbone. Dinucleotide steps that tend to show sliding motions include G/C steps, which (as indicated in Fig. 1.9c) usually do so to optimise electrostatic interactions. However, by sliding, these steps alter the backbone conformation, and these steps are therefore controlled by the stacking properties of neighbouring base pairs. For instance, if one pair tends towards positive slide and the adjoining step slides in the opposite direction, strain is exerted on the backbone. Such steps are therefore incompatible and a compromise must occur, resulting in a less than optimal stacking arrangement. Similar interactions may be of importance to

TWJ conformation as well, although the situation is will be more complicated due to the presence of the branch point. Table 5.2 provides an indication of the slide flexibility of the low and high-energy sequences studied here and by others, using data from Packer *et al.* (2000b). This data expresses the slide flexibility of the central dinucleotide step by taking the sequence context into account (*i.e.* they were based on tetranucleotide analysis). Interestingly, most low energy sequences appear as rigid, as judged from the high-energy values. Thus, this table provides evidence for the involvement of slide flexibility in junction formation.

Further support for the involvement of slide is already apparent at the dinucleotide level without considering sequence context. Table 5.3 lists the central dinucleotide sequences of the different junctions as R/Y sequences. Although this is based on a limited number of dinucleotide combinations, it is at least clear that where the choice is between RR/YY and YR/YR steps, RR/YY is almost always favoured (J1V3, PJ42C_2a, J1V5, PJ16C_2a, YTT, PJ62C_2a, TWJ2, PJ41C_2a, PJ51-55C_2a, PJ61C_2a, PJ63C_2a), except in PJ46T_2a. YR steps are more prone to slide deformations than RR/YY steps (Dickerson, 1999), probably due to reduced ring overlap, thus supporting the importance of slide flexibility in junction formation.

5.4.2 Experimental assessment of the effect of slide flexibility of the central coaxially stacked dinucleotide on TWJ folding

To test the effect of slide, six junctions were designed based on the tetranucleotide data obtained by Packer *et al.* (2000b). These junctions are illustrated in Fig. 5.15. PJ64G_2a and PJ65G_2a were designed such that a conformer I was expected on the basis of their flexibility values (*i.e.* they contained a low energy sequence with low slide propensity). However, PJ65G_2a constituted a special case in that its R/Y sequence was based on TWJ1 and J3CC (table 5.3) and designed such that based on the R/Y hypothesis, a conformer II was expected. PJ66C_2a was designed such that a conformer II was expected on the basis of slide flexibility. PJ67G_2a and PJ68G_2a contained the same sequences as PJ64G_2a and PJ65G_2a respectively but were "conformer converted", *i.e.* the base pairs in the template probe/target and extension probe/template probe arms had been swapped. PJ67G_2a and PJ68G_2a were therefore expected to become a conformer II junction on the basis of slide flexibility values (and PJ68G_2a a conformer I on the basis of R/Y sequence), and to show the opposite conformation from PJ64G_2a and PJ65G_2a respectively.

Table 5.2. Slide flexibility of the central dinucleotide high and low energy sequences of various junctions, taking the effect of sequence-context into account. The table is based on calculations by Packer *et al.* (2000b). LE = Low Energy sequence, HE = High Energy sequence, ΔFlex = Flexibility LE - Flexibility HE. Flexibility is expressed in kJ mol^{-1} $Å^{-2}$. Positive ΔFlex values (*i.e.* where the low energy sequence is less flexible than the high energy sequence) are printed bold. Please note that the conformation was ambiguous for junctions $PJ44C_2a$ and $PJ45T_2a$.
Colours indicate unique tetranucleotide R/Y sequences: YRRY/RYYR, YRRR/YYYR, YRYY/RRYR, YRYR/YRYR, YYRY/RYRR, YYRR/YYRR, YYYY/RRRR, RYRY/RYRY, RYYY/RRRY.

Low energy	*High energy*	*Junction*	*Flexibility LE*	*Flexibility HE*	*ΔFlex*	*Reference*
TGGC/GCCA	TGCG/CGCA	TWJ-TC	10.3	9.6	**0.7**	Leontis *et al.* 1994
CAAC/GTTG	CACG/CGTG	J1V6	17.5	4.1	**13.4**	Welch *et al.*, 1995
CGTC/GACG	CGAG/CTCG	J1	17.3	12.6	**4.7**	Welch *et al.*, 1995
CGTC/GACG	CGTG/CACG	J1V1	17.3	4.1	**13.2**	Welch *et al.*, 1995
CGTC/GACG	CGCG/CGCG	J1V2	17.3	10.0	**7.3**	Welch *et al.*, 1995
CCAC/GTGG	CCAG/CTGG	J1V4	7.9	9.1	-1.8	Welch *et al.*, 1995
TCAC/GTGA	TCCC/GGGA	$PJ46T_2a$	9.9	13.8	-3.9	This thesis
CCTC/GAGG	CCAG/CTGG	J1V3	13.6	9.1	**4.5**	Welch *et al.*, 1995
CCCC/GGGG	CCAG/CTGG	J1V5	16.4	9.1	**7.3**	Welch *et al.*, 1995
CTCC/GGAG	CTGC/GCAG	$PJ16C_2a$	13.3	7.2	**6.1**	This thesis
CCTG/CAGG	CCGT/ACGG	$PJ42A_2a$	10.8	7.3	**3.5**	This thesis
TGAG/CTCA	TGCC/GGCA	$PJ45T_2a$	14.5	12.1	**2.4**	This thesis
GCCG/CGGC	GCGT/ACGC	TWJ2	14.5	11.7	**2.8**	Overmars *et al.*, 1996
GTTA/TAAC	GTGA/TCAC	$PJ55C_2a$	20.7	9.9	**10.8**	This thesis
GGCG/CGCC	GGGT/ACCC	TWJ1	8.6	18.3	-9.7	Overmars *et al.*, 1996
GACG/CGTC	GAGC/GCTC	J3CC	17.3	11.9	**5.4**	Rosen & Patel, 1993
CGTT/AACG	CGGA/TCCG	$PJ40C_2a$	14.9	7.9	**7**	This thesis
TACC/GGTA	TAAG/CTTA	$PJ44C_2a$	8.8	19.8	-11	This thesis
CTTT/AAAG	CTGT/ACAG	$PJ41C_2a$	24.5	11.2	**13.3**	This thesis
CTTC/GAAG	CTGC/GCAG	$PJ51C_2a$	24.9	7.2	**17.7**	This thesis
TTTC/GAAA	TTGT/ACAA	$PJ52C_2a$	19.3	9.3	**10.0**	This thesis
TTTT/AAAA	TTGC/GCAA	$PJ53C_2a$	23.8	8.7	**15.1**	This thesis
TTTC/GAAA	TTGC/GCAA	$PJ54C_2a$	19.3	8.7	**10.6**	This thesis
TCTG/CAGA	TCAC/GTGA	YTT	13.5	9.9	**3.6**	Zhong *et al.*,1994
GTTT/AAAC	GTGG/CCAC	$PJ61C_2a$	27.2	7.9	**19.3**	This thesis
GCTT/AAGC	GCAA/TTGC	$PJ62C_2a$	14.3	8.7	**5.6**	This thesis
ATCC/GGAT	ATGG/CCAT	$PJ63C_2a$	11.4	11.7	-0.3	This thesis

Table 5.3. High and low energy sequences of junctions studied previously and in this thesis. All junctions contained two unpaired bases at the branch point. LE-DS = Low energy dinucleotide step, HE-DS = High energy dinucleotide step. Each colour shade represents a specific tetranucleotide R/Y sequence (refer to table 6.2). The line in the table separates the conformer I (upper part) and conformer II (lower part) junctions. The third and fourth column shows the central dinucleotide of the low energy and high energy sequence respectively. Columns five and six show the same dinucleotides but as purines and pyrimidines.

Low Energy	*High Energy*	*Centre LE-DS*	*Centre HE-DS*	*Centre LE-DS*	*Centre HE-DS*	*Junction*	*References*
TGGC/GCCA	TGCG/CGCA	GG/CC	GC/GC	RR/YY	RY/RY	TWJ-TC	Leontis *et al.* (1994)
CAAC/GTTG	CACG/CGTG	AA/TT	AC/GT	RR/YY	RY/RY	J1V6	Welch *et al.* (1995)
CGTC/GACG	CGAG/CTCG	GT/AC	GA/TC	RY/RY	RR/YY	J1	Welch *et al.* (1995)
CGTC/GACG	CGTG/CACG	GT/AC	GT/AC	RY/RY	RY/RY	J1V1	Welch *et al.* (1995)
CGTC/GACG	CGCG/CGCG	GT/AC	GC/GC	RY/RY	RY/RY	J1V2	Welch *et al.* (1995)
CCAC/GTGG	CCAG/CTGG	CA/TG	CA/TG	YR/YR	YR/YR	J1V4	Welch *et al.* (1995)
TCAC/GTGA	TCCC/GGGA	CA/TG	CC/GG	YR/YR	RR/RR	PJ46T_2a	This thesis
CCTC/GAGG	CCAG/CTGG	CT/AG	CA/TG	RR/YY	YR/YR	J1V3	Welch *et al.* (1995)
CCCC/GGGG	CCAG/CTGG	CC/GG	CA/TG	RR/YY	YR/YR	J1V5	Welch *et al.* (1995)
CTCC/GGAG	CTGC/GCAG	TC/GA	TG/CA	RR/YY	YR/YR	PJ16C_2a	This thesis
CCTG/CAGG	CCGT/ACGG	CT/AG	CG/CG	RR/YY	YR/YR	PJ42	This thesis
TGAG/CTCA	TGCC/GGCA	GA/TC	GC/GC	RR/YY	RY/RY	PJ45T_2a	This thesis
GCCG/CGGC	GCGT/ACGC	CC/GG	CG/CG	RR/YY	YR/YR	TWJ2	Overmars *et al.* (1995)
GTTA/TAAC	GTGA/TCAC	TT/AA	TG/CA	RR/YY	YR/YR	PJ55C_2a	This thesis
GGCG/CGCC	GGGT/ACCC	GC/GC	GG/CC	RY/RY	RR/YY	TWJ1	Overmars *et al.* (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	This thesis
TACC/GGTA	TAAG/CTTA	AC/GT	AA/TT	RY/RY	RR/YY	PJ44C_2a	This thesis
CTTT/AAAG	CTGT/ACAG	TT/AA	TG/CA	RR/YY	YR/YR	PJ41C_2a	This thesis
CTTC/GAAG	CTGC/GCAG	TT/AA	TG/CA	RR/YY	YR/YR	PJ51C_2a	This thesis
TTTC/GAAA	TTGT/ACAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ52C_2a	This thesis
TTTT/AAAA	TTGC/GCAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ53C_2a	This thesis
TTTC/GAAA	TTGC/GCAA	TT/AA	TG/CA	RR/YY	YR/YR	PJ54C_2a	This thesis
TCTG/CAGA	TCAC/GTGA	CT/AG	CA/TG	RR/YY	YR/YR	YTT	Zhong *et al.* (1994)
GCTT/AAGC	GCAA/TTGC	CT/AG	CA/TG	RR/YY	YR/YR	PJ62	This thesis
GTTT/AAAC	GTGG/CCAC	TT/AA	TG/CA	RR/YY	YR/YR	PJ61	This thesis
ATCC/GGAT	ATGG/CCAT	TC/GA	TG/CA	RR/YY	YR/YR	PJ63	This thesis

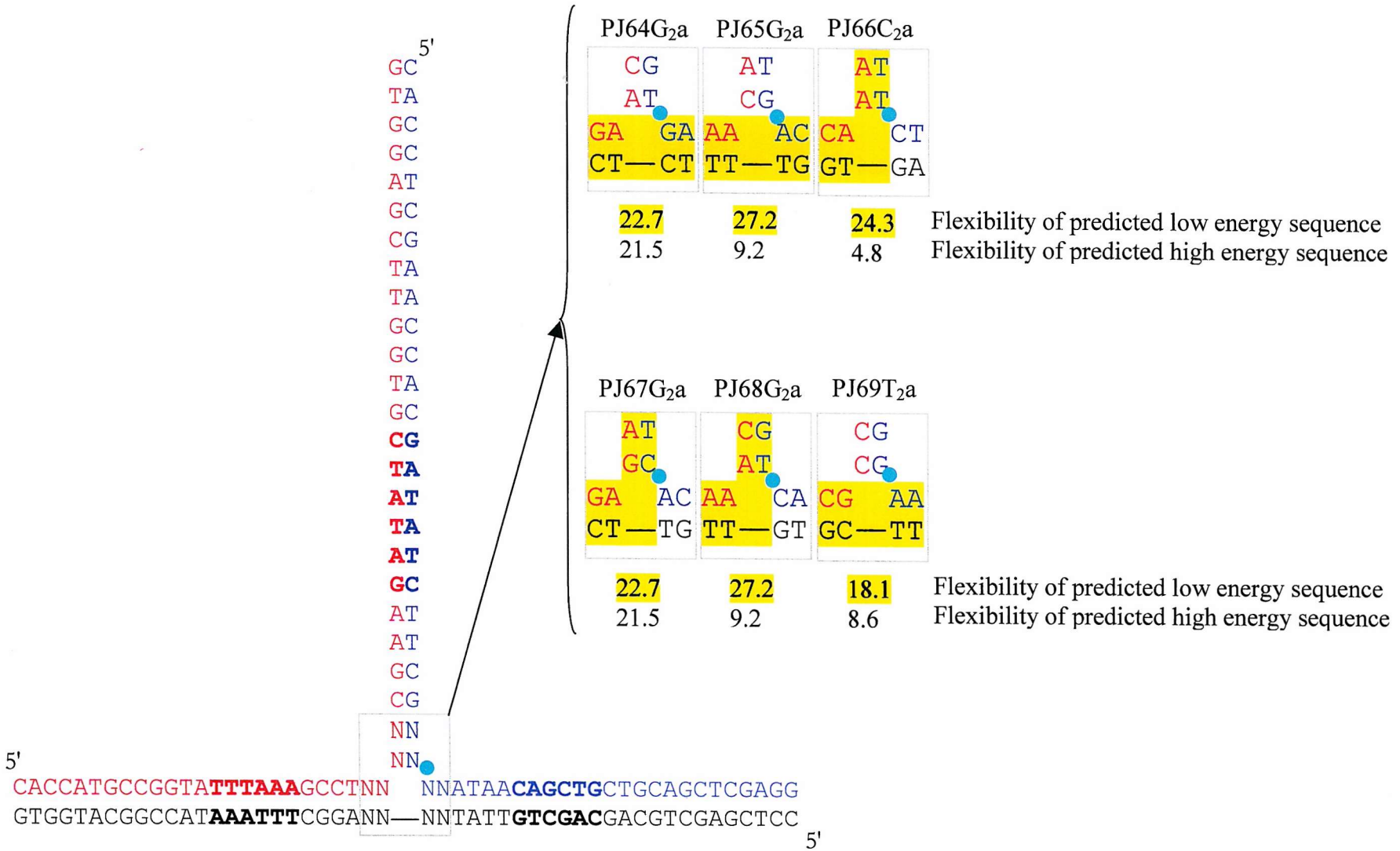


Fig. 5.16. Design of PJ64-69. The branch point sequences were chosen based on slide flexibility (table 6.2) of tetranucleotides. The hypothesis stated that tetranucleotide sequences with lower flexibility would be favoured over sequences with high flexibility with respect to coaxial stacking. The sequences presented here were chosen on that basis, with the lower flexibility sequences shaded yellow (predicted coaxial stacking, or low energy sequence). PJ67-69 are the same as PJ64-66 respectively but have the base pairs swapped as in the conformer conversion experiment. PJ65 contains the same R/Y sequence as PJ68, not the actual base sequence. Please refer to the text for a more detailed discussion on the design of PJ65 and PJ68. Flexibility is expressed as an energy term in kJ mol^{-1} $Å^{-2}$, and higher values reflect lower flexibility. The location of the unpaired bases is indicated by the green circle.

Finally, PJ69T_2a was, on the basis of slide flexibility, expected to form a conformer I.

The results for these six junctions are shown in Fig. 5.17. Highly unusual patterns are apparent for most junctions, except PJ69T_2a. The pattern for PJ64G_2a appears to correspond to the unfolded form as judged from the observation that the *Dra*I fragment migrated fastest, which is similar to observations for PJ44C_2a and PJ45T_2a. Another possibility would be that the junction was in fact a conformer II but with the coaxially stacked arms strongly bend as shown in Fig. 5.17. However, this seems like an unfavourable interaction due to the likelihood of stacking disruption at the branch, unless the bend was spread over a number of neighbouring base pair steps. The proposition of an unfolded form is problematic as well, since it suggests a lack of stabilising coaxial stacking interactions. Nevertheless, the unfolded form has been proposed as a transient intermediate in the conversion from one conformer to the other for FWJs where a structural equilibrium has firmly been established (Overmars & Altona, 1997). If a structural equilibrium underlies the observations for PJ64G_2a, this would mean that this transient intermediate should exist sufficiently long to produce the pattern observed. This could indicate that for PJ64G_2a the unfolded transition form constitutes a very shallow energy barrier between the two conformers and/or that the two conformers both were of relatively high energy (low stability). Both situations could lead to an apparent increase in the lifetime of the transient, unfolded intermediate, which could lead to the pattern observed in Fig. 5.17. This suggestion is similar to the one offered by Welch *et al.* (1995) who observed similar mobility patterns at moderate (50-200μM) magnesium concentrations for a number of junctions. However, magnesium concentration is probably not an issue here as all experiments were carried out in the presence of at least 5 mM $MgCl_2$ although a requirement for unusually high concentrations cannot be ruled out. The *Eco*RV fragment of PJ64G_2a further appeared as the slowest fragment (due to the short electrophoresis run this is not as clear in Fig. 5.17), thus suggesting that a conformer II conformation slightly dominated the equilibrium.

For PJ65G_2a a structural equilibrium biased towards a conformer II seems even more likely as judged from the observation that the *Eco*RV fragment was clearly the slowest fragment and the difference with the *Pvu*II fragment was increased relative to PJ64G_2a. Nevertheless, the *Pvu*II fragment would be expected to migrate faster

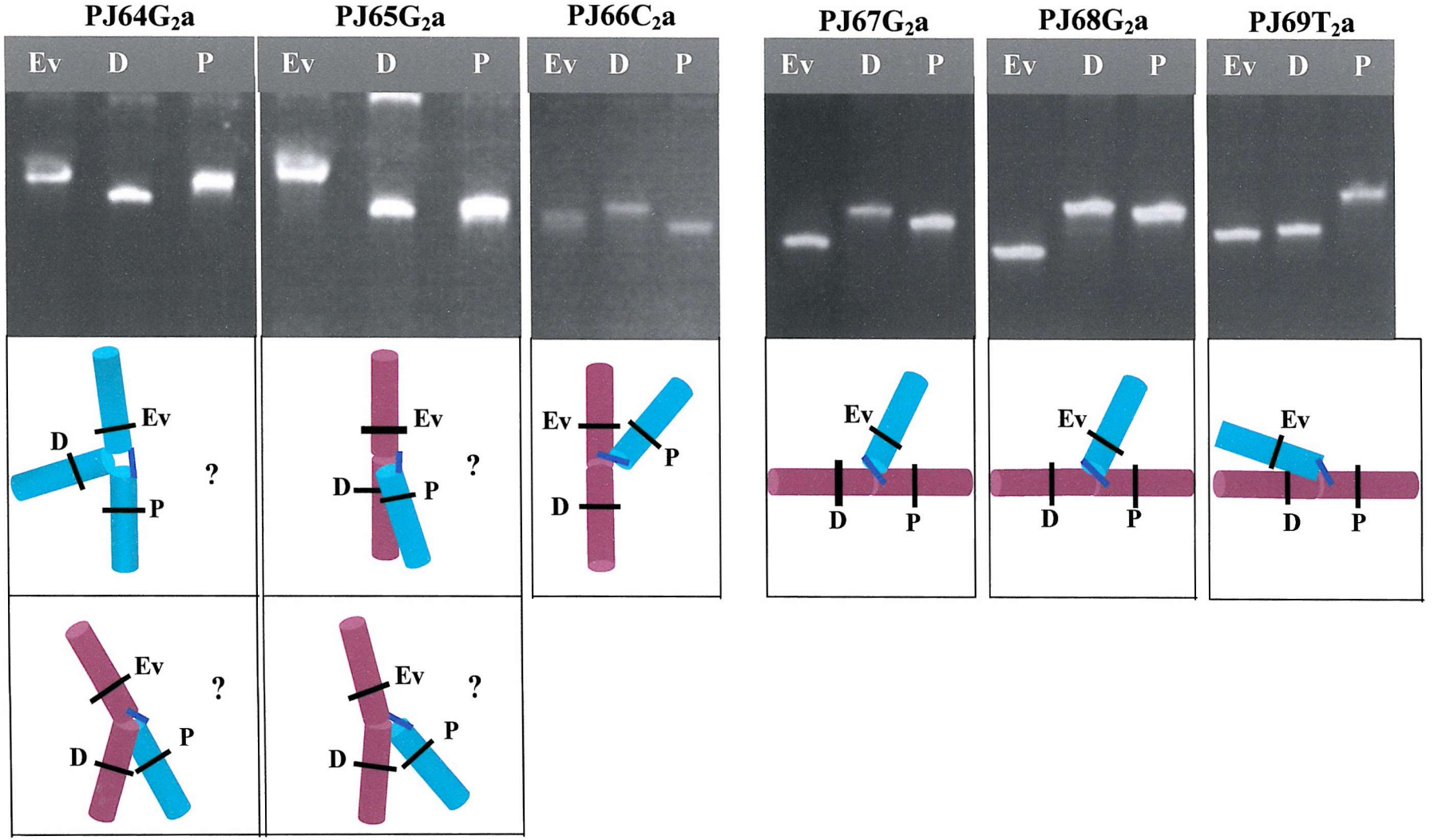


Fig. 5.17. LSA results for PJ64-69. Except for PJ69, the mobility patterns obtained do not conform to conventional mobility patterns, indicating that these junctions resided in an unusual conformation. Refer to the text for more details. Unstacked arms are drawn in light green, whereas coaxially stacked arms are drawn in dark red. Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II.

than the *Dra*I fragment in case of a conformer II. It is possible that if a dynamic structural equilibrium existed, a longer residence time in the conformer I state could reduce the mobility of the *Pvu*II fragment, explaining the somewhat unusual appearance of this fragment. As indicated in Fig. 5.17, an alternative explanation could be that a conformer II junction was indeed formed but one where the coaxial stack was slightly bend, which would also reduce the mobility of the *Pvu*II generated fragment but for reasons outlined above, this seems energetically unfavourable.

PJ66T_2a presents a problem in that the mobility pattern observed does not correspond to a conventional conformer I or II. It also cannot be explained on the basis of an equilibrium where the different states are rapidly interconverting. In stead, it suggests a novel variant of the conformer II as illustrated in Fig. 5.17. The fact that *Dra*I migrates slowest indicates that the extension probe/template probe arm formed a small angle with the template probe/target arm. *Pvu*II yields the fastest fragment suggesting that the junction adopted a conformer II-type coaxial stacking arrangement. However, the orientation of the unstacked arm appears highly unusual and doubts could be raised whether this conformation is actually feasible. Some experimental support for the hypothesised conformation does exist however. Ulyanov *et al.* (1998) investigated an unusual DNA conformation, named a pseudosquare knot by NMR. As illustrated in figure 5.18 this knot contained two (identical) three-way DNA junction domains. These domains are in fact, coaxially stacked as a conformer I, but show a reversed orientation of the unstacked arm relative to the coaxially stacked arms. We therefore tentatively name this conformation a "reverse conformer I". The NMR structure presented in Fig. 5.18 thus indicates that in principle this reverse conformer I is possible. Since there are no apparent arguments against the proposition that a reverse conformer II is also possible, this is the conformation proposed for PJ66G_2a. Surprisingly, PJ67G_2a and PJ68G_2a show a mobility pattern corresponding to a reverse conformer I, providing further support for the existence of this conformation. Interestingly, in accepting the possibility of a reverse conformer, the likelihood is raised that PJ42C_2a (Fig. 5.4) also forms a reverse conformer II. Nevertheless, for PJ42C_2a we suggested that the unusual pattern resulted from branch migration, which is not applicable to any of the junctions described in this section since no complementary bases to the unpaired bases are present at the branch points.

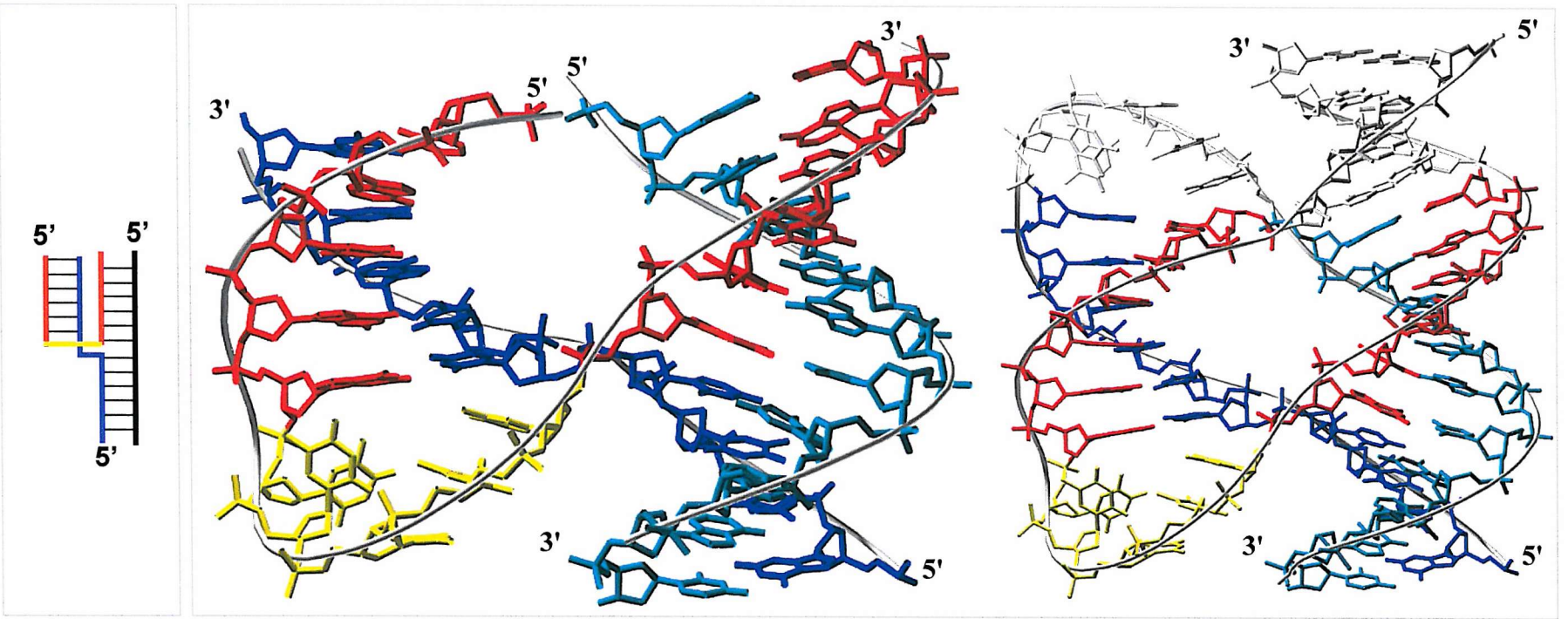


Fig 5.18. NMR structure of one of the two TWJs present in a DNA pseudosquare knot (Ulyanov *et al.*, 1998, PDB-entry: 1SLS, model 1). The TWJ is shown separately on the left and is shown as part of the pseudosquare knot on the right. Unpaired bases are coloured yellow (part of the red coloured strand). The schematic representation of the TWJ (with the green strand now coloured black) is shown to the left (compare this figure with the conformer I TWJ drawn in Fig. 1.8b) Refer to the text for further details.

An unambiguous result was obtained for PJ69T_2a, suggestive of a conformer I.

