UNIVERSITY OF SOUTHAMPTON

Protein Based Approaches to Heteroduplex Recognition for the High Throughput Detection of Unknown Mutations.

By Matthew James Smith

Human Genetics

PhD Thesis

February 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT
FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
HUMAN GENETICS

Doctor of Philosophy

Protein Based Approaches to Heteroduplex Recognition for the High Throughput Detection of Unknown Mutations
By Matthew Smith

The ability to screen PCR products from a large population of individuals for differences in genes could have significant research and clinical implications. Detection of unknown mutations is the first step in hopefully being able to characterise pathogenic and non-pathogenic forms of a particular gene. To further our understanding of human genetics it is essential that we can identify and characterise changes in the DNA sequence of a gene. DNA sequencing is the ultimate method for detection and definition of mutations. However the need for scanning methods exists due to the expensive and time-consuming nature of DNA sequencing. Scanning methods have arisen to avoid sequencing whole stretches of DNA, to reduce costs and increase throughput. It is not feasible when trying to detect and characterise mutations in a population to sequence the DNA of all the samples. Therefore the need for cost-effective simple scanning methods exists.

Endonucleases, Resolvases and DNA binding proteins all provide potential tools for the screening for mutations. Two such proteins MutS, a mismatch binding protein and T4 endonuclease VII, a resolvase with mismatch cleavage properties were studied to determine their suitability for mutation detection. Both proteins were His-tag labeled to allow convenient and high yield production of protein via affinity chromatography. The basis of a scanning technique using MutS or Endo VII for epidemiological studies is based upon heteroduplex analysis. The mismatch recognition of both these proteins has been used in conjunction with the high throughput capabilities of (MADGE) in order to develop novel mutation scanning strategies.

The work presented here investigates the suitability of the two enzymes MutS and T4 endonuclease VII for the high throughput identification of unknown mutations in epidemiological studies using short track electrophoresis. Following the initial characterisation of the proteins and application to short tract electrophoresis, the mutation detection capabilities of both proteins were tested using a variety of known SNPs and mutations. Of the two proteins studied T4 Endo VII demonstrated the greatest potential for high throughput mutation detection. It was successfully utilised to detect a number of mutations using MADGE. By combining the mismatch cleavage properties of Endo VII with the high throughput capabilities of MADGE a cost effective method for the high throughput detection of unknown or rare mutations has been created

Acknowledgements

I would like to take this opportunity to thank both my supervisors Professor Ian Day and Professor Keith Fox for their continued encouragement, support and advice throughout this project. To all members past and present of both Lab 18 and the Fox group, for making my 3 years thoroughly enjoyable. All their help and useful contributions is very much appreciated.

Special thanks go to my family, in particular to my parents without whom I would never have got this far.

Finally to Sarah for her unwavering support throughout the last 3 years and for reminding me that there is always light at the end of the tunnel.

Contents

Figure Index

Chapter1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Table Index

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Abbreviations

A	= Adenine
ADP	= Adenosine diphosphate
AP	= Aminopurine
APS	= Ammonium persulfate
ARMs	= Amplification refractory mutation system
ATP	= Adenosine triphosphate
BSA	= Bovine serum albumin
C	= Cytosine
CCD	= Charge coupled device
CCM	= Chemical cleavage of mismatch
Da	= Daltons
DGGE	= Denaturing gradient gel electrophoresis
DI	= Deionised water
DNA	= Deoxyribonucleic acid
DTT	= Dithiothreitol
EB	= Elution Buffer
EDTA	= Ethylene diamine tetraacetic acid
EMS	= Ethylmethanesulfonate
EndoVII	= T4 endonuclease VII
G	= Guanine
HA	= Heteroduplex analysis
HEPES	= N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
His	= Histidine
Hm	= Homoduplex
Hmz	= Homozygote
H-PAGE	= Horizontal polyacylamide gel elctrophoresis
Ht	= Heteroduplex
Htz	= Heterozygote
IPTG	= Isopropyl-β-D-thiogalactopyranoside
LB	= Lurie broth
MADGE	= Microtiter array diagonal gel electrophoresis
MMR	= Methylated Mismatch Repair
Mt	= Mutant
mRNA	= Messenger ribonucleic acid
NG	= Nitrosoguanidine
O.D	= Optical Density
ON	= Overnight
PAGE	= Polyacrylamide gel electrophoresis
PCR	= Polymerase chain reaction
Pol	= Polymerase
RB	= Reaction buffer
RE	= Restricition enzyme
RNA	= Ribonucleic acid
rRNA	= Ribosomal ribonucleic acid
SAP	= Shrimp alkaline phophatase
SSCP	= Single strand conformational polymorphism
SDS	= Sodium dodecyl sulphate
SNP	= single nucleotide polymorphism

T	= Thymine
Taq	= Thermus aquaticus
TBE	= Tris-boric acid EDTA
TE	= Tris- EDTA
TGGE	= Temperature gradient gel electrophoresis
TEMED	= N,N,N',N'-Tetramethylethylenediamine
T4PNK	= T4 phosphonuclease kinase
UHQ	= Ultra high quality
UV	= Utraviolet
WT	= Wildtype

CHAPTER 1
INTRODUCTION

1.1. General Introduction

With the near completion of the human genome combined with the rapid discovery of disease genes the need for high throughput mutation scanning methods increases. The ability to screen PCR products from a large population of individuals for differences in genes could have significant research and clinical implications. Detection of unknown mutations is the first step in hopefully being able to characterise pathogenic and non-pathogenic forms of a particular gene. From the perspective of research and diagnostic laboratories, quick throughput, cheap methods for detecting mutations are extremely useful.

DNA sequencing is often regarded as the ultimate method for detection and definition of mutations. However the need for scanning methods exists due to the expensive and time-consuming nature of DNA sequencing. Scanning methods have arisen to avoid sequencing of whole stretches of DNA and to reduce costs and increase throughput. It is not feasible when trying to detect and characterise mutations in large population samples, for example in epidemiology studies, to sequence the DNA of all the samples. Therefore the need for cost effective simple scanning methods exists.

1.2. Deoxyribonucleic Acid (DNA)

1.2.1. Structure

Deoxyribonucleic acid otherwise known as DNA was first discovered at the end of the 19th century. Its function and subsequent role as the molecule of hereditary did not become clear until the latter half of the 20th century with the discovery of the capacity of DNA to confer biological characteristics upon an organism (Avery et al., 1944)and the publication of the structure of DNA by Watson and Crick in 1953.

DNA consists of two complementary strands of nucleotides wound around each other to form a helix. A single nucleotide consists of a 2'-deoxyribose sugar, a phosphate group and one of four different bases. These bases are adenine (A), cytosine (C), guanine (G) and thymine (T). The variability of the bases is the only difference in the four nucleotides and

confers the genetic difference seen between organisms and individuals. The base is linked to the sugar residue at 1'C via its nitrogen group by a N-glycosidic bond. The phosphate group is attached at the 5' carbon of the sugar residue. Each separate nucleotide binds to a second nucleotide by a phosphodiester bond between the hydroxyl group of the 3' C and the phosphate group of the second nucleotide molecule. The subunits of DNA are arranged so that the sugar phosphate groups form the backbone of the molecule with the bases projected into the middle of the helix. Thus long chains of nucleotide molecules can be formed with each of the two strand that make a DNA molecule running anti parallel to one another i.e. the 5' and 3' ends run in opposite directions (Fig. 1.1).

Each base hydrogen bonds to a specific base on the opposite strand. The bases are categorised into pyrimidines T and C and purines A and G, a distinction determined by the number of nitrogenous rings they have (Figure 1.1). Within the DNA molecule A always pairs with T and G always pairs with C, forming complementary Watson & Crick base pairs. The whole structure is stabilised by hydrogen bonds between the bases, between G-C there are 3 H-bonds and between A-T there are 2 H-bonds. Stacking of the base pairs also bring in hydrophobic interactions and van der Waals interactions.

1.2.2. Packaging of DNA within the Nucleus

The DNA within a single human diploid cell comprises approximately 6 billion base pairs. The whole of an organism's genetic material is referred to as a genome. The human nuclear genome is distributed on highly packaged molecules known as chromosomes, which range in size from 55Mb to 250Mb. Human somatic cells contain 46 chromosomes consisting of 2 copies of each of the 22 types of autosomal chromosomes and 2 sex chromosomes (X &Y). Each set of autosomes and a single sex chromosome are inherited from each parent and cells containing two sets of chromosomes are known as diploid. The genes on the maternal and paternal chromosomes may or may not be identical and these variations are called alleles. DNA does not exist as a naked double helix because the total length of DNA within a single cell covers approximately 2 metres. Therefore for it to be accommodated within the

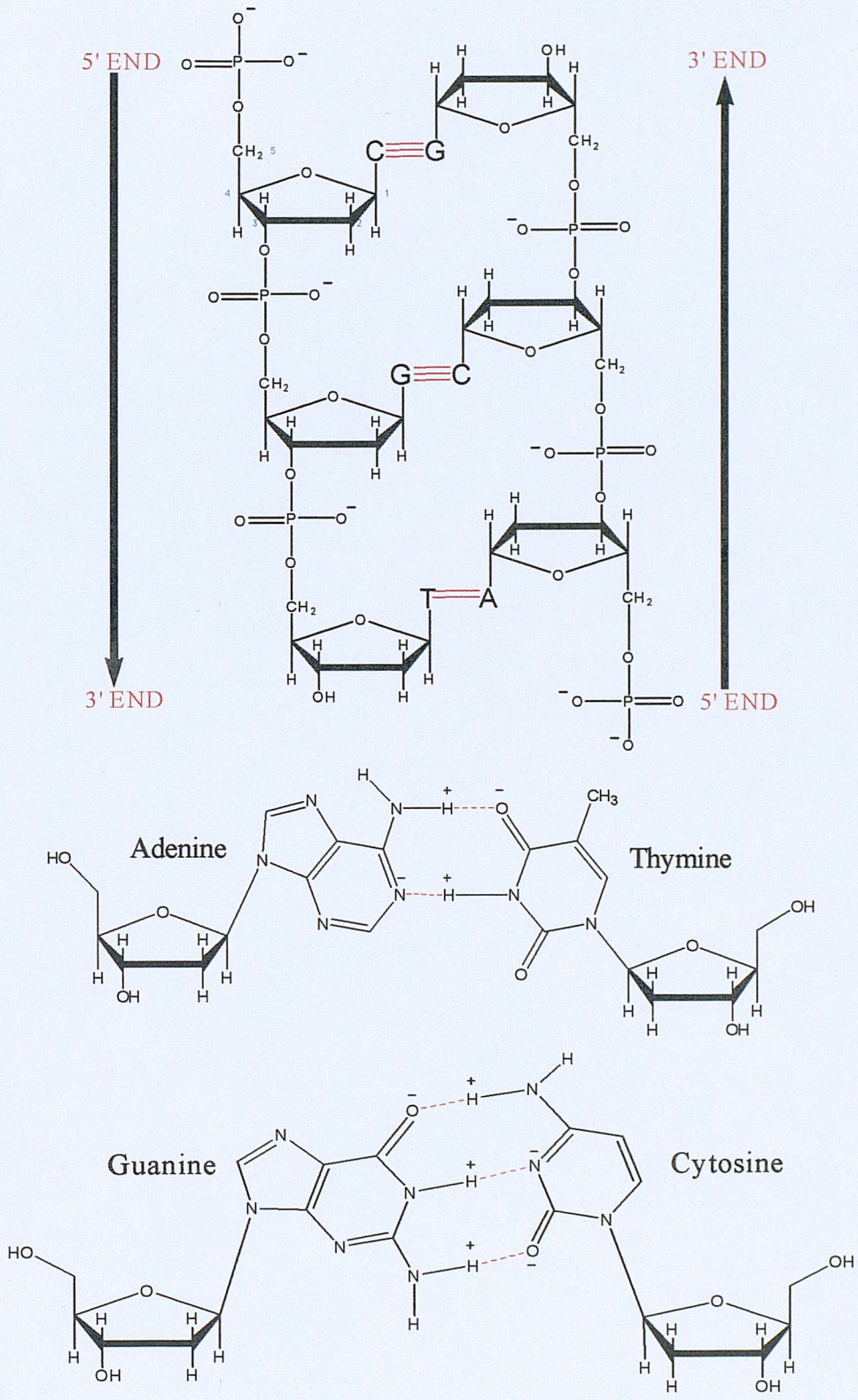


Figure 1.1. The antiparallel nature of double stranded DNA and the complementary base pairing exhibited by the 4 bases.

nucleus it undergoes packaging into higher ordered structures.

The principle packaging element within the nucleus is the nucleosome. At the core of the nucleosome is an octomer comprising of two copies of each histone proteins H2A, H2B, H3 and H4. The DNA within the nucleus wraps around this octamer forming the nucleosome core. Each octamer has between 145 – 147 base pairs of DNA wrapped around it (Luger et al., 1997). These nucleosome cores occur throughout the entire genome every couple of hundred base pairs. The nucleosome cores are linked via a 60 base pair linker sequence and the arrangement of the arrays of nucleosomes is known as chromatin. This is the primary determinant of DNA accessibility and generally inhibits the binding of non-histone protein such as transcription factors (Luger *et al.*, 1997).

Chromatin undergoes further packaging with the aid of a fifth histone molecule H1. Each nucleosome becomes associated with an H1 molecule; these H1 proteins in turn interact with one another forming a spiral known as a solenoid or 30nm chromatin fibre. This is the normal state of the DNA during the interphase stage of the cell cycle and allows the DNA to stay genetically active. The condensing of the chromatin for cell division or production of gametes gives rise to the higher ordered structures that give the characteristic karyotyping images associated with the appearance of chromosomes.

1.2.3. Role of DNA

DNA acts as the genetic store of the cell and provides the blue print for the synthesis of RNA, which in turn is responsible for the synthesis of polypeptides, the fundamental building blocks of all proteins. It is only a small percentage of the human genome (3%) that is responsible for building, running and maintaining the organism (Strachan and Read, 1999). These coding areas of DNA are called genes and control the hereditary characteristics of the cell. The total number of genes has been estimated to be between 30,000 – 40,000 by the international human genome sequencing consortium (2001).

1.2.4. Genome Variation

It is apparent that no two individuals (except identical twins) are exactly the same. These phenotypic traits are a result of the under lying genetic differences found between any two individuals and arise due to sequence variations found within the genome. This is referred to as nucleotide diversity. Nucleotide diversity varies greatly across the genome and although these variations are found on all the chromosomes their distribution is not uniform. Within the genome there are known genetic hotspots where there is a great deal of genetic variability, whilst other regions do not vary much between individuals and are described as stable. The vast majority of sequence variation between individuals is attributed to single base changes in the sequence known as single nucleotide polymorphisms (SNP), with the rest attributable to insertion/deletions, repeat length polymorphisms and rearrangements. Classification as an SNP occurs if the variant rarer allele is present within a population with a >1% frequency. Typically base changes that have a frequency of <1% are termed mutations. This terminology is based upon the frequency of the allele within the population. Mutation also often refers to a harmful genome variation that is associated with a specific human disease whilst polymorphism often infers neither harmful nor beneficial effect, however this is not entirely correct as it is now becoming apparent that SNP combinations play an important role in disease manifestation.

Numerous studies have been performed to estimate the sequence variation within the human genome. Screening of a 9.7kb region of the human lipoprotein lipase gene from three diverse populations gave a total of 88 sites showing sequence variation, of which 79 were SNPs and 9 were insertion-deletion variation (Nickerson et al., 1998). A second study on a 4-Mb high density SNP based map around the human APOE detected an SNP at a frequency of 1 SNP / 1.1 kb of genomic sequence (Lai et al., 1998). A third study which used all the publicly available SNPs spanning the whole human genome sequence indicated a frequency of one SNP every 1.9kb (Sachidanandam et al., 2001). Therefore a range of values has been postulated but, as mentioned earlier, the area chosen to evaluate will have a bearing on the sequence variation. The most commonly cited value for single

base sequence variation between two non-related individuals is a difference on average of 1 every 1kb. The exact meaning of this figure is that there is an average 0.1% chance of any base being heterozygous in an individual. This translates into several million sequence variations between two non-related individuals, which results in approximately 100,000 amino acid differences between their proteomes (Brookes, 1999). The 0.1% difference between these individual genomes is also seen if a third genome is compared although the sequence variation in the majority of cases would be located in different positions. Usually these changes are relatively small and it is very unlikely that individual genomes contain long stretches others lack. Although human genomes share 99.9% homology with one another it is the 0.1% difference that gives rise to us as individuals. In respect to the human genome project the sequence produced is a representative sequence based upon a few individuals. However the vast majority of the sequence is the same from person to person with the location and identity of the genes the same. To understand the consequences of sequence variation it is first useful to know the relationship between the DNA sequence of a gene and the subsequent formation of protein.

1.3. Genes and Gene Expression

1.3.1. The Genetic Code

The relationship between the sequence of bases in DNA and the sequence of amino acids in proteins is the genetic code, where three nucleotides (codon) code for one amino acid. The existence of four bases and three nucleotides per codon means that there are 64 possible codon combinations for the 20 amino acids that make up proteins. The code is therefore said to be degenerate because most amino acids have more than one code. The genetic code is read stepwise in groups of three so each group of three specifies only one amino acid and the code is said to be non-overlapping.

1.3.2. Genes

The majority of DNA found within in a nucleus as yet has no determinable function. Found throughout this "junk" DNA are the short sequences that are used to build proteins and RNA molecules. Broadly speaking there are two types of genes, non-coding RNA genes and the protein coding genes. Non-coding RNA genes represent approximately 2-5% of the total coding region of the human genome. They encode functional RNA molecules involved in the control of gene expression in particular protein synthesis.

Genes vary greatly in size and organisation and there is no typical structure, however there are several conserved features that are common among genes. Coding regions begin with an initiation codon, normally ATG, which codes for the amino acid methionine. A typical human gene contains both areas that code for amino acids (exons) and areas that are non-coding (introns). Initially the whole of the gene sequence, exons and introns are transcribed to form the primary transcript. Prior to export to the cytoplasm and subsequent translation the introns are removed, a process termed splicing, and the exons are spliced together to form a continuous coding mature RNA strand (Fig. 1.2). A gene includes more than the nucleotides encoding the amino sequence of a protein, it also includes the DNA sequences on either side of the actual coding region required for RNA transcript production.

Immediately prior to the transcriptional start are short sequence elements required for the recruitment of transcription factors and collectively constitute the promoter. A highly conserved AT rich sequence, the TATA box, is a common element of the promoter. The TATA box, approximately 30bp upstream of transcription start, is the initial target for a transcriptional complex that recruits RNA polymerase II. In the absence of a TATA box other promoter elements like GC box or CAAT box act as the site for transcriptional complex binding.

Located further away from the origin of transcription are enhancer sequences. They can occur at the beginning, middle or end of a gene. They act as binding sites for specific protein factors that act to recruit further transcription factors or interact directly with RNA polymerase. These

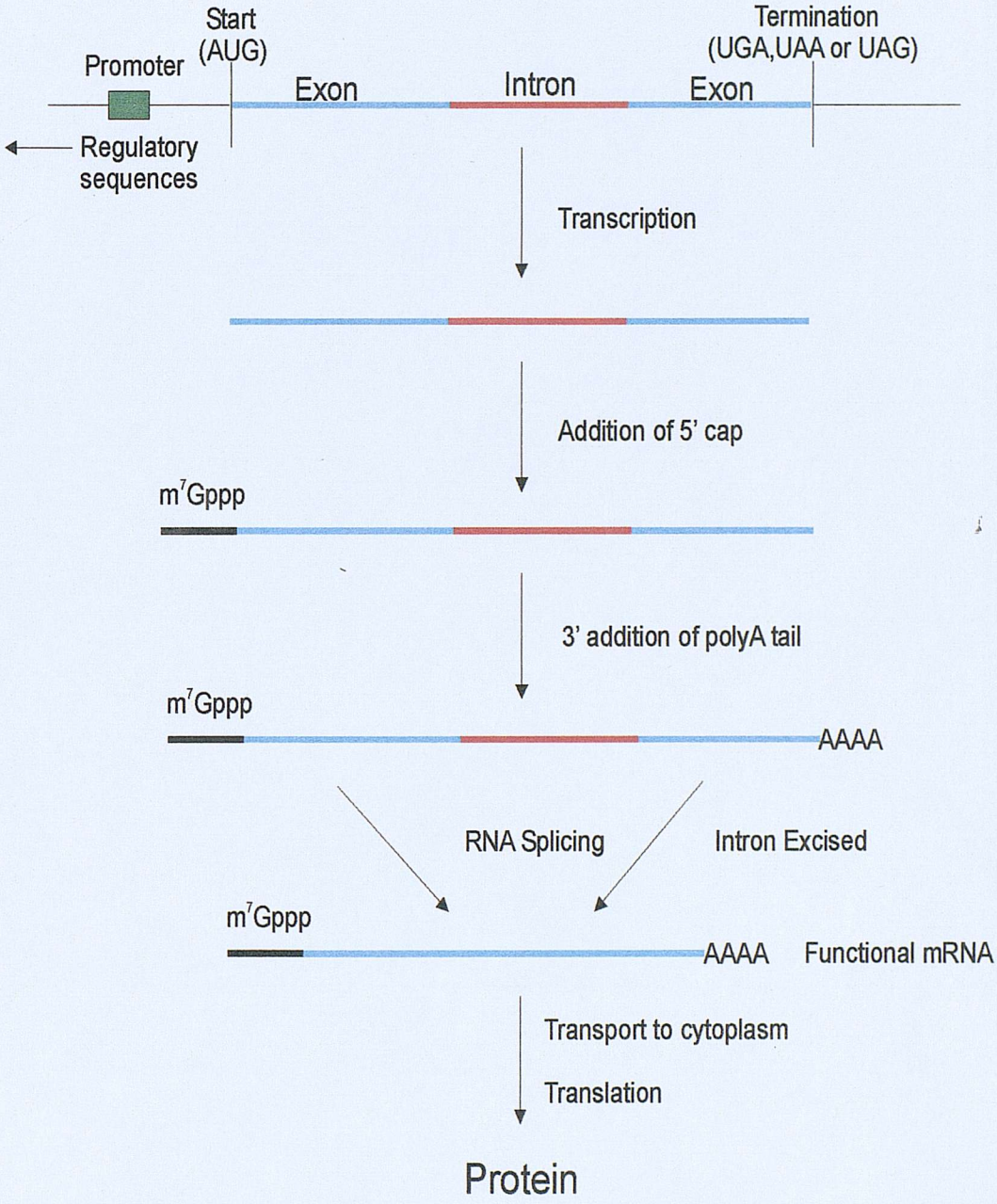


Figure 1.2. Representation of a typical eukaryotic gene and its subsequent conversion into a functional protein.

enhancer sequences can lie 50kb or more from the coding region. RNA processing signals are critical non-coding sequences, which are not involved in the recruiting of RNA polymerases or transcription factors. These sequences direct 3' cleavage and polyadenylation processes essential for protein expression whilst specific sequences in the introns act to direct splicing.

1.3.3. Gene Expression

The first step in gene expression is the production of a primary transcript. The gene is copied (transcribed) into a complementary ribonucleic acid (RNA) strand. The enzymes responsible for the production of the primary transcript are the RNA polymerases. Within eukaryotes there are three classes of RNA polymerases, I, II, III.

As RNA polymerase II is responsible for transcribing all protein-coding genes the transcription and translation processes will be described from the perspective of RNA Pol II. The successful recruitment and binding of RNA Pol II and the transcription of the DNA leads to the production of the primary transcript. To prepare the primary transcript for exportation to the cytoplasm for translation it undergoes a number of modifications. The 5' end of the nascent RNA molecule is capped by the addition of a guanine residue that is not coded for in the DNA. The G residue is separated from the RNA chain by three phosphates as opposed to one like the RNA molecules in the chain. Also the G is joined by a 5' to 5' bond unlike 3' to 5' usually seen between linked nucleotides (Latchman, 2002). The function of this cap is to protect the RNA molecule from exonuclease digestion and to act as a guide to ribosome binding and recognition. A second modification made to the primary transcript is the addition of up 200 adenosine residues to the 3' of the mRNA. These are added post transcriptionally and protect the mRNA from degradation by exonucleases (Barabino and Keller, 1999). The splicing of the exons and removal of the non-coding introns achieve the formation of a mature RNA molecule.

The final step in the conversion of the DNA sequence into a polypeptide is the translation of the mature mRNA into a sequence of amino

acids. This process is carried out by cytoplasmic organelles called ribosomes. Ribosomes contain two subunits consisting of a mixture of RNAs and protein and are designated the 40S and 60S subunits.

Translation is initiated by the binding of eukaryotic initiation factors to the mRNA molecule (Pestova et al., 2001). This is followed by the recruitment of the 40S subunit of the ribosome, tRNA carrying methionine and a further initiation factor (Latchman, 2002). The 40S subunit migrates along the RNA until it encounters the universal start codon AUG. The initiation factor is released and the 60S subunit recruited and the complete ribosome formed (Pestova et al., 2000). Amino acids are transported to the ribosome by transfer RNA (tRNA) molecules. Each tRNA contains an anticodon of 3 bases that pairs complementary with the mRNA codons, and each tRNA with a specific anticodon binds a specific amino acid. Therefore tRNA delivers a particular amino acid to the ribosome depending upon the sequence of the three base codon on the mRNA molecule. Within the ribosome are two $tRNA_i$ binding sites. The ribosome catalyses the formation of a peptide bond between the 2 amino acids of the adjacent tRNA contained in the binding sites. The ribosome then shifts to the subsequent codon and the complementary tRNA is bound in the free binding site made available by the movement of the ribosome and a peptide bond formed between the two amino acids. The information in the mRNA is gradually translated into protein according to the genetic code and continues until a stop codon is encountered upon which the ribosome leaves the mRNA and the polypeptide is released. Genes in the human genome are not all expressed in the same way or at the same time. Genes that are required by all cells and therefore are expressed throughout the body are known as house keeping genes. Some genes are only expressed in certain cell types or tissues, for example muscle proteins such as actin are only expressed in muscle cells where they are required. Other genes are activated or inhibited by signals such as hormones circulating in the body. Therefore differential gene expression is achieved by the regulation off transcription, translation and the availability and activity of numerous transcription factors.

1.4. Effects of Sequence Variation on the Genetic Code

1.4.1. Mutation Classification

Mutations can be classified initially on the basis of the sequence alteration they cause:

1. Point mutations – substitution of one base for another
2. Insertion/deletions – addition or deletion of bases from the DNA sequence
3. Larger Scale mutations - duplication and inversion of DNA sequence and recombination errors.

Within the confines of this project we are interested with point mutations, therefore the most common reasons behind their occurrence and the effects on DNA fidelity and cell biochemistry are looked at in more depth. The molecular consequence of a base substitution depends upon its location within the genome. The following list adapted from Friedberg, 1995, summarises the molecular effects of base substitutions dependant upon their location

1. Promoter – reduced or increased gene expression
2. Regulatory Sequence of a gene – alteration of regulation of gene expression
3. 3' of protein-coding region – defective transcription termination or alteration of mRNA stability
4. Certain areas within introns – defective mRNA splicing
5. Origin of DNA replication – defect in initiation of DNA replication
6. Protein coding regions – beneficial, no effect, harmful
7. Junk DNA – no discernable effect.

Point mutations are commonly divided into two groups: transitions that are purine-to-purine or pyrimidine-to-pyrimidine changes and transversions that are purine to pyrimidine (Fig. 1.3). The transition rate in mammalian

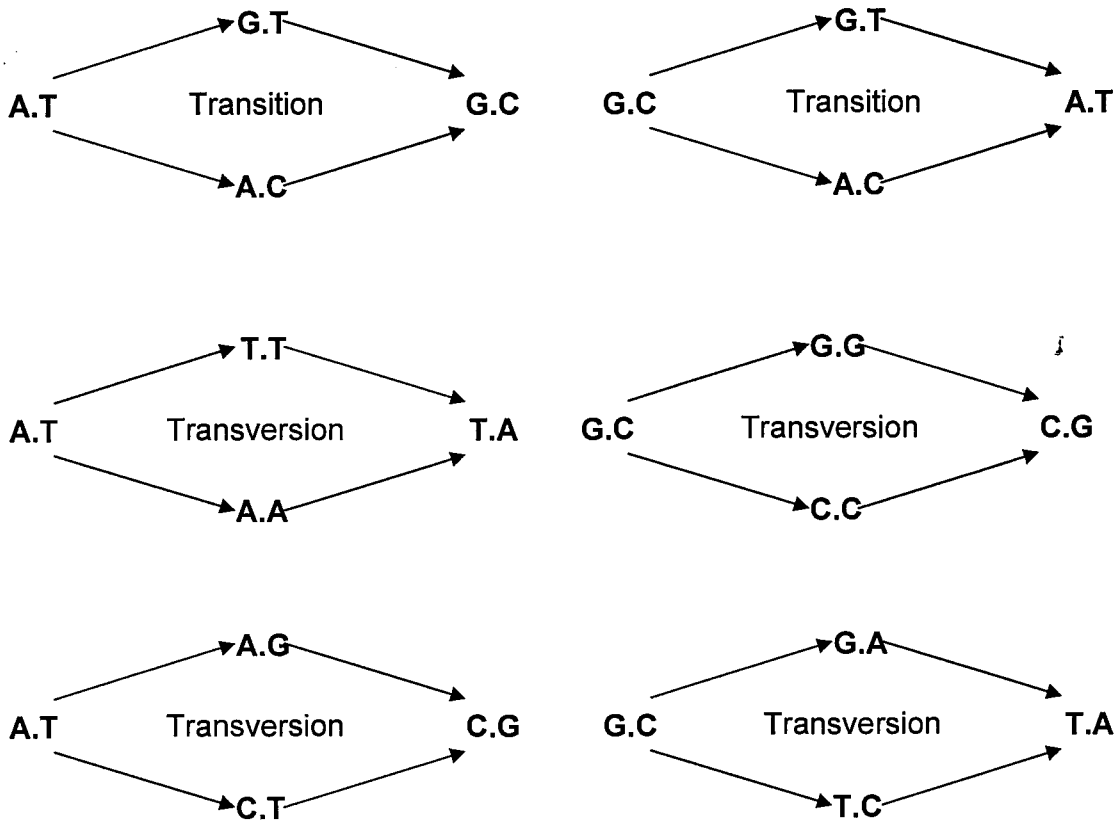


Figure 1.3. Mutagenic pathways. Result of mispairing to the genetic code and the classification of the mispairing (Brown and Hunter, 1997)

genome is higher than that of transversions, even though for each base there are two possible transversions and only a single transition. This disproportion in rates could be due to the fact that transitions in coding regions tend induce a more conservative change in the genetic code compared to a transversion (Strachan *et al.*, 1999). Due to the degeneracy of the genetic code not all mutations will lead to a change in identity of an amino acid incorporated into a polypeptide. These are called silent mutations and convey no advantage or disadvantage to the cell. Many of the codons that code for the same amino acid differ at the third base in the triplet code (wobble position) so a base substitution at this third base in many cases still leads to the incorporation of the original amino acid.

A mutation that causes the incorporation of a different amino acid is described as non-synonymous and can induce a deleterious effect, have no effect or less frequently a beneficial effect. The exchange of an amino acid for an amino acid with a similar chemical structure is a conservative change and is aided by the fact that many similar amino acid share similar codons. Such a change is intended to reduce any deleterious effects a mutation may cause. Non-conservative amino acid incorporation refers to the inclusion of an amino acid, which has very little structure similarity with the original amino acid. A non-conservative substitution, due to the dissimilarity between the side chains, means an adverse effect is more probable. The final possible outcome of a base change is the formation of a nonsense mutation. This occurs when the base change converts the original codon to one of the three stop codons (TAG/TAA/TGA). This frequently results in a partial or complete loss of protein function due to the premature termination of protein synthesis.

From the list it is apparent that not all base changes will directly affect the protein on a structural level. Mutations in promoter regions, transcription factor binding sites etc can all affect the level of protein expression. Either higher levels or lower levels of expression can alter the biochemistry of the cell dramatically. Similarly mutations that alter exon splicing or the mRNA stability will ultimately result in a gene product with an altered structure or expression level.

1.4.2. Germline Mutation

The location and the global effects of a mutation also depend upon when and where the mutation arose. If the mutation occurs within a germ cell this then becomes a hereditary change and is present within all cells of that particular organism. This gives rise to a new allele, which can enter the population, as long the mutation is not pathogenic. The mutation rate per generation has been calculated to range from 93-248 new sequence variations (Nachman and Crowell, 2000). Other variations arose many generations ago and have subsequently been passed on and are shared with many other people.

1.4.3. Somatic Mutations

A mutation that arises after the first round of replication will only be found in a certain proportion of cells, those that derive from the cell carrying the mutation and will become a permanent mutation within the lineage of that cell line. The proportion of cells carrying the mutation depends upon the stage at which the mutation arose. If mutations occur early in embryonic life then the effect of the mutation can be wide ranging and present in the majority of cells. Individuals whose genetic material varies between cells are termed mosaic.

Sporadic mutations that arise in gonadal cells can be passed onto successive generations, whereas mutations in somatic cells cannot be inherited. Somatic mutations in certain cells are known to play a key role in many common cancers. Cancer is caused by the uncontrolled division of a cell that eventually leads to the formation of a tumour. The fault occurs after birth in a restricted number of somatic cells and frequently involves mutations in onocogenes, tumour-suppressor (TS) genes and mutator genes. Oncogenes are genes that are involved in promoting cell proliferation and participate in a variety of normal cellular functions. Mutations in onocogenes lead to excessive and inappropriate cell proliferation, which leads to tumour formation. TS gene products act in the opposite way to oncogenes and their products are involved in inhibition of cell proliferation. In their absence cells

are able to proliferate and form tumours. The products of mutator genes are involved in the pathways utilised for repairing mutations and damage to the genome. Mutations in these genes reduce the efficiency of DNA repair. The consequence of this is the increased likelihood of cancer causing mutations occurring and remaining unnoticed and unrepaired.

1.4.4. Influence of Single Base Variations on Human Disease.

As our knowledge of DNA has progressed it has be come apparent that many human diseases and human traits have an underlying genetic determinant that depends on the sequence of nucleotides found within our cells. It is this observation that has fuelled the need for discovery of both common and rare alleles that could be linked to disease. Human diseases can be broadly divided into two categories. Monogenic diseases are caused by a defect in a single gene arising from a single base change. Due to the severity that these single bases have on the overall protein function, they are genetically rare, with a low frequency in the population. An example of a well-known single gene disorder and effects that a single point mutation can have is sickle cell anaemia. A single base change in the gene for β-globin results in the substitution of a valine for a glutamic acid. The consequence of this single polymorphism is the alteration of the shape of the red blood cells. These red blood cells are able to polymerise forming painful blood clots. Thrombosis and anaemia are common symptoms of heterozygotes whilst homozygotes suffer from congestive heart failure. Cystic fibrosis and Huntington disease are two other commonly known single gene disorders.

There are also more complex diseases where the phenotypic trait is a combination of alleles that have a cumulative effect. These are referred to as polygenic and no one gene has the overall determining factor as to whether the disease manifests itself. The disease is not a result of a single mutant allele but requires a number of genes each making a subtle contribution for disease manifestation. This complex interaction of multiple genes can be complicated further by the influence of environmental and lifestyle factors. Diabetes, heart disease and asthma are all examples of polygenic diseases.

Single base variations can also influence the susceptibility to disease or certain phenotypes like obesity. This type of gene variation is not sufficient to cause the disease but could influence the age of onset, determine how the disease progresses and the severity of the disease.

The detection of single base changes is the first step in identifying potential disease causing mutants, as well as determining susceptibility of individuals to a whole host of genetic disorders. Although environmental and lifestyle factors will influence the development of a disease it is hard to measure their effects, therefore it is the genetic predisposition that is the major determinant of the risk to developing particular traits.

1.5. DNA Damage

1.5.1. Sources of DNA Damage

It is surprising that as the fundamental building block of life, how dynamic DNA actually is. As the blue print for all cellular activity it might be expected that DNA would be extremely stable in order to maintain the high degree of fidelity needed to maintain correct functioning. However the integrity of the DNA sequence is challenged constantly due to the Inherent instability of specific chemical bonds that constitute the normal chemistry of nucleotides. Mutations (changing of the genetic code), whilst necessary to drive evolution, more often than not lead to adverse changes in the cell's biochemistry via non-sense or incorrect gene products. Many errors occur during DNA replication, recombination and repair itself. Furthermore DNA is constantly exposed to a variety of chemical and physical insults from both internal and external sources that attack and alter DNA.

1.5.2. Intracellular

1.5.2.1. Base Misincorporation

DNA replicates semi conservatively, each strand acting as a template for the synthesis of new (daughter), it is during this process that mispairing can occur. If these mismatches are left uncorrected then they will lead to an

alteration in the genetic code. Based on free energy analysis the difference between a complementary and non-complementary base pairing is only 2/3 kcal/mol (Loeb and Kunkel, 1982) and would give an unacceptable error rate of 1 - 10 % for each nucleotide . However the spontaneous error rate is $10^{-5}/10^{-6}$ due to proof reading properties of DNA polymerase. When coupled with the other mechanisms for retaining genetic stability the probability of misincorporation is reduced 10^{-10}.

1.5.2.2. Genetic Recombination

Genetic recombination is performed by the cell during the preparation of gametes and contributes to the uniqueness of each human individual. During recombination parental DNAs align at regions of sequence similarity and DNA is exchanged between the chromosomes. Recombination plays a key role in the repair of DNA, generates new combinations of genes and regulates the expression of DNA (Stryer, 1988). The formation of a heteroduplex between two homologous DNA molecules as part of a genetic recombination process can generate mismatches if the two DNA's differ slightly in their sequence.

1.5.2.3. Base Analogs

Chemical compounds that are sufficiently similar to the normal nitrogen bases of DNA can be occasionally incorporated in their stead. These analogues can have pairing properties that are different to the normal base. During replication this can lead to incorrect nucleotides being incorporated into DNA.

An analog of adenine, 2-aminopurinne (2-AP) can base pair with thymine but can also base pair with cytosine when protonated (Griffiths, 1999). Therefore if 2-AP is incorporated into DNA by pairing with thymine, upon subsequent rounds of replication it can mispair with cytosine thus causing an AT-GC transition. Alternatively 2-AP can mispair with cytosine on the template strand which will eventually result in a GC-AT transition.

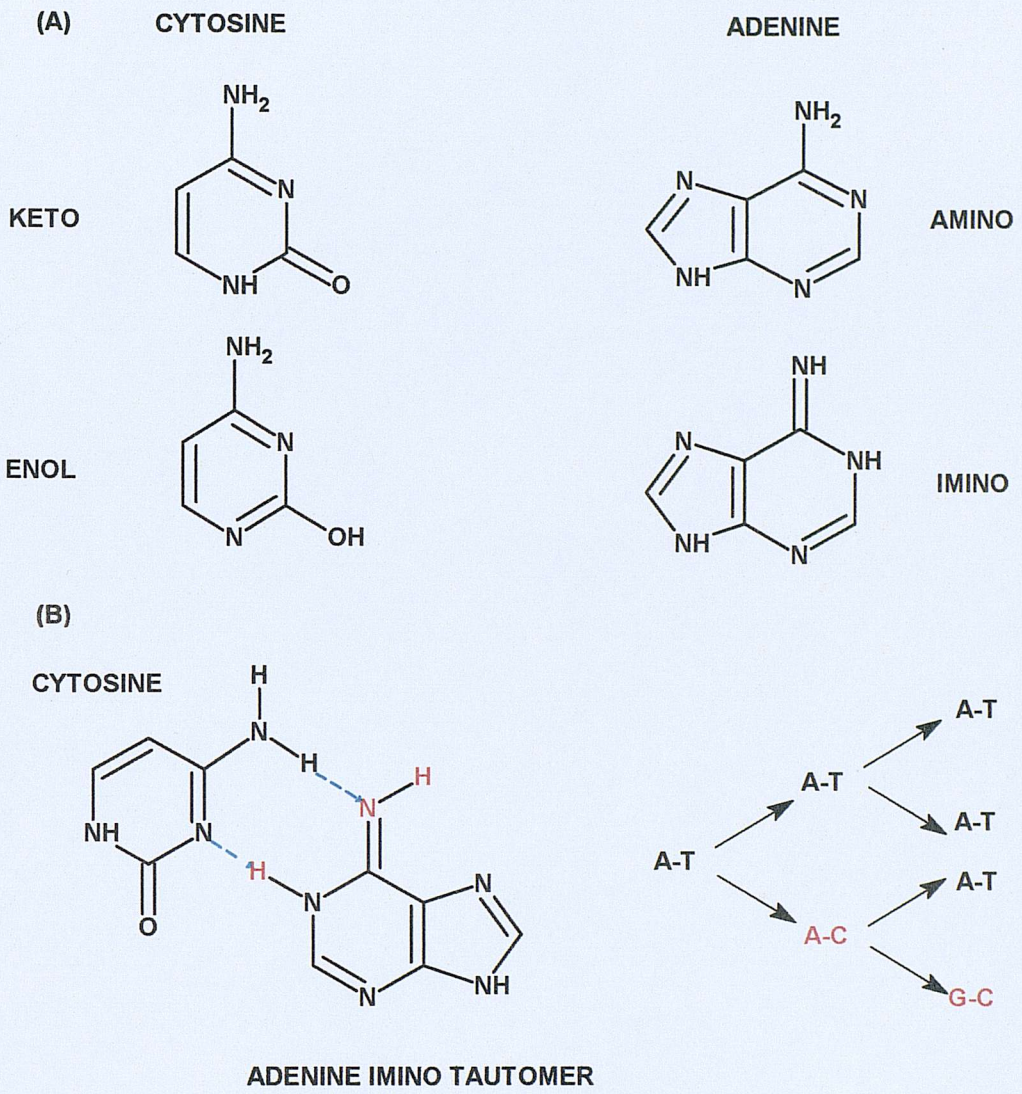


Figure 1.4. Tautomeric shift. (A) The normal keto and rarer enol tautomer of cytosine and the normal amino and rarer imino tautomer of adenine. (B) The mispairing of cytosine with the adenine imino tautomer, which leads to a GC transition.

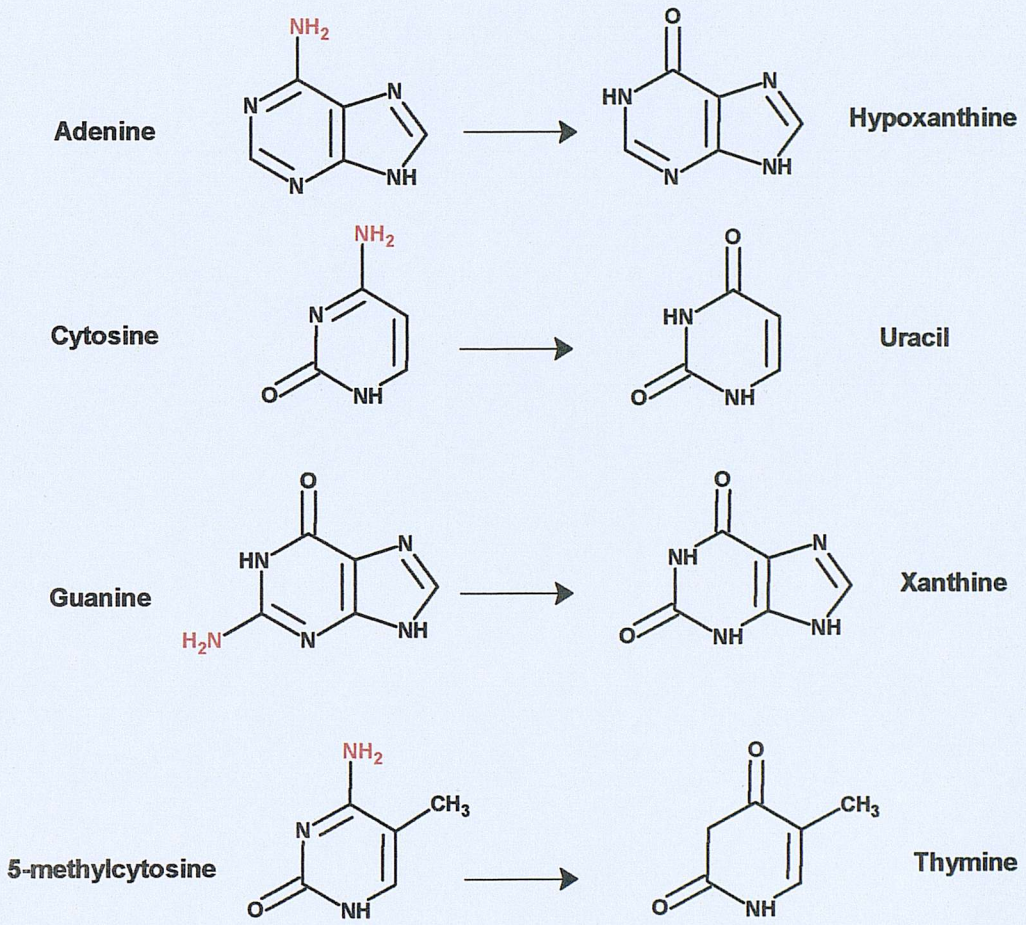


Figure 1.5. Base deamination. Deamination adenine, guanine, cytosine and the cytosine derivative 5-methyl cytosine to uracil, hypoxanthine, xanthine and thymine respectively.

1.5.2.4. Tautomeric Shifts

It is assumed that the majority of bases exist in the most favourable tautomeric form which gives rise to the correct base pairings (Watson and Crick, 1953). However the four nucleotide bases adenosine, cytosine, guanine and thymine can undergo a spontaneous rearrangement of bonding. This tautomeric shift forms a structural isomer of the base, which in turn affects the base pairing properties of the base allowing mispairs to arise. If either a cytosine or adenine shifts in to their imino (NH) tautomeric form they can mispair with the other in the more frequent amino (NH_2) form via two hydrogen bonds (Fig.1.4). A similar mispairing reaction can occur between thymine and guanine involving the oxygen on the C-6 atom of the base. Normally the C-6 oxygen is in the keto (C=O) form, however the oxygen can undergo a tautomeric shift to the less frequent enol (C-OH) configuration. Similarly to the cytosine and adenine mispairing it is possible for a guanine in the keto form to mispair with thymine in the enol configuration or vice versa via three hydrogen bonds. If these shifts occur during replication misincorporation into the daughter strand can occur. For example if in the template strand a guanine changes to its rare enol form just prior to it being copied in replication, an incoming thymine will base pair instead of the correct cytosine. This leads to a GC-AT transition. Secondly if an incoming guanine temporarily enolizes, it will base pair with thymine rather than cytosine causing an AT-GC transition. The fraction of each base being in its less favourable imino or enol tautomer is 10^{-4} (Stryer, 1988).

1.5.2.5. Base Deamination

The loss of exocyclic amino groups from adenine, guanine, cytosine and the cytosine derivative 5-methyl cytosine (control element in higher eukaryotes), known as deamination, also occurs spontaneously and results in the conversion of these bases to uracil, hypoxanthine, xanthine and thymine respectively (Lindahl, 1993)(Fig. 1.5). These bases can cause base mispairing during DNA replication resulting in the alteration of the genetic code. Cytosine deaminates to uracil, a base usually found in RNA molecules.

It has been suggested that using thymine in DNA enables the cell to determine the deamination product of cytosine as inappropriate (Friedberg et al., 1995). The cytosine analogue 5-methyl cytosine forms thymine on the loss of its exocyclic amino group. These T-G lesions are less repairable than U-G due to the fact in U-G lesions it is apparent which base is the inappropriate one. The deamination of cytosine and 5-methyl cytosine is far more common than deamination of DNA purines, adenine deamination occurs at 2-3% of the rate of cytosine deamination (Karran and Lindahl, 1980).

1.5.2.6. Uracil Incorporation

Typically thymine is incorporated into DNA to base pair with adenine. The relative pool of dUTP is significantly smaller than the dTTP pool, however it is possible for a small amount of uracil to be accidently incorporated in DNA during replication. The presence of uracil will not affect the fidelity of replication because uracil will still correctly base pair with adenine. However it will affect the recognition of specific sequences in the DNA required by various enzymes and DNA binding proteins and therefore can be deleterious.

1.5.2.7. Depurination and Depyrimidination

Thermal fluctuations and alterations in pH disrupt the N-glycosyl linkage between DNA base and the deoxyribose sugar. The result of this disruption is the loss of purines (depurination) and pyrimidines (depyrimidination).

Depurination and depyrimidination follow the same mechanism, with base protonation followed by direct cleavage of the glycosyl bond. In the case of pyrimidines the glycosyl bond between the base and the sugar is more stable hence the rate of loss of pyrimidines is lower than that of purines (Friedberg *et al.*, 1995). Based on rates of depurination in *E.coli* and taking into consideration the longer replication and size of the human genome it has been estimated that there are between 2000-10000 depurination events per generation (Lindahl and Nyberg, 1972).

1.5.2.8. Oxidative Damage to DNA

DNA is prone to damage from the attack of reactive oxygen species. These reactive oxygen species arise from both intra and extracellular sources. Leakage of reactive species associated with the reduction of oxygen during respiration from mitochondria is thought to be the major source of oxygen radicals intracellularly (Friedberg *et al.*, 1995). Extracellular sources include ionising radiation, ultra violet light and numerous chemicals.

Oxygen readily accepts unpaired electrons giving rise to a series of partially reduced oxygen species. These oxygen species include singlet O, peroxide radicals ($\cdot O_2$), hydrogen peroxide (H_2O_2) and hydroxyl radicals ($\cdot OH$) (Riley, 1994). Of these species the most reactive and therefore potentially the most damaging is the hydroxyl radical. H_2O_2 is the most widely diffused oxygen species and has relatively mild oxidative capabilities. Its significance to DNA damage is that it ultimately leads to the generation of the more damaging $\cdot OH$ (Imlay and Linn, 1988). The attack of oxygen radicals on DNA can produce a variety of products depending upon point of attack. $\cdot OH$ attack on C-8 of guanine forms the derivative 8-hydroxyguanine that can generate a mispairing legion. 8-hydroxyguanine potentially mispairs with a number of bases. In one study the presence of the guanine derivative as template produced G-T substitutions whilst misincorporation of the analogue as a substrate caused A-C substitutions (Cheng et al., 1992). In both cases 8-hydroxyguanine mispaired with adenine but in the case of the A-C substitutions, 8-hydroxyguanine subsequently paired with cytosine. Other effects of radical attack include base loss and strand fragmentation.

1.5.3. Extracellular Sources of DNA Damage

External or environmental factors can bring about damage to DNA. These factors can be split into physical agents like ionising radiation or ultra violet light or chemical agents such as alkylating agents and intercalating agents.

1.5.3.1. Ionising Radiation

We are all exposed to ionising radiation both from natural sources and increasing from man made sources. Ionising radiation causes cellular molecules to move from their ground state to an excited state. This can have direct and indirect effects on DNA. Direct effects involve the interaction of radiation energy directly with DNA. Indirect effects involve the interaction of DNA with reactive species. Due to the prevalence of water in biological systems radicals produced from the ionisation of water form the main reactive species. These oxygen radicals affect the DNA as described in 1.5.2.8.

1.5.3.2. Ultra Violet (UV) Light

UV radiation causes cross-linking of DNA to proteins, strand breaks and numerous DNA lesions (Friedberg *et al.*, 1995). The most common DNA lesion associated with UV damage is the dimerisation of adjacent pyrimidine bases, with the majority being thymine dimers. These lesions can block DNA replication and affect the binding of sequence specific DNA binding proteins. UV damage that can result in the alteration of the genetic code is the UV irradiation of cytosine which can form cytosine hydrate via the addition of water across the 5,6 double bond of cytosine. A small proportion of the irradiated cytosines will form uracil hydrate. These hydrate derivatives undergo dehydration to yield cytosine or uracil respectively (Boorstein et al., 1990). Therefore in the small proportion that formed uracil a mispairing has occurred that will lead to C-T transition. Similarly the photoproduct of 5-methylcytosine, 5-methylcytosine hydrate can undergo deamination to thymine hydrate which upon dehydration will form thymine (Vairapandi and Duker, 1994).

1.5.3.3. Alkylating Agents

The nucleophilic centre of macromolecules like DNA can come under attack from electrophilic compounds termed alkylating agents. All four bases contain many potential sites for alkylation, although the reactivity for this

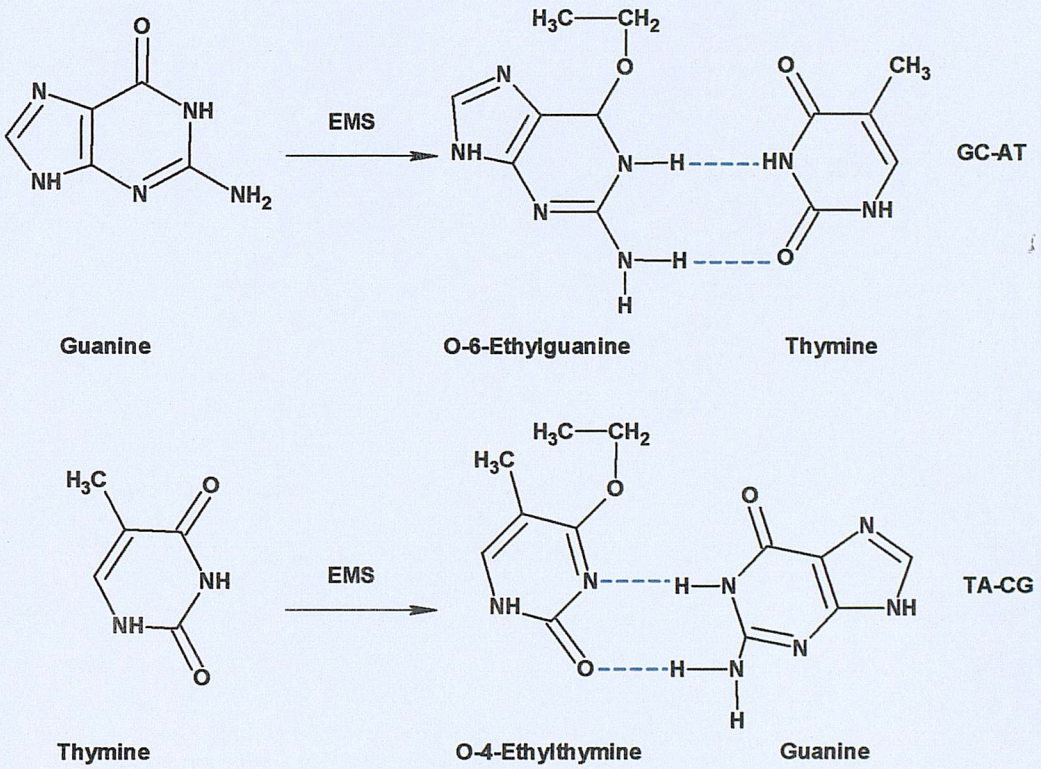


Figure 1.6. Effect of the alkylating agent ethylmethanesulfonate (EMS) that adds an ethyl group to O-6 of guanine and O-4 of thymine. The modified bases can mispair with thymine and guanine respectively producing transition mutations. The hydrogen bonds are coloured blue.

reaction may differ between bases and the extent and distribution of the modifications depends markedly upon the chemistry of the alkylating agent. Alkylating agents come from both environmental sources and as products of metabolic activities. They are divided into monofunctional agents and bifunctional agents depending on the number of nucleophilic sites they interact with. Monofunctional agents tend to cause mispairing lesions and are efficient DNA modifying compounds. Two such alkylating agents ethylmethanesulfonate (EMS) and nitrosoguanidine (NG) add an ethyl group or methyl group respectively to various sites in DNA bases. In terms of altered base pairing modification of the O-6 of guanine and O-4 thymine lead to direct mispairing (Griffiths, 1999)(Figure 1.6). For example treatment of DNA with the alkylating agent N-ethyl-N-nitrosourea produces the guanine derivative O-6-ethylguanine. This not only forms G-A transitions but can also cause frameshift mutations (Eckert and Hile, 1998).

Therefore it can be seen that alkylating agents that cause mutations to the genome, which alter the genetic code, arise by one of two processes. An alkylated base such as guanine forms an abasic site following replication upon further rounds of replication the genome loses a GC base pairing at the original site of the alkylated base causing a frame shift mutation. The second process involves the mispairing of thymine with the alkylated guanine causing a transition mutation with the original GC base pairing being replaced with AT.

1.5.3.4. Intercalating Agents

These agents are flat planar molecules such as ethidium bromide, a widely used DNA staining dye, that slip between the stacked nitrogenous bases in the DNA helix (intercalate). The presence of an intercalating agent disrupting the structure of the DNA typically causes the insertion or deletion of one or more bases. This can cause a frame shift mutation.

1.5.4. Repair Processes

Although it may seem that our DNA has little chance against the wealth of insults it is exposed to, the cell is well equipped to deal with DNA damage

as it occurs. Fortunately cells have developed numerous mechanisms to combat and correct errors to maintain the fidelity of the genetic code otherwise the great variety of DNA damaging agents would be an overwhelming problem for cells.

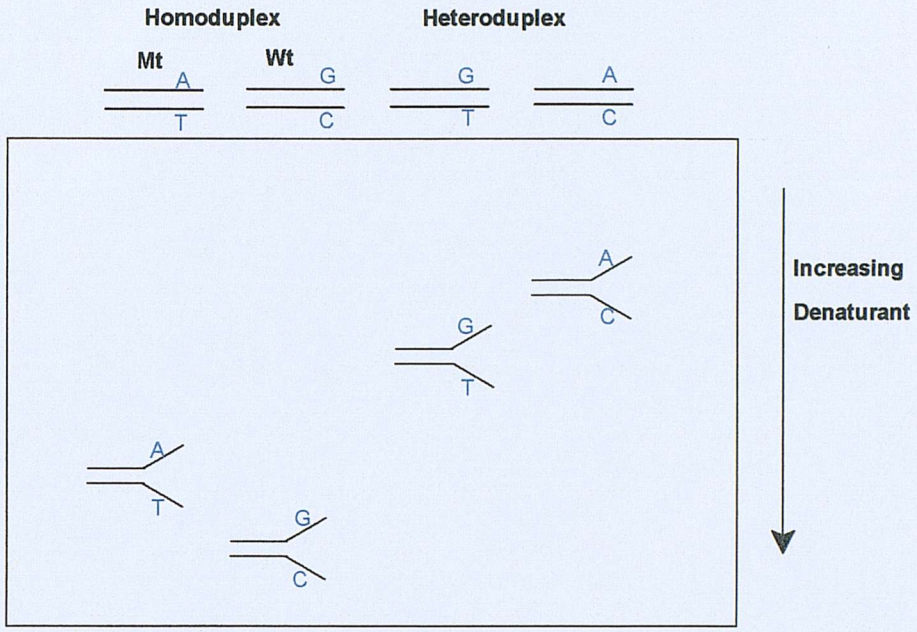
A recent compilation of the number of human DNA repair genes, taken from the draft sequence of the human genome, whose function has been linked to the repair of DNA and those showing strong sequence homology to repair genes in other organisms described 130 known genes (Wood et al., 2001). Their gene products are involved in the distinct repair pathways, base excision repair, nucleotide excision repair and mismatch repair and numerous other processes linked to direct reversal of DNA damage, recombination and rejoining pathways.

1.6. Screening Methods for Detection of Unknown Point Mutations

As outlined above the ability to detect mutations is vitally important. To further our understanding of human genetics it is essential that we can identify and characterise changes in the DNA sequence of a gene that produce both common polymorphisms and rare mutations. It is important that these changes are classified and the phenotypic changes induced are identified. This need has lead to the development of a number of screening methods to avoid the repetitive sequencing of large sections of DNA. There are numerous techniques for the detection of known or unknown sequence variations (Kirk et al., 2002). The choice of which is influenced by a number of factors including throughput, cost, sensitivity and specificity.

Current techniques for detection of unknown mutations can be broadly classified into three groups. The first group rely upon the physical properties of DNA and the differing migration of homoduplex DNA versus heteroduplex DNA or DNA with sequence variations. The second group involves chemical modification in the identification of mutations. The third group utilises mismatch cleavage or binding properties of a number of proteins to identify mutations. Ultimately to classify the mutation it is necessary to confirm the change by sequencing, however by screening prior to sequencing unnecessary work and cost is avoided.

DGGE



SSCP

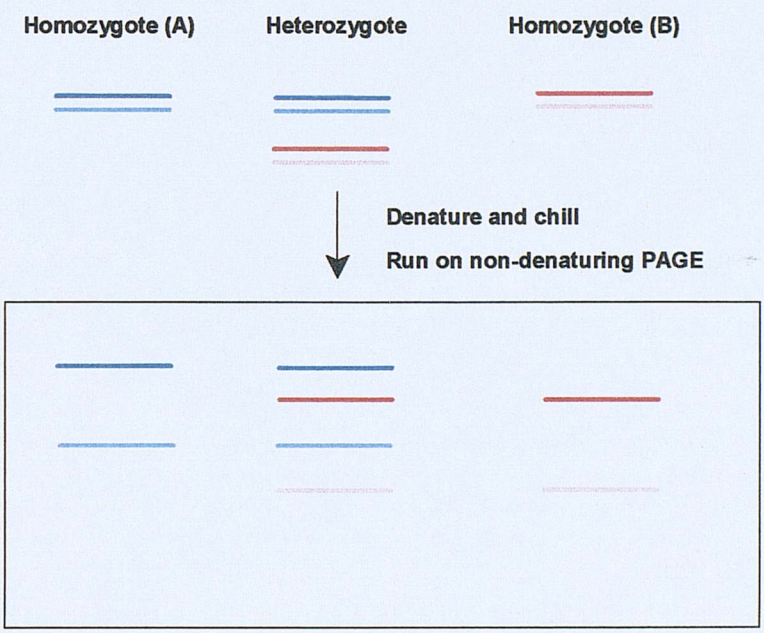


Figure 1.7. Principle of DGGE and SSCP. DGGE- In a Heterozygous individual there will be a homozygous normal and a homozygous mutant. During PCR two heteroduplexes and two homoduplexes will form. Subjecting them to a increasing denaturant will cause the DNA to melt dependent upon there sequence. Diagram adapted from Cotton, 1997. SSCP – Homozygote wildtype (A) and homozygote mutant (B) when denatured form two distinct bands when run on a non-denaturing gel. A heterozygote forms four bands.

1.6.1. Physical Methods

1.6.1.1. Denaturing Gel Electrophoresis (DGGE)

Since its conception (Fischer and Lerman, 1979) DGGE has become a commonly used technique in the identification of unknown mutations. The two strands of double stranded DNA, when subjected to electrophoresis through a gradient of increasing denaturant, are forced apart. This melting of the duplex is dependant upon the DNA sequence, with GC base pairs more strongly held together than AT. Therefore within a given DNA sequence there will be a region that melts at a lower temperature compared to the rest of the duplex, this is referred to as the lower melting domain. As one end of the DNA comes apart the un-melted (higher melting domain) portion holds the two DNA strands together. As soon as one end of the duplex melts, mobility is reduced through the gel. If the area, which melts first differs from an identical sequence by a single base pair the melting points of the two duplexes are different and thus mobility is different. Therefore two duplexes can be separated on a gel that has an increasing gradient of denaturant (Fig. 1.7).

Since its original conception the technique has undergone many developments most notably the addition of a GC clamp (Myers et al., 1985a). The GC clamp is a stretch of GpC attached to the end of the duplex to create an artificially high melting domain, so that the whole region of the DNA duplex of interest becomes the low melting domain and thus improves the accuracy of the technique.

DGGE can detect close to 100% of point mutations when used in conjunction with a GC clamp. The maximum fragment size suitable for DGGE is 1000bp, as increasing length gives rise to an increasing number of melting domains, decreasing the mobility shift. Factors that work against DGGE being the detection tool of choice are the commitment to set up, costly GC clamps and no mutation localisation information.

1.6.1.2. Single Strand Conformation Polymorphism (SSCP)

SSCP is one of the most widely used mutation scanning techniques

mainly due to its simplicity. First described in 1989 (Orita et al., 1989) this method is based on the observation that under certain conditions single stranded DNA folds in a sequence specific manner. The flexibility of ssDNA coupled with the sequence of the ssDNA allows it to adopt a conformation determined by intramolecular interactions and base stacking. A single base change can therefore cause a difference in these interactions causing the DNA folding to be altered. This affects its electrophoretic mobility through polyacylamide gels run under non-denaturing conditions.

The sample is denatured to form ssDNA and chilled on ice to prevent the single stranded DNA reannealing and reforming dsDNA. On a non-denaturing gel a homozygous wildtype sample will generate two distinct bands the location of which is determined by the folding of the ssDNA. Similarly a homozygous mutant sample will generate two bands that migrate differently compared to the wild type sample. A heterozygote sample will therefore form four bands corresponding to the two wildtype and two mutant bands (Fig. 1.7). Bands of the single-stranded DNA at positions in the gel different from the wildtype indicate the presence of a mutation.

The successful detection of sequence differences depends upon finding optimal conditions; therefore this technique is dependent on experimental conditions. SSCA has a detection rate of 75-95% under optimal conditions, with an optimal fragment size of 150-200bp (Cotton Richard, 1997). To achieve 95% detection, requires the gels to be run under several conditions due to the conformation of ssDNA being sensitive to a variety of parameters including ionic strength, gel matrix, fragment length and temperature. Perhaps the major advantage of this technique is the lack of special equipment or reagents. In common with DGGE this technique does not give any localisation information.

1.6.1.3. Heteroduplex analysis (HA)

This technique is often used in conjunction with SSCA to overcome the deficiencies of both techniques. Like SSCP there is no complex theory involved and carrying out HA is relatively simple. Both techniques involve mobility shifts of DNA running on non denaturing gels but whereas SSCA

uses ssDNA, HA utilises dsDNA that has been manipulated to form homoduplex or heteroduplex DNA. Due to a more opened double strand configuration surrounding the mismatches in heteroduplex DNA in comparison to the corresponding homoduplex, heteroduplexes will demonstrate a reduced mobility when run on non denaturing gels.

Heteroduplex DNA is generated from a heterozygote by heat denaturation and reannealing of the wild type and mutated DNA. When run under non-denaturing conditions homoduplexes and heteroduplexes exhibit distinct electrophoretic mobilities and thus a mutation can be detected. If the sample is not heterozygous it is necessary to mix the homozygous mutant allele with wild type to generate the necessary heteroduplexes. The optimal fragment size varies between 200-600bp and HA has a detection sensitivity of 80% (Cotton Richard, 1997). As with all the physical methods the location of the mutation cannot be determined.

1.6.2. Chemical Methods

1.6.2.1. Chemical Cleavage of Mismatch (CCM)

Not all screening methods rely upon the difference in physical properties between heteroduplex and homoduplex DNA. Chemical cleavage of mismatches (CCM) utilises two chemicals hydroxylamine (H) and osmium tetroxide (OT) and a cleavage reaction catalysed by piperdine to detect the presence of mutations (Cotton et al., 1988). Following heteroduplex formation hydroxylamine (H), reacts with mismatched C bases and osmium tetroxide (OT), reacts with mismatched T bases. The reaction of the two chemicals is detected using piperidine, which cleaves the modified duplex at the point of the mismatch and separation of the cleavage fragments via electrophoresis on a denaturing gel. To detect all mutations it is necessary to use both strands of the wild-type as probe as mismatched A and G in the sense probe are then detected as T and C bases in the anti sense probe.

Adaptations to this technique have included the replacement of the hazardous osmium tetroxide with potassium permanganate. A second disadvantage of this technique is the numerous ethanol precipitations required

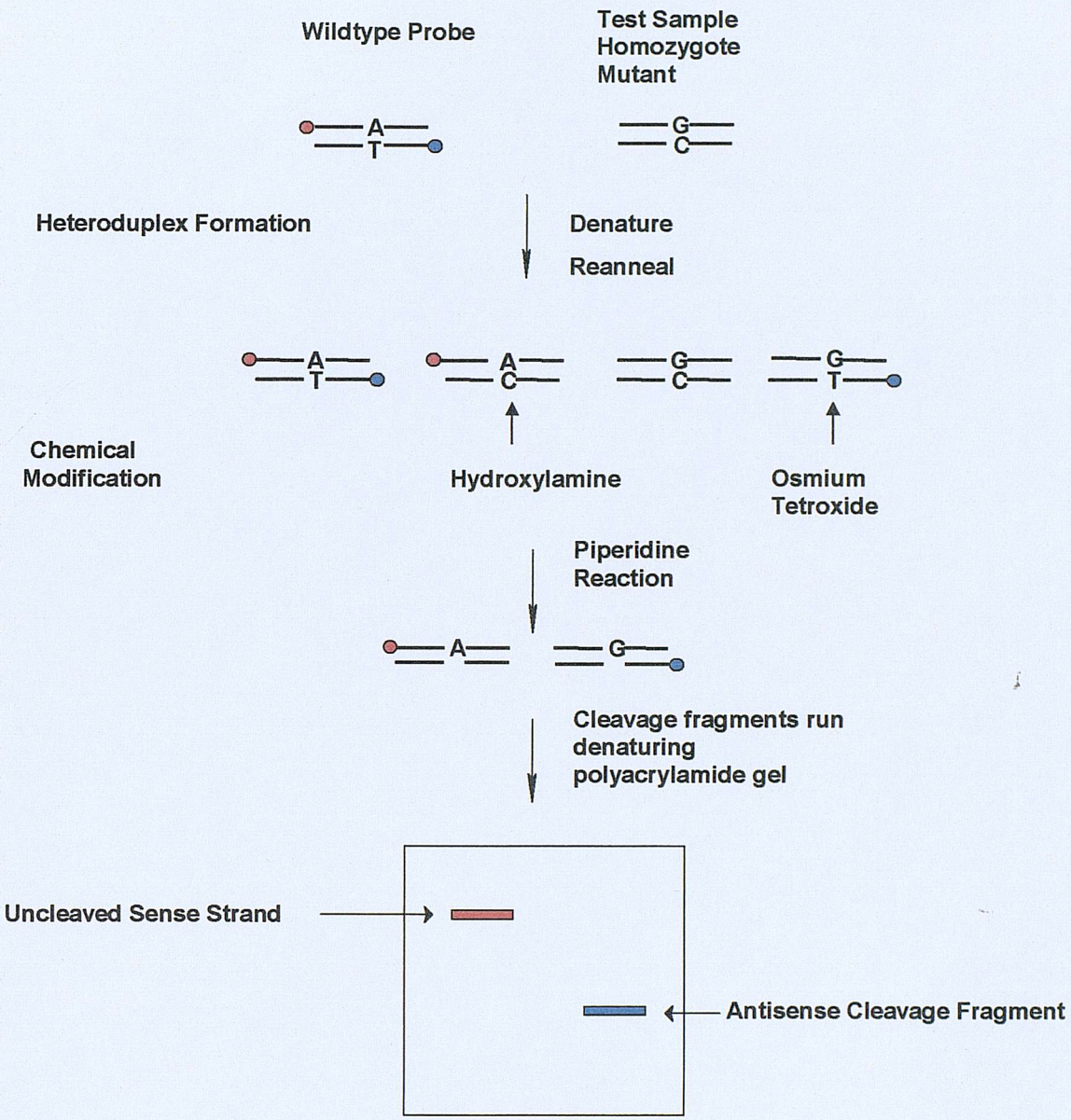


Figure 1.8. Schematic representation of chemical cleavage of mismatch DNA. In the example above a wildtype probe labelled on both the sense and antisense strand is hybridised with a mutant homozygote test sample. Heteroduplex formation results in two heteroduplex species each with a single labelled strand. Following chemical modification and cleavage reaction, the resulting fragments are run on a denaturing polyacrylamide gel. The A•C mismatch fails to produce a specific cleavage fragment, as the labelled strand remains unmodified, whilst the G•T mismatch produces a specific cleavage fragment as the labelled strand has been chemically modified. If the test had been a wild type homozygote then all the DNA after hybridisation would have been homoduplexes therefore the specific cleavage fragment, observed for the G•T mismatch would not be seen.

(3). Labelling the DNA with biotin and binding this to streptavidin-coated magnetic beads has reduced the number of precipitations (Hansen et al., 1996). CCM shows high sensitivity and specificity as well as giving localisation information however it is technically demanding and the use of hazardous chemicals has restricted its use.

1.6.3. Protein Based Mutation Techniques

Protein based techniques offer the prospect of simplified protocols and less hazardous manipulations, which makes them an attractive option for mutation detection. Typically these methods tend to be based on either cleavage or binding to a DNA duplex containing a mismatch base pair or small insertions or deletions. Mismatched DNA is usually generated by hybridisation of mutant and wild type DNA. The difference in sequence generates heteroduplex DNA that induces structural distortions providing the protein with a suitable target site that it either cleaves or binds to. Techniques employing enzymes are often limited by enzyme activity or mismatch recognition specificity and also the availability of some of these proteins.

Outlined below is a brief description of proteins that have shown potential for mutation scanning for a more in depth review see (Taylor, 1999). Proteins that have been reported in mutation methods include ribonucleases, Cleavase, DNA glycosylases, Cel I, MutS and T4 endonuclease VII. Although to date a number of proteins have been described and successfully used for mutation detection, such techniques are not as widely used as SSCP or DGGE.

The two proteins chosen for investigation to develop mutation scanning using short-track electrophoresis were MutS (mismatch binding) and T4 endonuclease VII (mismatch cleavage) and therefore a more in depth look at them is carried out in section 1.7 & 1.8. There were several reasons for the choice of these two proteins. Firstly was their availability, clones were available for both and so the quantities required for high throughput were available. Secondly previous published work had met with some success at applying them to mutation detection. Thirdly the choice of a mismatch binding

protein and a mismatch cleavage enzyme enabled a comparison between cleavage assays and band shift assays and the suitability for use with MADGE.

1.6.3.1. Restriction Digests

Endonucleases such as restriction enzymes have already been successfully utilised for mutation detection. Restriction fragment length polymorphism involves cutting the DNA with a restriction enzyme and observing the pattern of bands. Mutations in the recognition sequence of the restriction enzyme will prevent cutting or create a restriction site. Obviously these enzymes are severely restricted in their application due to the dependence on specific sequence recognition.

1.6.3.2. Ribonuclease Cleavage of Mismatch

Ribonuclease A cuts RNA at single base pair mismatches with the cuts occurring 3' to pyrimidines. This cleavage occurs whether the strand is hybridised to DNA or RNA and has been utilised for mutation detection (Myers et al., 1985b). Annealing of normal and mutant DNA with an RNA probe produces homoduplex and heteroduplex. This is followed by digestion with RNase A that cleaves at mismatch sites in the heteroduplex DNA. The cleavage products are then separated and characterised on a denaturing polyacrylamide gel. Mutations are detected and localised by the presence and size of the RNA fragment generated by cleavage at the mismatch. The Ribonuclease cleavage method detects 60-80% of point mutations if both strands are cleaved. The major disadvantage is that RNA needs to be produced from DNA to be scanned.

1.6.3.3. Cleavage Fragment Length Polymorphism

Cleavage fragment length polymorphism utilises an enzyme termed Cleavase that cleaves stem-looped structures. Cleavase has a 5'-3' endonucleolytic action, which cleaves on the 5' side of the junction.

Denaturing DNA, then cooling quickly to prevent reannealing of the ssDNA, forms stem loops. (Rossetti et al., 1997). The primary sequence of the DNA determines the stem loop profile, thus cleavage patterns are sequence dependant. Unlike the other methods this method does not require the formation of heteroduplex DNA to act as a substrate.

1.6.3.4. CEL I

A novel nuclease termed CEL I, purified from celery, displayed mismatch cleavage specificity (Yang et al., 2000). CEL I introduces an incision on the 3' –side of the mismatch in one of the two strands in heteroduplex. This cleavage was applied in a simple mutation detection assay, where *BRAC1* exons were treated with CEL I (Oleykowski *et al.*, 1998). In this study 100% of the sequence variants present were detected.

1.6.3.5. Mut-Y

Mut-Y is a member of the DNA glycosylase family whose function is to repair mismatched GA base pairs. Cleavage occurs on the adenine containing strand of the GA mismatch. E.*coli* MutY has been used in the detection of mismatched GA in p53 (Lu and Hsu, 1992). The limitation of this enzyme is that it only detects GA mismatches. To increase mismatch recognition MutY was used in conjunction with a second enzyme a thymine glycosylase which cleaves T at GT mismatches (Hsu et al., 1994). This method however does not offer full mismatch detection, as it is unable to detect CC or GG mismatches.

1.7. MutS

MutS and its homologues are highly conserved proteins found in both prokaryotic and eukaryotic cells. MutS is a component of the methyl directed mismatch repair pathway along with a series of other proteins MutL and MutH. MutS's role within this pathway is to provide the initial mismatch recognition in mismatch repair. Although the recent publication of the crystal structure of *E.coli* MutS (Lamers et al., 2000)and *Thermus aquaticus* (Taq) *MutS* (Obmolova et al., 2000) bound to heteroduplex DNA has provided great insight between the interaction of MutS with DNA, many areas of the molecular mechanism remain unclear and are the subject of much debate.

1.7.1. Structure

In solution MutS readily forms dimers and tetramers (Su and Modrich, 1986) with higher order aggregates seen with high protein to DNA ratio (Allen et al., 1997). *E.coli* MutS is a 95kDa 800 amino acid protein whose shape resembles a comma. The crystal structure revealed that MutS comprises five distinct structural domains (Fig. 1.9). In the functional dimer each of these five domains is structurally identical between the two subunits. Domain I is located at the N-terminal region of the MutS protein and deletion studies (Wu and Marinus, 1999) have shown this domain to contain the amino acid motif required for mismatch recognition.

Domain VI or the 'clamp' is also involved in DNA binding. It is believed to be involved in the initial recognition and binding of DNA rather than mismatch recognition. Data from the crystal structure indicates that domain I and IV are independently folded and attached to the core of MutS by flexible peptide linkages. In the absence of bound DNA domain I and IV were too mobile to be discerned in the crystal structure (Obmolova *et al.*, 2000). It was only upon binding DNA that both domains were brought into the dimer interface.

Domain II, III and V are more rigid, retaining similar structures in the presence or absence of DNA. Domain III shares the most interactions with any of the other domains and is central to the MutS structure. Deletion

Figure 1.9. MutS monomeric protein (Lamers *et al.*, 2000)(PDB file accession number 1E3M). The different colours correspond to the different distinct domains of MutS. Domain I (blue), domain II (orange), domain III (red) domain IV (dark green), domain V (light green) and the helix-turn-helix motif (yellow).

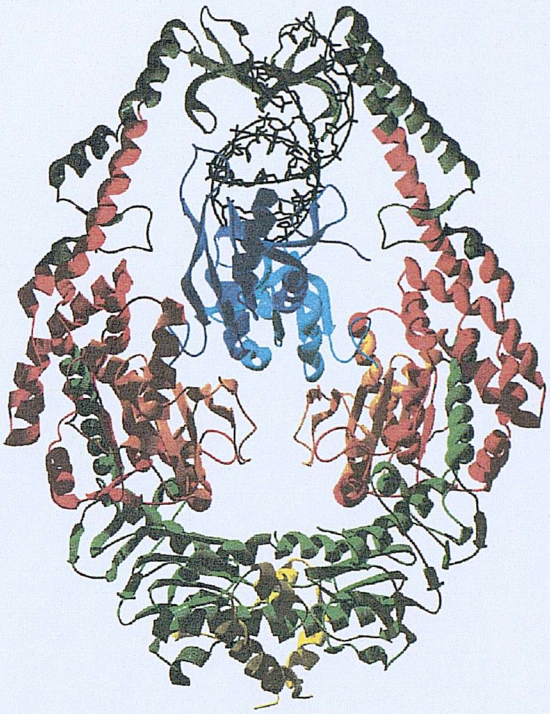
experiments indicated that the C- terminal end of MutS was responsible for efficient dimerisation (Wu *et al.*, 1999). Subsequently it was shown that Domain V situated at the carboxyl terminal along with domain I and IV formed the interface between the two MutS subunits. Within domain V a pair of antiparallel helices assume a helix-turn-helix conformation, which is the prime component of the dimer interface. Proteins lacking this helix-turn-helix motif were found to be predominantly monomers in solution (Biswas et al., 2001) The helix-turn-helix from one domain V reaches across to the other subunit domain V and interacts with its helix-turn-helix mainly through hydrophobic interactions. This interaction comprises 70% of the total dimerisation interaction (Obmolova *et al.*, 2000) between MutS subunits, with domain I and IV contributing the remainder. Disruption to the helix-turn-helix motif (Biswas et al., 2001) leads to a loss of dimerisation suggesting the remaining dimer interaction only occurs after DNA binding.

1.7.2. DNA Binding

Upon dimerisation the structure of MutS has been described as a pair of praying hands an analogy that helps visualize the way in which the subunits come together and their interaction with DNA. The thumbs are folded inwards creating a passage into which DNA passes. The DNA is held in the ' fingers' (domain IV) and the DNA is probed by domain I (Figure 1.10). MutS forms a dimer in the presence or absence of DNA however the MutS dimer in the absence of DNA is less structured with the domains disordered (Lamers *et al*, 2000). Referring to the praying hands analogy the observed disorder can be used to offer an explanation as to how DNA is bound.

Observations from the crystal structure show limited interactions between the two clamp domains on both subunits. Situated around the interface are a high number of positive charges that prevent a tight clamp interaction to occur in the absence of DNA. The electrostatic repulsion forces an open conformation providing a docking area for DNA. The phosphate backbone of the DNA upon entering the protein complex would neutralise the repulsion generated by the basic residues and the two clamp domains would fold around and bind the DNA.

(A)



(B)

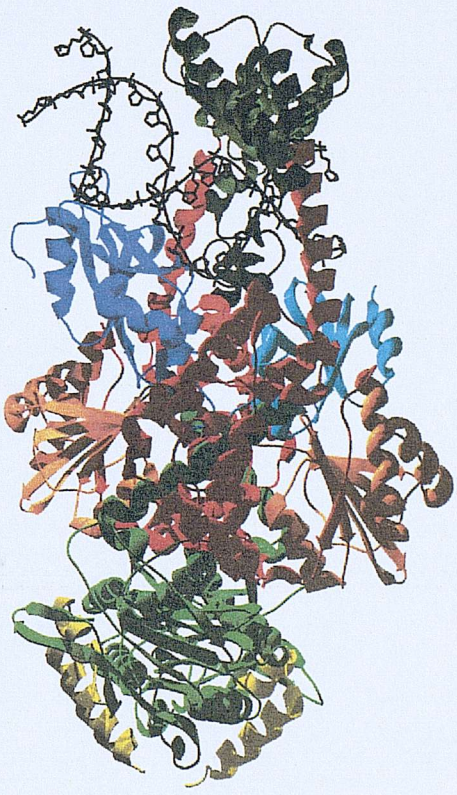


Figure 1.10. MutS-DNA complex front view (A) and side view (B). The MutS dimer with bound DNA (black) held in the DNA binding site. The separate domains are coloured as in figure except for domain I of the right hand monomer (blue) and domain I of left hand monomer (pale blue). The lighter helix-turn-helix motif corresponds to right hand monomer indicates the area of major interactions required for dimer formation. It can be seen that domain I of one subunit is positioned much closer to the DNA. Also of interest is the bend in the DNA caused by MutS

Mutation studies (Yamamoto *et al*, 2000) upon Phe36 of *E.coli* have shown this residue to be important in mismatch binding. Replacement of this amino acid resulted in a loss of binding of MutS to heteroduplex and homoduplex DNA. Bound DNA is kinked by 60° allowing Phe39 residue in the 'thumbs' to probe the DNA for mismatches. The Phe residue scans the DNA for mismatches by inserting at every base step. The energy needed to bend the DNA is lower at the location of a mismatch when compared to that of normal B DNA. Therefore the ability of MutS to detect mismatches or insertions/deletions is reliant upon weakened hydrogen bonding and stacking interactions.

The DNA binding site is a composite site consisting of domains from each of the two MutS subunits. Although both subunits are involved in the binding of DNA only a single Phe residue from domain I is used to probe for a mismatch. The domain I from the first subunit contacts the mismatch directly and bridges the widened minor groove formed by the bending of the DNA. Domain IV from the second subunit spans the major groove underneath the bound duplex. Domain IV from the first subunit and domain I from the second subunit act to stabilise the kinked DNA.

1.7.3. ATPase Activity

MutS and its various homologues have been shown to have weak ATPase activity (Haber and Walker, 1991). The ATPase activity is located in domain V where there is a highly conserved Walker-type-A nucleotide motif. Although the binding of MutS to duplex DNA is ATP independent (Allen *et al.*, 1997) ATP plays a key role in the subsequent steps after the initial mismatch recognition.

ATP binding reduces the affinity of both bacterial MutS (Allen *et al.*, 1997)and the human homologue MuTSα (Blackwell et al., 1998) for mismatch DNA. The use of terminally blocked DNA has shown that, although ATP reduces the affinity of MutS for mismatch DNA, it still remains associated with the DNA duplex (Blackwell *et al.*, 1998). The ATPase activity of MutS is modulated by the presence of DNA within the DNA binding site. The presence of DNA was observed to alter the structures of the ATPase domains

suggesting an allosteric connection between the ATP and DNA binding site (Obmolova *et al.*, 2000). Observations into the rate of nucleotide hydrolysis (Bjornson et al., 2000) in protein-DNA complexes found that for homoduplex bound MutS the rate limiting step occurred after hydrolysis and therefore was modulated by product release. Conversely for heteroduplex bound DNA the initial ATP hydrolysis event is a lot slower. Therefore the MutS-ATP-mismatch is longer lived than the MutS-ADP-homoduplex complex, a factor that could be important in the subsequent recruitment of MutL. The conformation of MutS has been shown to be influenced by the presence of ADP and ATP (Kato et al., 2001), leading to the suggestion that the DNA-binding activity of MutS may depend on conformational changes triggered by both binding and hydrolysis.

The conformational changes and the release of the mismatch is generally interpreted as the beginning of the search for the strand discrimination signal. The involvement of ATP in this process is hotly debated. The first school of thought is that ATP hydrolysis drives the movement of MutS and is called the translocation model. The translocation model is based upon electron microscopy observations that show the formation of an α-loop on heteroduplex DNA in the presence of MutS (Allen *et al*, 1997). The loop size increased linearly with time suggesting that the MutS dimer was migrating along the helix. The translocation model suggests that ATP hydrolysis provides the energy needed for this unidirectional translocation of MutS away from the mismatch in search of other MMR proteins.

The second model (the switch model) postulates that ATP, upon binding of MutS to a mismatch, acts as a molecular switch replacing ADP(Gradia et al., 2000). This transforms MutS from a mismatch binding protein to a sliding clamp protein. In this state MutS diffuses away from the mismatch in search of other repair proteins. In this case ATP hydrolysis serves to cause a conformational change in MutS reverting it back into a mismatch binding state (Joshi *et al*, 2000).

In both cases MutS leaves the mismatch however if this was the case then knowledge of the mismatch site would be lost. Indeed it has been

observed that in the presence of MutL, MutS is prevented from leaving the mismatch site (Galio et al., 1999). Thus in this case MutS acts as a marker for the position of the mismatch. It has been suggested that MutS remains bound to the mismatch site to invoke MMR (Obmolova *et al.*, 2000). Further theories postulate that part of the MutS holds onto the DNA whilst the rest of the protein scans the DNA for the strand discrimination signal by passing the DNA through the holes in the centre of the dimer (Fig. 1.11)(Sixma, 2001). Alternatively although dimers initially bind DNA the active MutS complex could involve tetramers. In this scenario a MutS dimer binds to the mismatch whilst the other searches for the strand discrimination signal (Lamers *et al.*, 2000).

1.7.4. MutS Specificity

The ability of MutS to discriminate between heteroduplex and homoduplex has been estimated to be 10 - 20 fold (Feng and Winkler, 1995). Observations using the Phe39 mutant indicated that homoduplex binds at the same site as heteroduplex. Upon mutation a weak signal for homoduplex was transformed into an undetectable signal (Yamamoto et al., 2000). The ability of MutS to recognise different mismatches also varies; with G-T being recognised the strongest whilst C-C is very poorly recognised (Su *et al.*, 1986)

1.7.5. Methylated mismatch Repair

MMR role is to correct biosynthetic errors that occur during replication on the newly synthesized strand. This is possible because the newly synthesized DNA strand is for a short period in a transiently unmodified state. Newly formed DNA is not methylated at the strand signal site d (GATC) and it is this that directs repair towards them. As well as recognising single base pair mismatches MMR has been shown to recognise insertion/deletion mutations up to 4bp in length (Parker and Marinus, 1992).

In *E.coli*, upon binding of MutS to a mismatch, a second protein MutL is recruited, a reaction that is ATP dependent. The MutS/MutL complex activates an endonuclease associated with a third protein MutH, a step that very little is known about. The unmethylated strand is cleaved by the

endonuclease MutH at the d(GATC) site. Excision proceeds towards the mismatch where it is terminated beyond the mismatch. The exonuclease involved in the excision of nucleotides depends upon which side of the mismatch excision is initiated. Excision from 5' end depends on RecI or exonuclease VII, whereas from 3' end depends upon exonuclease I. The final step is the filling of the gap left by the exonuclease by DNA polymerase III holoenzyme and ligation of the strands.

1.7.6. Mutation Detection

Isolating MutS has already led to the successful application of MutS to mutation detection. The ability of MutS to protect heteroduplex DNA from exonuclease digestion has been used to detect point mutations and small insertions or deletions (Ellis et al., 1994). The site of the point mutation can be localised to within 20 bases.

Band-shift in polyacylamide gels as a result of electrophoresis has also been applied successfully in detection of heterozygosity at a locus in the cystic fibrosis gene (Lishanski et al., 1994). However the methods described do not meet the demands of a high throughput system. For example MutS protection of heteroduplex DNA from exonuclease digestion requires analysis on a automated DNA sequencer which means that sample numbers are restricted and the time to analyse the samples does not meet the demands of a high throughput system. The application of band shift to the cystic fibrosis gene used conventional gel electrophoresis, which restricts the sample number that can be run per gel, coupled with the radio labelling of the DNA to achieve visualization reduces this methods high throughput capability.