Finally, the conformer conversion experiments appeared to work for PJ67-69 as all three showed (reverse) conformer I patterns whereas a conformer II biased conformation for PJ64-66 fits the LSA data for these junctions. This supports the hypothesis that the two conformers share the same stacking preferences.

5.5 Discussion

The data presented in this chapter provide an indication that junction structure has a substantial influence on the SMART signal. The R/Y hypothesis proposed here provides reliable predictions of TWJ structure based solely on its sequence. Evidence has been obtained, that the purine/pyrimidine distribution around the branch point plays a prominent role in junction formation. However, not all the sequences tested support this hypothesis (e.g. PJ44C_2a, PJ45T_2a and PJ46T_2a). This may not be surprising since it has been shown that in B-DNA certain sequences adopt unique geometries that are likely to be caused by base-specific differences (Packer *et al.*, 2000a and 2000b). It is important to note that the junctions designed based on the slide flexibility values presented in table 5.2 (PJ64-69, Fig. 5.16) provide an uncertain picture on the effect of slide on TWJ structure. Assuming PJ64G_2a and PJ65G_2a are biased towards a conformer II, it seems that the low energy sequences of these junctions were the most flexible rather than the most rigid, in contrast to expectation. The same was observed for PJ67G_2a and PJ68G_2a (based on PJ64G_2a and PJ65G_2a respectively via conformer conversion). For PJ65G_2a and PJ68G_2a it appears therefore, that the R/Y stacking hypothesis allows for better prediction of structure. Since R/Y sequence and slide flexibility are only partially related, this may indicate that, at least for the sequences of these junctions, other base step parameters contribute to junction stacking. On the other hand, PJ66C_2a and PJ69T_2a did show a conformer II where the most rigid dinucleotide became coaxially stacked. Taken together, these experiments suggest that slide is important in coaxial stacking interactions but probably not the only parameter involved.

The question how slide could affect junction formation remains speculative but could be related to the way in which slide affects backbone conformation. Packer & Hunter (1998) established that slide is coupled mainly to the χ, δ, ζ and P torsion

angles of the DNA backbone. Although only a limited number of high-resolution structures of FWJs and TWJs are available, they appear to indicate that the exchanging strands are characterised by unusual χ and ε torsions (for a detailed comparison between B-DNA, FWJs and TWJs, see Fig. 5.2). Since slide affects χ, it is possible that base steps with a strong propensity for sliding limit the conformational freedom of the backbone in establishing the appropriate cross-over conformation at the branch point.

It is important to realise that most experiments described in this chapter, tested the influence of only the first two base pairs in each arm. It is quite possible that the identity of base pairs located further away also affect the conformational equilibrium, which could explain why certain junctions behaved differently than expected based on the proposed hypotheses. Packer *et al.* (2000b) noted that dinucleotides could be classified as context-independent, weakly context-dependent and strongly context-dependent. The latter included CG/CG, GC/GC and CC/GG dinucleotides, which, for instance, could explain the different results obtained for PJ40C_2a, PJ44C_2a and PJ45T_2a. These junctions contained the same R/Y sequences around the branch point (the first two base pairs in each arm), but contained GG and GC dinucleotides in different locations. PJ40C_2a for instance, contained two A/T pairs in the extension probe/template probe arm, which were changed to two C/G pairs in PJ44C_2a and PJ45T_2a (Fig. 5.11). It is therefore possible that the sequence in this arm at the third position or even further away from the branch point, affected the conformer distribution differently in the different junctions.

The results in this chapter have provided further evidence that most but not all conformer I and II junctions share the same stacking preferences, as judged from the conformer conversion experiments. The observation that not all junctions showed complete conversion suggest that the solvent and/or backbone dynamics contributed to the conformation, and/or that sequences beyond the first two base pairs in each arm contributed to TWJ conformation. Importantly, all conformer conversion experiments used for assessment of the TWJ-structure/RNA yield relationship confirmed that the predicted conformer II junctions yielded more RNA (even if an ambiguous structure was apparent on gel). These results are not only important for the design of SMART junctions but also have implications for DNA structure in general. The finding of structure consensus sequences allows comparison with their

structure behaviour in a normal B-DNA setting, which ultimately may improve the understanding of base-stacking interactions and DNA-solvent interactions in general.

Finally, this chapter contains evidence for a novel TWJ conformation, tentatively called a "reverse conformer" to highlight the unusual orientation of the unstacked arm relative to the stacked arm.

5.5.1 A case for a dynamic conformational equilibrium in TWJ folding

Most of the anomalous results presented in this chapter (*i.e.* those showing an unusual mobility pattern, different from that expected for conformer I or II) can be explained on the basis of a structural equilibrium. This is schematically summarised in Fig. 5.19 and is based on the assumption that the unfolded form constitutes the conformation of lowest stability and acts as a transient intermediate in the conversion from conformer I to II as formulated by Overmars & Altona (1997) for FWJs. There are two extreme possibilities: either one or both of the conformers show a substantial difference in free energy from the unfolded form or the unfolded and folded states have very similar free energies. If the energy difference is large, again two possibilities can be postulated but both are characterised by a short lifetime of the unfolded form. One possibility is that both conformers have (approximately) the same free energy. This could explain mobility patterns where no difference in fragment mobility is observed. The other possibility is that one conformer is of significantly lower energy. In this case, a clear conformer I, II or reverse conformer pattern will emerge. On the other hand, if the energy difference between the two conformers and the unfolded form is small, it is likely that the lifetime of the unfolded form increases due to increased interconversion. This would bias the resulting mobility patterns towards that of the unfolded form, even if a conformer preference exists. These are the mobility patterns observed for PJ64G_2a and PJ65G_2a but also PJ44C_2a, PJ45T_2a and PJ71C_2a. A structural equilibrium fits these junctions well, since if a static junction is invoked the patterns suggest a conformation in which the unstacked arm must lie at an angle smaller than 90°. This would probably strain the DNA backbone at the branch point, unless other distortions such as loss of base pairing at the branch point would compensate for this. The latter would however, reduce the stability of the junction and does not appear to be an energetically favourable situation. A biased conformational equilibrium with an

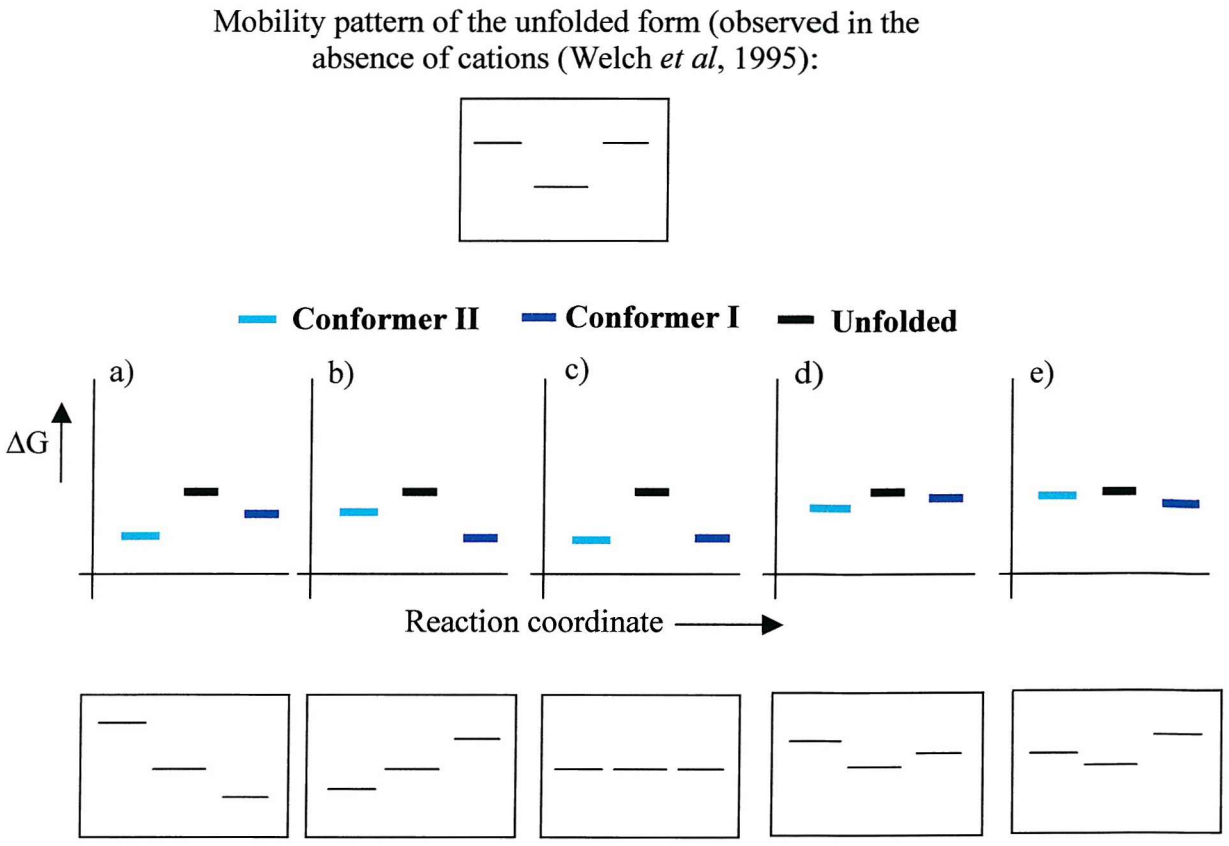


Fig. 5.19. Possible explanation for the observed mobility patterns in this thesis, based on a dynamic conformational equilibrium. The top graphs show examples of hypothetical differences in ΔG between the different conformational states, with the expected gel mobility pattern shown directly below each graph. In case large energy differences exist between the unfolded state and the folded states, the three gels shown to the left may arise (a-c), depending on the difference in energy between the two conformers. The transient unfolded form is short lived and is therefore not reflected in the mobility patterns. However, if the energy difference between the unfolded and folded forms is small as shown in d) and e), the lifetime of the unfolded form may increase since interconversion occurs more often. On a gel, the mobility pattern is then more likely to reflect the unfolded form (which for clarity is reproduced at the top). Schematic gels are shown as used in this thesis (following the order *Eco*RV - *Dra*I - *Pvu*II)

increased apparent lifetime for the unfolded form provides an energetically more appealing solution.

The previous suggestions can also be applied to the results obtained for junctions PJ70-75. The LSA data obtained for PJ70A_2a and PJ74C_2a could be interpreted as reflecting an increased lifetime of the conformer II conformation when compared to the conformer converted counterparts of these junctions (PJ71A_2a and PJ75C_2a respectively). Even so, a conformer I still appears to dominate the equilibrium as judged from the observation that the *Eco*RV fragments migrated faster than the *Pvu*II fragments for PJ70A_2a and PJ74C_2a (Fig. 5.10a). This means that the conformer conversion experiment did not lead to the expected conformer preference and thus, forces other than stacking interactions control the conformer equilibrium. Nonetheless, a dynamic conformational equilibrium explains the higher yields observed for the proposed conformer II junctions. The increased life-time of the conformer II state (or a more rapid interconversion rate) could increase accessibility and binding by the DNA polymerase, leading to higher RNA yields. It is interesting to note that no significant difference in yield was apparent between PJ72C_2a, which showed a clear conformer II mobility pattern, and junctions PJ70A_2a and PJ74C_2a, both which showed an unfolded mobility pattern. This supports the existence of a conformer II state for PJ70A_2a and PJ74C_2a since all three yielded similar amounts of RNA, and suggests that only a marginal increase in conformer II life-time may be required for efficient DNA polymerase binding.

5.6 Concluding remarks

The experiments described in this chapter used a novel approach to improve the understanding of TWJ folding with the aim to improve junction design in SMART. Previous studies that have assessed the sequence-structure relationship in junction folding adapted a non-systematic approach, *i.e.* random sequence alterations were probed on their effects on junction folding (Lu *et al.*, 1991, Welch *et al.*, 1995). One exception concerns the NMR study by Overmars *et al.* (1996) who assessed whether R/Y stacking rules applicable to B-DNA were also applicable to TWJ folding. However, this study only probed the first base pair of each arm at the branch point and ignored the potential influence of sequence-context and as such did not reveal the clear consensus found here, that R/Y sequence in many cases appears an

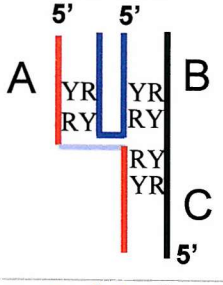
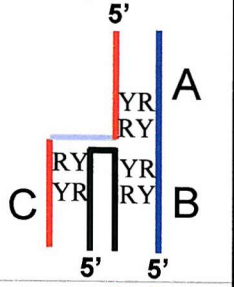
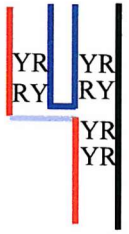
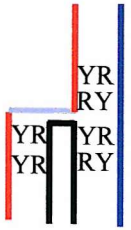
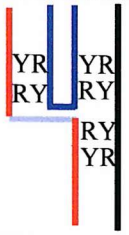
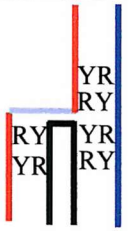
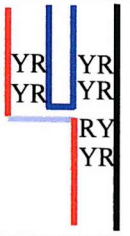
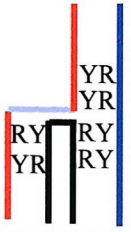
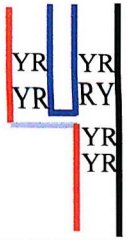
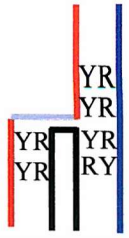
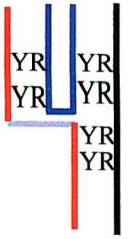
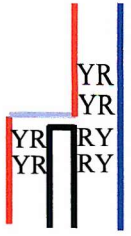
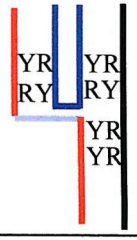
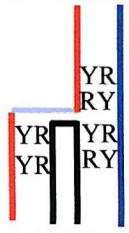
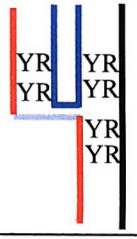
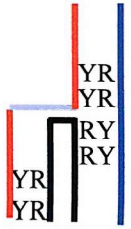
important factor in junction folding. The approach used in this chapter entailed a much more rigorous analysis and comparison with available crystallographic data and theoretical analysis performed on numerous B-DNA structures than other studies have done so far. Since TWJ conformation shares many similarities with B-DNA, rules that dictate B-DNA conformation were postulated to apply to TWJs as well, including the similar stacking behaviour of purines and pyrimidines, and a constraining influence by the DNA backbone on the number of conformations that can be adopted by a given base pair. Evidence was presented showing that structure prediction of TWJs based on sequence was most reliable when this was based on the R/Y stacking hypothesis. The latter illustrates the importance of base-base stacking interactions in TWJ folding. Nevertheless, the conformer conversion experiments indicated that most but not all junctions share the same stacking preferences between the two possible conformers. This shows that stacking interactions, although important, are probably not the only factors involved in dictating TWJ conformation. Instead, solvent-DNA or backbone conformation must be considered as potential factors affecting junction structure as well as the influence of sequences beyond the first two base pairs at the branch point. Together these studies reveal that junction folding is much more complex than hitherto assumed and have exposed some of the rules that underlie the sequence-dependency of this process.

From the viewpoint of SMART design, these results allow, based on the DNA sequence, better predictions to be made on the conformation of TWJs. That this is important is illustrated by the observations that the TWJ conformation can have dramatic effects on the RNA yield.

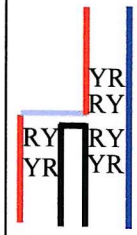
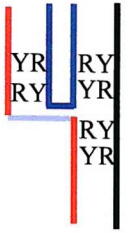
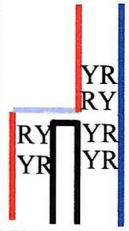
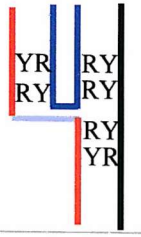
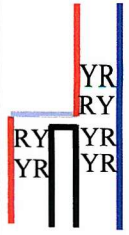
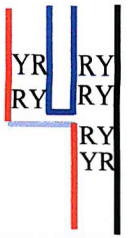
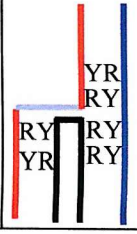
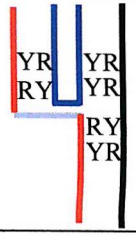
Appendix

Table A5.1. The junctions below illustrate the preferred and rejected conformations of the junctions listed in table 5.1. The low-energy sequences listed in table 5.1 are equivalent to the coaxial stacks shown here in the preferred conformation columns. The high-energy sequences listed in table 5.1 are equivalent to the coaxial stacks shown in the rejected conformation columns. Junctions are drawn as in Fig. 1.8.

Conformer I

Junction	Preferred	Rejected	Junction	Preferred	Rejected
TWJ-TC			**J1V2**		
J1V6			**J1V4**		
J1			**J1V3**		
J1V1			**J1V5**		

Conformer II

Junction	Preferred	Rejected
TWJ2		
TWJ1		
J3CC		
YTT		

6 Effect of non-nucleosidic linkers on three-way DNA junction conformation and interaction with restriction endonucleases

6.1 Introduction

The conformational analysis of three-way DNA junctions presented in this thesis, has focused on junctions containing unpaired bases at the branch point to facilitate and control coaxial stacking. However, as outlined in chapters 1-3, most junctions used in SMART contain synthetic non-nucleosidic linkers such as hexaethylene glycol or octanediol at this position. An important question that therefore remains to be resolved is whether these linkers act in a similar fashion as unpaired bases. It is possible that the different chemical structures of these linkers have a different effect on junction structure, and in the most extreme case affect the coaxial stacking interactions. This chapter will therefore compare junctions containing unpaired bases or chemical linkers to assess the effect of linker composition on junction structure.

6.2 Effect of deletions at the branch point and introducing synthetic linkers in a PJ1-derived junction.

Chapter 4 discussed the conformation of a junction (PJ1C_2a, section 4.2) that was based on the CFTR target. To assess the effect of different linkers on the conformation of the junction, the two unpaired cytosines were replaced with either two hexaethylene glycol (HEG) moieties (PJ36H_2a), two octanediol linkers (PJ36Oct_2a) or two abasic site analogues (PJ36A$basic_2$a) (Fig. 6.1). It should be noted that PJ36 and PJ1 also differ in a few details. Firstly, sites for blunt cleaving R-ENases were used (for reasons discussed in chapter 4) and secondly, a subtle modification of the sequence in the overlap arm was introduced. The latter was a CG change to TA, thereby retaining the same R/Y sequences. The reason for this change was that PJ1 represented an early, unoptimised junction, with respect to the overlap region. Experiments carried out by others at Cytocell Ltd. indicated that the 8 bp overlap as used in J-LV1A (chapter three) improved the RNA yield. In addition, as

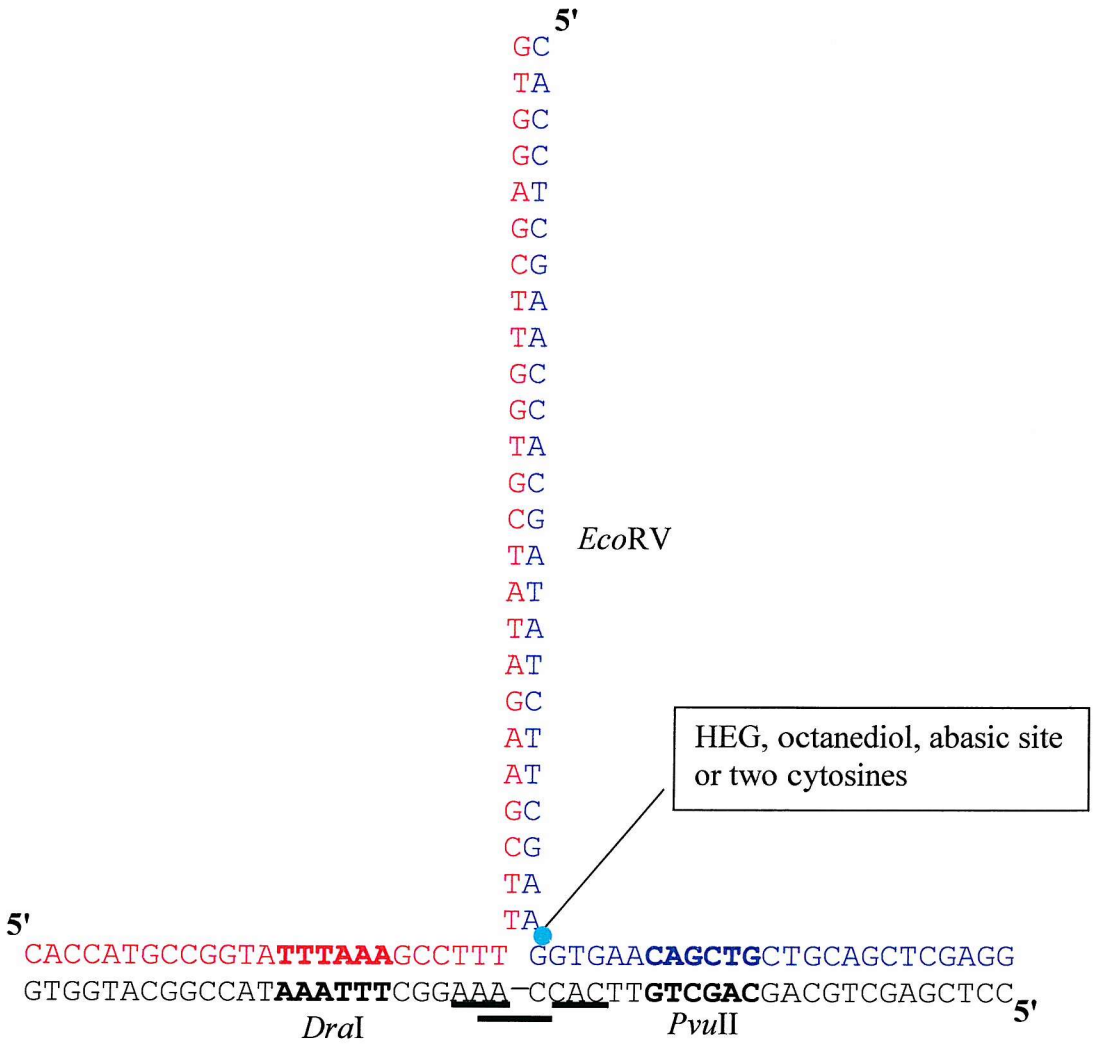


Fig. 6.1. Design of PJ36 containing different linkers. The deletions tested are underlined.

discussed below (section 6.3), inclusion of HEG linkers in PJ1 yielded anomalous results.

As in chapter 3, different target deletions were assessed. These allowed a more comprehensive analysis of the effect of chemical linkers on the junction conformation. Fig. 6.2 shows the effect of different linkers on the conformation of PJ36 junctions based on the wild-type or deletion containing CFTR gene. The figure illustrates that virtually all junctions showed the same conformation regardless of which linker was included at the branch point. The only exception is found for the junction containing a CAC deletion (second row in Fig. 6.2) where the double cytosine-containing junction showed a slightly different mobility pattern to junctions containing a non-nucleosidic linker. However, the HEG, octanediol and abasic site derivatives of this junction showed the same mobility pattern. It is interesting that the ΔCAC based junctions are the only junctions that show a difference in conformation. The ΔAAA and ΔAAC junctions show the same conformation as the wild-type junctions, regardless of linker identity. Yet, the ΔCAC junction shows a different mobility when two unpaired cytosines are present and again a different mobility pattern when a non-nucleosidic linker is present. The reason for this remains unclear, but suggests that in some cases these linkers are able to influence the conformation of TWJs differently than unpaired bases.

6.3 Unconventional effect of non-nucleosidic linkers on the cleavage activity of *Mse*I

PJ1 (refer to Fig. 4.1 for the sequence) is a junction based on J-CF1A, and its conformation was assessed in section 4.2. To establish the effect of HEG on the conformation of PJ1, HEG moieties were inserted in various strands but always at the branch point. The designs of the junctions used (with the HEG represented by the red-filled circles) and the LSA results obtained for these junctions are shown in Fig. 6.3. Comparing Fig. 4.3 (and Fig. 6.2) with Fig. 6.3 an interesting difference in PJ1 conformation can be observed. Fig. 6.3 clearly shows that introducing one or two HEG molecules in the template probe strand results in a conformer I. Although this conformation was also apparent for PJ1 lacking a linker, it is clearly different from the structures formed in the presence of one or more unpaired cytosines. These results therefore suggest that HEG has a different effect on the TWJ conformation than

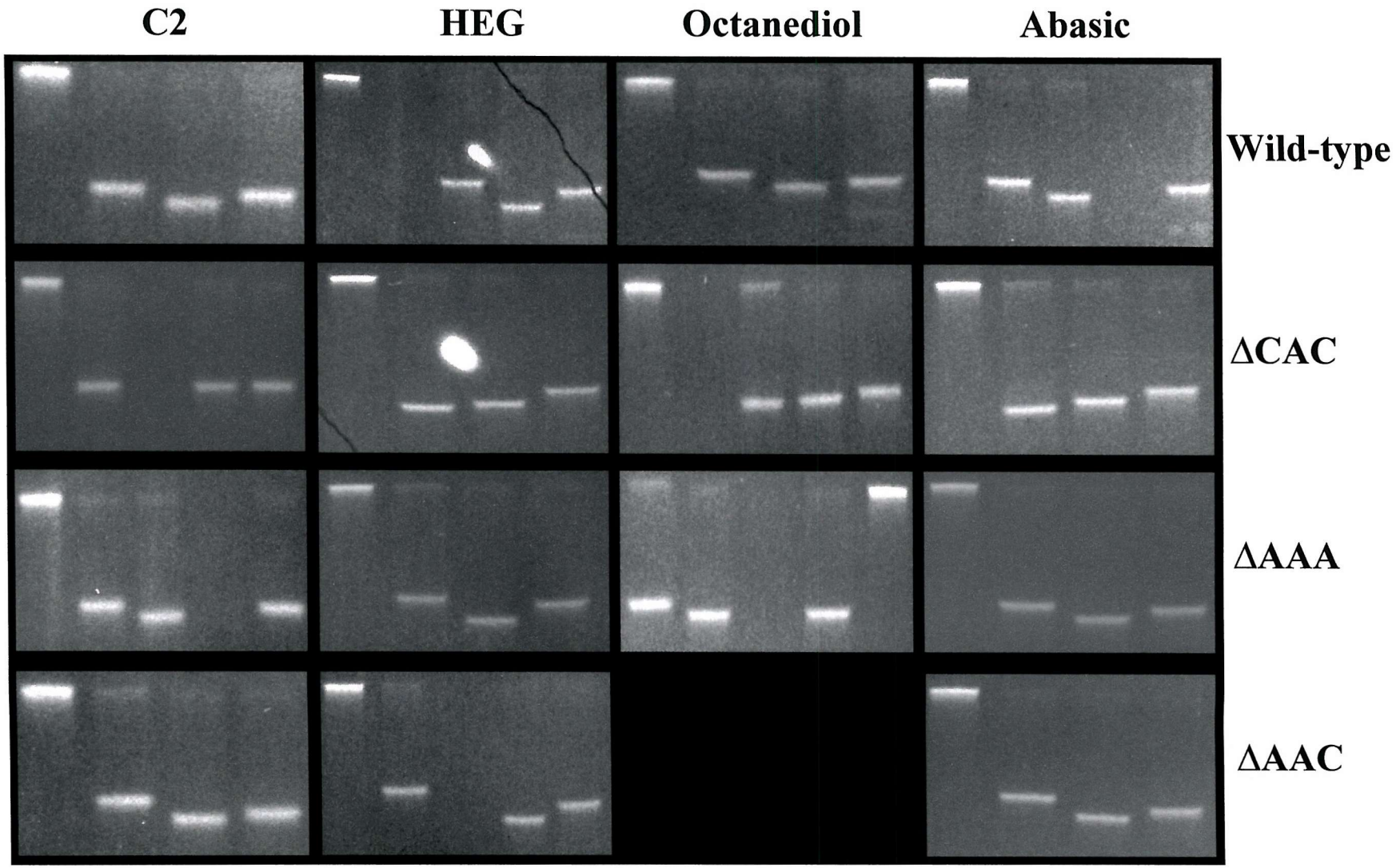


Fig. 6.2. Effect of different deletions on the TWJ structure of PJ36 (CFTR-derived junction). The junctions are sorted left to right on the type of linker present at the branch point and from top to bottom on the type of target. The gel for PJ36Oct$_2$a with the ΔAAC target failed and was not repeated. Each gel follows the order: no R-ENase, *Eco*RV, *Dra*I, *Pvu*II, except where the octanediol linker and ΔAAA deletion were tested (3rd column, 3rd row) For the latter, the order was: *Eco*RV, *Dra*I, *Pvu*II, no R-ENase.