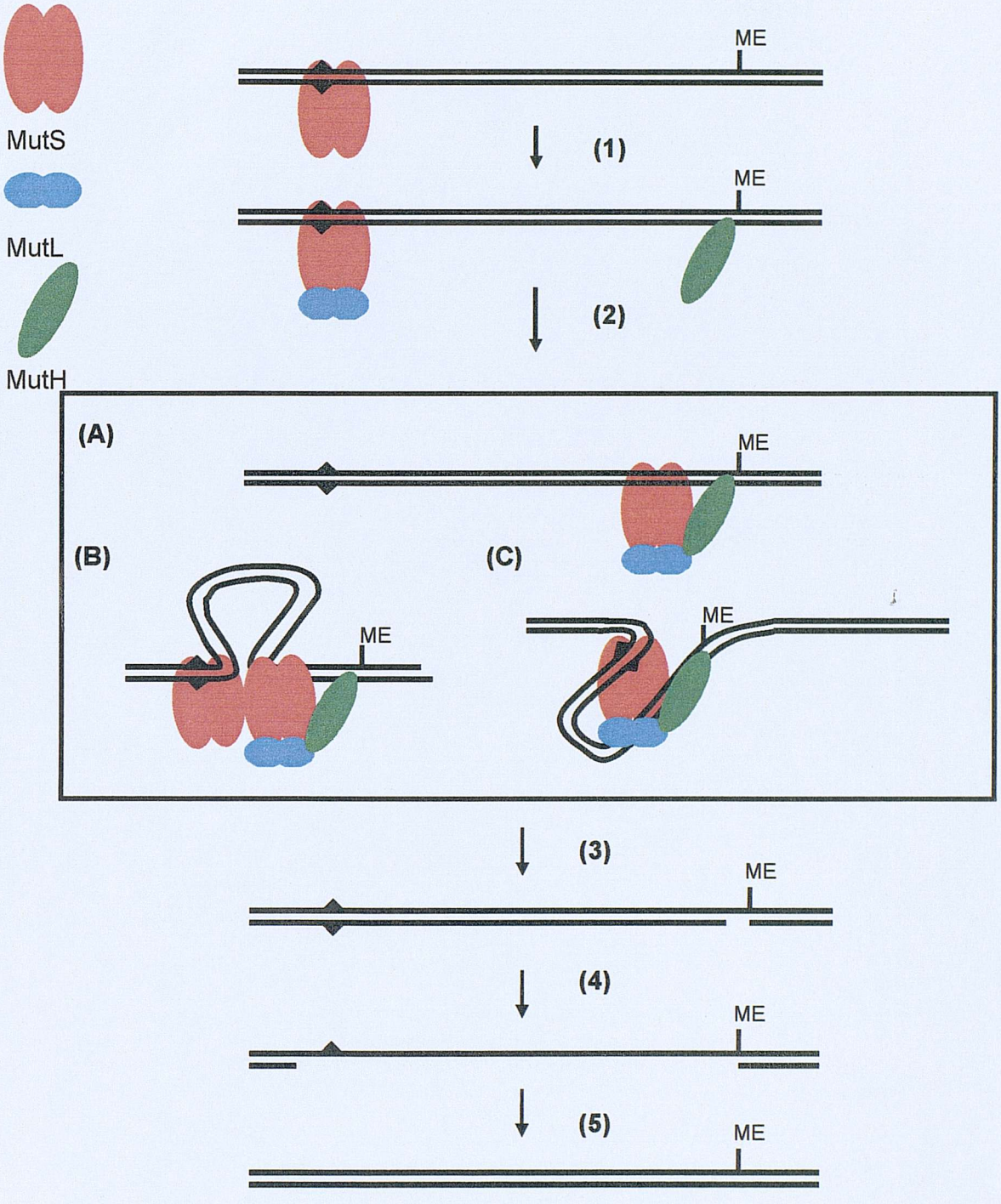


Figure 1.11 Methylated mismatch repair adapted from (Sixma, 2001). After initial recognition of mismatch, MutL is bound a step requiring ATP (1-2), MutH binds to the strand discrimination signal. Three possible modes of searching DNA strand are shown (A, B, C) (A) scanning DNA for the signal, (B) MutS tetrahmer scanning for strand discrimination, (C) threading DNA through MutS whilst maintaining mismatch binding. Newly synthesised strand is then nicked by MutH (3), DNA helices and exonuclease remove a portion of the strand including the mismatch (4). DNA polymaerase and DNA ligase complete the repair (5).

1.8. T4 Endonuclease VII (Endo VII)

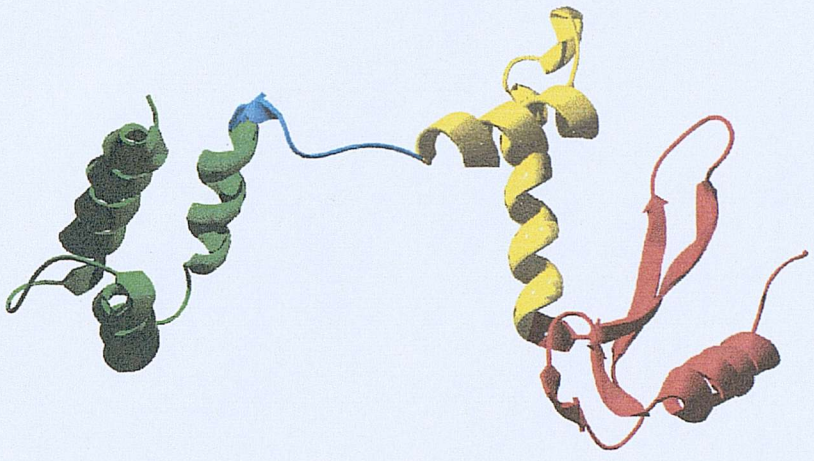
Gene 49 of the bacteriophage T4 produces a 157 amino acid protein termed T4 endonuclease VII (Endo VII). Within the bacteriophage it is produced during the early and late stages of bacterial infection. Its main function is believed to be involved in the resolution of branch points in DNA prior to packaging of the DNA into the phage head (Kemper and Brown, 1976).

An interesting aspect of Endo VII is its broad range of substrates. It was the first enzyme shown to resolve Holliday junctions (Mizuuchi et al., 1982) and its broad range of substrates include Y-junctions, heteroduplex loops, single stranded overhangs (Pottmeyer and Kemper, 1992) and of particular interest for mutation detection single base mismatches (Solaro et al., 1993). Both the high resolution crystal structure and the use of truncated or mutant proteins have given a wealth of information concerning the structure, DNA binding and catalytic properties of the Endo VII protein.

1.8.1. Structure

Endo VII is a functional dimer (Pohler et al., 1996) with a highly unusual dimerisation motif. The crystal structure of an individual wild type Endo VII 157 amino acid monomer indicated that the Endo VII comprised of two distinct flexibly linked domains. These are a 50 residue C-terminal domain (helices 4/5/6) and the 60 residue N-terminal domain (helices 1, loop region and β finger) which are separated by a central helix-loop-helix motif (Raaijmakers et al., 1999). Due to a number of hydrophobic residues found in this central portion of the protein the monomer does not represent a stable form. Therefore Endo VII dimerises to help stabilise this molecule. At the centre of this dimer there is the dimerisation domain. Truncating the protein from both ends to find the smallest dimerising peptide indicated that amino acids 55 to amino acids 105 were needed for successful dimer formation. Mutational studies identified a number of mutants within this stretch of amino acids, that if removed, reduced the dimerisation of the protein, in particular one mutation W87R reduced dimer formation by 95% (Birkenbihl and Kemper, 1998a).

(A)



(B)

Figure. 1.12. Ribbon diagrams of Endo VII (PDB file accession 1E7D). (A) Endo VII monomer with coloured domains, C-Terminal domain (green), N- terminal domain (red), helix-loop-helix motif (yellow) and extended chain segment (blue). (B) Endo VII dimer, individual Endo VII subunits coloured green and red respectively. The N-terminal of one subunit interacts with the C-terminal of the other subunit forming the DNA binding domains.

Upon dimerisation a unique motif is observed within the Endo VII dimer. The C-terminal domain of one monomer interacts with the N-terminal of the other monomer at one end of the dimer, with the same happening on the opposite end (Raaijmakers *et al.*, 1999). The C terminal within this dimer is a compact entity compared to the N-terminal whose stability depends upon the presence of a zinc ion located tetrahedral to four cysteine residues (Giraud-Panis et al., 1995).

1.8.2. DNA Binding

Single Endo VII monomers do not have the ability to bind DNA (Birkenbihl and Kemper, 1998b), it is only upon dimerisation that the protein is able to bind DNA. This is achieved by both ends of the protein forming two DNA binding sites between the N / C domains of the different Endo VII subunits. Both the C-terminal domain (Golz et al., 1997) and the N-terminal domain (Birkenbihl and Kemper, 1998a) are essential for DNA binding. The removal of more than 3 amino acids from the C - terminus abolished the binding of Endo VII. Similarly a removal of 10 amino acids from the N-termini reduced binding drastically whilst the removal of 20 amino acids caused the EndoVII to completely lose the ability to bind DNA. The removal of residues from either end of the Endo VII peptide disrupts the stabilising interactions between the intertwining domains as well as the removal of residues involved in binding DNA.

The close interaction of the C terminal domain and the N terminus domain forms two bays at either end of the molecule for DNA interactions. These are lined with a number of positive residues. In the case of binding to its native substrate (Holliday junctions) the phosphate backbone of the minor groove reaches into these bays, due to the positive charged face of the binding sites. This exposes the phosphates to the catalytic residues and the divalent metal ion (Raaijmakers *et al.*, 1999).

1.8.3. Catalytic Activity

In common with almost all other nucleases Endo VII requires divalent

metal ions for catalysis. These catalytic metal ions in enzymes are generally liganded by acidic amino acids and participate in the cleavage of the phosphodiester bond. The putative active sites were revealed by the location of the divalent ion required for Endo VII activity from the X-ray crystal structure (Raaijmakers *et al.*, 1999). A cleft formed between the β-finger and helix 2 which is lined with a number of residues that, through mutation studies, have been shown to be essential for activity but not for DNA binding. Asp-40, His-41 and His-43 were all identified as being needed for cleavage of DNA junctions (Giraud-Panis and Lilley, 1996). Other essential amino acids are asn-62 (Golz et al., 1997) glu-65 and glu-86 (Pohler *et al.*, 1996). The acidic residues asp-40, asn-62 and glu-65 co-ordinate the divalent metal ion. The precise roles of the other residues are as yet unclear. Endo VII thus contains two catalytic sites, which contain amino acids from a single subunit and act completely independent of each other.

Data from the crystal structure puts the two active sites - 24Å apart. Even with optimal binding, due to steric hindrance it is unlikely that both catalytic sites would contain a phosphate group at the same time, this suggests that cleavage is in two steps, after the first cut there is some readjustment to allow cleavage of the opposite strand (Raaijmakers *et al.*, 1999).

1.8.4. Substrate Specificity

The dimeric Endo VII meets its target DNA as a dimer (Pohler *et al.*, 1996). To investigate the cleavage properties of Endo VII, four cruciform DNA products were formed to act as Holliday junction analogues, each with a different oligonucleotide labelled and used in Endo VII cleavage studies. Resolution of the cruciform DNA indicated that the DNA was nicked twice in a nick and counter nick reaction occurring to the strands across the junction (Pottmeyer *et al.*, 1992). To determine whether this 2 nick reaction is achieved by a single Endo VII dimer or is the result of two independent Endo VII proteins a super coiled stabilized cruciform substrate was used to probe the nicking of the two sites (Giraud-Panis and Lilley, 1997). Fully active Endo

VII converted this super coiled DNA into a linear product whilst a mixture of active and inactive subunits produced a circular nicked product. The formation of circular nicked product indicated that the subunits work independent of one another. Secondly both reactions must take place within the lifetime of the protein-DNA complex. In the case of the fully active protein a single reaction per protein-DNA lifetime would result in a nicked circular product not the linear product observed. Endo VII has a high affinity for branched DNA. An 1000-fold excess of competitor DNA was found to be unable to displace the Endo VII dimer from DNA junctions (Pohler *et al.*, 1996). A heterodimer Endo VII mutant containing one active site was observed to remain tightly bound to the nicked junction thus indicating a high affinity for the nicked intermediate before junction resolution (Birkenbihl and Kemper, 2002). This high affinity would ensure that the Endo VII could readjust to allow the opposite DNA strand phosphate to enter the second active site so that the second nick occurs within the protein-DNA lifetime. The two single stranded nicks must be generated within the lifetime of the protein-DNA complex otherwise the DNA would be left with potentially damaging lesions in its DNA.

The requirement for dimer formation for DNA binding ensures that the two catalytic sites are positioned for the nicking of the opposite strands flanking the junction. Upon binding the Endo VII searches the DNA for structural distortions. For branched DNA substrates the Endo VII is thought to recognise the angle (bend) on either side of the branch point similarly to other DNA resolvases (Bhattacharyya *et al.*, 1991). Further support for kinking to act as the signal for cleavage came from the cleavage of DNA with an apyrimidinic site (Greger and Kemper, 1998). These AP-sites kink DNA and as such were cleaved more efficiently than mismatched DNA and at the same efficiency as cruciform DNA.

The recognition of mismatches is based on flexibility of the DNA duplex induced by the presence of a mismatch rather than the pronounced bending as seen for cruciform DNA. The observation that binding of Endo VII alters the conformation of the DNA structure (Pohler *et al.*, 1996) suggests that a shallow kink or bend could be enhanced upon first contact with the protein. Cleavage occurs 2-6bp on the 3' side of the distortion and is influenced by

both sequence context (Pottmeyer *et al.*, 1992) and mismatch type (Solaro *et al.*, 1993). Cleavage studies have shown that Endo VII will cleave perfectly matched DNA, however the digestion for mismatched DNA is far greater and is cleaved far more efficiently (Surdi et al., 1999).

1.8.5. Mutation Detection

The ability of Endo VII to cleave mismatched DNA has met with some success in mutation detection. Endo VII detected 81 known single base pair mutations in the mouse β-globin promoter (Youil et al., 1996). It was observed that 49 of the 81 samples relied upon cleavage of the sense strand, 17 of the 81 upon cleavage of the antisense strand and 15 showed cleavage on both strands. In all samples there were non-specific bands that were common in both wild-type and mutant, this has been attributed to sequence induced bends in the DNA acting as substrates for Endo VII (Youil et al., 1995). It was observed that cleavage could result in two or more bands in the area of the mismatch or the reduction of a background band. Therefore the detection of mutations is based upon a range of cleavage fragments and the existence of extra bands that do not exist in wild-type digestion. Endo VII digestion has been used for mutation detection in the p53 gene (Inganas et al., 2000) and the MLH1 and MSH2 genes (Otway et al., 2000) with positive results.

1.9. Microtiter Array Diagonal Gel Electrophoresis (MADGE).

The success of developing a high throughput system depends upon being able to run large number of samples whilst retaining high accuracy along with keeping things as simple as possible. Typically electrophoresis of DNA is performed in either an agarose or polyacrylamide gel matrix. Of the two matrices polyacylamide offers the best resolution and therefore is more suitable for smaller fragment analysis, however agarose gels can be easily prepared in an open-faced format allowing the convenience of horizontal electrophoresis. Polyacylamide does not polymerise in air therefore polyacylamide gels are usually run vertically between two glass plates.

Vertical polyacylamide gels are therefore limited in their application to high throughput due to the fact that each gel only has one set of rows. The remainder of the gel is inaccessible for sample loading. Conventional techniques for mutation detection like DGGE or SSCP are not suited for high throughput mutation detection due to the relatively small number of samples on each gel.

For high throughput mutation detection it would useful to combine the high-resolution properties of a polyacylamide gel with the convenience of an open faced horizontal format. A technique pioneered by our group Microtiter Array Diagonal Gel Electrophoresis (MADGE) (Day and Humphries, 1994) meets this criteria. MADGE gels have been used for PCR checking, SNP typing (restriction digest and amplification refractory mutation system), and detection of mutations (Melt-MADGE)

The basic MADGE format consists of a two dimensional plastic former and a single glass plate (Fig. 1.13a). The plastic former's dimensions are 2mm deep by 100X150mm. Placed within this are 96X2mm cubic teeth in a standard 8 X 12 array that is compatible with commercially available 96 well plates. The gel mix is poured into this format and the glass plate coated with γ-methacryloxypropyltrimethoxy-silane (sticky silane) placed on top. Upon polymerisation the sticky silane chemically cross-links with the gel thus causing the gel to adhere to the glass plate. Upon removal of the glass plate an open faced 96 well gel supported on a glass plate is achieved. The 96 well array is set at an angle of 71.6° thus enabling tracks to pass through two succeeding rows of wells with track lengths 26.5mm (Figure 1.13b). This system allows rapid loading with a multi-channel pipette or 96-pin passive replicator. A standard gel can be prepared and set within 10 minutes and on average takes between 10-30minutes to resolve depending on gel percentage and resolution required. The stretch MADGE system puts the angle at 78.7° allowing an increase in track length to 46mm at the expense of lane width. Recent developments have seen the number of wells successfully increased to 384 and 756 (Gaunt et al., 2003).

MADGE gels can either be run dry or submerged. Dry gels are run in buffer less dry electrophoresis boxes as seen in figure 1.14. Samples can be loaded straight onto these dry gels and subjected to electrophoresis where the

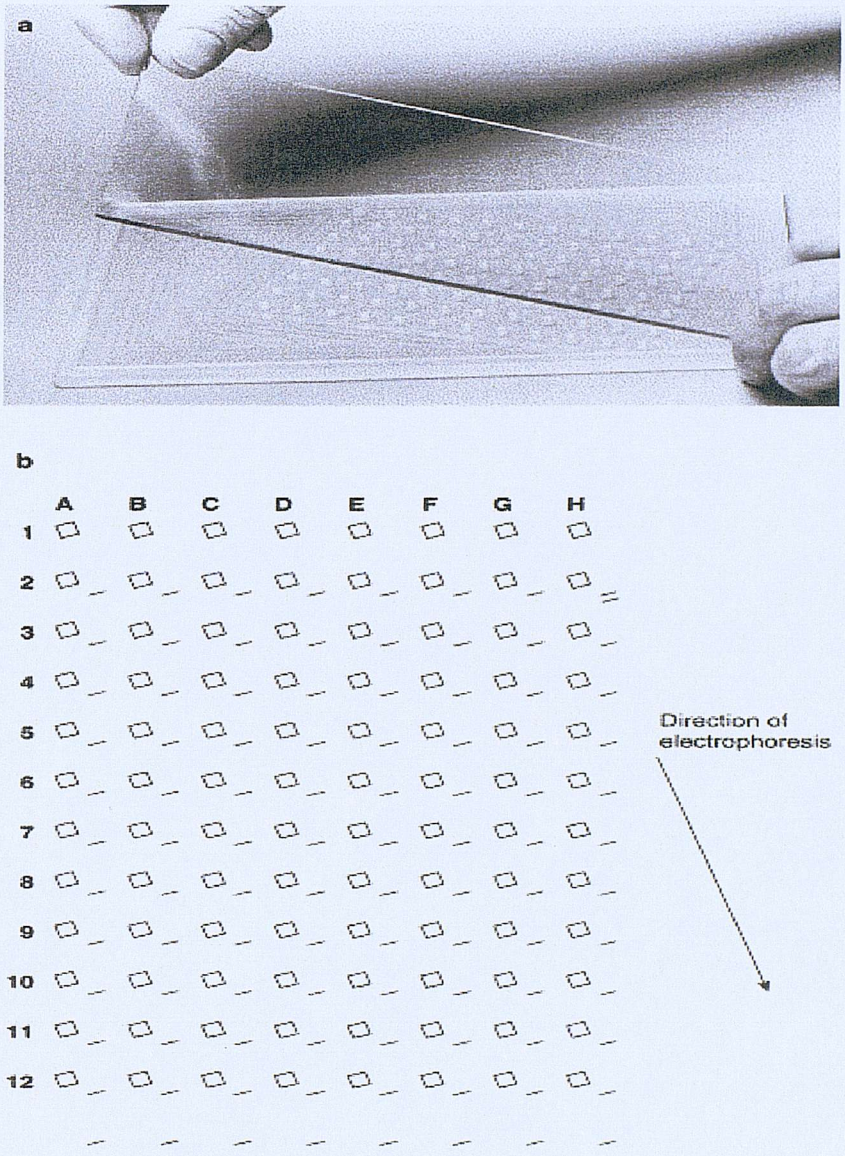
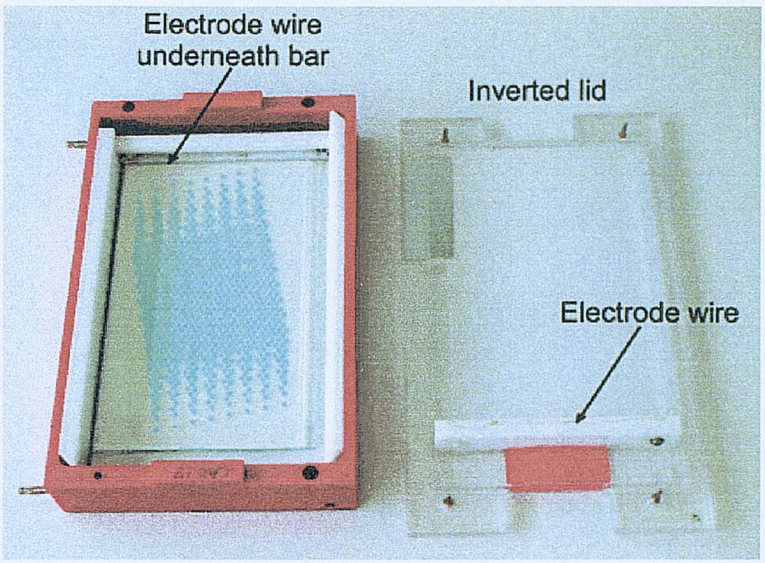


Figure 1.13. Microplate-array diagonal-gel electrophoresis (MADGE) taken from (Day et al., 1998) (A) MADGE former and silanised glass plate. The 96 well array compatible with microtiter arrays is placed on an angle of 71.6° to allow rows to pass through each other. (B) Schematic diagram of a MADGE-gel image.



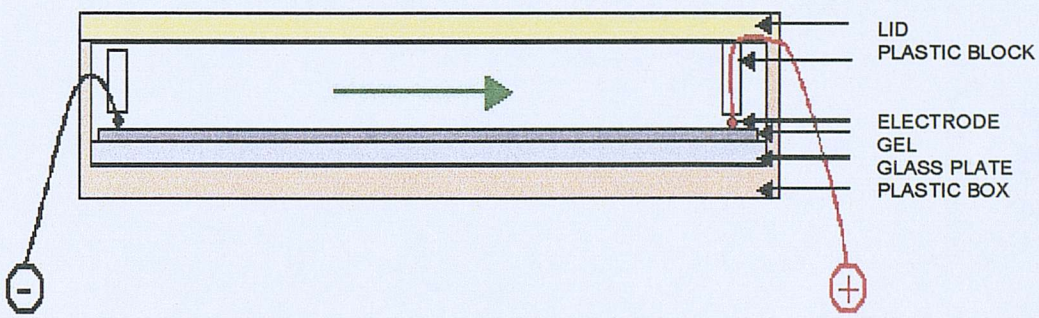
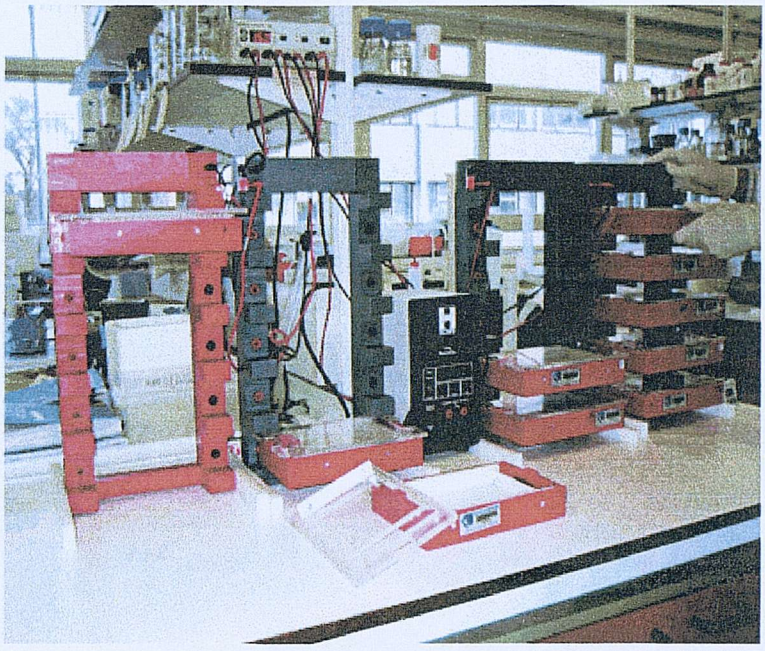


Figure 1.14. Dry electrophoresis box for MADGE gels. The box is attached to a stacking frame (Fig. 1.15).

electrodes make direct contact with the gel. Dry boxes are plugged directly into a stacking frame (Figure 1.15a), which can each hold 5 dry boxes, which can be run concurrently.

By placing a glass plate over the gel after loading of the samples it is possible to run MADGE gels submerged in specially constructed 2 litre gel tanks. These gel tanks can be connected to a programmable thermal ramp as in the case of Melt-MADGE (Day et al., 1995). Melt-MADGE is a combination of DGGE and MADGE. The denaturing variable is thermal rather than chemical and the independent variable is temporal rather than spatial. This means that during electrophoresis the temperature of the entire gel is raised rather than areas of the gel being subjected to various denaturant or temperature as in DGGE or TGGE. If two sequences differ by a single base the one with the lower melting domain at the lower temperature will display reduced mobility at an earlier stage of the run compared to the other DNA fragment. Obviously Melt-MADGE offers a significant increase in throughput when compared to conventional DGGE on vertical polyacylamide gels but at the cost of track length. The size of the gels enables direct stacking of the gels in one tank allowing the running of a number of gels in a tank. Alternatively the temperature of the buffer in the gel tank can be set to constant to allow standard denaturing gels to be run in the MADGE format. The set-up of apparatus can be seen in figure 1.15b. MADGE offers short run electrophoresis and is compatible with industrial produced Microtiter arrays. Therefore PCR products can be easily and quickly loaded. The gels themselves are convenient to handle due to the robustness afforded to them by the glass plate support and due to their small size. Stacking of a number of gels either in dry electrophoresis boxes or in gel tank enables the running of significant numbers of samples in a short time interval. Imaging of the gels can be achieved using intercalating reagents removing the need for isotope labelling of the DNA. There are available specifically designed software packages that enable convenient data analysis of large sample numbers. Theoretically MADGE offers the potential to run thousands of samples per day an attractive scenario when screening large numbers of samples. As this technique fundamentally relies upon electrophoresis, a technique that is commonplace in both large and small laboratories, it is potentially widely

(A)



(B)

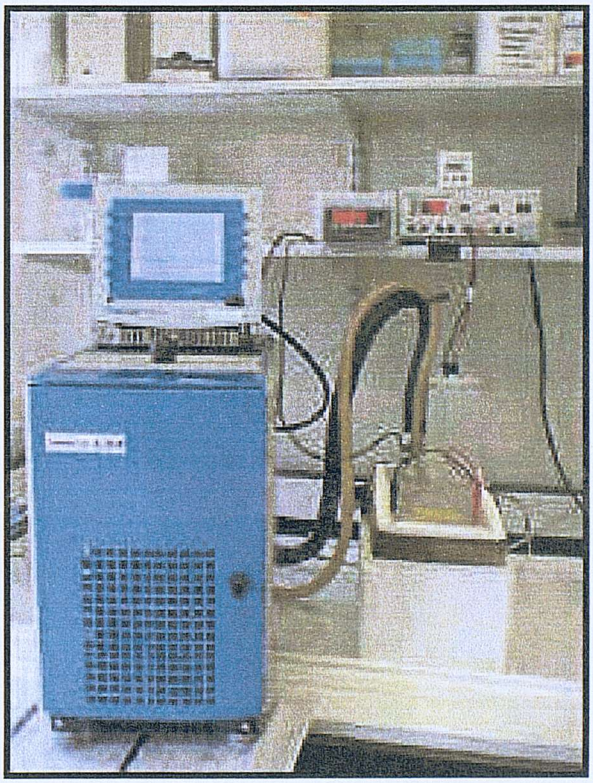


Figure 1.15. Electrophoresis Apparatus. (A) Dry electrophoresis stacking frame, (B) Melt-MADGE apparatus consisting of a programmable thermal cycler connected to 2l gel tank and standard power pack

accessible. The combination of many wells, relatively short run times when compared to other electrophoresis methods and ease of imaging make MADGE a very suitable tool to employ with mutation detection.

1.10. Project Aims

The primary aim of this project is the investigation of protein-based recognition of heteroduplex DNA for the development of a novel high throughput mutation detection assay for unknown point mutations by combining enzymatic heteroduplex recognition and the high throughput capabilities of MADGE. The two proteins chosen to investigate are a mismatch binding protein MutS and a mismatch cleavage enzyme T4 endonuclease VII

As with the development of any new assay there will be a number of issues to address before MutS or Endo VII can be used for mutation scanning. Firstly a mutation detection assay utilizing these enzymes requires a significant quantity of enzyme. Therefore the first step is the establishment of an in house source of protein followed by characterisation of the purified proteins in terms of mismatch specificity. Following this will be to determine suitable conditions where either a band shift or cleavage fragment can be identified using short track electrophoresis.

The sensitivity and specificity of a mutation detection assay using MutS or EndoVII in conjunction with MADGE will be examined using a variety of known SNPs and mutations. The successful development of such a detection tool will enable the cost effective, straightforward screening of large populations for unknown mutations.

CHAPTER 2
MATERIALS & METHODS

2.1. Materials

2.1.1. Enzymes and Chemicals

All chemicals were purchased from BDH, DIFCO, Fisher, ICN Biochemicals, Malford laboratories and Sigma. Accugel, Protogel and Sequagel polyacrylamide solutions were acquired from National Diagnostics. All oligonucleotides were purchased from MWG or Oswel. Taq polymerase, *Pfu* polymerase, *Nde*I, *Bam*HI was purchased from Promega. The *Apa*I and *Xmn*I restriction enzymes and T4 polynucleotide kinase (T4PNK) were from New England Biolabs. The shrimp alkaline phosphatase was from Sigma. The pET15b vector was from Novagen. α-^{32}P-ATP from Amersham international at a concentration of 3000 Ci/mmol.

2.1.2. Water

Two grades of water were used during this project to prepare solutions and reaction mixes. For buffers and general solutions deionised (DI) water was prepared by reverse osmosis on a Purite Prestige system. For a higher quality water (UHQ) for use in DNA preparations and PCR, water was further purified by a Purite still plus.

2.1.3. Bacterial Growth Medium

For bacterial growth two growth mediums were used. For initial bacterial growth and the propagation of recombinant *E.coli* for plasmid production a rich medium was used 2YT (16g tryptone, 10g yeast extract, 5g NaCl per litre). The second medium Laurie broth (LB) (10g tryptone, 10g yeast extract, 5g NaCl per litre) is a standard medium used widely in bacteria cultivation and was the chosen medium for induction. All medium was sterilised before use by autoclaving at 151lbs/in^2.

2.1.4. Antibiotic Preparation

Carbenicillin solution was prepared by dissolving carbenicillin in UHQ water to a produce a stock concentration of 100mg/ml. The antibiotic solution was sterilised using Schleicher & Schuell 0.2um filters. Aliquots were stored at -20°C until use.

2.1.5 Agar Plates

Bacterial colonies were grown on agar plates. 39.5g of agar was added to 1 litre of DI water and heated until the agar had fully dissolved. The agar was sterilised by autoclaving at 15 p.s.i (121°C) for 15 minutes. The agar was stored at 4°C until required. When needed the solidified agar was heated until it was liquid and if required the agar was supplemented with carbenicillin, once the temperature had fallen below 50°C. The agar was poured into Petri dishes and any air bubbles removed by flaming the surface with a Bunsen burner. The plates were stored at 4°C.

2.1.6. DNA Purification Kits.

Two kits were routinely used for the purification of PCR amplicons. These were the Qiagen QIAquick PCR purification kit and the Promega Wizard PCR Prep Kit. Both kits were used as described in the manufacturers handbook.

2.1.7. Oligonucleotides.

Oligonucleotides were used for both the formation of synthetic duplexes and as primers for PCR reactions (Table 2.1).

2.1.8. General Solutions and Buffers

A list of the solutions and their constitution can be found on pages 75-76

2.1.9. Primers

Table 2.1. PCR primers. Oligonucleotides used in (A) generation of PCR amplicons for mutation detection; (B) Sub-cloning of Endo VII; (C) BRCA1 ARMS assay. Also included are the optimum PCR temperatures and Mg^{2+} for the corresponding PCR.

(A)

Gene	Primer	Primer Sequence	Optimum PCR temp	Mg^{2+} Conc	Product Size
IGF2	URT 7 URT 10	5'-TATTAAAAACGAATTGGCTG -3' 5'-TCCTTTGGTCTTACTGGGTC-3'	53.9°C	1.5	313bp
	MS3 MS4	5'-TCTTCTCCCTTGGACTTTGA-3' 5'-AGAAGGGAGATGGCGGTA-3'	56.9°C	2	120bp
MMP-1	Mmp1-F Mmp1-R	5'- TCGTGAGAATGTCTTCCCATT-3' 5'-TCTTGGATTGATTTGAGATAAGTGAAA TC-3'	55°C	1.5	118bp
BRCA1	BRCA1-F BRCA1-R	5'-TGAAAGATGTGTAGAGAGTA –3' 5'-TTTCTATGCTTGTTTCCCGA –3'	51.3°C	2.5	260bp
	BRCA1-281F BRCA1-281R	5'- TGTCAATCCTAGCCTTCCAA –3' 5'- TTAGTCCCTTGGGGTTTTCA –3'	59°C	2	281bp
LDLR	LDLrx3-F LDLrx3-R	5'-GCCTCAGTGGGTCTTTCCTT-3' 5'-CCAGGACTCAGATAGGCTCAA-3'	59°C	1.5	222 bp

(B)

T4endoVII-F T4endoVII-R	5'-GCGCCATATGATGTTATTGAC-3' 5'-CAGCGGATCCTCATTTTAAACT-3'	50.2°C	2

(C)

2430CF	5'-GATCTGTAGAGAGTAGCAGTATTTAAC-3'	57°C	2
2430TF	5'-GATCTGTAGAGAGTAGCAGTATTTAAT-3'		
2430U/C	5'-AATCCTAGCCTTCCAAGAGAAGAA-3'		
2430D/C	5'-AACAACCATGAATTAGTCCCTTGG-3'		

2.1.10. Oligonucleotides for Synthetic duplexes

Table 2.2. Oligonucleotides for construction of 50bp synthetic duplexes (A) IGF2 ApaI synthetic Duplex (B) Control Synthetic Duplex

(A)

Identity	Sequence
IGF2G-50	5'-CCCTGAACCAGCAAAGAGAAAAGAAGG**G**CCCCAGAAATCACAGGTG GGCA-3'
IGF2A-50	5'- CCCTGAACCAGCAAAGAGAAAAGAAGG**A**CCCCAGAAATCACAGGTG GGCA-3'
IGF2C-50R	5'-TGCCCACCTGTGATTTCTGGGG**C**CCTTCTTTTCTCTTTGCTGGTTCAG GG –3'
IGF2T-50R	5'-TGCCCACCTGTGATTTCTGGGG**T**CCTTCTTTTCTCTTTGCTGGTTCAG GG –3'

(B)

Identity	Sequence
TBO683	5'-CTGACGATGTGAACTATGACGCTCC**T**AGGCTGATTACTGACTGCTGGA CT-3'
TB0684	5'- AGTCCAGCAGTCAGTAATCAGCCT**G**GGAGCGTCATAGTTCACATCGTC AG-3'
TB0709	5'- AGTCCAGCAGTCAGTAATCAGCCT**T**GGAGCGTCATAGTTCACATCGTC AG -3'
TB0710	5'- AGTCCAGCAGTCAGTAATCAGCCT**C**GGAGCGTCATAGTTCACATCGTC AG 3'
TB0711	5'-AGTCCAGCAGTCAGTAATCAGCCTAGGAGCGTCAT**A**GTTCACATCGTC AG-3'
TB0964	5'- CTGACGATGTGAACTATGACGCTCC**A**AGGCTGATTACTGACTGCTGGA CT-3'
TB0966	5'- AGTCCAGCAGTCAGTAATCAGCCT**A**GGAGCGTCATAGTTCACATCGTC AG 3'
TB0978	5'- CTGACGATGTGAACTATGACGCTCC**C**AGGCTGATTACTGACTGCTGGA CT-3'
TBO970	5'- CTGACGATGTGAACTATGACGCTCC**G**AGGCTGATTACTGACTGCTGGA CT-3'

2.1.11. DNA Duplexes.

IGF2 ApaI Synthetic Duplexes

GC Homoduplex IGF2G-50 + IGF2C-50R

5'-CCCTGAACCAGCAAAGAGAAAAGAAGG**G**CCCCAGAAATCACAGGTGGGCA
GGGACTTGGTCGTTTCTCTTTTCTTCC**C**GGGGTCTTTAGTGTCCACCCGT-5'

AT Homoduplex IGF2A-50 + IGF2T-50R

5'-CCCTGAACCAGCAAAGAGAAAAGAAGG**A**CCCCAGAAATCACAGGTGGGCA
GGGACTTGGTCGTTTCTCTTTTCTTCC**T**GGGGTCTTTAGTGTCCACCCGT-5'

GT Mismatch IGF2G-50 + IGF2T-50R

5'-CCCTGAACCAGCAAAGAGAAAAGAAGG**G**CCCCAGAAATCACAGGTGGGCA
GGGACTTGGTCGTTTCTCTTTTCTTCC**T**GGGGTCTTTAGTGTCCACCCGT

AC Mismatch IGF2A-50 + IGF2C-50R

5'-CCCTGAACCAGCAAAGAGAAAAGAAGG**A**CCCCAGAAATCACAGGTGGGCA
GGGACTTGGTCGTTTCTCTTTTCTTCC**C**GGGGTCTTTAGTGTCCACCCGT

Control Synthetic Duplexes

Watson & Crick homoduplex TB0683 + TB0711

5'-*CTGACGATGTGAACTATGACGCTCC**T**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**A**TCCGACTAATGACTGACGACCTGA

Watson & Crick homoduplex TB0978+ TB0684

5'-*CTGACGATGTGAACTATGACGCTCC**C**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**G**TCCGACTAATGACTGACGACCTGA

Single (TT) mismatch TB0683 + TB0709

5'-*CTGACGATGTGAAXTATGACGCTCC**T**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**T**TCCGACTAATGACTGACGACCTGA

Single (TG) mismatch TB0683 + TB0684

5'-*CTGACGATGTGAACTATGACGCTCC**T**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**G**TCCGACTAATGACTGACGACCTGA

Single (TC) mismatch TB0683 + TB0710

5'-*CTGACGATGTGAACTATGACGCTCC**T**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**C**TCCGACTAATGACTGACGACCTGA

Single (AA) mismatch TB0964 + TB0966

5'-*CTGACGATGTGAACTATGACGCTCC**A**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**A**TCCGACTAATGACTGACGACCTGA

Single (AC) mismatch TB0964 + TB0710

5'-*CTGACGATGTGAACTATGACGCTCC**A**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**C**TCCGACTAATGACTGACGACCTGA

Single (AG) mismatch TB0964 + TB0684

5'-*CTGACGATGTGAACTATGACGCTCC**A**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**G**TCCGACTAATGACTGACGACCTGA

Single (GG) mismatch TB0970 + TB0684

5'-*CTGACGATGTGAACTATGACGCTCC**G**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**G**TCCGACTAATGACTGACGACCTGA

Single (CC) mismatch TB0978 + TB0710

5'-*CTGACGATGTGAACTATGACGCTCC**C**AGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGG**C**TCCGACTAATGACTGACGACCTGA

2.1.12. General Solutions

Bradford Reagent

0.01% Coomassie G-250
4.7% ethanol
8.5% phosphoric acid

Buffer A

50mM NaH_2PO_4
500mM NaCl
10mM Imidazole

Buffer B

50mM NaH_2PO_4
500mM NaCl
500mM Imidazole

Diluent (1 Litre)

8mM Urea

10X Denaturing Running buffer

900mM Tris base
900mM Boric Acid
20mM EDTA
8mM Urea

Ficol Loading Dye

0.25% Bromophenol Blue
0.25% Xylene Cyanol FF
15% Ficol

MADGE Dye

48% Formamide
10mM EDTA
0.025% Xylene cyanol FF
0.025% Bromophenol Blue

MutS Dilution Buffer

50mM HEPES pH 7.2/ 20mM
Tris.HCl pH 7
1mM DTT
13.3% Glycerol

5X MutS Reaction Buffer

250mM HEPES pH 7.2 /100mM
Tris.HCl. pH 7
25mM $MgCl_2$
0.5mM DTT
0.05mM EDTA
500μg/ml BSA

SDS PAGE Running Buffer

25mM Tris Base
190mM Glycine
0.1% SDS (w/v)

2X SDS Loading Buffer

100mM Tris.HCl. pH 6.8
200mM DTT
4% SDS
0.2% Bromophenol Blue
20% Glycerol

SDS Gel Stain

40% methanol

10% Acetic Acid

0.25% cosmaine brilliant blue

SDS Destain

40% Methanol

10% Acetic Acid

Sticky Silane

0.5% glacial acetic acid

0.5% methacryloxypropyltrimethoxy silane

Stop solution

10mM NaOH

50mM EDTA

80% Formamide

0.25% Bromophenol blue

Sucrose loading dye

50% Sucrose

0.1% Bromophenol blue

10X TBE

900mM Tris Base

900mM Boric Acid

20mM EDTA

TE Buffer

10mM Tris.HCl pH 8

1mM EDTA

Transformation Buffer

10mM Tris.HCl pH8

50mM $CaCl_2$

T4 Endo VII dilution buffer

50mM Tris.HCl pH 8

1mM DTT

5X T4 Endo VII Reaction Buffer

250mM Tris.HCl pH 8

50mM $MgCl_2$

5mM DTT

0.5ug/μl BSA

Bradford Reagent

0.01% Coomassie G-250

4.7% ethanol

8.5% phosphoric acid

2.2. Methods

2.2.1. Radiolabeling and Formation of a Short 50 base Pair Duplex

2.2.1.1. 5' End Radiolabeling of Oligonucleotides

Radiolabeled DNA was generated for use in band shift and cleavage gels. The first step in the production of radiolabeled DNA is the labelling of a single strand of the DNA duplex. α-^{32}P-ATP was used to label the 5' end of an oligonucleotide. A typical radiolabeling reaction contained 1ul of oligonucleotide (100pmoles), 2µl polynucleotide kinase buffer to give a final concentration of 70mM Tris.HCl (pH 7.6), 10mM $MgCl_2$, 5mM DTT 1µl α-^{32}P-ATP, 16µl water and 1µl polynucleotide kinase (1-3units). This was left at 37°C for at least 30mins. Samples were then denatured by adding 12µl stop solution and heating to 100°C for 3 minutes before being placed on ice. The labelled oligonucleotide were run on a 10% denaturing acrylamide gel containing 7% urea for 2-3 hours. The gel was exposed to X-ray film, which was used to determine the position of the oligonucleotide so that it could be cut from the gel. The gel band was then left overnight in 300µl 20mM Tris, 0.1mM EDTA (TE) buffer so that the oligonucleotide would diffuse from the gel into the buffer.

2.2.1.2. Ethanol Precipitation

To precipitate labelled oligonucleotides or labelled duplexes 3M sodium acetate equivalent to 1/9 of the total volume was added to the DNA solution. 3 volumes of 95% ethanol were added. This was left on dry ice for at least 10 min and then spun at 21,000xg for 10 minutes. The supernatant was removed and the pellet washed with 2X 200µl of 70% ethanol with spinning at 21,000xg for 5 minutes after each wash. The pellet was dried on Speed Vac for 5 minutes and resuspended in TE or UHQ water.

2.2.1.3. Formation of Synthetic Duplexes

The labelled oligonucleotide was redissolved in TE buffer. 5M NaCl was added to give a final concentration of 50mM. An excess of unlabelled strand was added and the solution heated to 95°C then allowed to cool to room temperature. After the addition of Ficol loading dye the newly formed duplexes were run on a 10% non-denaturing gel for 1 hour at 800volts. The gel was exposed to Kodak X-ray film and the bands excised with the aid of the film. The gel bands were left in 300µl TE overnight. Overnight the DNA within the gel diffuses into the TE buffer. The DNA in solution is ethanol precipitated and the resulting pellet resuspended in TE buffer to desired counts per µl.

2.2.1.4. Non-labelled and Fluorescent Labelled Synthetic Duplex Formation

For formation of duplexes without an end label the two complementary strands were mixed in equal molar concentrations, heated to 95°C and allowed to cool gradually to room temperature. To check duplex formation, duplexes were run along side oligonucleotide on a non-denaturing H-PAGE, a difference in mobility indicated duplex formation.

For fluorescently labelled duplexes the labelled oligonucleotide was mixed with an excess of unlabelled complementary oligonucleotide and hybridised as above. If required, the synthetic duplexes were purified using ethanol precipitation or Promega Wizard PCR Prep Kit.

2.2.2 Generating PCR Amplicons for Mismatch Detection

2.2.2.1. Polymerase Chain Reaction

Since its conception in the mid 1980s the polymerase chain reaction (PCR) has become an invaluable tool in molecular biology. It uses enzymatic synthesis of specific DNA sequences directed by two oligonucleotide prlmers. The two primers hybridise to opposite strands flanking a chosen region of interest within a DNA molecule. Subjecting this to a series of temperature changes for varying lengths of time in the presence of a polymerase enzyme

leads to the amplification of a target site within a template DNA.

A typical PCR reaction contains 10ng of DNA for every 10ul of PCR mix. The PCR mix contains a final concentration of 1X PCR buffer (50mM KCl, 10mM Tris-HCl & 0.1% Triton X-100) 1-3mM $MgCl_2$, 0.2mM dNTPs, 0.4µM of each primer and 0.02u/µl polymerase. The $MgCl_2$ concentration depends on the primer set used.

The first step in the cycle is the denaturation of the DNA template to form ssDNA. This is achieved by heating the template DNA at 95°C for between 15secs-2mins. The temperature is then dropped usually to between 40°C- 60°C to allow the oligonucleotide primers to anneal with the denatured ssDNA. The annealing temperature depends on primer size and length and lasts between 30-60secs. The final step sees the raising of the temperature. to the optimum temperature of the polymerase. The time of this step depends on the size of the fragment being synthesized. These steps are collectively known as a cycle and a typical PCR consists of between 20-30 cycles. Theoretically each single cycle doubles the amount of target DNA synthesis. After 20 cycles the target DNA has been amplified in the region of a million fold in a matter of hours.

2.2.2.2. Primer Design

Primers were designed using Primer3 output software, a software programme freely available on the Internet at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi

2.2.2.3. Optimising PCR conditions

To maximise yield and reduce non-specific amplification during PCR it is necessary to determine the optimal $MgCl_2$ concentration and the optimal annealing temperature for the primers. This was achieved by running a gradient PCR. A typical optimising reaction consisted of a series of reaction mixes each containing a different Mg^{2+} concentration. Each reaction mix contains enough PCR mix for 12 PCR reactions. A standard 96 well tetrad

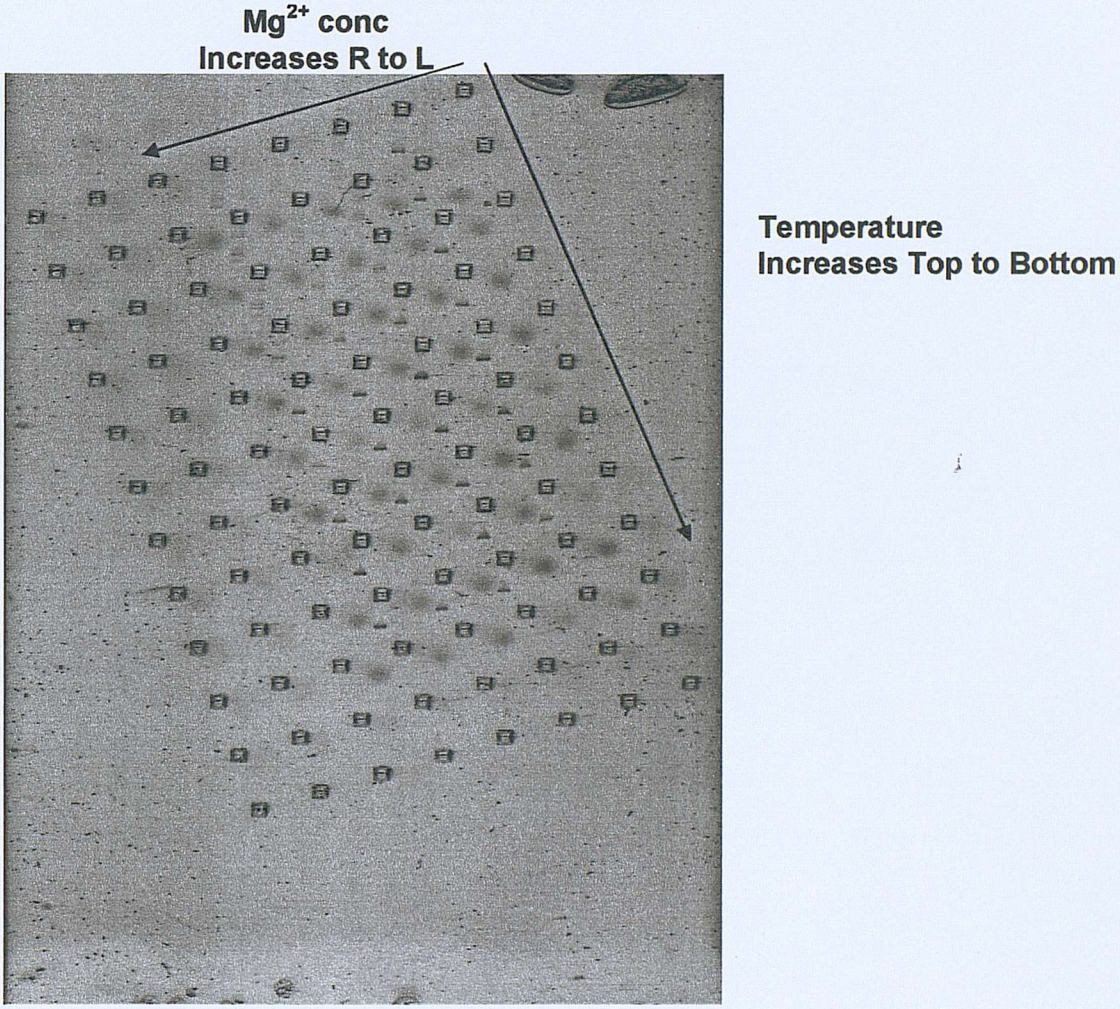


Figure 2.1. PCR optimisation. From right to left Mg^{2+} increases and from top to bottom the temperature increase. From this the optimum Mg^{2+} and temperature can be determined.

plate contains 8 rows of 12 wells. Using one row for each Mg^{2+}conc it is possible to set up to 12 identical PCR reactions in the same row. It is possible using a Thermocycler to set a temperature gradient so that each individual reaction can have a different annealing temperature starting at the lowest temperature and progressing to the highest. Running the PCR product on a MADGE gel enables the comparison of amplicon yield at different Mg^{2+} concentrations over a range of annealing temperatures.

2.2.2.4. DNA Quantification

DNA quantification was achieved by measuring the absorption spectra at 260 and 280nm on a Beckman Du-7000 spectrophotometer. The O.D value at 260nm allows the calculation of the concentration of nucleic acid assuming 1 O.D at 260nm for dsDNA is equal to 50µg/ml. The value at 280nm provides an indication of the purity of the DNA sample. The O.D at 260nm is twice that at 280nm if the solution contains pure DNA. This O.D ratio is affected by contaminants, which act to reduce the O.D ratio. Pure DNA has an OD260/OD280 ratio of between 1.8 and 2.

Typically the spectrophotometer was blanked using DI water. The DNA samples were diluted in DI water and the O.D values at 260nm and 280nm taken. The DNA concentration was then calculated from the O.D at wavelength 260nm adjusting for the dilution factor.

2.2.2.5. PCR Amplicon Quantification

To determine the concentration of product from a PCR, picogreen analysis was used. 2µl of PCR amplicon and 1/200 dilution of picogreen dye were mixed in a final volume of 100µl TE buffer. This was left in the dark for 15minutes before the fluorescence was measured on FluorImager at excitation 488nm using emission filter 530nm. Samples were compared to a standard curve ranging from 0-36ng/µl. ImageQuant (Molecular dynamics) and Microsoft excel was used to calculate dsDNA concentration In ng/ul.

2.2.2.6. Amplification of Target Sites using PCR

A number of target sites from different genes were used to generate substrates for use in mutation detection. They were chosen due to the fact that they contained a sequence change or a single base insertion, which could be detected using restriction digest, ARMS assay or melt-MADGE.

PCR amplification was carried out on an MJ research thermocycler. Genomic control samples were used as target DNA. A standard short PCR protocol was followed, this consisted of a 2-minute denaturation at 95°C followed by 30-35 cycles of 30s at 95°C, 30s at the annealing temperature of each primer pair (see section on determining annealing temperature of each primer set), 30s at 72°C with a final 10-minute extension at the end of the cycles at 72°C. The primer sets and optimum conditions as well as fragment length are displayed in Table 2.1.

2.2.3 Typing PCR Amplicons from Control DNA

2.2.3.1. I*GF2 Apa*I Restriction Digest

In order to characterise PCR product as heterozygote or homozygote 5µl of PCR product (50-100ng) was subjected to restriction digest using the appropriate restriction enzyme (RE). A 10µl *Apa*I digestion reaction contained 50-100ng PCR amplicon, 50mM potassium acetate, 20mM Tris acetate, 10mM magnesium acetate, 1mM DTT, 100 µg/ml Bovine Serum Albumin BSA and 10 units of *Apa*I and was incubated for 2hrs at 37°C. Digest fragments run on a 12.5% MADGE gel, which was pre-stained with ethidium bromide for 30minutes. Gels were imaged using a Molecular Dynamics Fluorimager-595.

2.2.3.2. *MMP4-Xmn*I Digestion

A 10 µl *Xmn*I digestion contained 5µl PCR 50mM NaCl, 10mM Tris.HCl pH 7.9, 10mM $MgCl_2$, 1mM DTT, 100 µg/ml BSA and 10 units *Xmn*I. The reaction was incubated for 8hrs at 37°C with digest fragments run on pre-stained ethidium bromide 12.5% H-PAGE gel for 30minutes.

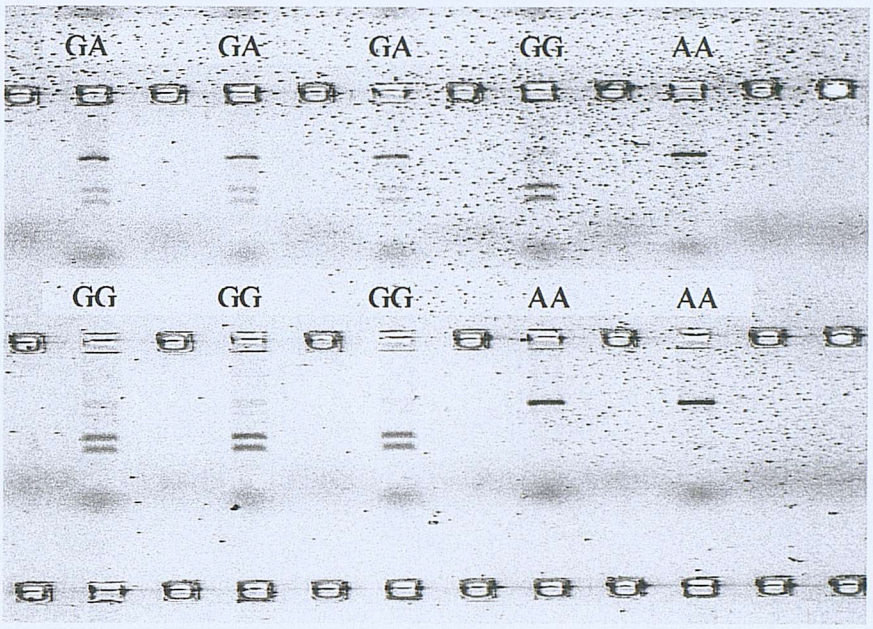
2.2.3.3. *BRCA1* ARMS Assay

The amplification refractory mutation system (ARMS) is a method used for detection of point mutations or small deletions in test DNA and is based upon a modified PCR protocol. The ARMS assay for BRCA1 consists of 2 complementary PCR reactions with which to test sample. Both contain a common forward and reverse primer that acts as a control to indicate that amplification has taken place. Each of the reactions also contains a specific ARMs primer for the target DNA for the T allele or the C allele. The technique is based upon the observation that ARMs primers with mismatched 3'termini will not function as PCR amplimers under appropriate conditions.

These specific ARMs primers will allow the amplification of test DNA if the correct target sequence is present, whereas if the target sequence is not present then amplification will not take place. Therefore if a CC homozygote is tested with both T and C specific primers we will only see a PCR band in the C allele reaction and not the T allele reaction. The opposite happens for a TT homozygote, whereas for a heterozygote we see a PCR band in both T and C primer reactions.

Two primer mixes were prepared. They contained 2μl 2430 U/C, 10 μl 2430 D/C and 10 μl 2430CF or 2430TF. A 10μl reaction mix contained 2μl of template DNA, 0.4μl $MgCl_2$ (2mM), 0.2μl dNTPs, 0.2μl of primer mix, 2.6μl betaine (3.8mM), 0.1μl Taq and 3.5μl UHQ water. Each DNA template was subjected to two PCR reactions. A standard short PCR protocol was followed, this consisted of a 2-minute denaturation at 95°C followed by 20 cycles of 30s at 95°C, 30s 57°C 30s at 72°C with a final 10-minute extension at the end of the cycles at 72°C.

(A)



(B)

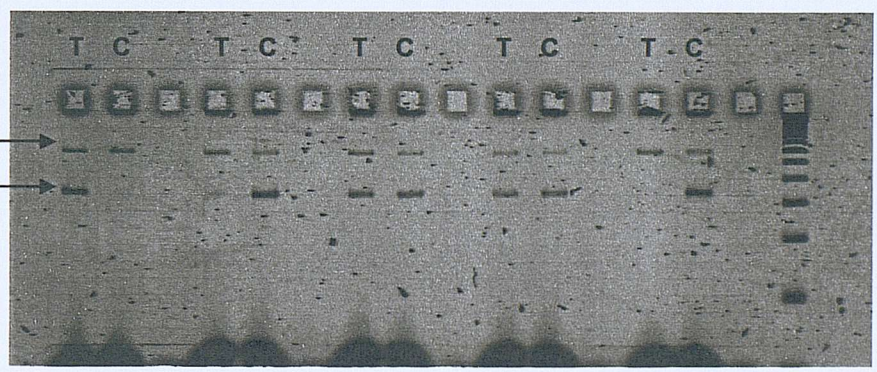


Figure 2.2. Characterising PCR samples. (A) A typical digestion of the PCR amplicon using the *Apa*I restriction enzyme. A GA heterozygote produced 3 bands as 50% of the PCR product contained correct recognition site whilst the remaining 50% had lost the recognition site. GG homozygotes containing correct sequence all underwent digestion-giving rise to two bands. AA homozygotes containing incorrect sequence under went no digestion giving rise to a single band.
(B) Example of the BRCA1 ARMS assay. Each sample DNA subjected to two PCR reactions each containing an allele specific primer. The top band represents the control band indicating a successful PCR reaction and the bottom band represents the specific PCR product. The presence of the lower band indicates the identity of the allele.

2.2.4 Expression and Purification of His-tagged Proteins

2.2.4.1. pET15b Vector

The vector chosen for the expression of both the MutS and T4 endonuclease protein was the pET15b vector from Novagen. In common with the other vectors of the pET series pET15b uses the bacteriophage T7 promoter to direct expression of target genes. This enables stringent control of target gene expression as *E. coli* RNA polymerase does not recognise the T7 promoter therefore in the absence of any T7 polymerase there is virtually no gene expression. This is important when trying to express toxic proteins such as T4 endonuclease VII as leaky expression can cause cell death. The pET vectors are used in conjuncture with λDE3 lysogen strains. The λDE3 strains bear the T7 RNA polymerase gene within their genome, which is expressed upon the addition of isopropyl-β-D-thiogalactopyranoside (IPTG). Both the λDE3 and pET15b vector contain LacI gene that encodes a lac repressor. In the bacteria this binds to a lac operator site in lacUV5 the promoter responsible for the expression of the T7 RNA polymerase. The binding of the lac repressor to this site prevents *E. coli* RNA polymerase from binding thus preventing expression of the T7 RNA polymerase. On the vector the lac repressor binds to a 25bp lac operator sequence immediately downstream of the T7 promoter preventing T7 RNA polymerase binding and transcribing the target gene. Thus control of gene expression is controlled on two sites. The addition of IPTG facilitates the expression of the target protein by binding to the lac repressor. This prevents it from repressing the expression of the T7 RNA polymerase in the bacterial genome. The removal of the lac repressor from the lac operator on pET15b enables the T7 RNA polymerase to transcribe the target gene.

The high level of control provided by the combination of pET vector and the λDE3 lysogen are not the only advantage to choosing pET15b for protein expression. Within the cloning region of pET15b is a string of 6Xhisidine residues. This string of His residues, depending upon where the clone is inserted within the cloning region, adds a His-tag to either end of the target

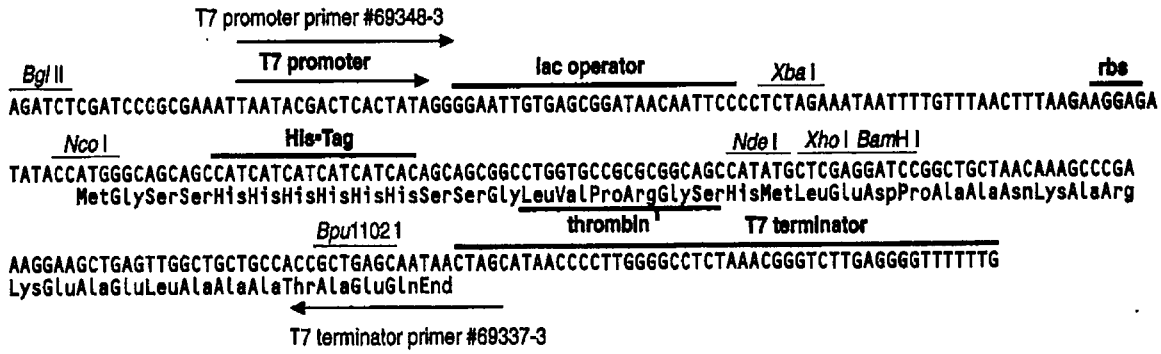


Figure 2.3. The cloning region of pET15b.

protein. The attachment of the His-tag enables the protein to be purified using metal chelation chromatography. The advantages of the His-tag are that it enables single step purification and therefore reduces manipulation of the target protein. In most cases the His-tag does not interfere with protein function but can be easily cleaved using Thrombin. As the majority of bacterial proteins do not contain sequential histidine residues very high purity of target proteins can be achieved.

2.2.4.2. Preparation of Competent Cells

A single colony from a freshly plated agar plate of BL21 DE3 was used to inoculate a 5ml 2YT media. This was grown overnight at 37°C. A fresh 50ml LB culture was inoculated with 0.5-1mls of the overnight culture. Cells were grown to an optical density (O.D) of 0.6 and harvested by centrifuging at 5000xg for 10 minutes. The cell pellet was resuspended in 20ml ice-cold transformation buffer and kept on ice for 30 minutes. Spinning at 5000xg for 10 minutes formed the cell pellet and this was suspended in 4ml transformation buffer.

2.2.4.3. Measuring the O.D

Under sterile working conditions 1ml of the bacterial culture was placed in a 1ml plastic curvete. The O.D was measured at a wavelength of 600nm on a spectrometer that had previously been blanked with water.

2.2.4.4. Transfection

1µl of plasmid was incubated with 200µl of competent cells on ice for 30 minutes. The solution was heat shocked by placing at 42°C for 60 seconds and then returned to the ice. 100µl of the transfected competent cells plated onto a LB agar plate containing 100ng/µl carbenicillin. The plate was then placed at 37°C overnight to allow colony formation and then stored at 4°C.

2.2.4.5. Over Expression of His-tagged Protein.

Cells from a glycerol stock or from freshly transfected cells were streaked on an agar plate containing 100ng/µl carbenicillin and grown overnight at 37°C. For MutS expression a single colony was used to inoculate a 5ml LB (10g tryptone, 10g yeast extract, 5g NaCl) culture containing 100ng/µl carbenicillin and grown overnight with continuous shaking at 300rpm at 37°C. It was observed that for T4 endonuclease VII a single colony grown at 37°C in a 5ml LB culture overnight lead to cell death. Therefore for T4 endonuclease VII a single colony was used to inoculate a 1litre LB culture containing 100ng/µl carbenicillin and this was grown overnight at 30°C. Following overnight growth the cultures were used to inoculate fresh 500 ml LB media containing 100ng/µl carbenicillin. The cultures were grown to mid-log phase of growth (O.D 0.6-0.8) and protein expression induced with the addition of IPTG to a concentration of 1mM. After a further 3 hours growth in the case of MutS and 2 hours growth in the case of T4 endonuclease VII the cells were harvested by centrifuging at 5000xg for 10 minutes. The

supernatant was discarded and the cell pellet stored at -70°C prior to His-tag purification.

2.2.4.6. Preparation of Cell Pellet for Purification

Once thawed cell pellets were resuspended in buffer A. Cells were lysed by sonication; the cell debris and intact cells were removed by centrifuging at 100,000xg for 40 minutes. All steps were carried out at 4°C. The cell lysate was filtered by passing through a Schleicher & Schuell 0.2µm single use filter.

2.2.4.7. Purification of His-tagged Protein

The His-tagged proteins were purified via affinity chromatography. This technique separates the His-tagged protein from the other cellular proteins on the basis of a reversible interaction between the His-tagged protein and immobilised Ni^{2+} attached to a chromatographic matrix. The His-tagged protein is applied under conditions that favour binding to the Ni^{2+} and unbound material is washed away. Recovery of the bound protein is achieved by introducing a competitive ligand (imidazole) that competes for binding to the Ni^{2+} causing the his-tagged protein to dissociate. The dissociated protein is collected in fractions in a pure and concentrated form.
Throughout this project both Ni-NTA Superflow resin (Qiagen) and Hitrap™ Chelating HP 1ml columns (Amersham) were used in the trapping and purification of the His-tagged fusion protein. Both medias can be used for small-scale purification or can be used in conjunction with a chromatography system to enable larger scale purifications.

Both methods involved loading the cell lysate to the column and washing with 20mM NaH_2PO_4 pH 7.5, 500mM NaCl and 10mM imidazole to reduce non-specific binding. The fusion protein was eiuted using a gradient of imidazole from 0 to 500mM. Analysis of fractions and purity of eluted proteins were determined using SDS gels.

2.2.4.7.1. Small Scale Prep

Ni-NTA resin was mixed thoroughly before loading 0.2ml in a 5 ml syringe. Placed in the tip of the syringe was a bung of glass wool to prevent resin from leaking out of the syringe. The resin was allowed to settle for 10 minutes. The resin was washed with 5ml of buffer A before the supernatant was applied to the resin. The flow through was collected and discarded. The column was washed with 5mls of buffer A. The protein was eluted by washing the column with 1ml buffer B and the fractions were collected in 1.5ml Eppendorfs. This method was used for small preps used in initial expression studies and to determine that His-tag was functioning as expected.

2.2.4.7.2. Large Scale Prep

1ml HiTrap chelating columns were used in conjunction with the AkTaprime FPLC chromatography system (Amersham). This system allowed the gradient elution of His-tagged proteins. The column, machine and buffers were prepared as outlined in the manufacturer's handbook. The AkTaprime has a His-tag protein purification protocol pre programmed and this was used to purify the His-tagged proteins. This program charges the column with Ni ions and washes the column with buffer A. The sample is applied to the column where an increasing concentration of imidazole washes through the column to elute the protein. The gradient is achieved by mixing increasing volumes of buffer B with decreasing volumes of buffer A. The machine collects 40 1ml fractions corresponding to the start of the gradient and the end of the gradient. The His-tagged protein purification can be followed in real time using primeview software so that fractions corresponding to the peak of interest can be collected.

2.2.4.8. His-tagged Cleavage

The removal of the His-tag was achieved using the endoprotease thrombin. 0.04u/µl of thrombin was sufficient to cleave 10µg of His-tagged protein over a 2hr time period. Thrombin and His-tagged protein were

incubated for 2hrs in 20mM Tris.HCl, pH 8.4, 150mM NaCl and 2.5mM $CaCl_2$ at room temperature. The thrombin was removed from the reaction using Streptavidin agarose. For each unit of thrombin present 32µl of the 50% streptavidin agarose was used to capture the thrombin. After 30 minutes the target protein was removed by spin-filtration. Conformation of cleavage and recovery of target protein was determined by SDS-PAGE analysis.

2.2.4.9. Dialysis

To remove the elution buffer and replace it with a suitable storage buffer the elute was subjected to dialysis. Dialysis tubing was cut to produce a 10cm-20cm strip. This was boiled for 10minutes in sterile water containing 1mM EDTA (pH 8). After allowing too cool, one end of the tubing is clamped using a plastic clip. The protein elutes were carefully syringed into the dialysis tubing. Any air was expelled from the tubing and the open end clamped. The dialysis tubing was placed in a litre of the required buffer and left at 4ºC with gentle stirring until the buffers had exchanged.

2.2.4.10. Protein Quantification

To determine the concentration of purified protein the Bradford method was employed. This involves the addition of coomassie blue dye that binds to the basic and aromatic amino acids found in proteins resulting in a shift of the absorbance maximum from 465 nm to 595nm, a resulting colour change of brown to blue. OD_{595} values for samples are compared to a standard curve generated from a set of known protein concentrations.

A standard curve was created by diluting known stock concentration of BSA. The stock concentration of the BSA was 1mg/ml. This was diluted in DI water to give a range of protein concentrations from 0-100 µg/µl. The BSA dilutions were mixed with 1ml of Bradford reagents and left for 2 minutes. The absorbance at 595nm was measured using a spectrometer and the values collected used to plot a standard of Abs against ug/ul of protein. Purified protein samples were then measured and compared to the standard to

determine the protein concentration.

2.2.4.11. Plasmid Minipreps

For transfection and analytical studies it is necessary to purify the plasmid so that it is free of any contaminating bacterial DNA. The purification of plasmid DNA was achieved using the QIAprep miniprep kit. A single colony was selected from an agar plate and grown overnight in 5ml LB media containing 100µg/ml carbenicillin. A 1ml sample was taken from the overnight culture and spun at 5000xg for 10 minutes in a bench top centrifuge. The cell pellet was resuspended in 250µl P1 buffer. To this was added 250µl of buffer P2 and the solution gently mixed. Cell lysis was stopped by the addition of 350µl N3 buffer. The solution was then spun for 10 minutes at 17,900xg. The supernatant was decanted into a QIAprep spin column and spun for 30-60 seconds at 17,900xg. The flow-through was discarded and 500µl of buffer PB applied to the column and spun as previous. After discarding the flow-through the column was subjected to a final wash using 750µl of buffer PE and centrifuged as before for 1 minute. To remove any residual wash buffer after the discarding of the flow-through the column was spun at 17,900xg for 1 minute.

To elute the DNA the column was placed in a clean 1.5ml Eppendorf and 50µl of buffer EB (10mM Tris.HCl, pH 8.5) applied to the column. This stood for 1 minute before centrifuging for 1 minute. The elute in the Eppendorf contained the purified plasmid. Running on 0.6% agarose gel containing ethidium bromide checked plasmid characterisation and yield.

2.2.5 Sub-cloning of T4 endonuclease VII.

2.2.5.1. Clone Production

The plasmid pRB210 (T4 endonuclease VII in pET11a) was a kind gift from Dr S. Golz. As pET11a does not contain a His-Tag it was decided to clone the T4 endonuclease gene into pET15b to simplify the purification of T4 endonuclease VII. A set of primers was designed containing the *Nde*I and the

*Bam*HI restriction sites called T4VIIFOR and T4VIIREV. These restriction sites were included because within the pET15b cloning region is a single NdeI and BamHI site in close proximity to one another and this would provide a good site for insertion of the T4 endonuclease VII gene.

The T4 endonuclease VII gene was amplified from pBR210 via PCR using the above primers. The PCR reaction mix contained 10ng pBR210, 20mM Tris.HCl, 10mM KCl, 10mM $(NH_4)_2SO_4$, 20mM $MgSO_4$, 0.1% Triton X-100, 100µg/ml BSA, 0.2mM dNTPs, 0.4um T4VIIFOR, 0.4um T4VIIREV and 2-3 units of Pfu polymerase enzyme. The PCR mix was heated to 95°C for 2 minutes, followed by 30 cycles of 30s at 95°C, 30s at 50.2°C and 30s at 72°C with a final 10-min extension at 72°C at the end of the cycles.

2.2.5.2. Production of Complementary Overhangs.

To prepare PCR and pET15b for ligation both were treated with *Nde*I and *Bam*HI restriction enzymes to produce complementary overhangs on both the PCR product and within the cloning region of the pet15b vector.

PCR	pET15b
10µl PCR product	1µl pET15b 10ng
10µl Buffer D	10µl Buffer D final
(6mM Tris.HCL, 6mMgCl, 50mM NaCl, 1mM DTT)	
5µl 10mg/ml BSA	5µl 10mg/ml BSA
1µl *Nde*I 10u/µl	1µl *Nde*I
1µl *Bam*HI 10u/µl	1µl *Bam*HI
23µl Nuclease free water	32µl Nuclease free water

The digestion mix was incubated for 2hrs at 37°C and was followed by 10mins at 65°C to denature the RE. The cut PCR product and pET15b purified using QIAQuick PCR purification kit.

2.2.5.3. Dephosphorylation of 5' overhangs of pET15b

To prevent religation of vector with itself the 5' ends were

dephosphorylated using Shrimp Alkaline Phosphatase (SAP). 1 unit per µg DNA of SAP is incubated with restriction–digested pET15b in 1XSAP buffer (50mM Tris.HCL, 10mM $MgCl_2$). The reaction was left for 15mins at 37°C and then placed at 65°C for 15 mins to inactivate the SAP.

2.2.5.4. Ligation of T4 endonuclease VII Gene with pET15b

Ligation reaction consisted of 1µl T4 DNA ligase (20u/µl), 1X T4 DNA ligase buffer (30mM Tris.HCL pH 7.8, 10mM $MgCl_2$ 1mM ATP, 5% PEG and 10mM DTT) 1µl PCR product, 3µl pET15b made up to final reaction volume of 10µl with UHQ water. After 3hrs incubation at room temperature 200µl of competent TG2 or BL21 (DE3) Gold cells were added and a transfection reaction performed. TG2 lack expression apparatus and thus act as a good host strain. 100µl of transfected TG2 used to inoculate agar plate containing 100ng/µl carbenicillin.

2.2.6. Protein and DNA Gel Systems

2.2.6.1. Agarose Gels

Agarose gels were used to determine the presence and size of plasmids and PCR products. Typically 0.3g of agarose was dissolved in 40ml of boiling 1XTBE to give a 0.6% gel. Staining was achieved by adding ethidium bromide to a 1/10000 dilution of the 500µg/ml stock, once the solution had cooled to below 65°C. 5µl of sample and 1µl of gel loading buffer (0.09% bromophenol blue, 0.09% xylene cyanol FF, 60% glycerol and 60mM EDTA) were loaded onto the gels and the gels were run at 100volts in 1X TBE buffer. The length of time the gels were run depended upon the size of the DNA and the resolution required. To aid identification of bands a Gene Ruler 1kb DNA ladder was run along side the samples.

2.2.6.2. Polyacrylamide Gel Electrophoresis

Both non-denaturing and denaturing PAGE were used extensively

throughout this project. Non-denaturing PAGE are typically poured and run in 1XTBE. They are run at a low voltage to prevent denaturation of the DNA by heat generated by the electric current. Denaturing PAGE have addition agents usually urea that suppresses nucleic acid base pairing. By denaturing the DNA prior to loading on the gel the urea effectively prevents the DNA reannealing as it passes through the gel.

2.2.6.3 Preparation of Non-denaturing PAGE

Two 20cm X 40 cm glass plates were cleaned with ethanol and placed together with 0.35mm thick spacers placed down the left and right edges to separate the plates. Yellow electrical tape was placed around the sides and bottom to hold the plates together and to prevent leakage of the gel mix before it had set. A 50 ml gel mix was prepared as follows with different gel percentages achieved by varying the volume of acrylamide in the 50ml gel mix.

Non-denaturing	
30% Acrylamide (protogel)	15ml
5X TBE	10ml
DI water	25ml
APS	200µl
TEMED	40µl

The gel mix was poured between the plates and any air bubbles removed by gently tapping the plates on the bench surface. A comb was placed in the top of the gel with the tops of the teeth slightly higher than the top of the gel. Bulldog paper clips were placed down the side of the gel to act as clamps and the gel was left at an angle of 10° to set for 60 minutes. Once set, the tape was removed from the bottom of the gel but left on the sides to direct electrophoresis down the gel. The gel was attached to the

electrophoresis tank with bulldog clips and the comb was removed and the wells rinsed with 1XTBE. The buffer reservoirs of the tank are filled with 1X TBE and the gel is pre run before DNA samples are loaded.

2.2.6.4. Denaturing PAGE

Denaturing PAGE was prepared as above with a few adaptations to the technique. Firstly the gel mix contained urea. Secondly due to the heat generated by the high voltage used in denaturing gels, a metal plate was attached to the front of the glass plates to aid in the dissipation of heat and prevent smearing of the DNA. Prior to use the gels were pre-run for an hour in 1XTBE buffer

Denaturing	
30% Acrylamide (Sequagel)	15ml
5X Denaturing RB	5ml
Diluent	30ml
APS	200µl
TEMED	40µl

2.2.6.5. MADGE Gel and H-PAGE

The gels were prepared and poured as described in Day & Humphries, 1994. MADGE gels and H-PAGE formers differ only in the arrangement of the wells. MADGE gels have the wells arranged in 12 rows of 8 set at an angle of 71.6° compatible with Microtiter arrays. The wells in a H-PAGE are more conventional with 6 rows containing 16 wells. By setting the array at 71.6° a longer track run is available using MADGE gels when compared to H-PAGE.

The MADGE/H-PAGE gel system uses a 2D former and a glass plate coated with sticky silane. The Gel mix contains TBE buffer, acrylamide from a

liquid source, TEMED and ammonium persulphate. Varying the volume of acrylamide in 50ml gel mix alters the percentage of the gel.

Acrylamide (30%)	Volume depending on required %
10X TBE	5ml
25% APS	150µl
TEMED	150µl

Made up to 50ml with DI water

The gel mix was poured into H-PAGE or MADGE formers and a silanised glass plate acts as the support for the gel. The gels take 5 minute to set and once set can be removed from the former using a spatula and remains attached to the glass plate. Gels can be prestained with 1/10000 dilution of ethidium bromide or post stained with 1/10000 vistra green or ethidium bromide depending upon experiment. Dry gels were run in dry gel electrophoresis box at 150V, the length of the run depending on DNA size and migration required. Submerged gels run in a horizontal electrophoresis tank in 1X TBE buffer at 150V.

2.2.6.6. Denaturing MADGE/ H-PAGE

A slight adaptation is needed in order to run MADGE/H-PAGE under denaturing conditions. The gel mix contains urea

Acrylamide (30%)	Volume depending on required %
10X Denaturing RB	5ml
25% APS	150µl
TEMED	150µl

Made up to 50ml with diluent.

The second adaptation arises once the samples have been loaded onto the gel. As the gels are run submerged a glass plate is slid over the top of the gel to prevent samples diffusing out of the wells once placed in the gel tank. To prevent air bubbles DI is pipetted in front of the glass plate. The

plates are held together with elastic bands and rubber tubing is placed down the side of the plates to direct the electrophoretic flow. The gels are run in a specially constructed gel tank that consists of a water heater connected to a thermoregulator and a gel tank (Figure 1.14b)

2.2.6.7. SDS PAGE

Sodium dodecyl sulphate is an anionic detergent that wraps around the hydrophobic tail of the polypeptide backbone. SDS binds to protein in a ratio of 1.4g: 1g and confers a negative charge to the protein in proportion to the length of the protein. The disruption to hydrogen bonds and hydrophobic interactions also minimalise differences in secondary and tertiary structures. Therefore in combination with a reducing agent like dithiothreitol (DTT) the proteins have a uniform negative charge per unit length. When run on a denaturing gel protein treated with SDS and DTT separate dependant on molecular weight and not due to the electrical charge of the polypeptides. SDS-PAGE is a very useful tool for protein expression work. It can be used to determine expression of a protein of interest when run along side a molecular weight marker. The intensity of the protein band of interest after staining can give some indication to expression levels and after purification to the purity of eluted protein.

The SDS gels used in this project consisted of two layers. The first layer the stacking gel concentrates the proteins into a compacted zone. The second layer the resolving layer causes the proteins to unstack and separate according to their molecular weight.

Throughout this project all gels were poured using protogel (acrylamide: bisacrylamide 37.5:1) and run in a mighty Small II gel tank from Hoefer filled with SDS running buffer. The stacking gel was a 5% gel and the percentage of the resolving gel ranged from 10-15% depending on the molecular weight of the protein. All samples loaded onto the SDS gels were mixed with SDS loading buffer. The volume of SDS loading buffer added depended upon the sample type and the nature of the analysis. To fully denature the samples before loading they were boiled for 5 minutes. Typically 10-15µl of sample was loaded onto the stacking gel and run at 80Vs. Once

the dye front reached the resolving gel the voltage was increased to 150V and the gel run until the dye front reached the bottom of the gel. The gel mix can be seen below.

10% Resolving Gel		5% Stacking Gel	
Acrylamide (30%)	2ml	Acrylamide (30%)	330µl
1.5M Tris pH 8.8	2.25ml	1M Tris pH 6.8	250µl
DI Water	1.65ml	DI Water	1.395µl
10% SDS	60µl	10% SDS	20µl
10% APS	60µl	10% APS	20µl
TEMED	5µl	TEMED	5µl

2.2.7. MutS Band Shift Gels

2.2.7.1. Band Shift Assays using ^{32}P Labelled Duplex DNA

Band shift assays were used to assess the mismatch specificity MutS. Typically the 10ul reaction mix contained 2µl labelled DNA and varying concentrations of MutS in 20mM Tris.HCl pH 7.5, 5mM $MgCl_2$, 0.1mM DTT and 0.01mM EDTA. Reactions were left on ice for 30 minutes. Samples were loaded onto a 6% non-denaturing PAGE gel after addition of 5µl of Ficol loading dye. Gels were run at 200volts for between 1-2hrs. The gels were fixed using a 10% acetic acid solution. Gels were transferred to 3mm Whatman paper and dried under a vacuum at 80°C for 1hour. A Kodak Phosphor Screen was used to expose the gel and a Molecular Dynamics Storm 860 phosphorimager to scan the gel.

2.2.7.2. Band shift Assays using Short Run Electrophoresis (H-PAGE& MADGE)

Standard incubation mix contained 1-2µl PCR product (depending on PCR conc), MutS protein 20mM Tris.HCl pH 7.5, 5mM $MgCl_2$, 0.1mM DTT and 0.01mM EDTA. The reaction mix was incubated on ice for 30minutes. After addition of 50% sucrose loading dye the samples were loaded onto the gel and the gels were run in a dry gel electrophoresis box at 150V, the length

of the run depending on DNA size and migration required. Submerged gels were in a horizontal electrophoresis tank in 1X TBE buffer at 150V.

2.2.7.3. MutS ATP Titration

2µl of 120bp *IGF2 Apa*1 PCR fragment (final conc in incubation mix 0.05pmole\µl) was incubated for 20minutes on ice with 6ul of MutS (final conc in incubation mix 1pmole/µl, ratio of MutS to DNA fragment was 20:1). After 10minutes incubation 2µl of ATP in MutS buffer was added to the incubation mix. This was left for a further 10 minutes before 3µl of the sucrose-loading buffer was added. After mixing, 5µl was loaded onto a H-PAGE and run under dry conditions at 150V for 15mins. The gel was post stained in vistra green and analysis carried out using Phoretix.

2.2.8 T4 endonuclease VII Cleavage Gels

2.2.8.1 T4 endonuclease Cleavage of ^{32}P labelled duplex DNA.

Cleavage assays were used to assess the mismatch cleavage properties of Endo VII. Samples contained 2µl of labelled DNA (<100pmole) and varying amounts of Endo VII depending on the cleavage reaction. A typical reaction volume was 10µl and this contained 2µl labelled DNA, 1µl Endo VII (diluted to desired concentration in 20mM Tris pH8, and 1mM DTT) in 20mM Tris-HCl pH 8, 10mM $MgCl_2$, 1mM DTT and 0.5µg/µl BSA. Cleavage was allowed to proceed for 5mins at 37°C before the addition of a stop solution (95% formamide, 20mM ethylene diamine tetracetic acid (EDTA), 0.05% bromophenol blue and 0.05% xylene cyanol). Samples were then heated for 3mins at 100°C before 5-7µl of sample was loaded onto a 15% denaturing gel (sequagel). The gel was run until the dye front reached the bottom of the gel. The gel was dried and imaged as previously described.

2.2.8.2. GA Marker

GA markers were used in Endo VII cleavage gels to identify where

DNA cleavage was occurring. Boiling of sample DNA with a formamide mixture resulted in the depurination of the DNA. When run on a denaturing gel the bands correspond to the G's and A's of the DNA sequence and thus provide a way of localising where DNA cleavage has occurred.

2.2.8.2. PCR Probe Formation for Endo VII digests

Amplifying a region of interest with the wild type sequence using fluorescently end labelled primers generated a probe. The identity of the fluorescent tag and which strand was labelled varied upon experiment. Probes were purified using the Promega Wizard PCR Prep Kit. Fluorescently end labelled wildtype probe was mixed with a 5X excess of unlabelled test DNA. Probe and test were denatured for 5mins at 95°C and hybridised by cooling to 65°C for 30mins and room temperature for a further 30mins on a tetrad.

2.2.8.3. Cleavage Assays using Short track Electrophoresis

A typical 10µl reaction mix contained 2-8µl DNA sample (depending on sample type and concentration), 1µl Endo VII reaction buffer and 1µl Endo VII protein solution. This was incubated at 37°C for 20 minutes before addition of stop solution. The samples were then denatured by placing at 95°C for 5 minutes and kept on ice until gel loading. 5-6µl was loaded onto a denaturing H-PAGE or MADGE gel. These were sealed by placing a glass plate over the top of the gel and rubber tubing placed along each edge of the gel. Gels were run at 150 volts in a submerged gel tank. Run times depended upon percentage of gel and size of amplicons. Gels were imaged on Molecular Dynamics FluorImager-595.

2.2.9. Quantitative Data Analysis of Band Shift Gels and Cleavage Assays

Phoretix software was used to quantify the fraction of DNA retarded in a MutS band shift assay or the fraction of DNA present as a specific cleavage product for Endo VII cleavage assays. This program allows the calculation of

band intensities and can represent this as a % of the total intensity within a lane or from selected bands. The quantifying of the bands to give a numerical value to represent the amount of DNA shifted or cleaved enabled the representation of the data collected in graph format using analysis software such as Excel (Microsoft).

2.2.10. ABI 310 Genetic Analyser

The ABI 310 genetic analyser from Applied Biosystems is a single capillary genetic analyser, which is suitable to carry out a range of sequencing and fragment analysis applications. The ABI 310 offered an alternative to gel based detection and was used for fragment analysis of Endo VII digestion.

The system is based around a single glass capillary 50um X 47cm filled with Performance Optimised Polymer 4% (POP-4) and fluorescent labelled DNA. The sample and capillary are brought into contact via the instruments auto sampler where the sample enters the capillary via electrokinetic injection. The samples are electrophoretically separated as they pass through the polymer filled capillary. Located along the capillary is a detector window that lines up with a laser located within the instrument. As the DNA passes through this window the fluorescent dye is excited. A CCD camera collects the emitted fluorescence. With the inclusion of an internal size standard the information collected can be interpreted by Genescan software, which calculates the size and quantity of the DNA fragments.

2.2.10.1 Sample Preparation

The instrument was set up as outlined by the manufacturer. A typical reaction mix contained 0.5µl Genetix TAMRA labelled genetic marker, 12µl deionised formamide and 1-5µl of reaction mix depending on initial DNA concentration. The reaction mix was denatured at 100°C for 5 minutes and then placed on ice until ready to place in the auto sampler tray in the instrument.

CHAPTER 3
STUDIES OF MUTS BINDING PROTEIN AND ITS APPLICATION TO HIGH THROUGHPUT MUTATION SCANNING USING MADGE

3.1. Introduction

Presented in this chapter is the work carried out to investigate the suitability of MutS for use in mutation scanning using MADGE. To begin with the specificity of MutS for heteroduplex DNA over homoduplex DNA is investigated. This is then followed with the application of MutS band shifts to the MADGE format and its potential use for detection of mutations in relation to a high throughput technique.

3.1.1. Experimental Strategy

A mutation detection assay using MutS is based upon testing for heterozygosity at a given locus. If an individual is heterozygous (Htz) at a particular locus then there is a difference in the genetic code between both copies of the genes whereas a homozygote (Hmz) contains an identical sequence at that locus. The difference between alleles of the heterozygote means that heteroduplex DNA can be produced experimentally via PCR amplification. In the case of heterozygotes reannealing of the PCR product gives rise to four different species of DNA: 2 heteroduplexes where single nucleotide strands from the two genes have annealed to one another and formed a base mismatch: 2 homoduplexes where strands from the same gene have annealed and the sequences are complementary. For homozygotes the PCR product would consist of only homoduplex product, as the strands are identical. The difference between the alleles determines the extent of the mismatch be it a single base mismatch to small loops in one of the strands. MutS can therefore potentially be used to detect individuals who are heterozygous for a particular gene by binding to the heteroduplex PCR product and causing a mobility shift (Fig. 3.1).