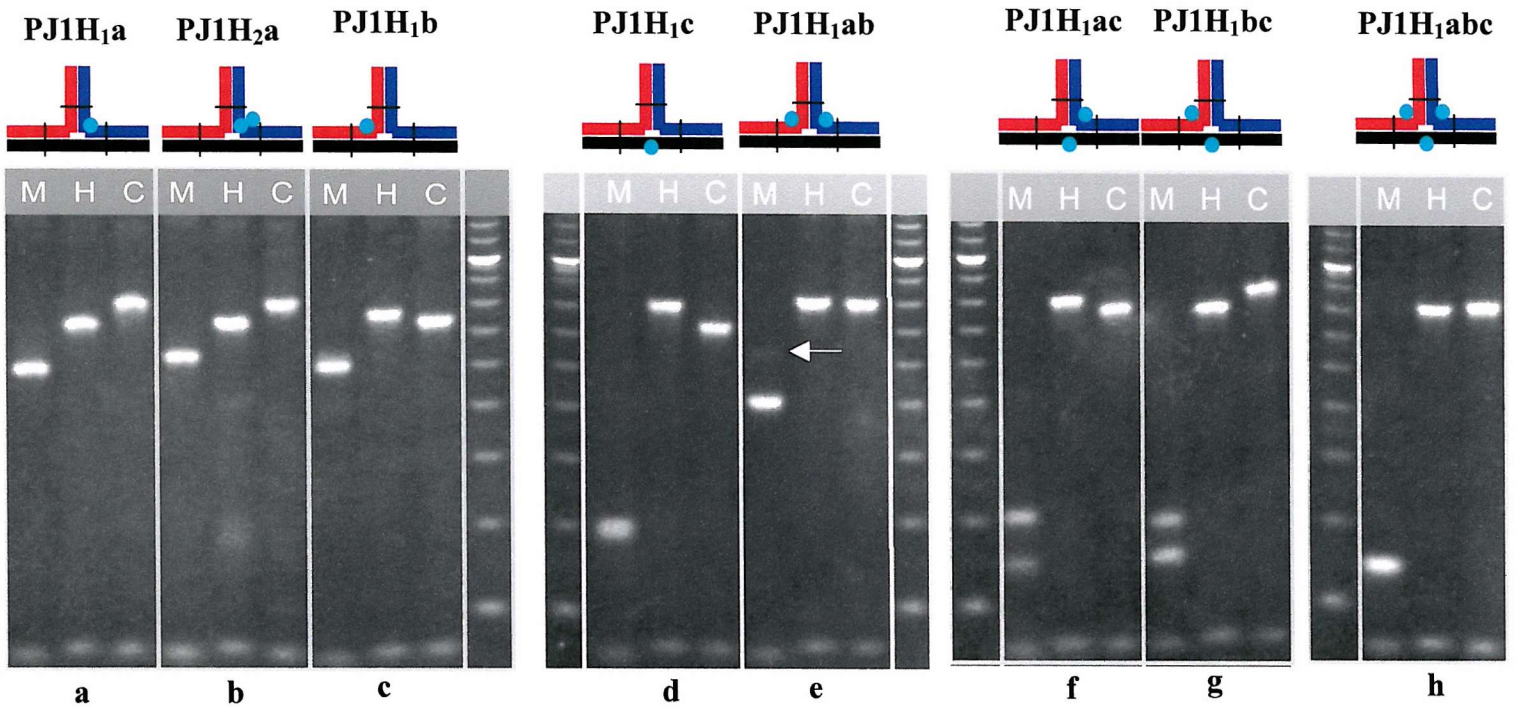


Fig. 6.3. Cleavage patterns for PJ1-based junctions with HEG linkers at different locations around the branch point (HEG linkers are indicated by the red circles). Note that the second junction from the left contains two HEG linkers in the template probe strand. M = *Mse*I, H = *Hin*PI, C = *Csp*45I. Ladders are 10bp ladders, purchased from Gibco LifeTechnologies Ltd., UK, with the 100bp band being brighter than other bands. The arrow is explained in the text.

unpaired cytosines. However, the junctions containing at least one HEG in the target strand (Fig. 6.3d, f, g, h) show unusual fragment mobility patterns. PJ1H$_1$c for instance (Fig. 6.3d), shows the appearance of an unusually fast migrating *Mse*I fragment. In addition, PJ1H$_1$ac (Fig. 6.3f) and PJ1H$_1$bc (Fig. 6.3g) each display two high mobility *Mse*I fragments and PJH$_1$abc (Fig. 6.3h) shows one high mobility *Mse*I fragment, but with a mobility that is different to the fragment observed for PJ1H$_1$c. These bands migrate with a speed comparable to the 20 and 30 bp bands of the DNA ladder, which is clearly much more rapid than the expected minimum of 57 bp (the actual size of the *Mse*I fragment). The mobility of these fragments is therefore too fast to be explained on the basis of a conformational change in the junction. Instead, it is possible that *Mse*I has cleaved the junction at additional sites. To assess whether this was the case, and if so, whether the additional cleavage activity depended on the presence of the HEG, a control experiment was conducted. A number of DNA constructs, shown in Fig. 6.4, were analysed that were based on PJ1H$_1$c. One was a duplex containing the target strand of PJ1H$_1$c and its complement and one was a synthetically prepared PJ1H$_1$c *Mse*I fragment. In addition, derivatives of these three constructs were prepared lacking the presence of a HEG. Note that none of these fragments contained the *Mse*I site (TTAA). These fragments were annealed as described in section 2.2 and incubated in the presence or absence of *Mse*I for 4 hours. Following restriction, the mixtures were subjected to denaturing and native polyacrylamide gel electrophoresis. Fig. 6.4 shows the results of this experiment and reveals that the synthetic PJ1H$_1$c *Mse*I fragment incubated in the presence of *Mse*I shows a fast migrating species on both denaturing and non-denaturing gel (compare Fig. 6.4d and 6.4h). In the presence of *Mse*I, both types of gel show the appearance of a fast migrating fragment. In contrast, in the presence of *Mse*I the HEG-containing duplex shows a fast migrating species only on the denaturing gel (Fig. 6.4b & f). These results strongly suggest that *Mse*I has nicked the target strand in the vicinity of the HEG. This is further explained in Fig. 6.5, illustrating that a nick in the target strand at the branch point in the synthetic *Mse*I fragment of PJ1H$_1$c leaves only a very short duplex connecting the extension probe/target and template probe/target arms. Melting of this duplex region would yield two separate 25 bp duplexes with 10-base overhanging ends. Thus, the high mobility fragment observed on the native gel for the synthetic *Mse*I fragment of PJ1H$_1$c, probably represents these two 25 bp duplexes. The high mobility band visible on the denaturing gel on the other hand, probably

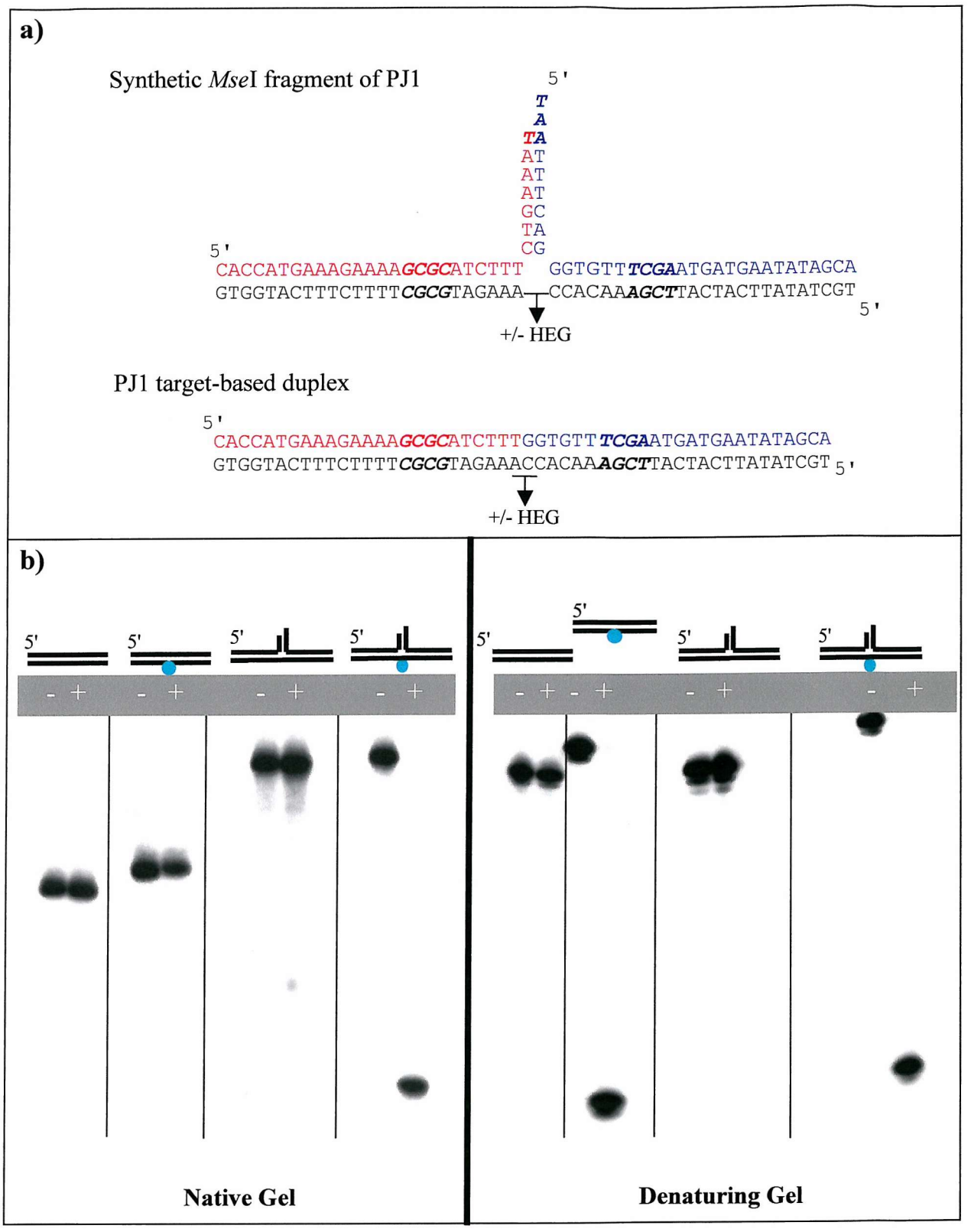


Fig. 6.4 Effect of HEG on the endonuclease activity of *Mse*I. The duplex constructs were the target strand of PJ1 annealed to its complement (the sequence is shown above), either containing or lacking a HEG in the target strand at the position corresponding to the branch point in PJ1. The TWJ constructs were synthetically prepared *Mse*I fragments of PJ1 (shown above), either containing or lacking a HEG in the target strand at the branch point. For further explanation, please refer to the text. The green circles indicate HEG. + = Incubated in the presence of *Mse*I, - = incubated in the absence of *Mse*I. All constructs had the target strand (bottom strands in the constructs shown above the gels) labeled.

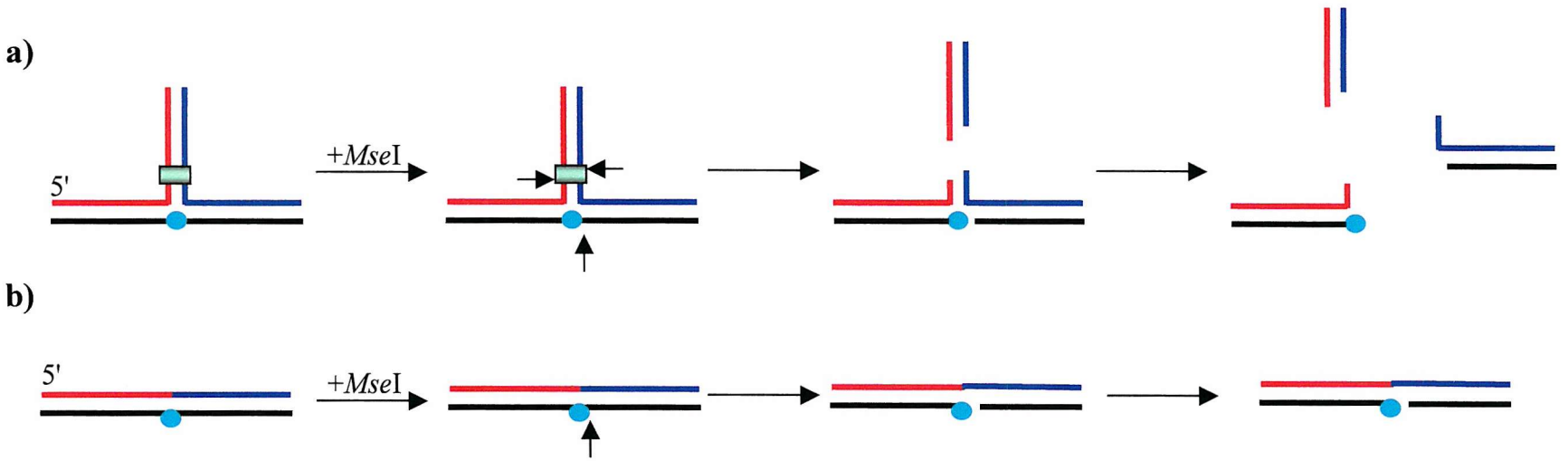


Fig. 6.5. Explanation for the results obtained for PJ1-junctions with a HEG in the target strand shown in Fig. 6.3 and Fig. 6.4, assuming *Mse*I nicked the target strand. (A) When a TWJ with a HEG at the branch point in the target strand is subjected to *Mse*I cleavage, both the canonical *Mse*I site and the target strands are cleaved. The resulting fragment is probably highly unstable due to the short (9 bp) extension probe/template probe arm (red/blue arm) and the nick in the target strand. Due to the low stability, this fragment probably dissociates into two discrete fragments even under native conditions, with each fragment containing 25 bp and a short single-stranded overhang. (B) The control duplex analysed in Fig 6.4b and 6.4f contains only a single nick after *Mse*I incubation. Hence, the target-complementary strand remains intact and the duplex does not dissociate (unless when subjected to a denaturing gel).
The green shaded box: *Mse*I restriction site, filled circle: HEG, red strand: extension probe, blue strand: template probe, black strand: target.

represents the cleaved target strand. The results for the duplex in Fig. 6.4b and 6.4f can also be explained on the basis of a HEG-dependent nicking activity of *Mse*I. In contrast to the situation for the synthetic *Mse*I fragment of PJ1H$_1$c, nicking of the duplex is unlikely to destabilise the duplex sufficiently to cause it to dissociate under native conditions. This explains why the native gel does not show a high mobility band, whereas the denaturing gel does.

This finding is highly unusual since the fragments analysed in Fig. 6.4 do not contain an intact *Mse*I restriction sequence. In other words, the *Mse*I nicking activity appears to be independent of the presence of canonical substrate, but dependent on the presence of HEG.

The observation that unusually fast migrating species were only apparent for PJ1-derived junctions with a HEG present in the target strand (Fig. 6.3), could imply an important role for junction conformation and/or the DNA sequence around the HEG. If the presence of HEG alone would be sufficient to induce cleavage, fast mobility fragments would be expected for junctions containing a HEG in other strands. Nevertheless, PJ1H$_1$ab (Fig. 6.3e), containing a HEG in the extension probe and template probe strands, shows two fragments of increased mobility of which one is much fainter (indicated by the arrow in Fig. 6.3e). The faint band migrated normally, between the 70 and 80 bp bands of the DNA ladder (which is comparable to the *Mse*I fragments of PJ1H$_1$a and PJ1H$_1$b). In contrast, the bright band migrates close to the 60 bp band of the DNA ladder, which is close to its molecular weight. This is unusual since all junctions studied in the previous chapters showed mobilities at or above the 70 bp band of the DNA ladder. The fact that two bands rather than a single band was observed for PJ1H$_1$ab incubated with *Mse*I, could mean that *Mse*I also nicked this junction in manner similar to that observed for PJ1H$_1$c.

In summary, the results provide evidence for an unusual cleavage activity of *Mse*I in the presence of DNA with a HEG linker, which is independent of the presence of canonical substrate. Thus far, the results suggest that this activity is restricted to *Mse*I as other restriction endonucleases such as *Hin*PI and *Csp*45I did not appear to cleave the junctions studied here in an unusual fashion.

6.4 Analysis of PJ16 with unpaired bases, HEG, octanediol or abasic sites

6.4.1 *Bam*HI also cleaves HEG-containing DNA in an unconventional manner

A similar result to that described for PJ1 in the previous section was also observed for an unrelated junction, which revealed anomalous cleavage by *Bam*HI when a HEG linker was present in the template probe strand. This junction, PJ16H_2a (Fig. 6.6), appeared to be a conformer I junction when it contained two unpaired bases in the template probe strand. In addition, PJ16 contains cleavage sites for different restriction endonucleases to PJ1, namely: *Eco*RI, *Bam*HI and *Ava*I. Replacing the two cytosine bases with two HEG linkers in the template strand of PJ16C_2a (PJ16H_2a) resulted in the fragment mobility pattern shown in Fig. 6.7. This figure clearly shows the appearance of an unusually high mobility band when PJ16H_2a was incubated in the presence of *Bam*HI, comparable to the band observed for *Mse*I-cleaved PJ1H_1c although the *Bam*HI fragment migrated slower. Since PJ16C_2a did not show the appearance of a high mobility band when incubated in the presence of any of the three restriction fragments, the presence of HEG again appears to mediate unconventional cleavage, this time by *Bam*HI. These results therefore suggests that the HEG-mediated cleavage is not a unique property of a single restriction endonuclease, but possibly an inherent property of several restriction endonucleases.

6.4.2 Unconventional cleavage of PJ16H_2a by *Bam*HI is inhibited by phosphorothioate substitution

Further evidence for the involvement of HEG in the cleavage activity of *Bam*HI was obtained by phosphorothioate interference mapping. Previous studies have shown that restriction endonucleases show a (much) reduced cleavage rate when a phosphorothioate molecule replaces the scissile phosphate (Pingoud & Jeltsch, 1997 and references therein). Fig. 6.7 shows that when the phosphate located between the HEG and the neighbouring 5' base was replaced by a phosphorothioate, the unusual high mobility band disappeared, strongly suggesting that this phosphate group in PJ16H_2a was involved in the cleavage by *Bam*HI.

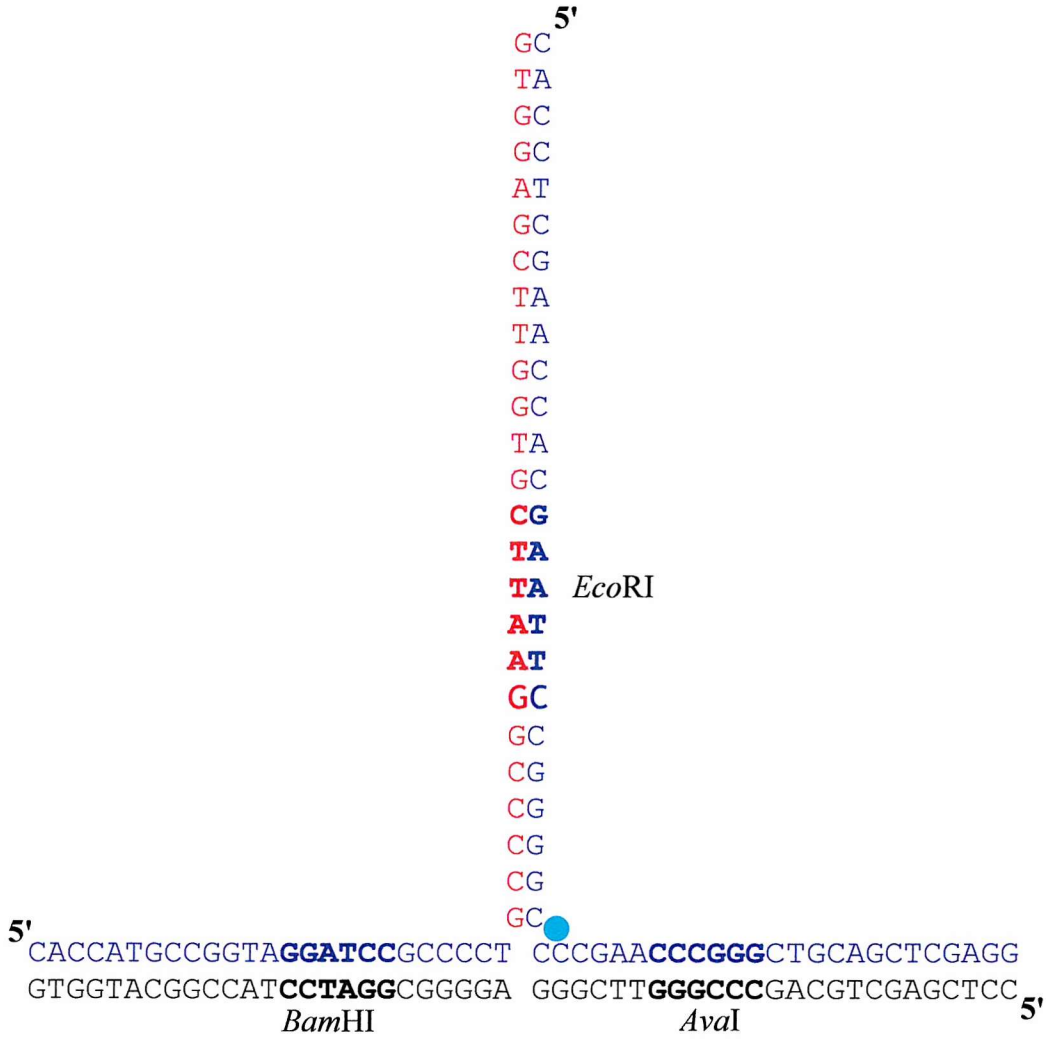


Fig. 6.6. Design of PJ16C_2a/H_2a. The filled circle indicates the location of the linker.

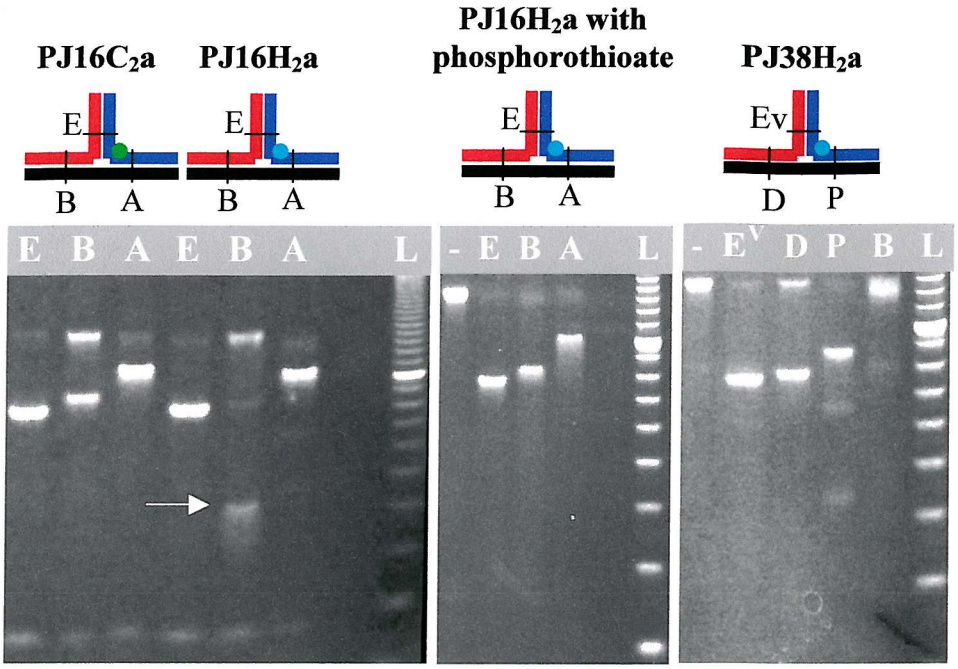


Fig. 6.7. Effect of HEG in PJ16 on the cleavage of activity of *Bam*HI. The first, left most junction contains two unpaired cytosines at the branch point (green circle). The second junction contains a HEG linker (green circle) at branch point and the third a HEG linker plus a phosphorothioate located between the 5' end of the HEG and the adjacent base in the template probe strand. The fourth, right most gel shows the result of a junction ($PJ38H_2a$), containing the same branch point sequences as PJ16 and a HEG at the branch point, but lacking the *Bam*HI site.
- = Uncleaved junction. A = *Ava*I, B = *Bam*HI, D = *Dra*I, E = *Eco*RI, Ev = *Eco*RV, P = *Pvu*II. The ladder is as in previous figures. The arrow indicates the anomalous fragment, which is absent in the unpaired cytosine and phosphorothioate-containing junctions. Note also that *Ava*I appears to generate two faint, faster bands, with one migrating closely to the anomalous *Bam*HI fragment. Both the extra *Ava*I fragments are very similar to the extra bands appearing for *Pvu*II cleavage.

6.4.3 Unconventional cleavage of PJ16H$_2$a by *Bam*HI is independent of the presence of canonical substrate

To assess whether the HEG-dependent cleavage activity of *Bam*HI was dependent on the presence of canonical substrate an additional junction was designed, named PJ38 (Fig. 6.8). This junction contained the same sequence as PJ16 except that the restriction sequences were replaced by those for the blunt cleaving restriction endonucleases *Eco*RV, *Dra*I and *Pvu*II (these replaced *Eco*RI, *Bam*HI and *Ava*I respectively). In other words, the *Bam*HI site had been eliminated. Incubation of PJ38H$_2$a with *Eco*RV, *Dra*I, *Pvu*II and *Bam*HI yielded a fragment mobility pattern shown in Fig. 6.7. *Eco*RV, *Dra*I and *Pvu*II restrictions were comparable to the respective *Eco*RI, *Bam*HI and *Ava*I restrictions of PJ16C$_2$a. However, incubation of PJ38H$_2$a with *Bam*HI showed the appearance of two closely migrating species. The two bands migrate close to that of the uncleaved junction. The top most band appears to correspond to the uncleaved junction. The lower band therefore could have resulted from a single nicking event by *Bam*HI. This was confirmed by subjecting PJ16C$_2$a, PJ16H$_2$a , phosphorothioate-substituted PJ16H$_2$a and PJ38H$_2$a restriction cleavage reactions to denaturing gel electrophoresis (Fig. 6.9-6.12). Different versions of these junctions were assessed, with each version having a different strand labelled. Only PJ16H$_2$a (Fig. 6.10) and PJ38H$_2$a (Fig. 6.12) with the template probe strand labelled show an unusual (because this strand is not part of the canonical *Bam*HI site) cleavage product by *Bam*HI. Fig. 6.11 clearly shows that this cleavage activity was inhibited by phosphorthioate substitution (see also section 4.3.2), and Fig. 6.9 shows that replacing the HEG with two cytosines also inhibited the unconventional *Bam*HI cleavage of the template probe.

Junctions with the extension probe labelled also yielded a high mobility band in the presence of *Bam*HI (Fig. 6.9-6.11), but this is expected as the canonical *Bam*HI site is positioned close to the 5' label (which explains the absence of this band for PJ38H$_2$a, Fig. 6.12, as this junction lacked a *Bam*HI site). Similar arguments explain the higher mobility of extension probe-labelled PJ16H$_2$a-derived junctions (Fig. 6.9-6.11), incubated in the presence of *Eco*RI (which shortens the extension probe by cleavage of its recognition site), and target-labelled junctions incubated in the presence of *Bam*HI.

Although one could argue that unusual cleavage of the extension probe of PJ16H$_2$a by *Bam*HI could have occurred (which would be masked on a denaturing gel

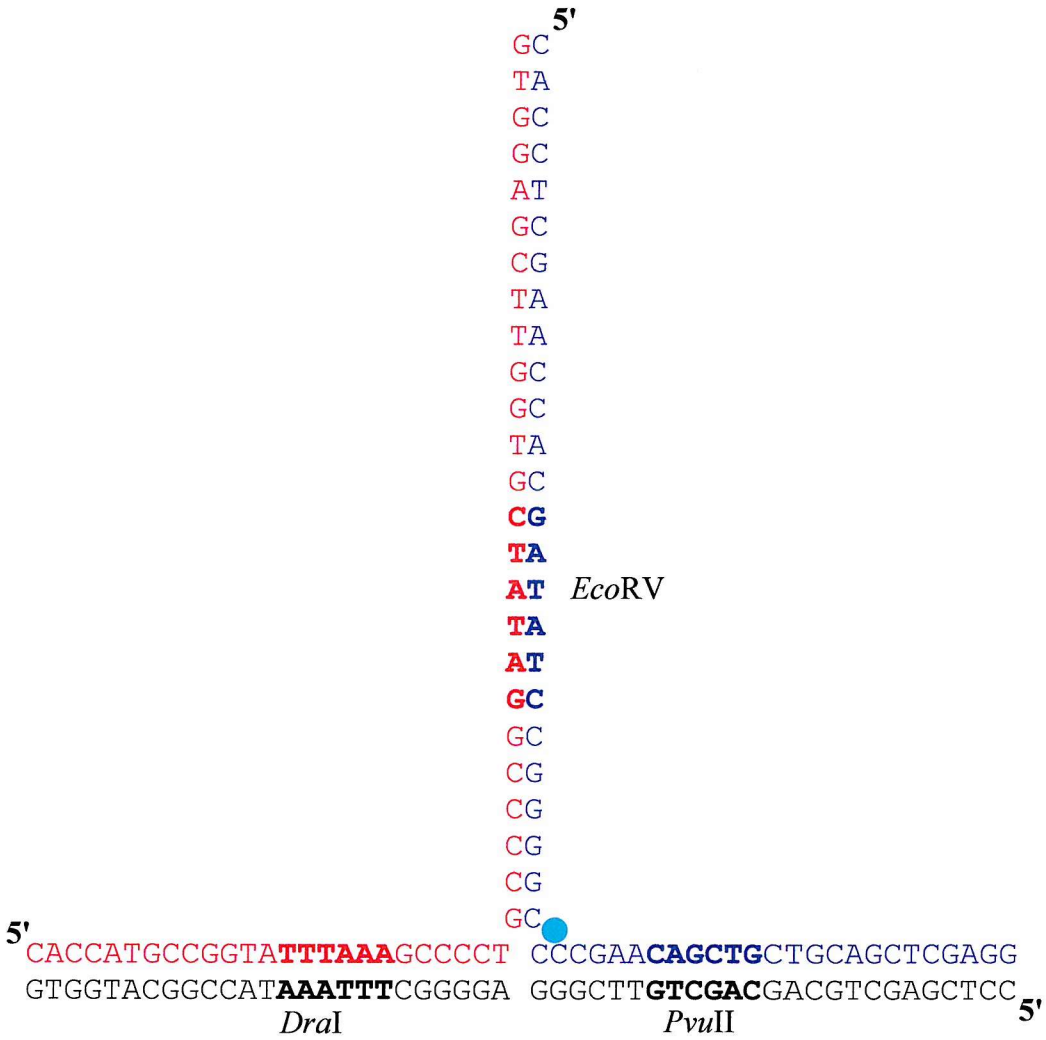


Fig. 6.8. Design of PJ38H_2a. The filled circle indicates the location of the HEG linker.

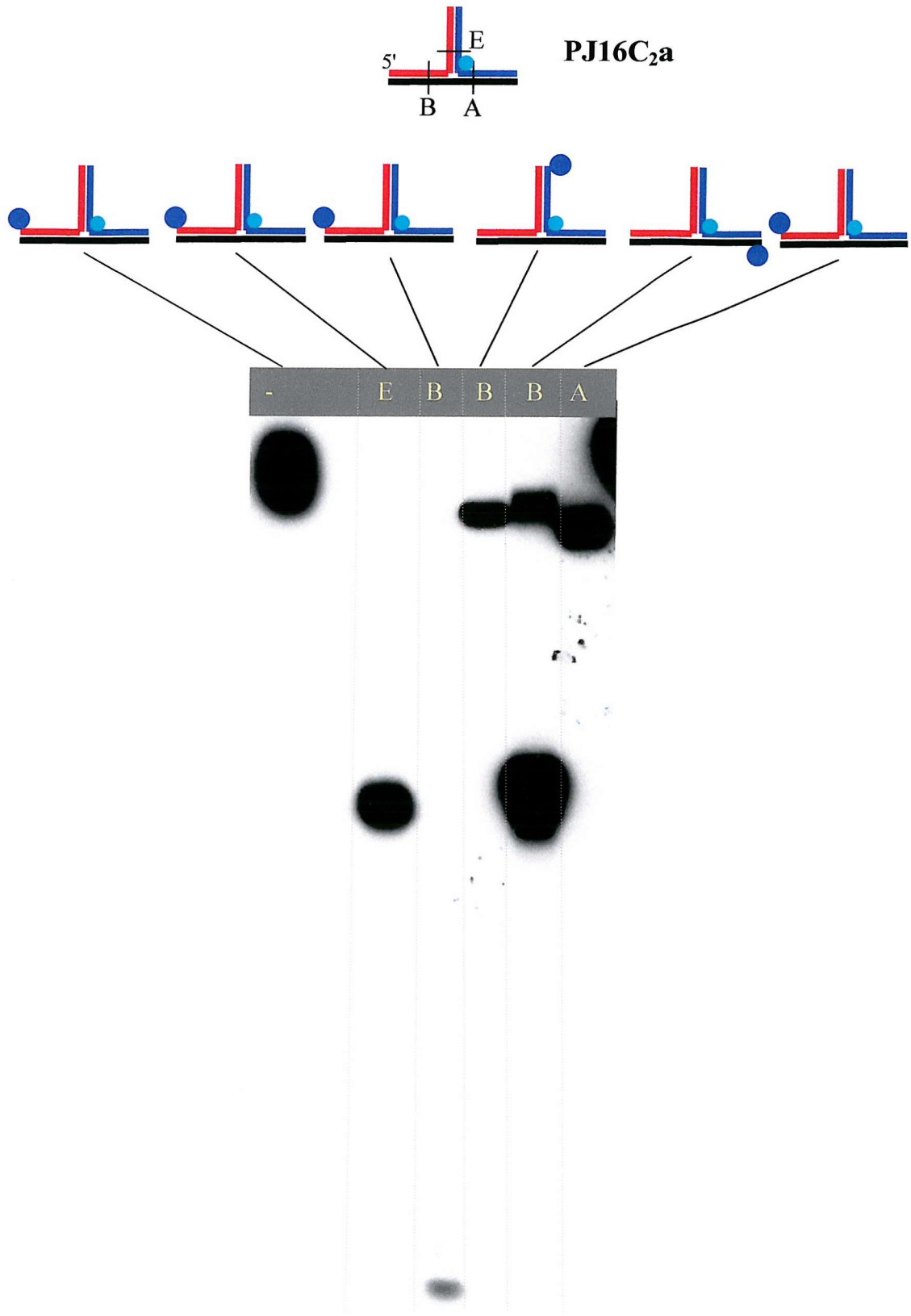


Fig. 6.9. Denaturing gel electrophoresis of PJ16C_2a containing different labelled strands and cleaved with *Eco*RI, *Bam*HI or *Ava*I. The 5' γ-^{32}P-ATP label is represented by the blue circles and the linker by the green circles. E = *Eco*RI, B = *Bam*HI, A = *Ava*I.

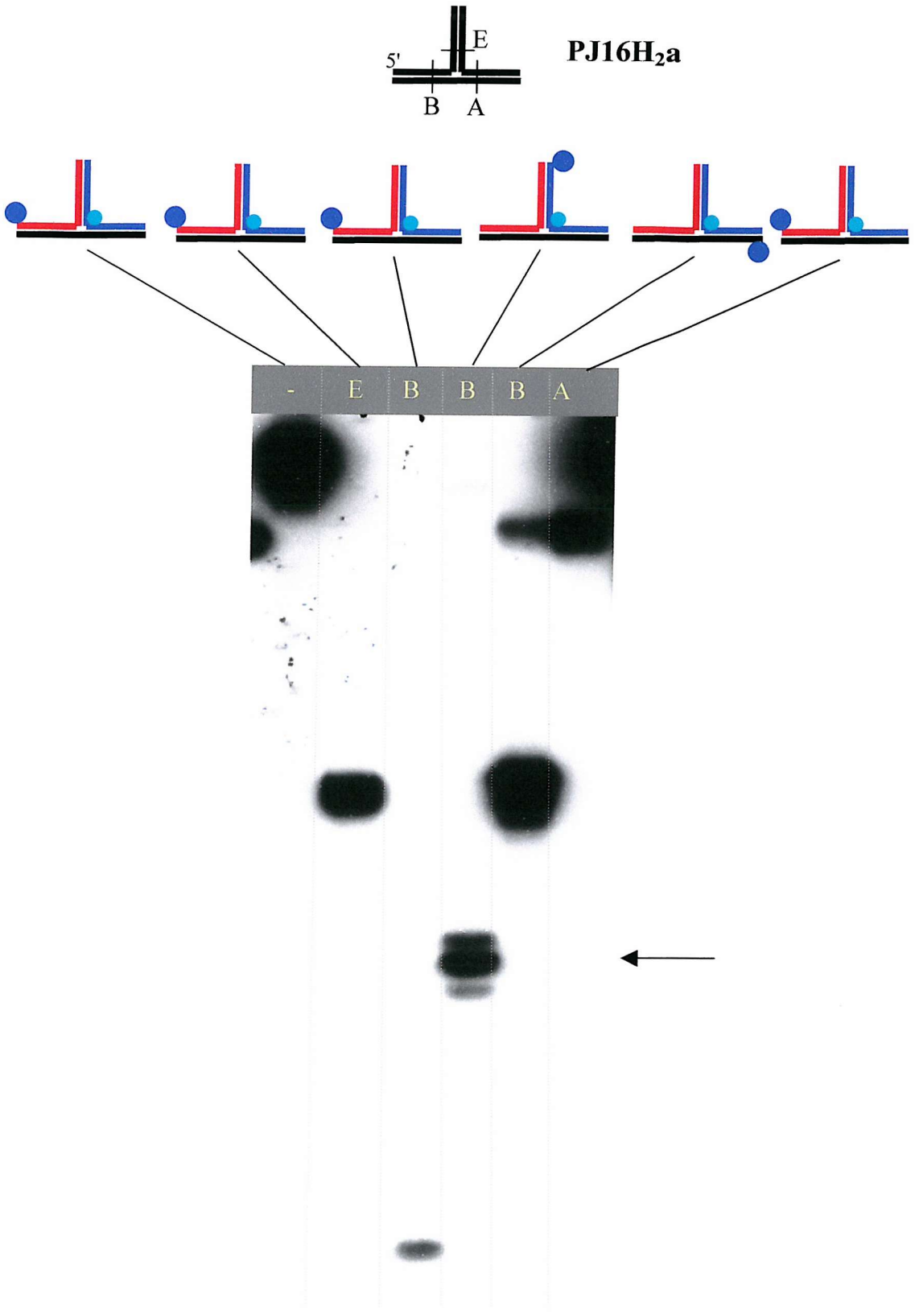


Fig. 6.10. Denaturing gel electrophoresis of PJ16H$_2$a containing different labelled strands and cleaved with *Eco*RI, *Bam*HI or *Ava*I. The 5' γ-^{32}P-ATP label is represented by the blue circles and the linker by the green circles. E = *Eco*RI, B = *Bam*HI, A = *Ava*I. The arrow indicates the anomalous cleavage product generated by *Bam*HI.

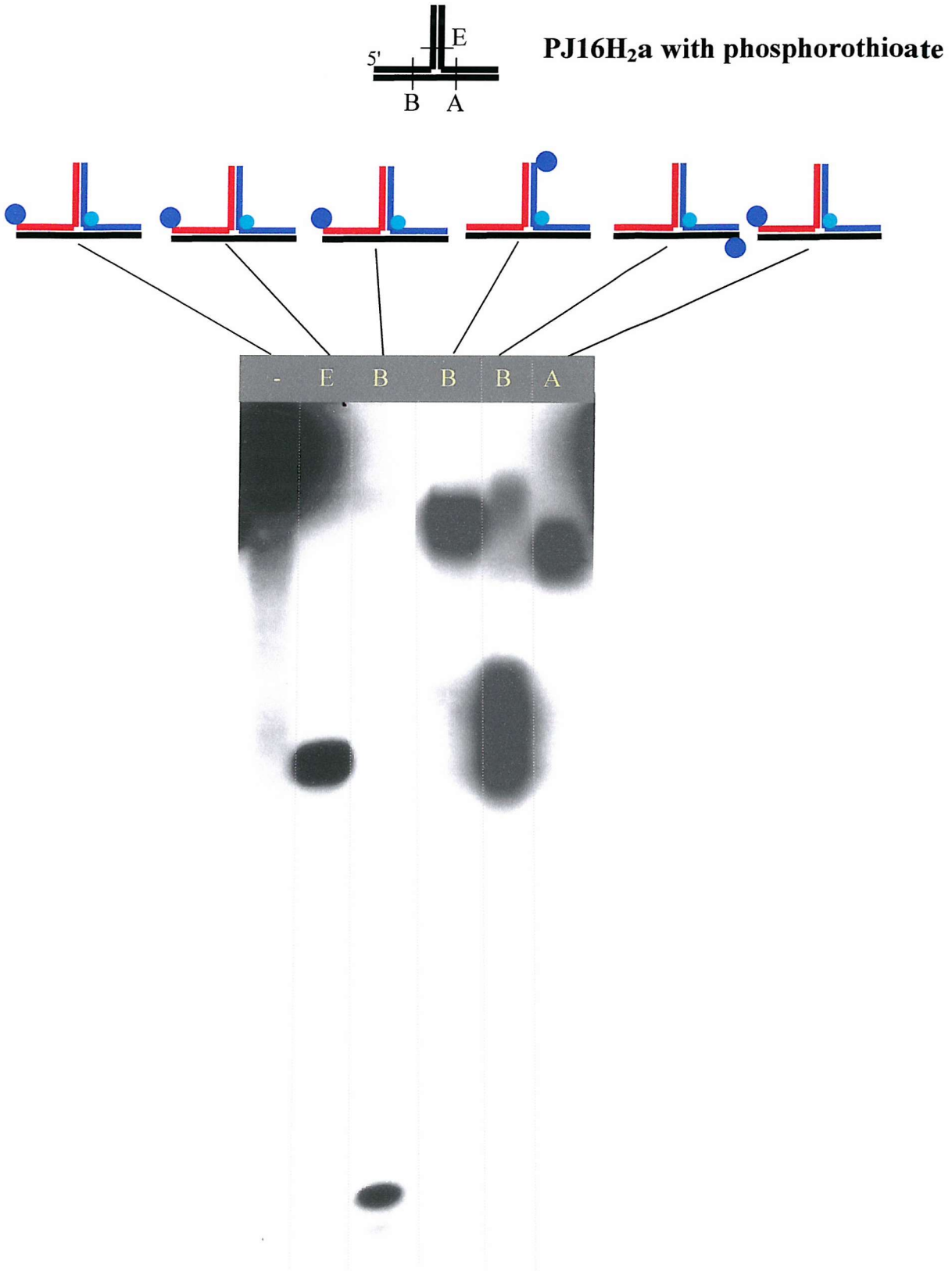


Fig. 6.11. Denaturing gel electrophoresis of PJ16H$_2$a with a phosphorthioate located between HEG and the first 5' base from HEG, containing different labelled strands and cleaved with *Eco*RI, *Bam*HI or *Ava*I. The 5' γ-^{32}P-ATP label is represented by the blue circles and the linker by the green circles. E = *Eco*RI, B = *Bam*HI, A = *Ava*I. The arrow indicates the anomalous cleavage product generated by *Bam*HI.

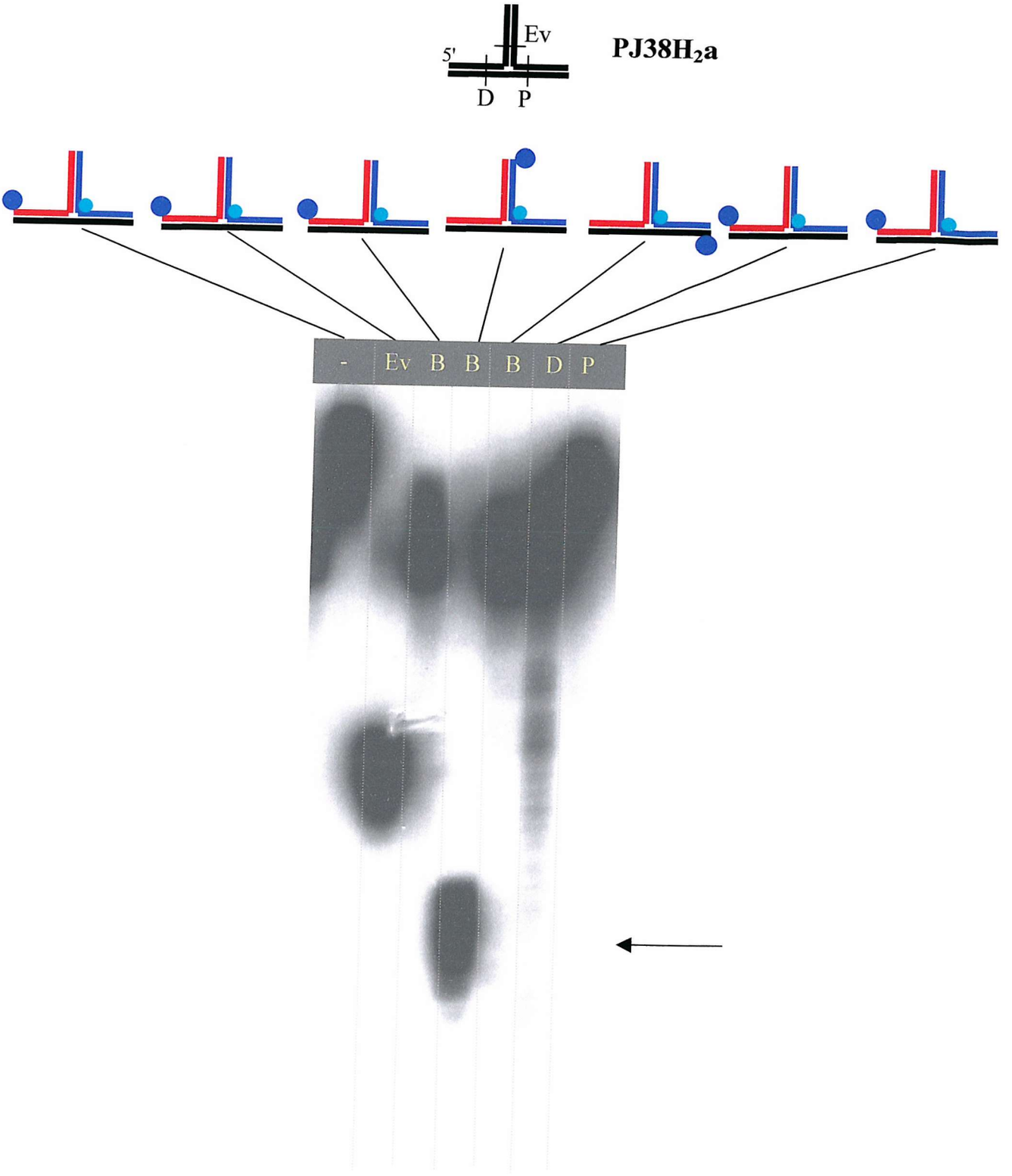


Fig 6.12. Denaturing gel electrophoresis of PJ38H$_2$a, containing different labelled strands and cleaved with *Eco*RV, *Dra*I, *Pvu*II or *Bam*H. The 5' γ-^{32}P-ATP label is represented by the blue circles and the linker by the green circles. Ev = *Eco*RV, B = *Bam*HI, D = *Dra*I, P = *Pvu*II. The arrow indicates the anomalous cleavage product generated by *Bam*HI.

as in Fig. 6.10 by cleavage of the canonical *Bam*HI site), it is at least clear that *Bam*HI did not cleave the extension probe in PJ38H$_2$a (Fig. 6.12). In other words, if unconventional cleavage of the extension probe in PJ16H$_2$a had occurred, it must have been dependent on the presence of canonical substrate. The latter appears however, a highly unlikely situation and absence of extension probe cleavage seems a more likely explanation.

Thus, no evidence for unconventional cleavage of other strands than the template probe by *Bam*HI was detected, suggesting that the cleavage involved only one strand, which supports the HEG-dependency of the reaction.

The results presented in this section are very similar to those obtained for HEG-containing PJ1 derivatives incubated in the presence of *Mse*I and a similar explanation can be invoked for the high mobility bands observed for PJ16H$_2$a incubated with *Bam*HI, illustrated in Fig. 6.13. The combined cleavage of the canonical *Bam*HI site in the extension probe/target arm and the nicking of the template probe at the branch point, yields a product very similar to *Mse*I-cleaved PJ1H$_1$c (Fig. 6.5). As argued before, this product is likely to be unstable since it involves only a short duplex region (9 bp).

6.4.4 Synthetic methylation of the *Bam*HI site of PJ16H$_2$a does not inhibit unconventional cleavage by *Bam*HI

To further ascertain the importance of the cognate *Bam*HI sequence in PJ16H$_2$a, a number of 5-methylcytosine substituted versions of this junction were designed (Fig. 6.14). The 5-methylcytosine groups were incorporated into the canonical *Bam*HI sequence at sites where the natural *Bam*HI methyltransferase normally acts. Note that the latter methylates cytosines at the N4 position. However, since 5-methylated cytosine is readily available and 4-methylated cytosine is not, the former was assessed in these studies. Also, as the results presented in Fig. 6.14d show, 5-methylation inhibits *Bam*HI cleavage of its canonical substrate, which was the main purpose of methylation in these experiments. Synthetically methylated PJ16H$_2$a, with methyl groups present in both strands, inhibits cleavage of the *Bam*HI site as judged by the absence of a high mobility band (Fig. 6.14d). However, the faster band migrating close to the uncleaved junction band (similar to the band appearing when PJ38H$_2$a was subjected to *Bam*HI treatment, Fig. 6.7) does appear for this junction, indicating

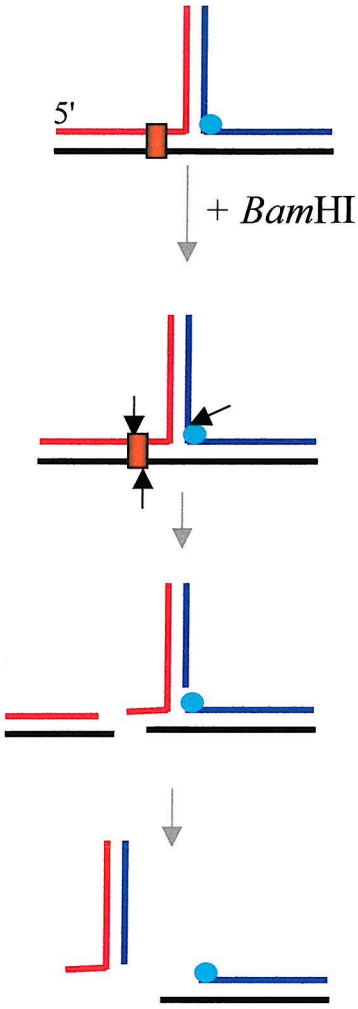


Fig. 6.13. Explanation for the results obtained for PJ16-junctions with a HEG in the template probe, analogous to the explanation for *Mse*I-cleavage of PJ1 junctions shown in Fig. 6.5. *Bam*HI cleavage of canonical substrate in conjunction with nicking the template probe resulted in a similar fragments as observed for *Mse*I cleavage of $PJ1H_1c$, which probably resulted from dissociation of the junction.
The red shaded box: *Bam*HI restriction site, filled circle: HEG, red strand: extension probe, blue strand: template probe, black strand: target.

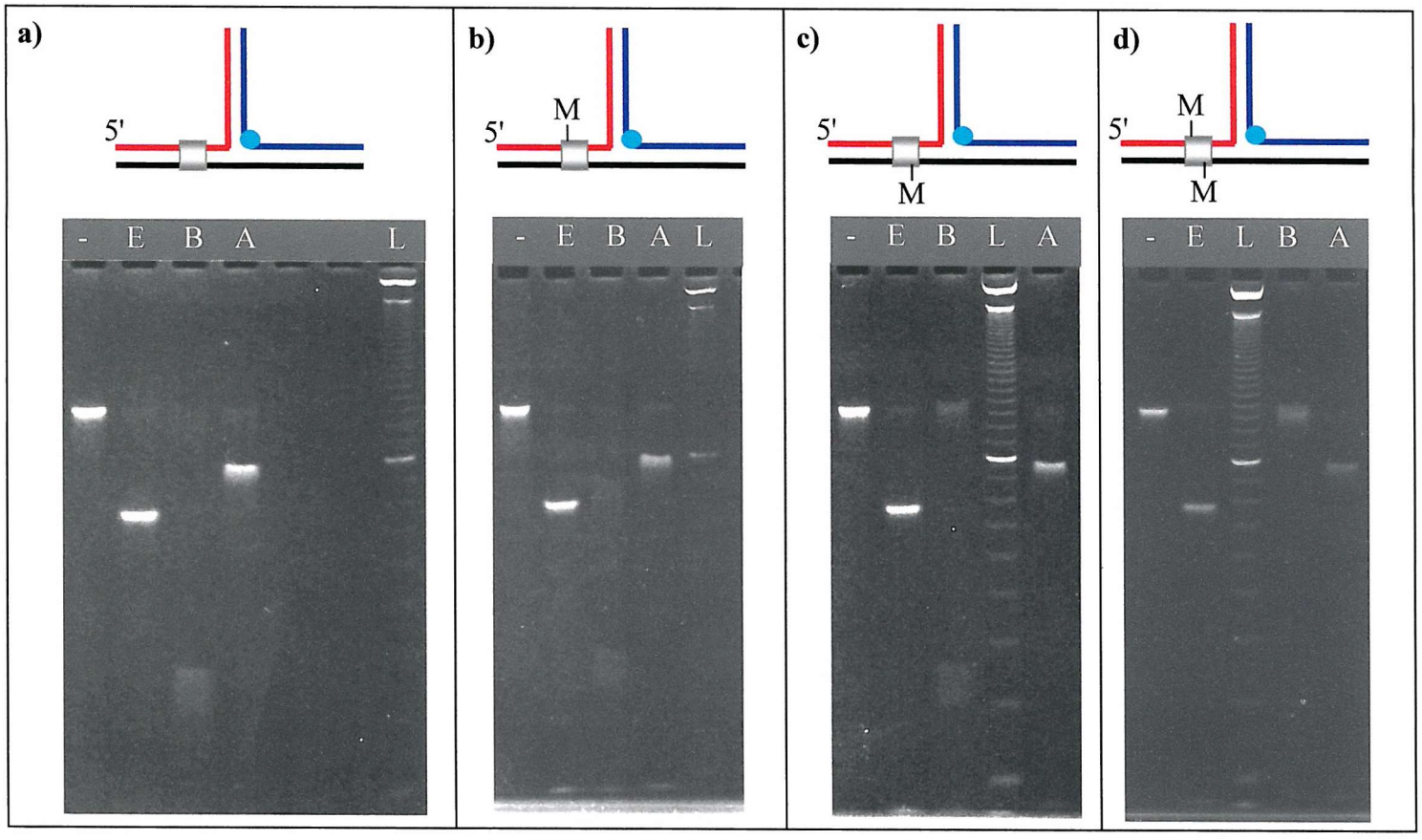


Fig. 6.14. Effect of synthetic methylation on the unconventional cleavage activity of *Bam*HI. E = *Eco*RI, B = *Bam*HI, A = *Ava*I, L = 10 bp ladder (Gibco/Life Technologies Ltd.), M = 5-methyl cytosine. The filled circle represents the HEG linker. *Bam*HI site is represented by the grey shaded box.

that HEG-mediated cleavage occurred. This result further supports the conclusion that canonical substrate is not required for the HEG-dependent cleavage by *Bam*HI.