3.1.2. MutS Purification

MutS was purified using a single step purification using His-tag and affinity chromatography (Feng and Winkler, 1995). Previous work under taken by members of the Fox group had established both pETMutS and

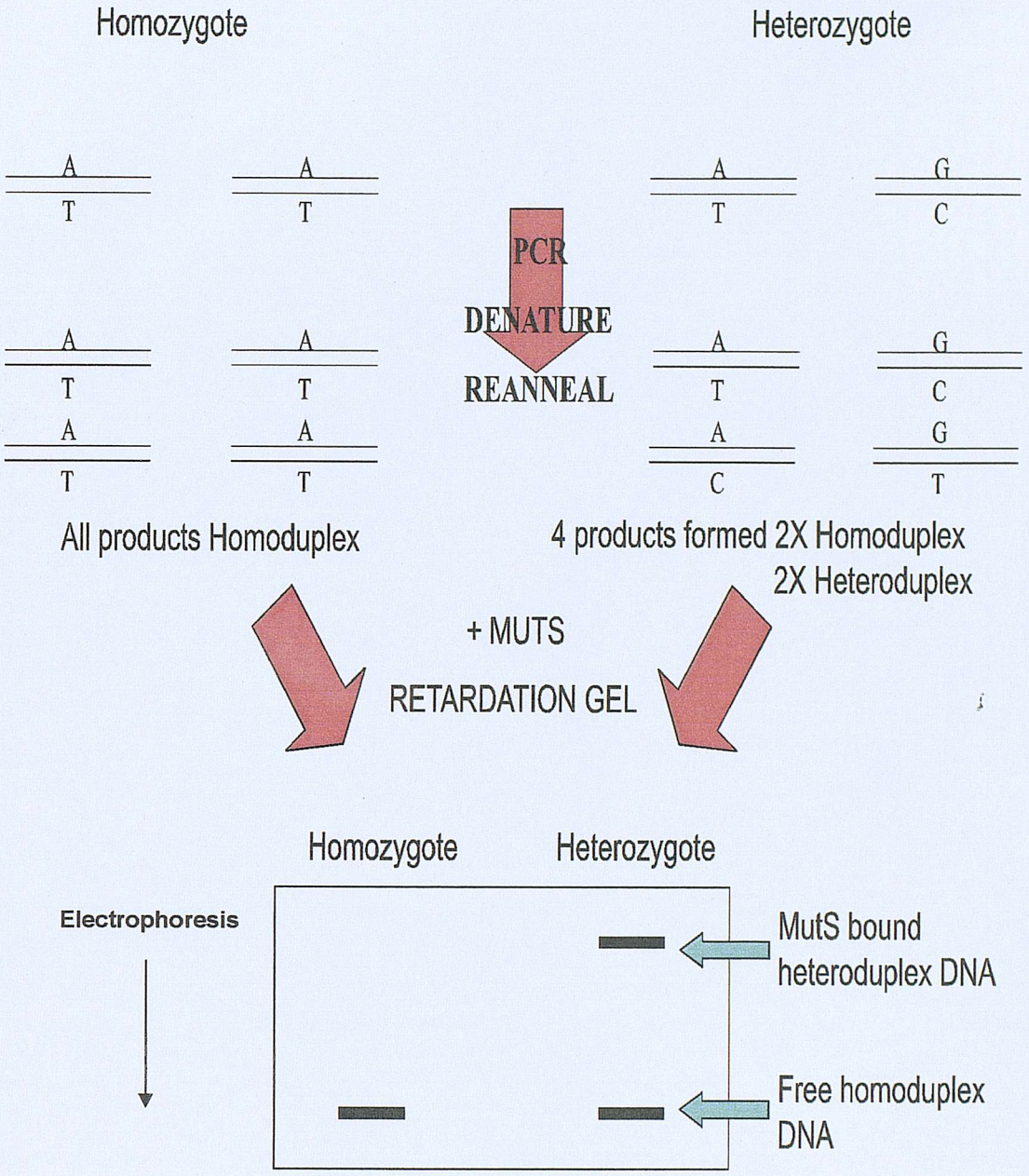


Figure 3.1 Experimental Strategy for detection of heterozygosity using the mismatch binding properties of MutS.

glycerol stocks for high expression and purification as well as purified protein (Brown et al., 2001).

3.2. Results

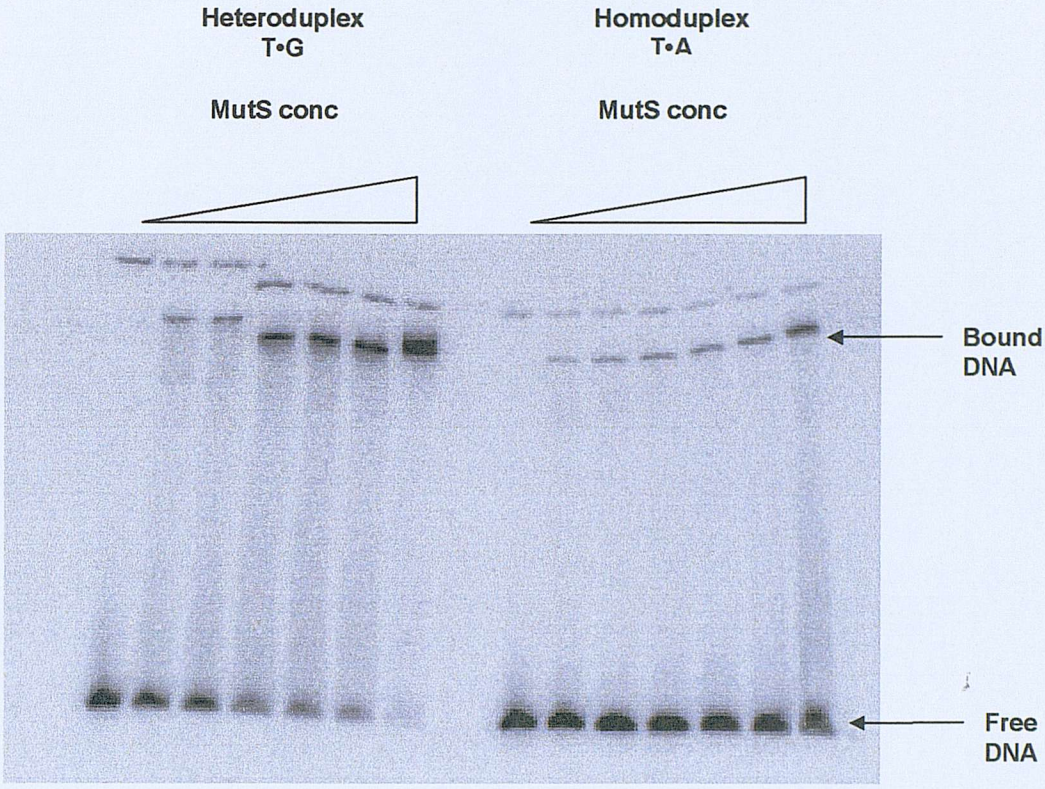
3.2.1. Mobility Shift of 50bp T•G Heteroduplex and T•A Homoduplex

To examine the mismatch binding specificity of MutS for a heteroduplex over a homoduplex, a range of MutS concentrations were incubated with ^{32}P radiolabeled DNA (Chapter 2 pg 72 for duplex sequence and mismatch location prefix TB0). Two 50mer DNA duplexes one containing a single T•G mismatch and one with the correct Watson and Crick base pairing at the mismatch site were incubated with MutS and the resulting MutS-DNA complexes run on a conventional non-denaturing gel.

As indicated in Figure 3.2, as the MutS concentration increased so does the quantity of DNA contained in the mobility-shifted band. The absence of MutS acts as a control and demonstrates that the band retardation seen on the gel is the result of the presence of MutS. By comparing the control (-MutS) with the other samples, it can be observed that there is a loss of signal from the free DNA, which is transferred to the bound band as MutS concentration increases. This pattern is present not only in the heteroduplex but also to a lesser extent in the homoduplex indicating non-specific binding of MutS. Quantitative analysis indicates that at its peak the discrimination between the two duplexes was approximately 8 fold in favour of the heteroduplex. These results indicate that MutS binds to the T•G mismatch stronger than the correct Watson and Crick base pairing and that there is a discernable discrimination between heteroduplex and homoduplex DNA.

3.2.2 Mismatch Binding Specificity of MutS

MutS was incubated with identical DNA duplexes each containing one of the eight mismatches (Fig 3.3). The final concentration of MutS within the reaction mix was 0.6pmole/µl and was based upon the MutS concentration that gave the greatest discrimination between Ht and Hm duplex DNA in the



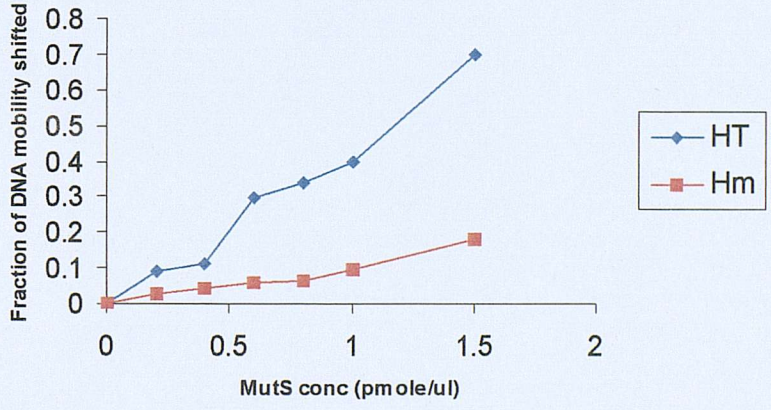


Figure 3.2. MutS concentration titration (final concentration in reaction mix was 0, 0.2, 0.4, 0.6, 0.8, 1 and 1.5pmole/ul) using a ^{32}P labelled 50bp duplex. DNA duplex containing a single T•G mismatch and a DNA duplex with correct base pairing were incubated on ice for 30minutes before loading onto a 6% non-denaturing PAGE. Gel was imaged as described on page 98 and the bands analysed using Phoretix. The graph displays the fraction of DNA present in the band shift from the total DNA within the well for both the heteroduplex and the homoduplex as determined by phoretix

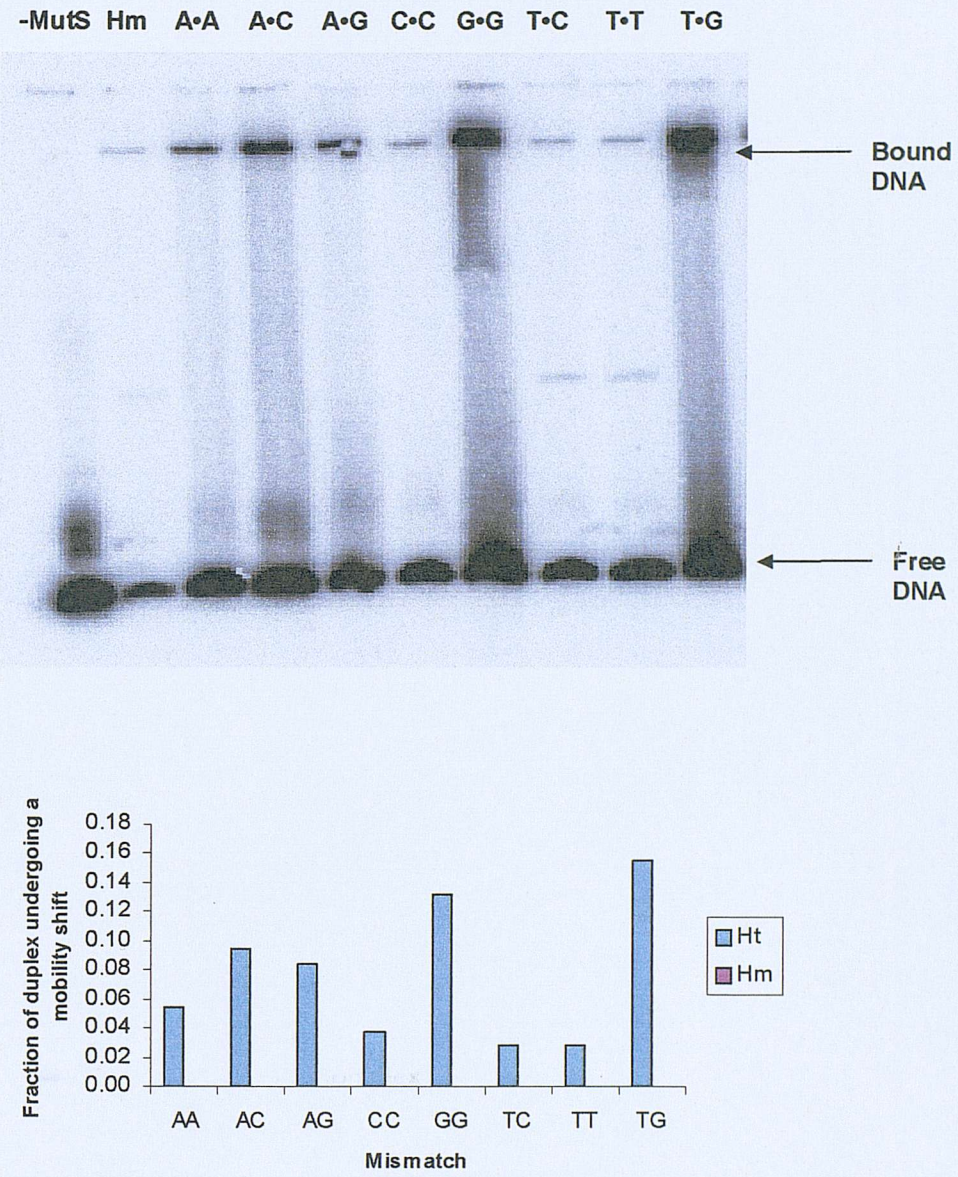


Figure 3.3. Comparison of the binding specificity of MutS for the different mismatches. The final MutS concentration within the reaction was 0.6pmole/µl. MutS-DNA complexes were run on a 6% non-denaturing PAGE and the bands quantified using ImageQuANT. The results are displayed as a bar chart with the fraction of DNA present in the band shift for each different mismatch with the homoduplex mobility shift displayed along side for comparison.

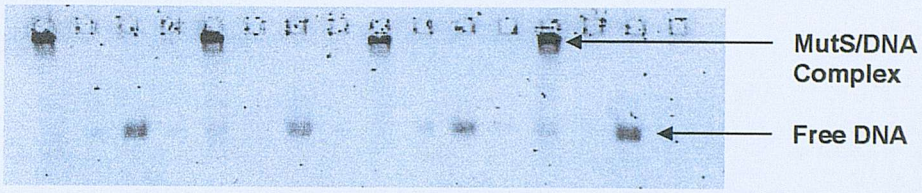
titration experiment. The fraction of DNA present in the mobility shifted fragment varied dependent upon the identity of the mismatch. Based upon the values obtained from the best experiment, G•T and G•G displayed the greatest mobility shift with 7-8 fold greater specificity over the homoduplex. A•A, A•C and A•G displayed a 3-5-fold greater mobility shift than the homoduplex. The remaining mismatches C•C, T•C and T•T all displayed very poor binding with very little difference between them and the homoduplex. G•T, G•G, A•A, A•C and A•G all displayed various degrees of smearing down the gel. This is the result of dissociation of the duplex from MutS during the electrophoretic run.

3.2.3. Gel Conditions for Short Run Electrophoresis

In previous literature MutS has been shown to be capable of producing a mobility shift under a number of conditions, but has never been examined under MADGE conditions. In order to investigate how MutS and the MutS/DNA complex would respond to being electrophoresied on MADGE, a range of variables were looked at that could all have some influence on creating a successful band shift. For preliminary studies H-PAGE gels were used. The only difference between these and MADGE gels is the angle the wells are set at. Both H-PAGE and MADGE gels can be run either submerged or dry. For rapid mutation studies it would be more cost effective and less labour intensive to be able to run the gels dry. Comparison of gels run under wet and dry conditions using 3ul samples from the same incubation mix show that there is little difference between submerged and dry gels and that the images produced are similar (Fig. 3.4).

Incubation of MutS and DNA is carried out on ice so as to maintain the stability of the MutS protein, although for the short incubation times used in the MutS binding reactions this may not be entirely necessary. Contrasting gels where the temperature has been controlled at 4°C with gels that have undergone no stringent temperature control except for the incubation being at 4°C indicates that the stability of the MutS/DNA complex is not affected by being run at 4°C or room temperature (Fig. 3.5). At both temperatures distinct band shifts are observed with both gels displaying a similar band shift pattern.

(A) Submerged Gel



(B) Dry Gel

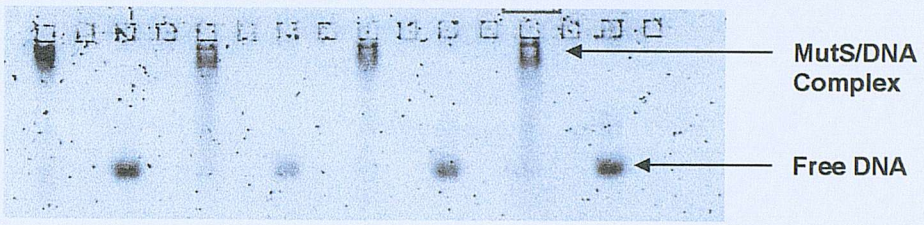
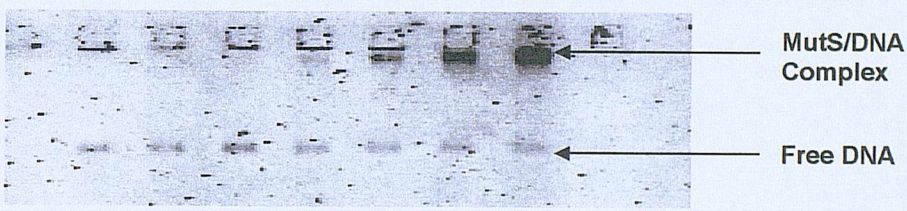


Figure 3.4. Comparison of MutS/DNA band shift gel run in a submerged in 1XTBE (A) or run dry (B) in a dry electrophoresis box.

(A) 4°C



(B) Room Temperature

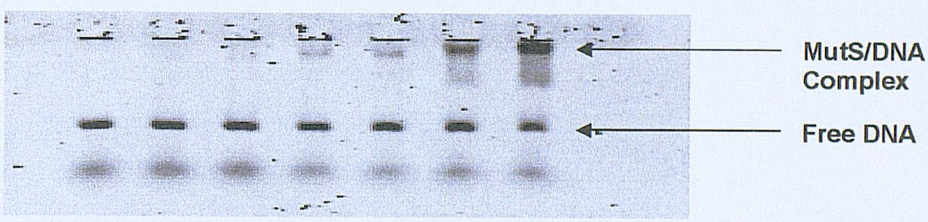
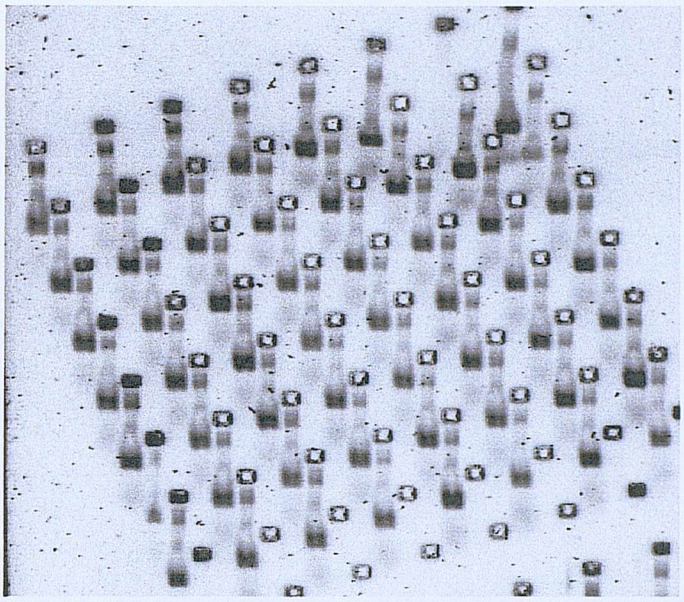


Figure 3.5. Comparison of MutS mobility shift gels run at 4°C and room temperature.

(A) HEPES



(B) TRIS

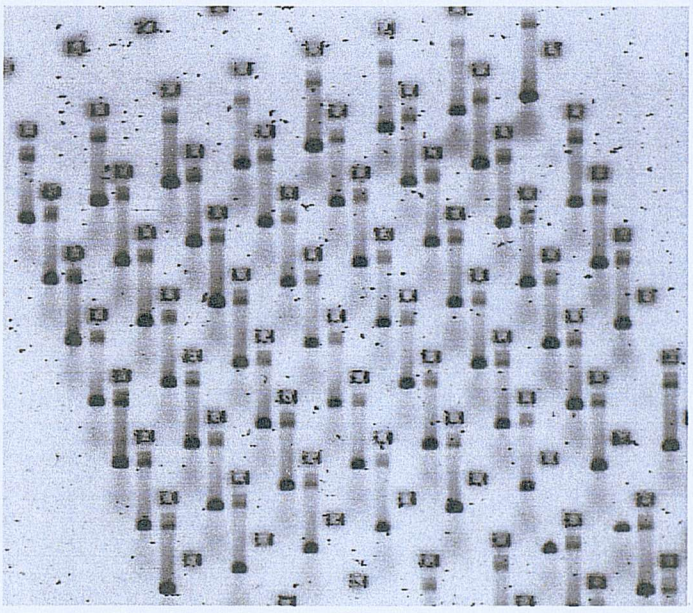


Figure 3.6. Comparison of gel images produced when using HEPES or Tris in the MutS buffer. The DNA samples and conditions used were the same the only difference being that MutS was diluted in either a buffer containing HEPES or Tris. Gel (A) is buffer containing HEPES, (B) Tris. This crisper image produced using Tris buffer makes it easier to quantify the bands using Phoretix as opposed to the HEPES buffer where DNA from one well would appear in the neighbouring well (Figure 3.6).

3.2.4. MutS Dilution Buffer

Comparison of the two most frequently used buffers to dilute MutS indicated that MutS binding was not dependent upon the buffer present as use of both gave similar band retardation. Of the two Tris produced a cleaner gel image and it was observed (Fig 3.6) that the use of HEPES buffer caused a teardrop affect.

3.2.5. Binding of MutS to 120bp *IGF2 Apa*I Amplicon

Figure 3.7 shows a MutS concentration titration using a 120bp DNA fragment generated from the *IGF2 Apa*I site. The MutS dilution range was 0-0.8pmole/µl in the final 10µl reaction mix. Confirmation of the identity of the PCR amplicons was achieved via restriction digest as explained in chapter 2. After denaturing and reannealing the resulting DNA species of the G•C/A•T heterozygote would include two homoduplexes with G•C or A•T base pairing at the polymorphic site and two heteroduplexes with a G•T or an A•C mismatch. The resulting gel images (Fig. 3.7) indicate that as the concentration of MutS increases then the retardation of the DNA band increases. The resulting band shift is not the result of non-specific protein interaction as replacement of MutS with bovine serum albumen (BSA) failed to produce any band mobility retardation. As with synthetic DNA duplexes both the heteroduplex species and homoduplex species undergo band shifts making if very difficult to distinguish accurately between the two using the gel images. As the human eye finds it hard to distinguish shades of grey the bands may appear very similar but there could be a considerable quantitative difference, for this reason analysis using Phoretix was used. Phoretix enables the calculation of the fraction of material shifted from the total amount of material within the track by quantifying the bands. When applied to titration reactions a trend can be seen with the fraction of DNA mobility shift for heterozygous slightly greater compared to the homozygous at equal MutS and DNA concentration. The data suggests that the clearest distinction between heteroduplex DNA and homoduplex DNA is between 0.4-0.5pmole/µl

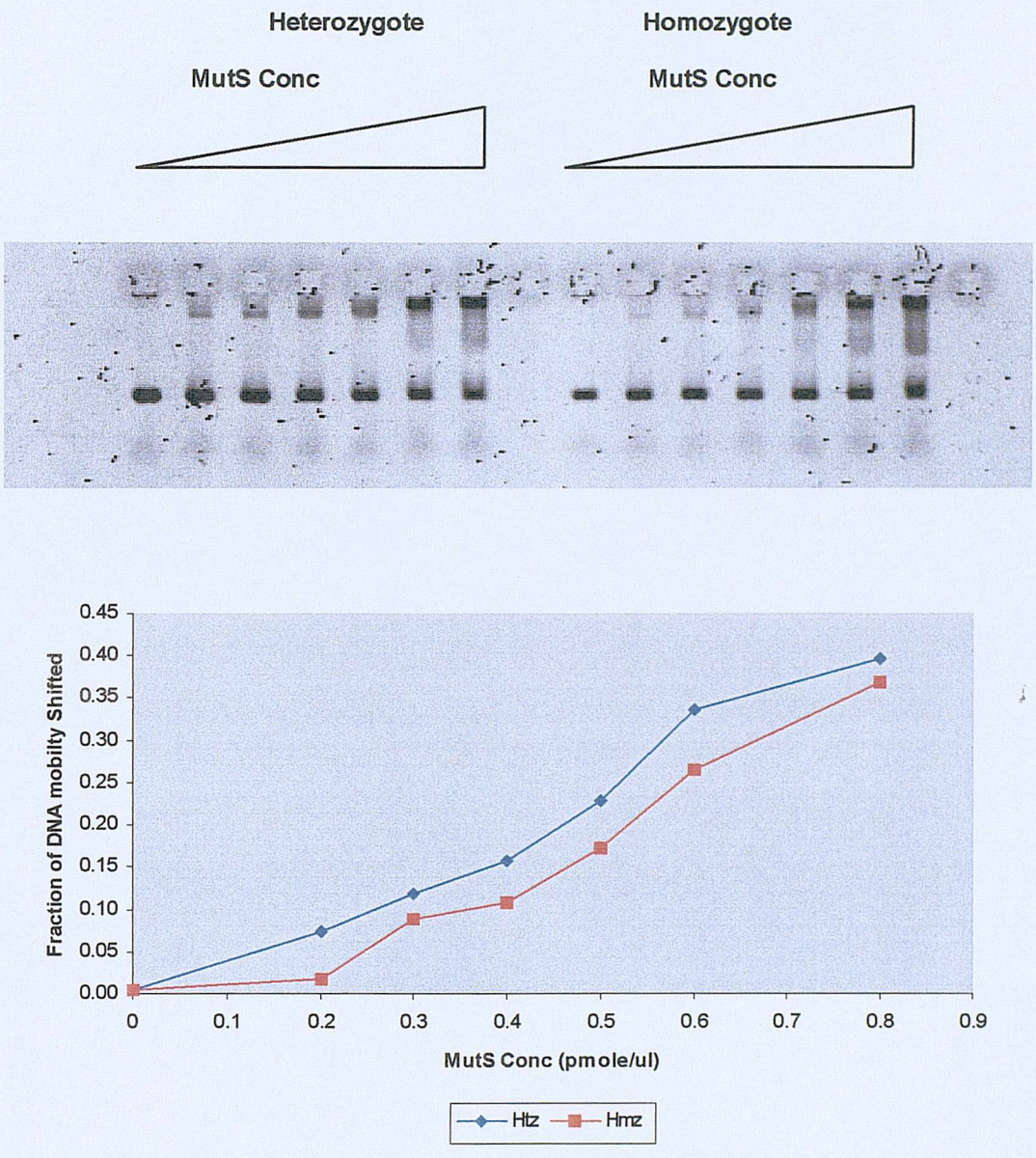
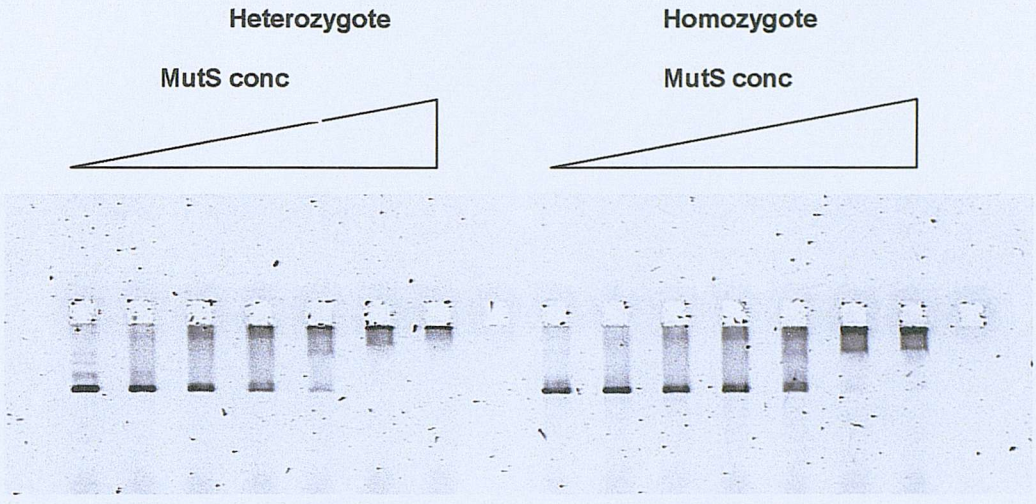


Figure 3.7 MutS titration (the final concentration of MutS in reaction mix was 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 pmole/µl) using *IGF2 Apa*I 120bp fragment. Band analysis performed using Phoretix and the amount of DNA retarded calculated as a fraction of the total amount of DNA present in each lane.



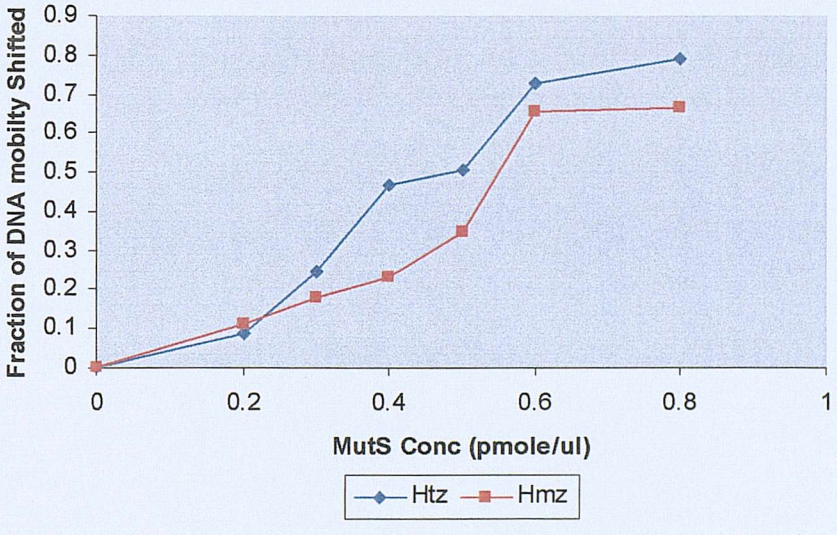


Figure 3.8 MutS titration (the final concentration of MutS in reaction mix was 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 pmole/µl) using *IGF2 Apa*I 313bp fragment. Band analysis performed using Phoretix and the amount of DNA retarded calculated as a fraction of the total amount of DNA present in each lane.

of MutS a molar ratio of 20-25: 1 to the target DNA. However the discrimination is not good.

3.2.5.1. Binding of MutS to 313bp *IGF2 Apa*I Amplicon

A second larger amplicon of 313bp from the same site was also subjected to band shift analysis using MutS (Fig. 3.8). This showed a similar trend in comparison with the shorter amplicon. An interesting observation is that at the higher MutS concentrations almost all the DNA was mobility shifted whereas for the shorter amplicon only approximately half the DNA was bound. Picogreen analysis indicated that the concentrations of the two amplicons used were similar, therefore this increased mobility shift was probably due to the increase in amplicon length.

3.2.5.2. MutS Band Shift analysis of 55 *IGF2 Apa*I amplicons using MADGE

A collection of 80 120bp PCR amplified *IGF2 Apa*I fragments from a DNA bank of previously typed samples were subjected to MutS mobility shift as described previously. The final concentration of MutS in the reaction was 0.5pmole/µl. The concentration of the PCR amplicon was determined by picogreen analysis and ranged from 5ng/µl → 17ng/µl for the 80 samples. 56 samples were chosen as highlighted by the blue box in the picogreen image (Fig 3.9). This provided 8 rows containing 7 samples each. The rows were labelled A→ H and the samples 1-7 as shown. The concentration of the samples chosen ranged from 10ng/µl → 17ng/µl. The only sample excluded was H6 due to the low concentration. From the initial 56 samples selected 55 samples were treated to MutS band shift consisting of 40 homozygotes and 15 heterozygotes. Figure 3.10 compares the gel images for the restriction digest and the MutS band shift of the same samples. For each assay a heterozygote and one of each of the homozygotes have been high lighted. In the case of the *Apa*I digest this displays the cleavage pattern seen for the genotypes. Homozygotes samples either had a single undigested band (A/A) or 2 smaller complete digest fragments (G/G) whilst the heterozygotes contained 3 bands. In terms of the restriction digest the genotype of the

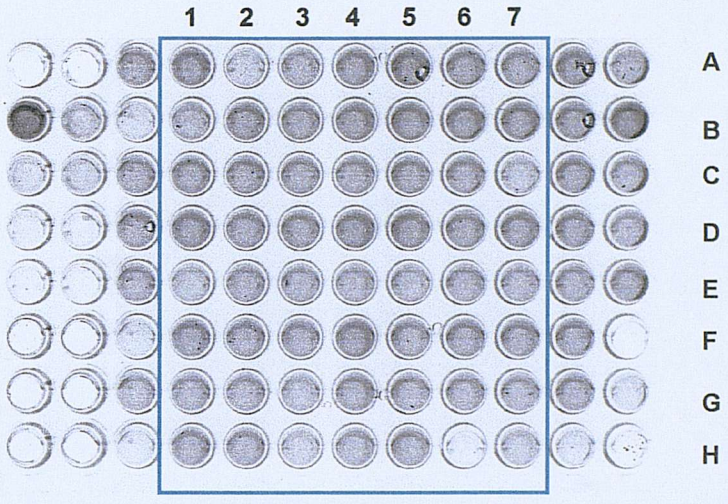


Figure 3.9. Picogreen PCR yield quantification image. The samples chosen for MutS band shift analysis are highlighted in the blue box with the columns numbered 1-7 and the rows A-H.

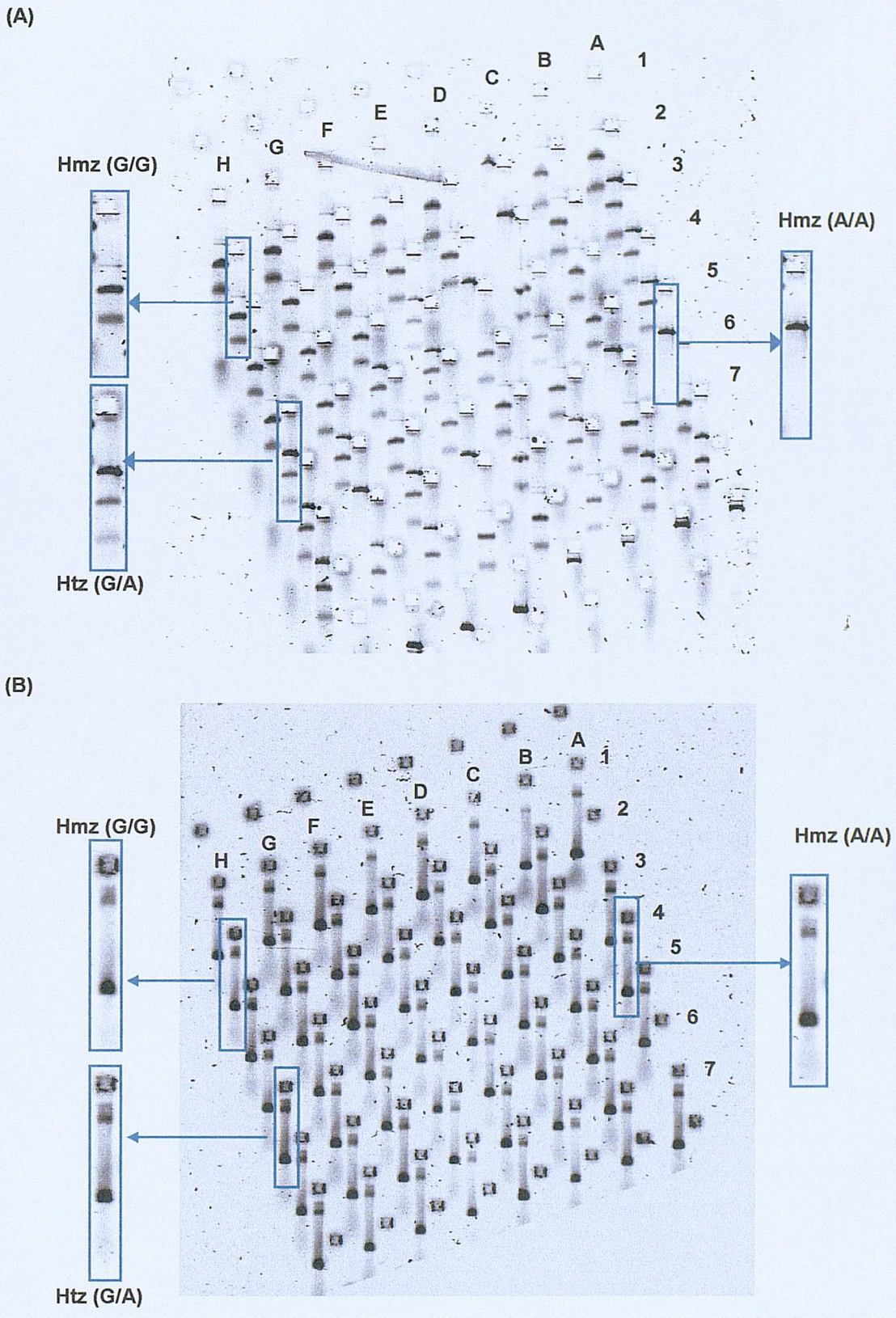


Figure 3.10. Collection of *IGF2 Apa*I amplicons subjected to restriction digest (A) and MutS mobility shift (B). The blue boxes highlight the same heterozygote and homozygotes in each assay as an example of the restriction digest pattern and the retarded DNA species. The final concentration of MutS in the reaction was 0.5pmole/µl)

Sample	Genotype	PCR Conc ng/µl	% Band shift	Sample	Genotype	PCR Conc ng/µl	% Bandshift
A1	GG	16	12	E1	GG	11	13
A2	GA	10		E2	GG	14	21
A3	GA	13	24	E3	GA	13	17
A4	GA	15	17	E4	GA	13	23
A5	AA	17	20	E5	GG	13	16
A6	GG	15		E6	AA	13	15
A7	GA	14	21	E7	GG	13	13
B1	GG	12	6	F1	GG	15	17
B2	GG	15	19	F2	GG	14	21
B3	GG	14	21	F3	GG	14	17
B4	GG	15	22	F4	GA	16	21
B5	AA	16	18	F5	GG	15	16
B6	GG	16	19	F6	GA	15	20
B7	GA	16	17	F7	GA	15	19
C1	AA	14	16	G1	GG	13	20
C2	AA	14	22	G2	GG	13	17
C3	GG	14	18	G3	GG	12	19
C4	GA	14	19	G4	GG	15	20
C5	GG	15	16	G5	GA	14	19
C6	GA	14	20	G6	AA	14	18
C7	GG	11	18	G7	GA	13	19
D1	GG	15	17	H1	GG	14	23
D2	GG	14	18	H2	GG	13	18
D3	AA	14	18	H3	GG	10	25
D4	GG	16	15	H4	GG	13	20
D5	GG	16	18	H5	GA	15	24
D6	GG	15	16	H6	GA	7	23
D7	GG	15	12	H7	GA	11	19

Table 3.1. Concentration of PCR amplicons and the % of DNA as a band shift in the MutS assay. Homozygotes are not high-lighted, heterozygotes blue and dropouts red.

samples is clearly determinable from the cleavage pattern, however all the samples in the MutS assay display mobility shift. Two tracks A2 and A6 failed to show any DNA, which was most likely a loading error. Phoretix was used to calculate the proportion of DNA in each track that had undergone mobility shift (Tab. 3.1) to see if the genotype correlated with the proportion of DNA retarded. The heterozygotes are highlighted in blue and the percentage of DNA present as a retarded band ranged from 17%→ 24%, whilst the homozygotes (clear) ranged from 6% → 25%. The difference between Htz and Hmz is not great enough to allow positive identification reliably and there appears to be a significant amount of cross over between the two.

One explanation for this cross over could be due to variation in DNA concentration due to the PCR reaction. Picogreen analysis however indicated that the amplicon concentrations were consistent. Comparisons of PCR concentration and band retardation failed to display any significant correlation. A higher concentration of PCR product did not automatically lead to an increase in DNA bound by the MutS. For example the heterozygote sample H5 had a calculated band retardation of 24% with a concentration of 15ng/µl, whilst the values for B7 were 17% and 16ng/µl. A3 had a lower PCR yield of 13ng/µl but the band retardation was measured at 24%. A similar situation existed for the homozygotes with no correlation between PCR yield and quantity of DNA mobility shifted.

3.2.5.3. Effect of Polymerase Fidelity on MutS Binding

A second possible reason that could contribute to the binding of MutS to homozygote PCR amplicon is the generation of mismatches during PCR. To investigate this possibility a two polymerase system was employed (Barnes, 1994) which provides optimal levels of both processive and proofreading activity. The PCR reaction was supplemented with the hyperthermophilic archaebacterium Pyrococcus woesei DNA polymerase Pwo which displays a 10 fold increase in fidelity over Taq, attributed to its 3'-5' exonuclease activity. The result of the mobility shift assay performed on the reannealed 120bp *IGF2* amplicons generated using Taq supplemented with Pwo is shown in figure 3.11. As with previous MutS concentration titrations

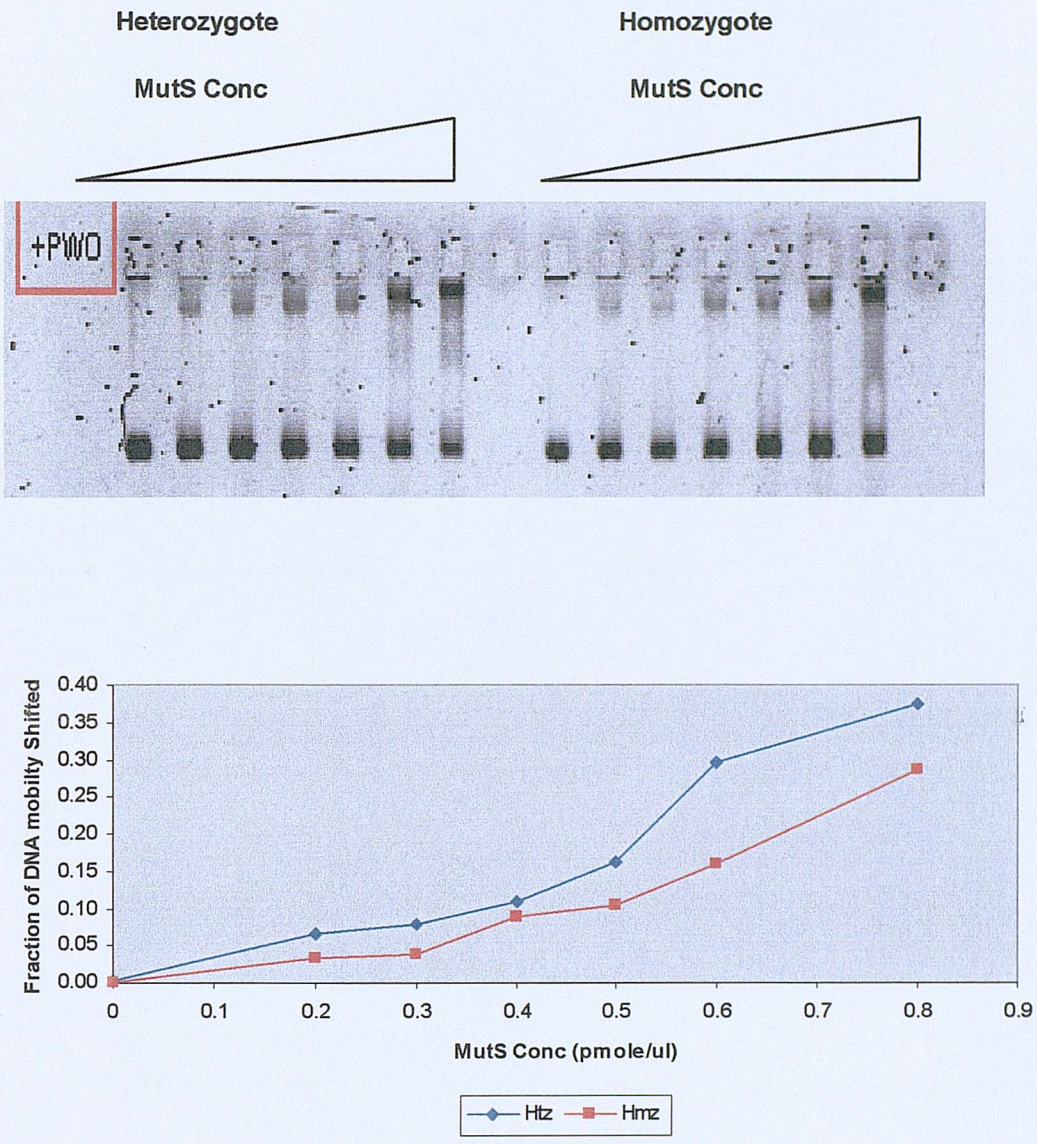


Figure 3.11 MutS titration (the final concentration of MutS in the reaction was 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 pmole/ul) using *IGF2 Apa*I 120bp fragment generated with Taq and Pwo. Band analysis performed using Phoretix and the amount of DNA retarded calculated as a fraction of the total amount of DNA present in each lane.

both the heterozygote and homozygote displayed mobility shift, with the heterozygote demonstrating a slight increase in mobility shift over the homozygote. However the inclusion of Pwo has not reduced the non-specific binding seen in the homozygote. This would indicate that the non-specific binding is not due to mismatches introduced due to Taq polymerise errors.

3.2.9. Influence of ATP upon MutS/DNA Binding and Specificity

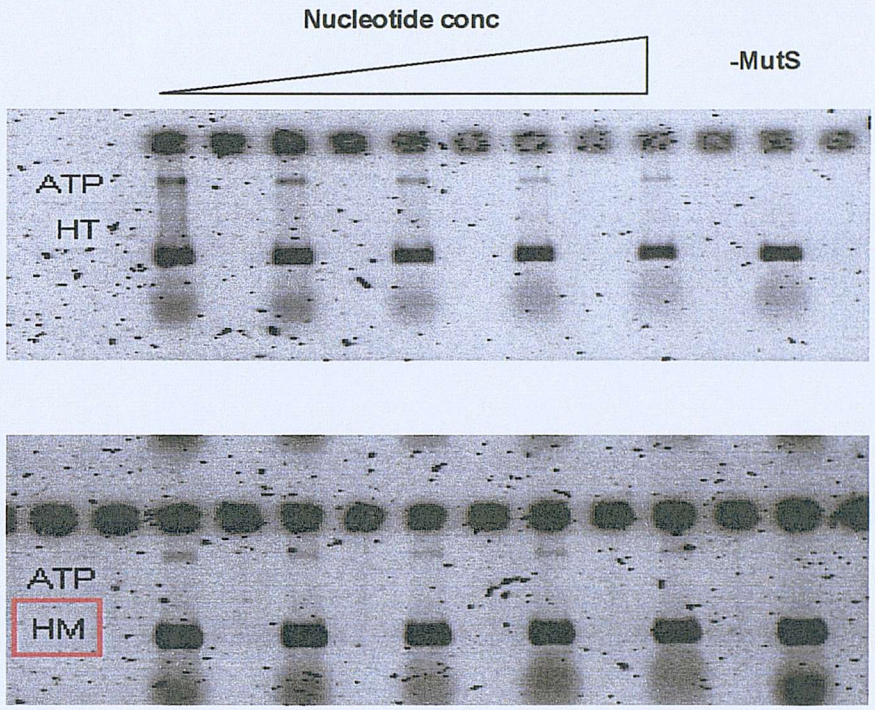
ATP affects the binding of MutS as described in chapter 1. To determine if the MutS was functioning normally the MutS/DNA incubation mix was supplemented with ATP to determine if it would influence binding. ATP concentrations ranged to give a final concentration of 1mM, 2mM, 5mM and 10mM in the reaction mix. The resulting data (Fig. 3.12) indicated that the presence of increasing ATP led to a reduction in the amount of DNA retardation in heterozygotes. Homozygotes also displayed loss of signal in the band shift but this appeared to be far less dramatic than the one seen heterozygotes. This data suggests that the presence of ATP causes the release of MutS from the mismatch.

3.2.10. Influencing MutS Binding

Although not displayed in this thesis a number of reaction conditions were investigated to try and influence the binding of MutS to DNA too improve heterozygote and homozygote discrimination. This included $MgCl_2$ and salt titrations as well as varying incubation time and temperature. All met with little success in improving the discrimination of mismatch DNA from Watson and Crick DNA.

3.2.11. *MMP-1* Amplicon MutS Band Shift

As an alternative target a short PCR amplicon was generated from the *MMP-1* gene containing an *Xmn*I restriction site (AF023338). Similar to the *IGF2 Apa*I amplicon the amplified region contains a restriction site in which a sequence alteration occurs. One allele contains the intact restriction site



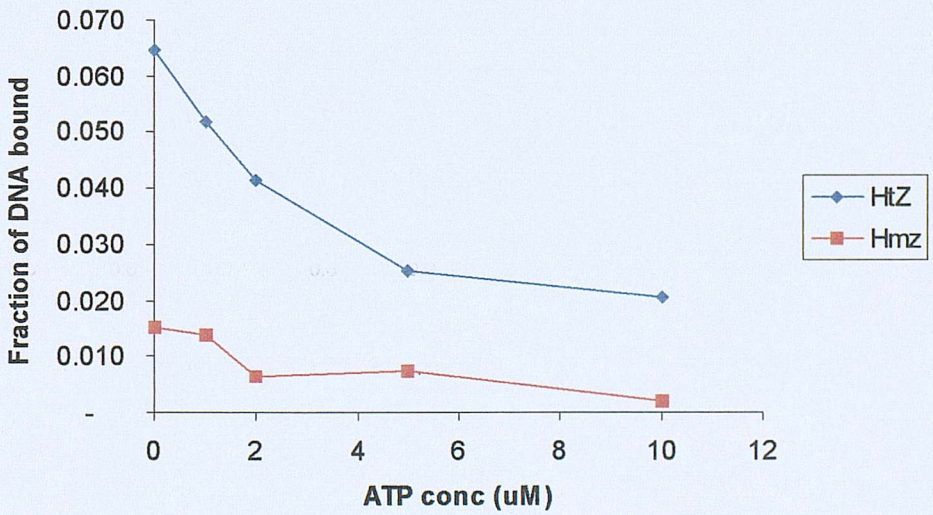


Figure 3.12. Affect of ATP on the fraction of DNA undergoing a mobility shift. The concentration of ATP used was 0, 1μM, 2μM, 5μM and 10μM.

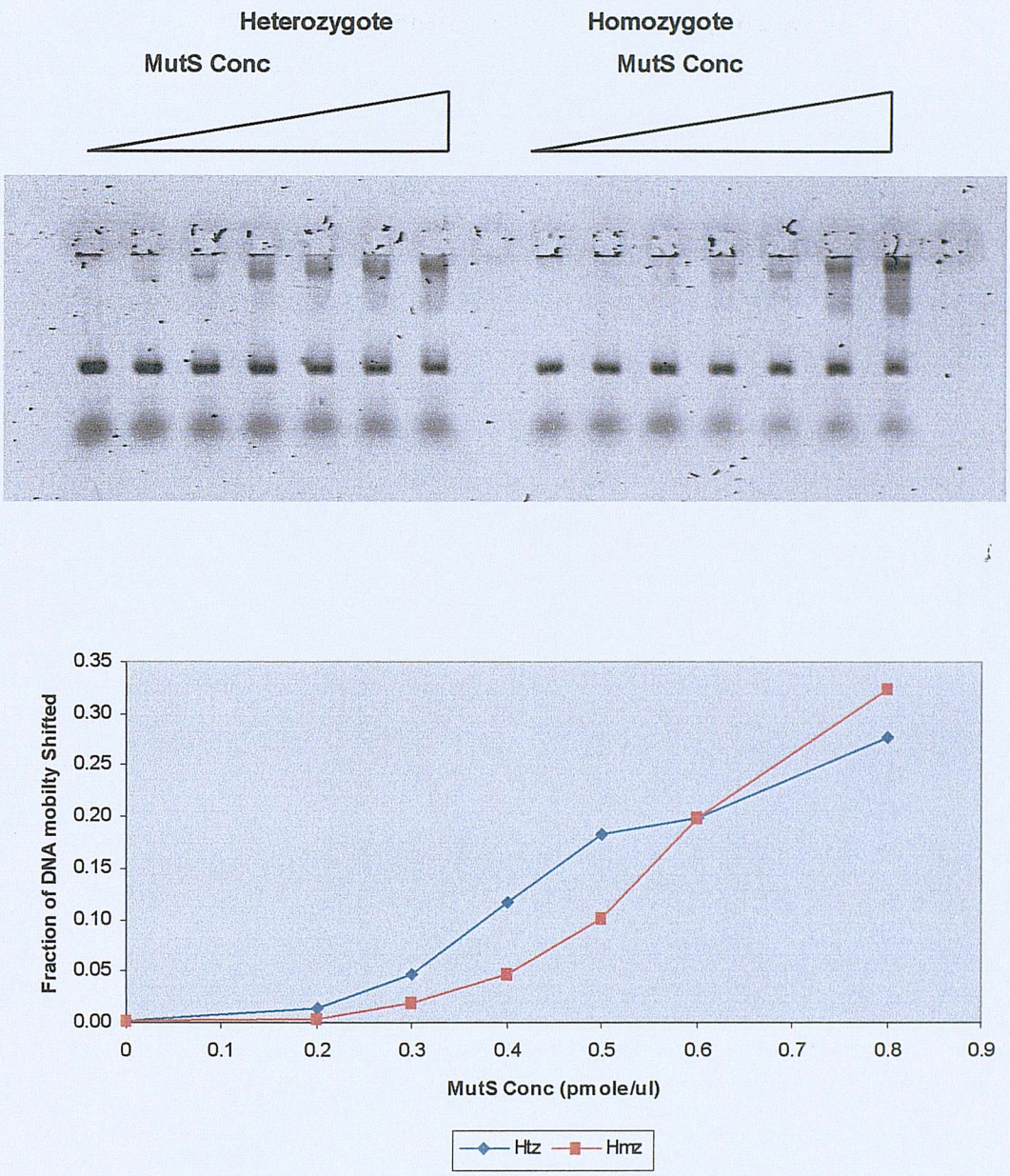


Figure 3.13 MutS titration (the final concentration of MutS in reaction mix was 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8 pmole/µl) using *MMP1 Xmn*I 118bp fragment. Band analysis performed using Phoretix and the amount of DNA retarded calculated as a fraction of the total amount of DNA present in each lane.

whilst the other contains a G insertion, which removes the restriction site. As with *IGF2* this provides a simple method to identify the alleles present in sample DNA via restriction digests. Reannealing of heterozygous PCR products will generate a single unpaired nucleotide within the DNA molecule.

In house control DNA was amplified using MMP1-F & MMP1-R primers creating an 118bp amplicon. The genotype was determined via *Xmn*I restriction digest and a heterozygote and wildtype homozygote subjected to MutS band shift following heteroduplex formation (Fig 3.13). Similar to the *IGF2* amplicon both the heterozygote and homozygote displayed increased DNA retardation as the MutS concentration increased. At the highest MutS concentration tested the homozygote displayed greater band retardation than the heterozygote.

3.3. Discussion.

3.3.1. Mismatch Recognition

The discrimination of MutS for heteroduplex over homoduplex DNA has been approximated to a 10 – 20 fold difference (Feng *et al.*, 1995). Band shift assays using 100% heteroduplex whilst displaying a difference in the fraction of DNA mobility shifted when compared to the complementary homoduplex, did not display as great a difference as previously reported. The value determined in this study placed the discrimination at approximately 8 fold for the best-recognised mismatch. It has however been reported (Brown *et al.*, 2001) that the length of the DNA fragment could influence this value. As the synthetic DNA fragment used here was larger than that used previously this could explain the lower discrimination observed. The order of mismatch recognition determined by this study was T•G > G•G > A•A≈ A•C > A•G > C•C ≈T•C ≈T•T ≈T•A, which correlates well with previous studies (Kramer et al., 1984).

3.3.2. Detection of Heterozygotes from Homozygotes using MutS Band Shift

When compared to the discrimination between 100% heteroduplex and

homoduplex the discrimination between heterozygotes and homozygotes has been reduced. This however is not surprising as the maximum amount of heterodupiex in any heterozygous sample would be 50%. When coupled with the fact that there are two heteroduplex species for each heterozygote, and the poor specificity of MutS for some mismatches, this further reduces the difference between the band shifts observed for both homozygotes and heterozygotes.

Both the 120bp, 313bp *IGF2* amplicon and *MMP-1* amplicon displayed significant non-specific binding. The increased mobility shift seen in the 313bp amplicon compared to the 120bp amplicon further supports the idea that MutS binding increases with fragment length. As the fragment length increases the number of available wt homozygotes sites must also increase. Mismatch correction of transition mismatches by MutS and the methylated mismatch repair pathway has been observed to be better than that of the correction of transversion mismatches (Jones et al., 1987). The transition mismatches are A•C and T•G and so it can be seen that repair efficiency does not necessary correlate to the mismatch recognition of MutS. This has lead to suggestions that other factors must influence the efficiency of mismatch repair other than the mismatch recognition of MutS (Brown *et al.*, 2001).

The *IGF2* amplicon G→ A transition mutation generates a G•T and A•C heterodupiex. In terms of the possible mismatch pairs that can be generated between two differing alleles this produces the two most favourable mismatch heteroduplexes. Therefore the poor discrimination between homozygotes and heterozygotes cannot be attributed to the identity of the mismatches involved. There is some evidence to suggest that the immediate sequence flanking the mismatch could play a role in the efficiency of mismatch repair by MutS. Studies have shown that mispair repair is directly proportional to the GC content 4bp on each side of the mismatch (Fazakerley et al., 1986). The *IGF2 Apa*I site mismatch is within a high GC content, where 6 or the 8 bases flanking the mismatch are GC, therefore the immediate flanking region should favour efficient MutS recognition.

MutS has been shown to bind to oligonucleotide duplexes containing up to 4 unpaired bases (Parker and Marinus, 1992). Therefore as a second substrate to test MutS mutation detection the MMP-1 amplicon containing a

single G deletion is suitable. Similar to the oligonucleotide duplexes a reannealing of a heterozygote will form 2 different species, a perfectly matched species and a species with a single unpaired base. Similar to the *IGF2* amplicon both heterozygotes and homozygotes both displayed mobility shifts.

Generating heteroduplex DNA using the method outlined in this chapter does not differentiate between wild-type or mutant homozygotes. The use of a wildtype probe to generate heteroduplex would enable both heterozygous and homozygous mutations to be determined. In the case of mutant homozygotes this may provide a quantitative difference in DNA mobility shift, as the relative proportion of heteroduplex species to homoduplex species would increase. In terms of heterozygotes the relative proportion of each DNA species remains the same with 2 homoduplex species and 2 heteroduplex species. However for screening purposes a mutation detection tool needs to be able to detect for heterozygosity.

3.3.3. Non-Specific MutS Binding

For both synthetic homoduplexes and PCR generated homoduplexes there was significant non-specific binding. Structural studies do offer some explanation to the observed binding of MutS to homoduplex. Numerous groups have postulated non-specific binding domains for MutS (Allen et al., 1997;Tachiki et al., 1998). Furthermore the purposed mechanism of mismatch scanning infers the initial binding to homoduplex DNA prior to mismatch detection. Before mismatch detection the entire DNA molecule is engulfed by the two MutS subunits where it is held as the DNA is probed for mismatches. The presence of a mismatch enables the correct conformational change required for the ATP dependent recruitment of MutL.

3.3.4. Effect of ATP on MutS-DNA Complex

Observations have indicated that binding of ATP reduces the stability of the MutS/DNA complex resulting in the release of MutS from the mismatch. The presence of DNA and ATP causes a conformational change in *E.coli*

MutS (Joshi et al., 2000). Addition of ATP to MutS/DNA complexes caused a reduction in the amount of DNA in the band shift indicating a release of MutS from the DNA. It is believed that this release is dependent on the correct DNA binding of MutS to a mismatch, which would explain the reason why the heterzygote sample showed a much larger effect upon the presence off ATP than the homozygote. In vitro studies suggest MutL prevents the release of MutS from the mismatch (Galio et al., 1999). This could stabilise the mismatch binding and increase the efficiency of mismatch detection (Lamers et al., 2000).

3.3.5. MutS and Short Track Electrophoresis

MutS successfully bound to and caused a visible mobility shift of PCR product within the MADGE format. This mobility shift is easily observable using short run electrophoresis and DNA staining dyes. The resulting band-shifts are suitable for analysis using Phoretix enabling quick and easy analysis of multiple samples. However as a mutation detection tool MutS proved to be very disappointing. In the scanning of 55 *IGF2* samples, MutS could not consistently produce a stronger band shift for heterozygotes than in the amplified DNA of both homozygotes. In this format MutS displayed significant non-specific binding to homozygous DNA where there was no mismatch. Binding studies of synthetic duplexes demonstrated that MutS does recognise some mismatches specifically over the corresponding homoduplex DNA. However it appears that MutS does not have the sensitivity to detect lower levels of heteroduplex within a background of homoduplex and translate this into a consistent quantative difference. This made analysis difficult and reduces the effectiveness of MutS as a mutation-scanning tool as identification of heterozygotes from homozygotes was extremely difficult. This non-specific binding to homoduplex reduces its overall effectiveness for mutation detection. When this is coupled with the poor recognition of some of the mismatches the use of MutS for high throughput mutation detection in this form observed to be severely limited.

CHAPTER 4
CLONING, HIGH YIELD EXPRESSION AND INITIAL CHARACTERISATION OF THE MISMATCH CLEAVAGE ENZYME T4 ENDONUCLEASE VII

4.1. Introduction.

As a mismatch cleavage enzyme T4 endonuclease VII is potentially an ideal candidate for mutation detection. This chapter presents the results of simplifying the purification by reducing a two-step purification (Golz et al, 1995) to a one step purification utilising His-tag technology. This chapter also covers the initial characterisation of the protein and the detection of cleavage fragments using short track electrophoresis.

4.2. Results

4.2.1. Cloning of Endo VII into pET15b

Figure 4.1 illustrates the resulting difference in plasmid size when a successful ligation has occurred when compared with a control vector that has not undergone a ligation reaction. This was achieved by introducing an *Nde*I site at the 5' end (C^ATATG) and a *Bam*HI site (G^GATCC) at the 3' end of the gene in pRB210 using the primers outlined on pg71. The successful ligation resulted in a recombinant His-tagged Endo VII protein. In figure 4.1 the first two samples following the size standard both contain the 474bp Endo VII gene insert. The inclusion of the Endo VII gene results in larger fragments being observed when compared with the control sample (pET15b with no insert). On an agarose gel these larger fragments do not migrate as quickly as the control therefore a difference in mobility indicated successful ligation. The vector can exist in three formations as indicated by the three bands in each lane. These forms are supercoiled, linear and circular. The predominant species in figure 4.1 is circular. The resulting plasmid was termed pETEndoVII.

Confirmation that the extracted plasmids contained Endo VII gene at the correct site was achieved by amplifying the cloning region of pETendoVII using T7 primers. The bands displayed in figure 4.2 confirm that the EndoVII gene is within the cloning region of pET15b. A vector that has undergone a successful ligation reaction produces a 694bp PCR product whilst amplification of a vector minus an insert produces a 220bp product as

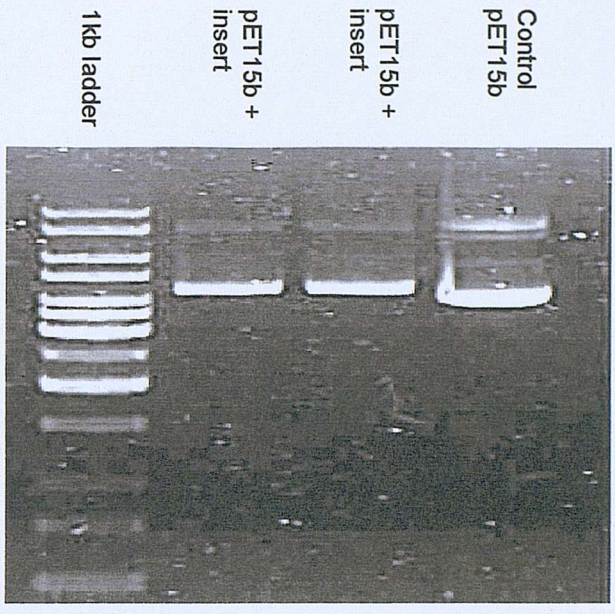


Figure 4.1. Agarose gel electrophoresis of plasmids with and without Endo VII insert. Plasmids from TG2 cells transfected with pETendoVII vector were extracted and run on a 0.6% agarose gel stained with ethidium bromide. These were run alongside a 1kb marker and a control pET15b without EndoVII insert.

Figure 4.2. Agarose gel of pET15b, pETendoVII and PCR fragments generated from cloning region of pET15b with or without EndoVII insert. Universal T7 primers used to amplify cloning region of pETendoVII and the cloning region of pET15b. The modified cloning region produced a 694bp product and the unmodified cloning region a 220bp fragment.

indicated by the blue arrows. The difference in product size is due to the presence or absence of the Endo VII gene as seen in figure 4.2. The relative size of the PCR amplicons was confirmed via comparison with the size standard.

4.2.2. Sequencing of pETEndoVII

The successful ligation of the Endo VII gene as indicated above was further verified by DNA sequencing of the cloning region of pET15b using the universal T7 primers (Fig. 4.3). Sequencing not only confirmed the presence of the Endo VII at position and orientation required but comparison of the sequenced gene and the published Endo VII genbank (X12629) sequence demonstrated 100% identity (Fig. 4.4). Sequence reactions were performed by Oswel DNA services, which is now part of the Eurogentec group. The successful sub cloning of the Endo VII gene and the verification that all the major features of the cloning site were all correct, permitted the next step of progression to expression and purification of the Endo VII protein.

4.2.3. Conditions for Growth of BL21 DE3 Transfected with pETEndoVII

For each purification a fresh transformation reaction had to be carried out because transformed cells kept on agar plates suffered from substantial loss of cells in a very short space of time (2-3days) and were unsuitable for culture and Endo VII expression. The expression host BL21 Gold DE3 has high expression upon induction but it is known to have "leaky expression", which can lead to low levels of un-induced expression. It was therefore necessary to determine to what extent leaky expression would affect the growth of the host cells.

As an indirect measurement of bacteria number the optical density can be measured at wavelength 600nm. The optical density measures the amount of light that can pass through a bacterial sample. The greater the $O.D_{600}$ the more bacterium there are. It was observed that $O.D_{600}$ values of BL21 Gold DE3 transfected with pETendoVII failed to increase beyond 1.4-1.8. Subsequently it was observed that cultures with an $O.D_{600}$ of greater

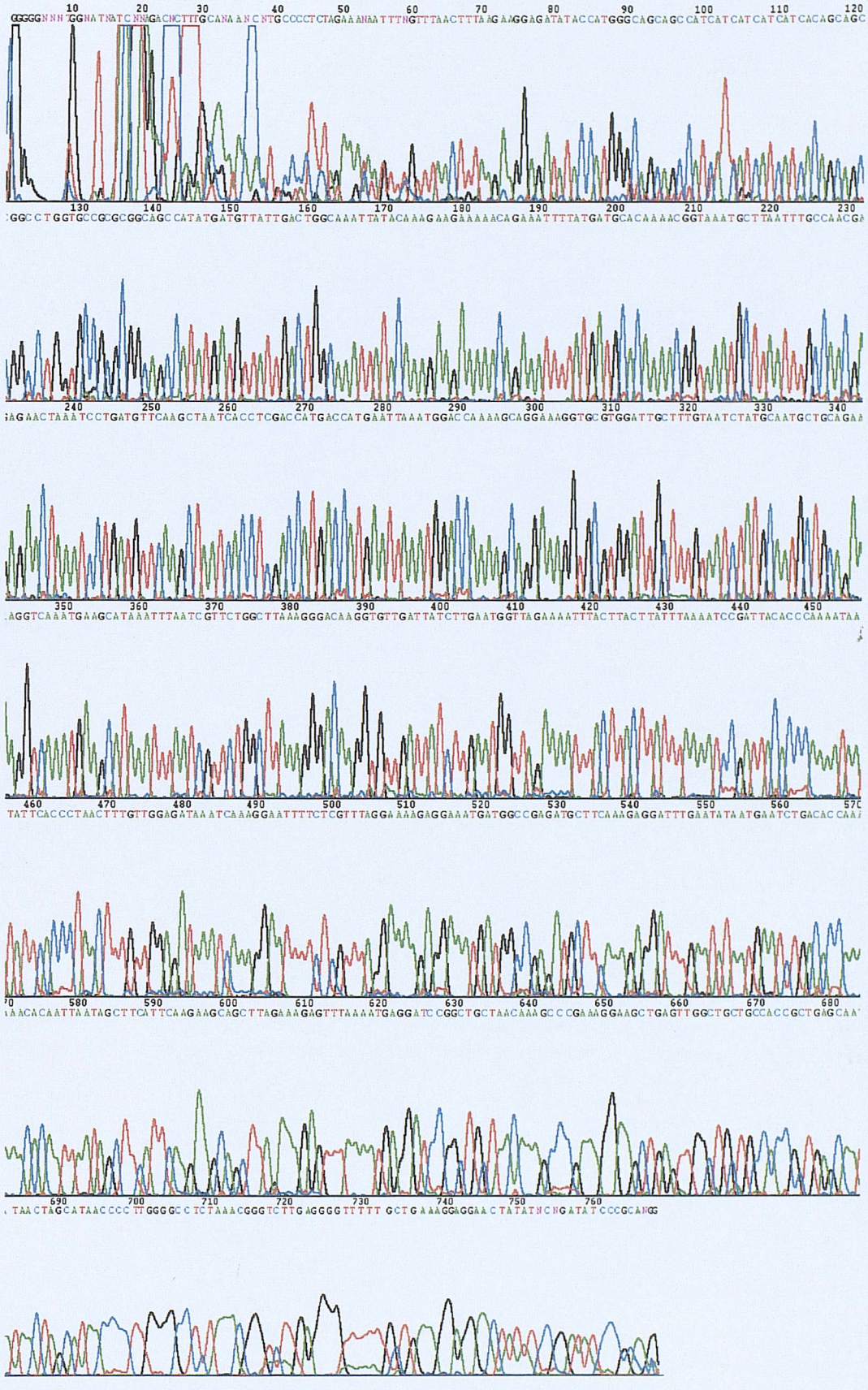


Figure 4.3. EndoVII sequence as supplied by Oswel DNA services.

GGGGGNNNTGGNATNATCNNAGACNCTTTGCANAANCNTGCCCCTCTAGAAANAATT

TNGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCA

CAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGATGTTATTGACTGGCAAATTATA

CAAAGAAGAAAAACAGAAATTTTATGATGCACAAAACGGTAAATGCTTAATTTGCCA

ACGAGAACTAAATCCTGATGTTCAAGCTAATCACCTCGACCATGACCATGAATTAAA

TGGACCAAAAGCAGGAAAGGTGCGTGGATTGCTTTGTAATCTATGCAATGCTGCAGA

AGGTCAAATGAAGCATAAATTTAATCGTTCTGGCTTAAAGGGACAAGGTGTTGATTA

TCTTGAATGGTTAGAAAATTTACTTACTTATTTAAAATCCGATTACACCCAAAATAA

TATTCACCCTAACTTTGTTGGAGATAAATCAAAGGAATTTTCTCGTTTAGGAAAAGA

GGAAATGATGGCCGAGATGCTTCAAAGAGGATTTGAATATAATGAATCTGACACCAA

AACACAATTAATAGCTTCATTCAAGAAGCAGCTTAGAAAGAGTTTAAAATGAGGATC

CGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAAT

AACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGCTGAAAGG

AGGAACTATATNCNGATATCCCGCANGG

= T4 endonuclease VII gene

= His-tag

= Restriction sites

= Translation start

= Ribosome binding site

Figure 4.4. Sequence of the pETEndoVII cloning region containing Endo VII gene. The key sequences highlighted according to the key given.

than 1.4 failed to produce fresh growth when used to inoculate fresh LB media supplemented with carbenicillin indicating that cell death had occurred. The question that arose was, was this due to the toxicity of Endo VII or had the culture reached the natural death phase where the build up of waste products and exhaustion of nutrients causes cell death.

The work performed using MutS meant that we had plasmid containing a protein that was known to be non-toxic to its host. This fact could be used to determine if cell death was arising from the presence of Endo VII. TG2 and BL21 DE3 Gold containing Taq MutS or Endo VII were used to inoculate 50ml LB media. These were grown under identical conditions. When the cultures had reached an $O.D_{600}$ of 1.4-1.8 a 1ml sample was taken and under went a plasmid extraction. Plasmids were successfully recovered from all the samples except for BL21 DE3 Gold transfected with pETendoVII. The presence of plasmid in the TG2 transfected with pETendoVII and of plasmid in the BL21 DE3 Gold transfected with pETaqMutS indicated that cell death in BL21 DE3 Gold containing pETendoVII was due to leaky expression of Endo VII protein. The Endo VII attacks both its own DNA and the pETendoVII vector causing cell death and the loss of plasmid. In the case of BL21 DE3 Gold transfected with pETendoVII successful recovery of pETendoVII was achieved up to an $O.D_{600}$ of 0.8 (Fig. 4.5). Further more cultures with an $O.D_{600}$ up to 0.8 were found to successfully inoculate fresh LB media.

It was observed that inoculating a 500ml LB culture with a single colony from a freshly plated transformation reaction and leaving overnight (O/N) at 30°C gave a satisfactory growth rate without leading to cell death. Smaller media volumes or higher temperatures produced to rapid growth overnight leading to cell death. This O/N culture was then used to inoculate fresh 500ml LB cultures. IPTG induction was carried out at 0.6 $O.D_{600}$ -0.8 $O.D_{600}$ and the success of expression was analysed using SDS-PAGE.

The addition of IPTG resulted in cessation of the increase in bacteria number as observed by monitoring the $O.D_{600}$. After induction the $O.D_{600}$ only increased by 0.1. This is as expected because the host cell directs all its energy to producing Endo VII. Secondly the high expression of a toxic protein would obviously be detrimental to the cell. This fact can act as an indicator that successful over expression is occurring.

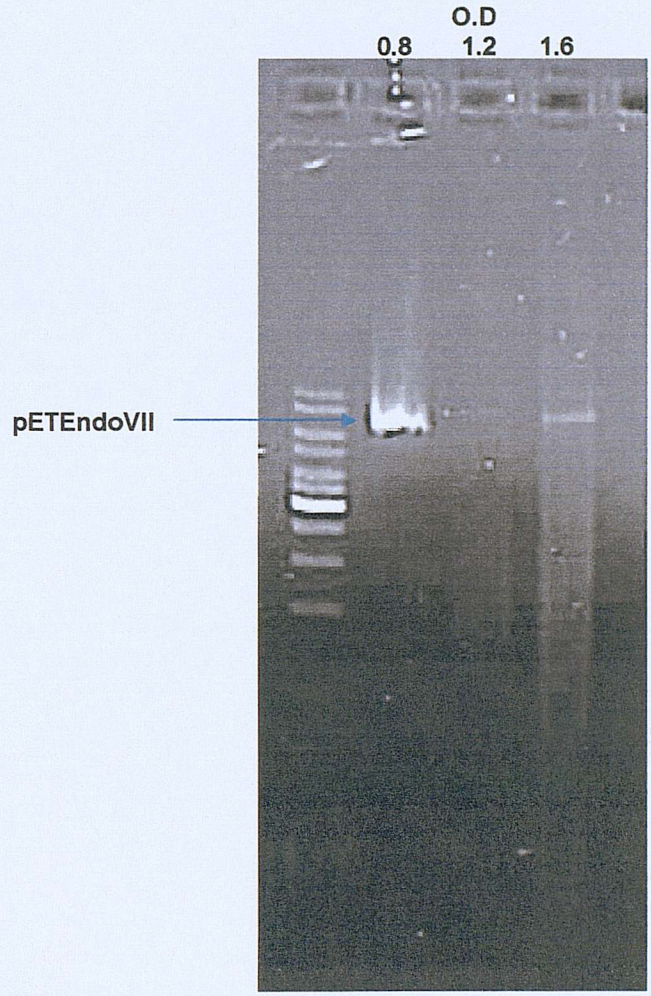


Figure 4.5. Toxicity of Endo VII. pETEndoVII plasmid extraction from inoculated media with an O.D of 0.8, 1.2 and 1.6. The 1.2, and 1.6 O.D samples show significant loss of a plasmid and smearing indicating digestion by Endo VII.

4.2.4. SDS PAGE Analysis of Endo VII Expression

BL21 (DE3) Gold were transfected with the plasmids pET15b (without EndoVII insert to act as a control) and pETEndoVII. Samples from pre-induced and 2hrs post induction were analysed for amount and identity of the over expressed protein (Fig. 4.6). In the strain containing the pETEndoVII plasmid an extra protein band appears post induction (4). This additional protein band is absent in the post-induced control culture and both pre induced samples. Comparison with the size standard indicates the size of the protein is between 15kDa and 25kDa. The size of the His-tagged Endo VII is approximately 20kDa. Therefore the size of this protein and its presence in only the post induced pETendoVII sample indicates successful expression of the Endo VII protein. Further observations indicated that post induction incubation need not be greater than 2hrs (Fig. 4.7). There appears to be little increase in Endo VII expression levels post induction between 2hr and 3hr incubation.

4.2.5. Purification of over Expressed Endo VII

Figure 4.8 compares the different chromatographs generated using Primeview software for a control purification and an Endo VII purification. The control purification differed from the Endo VII purification in only one single aspect and that was the vector transfected into the expression host lacked the Endo VII gene. Comparison of the two chromatographs (Fig. 4.8), reveals that both display a substantial amount of non-bound host proteins that come through in the initial wash stages of the purification. Upon addition of the imidazole gradient both display a peak of non-specifically bound proteins between fractions 6-8. A second peak is observed between fractions 17-19 in the pETEndoVII containing culture, which is absent in the control purification. SDS-PAGE of samples taken from these fractions indicates that this second peak corresponds to the elution of a high purity protein (Fig. 4.9). The size standard on the SDS gel places the single protein seen in fractions 17-19 between 15kDa and 25kDa. This corresponds to the approximate size of a His-tagged Endo VII protein. Fractions 6-8 did not contain a protein

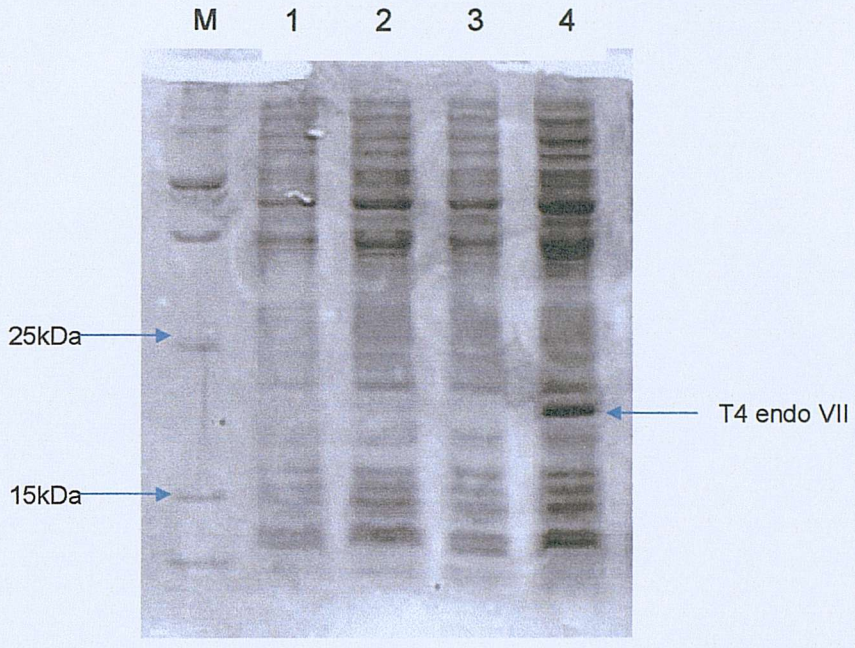


Figure 4.6. Analysis of EndoVII expression using SDS-PAGE. Samples were treated with SDS loading dye and run on a 15% SDS-PAGE. Lane 1 pre-induced BL21 Gold (DE3) + pET15b. Lane 2 2hr post induction BL21 Gold (DE3) + pET15b. Lane 3 pre induced BL21 Gold (DE3) + pETEndoVII. Lane 4 2hr post induction BL21 Gold (DE3) + pETEndoVII. (M) Molecular weight standard.

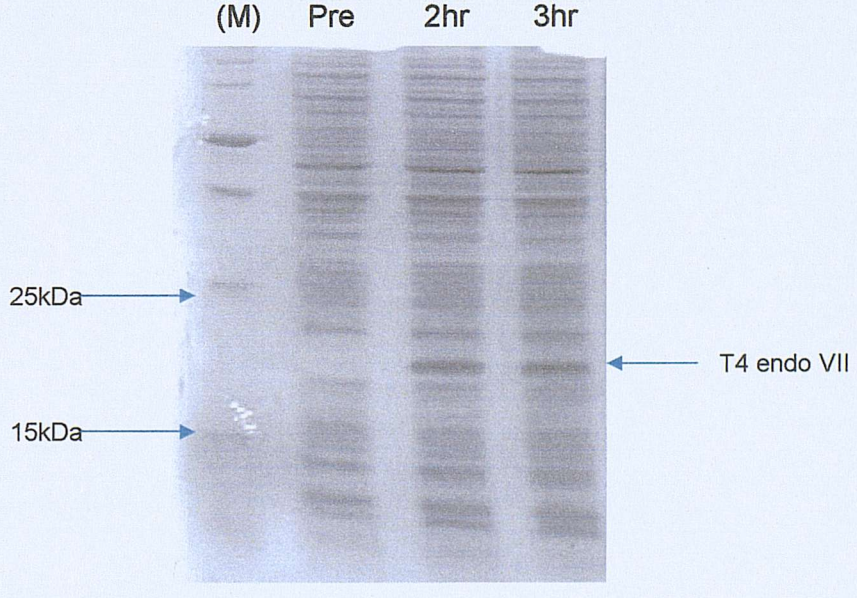


Figure 4.7. Analysis of expression levels post induction using SDS-PAGE. Starting from the left the samples are as follows, protein molecular weight marker, pre induced culture, EndoVII expression after 2hrs and EndoVII expression after 3hr.

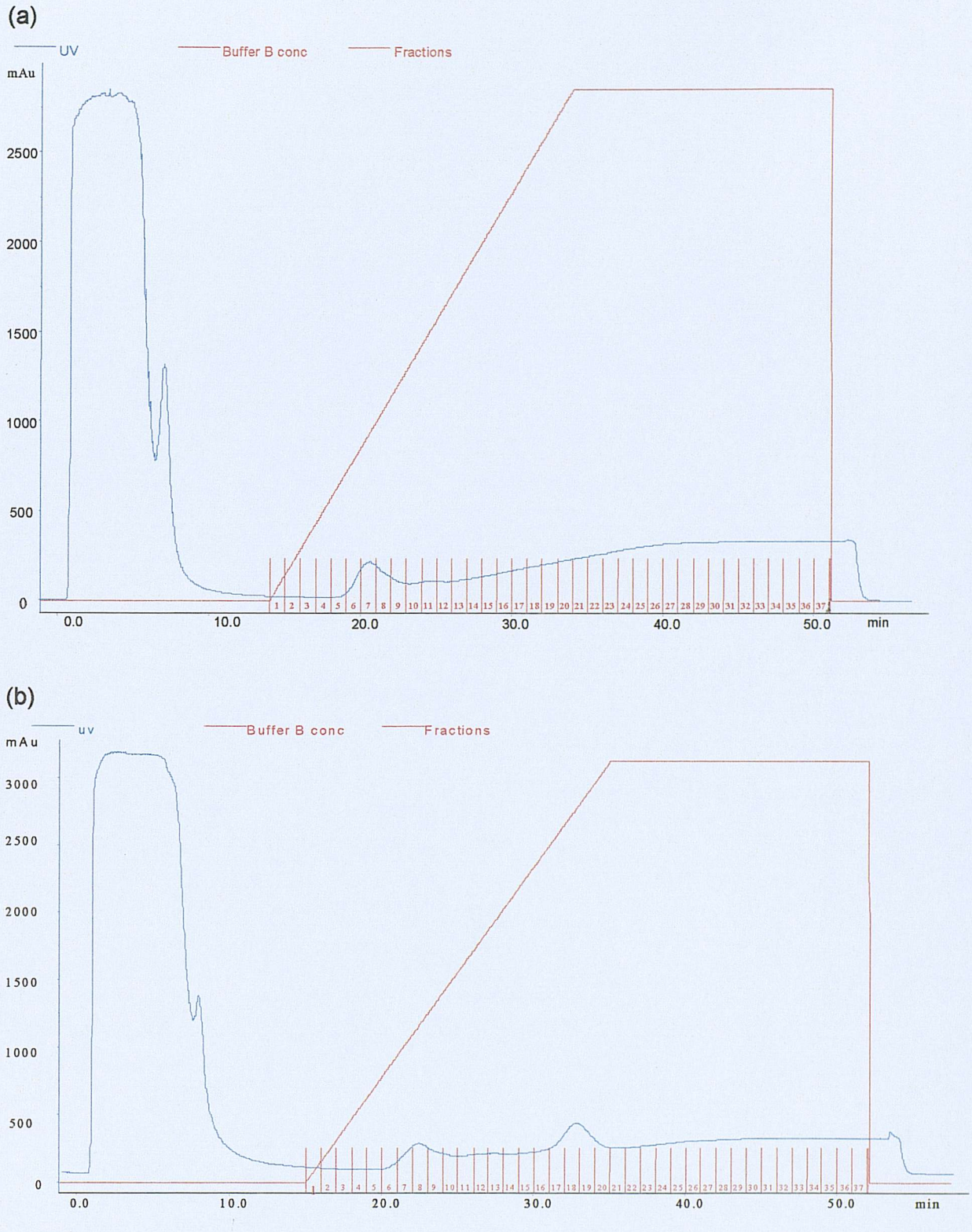


Figure 4.8. EndoVII purification. Chromatographs of purification of Bl21 (DE3) Gold transfected with (a) pET15b or (b) pETendoVII. Chromatographs produced on Primeview software (Amersham).

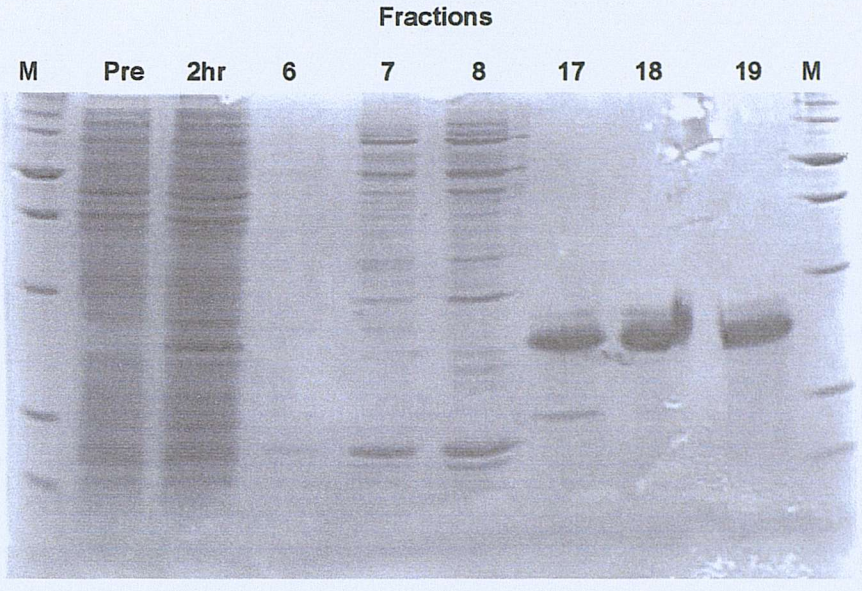


Figure 4.9. SDS-Page analyse of expression and purification of T4 endonuclease VII. The marker lane (M) contains a Biorad 100kDa protein molecular weight marker. Samples taken from fractions corresponding to the two peaks seen on the chromatograph (Figure 4.8b) after addition of the imidazole gradient for the strain containing the pETEndoVII plasmid.

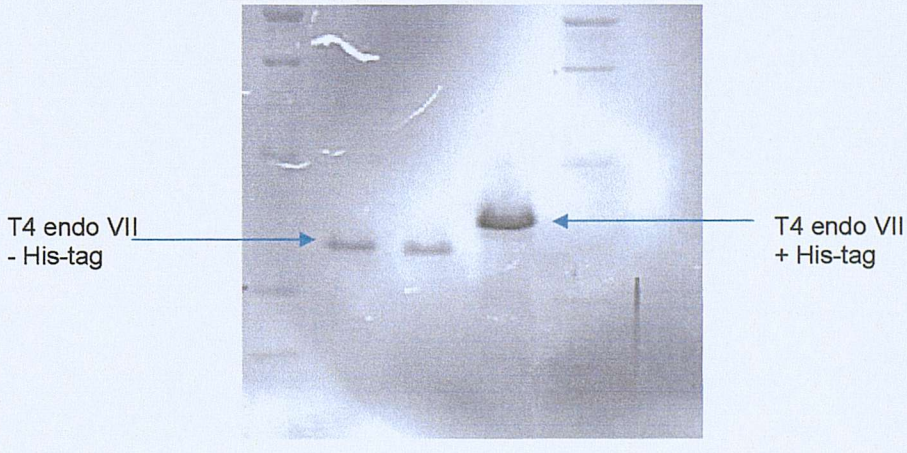


Figure 4.10. His-tag cleavage using thrombin. Endo VII treated with thrombin and untreated Endo VII analysed using a 15% SDS-PAGE.

band in the same place as the single band in fractions 17-19 indicating that these are weakly bound non specific proteins and no Endo VII has been lost in these fractions.

4.2.6. Effect of the His-tag on Endo VII activity

Figure 4.10 confirms the removal of the His-tag via thrombin digestion. The absence of the His-tag results in a protein with a lower molecular weight, which migrates further in the SDS gel than the corresponding His-tagged Endo VII.

To determine if the addition of a His-tag to the N-terminus of the Endo VII protein had a detrimental effect on the function of the Endo VII protein resolution of a three-way junction (3WJ) was investigated using the protein with or without a His-tag. Cleavage of the 3WJ using Endo VII with or with out a His-tag revealed very similar cleavage patterns when run on a non-denaturing PAGE (Fig. 4.11). As Endo VII concentration increases the undigested fragment diminished to be replaced further down the gel by cleavage fragments generated by the cleavage properties of Endo VII. At low concentrations of Endo VII two cleavage fragments can be observed, as Endo VII concentration increase the fragment closest to the undigested 3WJ disappears. This suggests that this fragment may represent an intermediate structure in the resolution of the 3WJ. As the His-tag did not affect the activity of Endo VII all subsequent experiments were performed using Endo VII enzyme with a His-tag.