Hemi-methylated substrate (Fig. 6.14b and 6.14c) did not appear to inhibit the cleavage of canonical substrate as *Bam*HI cleavage yielded a high mobility band. This is in accordance with other experiments, showing that hemi-methylated substrates are cleaved by R-ENases but at a reduced rate (Jen-Jacobson *et al.*, 1996). The results presented here show that both full and hemi methylated substrates had no inhibitory effect on the HEG-dependent cleavage by *Bam*HI.

6.4.5 Effects of prolonged incubation of PJ16H$_2$a and PJ38H$_2$a with R-ENases

One curious observation, not mentioned so far, are the faint bands visible for *Pvu*II cleavage of PJ38H$_2$a in Fig 6.7 and for *Ava*I cleavage of PJ16H$_2$a. Although these bands could have been artefacts, increasing the incubation time from 4 hours to 10 hours showed a marked increase in the intensity of the anomalous bands (Fig. 6.15). These results might imply that *Pvu*II (and possibly *Ava*I) cleaved HEG-containing junctions in the same fashion as *Bam*HI but at a much-reduced rate. The following section provides support for this suggestion.

6.4.6 Interactions between *Bam*HI methyltransferase and PJ16H$_2$a and PJ38H$_2$a

In vivo, bacterial restriction endonucleases collaborate with methyltransferases to protect the host's genome from anomalous endonuclease cleavage (by methylation) and eliminate foreign DNA (by restriction cleavage). Since methyltransferases recognise the same sequence as their restriction endonuclease partner, albeit in a different manner, it was of interest to assess whether the naturally occurring similarity in target selectivity between R-ENases and methyltransferases extends to HEG-containing junctions. To assess this, PJ16H$_2$a was subjected first to *Bam*HI methyltransferase treatment (in the presence of S-adenosylmethionine) for three hours, followed by a three hour incubation with *Eco*RI, *Bam*HI or *Ava*I R-ENase and subjection to non-denaturing gel electrophoresis. The results, shown in Fig. 6.16, indicate that PJ16H$_2$a pre-treated with *Bam*HI methyltransferase (henceforth designated as PJ16H$_2$a-M) shows a different pattern from untreated PJ16H$_2$a and PJ16C$_2$a. Both *Eco*RI and *Ava*I now show the appearance of two bands, whereas

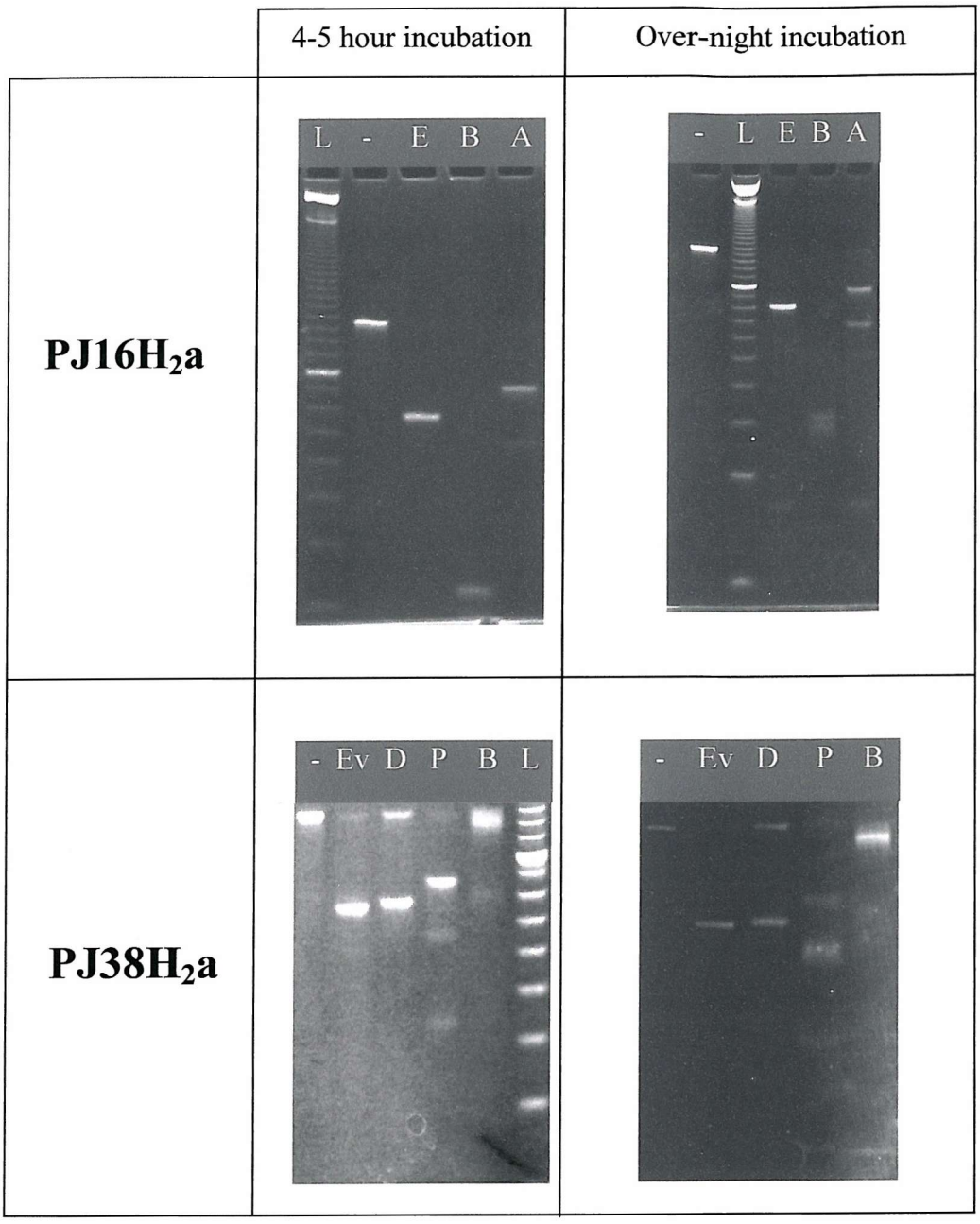


Fig. 6.15. Effect of incubation time on the appearance of anomalous bands for PJ16H_2a and PJ38H_2a. E = *Eco*RI, B = *Bam*HI, A = *Ava*I, Ev = *Eco*RV, D = *Dra*I, P = *Pvu*II, L = 10 bp ladder (Gibco/Life technologies Ltd., UK), - = no R-ENase.

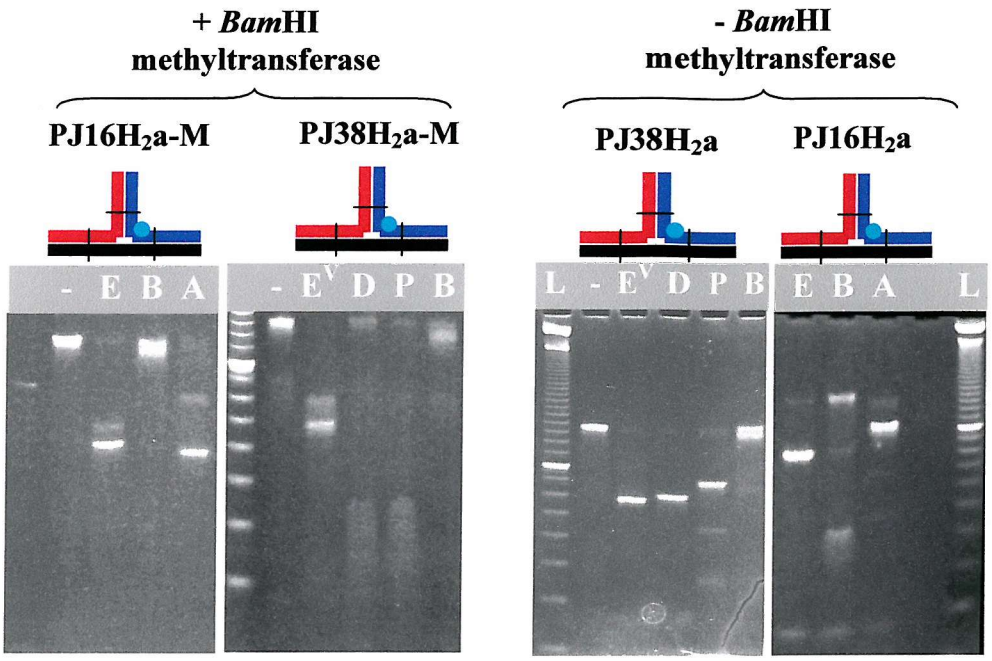


Fig. 6.16. Effect of *Bam*HI methyltransferase treatment of junctions PJ16H$_2$a and PJ38H$_2$a. The right two gels were included for comparison and are shown also in Fig. 6. The methyltransferase treated junctions both show enhancement of the mobility of all fragments. The *Eco*RI and *Eco*RV patterns resemble each other closely, whereas the *Dra*I and *Pvu*II patterns are similar to that observed for *Bam*HI treated PJ16H$_2$a (right most gel). The second band in the *Ava*I lane of PJ16H$_2$a-M migrates close the slowest of the two faint bands observed for untreated PJ16H$_2$a *Ava*I cleavage (right most gel) and to the second fragment visible in the untreated PJ38H$_2$a *Pvu*II lane. Finally, *Bam*HI restriction endonuclease treatment of PJ16H$_2$a-M, PJ38H$_2$a-M and untreated PJ38H$_2$a yields two closely migrating fragments. - = Uncleaved junction. A = *Ava*I, B = *Bam*HI, D = *Dra*I, E = *Eco*RI, Ev = *Eco*RV, P = *Pvu*II, L = 10 bp ladder (Gibco/Life Technologies Ltd., UK).

*Bam*HI shows two closely migrating bands. However, the figure indicates that one of the two PJ16H$_2$a-M fragments produced by both *Eco*RI and *Ava*I correspond to the fragments observed for untreated PJ16H$_2$a. The second fragment in the *Eco*RI and *Ava*I lanes appear to migrate faster. The suggestion that these two additional bands originate from a change in conformational equilibrium to a conformer II is probably not correct, since if this were the case the additional *Eco*RI fragment would be expected to migrate slower than the original *Eco*RI fragment. As both extra fragments migrated faster than the original bands, an alternative possibility is that methyltransferase treatment facilitated extra *Eco*RI and *Ava*I cleavage. Nicking of PJ38H$_2$a by *Bam*HI (Fig. 6.7) was previously shown to cause migration enhancement, supporting the previous suggestion. Further support comes from an experiment where the incubation steps were reversed and prolonged: PJ16H$_2$a was first incubated with R-ENases for three hours, and then subjected to an over-night incubation in the presence of *Bam*HI methyltransferase and S-adenosylmethionine. The results, shown in Fig. 6.17, clearly show that the *Eco*RI and *Ava*I reactions yielded only a single band, corresponding to the enhanced mobility bands observed in Fig. 6.16.

PJ16H$_2$a-M treated with *Bam*HI showed a faster migrating species (Fig. 6.16) similar to the one observed for PJ38H$_2$a (Fig. 6.7), and denaturing gel electrophoresis confirmed that *Bam*HI nicked the template probe strand in PJ16H$_2$a-M (Fig. 6.18). Although these results indicate that enhancement of fragment mobility is possible due to a nicking event, it remains to be explained how *Bam*HI methyltransferase can stimulate nicking by *Eco*RI and *Ava*I. The fact that a different mobility pattern was obtained provides evidence for an interaction between *Bam*HI methyltransferase and the branch point. Thus, it is possible that this enzyme methylated residues close to branch point (a region which in the cases of PJ16 and PJ38 is very cytosine rich). It is worth noting that the results do not exclude the possibility that the methyltransferase cleaved the junction, although this appears unlikely in the light of its biological function and the fact that no endonuclease activity is known to be associated with this enzyme.

The proposed unconventional interaction between the methyltransferase and the TWJ is further supported by results obtained for *Bam*HI methyltransferase pre-treated PJ38H$_2$a (which, as pointed out previously, does not contain the canonical *Bam*HI site) shown in Fig. 6.16. The *Eco*RV lane shows two bands, very similar to the two bands observed for *Eco*RI cleavage of PJ16H$_2$a-M. However, *Dra*I and *Pvu*II now

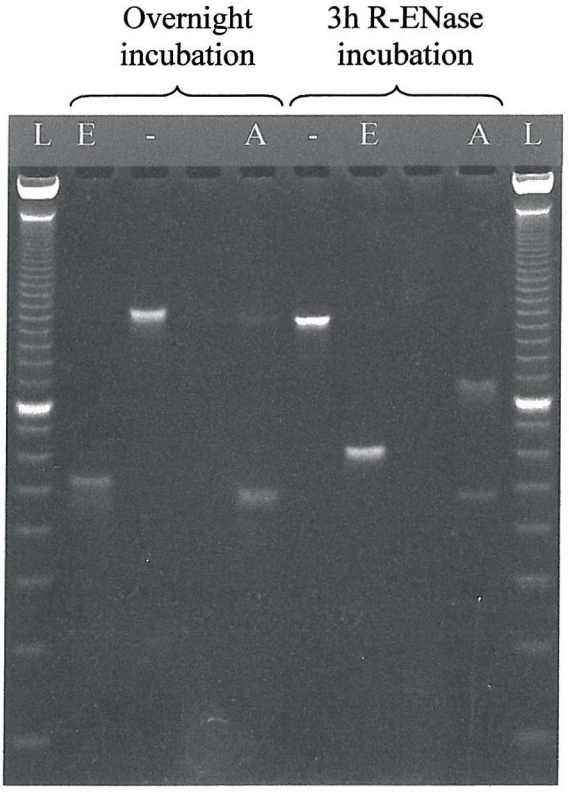


Fig. 6.17. Effect of incubation time on the effect of methyltransferase treatment of PJ16H$_2$a on *Eco*RI and *Ava*I fragment mobility. The overnight incubation consisted of a 3 hour incubation with R-ENase, followed by the addition of *Bam*HI methyltransferase and a further 12h incubation. E = *Eco*RI, A = *Ava*I, L = 10 bp ladder, - = no R-ENase.

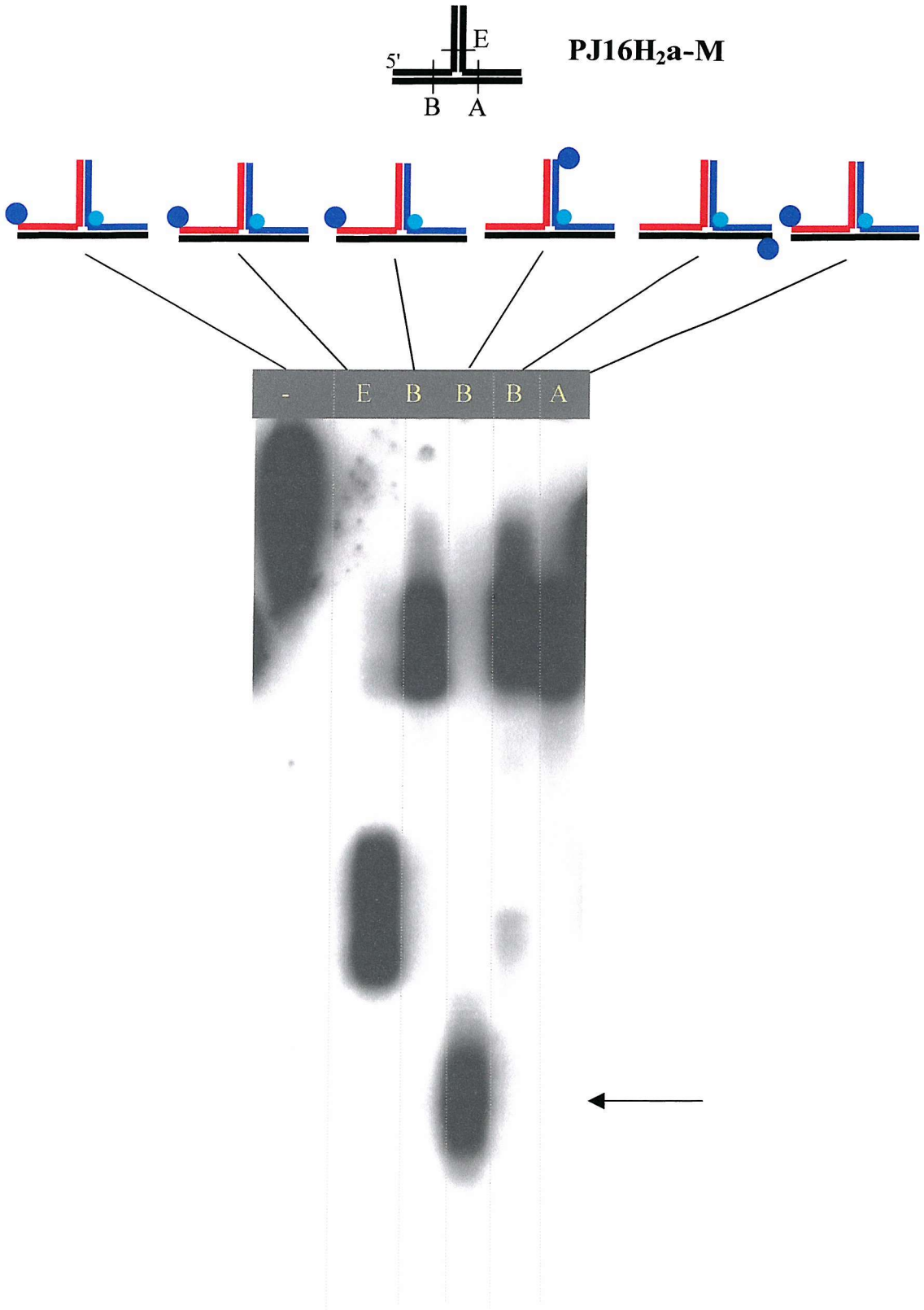


Fig. 6.18. Denaturing gel electrophoresis of *Bam*HI methyltransferase pre-treated PJ16H$_2$a, containing different labelled strands and cleaved with *Eco*RI, *Bam*HI or *Ava*I. The 5' γ-^{32}P-ATP label is represented by the blue circles and the linker by the green circles. E = *Eco*RI, B = *Bam*HI, A = *Ava*I. The arrow indicates the anomalous cleavage product generated by *Bam*HI.

show the appearance of fast migrating species similar to, yet migrating slower than the anomalous fragment produced by *Bam*HI cleavage of PJ16H$_2$a. Thus, although unusual, these observations indicate that *Bam*HI methyltransferase interacts with PJ38H$_2$a. Since *Ava*I and *Pvu*II (both cleaving the same arm) show different results when PJ16H$_2$a and PJ38H$_2$a respectively were applied to methyltransferase treatment, this makes the possibility of methyltransferase mediated cleavage of HEG-containing junctions less likely. Combined with the fact that its natural function is to methylate cytosines in *Bam*HI sites, cytosine methylation at or close to the branch point appears more likely, as suggested before.

6.4.7 Synthetic methylation of the *Bam*HI site in PJ16H$_2$a does not inhibit the unconventional interaction with *Bam*HI methyltransferase.

Section 6.4.4 showed that synthetic cytosine methylation of the *Bam*HI site in PJ16H$_2$a did not inhibit the cleavage of the template probe by *Bam*HI R-ENase (Fig. 6.14). The same synthetically methylated junctions used for this experiment were used to assess whether methylation of the *Bam*HI site would affect the activity of the *Bam*HI methyltransferase. As with the *Bam*HI R-ENase this would provide evidence for the importance of the presence of canonical substrate. Since PJ38H$_2$a appeared susceptible to *Bam*HI methyltransferase treatment in spite of lacking a *Bam*HI site, it was expected that synthetic methylation of the *Bam*HI site in PJ16H$_2$a would not inhibit the observed unconventional activity of the methyltransferase. Fig. 6.19 confirms this expectation as the *Eco*RI and *Ava*I fragments of all constructs increase in mobility in the presence of *Bam*HI methyltransferase, regardless of the methylated state of the *Bam*HI site.

6.4.8 HEG-substitution by other non-nucleosidic linkers does *not* inhibit the unconventional cleavage activity of *Bam*HI.

The previous sections provided evidence for a novel, unconventional cleavage activity of HEG-containing DNA by *Bam*HI R-ENase. To assess to what extent the chemical composition of this linker contributed to the observed *Bam*HI activity, a number of constructs were prepared as shown in Fig. 6.20. These constructs were PJ16H$_2$a-derivatives, truncated in the *Bam*HI site to eliminate any possible influence

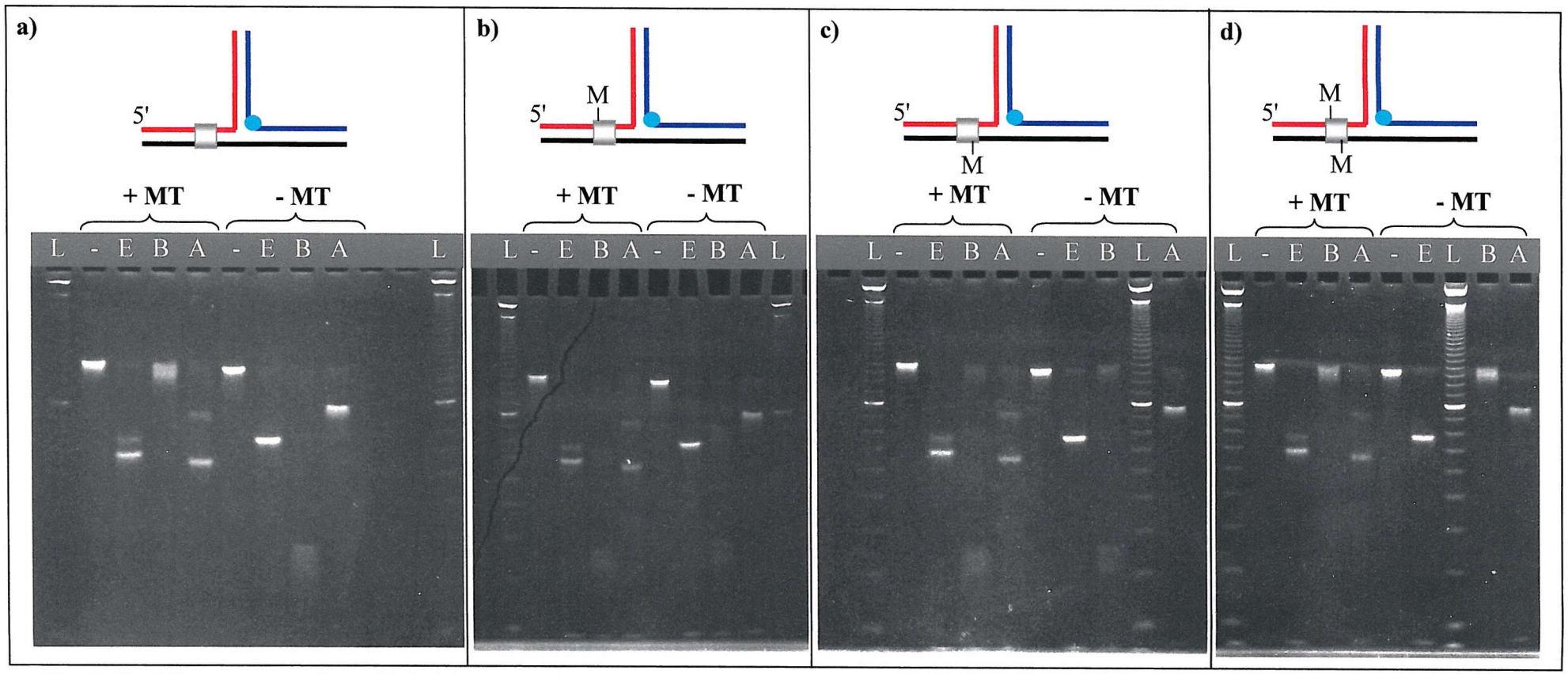
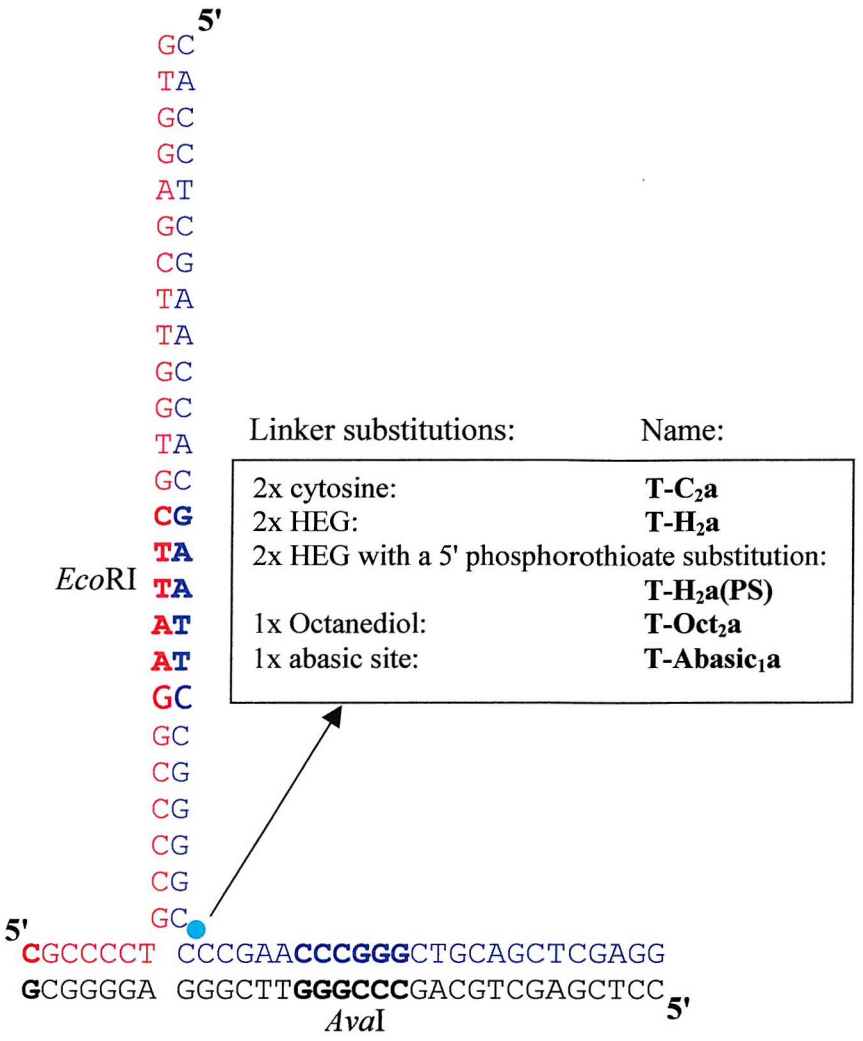


Fig. 6.19. Effect of synthetic methylation on the unconventional activities of *Bam*HI and *Bam*HI methyltransferase. E = *Eco*RI, B = *Bam*HI, A = *Ava*I, L = 10 bp ladder (Gibco/Life Technologies Ltd.), M = 5-methyl cytosine, MT = methyltransferase. The filled circle represents the HEG linker. *Bam*HI site is represented by the grey shaded box. Note that the part of the data presented here is also shown in Fig. 6.14.

a)



b)

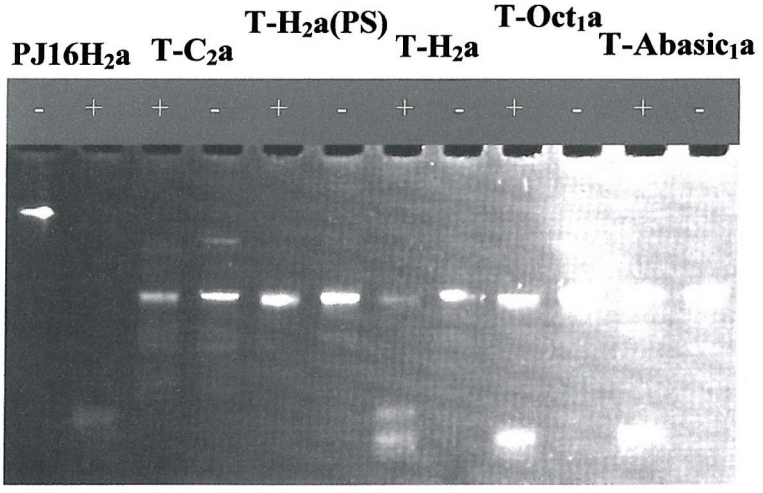


Fig. 6.20. Mobility of PJ16H_2a and truncated PJ16 (T) fragments containing different linkers and incubated in the presence (+) or absence (-) of *Bam*HI. The filled circle indicates the position of the linker. Note that some fragmentation is visible for T-C_2a. This was probably caused by the fact that the template probe (blue strand) solution used for this junction was more than two years old.

from its cognate substrate. In these junctions, the HEG linker was replaced either by two cytosines, an octanediol linker or an abasic site analogue (2'-deoxyribose). These constructs were incubated in the presence and absence of *Bam*HI and analysed on a gel. The results are shown in Fig. 6.20b. Surprisingly, all constructs except where two cytosines replaced the HEG-linker, showed the appearance of a high mobility band similar to the *Bam*HI fragment of PJ16H$_2$a. In other words, the unconventional cleavage activity of *Bam*HI is not specific for HEG, but appears to be specific for the absence of DNA bases in the linker.