4.2.7. Endo VII Cleavage of Synthetic DNA duplexes

A 50bp DNA duplex labelled on 5' sense strand with ^{32}P containing a single base mismatch (T•G) and a homoduplex with the correct base pairing were incubated with a range of Endo VII concentrations (1μM, 2.5 μM, 5 μM, 7.5μM 10 μM, 20 μM) to determine the specificity of Endo VII for heteroduplex DNA over the corresponding homoduplex. The cleavage fragments from the cleavage reaction were separated on a 15% denaturing PAGE. The results of the Endo VII cleavage are displayed in figure 4.12. Both heteroduplex and

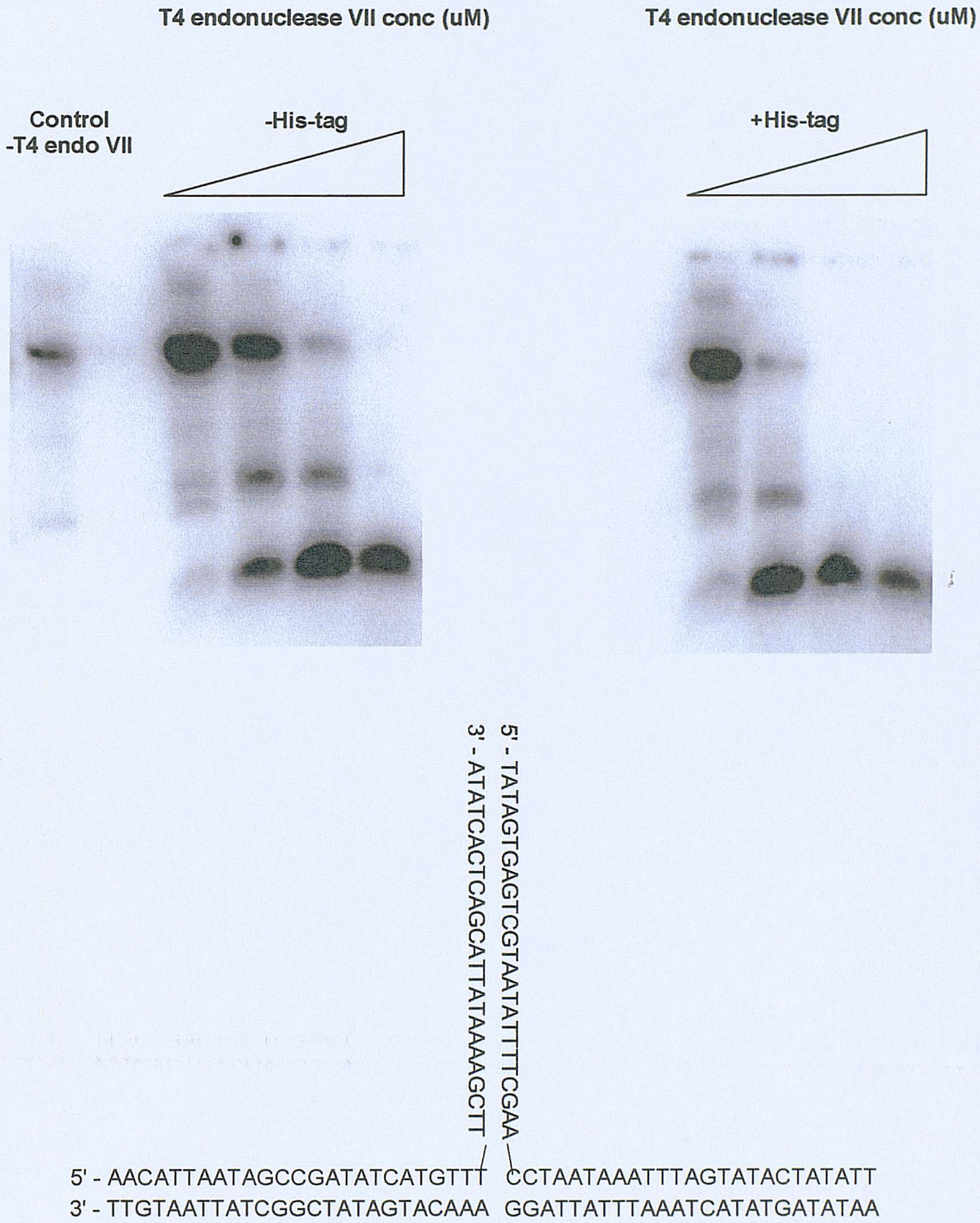


Figure 4.11. Resolution of a 3-way junction using recloned EndoVII with or without the His-tag. Stock EndoVII was diluted to the following concentrations 2.5μM, 5μM, 7.5μM and 10μM and incubated with the 3WJ. 3WJ was treated with a range of Endo VII concentrations and the cleavage fragments run on 10% non-denaturing polyacylamide gel.

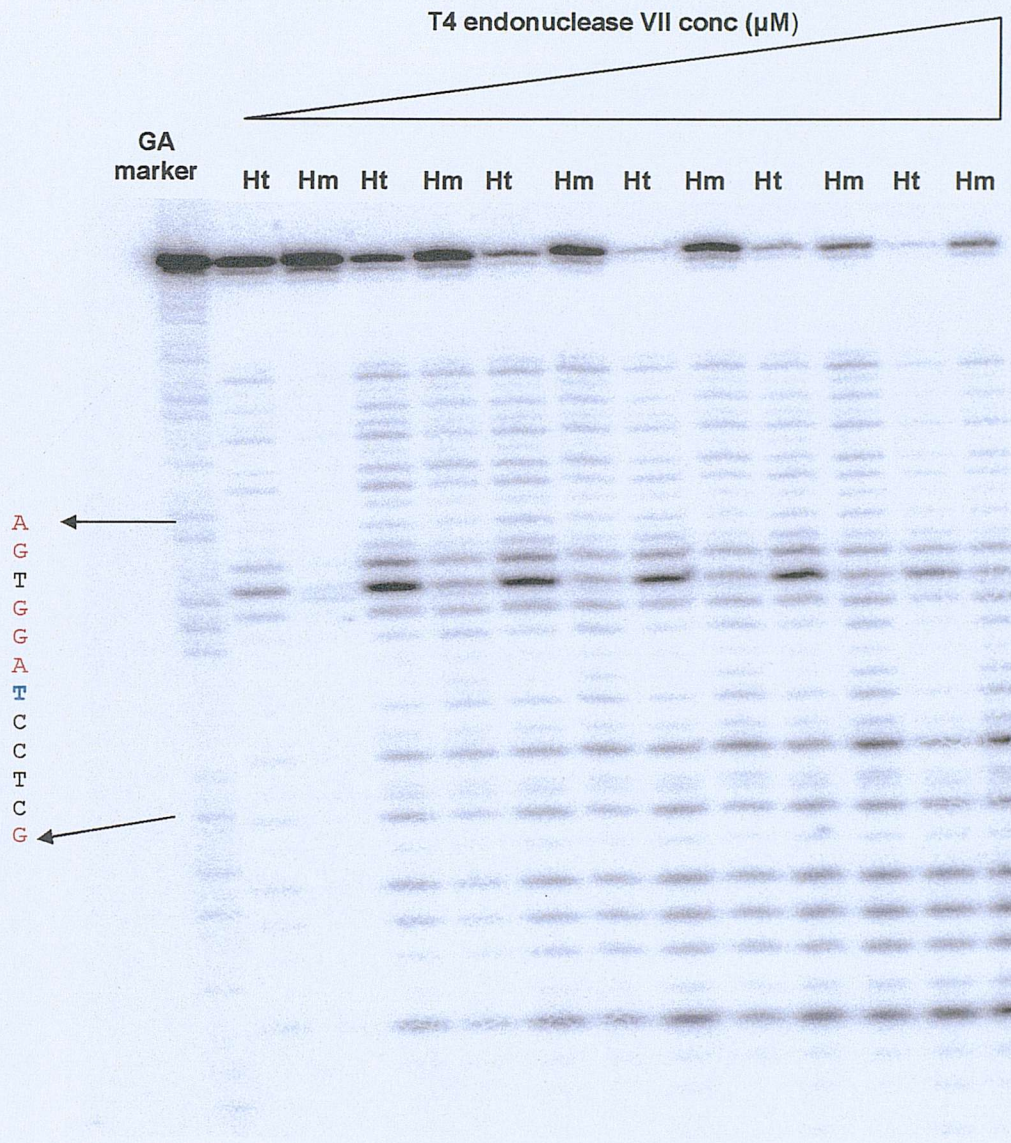
homoduplex DNA display non-specific background cleavage; however the heteroduplex DNA shows a distinctive cleavage fragment that is present as a faint band in all the homoduplex reactions. Comparison with the GA size ladder indicates that this fragment is the result of cleavage 3bp on the 3' side of the mismatch. This is consistent with previous work that has indicated that Endo VII cleaves from 2-6bp from the mismatch site on the 3' side of the distortion (Solaro et al., 1993). Therefore this fragment is the result of a specific cleavage by Endo VII due to the presence of a mismatch within the DNA.

As shown in figure 4.12 the concentration of Endo VII needs to be controlled. At the higher concentrations there appears to be secondary cleavage of the main cleavage fragment resulting in the loss of the cleavage fragment. It would appear that given excess enzyme any heteroduplex or homoduplex could be completely digested to very small fragments, obscuring any mismatch preferences of the Endo VII. This over digestion results in a decrease of the signal seen for the specific cleavage fragment (Fig 4.14 (A)). The optimum Endo VII concentration was 5μM. The fraction of the total DNA in the heteroduplex sample present as this specific cleavage fragment ranged from 0.05-0.2 depending on the concentration of Endo VII. In comparison the values determined for the homoduplex sample ranged from 0.01-0.03.

4.2.8. T4 endonuclease VII cleavage of all Eight Different Mismatches

To determine the mismatch specificity of Endo VII, Endo VII (5μM) was incubated with identical 50mers each containing one of the eight possible mismatches at the same location. The cleavage fragments were run on a 15% denaturing gel alongside a homoduplex control and a GA marker (Fig. 4.13). All eight mismatches displayed specific cleavage by Endo VII at the mismatch site but with variability in the relative cleavage specificity.

Comparison of the cleavage pattern of the duplex DNA containing a mismatch with that of the homoduplex demonstrates that Endo VII recognised all eight mismatches over Watson & Crick base pairing producing a specific cleavage fragment. Quantitative analysis of the cleavage fragment intensities revealed that the fraction of the total DNA present as the cleavage fragment



5'-*CTGACGATGTGAACTATGACGCTCCTAGGCTGATTACTGACTGCTGGACT
GACTGCTACACTTGATACTGCGAGGGTCCGACTAATGACTGACGACCTGA

Figure 4.12. Cleavage of heteroduplex (Ht) DNA with a single T•G mismatch and cleavage of homoduplex (Hm) DNA by Endo VII. Stock Endo VII protein was diluted to the following concentrations 1μM, 2.5 μM, 5 μM, 7.5μM 10 μM, 20 μM and incubated with the synthetic duplexes as described in Chp 2 pg 99. Each substrate is identical except for the single base mismatch. The sequence of the mismatch duplex is shown below the gel image with A/G corresponding to the GA marker highlighted in red.

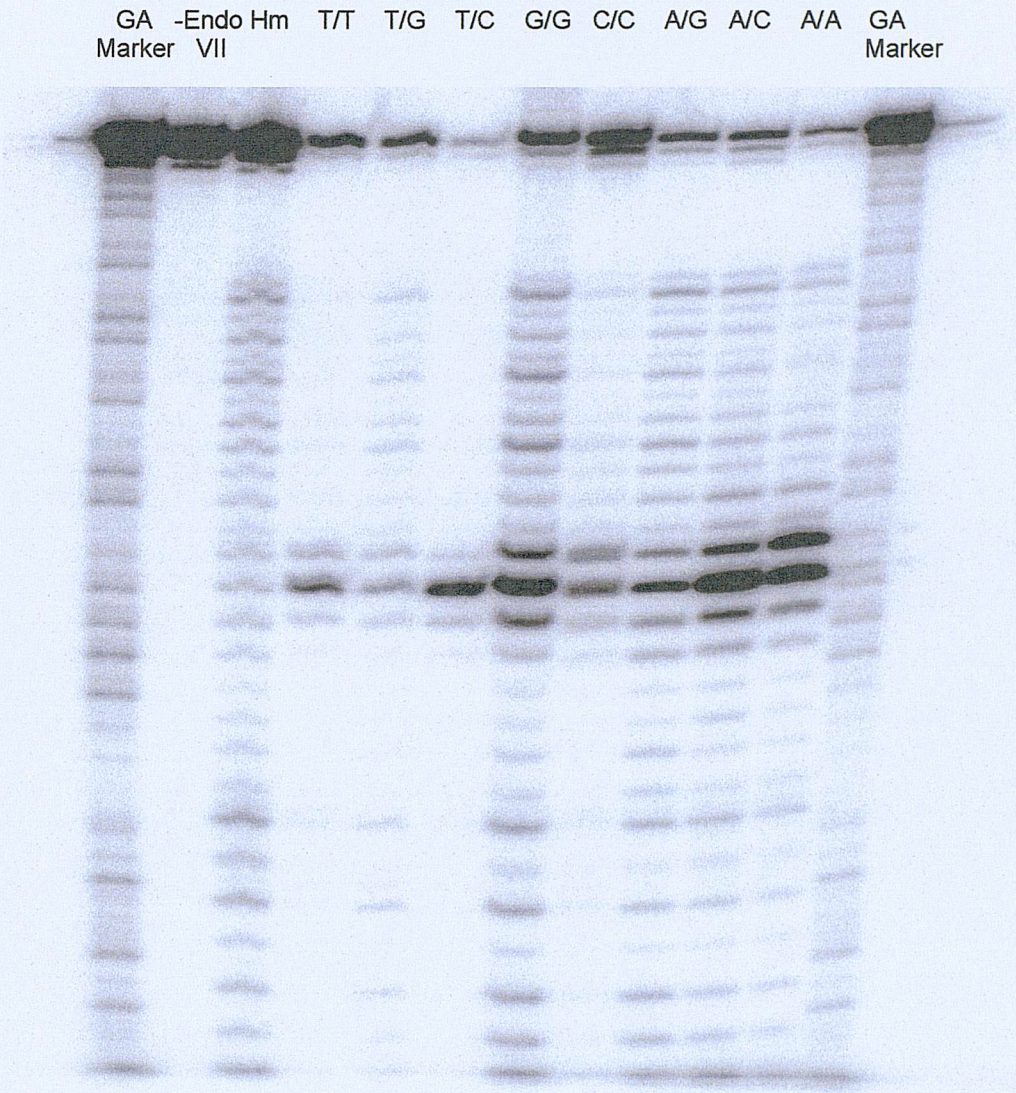
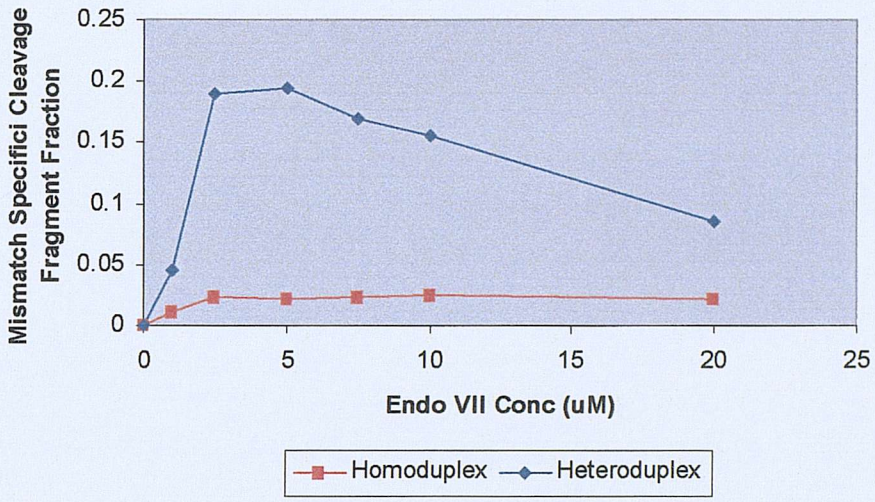


Figure 4.13. Mismatch cleavage of Endo VII. All eight mismatches and a homoduplex control were treated with Endo VII. The EndoVII concentration used was 5μM and was based upon titration experiments.

(A)



(B)

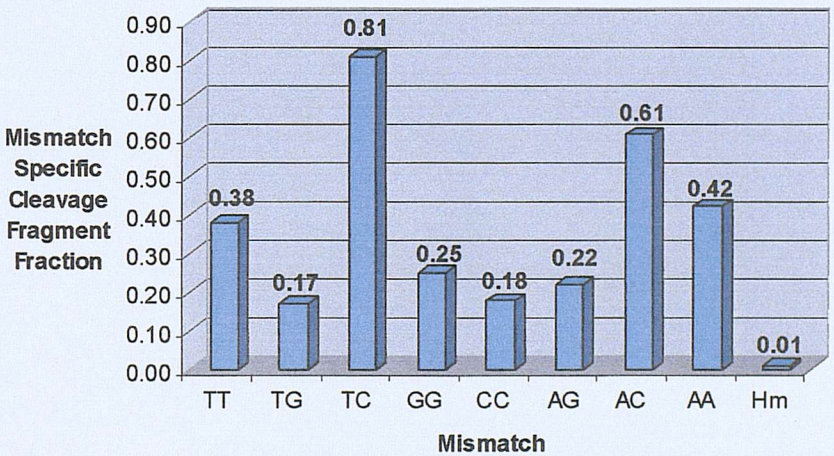


Figure 4.14. Quantitative data analysis of the main cleavage fragment for (A) G/T mismatch compared to Hm control and (B) all eight mismatches using Phoretix software.

ranged from approximately 0.17-0.70 depending on the mismatch. Based upon the values obtained from the best experiment the order of efficiency of cleavage was T•C <A•C <T•T <A•A <G•G <T•G <A•G <C•C but even the lowest fraction of cleavage was at least 10 fold greater than that seen for the equivalent homoduplex. The mismatches can be broadly categorised into three sub groups. The least efficiently recognised C•C, A•G, and T•G. The medium recognised G•G, A•A, T•T and the most efficiently recognised A•C and T•C.

4.2.9. Resolution of Cleavage Fragments using Fluorescent tags and Short Track Electrophoresis

Figure 4.15 shows the T•C synthetic duplex and a T•A homoduplex with a sense fluorescently labelled 5' end treated with an increasing concentration of Endo VII. The concentration of Endo VII used was 2.5μM, 5μM, 10μM and 20μM. A sample without any Endo VII was used as a control to demonstrate that the bands were the result of Endo VII cleavage and not an artefact in the duplex sample. The results of this cleavage were run on a 15% denaturing H-PAGE. A specific cleavage fragment for the duplex containing a mispaired T•C was clearly visible and separate from the undigested duplex. The FAM labelled oligonucleotide used to form the various duplex DNA migrated the same distance as the sample not treated with Endo VII. This indicated that the gel was fully denaturing, as a 50bp duplex would have a different mobility when compared to a 50 base oligonucleotide. As the untreated sample and oligonucleotide migrated the same distance the sample has been denatured and is in single stranded form.

Traces that plotted counts 10^3 against pixel position generated using phoretix clearly demonstrate the differing affects Endo VII had on the T•C mismatch duplex and the corresponding homoduplex. The first peak marked by the blue arrow 1 corresponds to undigested DNA, the second blue arrow 2 indicates a mismatch specific cleavage fragment while the third arrow seen in some of the traces indicates background fragments not related to the specific cleavage of Endo VII. In the case of heteroduplex DNA as the first peak reduces there is an increase in the second peak this occurs up to 10μM where

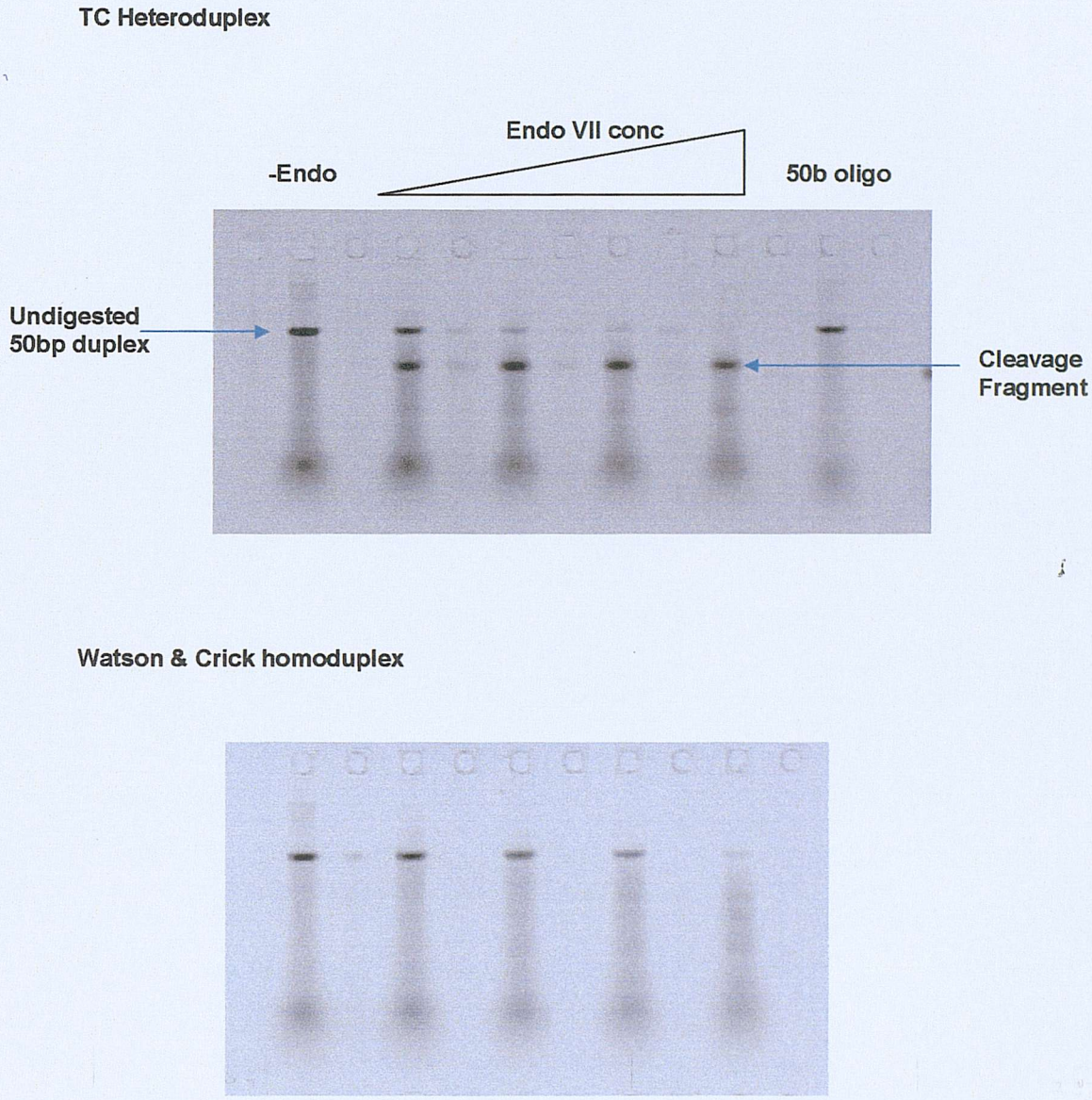


Figure 4.15. Mismatch Cleavage of synthetic T•C heteroduplex and Watson & Crick homoduplex with a range of Endo VII concentrations (2.5μM, 5μM, 10μM and 20μM). Cleavage reactions run on a 15% denaturing polyacylamide H-PAGE.

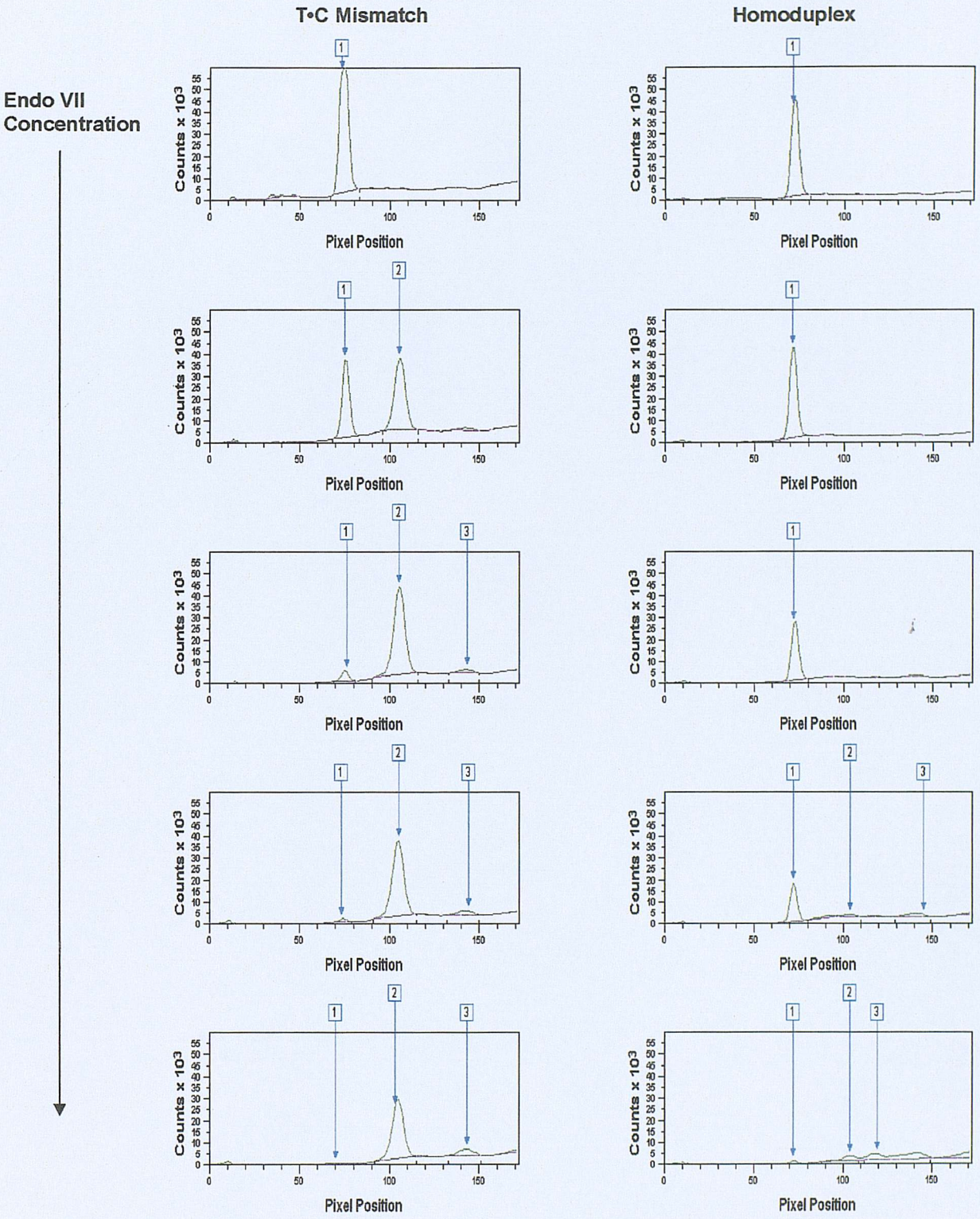


Figure 4.16. Phoretix analysis of cleavage fragments from Endo VII cleavage of T•C heteroduplex and the corresponding homoduplex. For the heteroduplex the arrows indicate (1) undigested duplex, (2) cleavage fragment, (3) background cleavage fragment. For the homoduplex sample the arrows indicate (1) undigested duplex, (2) & (3) background cleavage.

there has been complete digestion of the initial heteroduplex. At this concentration there is a reduction in cleavage peak (2) and the formation of a third peak (3) that was not visible before this point. In contrast the homoduplex sample shows very little background cleavage and maintains an undigested peak up to 20μM.

4.2.10. Reaction Parameters

A number of parameters were investigated that could affect the relative cleavage efficiency of Endo VII. These factors include incubation time, Mg^{2+} concentration and incubation temperature. To investigate their effects and determine the optimum conditions for Endo VII cleavage synthetic duplexes containing either a T•C , T•G mismatch or the correct base pairing were digested under a variety of conditions by Endo VII (5μM). The T•C and T•G mismatch represent both ends of Endo VII specificity spectrum. For the more poorly recognised T•G mismatch it is essential that the conditions be optimised so that we achieve the maximum amount of mismatch specific cleavage. This will improve the detection of specific cleavage fragments when used for mutation detection.

4.2.10.1. Temperature

The incubation of Endo VII with synthetic duplex was carried out at 20°C, 37°C and 55°C and the results are shown in figure 4.17. All temperatures produced a cleavage fragment for the TG and TC mismatched duplex and only weak non-specific background cleavage for the homoduplex. The change in incubation temperature has little effect on the efficiency of Endo VII activity with the proportion of DNA present as the cleavage fragment similar at all temperatures. Interestingly when the reaction mix was pre heated to the desired reaction temperature at 55°C the activity of Endo VII was abolished. This indicates that the majority of Endo VII cleavage must occur rapidly before the reaction has reached the incubation temperature or we would expect a reduction in Endo VII activity in the 55°C incubation

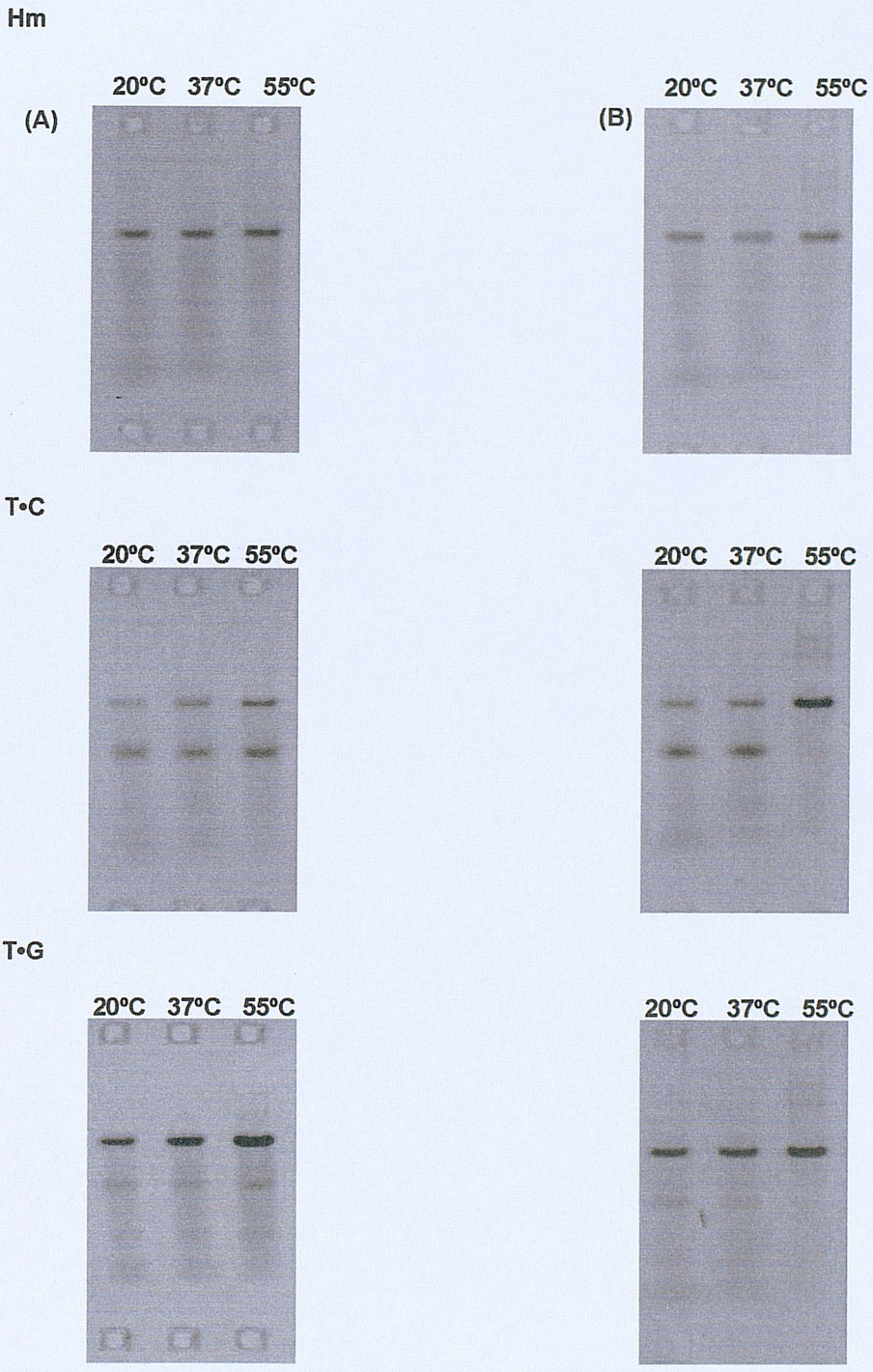


Figure 4.17. Cleavage reaction temperature. Endo VII cleavage of Watson & Crick homoduplex, T•C heteroduplex and T•G heteroduplex at 20°C, 37°C and 55°C. (A) Endo VII (5μM) added to reaction mix then placed at reaction temperature (B) Endo VII (5μM) added to pre heated reaction mix.

sample that has not been preheated.

4.2.10.2. Incubation Time

The 50bp synthetic duplex was incubated with Endo VII for various lengths of time at 37°C before the addition of the stop solution, which terminated the cleavage reaction. Phoretix analysis allowed the intensities of the undigested duplex and specific cleavage fragment to be plotted against time as seen in figure 4.18. For both the T•C and T•G heteroduplex DNA and the homoduplex the quantity of DNA remaining undigested decreased as reaction time increased. This reduction occurred over the 60 minute time period although the majority of Endo VII cleavage appears to occur within the first 20 minutes. From the data it is apparent that the optimum reaction time is between 10-20 minutes. After 20 minutes the formation of a specific cleavage fragment has plateaued.

4.2.10.3. $MgCl_2$ Concentration

The activity of Endo VII is dependent upon the presence of metal ions. The metal ion used was Mg^{2+}. Endo VII and the synthetic duplexes were incubated with reaction buffer that contained a range of Mg^{2+} from 0 to 100mM (Fig. 4.19). The concentration of Mg^{2+} that produced the cleavage fragment with the highest intensity as determined by phoretix for both the T•C and T•G mismatched duplex was 5mM &10mM. At the higher Mg^{2+} concentration (50mM & 100mM) the proportion of duplex present as the specific cleavage fragment was greatly reduced and in the case of 100mM was not present. This indicated that too greater a metal ion concentration has an inhibitory effect on the activity of Endo VII. It was found that when Mg^{2+} was removed from the reaction buffer there was still significant enzyme activity.

4.2.11. EndoVII Sensitivity

To investigate the proportion of heteroduplex DNA in a DNA sample

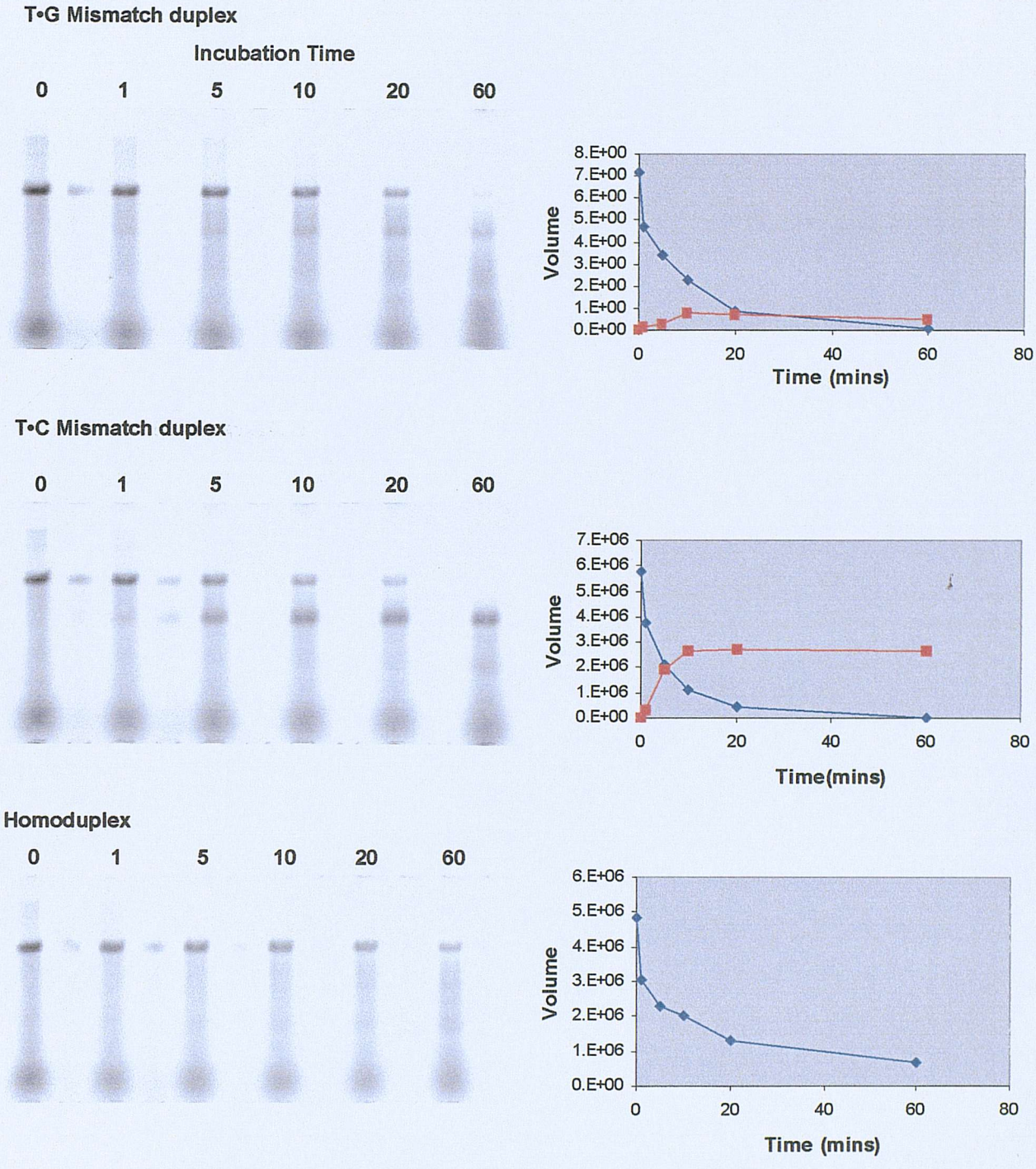


Figure 4.18. Endo VII time titration. T•C heteroduplex, T•G heteroduplex and corresponding homoduplex were subjected to Endo VII cleavage for varying reaction times ranging 1minute to 60minutes. Phoretix analysis was used to determine the volume of the cleavage fragments and this was plotted against reaction time. Both the volume of the undigested fragment (blue) and the volume of the cleavage fragment (red) were plotted for the T•C and T•G heteroduplex. For the homoduplex, only the undigested duplex was plotted (blue).

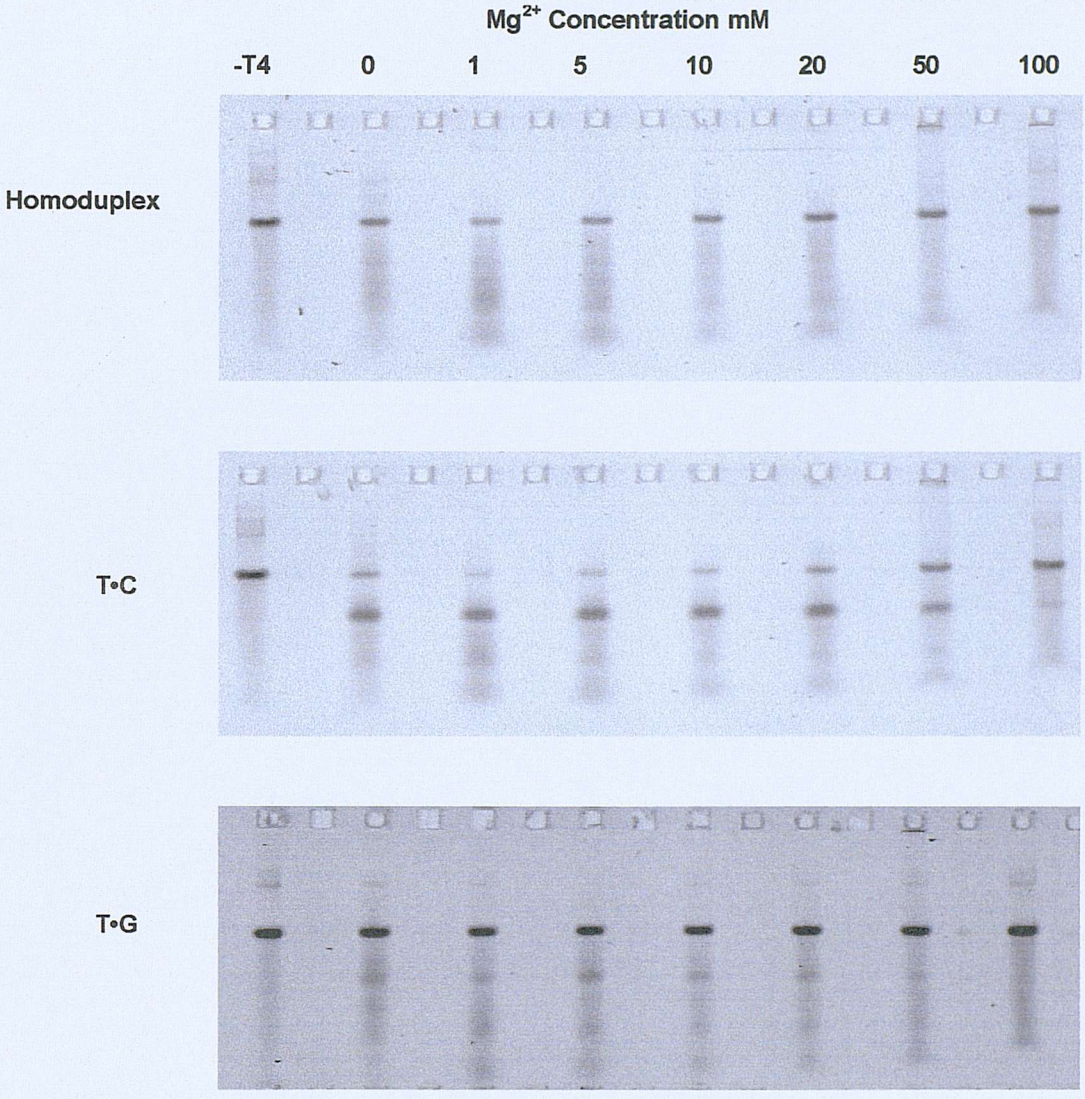


Figure 4.19. Endo VII reaction mix $MgCl_2$ titration. Watson & Crick homoduplex, T•C heteroduplex and T•G heteroduplex were incubated with Endo VII (5μM) in reaction buffers to give a final concentration of 0, 1mM, 5mM, 10mM, 20mM, 50mM and 100mM $MgCl_2$.

needed to produce a detectable mismatch specific cleavage fragment when compared to the same point in the homoduplex control, heteroduplex DNA was mixed with homoduplex DNA to produce DNA samples with a varying percentage of heteroduplex DNA. The percentage of T•C or T•G heteroduplex in the sample ranged from 5% to 50%. The resulting cleavage patterns of Endo VII digest of these samples is shown in Fig 4.20.

Endo VII digestion of the sample containing only homoduplex provided the background cleavage fragment pattern. In the case of the T•G heteroduplex a discernable cleavage fragment was seen with approximately 30% heteroduplex. This fragment was absent from the control homoduplex sample. In contrast the T•C heteroduplex produced a visible cleavage fragment at approximately 10% (Fig 4.20). Comparison of the phoretix traces for the T•G and T•C show that the sample containing 30% T•G displayed very similar cleavage to the 10% T•C sample (Fig 4.21).