6.5 Discussion

6.5.1 Effect of non-nucleosidic linkers on the conformation of PJ36.

The first section of this chapter compared the effect of non-nucleosidic linkers on the conformation of a CFTR-gene based junction (PJ36). The results showed that in most cases no difference was apparent between junctions containing unpaired bases or non-nucleosidic linkers (HEG, octanediol or abasic sites). However, in one case (PJ36 containing a ΔCAC deletion) it was observed that although HEG, octanediol and abasic site substituted junctions showed the same conformation, their conformations were different from the equivalent junction containing unpaired cytosines. It is worth noting that chemical probing of PJ1 (chapter five) showed that the branch point base pairs in the extension probe/target arm were reactive to DEPC and $KMnO_4$. This could explain why the ΔAAA and ΔAAC deletions did not yield an observable effect on PJ36 conformation as the region where this deletion resided was already destabilised. Since no chemical probe reactivity was observed for the arm where the ΔCAC deletion was located, this could also explain why this deletion had a more substantial effect on junction conformation. However, it does not explain the differences observed for different linkers, although the differences in conformation appear to be small as judged from the small mobility differences. Thus, the evidence presented in Fig. 6.2 suggests that in most cases non-nucleosidic linkers act in a similar fashion as unpaired bases, *i.e.* that they facilitate coaxial stacking interactions but do not affect the choice of stacking partners, although subtle differences may occur in some cases. The latter could be a function of linker flexibility as chemically different linkers showed the same conformation, except for (bulky) cytosine linkers in

case of the ΔCAC deletion. Finally, the PJ36 mobility patterns that were obtained when wild-type target was used, showed strong similarities with those obtained for PJ1C_2a, which again supports the R/Y stacking hypothesis discussed in the previous chapter as the R/Y sequence was the same between the two junction whereas the actual nucleobase sequence was not.

6.5.2 Non-conventional recognition and/or cleavage of HEG-containing DNA by restriction endonucleases and *Bam*HI methyltransferase

a) A summary of the results presented in this chapter and the questions they raise.

This chapter has shown that *Bam*HI and *Mse*I are able to cleave linker-containing DNA irrespective of the presence of canonical substrate and in a manner that is dependent on the presence of a flexible linker, lacking DNA bases. This was inferred from the observation that incubation of *Mse*I with radiolabelled PJ1H1c, and incubation of *Bam*HI with radiolabelled PJ16H_2a and PJ38H_2a showed a cleavage product on denaturing polyacrylamide gels when the strand that did not contain the cognate sequence of these enzymes was labelled. In addition, phosphorothioate substitution of the phosphate located directly 5' of the HEG in the template probe in PJ16H_2a, inhibited cleavage of this strand by *Bam*HI. The high mobility bands that resulted from *Mse*I cleavage of PJ1H_1c and *Bam*HI cleavage of PJ16H_2a can be explained on the basis of junction dissociated due to the combination of cleavage of cognate substrate and unconventional, HEG-dependent nicking as outlined in Fig. 6.5 and 6.13. The suggestion that the unconventional nicking activity was independent of the presence of cognate substrate, followed from the observation that *Bam*HI cleaved PJ38H_2a (lacking a *Bam*HI site, Fig. 6.12) and PJ16H_2a where the *Bam*HI site was either synthetically methylated or methylated by *Bam*HI methyltransferase (Fig. 6.18, 6.14).

In addition, indirect evidence for a similar activity by other R-ENases (*Eco*RI, *Ava*I, *Eco*RV, *Dra*I and *Pvu*II) was obtained, which appeared to be accelerated by *Bam*HI methyltransferase treatment. This conclusion was based on the observation that *Ava*I and *Pvu*II treatment of PJ16H_2a and PJ38H_2a respectively showed bands on

native gels that could not be explained on the basis of a conformational change (Fig. 6.7 and 6.16). Also, *Bam*HI methyltransferase treatment stimulated the appearance of these bands, as well as inducing the appearance of similar high mobility fragments for *Eco*RI (PJ16H2a), *Dra*I (PJ38H_2a) and *Eco*RV (PJ38H_2a) (Fig. 6.16). Most of these results can be explained on the basis of template probe nicking by these enzymes. This is schematically outlined in Fig. 6.21. As indicated in this figure, one problem remains the high mobility fragment observed for *Pvu*II treatment of PJ38H_2a-M. This fragment was also observed for non-methylated PJ38H_2a (Fig. 6.16) but was fainter than the other two fragments. In addition, *Ava*I (which cleaves the same arm as *Pvu*II) cleavage of PJ16H_2a yielded three very similar bands in the absence of *Bam*HI methyltransferase (Fig. 6.7, 6.16). The two lower mobility bands can be explained readily on the basis of conventional cleavage of the *Ava*I site and/or additional nicking of template probe strand as shown in Fig. 6.21. However, the highest mobility fragment, like the *Pvu*II fragment of PJ38H_2a and PJ38H_2a-M, cannot be explained on the basis of these criteria alone. Although the similarity with the *Bam*HI fragment of PJ16H_2a is clear from Fig. 6.16, dissociation into two separate 25 bp duplexes of the nicked *Ava*I fragment is unlikely unless additional nicking of the extension probe occurred. Denaturing gel electrophoresis however, clearly showed that this strand in PJ16H_2a was not cleaved by *Ava*I (Fig. 6.9) nor was the same strand in PJ38H_2a cleaved by *Pvu*II (Fig. 6.12). In other words, the explanation for the appearance of these high mobility bands must lie elsewhere. One possibility is that for some reason the remaining duplex slowly dissociates after cleavage and nicking by *Ava*I or *Pvu*II. In other words, the remaining target strand dissociates from the extension probe, leaving a 25 bp duplex with a much longer flap sequence (25 bases) then the *Bam*HI fragment of PJ16H_2a (with a flap of 10 bases). This could also explain why the *Pvu*II high mobility band was much more prominent than the corresponding *Ava*I fragment, since the PJ38H_2a extension probe/target arm was much AT-richer than in PJ16H_2a, and therefore significantly less stable. Nevertheless, in the absence of further analysis, the nature of this fragment remains unsolved.

An additional important issue is: how can branch point methylation induce HEG-dependent cleavage by restriction endonucleases? Since methylation usually inhibits cleavage by restriction endonucleases (or at least reduces the reaction rate substantially), it appears unlikely that methylation directly affected the interaction of *Bam*HI R-ENase with DNA since methyltransferase treated PJ16H_2a and PJ38H_2a

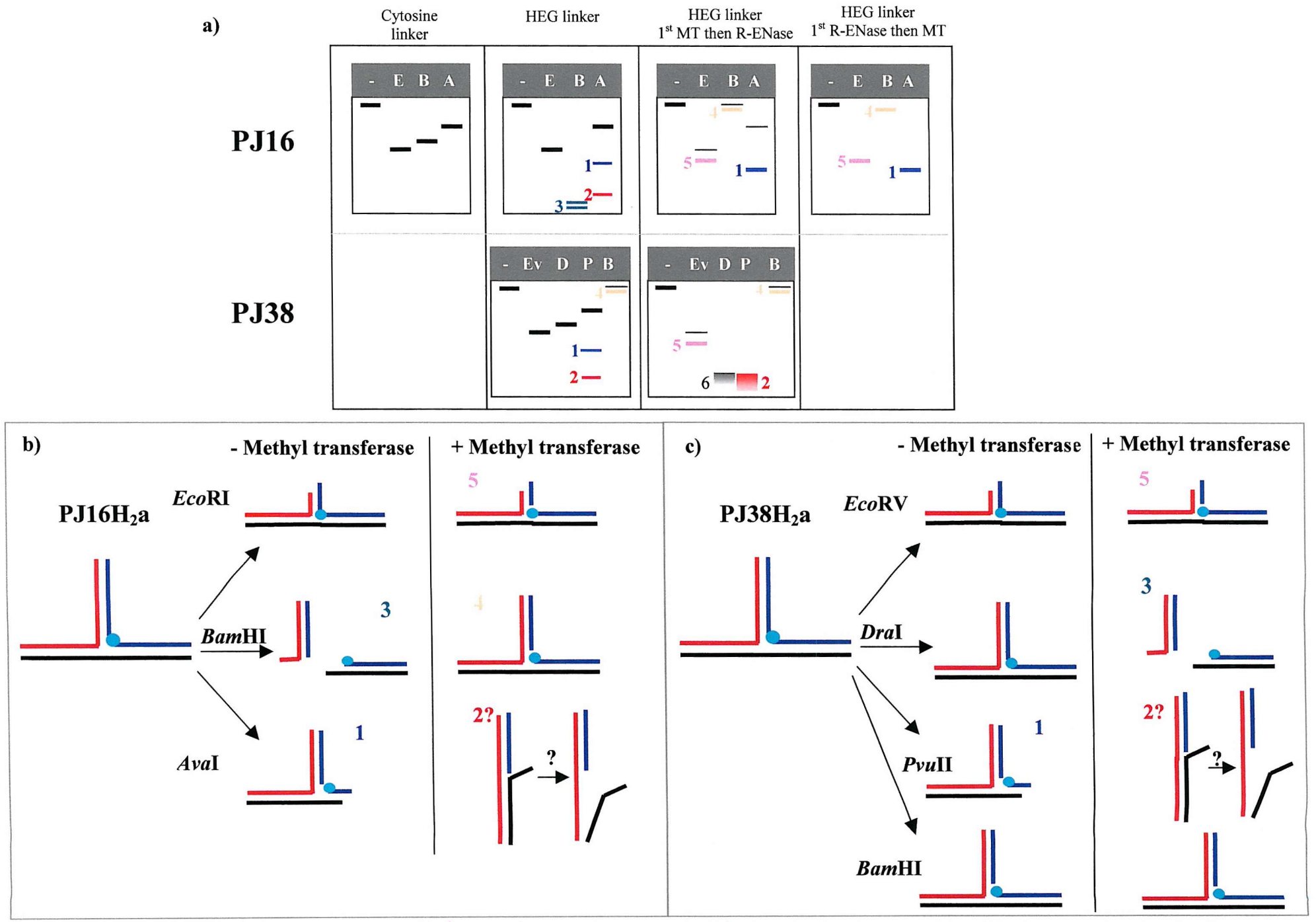

a)
Cytosine linker
HEG linker
HEG linker 1st MT then R-ENase
HEG linker 1st R-ENase then MT
PJ16
PJ38
- E B A
- Ev D P B
b)
- Methyl transferase
+ Methyl transferase
PJ16H2a
EcoRI
BamHI
AvaI
c)
PJ38H2a
EcoRV
DraI
PvuII
BamHI

Fig. 6.21. Possible explanation for the LSA results observed for PJ16H$_2$a and PJ38H$_2$a. (A) Schematic summary of the gel data presented in this chapter. (B) Possible effect of non-nucleoside linkers in PJ16 on the cleavage specificity of the R-ENases tested. In the absence of methyl transferase, only *Bam*HI nicks the template probe (blue) and its cognate sequence, resulting in dissociation of the junction. The latter yields two 25 bp duplexes with 10 base flaps. In the presence of methyl transferase, *Eco*RI, *Bam*HI and *Ava*I nick the template probe strand, yielding fragments of increased mobility (fragments 1, 2, 4 and 5). The nature of fragment 2 remains elusive although one tentative possibility is that it represents a flapped duplex resulting from further melting of the template probe/target (blue/black) arm. However, this fragment migrates between the 40 bp and 50 bp bands of the DNA ladder in Fig. 6.x whereas its actual size would be 50 bp plus 10 overhanging bases. It is possible that the flapped duplex is destabilised (or partially denatured) at the junction, which would give it a more single-stranded character, explaining the enhanced mobility. Additional cleavage of *Ava*I in the extension probe (red) strand could also explain the appearance of this band since this would yield two flapped 25 bp duplexes (similar to *Bam*HI cleavage of PJ16H$_2$a). However, the denaturing gel of PJ16H$_2$a treated with *Ava*I with the extension probe labelled at the 5' end did not show a cleavage product (Fig. 6.9).
(C) For PJ38H$_2$a, only *Bam*HI showed cleavage of the template probe strand when the junction was not treated with methyltransferase. Upon addition of methyl transferase, *Eco*RV yielded a fragment that strongly resembled the *Eco*RI fragment of PJ16H$_2$a-M (fragment 5). *Dra*I yielded a fragment similar to the fastest *Ava*I fragment of PJ16H$_2$a-M. This is surprising since cleavage of the DraI site in conjunction with template probe strand nicking is expected to give a fragment that migrates as the *Bam*HI fragment of PJ16H$_2$a. However, as suggested before, it is possible that the fastest AvaI fragment was highly destabilised and it therefore seems likely that the *Dra*I fragment also represented a dissociation product. Finally, *Pvu*II cleavage of PJ38H$_2$a (not treated with methyl transferase) yielded similar bands as *Ava*I cleavage of MT pre-treated PJ16H$_2$a (fragments 1 and 2). However, upon MT addition the mobility of the fragment increased so that it resembled the *Dra*I fragment. This further supports the suggestion of destabilisation of this fragment as proposed for the fastest *Ava*I fragment of PJ16H$_2$a-M. Finally, *Bam*HI treated of PJ38H$_2$a and PJ38H$_2$a-M yields a nicked TWJ of increased mobility similar to *Bam*HI-treated PJ16H$_2$a-M.

were still susceptible to HEG-dependent nicking of the template probe by *Bam*HI (Fig. 6.16 and 6.18). However, as noted in section 6.4.4 and by other studies (Jen-Jacobson *et al.*, 1996, Rice & Blumenthal, 2000), methylation of only one strand does not necessarily inhibit R-ENase cleavage. Therefore, it remains possible that the cytosine located directly 5' of the HEG was the target of the *Bam*HI methyltransferase. It is also worth noting the similarities in fragment patterns between methyltransferase-treated fragments and those resulting from prolonged R-ENase exposure (Fig. 6.15). These similarities suggest that methyltransferase treatment accelerated cleavage by *Ava*I and *Pvu*II (and possibly *Eco*RI and *Eco*RV) since the time required for the R-ENase to produce visible, enhanced mobility bands was reduced from 15 to less than 3 hours (no shorter incubations were assessed).

The explanations provided in Fig 6.5 and 6.21 can also be applied to the results obtained for PJ1H_1ab. This junction was analysed in section 6.3, and *Mse*I cleavage of this junction yielded two bands (Fig. 6.3f). Moreover, PJ1-derived junctions with a HEG present only in the template probe yielded *Mse*I fragments that migrated faster than the corresponding PJ1C_2a fragment (chapter 4, Fig. 4.2). This raises the possibility that *Mse*I also cleaved junctions with a HEG linker located in other strands than the target strand.

A number of important questions arise from these results:

First, although strong evidence was obtained for unconventional cleavage by *Mse*I and *Bam*HI, can this activity be reconciled with the known high sequence-specifity of R-ENases? Indeed, it appears that different R-ENases interact differently with non-nucleosidic linker containing junctions since some showed an absolute requirement for methyltransferase treatment (*Eco*RI, *Eco*RV), some were only stimulated by methyltransferase treatment (*Ava*I, *Pvu*II) and some did not seem to require methyltransferase at all (*Mse*I, *Bam*HI).

Second, methyltransferases and corresponding R-ENases recognise the same sequences yet there is little sequence or structural homology between these two types of enzymes. This raises the question how the presence of a non-nucleosidic linker is able to overcome these differences and induce binding/cleavage/methylation of nearby located residues.

Third, any explanation must also account for the observation that *Bam*HI methyltransferase can *stimulate/induce* cleavage by enzymes such as *Eco*RI, *Ava*I, *Eco*RV, *Dra*I and *Pvu*II, since this seems in contrast to its natural function, which is to *inhibit* R-ENase cleavage.

The following section addresses these questions by offering a number of possible mechanisms that could have caused the unusual activities observed for the R-ENases and *Bam*HI methyltransferase.

b) Star activity as an explanation for cleavage of non-nucleosidic DNA by R-ENases

One explanation for the observed unconventional cleavage could involve solvent rearrangement caused by the linkers tested, which in turn could induce a star-like activity. Previous studies have shown for instance, that the specificity of restriction endonucleases can be substantially altered by increases in osmotic pressure (Lynch & Sligar, 2000, Robinson & Sligar, 1995, Robinson & Sligar 1998, Sidorova & Rau, 1996, Sidorova & Rau, 2000). The latter however, appears less consistent with the observation that linkers of different chemical composition (HEG, octanediol and abasic site analogues) have the same effect on the cleavage. Nevertheless, star-activity remains an attractive explanation since contacts need to be established between the protein and DNA bases before the enzyme switches to a catalytically competent form. This is further illustrated in Fig. 6.22 showing the co-crystal structures obtained for *Bam*HI and *Eco*RV in the presence of cognate and non-cognate DNA (Newman *et al.*, 1995, Viadiu & Aggarwal, 2000, Winkler *et al.*, 1993, for a review refer to Pingoud & Jeltsch, 1997). What is clear from these structures is that a substantial conformational rearrangement accompanies the binding to specific DNA. Whereas the non-specific complexes of these enzymes are characterised by contacts only between the DNA backbone and the enzyme, the specific complex involves numerous direct and indirect (water and/or cation-mediated) contacts between the enzyme, the DNA backbone and specific bases. In the absence of a specific sequence, dehydration of the DNA surface has been postulated as a way to reduce the number of required specific contacts between the DNA and restriction endonucleases (Robinson & Sligar, 1995, 1998). With respect to the results presented here, dehydration could have been effected by

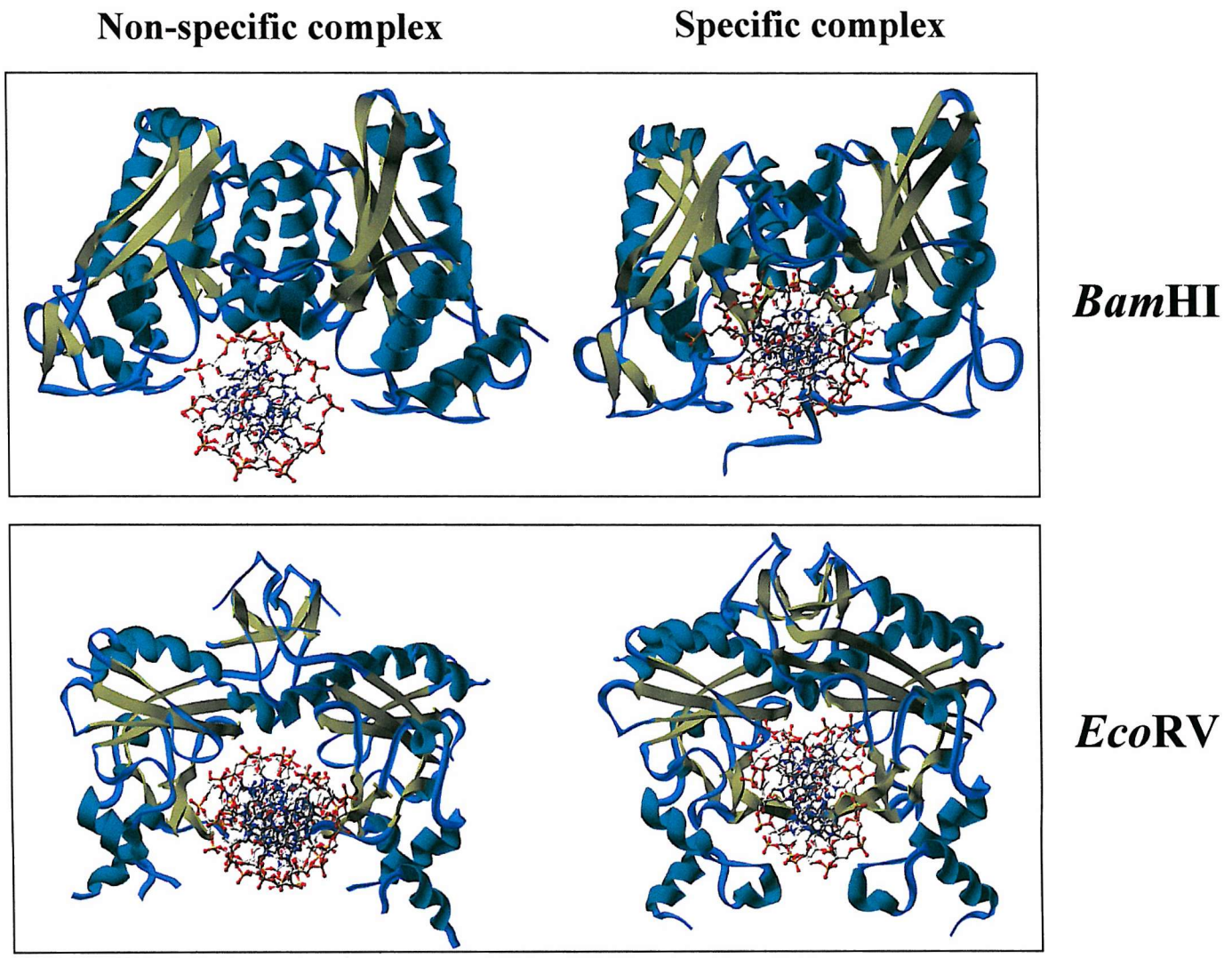


Fig. 6.22. Co-crystal structures of the specific and non-specific complexes of *Bam*HI and *Eco*RV with DNA (Newman *et al.*, 1995, Viadiu & Aggarwal, 2000, Winkler *et al.*, 1993, pdb-entries: 1esg, 2bam, 2rve, 4rve). Note that due to structural disorder parts of the *Eco*RV in the non-specific complex are missing. Both enzymes form a more open conformation when bound to non-specific DNA, and interact only with the phosphate backbone. Upon encountering a specific site, both enzymes assume a more closed conformation, which wraps round the DNA and is mediated by the formation of base-specific contacts.

the hydrophobic character of the different linkers tested. Cytosine methylation, due to the hydrophobic character of the methyl group, could also stimulate dehydration of the DNA (Tippin *et al.*, 1997), possibly explaining why some enzymes appeared to act only on *Bam*HI methyltransferase pre-treated junctions. An interesting experiment would be to incubate PJ16$\underline{C}_2$a with agents that increase the osmotic pressure (*e.g.* ethylene glycol) and then treat it with *Bam*HI. This could provide more answers as to whether a star-like activity is involved.

c) Amino-acid motifs other than those forming the natural catalytic centre as an explanation of non-nucleosidic DNA cleavage by R-ENases

This explanation involves a discovery made in the early 1980's by Behmoaras *et al.* (1981a, 1981b). They found that the tripeptide Lys-Trp-Lys (KWK) was able to bind and cleave apurinic sites in DNA, presumably mediated by tryptophan intercalation. Interestingly, the binding of this peptide was enhanced by methylation of the DNA, and it was postulated that both methylation and the absence of a DNA base facilitated the intercalation of the tryptophan (Behmoaras *et al.*, 1981b). Other short peptide motifs have been discovered that are able to bind and cleave abasic-DNA, including Lys-Tyr-Lys, Lys-Gly-Trp-Lys, Lys-Trp-Gly-Lys, Lys-Ala-Lys and Lys-Lys-Lys (Lhomme *et al.*, 1999 and references therein). Analysis of the protein sequence of *Bam*HI reveals that the C-terminus of this enzyme contains the motif Lys-Trp-Lys (residues 205-207). Fig. 6.23 shows the location of this tripeptide in both subunits in the dimeric *Bam*HI specific and non-specific complexes. The figure shows that the tripeptide is located in a region close to the DNA backbone right opposite the siscile phosphate, but only in one subunit (due to asymmetry in the dimeric structure of the enzyme). This raises the possibility that the observed cleavage by *Bam*HI of a abasic-site containing junction was mediated not by its natural catalytic centre, but by this specific tripeptide sequence. However, sequence analysis of other R-ENases (*Bgl*II, *Eco*RI, *Eco*RV and *Pvu*II) did not reveal the presence of this motif in these enzymes. Although this agrees with the observation that cleavage by several of these enzymes depended on methyltransferase treatment (suggesting that the cleavage proceeded differently for these enzymes), it remains uncertain whether an as yet uncharacterised peptide motif is present in these enzymes that is able to bind and cleave abasic sites in DNA.

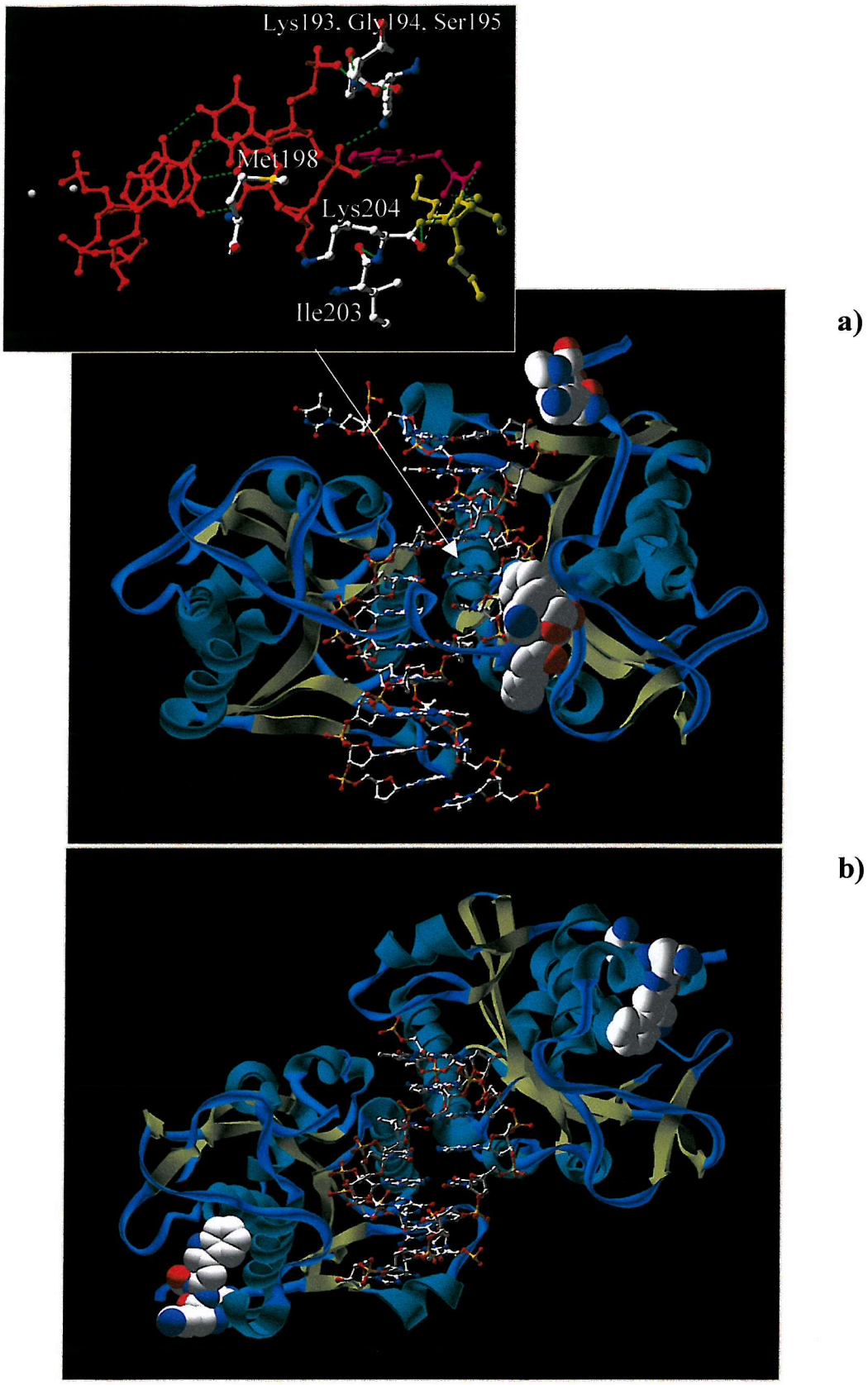


Fig. 6.23. Location of the KWK motif in *Bam*HI. (A) Location of the motif in the specific complex (pdb-entry: 2bam, Newman et al., 1995). The KWK motif is shown in a VDW representation. The inset shows the local neighbours of these three amino acids in a 5Å radius. DNA bases are drawn in red, the tryptophan in purple, the two lysines in gold and the two calcium ions bound to the scissile phosphate are drawn as grey spheres. (B) Location of the KWK motif in the non-specific complex.

It is also worth noting that the Lys-Trp-Lys motif in *Bam*HI is located close to the DNA when bound to specific DNA (Fig. 6.23a). However, in the non-specific complex, these two motifs (one in each subunit) are moved away (Fig. 6.23b). This is an important observation for it could mean that, if the proposed motifs are responsible for anomalous cleavage, formation of a specific protein-DNA complex may not be required. Instead, the enzyme may contact the lesion through non-specific interactions, either through diffusion or by dissociation from the DNA. The latter is a feasible scenario as studies have shown that R-ENases translocate along the DNA via linear diffusion either continuous (Pignoud & Jeltsch, 1997, and references therein) or through a hopping mechanism (*Eco*RV). Interruptions such as triplex DNA, nicks and gaps were shown to stall the enzyme. Thus, it is feasible that a lesion such as an abasic site (or the branch point of a TWJ) stalls the enzyme. Dissociation of the enzyme could then ensue, allowing the lesion-specific motif to bind and cleave the lesion.

Although the possibility presented here remains speculative, it fits most of the data in this chapter. It circumvents the requirement of formation of a specific complex, and explains why different R-ENases vary in their reactivity towards different linkers (since different motifs may be present in different enzymes). It also fits with the interference observed for phosphorothioate substitution as the Lys-Trp-Lys motif was found to cleave the phosphate group attached to the apurinic site (Behmoaras *et al.*, 1981a). The reason why cytosine linkers are not cleaved in PJ1 and PJ16 can be explained on the basis of steric interference with tryptophan intercalation. Finally, the stimulating effect of *Bam*HI methyltransferase treatment is also in line with the observed stimulation of KWK binding to abasic sites by methylation of adjacent residues. The latter was explained on the basis of improved stacking by methylation (Behmoaras *et al.* 1981a), although cytosine methylation can also bend the DNA (Marcourt *et al.*, 1999, Mayer-Jung *et al.*, 1998), which could facilitate tryptophan intercalation.

d) *Bam*HI methyltransferase?

One issue that does not fit easily with either of the two explanations provided, is the observed interaction between *Bam*HI methyltransferase and non-nucleosidic linker DNA. This puzzling observation is difficult to reconcile with the substantial

differences that are known to exist between prokaryotic methyltransferases and restriction endonucleases with respect to their catalytic centres and DNA binding domains. In other words, the observation that both R-ENases and a methyltransferase act on similar targets in an unconventional manner is highly surprising. The DNA binding domains and specific binding modes are substantially different for both types of enzymes. The structures of methyltransferases are not as well studied as those of restriction endonucleases. Nevertheless, it is clear that they contain two separate domains, a catalytic domain and a DNA binding domain. The catalytic domain consists of residues that are thought to facilitate the extrusion of the target base into a hydrophobic pocket of the enzyme (base-flipping) to allow methylation to occur. This striking activity has been demonstrated for a number of methyltransferases by crystallography (Kilmasauskas *et al.*, 1994, Reinisch *et al.* 1995, O'Gara *et al.*, 1998). Kinetic studies have further revealed that binding to sequences containing an abasic site or mismatch in the cognate sequence enhances binding of methyltransferases (O'Gara *et al.*, 1998, Wang *et al.*, 2000). However, little is known about their interactions with non-cognate sequences containing abasic sites, although some studies have been conducted assessing whether or not these methyltransferases can display a star-like activity (*i.e.* formation of a specific complex in the absence of a specific sequence). Smith *et al.* (1992) for instance, found that *Eco*RI methyltransferase is able to bind and methylate non-cognate sequences *in vivo*, but only when this enzyme was overexpressed. In addition, human (C5) methyltransferase has a high affinity for unusual DNA conformations, including cruciforms and slipped-strand structures, containing its cognate sequence (CG) at the deformation (Laayoun & Smith, 1995, Smith *et al.*, 1991, Smith *et al.*, 1992, Kho *et al.*, 1998). In conjunction with the observed higher affinity for abasic-site containing sites than for cognate sequences, this provides support for the possibility that a star-like activity could underlie the observations presented here for *Bam*HI methyltransferase.

Finally, no KWK motif was observed in the sequence for this enzyme, which appears unlikely anyway since no indications were obtained for the existence of a cleavage activity by *Bam*HI methyltransferase.

6.6 Concluding remarks

This chapter has revealed a novel, highly unconventional cleavage activity of restriction endonucleases, which is dependent on the presence of a non-nucleosidic linker but not on the presence of canonical substrate. In addition, evidence was obtained suggesting that *Bam*HI methyltransferase is also able to interact with DNA containing a non-nucleosidic linker but not its canonical substrate.

Although some suggestions were offered with respect to the underlying mechanism, including star-activity or the presence of a motif other than the natural catalytic sites of these enzymes, further experiments are required to substantiate these.

7 General Conclusions

Mutation detection to achieve better disease prognosis and diagnosis has important clinical value. Developments over the past five years have provided a wealth of new data regarding genetically controlled disease, highlighting that most common disorders such as cancer, are complex genetic diseases. The diagnosis of such diseases require highly accurate methods as they must be able to detect even a single base change within the 3 billion pairs present in the human genome. Since most genetic diseases probably involve multiple mutations, increasing emphasis on parallel screening methods is demanded. Chapter one outlines in detail the current short comings, from a clinical point of view, of most techniques available, and offers a novel alternative: Signal Mediated Amplification of RNA Technology (SMART). The development of this technology in order to achieve highly sensitive discrimination between wild-type and affected target sequences is the main theme that is pursued throughout this thesis. The technology involves two single stranded DNA probes, called the extension and template probe, which bind to a specific target sequence (DNA or RNA) and to each other such that a three-way junction is formed. Full details on the technology as well as a discussion on its potential advantages are provided in the first chapter.

This thesis has explored the role of the conformation of the three-way DNA junction on this assay. Previous studies have shown that the conformation is highly sequence-dependent, but that at the same time usually only two conformers exist, designated conformer I and II (Guo *et al.*, 1990, Leontis *et al.*, 1991, Lu *et al.*, 1991, Welch *et al.*, 1995, Overmars *et al.*, 1996). Structural investigations revealed that these conformers are characterised by coaxial stacking interactions, connecting two of the three arms. A requirement for these interactions to occur is however, the presence of unpaired bases and positively charged ions such as magnesium. Since SMART usually incorporates a linker at the branch point in the template probe and/or extension probe, and since the polymerases require cations for activity, coaxial stacking interactions are likely to occur in the SMART TWJs. This leads to the possibility of steric hinderance by the TWJ for the DNA polymerase. A co-

crystal structure of the DNA polymerase used in SMART, shows that the enzyme wraps around approximately 9 bp of DNA (Kiefer *et al.*, 1998). Since the overlap region is short, 7-8bp, the close proximity of two additional DNA arms could interfere with polymerase binding. If so, a reduction in RNA yield could be the result, which not only affects the sensitivity of the assay but also the potential levels of discrimination between wild-type and mutated targets. Finally, since coaxial stacking interactions occur in TWJs these may affect the thermodynamics, and thus mutation sensitivity in unpredicted ways since TWJs remain relatively poorly characterised. Therefore, the problem of the TWJ-yield and TWJ-mutation sensitivity relationships constituted the specific focus of this thesis.

The mutation sensitivity of TWJs was assessed in chapter three, by designing targets containing small mutations (three-base deletions and single base pair mismatches) and testing various probe designs based on these targets. The idea was that probes designed to detect wild-type CFTR gene would deliver a high RNA yield in the presence of wild-type target and a very low/absent yield when a mutation was present. In addition, DNA analogues were assessed that were reported to show increased mutation sensitivity. The two analogues assessed were PNA (peptide nucleic acids) and LNA (locked nucleic acids). Experiments revealed that a junction containing short target complementary sequences in the probes and with PNA substituting the DNA sequence in this region, yielded accurate discrimination between wild-type and three-base and single base variants of the cystic fibrosis transmembrane conductance regulator gene. These experiments further revealed that the branch point appeared the least sensitive, which is in accordance with previous studies showing that unpaired bases can have a stabilising effect on the structure (Leontis *et al.*, 1991, Welch *et al.*, 1995). However, a full DNA version of this junction showed substantially poorer discrimination, indicating that indeed, PNA is more sensitive for mutations (Eriksson & Nielsen, 1996). Nevertheless, reducing the size of the DNA template probe target complementary region was found to substantially improve the discrimination, close to the level observed for the PNA-containing junction. LNA on the other hand never showed a positive effect with respect to the level of discrimination, only equal or less than observed for DNA probes.

Finally, since the targets tested thus far, were synthetic polynucleotides, a more complex situation was created by adding DNA probes specifically designed for

detection of the wild-type CFTR gene to PCR-amplified homozygous and heterozygous genetic material. The mutation assessed was ΔF508, the most common Caucasian CFTR gene mutation. Accurate discrimination between wild-type, heterozygous and homozygous material was achieved. Together, these results established that SMART is sensitive for mutations as small as a single base change using both synthetic single stranded DNA as well as more complex, pre-amplified double stranded DNA as target. In addition, although PNA appeared more sensitive than DNA, by modifying the junction design DNA probes could made nearly as sensitive for mutations as PNA. The advantage of this is mainly economical as DNA is substantially cheaper to produce and exploit since additional licensing is required when PNA is commercially used.

Chapter three described a further improvement in the development of SMART. A single base change, causative of deep venous thrombosis, was assessed. The previous results already showed that discrimination between wild-type and single base changes was possible using SMART. These single base changes however were not common mutations, or even known to cause disease. They were simply generated to assess the feasibility of mutation detection with SMART. Now that the feasibility was established, it was of interest to apply SMART to assess whether it would be able to detect a clinically relevant single base pair mismatch. The mismatch chosen is usually referred to as Factor V Leiden, a G to A mismatch leading to deep venous thrombosis. The initial results showed very poor discrimination between wild-type and mutant target, using probes designed to detect the wild-type sequence. However, subsequent experiments revealed that a reduction in the overlap length from 8 to 7 bp substantially increased mutation sensitivity to the levels observed for the CFTR gene in chapter three. This finding illustrates the complex nature of the three-way DNA junction as a change in one arm can affect the stability of a different arm (noting that no apparent reduction in RNA yield occurred). Interestingly, the probes with PNA as the target complementary sequence showed sensitive discrimination even when using an 8 bp overlap. The reason for this difference remains unclear since both the higher sensitivity for mismatches as well a different conformational effect could have been the cause.

Finally, the DNA probes were redesigned so that they became specific for the Factor V Leiden target to show that mutations could be detected as readily as wild-type sequences. The results revealed a specific signal when the mutation was present but not when wild-type target was applied, verifying again the sequence-specificity of SMART.

Together the results presented in chapters three and four show that accurate discrimination between target sequences varying in only a few base pairs can be achieved with SMART. Probe design and TWJ conformation have been revealed as important parameters affecting the accuracy of SMART. Of these two parameters, probe design was assessed most thoroughly by introducing small, systematic changes that are known to affect standard DNA hybridisation. These optimisations are applicable to different target sequences, *i.e.* no or very little target-specific optimisations were required in order to achieve accurate discrimination between different target sequences. Nevertheless, most target sequences used were synthetic rather than drawn directly from affected and non-affected persons, and therefore represented an idealised situation. The latter constitutes the next major target and depends on improving the sensitivity of additional RNA amplification steps, outlined in chapter one, and current efforts are under way at Cytocell focussing on this issue to allow the direct detection of mutations in patient samples. Other important issues that need to be addressed include multiplexing and consistency improvements. Multiplexing is becoming increasingly important in mutation screening as it substantially reduces work-load and allows for high-throughput parallel screening of multiple samples at once (refer to the introduction). How SMART can be adapted to multiplexing remains to be seen. Although the transcription (=amplification) sequence in the template probe and amplification probes could be altered so that each target can be addressed with a unique set of probes, this seems a highly inefficient approach for the detection of multiple targets in a single tube. Other possibilities therefore, need to be explored such as (perhaps) chip-based assays.

With respect to consistency, chapter three revealed a substantial variation in absolute yields between identical experiments. Reflecting on the fact that only the first step of the assay was involved, it seems likely that any variation produced in this step will be amplified as soon as one or two additional amplification probes are added. Thus, it is clear that future efforts will have to overcome these problems.

Chapter three also showed that the conformation of the TWJ can play an important role, though this was not systematically assessed. The systematic analysis of junction folding was described in chapters four and five. The main objective was to confirm that junction structure can affect the RNA yield in SMART, and if so, to improve the understanding of the sequence-dependency of junction folding. In this respect it is useful that only two conformations are plausible. For reasons outlined in detail in chapter one, conformer I was hypothesised to yield less RNA in SMART due to steric interference with DNA polymerase binding, whereas conformer II TWJs were hypothesised to yield more RNA.

Previous studies revealed that only the sequences close to the branch point contribute to the coaxial stacking arrangement of the junction (Lu *et al.*, 1991, Welch *et al.*, 1995). The important, and probably most complex issue that needs to be resolved is how changes in sequence affect the conformation of the junction. The approach adapted in chapters four and five was one involving comparison of various junctions differing only by one or at most a few base pairs around the branch point. A comparative electrophoresis technique, called the long-short-arm assay first developed by Cooper & Hagerman (1987) was deployed. This technique has provided a wealth of structural information with respect to three-way junctions but also for its close structural cousin, the four-way or Holliday junction. In addition, it provided the relative high-throughput ability required for screening of multiple junction conformations in a relatively short period of time, an important advantage over for instance NMR and crystallography. To assess the influence of TWJ conformation on SMART, chapter four details the optimisation of the long-short arm assay. It shows that the assay as used here reproduced the conformations established by others for a number of different junctions. In addition, chapter five described the conformation of the CFTR-based junction discussed in chapter three, which appeared to reside in an unstacked, Y-shaped conformation or in a rapid equilbrium such as observed for FWJs. Both a Y-shaped conformation and a rapid equilibrium could explain the lower yields of this junction in comparison to the junctions used for the Factor V target. Nevertheless, the conformation did not appear to be substantially altered by an increase in the number of unpaired bases at the branch point, apart from decreasing the angle between the overlap and extension probe/target arms.

Chapter four further investigated a potential pit-fall inherent to the long-short arm assay. All applications of the long-short arm assay thus far incorporated sites for restriction endonucleases that leave a 5' four base overhang after cleavage. However, the results presented in chapter 5 revealed that these overhanging sequences can significantly affect the mobility pattern obtained for a TWJ. This effectively invalidates the interpretation of the assay, as one of its underlying assumptions is that these sequences do not affect mobility. Subsequent experiments confirmed that the change in mobility was not an artefact and appeared most severe when GC-rich sequences were involved and was temperature dependent. The results obtained corroborate those found by Muizniek & Doerfler (1998) and Woodcock *et al.* (1991) who noted that a number of duplexes with single stranded overhangs migrated anomalously through gels. The reason why these sequences caused the anomalous mobility remains unknown, although it appears to be dependent on temperature, cations and sequence. Despite these problems in interpreting the unusual gel migration behaviour, the anomalies can be avoided by using blunt cleaving restriction endonucleases, mung bean nuclease treatment and/or increasing the electrophoresis temperature. The first of these solutions was applied in the remainder of the thesis.

Chapter five addressed the relationship between TWJ conformation and RNA in the SMART assay. The conformation of a junction based on the junction used for detection of the Factor V target, described in chapter three, was assessed. The conformation appeared to correspond to a conformer II, which is in accordance with the higher yields produced when compared to the junction that was based on the CFTR target.

To further improve the relationship between TWJ structure and SMART yields, junctions were designed that were based on the hypothesis that conformer I and conformer II share the same stacking preferences. This hypothesis is in sharp contrast to suggestion made by Welch *et al.* (1995) who noted that the two conformers are stereochemically different and could therefore show differences in stability. To test this hypothesis a novel, "conformer conversion" experiment was performed. This experiment essentially creates two junctions both containing the same branch point sequences in each arm, but positioned such that in one junction a conformer I is expected whereas in the other a conformer II is expected to result. This is achieved

by simply swapping two arms. This simple experiment would provide a strong indication of the influence of stacking interactions over the influence of the backbone and solvent. Successful conformer conversion (*i.e.* the conformer I is converted to conformer II by arm swapping), would suggest that internal stacking interactions dominate the conformer preference. In relation to this, we tested the hypothesis that R/Y sequence rather than actual DNA base sequence governs the folding of TWJs. In B-DNA RR, RY and YR steps each have quite distinct structural properties (Hunter, 1993, Dickerson, 1999). In other words, AA is similar to GG (both are RR sequences), TA is similar to CG (both are YR sequences) *etc.*

Experiments designed for testing these hypotheses revealed a number of important insights. First, conformer I junctions do appear to yield significantly less RNA than conformer II junctions. Second, exploration of the R/Y stacking hypothesis has shown that most junctions are not affected by changing one purine to the other (*i.e.* A to G or G to A), and that at least two base pairs at the branch point, in each arm, contribute to junction conformation. An analysis of junctions described here and in the literature revealed a high level of consistency with the R/Y stacking hypothesis, and moreover, supported the hypothesis that conformer I and II junctions share the same coaxial stacking preferences. The conformer conversion experiments resulted in conversion in most cases. However, a number of junctions assumed an unusual conformation, and a number of junctions failed the R/Y stacking hypothesis. Interestingly, some of these conformations appeared to correspond to novel TWJ fold, termed "reverse conformer I" or "reverse conformer II" as the orientation of the unstacked arm with respect to the coaxial stack appeared to be reversed. This conformation has only been previously characterised as part of a DNA pseudosquare knot structure (Ulyanov *et al.*, 1999). A substantial A-type character was apparent for the knot structure, and appeared strained due to the presence of two closely spaced junction points. It nevertheless shows that in principle, the conformation is possible. Also, the results obtained are not consistent with a difference in the structural equilibrium between conformer I and II. A decrease in exchange rate would be expected to result in a smaller differences in fragment mobilities. An increase in exchange rate on the other hand, would be expected to increase the mobility differences but not to alter the fast-to-slow order of the fragments. Therefore, a reverse conformer appears the most satisfactory explanation.

A small number of other junctions appeared to show an unfolded form, but a case was made in chapter five, arguing for the existence of a dynamic conformational equilibrium as observed for FWJs (Miick *et al.*, 1995, Overmars & Altona, 1997) and suggested for TWJs (Welch *et al.*, 1995).

The structural similarities with FWJs may have important implication for our understanding of recombination, as it facilitates the prediction of junction conformation based on sequence. Moreover, the importance of purine/pyrimidine identity and the results for the conformer conversion approach, provide a mechanistic explanation for junction folding in much greater detail than before. There has only been one tentative attempt to predict FWJ structure based on sequence. Azaro & Landy (1997) postulated that FWJs tend to maximise the number of purines in the exchanging strands relative to the continuous strands but did not provide a mechanistic explanation for this apparent trend. The data presented in this thesis show that junctions containing more purines in the continuous strand than in the exchanging strand occur frequently. Nevertheless, the rules dictating FWJ structure may well be different from those controlling TWJs, even if only marginally, since a FWJ is effectively a result of two stacking optimisations whereas in TWJs of course, only one coaxial stack is formed.

In an attempt to increase our understanding of how R/Y sequence affects junction structure, we examined junctions containing predicted B-DNA structures. Recent studies on DNA conformation have revealed that B-DNA duplexes have two major degrees of freedom, one dictated by a slide and one by shift (Lu & Hunter, 1997). Knowing the values of these two parameters allowed Packer *et al.* (2000a, 2000b) to determine with reasonable accuracy, the values of the remaining base pair parameters (*e.g.* twist, roll). From a theoretical analysis of various three-way junction structures, and using data generated by Packer *et al.* (2000a, 2000b), evidence was obtained suggesting that low energy sequences with a rigid character (*i.e.* resistant to slide, positive or negative) appeared substantially more often than flexible sequences. Based on these observations, six junctions were designed to assess to what extend slide-flexibility contributed to junction conformation. However, the results revealed unusual conformations in most cases, with at least three appearing to correspond to a

"reverse conformer I or II". Nevertheless, these results indicate that slide is probably not the only factor involved in coaxial stacking.

Most evidence presented in chapter six supports the hypothesis that conformer I junctions yield less RNA in SMART than conformer II junctions. Combined with the analysis of sequence-dependent junction folding, this means that target sites can be more sensibly chosen to provide an optimal junction conformation for SMART assays. Nevertheless, the analysis of junction folding is far from complete. The proposed hypotheses fail to explain the results obtained for a number of junctions, and it is crucial to subject these to higher resolution analysis. Another persistent unknown remains the existence of a conformational equilibrium. Overmars *et al.* (1995) specifically noted that no evidence for an alternative conformer was found, in a NMR study on their TWJ1 and TWJ2, and neither have FRET studies (Stuhmeier *et al.*, 1997a, 1997b). Nevertheless, the strong conformational similarities between TWJs and FWJs and the results presented in this thesis make it likely that TWJs also reside in a conformational equilibrium.

In conclusion, the results presented in chapter five revealed a number of important structural rules underlying TWJ folding, by comparing TWJ stacking behaviour with detailed B-DNA stacking behaviour. Importantly, TWJ structure was found to have a significant influence on RNA yield in SMART assays.

Chapter six discusses an unusual observation regarding restriction endonuclease cleavage of sequences containing non-nucleosidic linkers. These results showed that junctions and duplexes, when containing a flexible linker such as hexaethylene glycol, octanediol or an abasic site, can be cleaved by a number of different restriction endonucleases. This cleavage was independent of the presence of canonical substrate, and involved the 5' phoshate group connecting the linker with the adjacent 5' base. However, details regarding the DNA binding and cleavage mechanism remain largely unsolved. Some similarities with star-activity are apparent and it is possible that the linker causes a change in the hydration of DNA, which is known to stimulate star-activity. Also, domains other than the catalytic centre could be involved in the unconventional cleavage as for *Bam*HI a tripeptide motif was discovered, which is known to be able to bind and cleave abasic DNA.

In addition to restriction endonucleases, *Bam*HI methyltransferase was also found to interact with linker-containing TWJs, again independently of the presence of canonical substrate. The results obtained did not reveal how the methyltransferase interacted with the junction, but methylation of one (or more) cytosine(s) seems a likely possibility. They revealed in stead that treatment with methyltransferase increased susceptibility of the junction for cleavage of the P5'-HEG bond by restriction endonucleases. Further kinetic and structural studies are essential to further substantiate the suggestions offered here.

References

Allawi, H. & SantaLucia, J. (1998). Nearest-neighbor thermodynamics of internal A-C mismatches in DNA: sequence dependence and pH effects. *Biochemistry* **37**, 9435-9444.

Altare, F., Jouanguy, E., Lamhamedi, S., Doffinger, R., Fischer, A. & Casanova, J. (1998). Mendelian susceptibility to mycobacterial infection in man. *Curr. Opin. Immun.* **10**, 413-417.

Altona, C., Pikkemaat, J.A. & Overmars, F.J.J. (1996). Three-way and four-way junctions in DNA: a conformational viewpoint. *Curr. Opin. Struct. Biol.* **6**, 305-316.

Azaro, M.A., Landy, A. (1997) The isomeric preference of Holliday junctions influences resolution bias by λ integrase. *EMBO J.* **16**(12), 3744-3755.

Bailly, C., & Waring, M. (1997) *In: Methods Mol. Biol.,* vol. 90 (Fox, KR, ed.), Humana Press Inc, Totowa, USA, 51-79.

Behmoaras, T., Toulmé, J. & Hélène, C. (1981a). A tryptophan-containing peptide recognizes and cleaves DNA at apurinic sites. *Nature* **292**, 858-859.

Behmoaras, T., Toulmé, J. & Hélène, C. (1981b). Specific recognition of apurinic sites in DNA by a tryptophan-containing peptide. *Proc. Natl. Acad. Sci. USA* **78**, 926-930.

Behn, M. & Schuermann, M. (1998a). Sensitive detection of p53 gene mutations by a 'mutant enriched' PCR SSCP technique. *Nucleic Acids Res* **26**, 1356-1358.

Behn, M. & Schuermann, M. (1998b). Simple and reliable Factor V genotyping by PNA-mediated PCR clamping. *Thromb. Haemost.* **79**, 773-777.

Bellamy, R. & Hill, A. (1998). Genetic susceptibility to mycobacteria and other infectious pathogens in humans. *Curr. Opin. Immun.* 10(1998), 483-487.

Berman, H.M. (1997). Crystal studies of B-DNA: the answers and the questions. *Biopolymers* **44**, 23-44.

Bertina, R., Koelman, B., Koster, T., Rosendaal, F., Dirven, R., de Ronde, H., Van der Velden, P. & Reitsma, P. (1994). Mutation in blood coagulation Factor V associated with resistance to activated protein C. *Nature* **369**, 64-67.

Biebricher, C. & Luce, R. (1996). Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase. *EMBO J* **15**, 3458-3465.

Biesecker, B. & Marteau, T. (1999). The future of genetic counselling: an international perspective. *Nature Genet.* **22**, 133-137.

Boucher, J., Yu, H., Mudd, M. & Deretic, V. (1997). Mucoid *Pseudomonas aeruginos*a in cystic fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infect. Immun.* **65**, 3838-3846.

Bradley, L. A., Johnson, D. A., Chaparro, C. A., Robertson, N. H. & Ferrie, R. M. (1998). A multiplex ARMS test for 10 cystic fibrosis (CF) mutations: evaluation in a prenatal CF screening program. *Genet. Test.* **2**, 337-341.

Brown, S. C., Thomson, S. A., Veal, J. M. & Davis, D. G. (1994). NMR solution structure of a peptide nucleic acid complexed with RNA. *Science* **265**, 777-80.

Bustin, S. & Dorudi, S. (1998). Molecular assessment of tumour stage and disease recurrence using PCR based assays. *Mol. Med. Today* **Sept**, 389-396.

Calladine, C. & Drew, H. (1997). *Understanding DNA: the molecule and how it works*. 2nd edit, Academic Press Ltd., London.

Cardy, D.L.N. & Delnatte, S.Y.J. (1991). Nucleic acids extension assay. Patent (WIPO): WO9306240A1.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C., Lim, E., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedlan, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G. & Lander, E. (1999). Characterisation of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* **22**, 231-238.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. *Science* **274**, 610-614.

Claverys, J., Prudhomme, M., Mortier-Barriere, I. & Marin, B. (2000). Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? *Mol. Microbiol.* **35**, 251-259.

Collins, F., Guyer, M. & Chakravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580-1581.

Compton, J. (1997). Nucleic acid sequence-based amplification. *Nature* **350**, 91-92.

Cookson, W. (1999). The alliance of genes and environment in asthma and allergy. *Nature* **402 (Supp)**, B5-B11.

Cooper, J. P. & Hagerman, P. J. (1987). Gel electrophoretic analysis of the geometry of a DNA four-way junction. *J. Mol. Biol.* **198**, 711-719.

Cremonesi, L., Carrera, P., Fumagalli, A., Lucchiari, S., Cardillo, E., Ferrari, M., Righetti, S. C., Zunino, F., Righetti, P. G. & Gelfi, C. (1999). Validation of double gradient denaturing gradient gel electrophoresis through multigenic retrospective analysis. *Clin. Chem.* **45**, 35-40.

Cronin, M. T., Fucini, R. V., Kim, S. M., Masino, R. S., Wespi, R. M. & Miyada, C. G. (1996). Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. *Hum. Mutat.* **7**, 244-255.

Cross, G. (1996). Antigenic variation in trypanosomes: secrets surface slowly. *BioEssays* **18**, 283-291.

Day, I. N., O'Dell, S. D., Spanakis, E. & Weavind, G. P. (1999). Microplate array diagonal gel electrophoresis (MADGE), CpG-PCR and temporal thermal ramp-MADGE (Melt-MADGE) for single nucleotide analyses in populations. *Genet. Anal.* **14**, 197-204.

Dean, M., Carrington, M., Winkler, C., Hutley, G., Smith, M., Allikmets, C., Goedert, J., Vittinghoff, E., Gomperts, E., Donfield, S., Vlahov, D., Kaslow, R., Saah, A., Rinaldo, C., Detels, R., study, H. g. a. d., study, M. A. c., study, M. H. c., cohort, S. F. c., study, A. & O'Brien, S. (1996). Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science* **273**, 1856-1862.

Deitsch, K., Moxon, E. & Wellems, T. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal and fungal infections. *Microbiol. Mol. Biol. Rev.* **61**, 281-293.

Diaz, J., Virji, M. & Heckels, J. (1984). Structural and antigenic differences between two types of meningococcal pili. *FEMS Microbiol. Lett.* **21**, 181-184.

Dickerson, R. (1999). Helix structure and molecular recognition by B-DNA. In *Oxford handbook of nucleic acid structure* (Neidle, S., ed.), pp. 145-197. Oxford University Press, Oxford.

Duckett, D. R. & 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.

Eichman, B.F., Vargason, J.M., Mooers, B.H.M., Shing Ho, P. (2000). The Holliday junction in an inverted repeat DNA sequence: sequence effects on the structure of four-way junctions. *Proc. Natl. Acad. Sci. USA* **97**, 3971-3976.

El Hassan, M.A. & Calladine, C.R. (1996). Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA. *J. Mol. Biol.* **259**, 95-103.

Eng, C. & Vijg, J. (1997). Genetic testing: the problems and the promise. *Nature Biotechn.* **15**, 422-426.

Eriksson, M. & Nielsen, P. E. (1996). PNA-nucleic acid complexes. Structure, stability and dynamics. *Q. Rev. Biophys.* **29**, 369-394.

Evans, W.E., Relling, M.V. (1999). Pharmacogenomics: translating functional genomics into rational therapeutics. *Science* **286**, 487-491.

Ferrie, R. M., Schwarz, M. J., Robertson, N. H., Vaudin, S., Super, M., Malone, G. & Little, S. (1992). Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *Am. J. Hum. Genet.* **51**, 251-262.

Fischer, S. G. & Lerman, L. S. (1979). Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell* **16**, 191-200.

Fox, J., England, J., White, P., Ellison, G., Callaghan, K., Charlesworth, N., Hehir, J., McCarthy, T., Smith-Ravin, J., Talbot, I. & al, e. (1998). The detection of k-ras mutations in colorectal cancer using the amplification-refractory mutation system. *Br. J. Cancer* **77**, 1267-1274.

Gandrille, S., Greengard, J., Alhenc-Gelas, M., Juhan-Vague, I., Abgrall, J., Jude, B., Griffin, J. & Aiach, M. (1995). French network on the behalf of INSERM: Incidence of activated protein C resistance caused by the ARG 506 GLN mutation in Factor V in 113 unrelated symptomatic protein C-deficient patients. *Blood* **86**, 219-224.

Gelfi, C., Righetti, S. C., Zunino, F., Della Torre, G., Pierotti, M. A. & Righetti, P. G. (1997). Detection of p53 point mutations by double-gradient, denaturing gradient gel electrophoresis. *Electrophoresis* **18**(15), 2921-2927.

Gilles, P. N., Wu, D. J., Foster, C. B., Dillon, P. J. & Chanock, S. J. (1999). Single nucleotide polymorphic discrimination by an electronic dot blot assay on semiconductor microchips. *Nature Biotechn.* **17**(4), 365-370.

Giunta, C., Youil, R., Venter, D., Chow, C. W., Somers, G., Lafferty, A., Kemper, B. & Cotton, R. G. (1996). Rapid diagnosis of germline p53 mutation using the enzyme mismatch cleavage method. *Diagn. Mol. Pathol.* **5**(4), 265-270.

Griffin, T. & Smith, L. (2000). Single-nucleotide polymorphism analysis by MALDI-TOF mass spectroscopy. *Trends Biotechn.* **18**, 77-84.

Gruenheid, S. & Gros, P. (2000). Genetic susceptibility to intracellular infections: Nramp1, macrophage function and divalent cation transport. *Curr. Opin. Microbiol.* **3**, 43-48.

Guo, Q., Lu, M., Churchill, M., Tullius, T. & Kallenbach, N. (1990). Asymetric structure of a three-arm DNA junctions. *Biochemistry* **29**, 10927-10934.

Guo, Z., Liu, Q. & Smith, L. (1997). Enhanced discrimination of single nucleotide polymorphisms by artificial hybridization. *Nature Biotechn.* **15**, 331-335.

Hacia, J. G., Brody, L. C., Chee, M. S., Fodor, S. P. & Collins, F. S. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nature Genet.* **14**(4), 441-447.

Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. (1997). Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089-1097.

Halushka, M., Fan, J., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. & Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood pressure homeostasis. *Nature Genet.* **22**, 239-247.

Hartmann, B. & Lavery, R. (1996). DNA structural forms. *Q. Rev. Biophys.* **29**, 309-368.

Henderson, B., Poole, S. & Wilson, M. (1998). *Bacteria-cytokine interactions in health and disease*, Portland Press Ltd., London.

Henderson, B., Wilson, M., McNab, R. & Lax, A. (1999). *Cellular microbiology: bacteria-host interactions in health and disease*, John Wiley & Sons Ltd., Chichester.

Hietala, M., Aula, P., Syvanen, A. C., Isoniemi, A., Peltonen, L. & Palotie, A. (1996). DNA-based carrier screening in primary healthcare: screening for aspartylglucosaminuria mutations in maternity health offices. *Clin. Chem.* **42**(9), 1398-1404.

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.

Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D. & Ke, H. (2000) Crystal structure of *Nae*I-an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J.* **19**, 3110-3118.

Humphries, S., Gudnason, V., Whittal, R. & Day, I. (1997). Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholestrolemia. *Clin. Chem.* **43**, 427-435.

Hunter, C. (1993). Sequence-dependent DNA structure: the role of base stacking interactions. *J. Mol. Biol.* **230**, 1025-1054.

Hunter, C. (1996). Sequence-dependent DNA structure. *BioEssays* **18**, 157-162.

Hunter, C. & Lu, X. (1997). DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotide X-ray crystal structures. *J. Mol. Biol.* **265**, 603-619.

Igloi, G. L. (1998). Variability in the stability of DNA-peptide nucleic acid (PNA) single-base mismatched duplexes: real-time hybridization during affinity electrophoresis in PNA-containing gels. *Proc. Natl. Acad. Sci. U S A* **95**(15), 8562-8567.

Jenison, R., Yang, S., Haeberli, A. & Polisky, B. (2001). Interference-based detection of nucleic acid targets on optically coated silicon. *Nature Biotechn.* **19**, 62-65.

Jen-Jacobson, L., Engler, L.E., Lesser, D.R., Kurpiewski, Yee, M.R. & McVerry, B. (1996) Structural adaptations in the interaction of *Eco*RI endonuclease with methylated GAATTC sites. *EMBO J.* **15**, 2870-2882.

Kerem, B., Rommens, J., Buchanan, J., Markiewicz, D., Cox, T., Chakravarti, A., Buchwald, M. & Tsui, L. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073-1080.

Kiefer, J., Mao, C., Braman, J. & Beese, L. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Science* **391**, 304-307.

Kilmasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369.

Klepser, M., Ernst, E. & Pfaller, M. (1997). Update on antifungal resistance. *Trends Microbiol.* **5**, 372-375.

Kopp, M. U., Mello, A. J. & Manz, A. (1998). Chemical amplification: continuous-flow PCR on a chip. *Science* **280**(5366), 1046-1048.

Koshkin, A.A., Nielsen, P., Meldgaard, M., Rajwanshi, V.K., Singh, S.K. & Wengel, J. (1998). LNA (Locked Nucleic Acid): an RNA mimic forming exceedingly stable LNA:LNA duplexes. *J. Am. Chem. Soc.* **120**, 13252-13253.

Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common diseases. *Nature Genet.* **22**, 139-144.

Kwoh, D., Davis, G., Whitefield, K., Chapelle, H., DiMichele, L. & Gingerase, T. (1989). Transcription-based amplification systems and detection of human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. *Proc. Natl. Acad. Sci. USA* **86**, 1173-1177.

Laken, S., Jackson, P., Kinzler, K., Vogelstein, B., Strickland, P., Groopman, J. & Friesen, M. (1998). Genotyping by mass spectrometric analysis of short DNA fragments. *Nature Biotechn.* **16**, 1352-1356.

Lander, E. & Schork, N. (1994). Genetic dissection of complex traits. *Science* **265**, 2037-2048.

Leontis, N. B., Hills, M. T., Piotto, M., Ouporov, I. V., Malhotra, A. & Gorenstein, D. G. (1995). Helical stacking in DNA three-way junctions containing two unpaired pyrimidines: proton NMR studies. *Biophys. J.* **68**, 251-65.

Leontis, N. B., Kwok, W. & Newman, J. S. (1991). Stability and structure of three-way DNA junctions containing unpaired nucleotides. *Nucleic Acids Res.* **19**, 759-66.

Levin, B. & Anderson, R. (1999). The population biology of anti-infective chemotherapy and the evolution of drug resistance: more questions than answers. In *Evolution in health & disease* (Stearns, S., ed.), pp. 125-137. Oxford University Press, Oxford.

Lhomme, J., Constant, J., Demeunynck, M. (1999). Abasic DNA structure, reactivity and recognition. *Biopolymers* **52**, 65-83.

Lilley, D.M.J. & Clegg, R.M. (1993). The structure of branched DNA species. *Q. Rev. Biophys.* **26**, 131-175.

Lilley, D.M.J., Clegg, R.M., Diekmann, S., Seeman, N.C., Von Kitzing, E., Hagerman, P. (1995) Nomenclature committee of the international union of biochemistry and molecular biology (NC-IUBMB): A nomenclature of junctions and branch points in nucleic acids. *Eur. J. Biochem.* **230**, 1-3.

Liu, Q., Li, X., Sommer, S. (1999). pK-matched running buffers for gel electrophoresis. *Anal. Biochem.* **270**, 112-122.

Liu, Q., Scaringe, W. & Sommer, S. (2000). Discrete mobility of single-stranded DNA in non-denaturing gel electrophoresis. *Nucleic Acids Res.* **28**, 940-943.

Lizardi, P., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. & Ward, D. (1998). Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nature Genet.* **19**, 225-232.

Lu, M., Guo, Q. & Kallenbach, N. (1991). Effect of sequence on the structure of three-arm DNA junctions. *Biochemistry* **30**, 5815-5820.

Lukacs, C.M., Kucera, R., Schildkraut, I., Aggarwal, A.K. (2000). Understanding the immutability of restriction enzymes: crystal structure of *Bgl*II and its DNA substrate at 1.5 Å resolution. *Nature Struct. Biol.* **7**, 134-140.

Lyamichev, V., Mast, A., Hall, J., Prudent, J., Kaiser, M., Takova, T., Kwiatkowski, R., Sander, T., Arruda, M., Arco, D., Neri, B. & Brow, M. (1999). Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechn.* **17**, 292-296.

Lynch, T.W. & Sligar, S.G. (2000). Macromolecular hydration changes associated with *Bam*HI binding and catalysis. *J. Biol. Chem.* **275**, 30561-30565.

Maiden, M. (1998). Horizontal exchange, evolution, and spread of antibiotic resistance in bacteria. *Clin. Infect. Dis.* **27 (suppl 1)**, S12-20.

Marcourt, L., Cordier, C., Couesnon, T. & Dodin, G. (1999). Impact of C5-cytosine methylation on the solution structure of d(GAAAACGTTTTC)2. *Eur. J. Biochem.* **265**, 1032-1042.

Marras, S. A., Kramer, F. R. & Tyagi, S. (1999). Multiplex detection of single-nucleotide variations using molecular beacons. *Genet. Anal.* **14**(5-6), 151-156.

Mayer-Jung, C., Moras, D. & Timsit, Y. (1998) Hydration and recognition of methylated CpG steps in DNA. *EMBO J.* **17**, 2709-2718.

McCormick, J. (1998). Epidemology of emerging/re-emerging antimicrobial resistant bacterial pathogens. *Curr. Opin. Microbiol.* **1**, 125-129.

Miick, S., Fee, R., Millar, D. & Chazin, W. (1997). Crossover isomer bias is the primary sequence-dependent property of immobilized Holliday junctions. *Proc. Natl. Acad. Sci. USA* **94**, 9080-9084.

Muiznieks, I. & Doerfler, W. (1998) DNA fragments with specific nucleotide sequences in their single-stranded termini exhibit unusual electrophoretic mobilities. *Nucleic Acids Res.* **26**(8), 1899-1905.

Neu, H. (1992). The crisis in antibiotic resistance. *Science* **257**, 1064-1073.

Neidle, S. (Ed.) (1999). *Oxford handbook of nucleic acid structure.* Oxford university press, Oxford, UK.

Newman, M., Strzelecka, T., Dorner, L., Schildkraut, I. & Aggarwal, A. (1995). Structure of *Bam*HI endonuclease bound to DNA: partial folding and unfolding on DNA binding. *Science* **269**, 656-663.

Newton, C., Graham, A., Heptinstall, L., Powell, S., Summers, C., Kalsheker, N., Smith, J. & Markham, A. (1989). Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503-2515.

Nickerson, D., Taylor, S., Weiss, K., Clark, A., Hutchinson, R., Stengard, J., Salomaa, V., Vartiainen, E., Boerwinkle, E. & Sing, C. (1998). DNA sequence diversity in a 9.7 kb region of the human lipoprotein lipase gene. *Nature Genet.* **19**, 233-240.

Nielsen, C. B., Singh, S. K., Wengel, J. & Jacobsen, J. P. (2000). Solution structure of an LNA hybridized to DNA: NMR study of the d(CTLGCTLTLCTLGC): d(GCAGAAGCAG) duplex containing four locked nucleotides. *Bioconjug. Chem.* **11**, 228-238.

Nielsen, P., Egholm, M., Berg, R. & Buchardt, O. (1991). Sequence selective recognition of DNA by strand displacement with thymidine-substituted polyamide. *Science* **254**, 1497-1500.

Nikaido, H. (1994). Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382-393.

Nowak, K., Wattanasirichaigoon, D., Goebel, H., Wilce, M., Pelin, K., Donner, K., Jacob, R., Hubner, C., Oexle, K., Anderson, J., Verity, C., North, K., Iannaccone, S., Muller, C., Nurnberg, P., Muntoni, F., Sewry, C., Hughes, I., Stuphen, R., Lacson, A., Swoboda, K., Vigneron, J., Wallgren-Pettersson, C., Beggs, A. & Laing, N. (1999). Mutations in the skeletal muscle alpha-actin gene in patients with actin myopathy and nemaline myopathy. *Nature Genet.* **23**, 208-212.

O'Gara, M., Horton, J.R., Roberts, R.J. & Cheng, X. (1998). Structures of *Hha*I methyltransferase complexed with substrates containing mismatches at the target base. *Nature Struct. Biol.* **5**, 872-877.

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

Orum, H., Jakobsen, M. H., Koch, T., Vuust, J. & Borre, M. B. (1999). Detection of the Factor V Leiden mutation by direct allele-specific hybridization of PCR amplicons to photoimmobilized locked nucleic acids. *Clin. Chem.* **45**, 1898-905.

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

Overmars, F. 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**, 702-713.

Packer, M., Dauncey, M. & Hunter, C. (2000a). Sequence-dependent DNA structure: dinucleotide conformational maps. *J. Mol. Biol.* **295**, 55-70.

Packer, M., Dauncey, M. & Hunter, C. (2000b). Sequence-dependent DNA structure: tetranucleotide conformational maps. *J. Mol. Biol.* **295**, 85-103.

Pecheniuk, N. M., Marsh, N. A. & Walsh, T. P. (2000). Multiple analysis of three common genetic alterations associated with thrombophilia. *Blood Coagul. Fibrinolysis* **11**, 183-189.

Perry, A., Hart, C., Nicolson, I., Heckels, J. & Saunders, J. (1987). Inter-strain homology of pilin gene sequences in *Neisseria meningitis* isolates that express markedly different antigenic pilus types. *J. Gen. Microbiol.* **133**, 1409-1418.

Pier, G.B. (2000). Role of the cystic fibrosis transmembrane conductance regulator in innate immunity to *Pseudomonas aeruginosa* infections. *Proc. Natl. Acad. Sci. USA* **97**, 8822-8828.

Pingoud, A. & Jeltsch, A. (1997). Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.* **246**, 1-22.

Rasmussen, H., Kastrup, J. S., Nielsen, J. N., Nielsen, J. M. & Nielsen, P. E. (1997). Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 Å resolution. *Nature Struct. Biol.* **4**(2), 98-101.

Reinisch, K.M., Chen, L., Verdine, G.L. & Lipscomb, W.N. (1995). The crystal structure of *Hae*III methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143-153.

Rice, M.R. & Blumenthal, R.M. (2000) Recognition of native DNA methylation by the *Pvu*II restriction endonuclease. *Nucleic Acids Res.* **28**, 3143-3150.

Robinson, C.R., Sligar, S.G. (1995). Heterogeneity in molecular recognition by restriction endonucleases: osmotic and hydrostatic pressure effects on *Bam*HI, *Pvu*II and *Eco*RV specificity. *Proc. Natl. Acad. Sci. USA* **92**, 3444-3448.

Robinson, C.R., Sligar, S.G. (1998). Changes in solvation during DNA binding and cleavage are critical to altered specificity of the *Eco*RI endonuclease. *Proc. Natl. Acad. Sci. USA* **95**, 2186-2191.

Rosen, M. A. & 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**(26), 6563-6575.

Rosen, M. A. & Patel, D. J. (1993b). Structural features of a three-stranded DNA junction containing a C-C junctional bulge. *Biochemistry* **32**(26), 6576-6587.

Saenger, W. (1984). *Principles of nucleic acid structure.* Springer advanced texts in chemistry (Cantor, C., Ed.), Springer-Verlag, New York.

Sambrook, J., Fritsch, E., & Maniatis, T. (1989) 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA,

Samson, M., Libert, F., Doranz, B., Rucker, J., Liesnard, C., Farber, C., Saragosti, S., Lapoumeroulie, C., Cognaux, J., Forceille, C. & al, e. (1996). Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR 5 chemokine receptor gene. *Nature* **382**, 722-725.

Sastry, S. & Ross, B. (1997). Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes. *J. Biol. Chem.* **272**, 8644-8652.

Schafer, A. & Hawkins, J. (1998). DNA variation and the future of human genetics. *Nature Biotechn.* **16**, 33-39.

Scriver, C. & Waters, P. (1999). Monogenic traits are not simple. *Trends Genet.* **15**, 267-272.

Seeman, N.C. & Kallenbach, N.R. (1993). DNA branched junctions. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 53-86.

Sidorova, N.Y. & Rau, D.C. (1996). Differences in water release for the binding of *Eco*RI to specific and non-specific DNA sequences. *Proc. Natl. Acad. Sci.* **93**, 12272-12277.

Sidorova, N.Y. & Rau, D.C. (2000). The dissociation rate of *Eco*RI-DNA-specific complex is linked to water activity. *Biopolymers* **53**, 363-368.

Stearns, S. C., Ed. (1999). *Evolution in Health & Disease*: Oxford University Press.

Steemers, F., Ferguson, J. & Walt, D. (2000). Screening unlabeled DNA targets with randomly ordered fiber-optic gene arrays. *Nature Biotechn.* **18**, 91-94.

Stuhmeier, F., Lilley, D. M. & Clegg, R. M. (1997a). Effect of additional unpaired bases on the stability of three-way DNA junctions studied by fluorescence techniques. *Biochemistry* **36**(44), 13539-13551.

Stuhmeier, F., Welch, J. B., Murchie, A. I., Lilley, D. M. & Clegg, R. M. (1997b). Global structure of three-way DNA junctions with and without additional unpaired bases: a fluorescence resonance energy transfer analysis. *Biochemistry* **36**(44), 13530-13538.

Subirana, J. & Faria, T. (1997). Influence of sequence on the conformation of the B-DNA helix. *Biophys. J.* **73**, 333-338.

Suzuki, Y., Sekiya, T. & Hayashi, K. (1991). Allele-specific polymerase chain reaction: a method for amplification and sequence determination of a single component among a mixture of sequence variants. *Anal. Biochem.* **192**, 82-84.

Svensson, P. & Dahlback, B. (1994). Resistance to activated protein C as a basis for venous thrombosis. *N. Engl. J. Med.* **330**, 517-522.

Syvanen, A. C. (1999). From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* **13**, 1-10.

Syvanen, A.C., Aalto-Setala, K., Harju, L. & Kontula, K. (1990). A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein. *Genomics* **8**, 684-692.

Taningher, M., Malacarne, D., Izzotti, A., Ugolini, D., Parodi, S. (1999). Drug metabolism polymorphisms as modulators of cancer susceptibility. *Mut. Res.* **436**, 277-261.

Terwilliger, J. & Weiss, K. (1998). Linkage disequilibrium mapping of complex diseases: fantasy or reality? *Curr. Opin. Biotechn.* **9**, 578-594.

Tippin, D., Ramakrishnan, B. & Sundaralingam, M. (1997). Methylation of the Z-decamer d(GC)5 potentiates the formation of A-DNA: the crystal structure of d(Gm5CGm5CGCGCGC). *J. Mol. Biol.*, **270**, 247-258.

Tomasz, A. (1999). New faces of an old pathogen: emergence and spread of multi-drug resistant *Streptococcus pneumoniae*. *Am. J. Med.* **107 (1A)**, 55S-62S.

Tully, G., Sullivan, K., Nixon, P., Stones, R. & Gill, P. (1996). Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* **34**, 107-113.

Ulyanov, N.B., Ivanov, V.I., Minyat, E.E., Khomyakova, E.B., Petrova, M.V., Lesiak, K., James, T.L. (1998). A pseudosquare knot structure of DNA in solution. *Biochemistry*, **37**, 12715-12726.

Viadiu, H. & Aggarwal, A.K. (2000) Structure of *Bam*HI bound to nonspecific DNA: a model for DNA sliding. *Mol. Cell.* **5**, 889-895.

Walker, G., Fraiser, M., Schram, J., Little, M., Nadeau, J. & Malinowski, D. (1992). Strand displacement amplification - an isothermal, in vitro DNA amplification technique. *Nucleic Acids Res.* **20**, 1691-1696.

Wang, D., Fan, J., Siao, C., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubell, E., Robinson, E., Mittmann, M., Morris, M., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T., Lipshutz, R., Chee, M. & Lander, E. (1998). Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-1082.

Wang, P., Brank, A.S., Banavali, N.K., Nicklaus, M.C., Marquez, V.E., Christman, J.K., MacKerrel, A.D. (2000). Use of oligodeoxyribonucleotides with conformationally constrained abasic sugar targets to probe the mechanism of base flipping by *Hha*I DNA (cytosine C5) methyltranserase. *J. Am. Chem. Soc.* (on-line paper, ahead of print at the time of writing, 19-12-2000).

Welch, J. B., Duckett, D. R. & Lilley, D. M. (1993). Structures of bulged three-way DNA junctions. *Nucleic Acids Res.* **21**, 4548-4555.

Welch, J. B., Walter, F. & Lilley, D. M. (1995). Two inequivalent folding isomers of the three-way DNA junction with unpaired bases: sequence-dependence of the folded conformation. *J. Mol. Biol.* **251**, 507-519.

Westin, L., Xu, X., Miller, C., Wang, L., Edman, C. F. & Nerenberg, M. (2000). Anchored multiplex amplification on a microelectronic chip array. *Nature Biotechnol.* **18**, 199-204.

Whitcombe, D., Newton, C. & Little, S. (1998). Advances in approaches to DNA-based diagnostics. *Curr. Opin. Biotechn.* **9**, 602-608.

Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K., Martin, P.D., Petratos, K., Wilson, K.S. (1993). The crystal structure of *Eco*RV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.* **12**, 1781-1795.

Woodcock, D.M., Crowther, P.J., Doherty, J.P., Linsenmeyer, M.E., Kruger, D.H. (1991) *Gene*, **102**, 79-81.

Wright, A., Carothers, A. & Pirastu, M. (1999). Population choice in mapping genes for complex diseases. *Nature Genet.* **23**, 397-404.

Yang, M. & Millar, D. P. (1996). Conformational flexibility of three-way DNA junctions containing unpaired nucleotides. *Biochemistry* **35**, 7959-7967.

Youil, R., Kemper, B. & Cotton, R. G. (1996). Detection of 81 of 81 known mouse beta-globin promoter mutations with T4 endonuclease VII--the EMC method. *Genomics* **32**, 431-435.

Youil, R., Kemper, B. W. & Cotton, R. G. (1995). Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. *Proc. Natl. Acad. Sci. U S A* **92**, 87-91.

Zhong, M. & Kallenbach, N. (1993). Conformation and thermodynamics of DNA "Necks". *J. Mol. Biol.* **230**, 766-778.

Zhong, M., Marky, L. & Kallenbach, N. (1997). Thermodynamics of dT-dT base pair mismatching linear DNA duplexes and three-arm DNA junctions. *Biochemistry* **36**, 2485-2491.

Zhong, M., Rashes, M. & Kallenbach, N. (1993). Effect of T-T mismatches on three-arm DNA junctions. *Biochemistry* **32**, 6898-6907.

Zhong, M., Rashes, M., Leontis, N. & Kallenbach, N. (1994). Effect of unpaired bases on the conformation and stability of three-arm DNA junctions. *Biochemistry* **33**, 3660-3667.

Zielenski, J., Corey, M., Rozmahel, R., Markiewicz, D., Aznarez, I., Casals, T., Larriba, S., Mercier, B., Cutting, G., Krebsova, A., Macek, M., Langfelder-Schwind, E., Marshall, B., DeCelie-Germana, J., Claustres, M., Palacio, A., Bal, J., Nowakowska, A., Ferec, C., Estivill, X., Durie, P. & Tsui, L. (1999). Detection of a cystic fibrosis modifier locus for meconium ileus on human chromosome 19q13. *Nature Genet.* **22**, 128-129.

Zimm, B.H., Levene, S.D. (1992) Problems and prospects in the theory of gel electrophoresis of DNA. *Q. Rev. Biophys.* **25**, 171-204.