4.3. Discussion

4.3.1 Endo VII Purification

A high throughput mutation detection based upon mismatch cleavage by Endo VII requires a source of active Endo VII. Although previously available as part of Amersham's Pharmacia Biotech Passport kit, Endo VII is no longer commercially available. The unavailability and potential costs of purchasing Endo VII had it been available for large-scale mutation screening meant that it was necessary for the Endo VII protein to be produced in house. The purification of Endo VII was initially based upon two chromatography steps and produced approximately 13mg of protein from 25g of induced cells routinely harvested from an 8 litres of inoculated media which took 5 days (Golz et al., 1995). An adaptation to this technique has been the fusion of a His-tag to pET11a creating a derivative pET11a-his (Birkenbihl and Kemper, 1998). Endo VII was purified under denaturing conditions on Ni-NTA-spin columns (Quigen). This was used for small-scale protein preparations of up 150µg of protein. The recloning of the EndoVII gene into pET15b removed the need to fuse an oligohistidine to the Endo VII gene prior to ligation of the

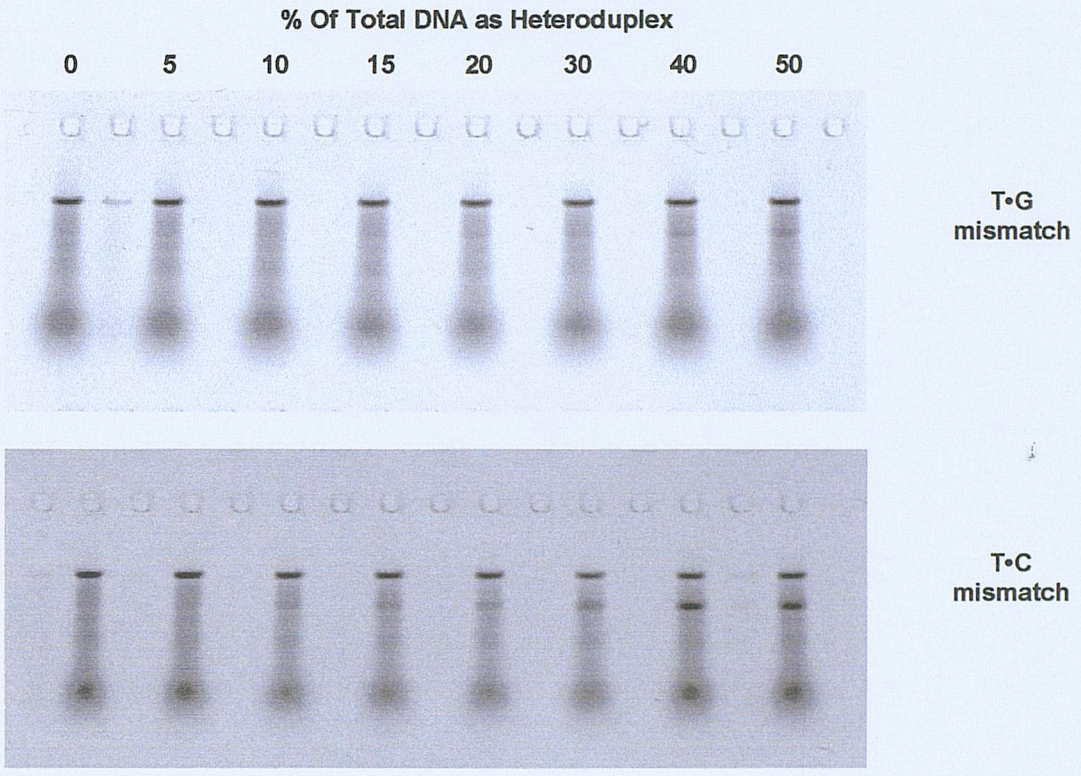


Figure 4.20. EndoVII cleavage sensitivity. Endo VII cleavage of heteroduplex DNA in a pool of heteroduplex and homoduplex DNA. Heteroduplex DNA increases as a percentage of the total amount of DNA present from left to right. The first sample contains only a 50 bp synthetic homoduplex treated with Endo VII. As the percentage of heteroduplex increases a mismatch specific cleavage fragment appears for both the T•C and T•G heteroduplex.

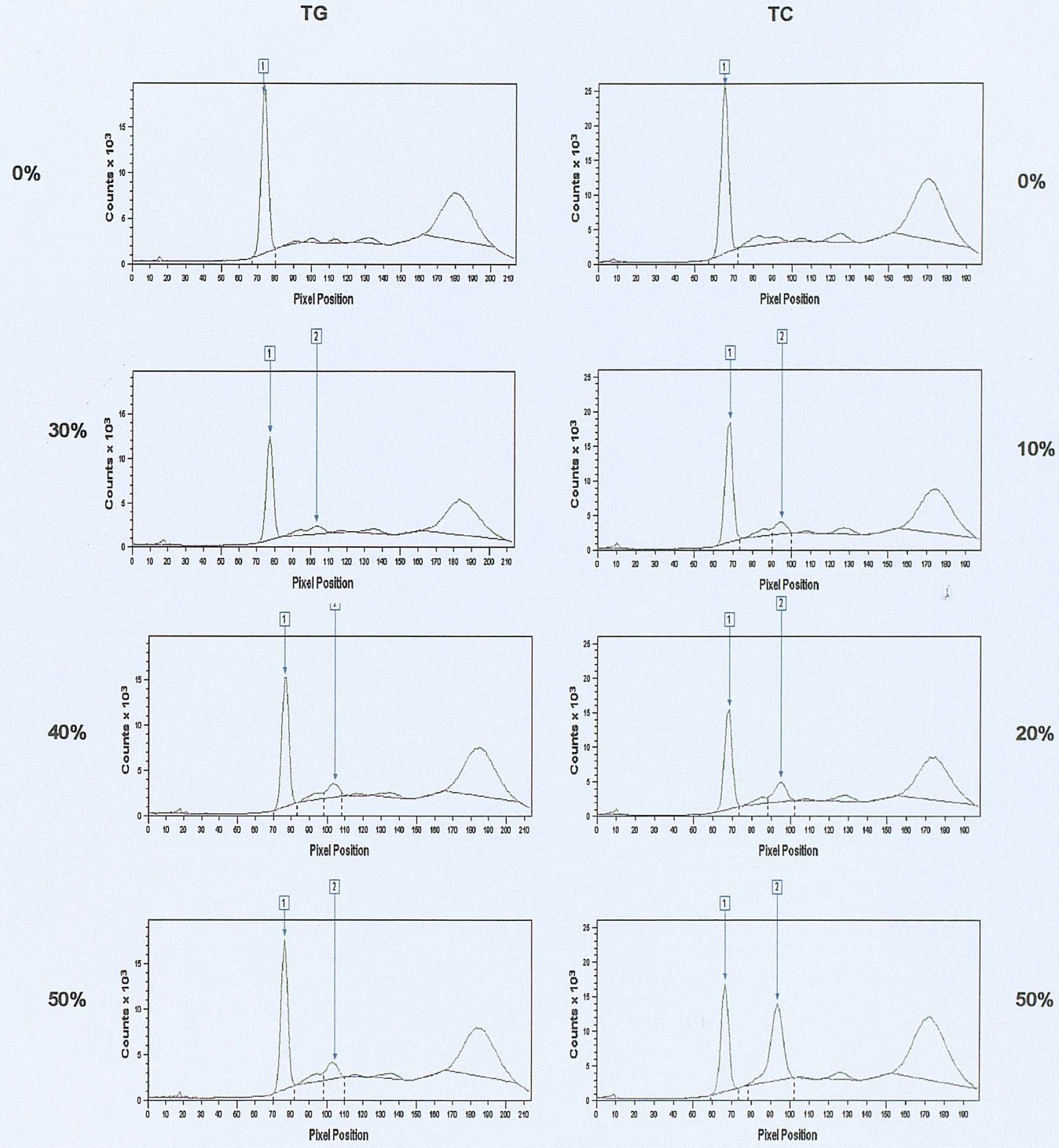


Figure 4.21. Phoretix analysis of the cleavage pattern for selected lanes from Endo VII sensitivity assay (Fig 4.20). Selected lanes for T•G heteroduplex include 100% homoduplex, 40% T•G and 50% T•G. Undigested duplex is indicated by arrow 1 and the cleavage fragment by arrow 2. Selected lanes for T•C heteroduplex include 100% homoduplex, 10% T•C, 20% T•C and 50% T•C. Undigested duplex is indicated by arrow 1 and the cleavage fragment by arrow 2

gene into the vector as pET15b has been constructed with a His-tag sequence in the cloning region. The His-tag attached to the N-terminal of the Endo VII protein allowed purification via ion exchange chromatography. Purification was achieved using Hitrap columns charged with Ni ions. Following sample loading the AKTAprime FLPC purification is fully automated and subjects the bound protein to a controlled gradient of increasing imidazole concentration. The purification from sample loading to elution of pure protein takes approximately 1½ hours. It was not necessary to denature the protein prior to loading thus reducing column washes. Inclusion of 10mM imidazole in the initial wash buffer and the application of a gradient of imidazole removed weakly bound proteins resulting in the elution of highly pure protein at around 500mM imidazole. The whole process from inoculation of overnight culture to dialysed protein routinely took two days, with a 1litre media yielding approximately 5mg of protein. This compares favourably to the original two-step purification where 13mg was purified from 8litres of media in five days. A one-step purification also reduces the manipulation of EndoVII thus helping to maintain a high activity of the protein. Further manipulation of the protein was avoided with the observation that the presence of the His-tag did not have a detrimental affect on Endo VII activity. Based upon cleavage assays this quantity of enzyme would be enough for 30,000 cleavage reactions. Endo VII stored in 50% glycerol at -20°C is still active after 6 months and Endo VII stored at -80°C has successfully been used after 12 months. The quantity of Endo VII produced was restricted by the available equipment and can easily be scaled up to produce higher Endo VII yields.

4.3.2. Bacteria Expression Host.

The pETendoVII plasmids were established in a non-expression host TG2. As TG2 lack the T7 RNA polymerase gene required to express the Endo VII gene they are a good way of obtaining a sufficient quantity of pETendoVII for storage and future purification. TG2 could be easily grown without being affected by the toxicity of Endo VII and the pETendoVII plasmid extracted and used to transfect the expression strain BL21 (DE3) Gold.

The λDE3 lysogen expression hosts provide a choice of three types of

expression host. Although the standard λDE3 lysogen can display leaky expression it was chosen as the expression host. More stringent control can be achieved using pLysS or pLysE strains of the λDE3 lysogens. The draw back of these more stringent strains is that they often lead to a reduction in expression levels. BL21-gold derived host cells also contain the Hte phenotype that results in 100-fold increase in transformation efficiency as well as the inactivation of the endonuclease I gene that rapidly degrades plasmid DNA from mini-preps. This not only helps produce high quality miniprep DNA but allows direct cloning of many constructs.

It became apparent that due to the toxicity of Endo VII it was essential to ensure that growth of the culture was closely monitored. This ensured that prior to induction there were sufficient quantities of cells available for satisfactory protein expression. By growing the cultures initially at a slow rate a simple control can be introduced that enables sufficient growth of the bacteria without excessive loss of cells and enables the higher yielding standard λDE3 lysogen to be used.

4.3.3. Endo VII Cleavage

When using Endo VII it is necessary not to over digest the substrate. It has been reported that a digestion of 10-50% is sufficient to detect mutations in the fragment (Babon et al., 1999). As seen in figure 4.12 using higher concentrations of Endo VII results in a loss of specific cleavage fragment. Given enough enzyme the substrate would be completely digested into very small fragments. The loss of cleavage fragment is likely to be the result of secondary cleavage of the cleavage fragment.

Endo VII recognises all eight mismatches with at least a 10 fold greater preferences over the homoduplex. This was determined by calculating the intensity of the mismatch specific cleavage fragment as a fraction of the total intensity in the lane. This was achieved using Phoretix software that can detect changes in signal produced by the DNA labelling from gel images and attribute an arbitrary unit to the these values. This allowed the mismatch specificity of Endo VII to be investigated, as the more efficiently recognised

mismatches would produce a stronger signal for the mismatch cleavage fragment translating into a larger fraction value.

The mismatch cleavage efficiency of Endo VII has been classed as low-efficiency, intermediate-efficiency and high-efficiency (Solaro et al., 1993). The relative cleavage efficiency of Endo VII for mismatch DNA was calculated by comparing the degradation of a substrate using limiting enzyme concentration and comparing this to the degradation of an 8nt heteroduplex loop, which was set to 100%. In their study all G-mismatches were classed as low efficiency (10%), A•A, A•C, C•A, T•C and T•T were classed as intermediate-efficiency (30%) with C•C and C•T as high efficiency (50%). A direct comparison between these values and the values obtained in this study cannot be made. However based indirectly upon their relative cleavage efficiencies the order of EndoVII mismatch specificity from worst to best was T•G/A•G/G•A/G•G< A•A/A•C/C•A/T•C/T•T< C•C/C•T. In all but one of the cases mismatches containing the same mispair but on different strands fell within the same specificity. Only T•C and C•T fell into different subgroups.

Endo VII mismatch specificity for the mismatches tested in this study from weakest to strongest was C•C<A•G<T•G<G•G<A•A<T•T<A•C<T•C. The major difference in mismatch specificity presented in this chapter is that of the C•C mismatch which displayed strong cleavage in the published data but weak cleavage in this study. All the other mismatches fell broadly into the same specificity, with the mismatches containing a mispaired G producing a weaker cleavage signal than the mismatches without a G. Mispaired Gs offer higher melting stability especially mispaired T•G which is the most thermally stable of all the mismatches. This could reduce the distortion induced by the presence of mispaired bases in the DNA molecule, giving rise to their lower cleavage specificity. C•C has a relatively low melting stability therefore would be expected to have a strong cleavage similar to T•C. The difference seen here and that of the published data could be due to sequence context.

4.3.4. Resolution of Cleavage Fragments using short track electrophoresis.

The successful cleavage of a radio labelled short synthetic duplex containing mispaired bases using conventional polyacylamide denaturing gels

demonstrated that the recloned Endo VII had the specific activity expected. This cleavage of DNA and resolution of cleavage fragments needed to be adapted for short track electrophoresis. The two major factors that would govern the transferral of this technique to short track electrophoresis were

1) The ability to resolve cleavage fragments from back ground cleavage and from undigested duplex over a significantly reduced track length. Reduction of track length would condense the background potentially obscuring specific cleavage fragments.
2) The loss of sensitivity that would result in replacing the radiolabel with an alternative label.

The removal of the radiolabel is desirable for obvious health reasons and for offering simplified working conditions. The other alternatives to radio labelling were to use a fluorescent end label or DNA staining dyes. Although both would result in a loss of sensitivity in comparison to a radiolabel, initial studies indicated that the cleavage was robust enough to deal with this loss of sensitivity. Although not presented in this thesis detection of cleavage fragments was carried out using Vista Green and SYBR Gold staining dyes. One of the advantages of Endo VII mutation detection is that it can give localisation information. The indiscriminate nature of DNA staining dyes means that this localisation information would be lost because it would be impossible to distinguish which bands where the result of nicking of the sense or antisense strand. It was observed that using DNA staining dyes also increased the background noise making fragment identification more difficult as specific cleavage fragments were masked by the background noise. Therefore the labelling of DNA with a fluorescent label located on the 5' end of DNA was chosen.

The analysis of Endo VII cleavage using fluorescently labelled DNA has been applied to PCR products using an ABI prism 377 (Babon *et al.*, 1999;Inganas et al., 2000). Here we successfully used short track electrophoresis with fluorescently labelled synthetic duplexes to detect cleavage fragments and using a FluorImager to image the gel. The short

tracks offered suitable resolution to separate a 50 base undigested oligonucleotide from a 25 base cleavage fragment. The non-specific background cleavage did not mask the specific cleavage fragment. In comparison with the radio labelled sample it can clearly be seen that the shorter track lengths and reduced sensitivity of the label reduce the number of non-specific or secondary cleavage sites that are observed. The loss of sensitivity by replacing the radiolabel with a fluorophore rather than being detrimental in fact helped by removing the weaker non-specific bands making band calling easier. The use of such gels offers a considerable reduction in preparation and run times. Subsequently a variety of reaction parameters were investigated to determine optimum reaction conditions.

4.3.5. Sensitivity of Endo VII Cleavage Assay

The detection of mutations will be based upon detection of heteroduplexes in a sample that in many cases contains both heteroduplex and homoduplex. This reduction in the available heteroduplex could have a bearing on the detection of a specific cleavage fragment. As the proportion of heteroduplex decreases the proportion of DNA as a specific cleavage fragment will also decrease, this results in a loss of signal. As homoduplex increases the homoduplex contamination and therefore background cleavage increases. This could have the potential effect of masking specific cleavage fragments at low heteroduplex levels.

The proportion of heteroduplex that produced a detectable specific cleavage fragment depended upon the mismatch. The weakly cleaved T•G mismatch required at least 30% heteroduplex for a detectable cleavage fragment to be observed. Conversely at the other end of the mismatch specificity spectrum only 10% T•C heteroduplex was required to produce detectable specific cleavage fragment from amongst the background cleavage.

A mutation detection assay based upon heteroduplex formation would result in the formation of the mismatch species listed below (Tab. 4.1). Based upon the mismatch specificity and sensitivity of EndoVII it can be seen

in the majority of cases a weaker cleavage, those mispairs containing a G has a strong cleavage alternative, those containing a mispaired C. This suggests that at least one of the mismatch pairs would produce a sufficient enough cleavage fragment for mutation detection.

Allele 1	Allele 2	Mismatch	Class
G•C	C•G	C•C G•G	Transversion
C•G	A•T	C•T A•G	Transversion
C•G	T•A	C•A T•G	Transition
G•C	A•T	G•T A•C	Transition
G•C	T•A	G•A T•C	Transversion
A•T	T•A	A•A T•T	Transversion

Table 4.1. Mismatch base pairs formed as the result of a transversion or transition.

CHAPTER 5
APPLICATION OF ENDO VII TO MUTATION DETECTION

5.1. Introduction

The successful cleavage of synthetic heteroduplexes and the confirmation of a broad mismatch specific range for the re-cloned EndoVII lead to the application of Endo VII to mutation detection. Endo VII cleavage was used to detect a number of known base alterations from short PCR products amplified from a variety of genes. This would assess the sensitivity, specificity and positional accuracy of Endo VII cleavage for different mutations. This chapter examines the ability of Endo VII to detect mutations using an ABI 310 PRISM Genetic Analyser as a detection platform and MADGE with a Molecular Dynamics FluorImager 595.

5.1.1. Experimental strategy

Mutation detection using EndoVII requires the formation of heteroduplex DNA. This was achieved using a wildtype probe that was mixed with an excess of unlabelled test DNA. The excess of unlabelled strand drives the formation of duplexes between the probe and the sample DNA rather than reannealing of the probe or sample forming the original homoduplexes. As the probe is the only species labelled the only visible DNA is that formed between the probe and test DNA. The excess test that cannot anneal to the probe does not create background, as it is not visible.

The identity of the visible duplexes formed between the probe and the test DNA depends upon the genotype of the sample. A homozygote wildtype will form homoduplex DNA when hybridised to the probe. A heterozygote will form 50% homoduplex DNA and 50% heteroduplex. A homozygote mutant will form only heteroduplex DNA. The identity of the mismatches in the heteroduplex depends upon the sequence difference between the two alleles. The addition of different fluorophores at either end of the probe allows the two strands to be visualised independently. As a sense labelled strand would form a different heteroduplex to the anti-sense labelled strand a double-labelled probe would form two heteroduplexes species offering double the chance of detecting the mutation. As explained in chapter 4 this will form a weak mismatch specific base pair and a stronger mismatch specific base pair. This

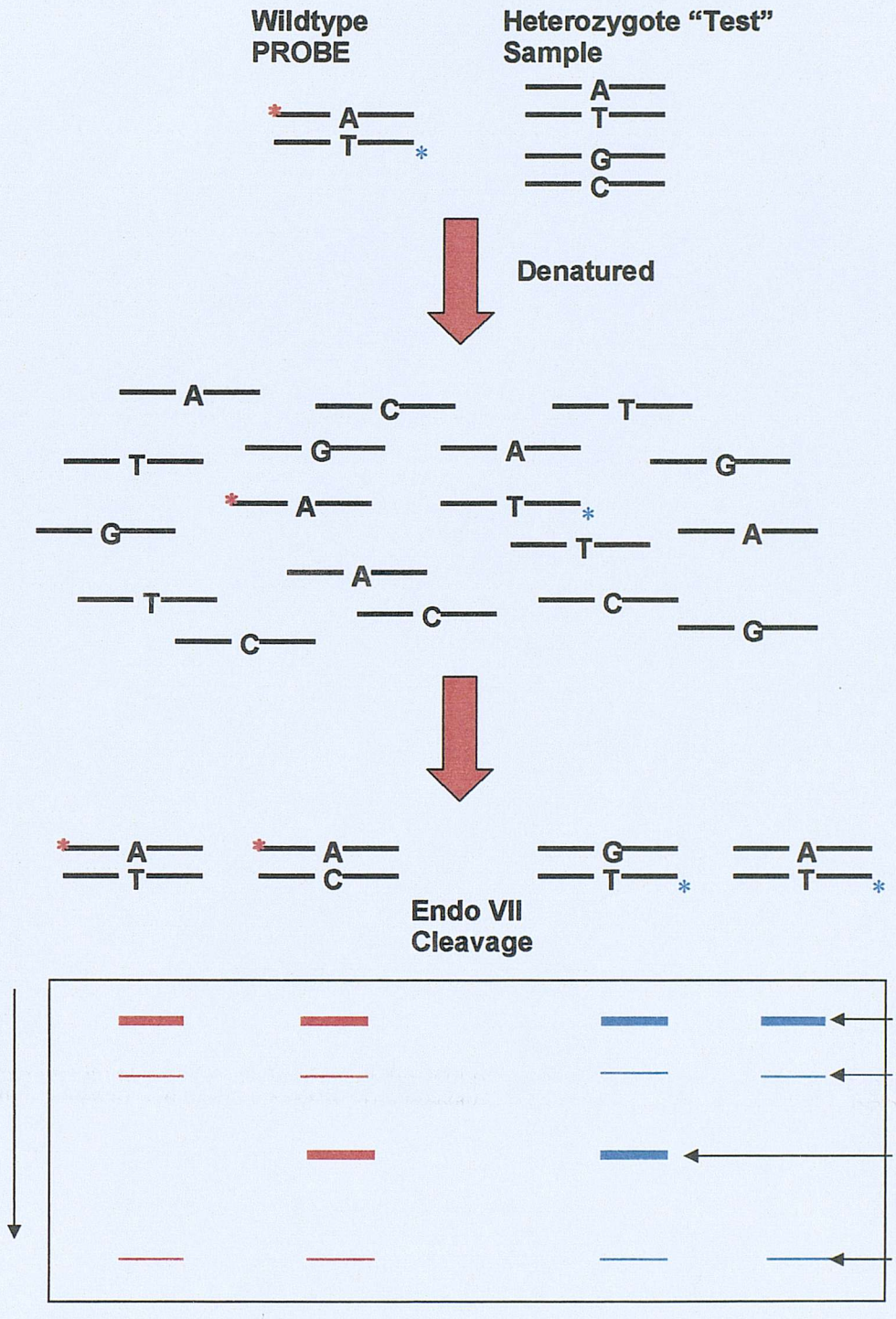


Figure 5.1. Experimental strategy to generate homoduplex and heteroduplex DNA for the detection of mutations in PCR amplicons utilising the mismatch cleavage enzyme T4 endonuclease VII.

is then subjected to Endo VII cleavage where the presence of a heteroduplex would produce a specific cleavage fragment (Fig. 5.1). For all the cleavage reactions the concentration of Endo VII added to the reaction was 5μM.

5.2. Results

5.2.1. *BRCA1* exon 11 Amplicon

A PCR amplicon was generated from a region of exon11 of the *Homo* sapiens breast and ovarian cancer susceptibility gene *BRCA1*, which contains a common polymorphism. The polymorphism, a T/C base substitution, occurs at position 2430 based upon the genbank sequence U14680. Two primer sets were used to generate PCR amplicons. A 260 base pair amplicon containing the polymorphic site situated 32 bases from the 5' end of the amplicon (BRCA1-F & BRCA1-R) and a 281 base pair amplicon where the base change was located centrally at nucleotide position 149 (BRCA1-281F & BRCA1-281R). In house control DNA samples were used to test the mutation detection of Endo VII. The DNA samples were typed as homologous (T/T), heterologous (T/C) or homologous (C/C) using ARMS (Table 5.1). The type of mismatch present after heteroduplex formation depended upon the choice of homozygote used for generation of the probe. The identities and relative portion of mismatch DNA using a T•A or C•G probe are listed in table 5.2.

5.2.1.1. Endo VII Cleavage of 260bp *BRCA1* Exon 11 Region Amplicon using C•G probe.

A 260 bp sense and antisense 5' FAM labelled probe was generated using the C/C homozygote as the template, this meant the probe DNA would contain a C•G base pairing at the polymorphic site. This was mixed with an excess of unlabelled PCR amplicon. Mixing the probe with a sample PCR that contained two C alleles generated a sample containing only homoduplex DNA. A second PCR amplicon containing two T alleles generated a sample containing only visible heteroduplex DNA. The identity of the mismatch that would be visible depended upon which strand the label was located. For the 260bp sense labelled reaction the mismatch formed would be C*•A whilst in

TGAAAGATCT GTAGAGAGTA GCAGTATTTC A**T**TGGTACCT GGTACTGATT ATGGCACTCA
GGAAAGTATC TCGTTACTGG AAGTTAGCAC TCTAGGGAAG GCAAAAACAG AACCAAATAA
ATGTGTGAGT CAGTGTGCAG CATTTGAAAA CCCCAAGGGA CTAATTCATG GTTGTTCCAA
AGATAATAGA AATGACACAG AAGGCTTTAA GTATCCATTG GGACATGAAG TTAACCACAG
TCGGGAAACA AGCATAGAAA

281bp

TGTCAATCC TAGCCTTCCA AGAGAAGAAA AAGAAGAGAA ACTAGAAACA GTTAAAGTGT
CTAATAATGC TGAAGACCCC AAAGATCTCA TGTTAAGTGG AGAAAGGGTT TTGCAAACTG
AAAGATCTGT AGAGAGTAGC AGTATTTCA**T** TGGTACCTGG TACTGATTAT GGCACTCAGG
AAAGTATCTC GTTACTGGAA GTTAGCACTC TAGGGAAGGC AAAAACAGAA CCAAATAAAT
GTGTGAGTCA GTGTGCAGCA TTTGAAAACC CCAAGGGACT AA

Figure 5.2. Sequence of PCR amplified regions of the *BRCA1* gene exon 11 using BRCA1-F/BRCA1-R (260bp) or BRAC1-281F/BRCA1-281R (280bp). The base highlighted by the bold font indicates polymorphic site.

SAMPLE	ALLELE
C23	T/T
C24	C/C
C25	T/C
C26	T/C
C27	C/C
C28	C/C
C29	T/C

Table 5.1. Identity of base at polymorphic site for the DNA samples as determined using ARMS assay

	T/T PROBE		C/C PROBE	
GENEOTYPE	SENSE STRAND LABELLED	ANTISENSE STRAND LABELLED	SENSE STRAND LABELLED	ANTISENSE STRAND LABELLED
T/T	T*•A (100%)	T•A* (100%)	C*•A (100%)	T•G* (100%)
T/C	T*•A (50%) T*•G (50%)	T•A*(50%) C•A* (50%)	C*•A (50%) C*•G (50%)	T•G*(50%) C•G* (50%)
C/C	T*•G (100%)	C•A* (100%)	C*•G (100%)	C•G* (100%)

Table 5.2. Duplex species formed upon hybridisation of sample DNA with probe generated from a T/T probe or C/C probe. The visible species observed depends upon the strand with the fluorescent label.

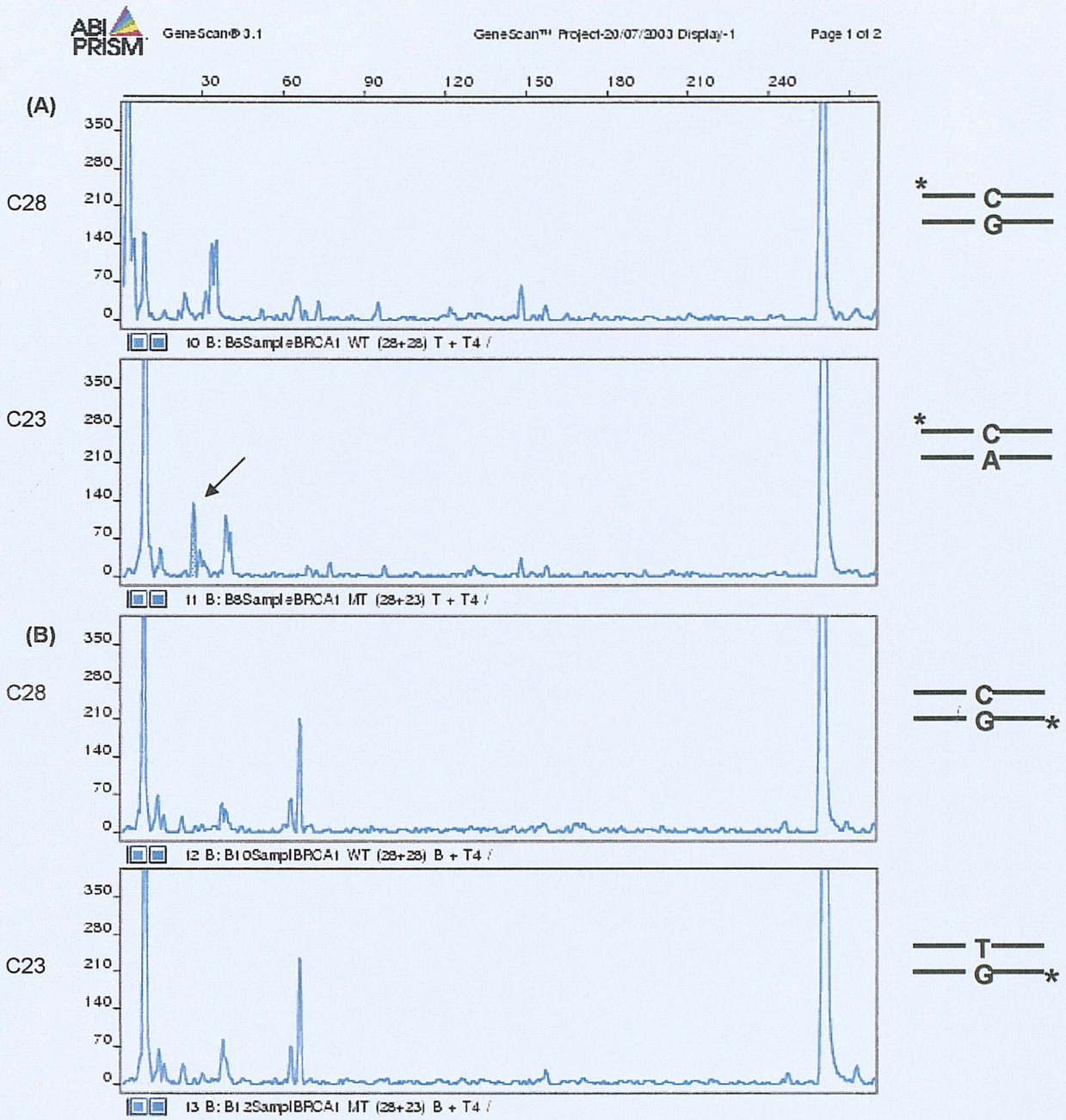


Figure 5.3. Comparison of the cleavage pattern generated by Endo VII cleavage of the 260bp *BRCA1* amplicon hybridised with C/C probe with a sense or anti sense FAM tag. The identity of the duplex that is formed and can be observed is indicated to the right of the ABI traces. (A) Endo VII cleavage performed on a homoduplex control and a heteroduplex containing A*•C mismatch labelled on the sense strand. An arrow indicates specific mismatch cleavage fragments. (B) Repeat of (A) save for the positioning of the FAM label on the antisense strand. Of the two strands only the sense strand displayed a detectable cleavage fragment.

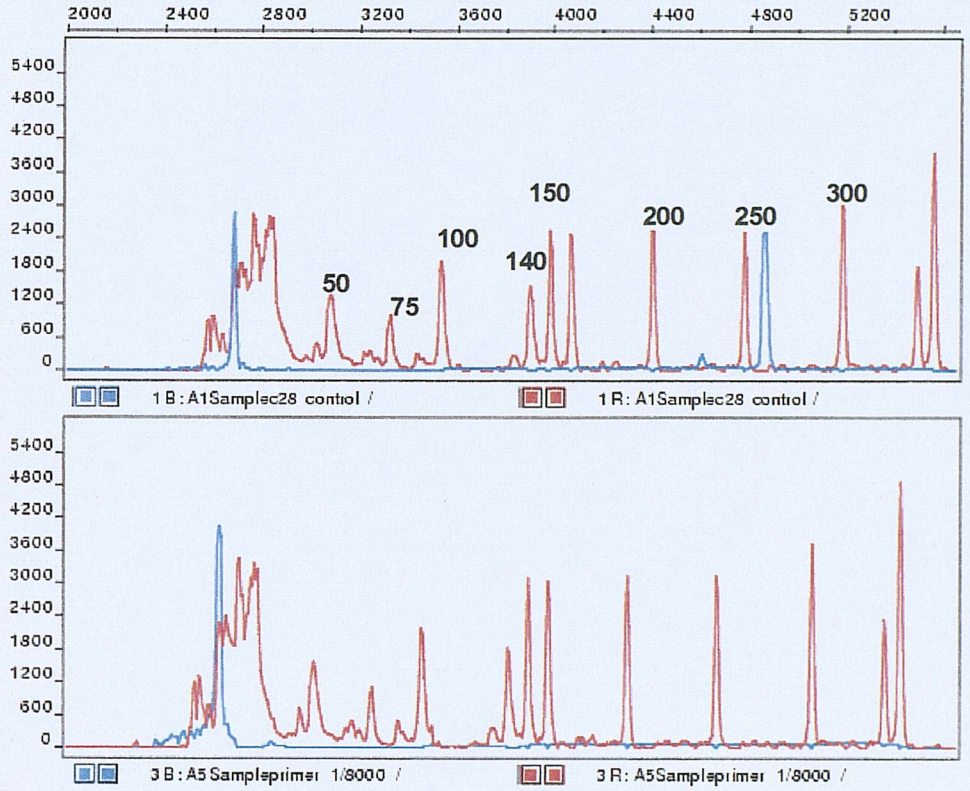


Figure 5.4. 260bp *BRCA1* amplicon (blue) and Genetix genSIZE TAMRA 50-350 labelled marker (red). (B) BRCA1-F 5' FAM primer (blue) run with a genSIZE T50-350 marker (red).

the antisense reaction the mismatch would be T•G*. The asterisk denotes that the base is present in the labelled strand. Comparison of the cleavage pattern using sense labelled probe for the wildtype and mutant sample displayed a unique cleavage fragment in the mutant sample that was not present in the wildtype sample as indicated by the arrow. This fragment was approximately 30 bases in length. The polymorphism occurs at position 32 within the amplicon. Based upon the EndoVII nicking the DNA 2-6bp on the 3' side of the distortion this would generate a fragment ranging from 33-38 bases in length. Although the observed fragment in the mutant sample is slightly smaller than expected inaccuracies in the standard can be used to explain this fractional difference.

To demonstrate these inaccuracies a 260bp sense labelled *BRCA1* amplicon was analysed on the ABI 310 with an internal size standard (Fig 5.4). In a separate experiment the BRCA1-F primer used to generate the 260bp sense labelled amplicon was analysed with the same internal standard. The internal standard as shown in figure 5.4 displays a peak of known size (red). The PCR amplified fragment (blue) is located to the right of the 250bp peak and the genetic analyser software sized the PCR amplicon at 259bp. In the *BRCA1* sample a second peak is located near to the origin of the trace. Comparison with the primer only sample indicates that this peak is unincorporated primer that has not been completely removed during purification. Based upon the size standard these primer peaks were approximately sized as 8 bases although the primer has a known length of 20 bases. It is only peaks that fall below the first standard marker that cannot be accurately sized however the relative position of the cleavage fragment and its absence in the wildtype sample indicates that Endo VII has successfully introduced a nick in the DNA due to the recognition of a mispaired base pair.

In contrast analysis of the cleavage pattern produced from the mixing of antisense probe with T/T homozygote failed to produce a specific cleavage fragment. Mismatch specific Endo VII cleavage would produce a fragment approximately 231-235 bases in length. As can be seen in the ABI-310 traces there are no cleavage peaks at or in the vicinity where a 230 base fragment would be and the control and mutant traces are identical.

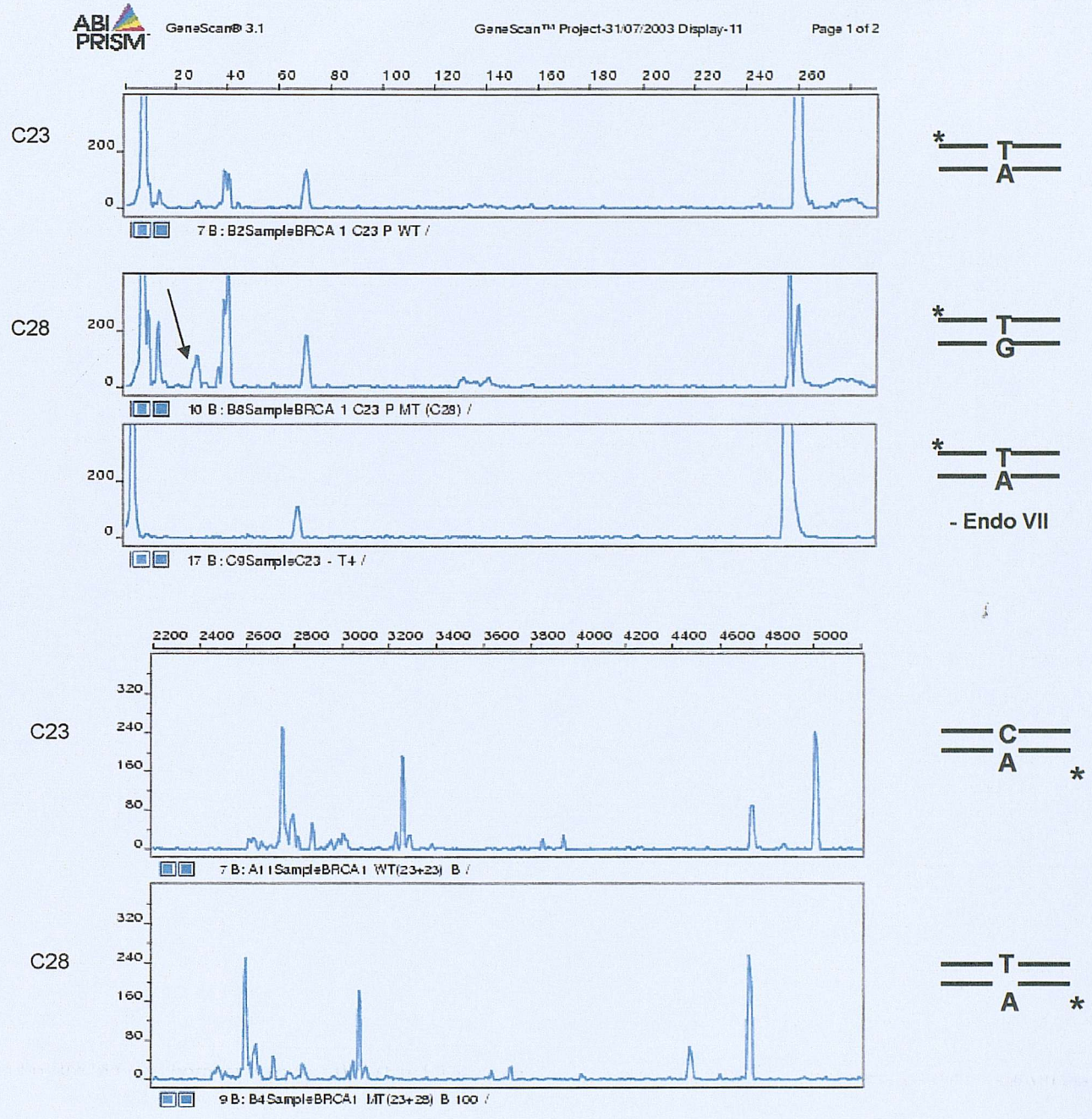


Figure 5.5. Comparison of the cleavage pattern generated by Endo VII cleavage of the 260bp BRCA1 amplicon hybridised to T/T probe with a sense or anti sense FAM tag. The identity of the duplex that is formed and can be observed is indicated to the right of the ABI traces. (A) Endo VII cleavage performed on a homoduplex control and a heteroduplex containing G•T mismatch labelled on the sense strand. A *BRCA1* amplicon not treated with EndoVII with a sense label is situated below these. An arrow indicates specific mismatch cleavage fragments. (B) Repeat of (A) save for the positioning of the probe on the antisense strand. Of the two strands only the sense strand displayed a detectable cleavage fragment.

5.2.1.2. Endo VII Cleavage of 260bp *BRCA1* Amplicon using a T•A probe

To determine whether the lack of specific cleavage seen on the antisense strand was due to the identity of the mismatch an alternative *BRCA1* probe was generated using the T/T homozygote as a template. For the sense labelled reaction the mismatch formed would be T*•G whilst in the antisense reaction the mismatch would be C•A*. These are the same mismatches as formed by C•G probe except the label has changed strands. Therefore whereas before on the antisense strand EndoVII cleavage was based upon the recognition of a G•T* the Endo VII is now trying to detect the more efficiently recognised C•A*. In contrast on the sense strand the C*•A mismatch is replaced by a T*•G.

As shown in Figure 5.5 Endo VII cleavage of the sense probe displayed specific cleavage at the same location as the C•G probe. The cleavage peak observed however was not as great as the one seen for the C*•A mismatch generated using the C•G probe. This is not unexpected because as shown with previous studies the C•A mismatch is more efficently recognised than the T•G mismatch. A cleavage peak that was not present when a C/C probe was used was attributed to a PCR artifact as a *BRCA1* amplicon untreated with EndoVII contained this peak. This ruled it out as an EndoVII cleavage fragment.

The presence of C•A* failed to produce a detectable cleavage fragment specific for mismatch cleavage. This implies that Endo VII is only introducing a single mismatch specific nick that is located upon the sense strand of the duplex. This confirmed the enzyme was only recognising one of the strands.

5.2.1.3. Cleavage of *BRCA1* exon 11 heterozygotes.

Figure 5.6 shows an example of four *BRCA1* fragments hybridised with C*•G sense probe and then digested by Endo VII. The four samples contained a homologous C•G/C•G (C28), T•A/T•A (C23) and two C•G/T•A heterozygotes (C25 & C26). Whereas hybridising the probe to the two homozygotes theoretically generates 100% heteroduplex or homoduplex, the heterozygotes will generate both homoduplex and heteroduplex in roughly

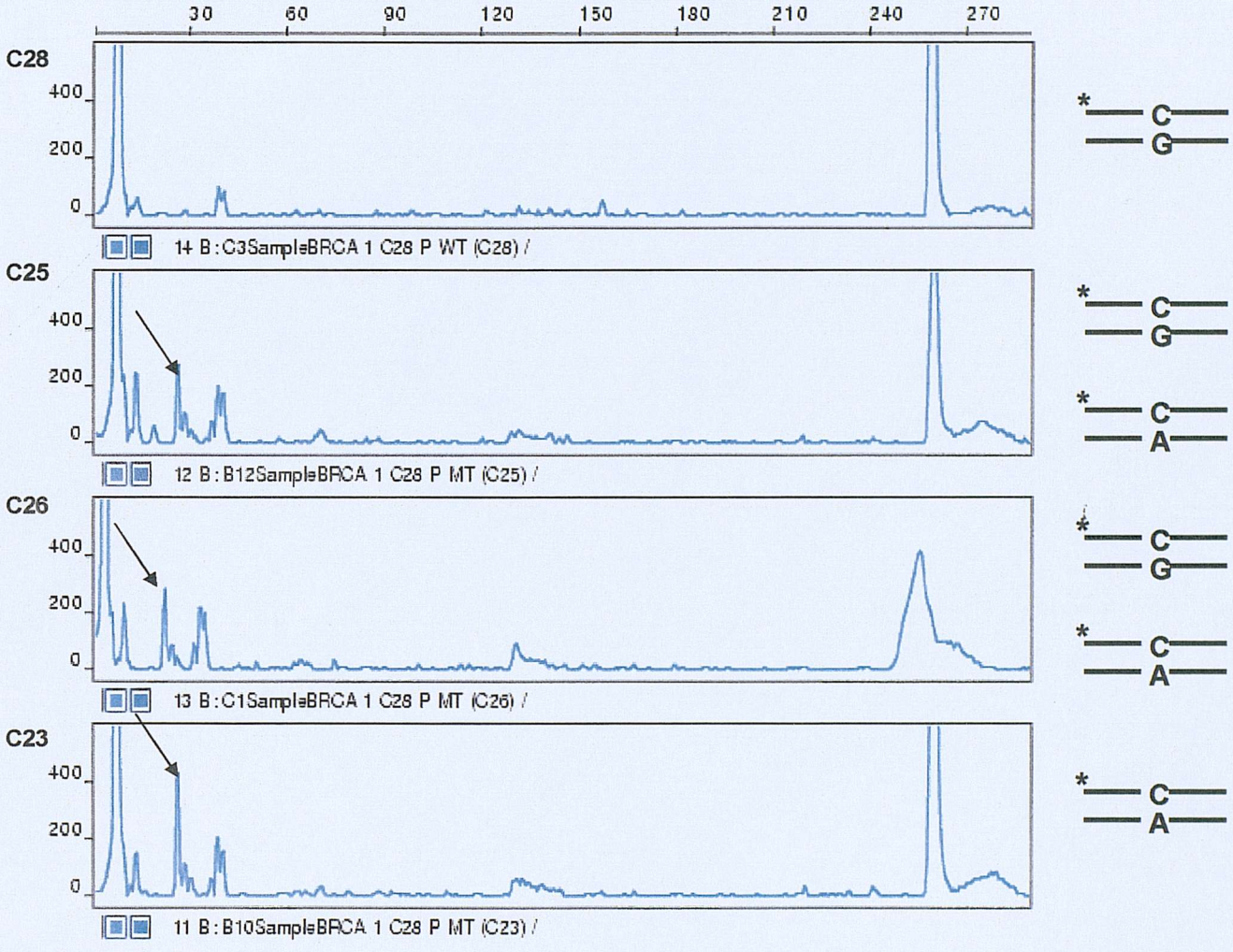


Figure 5.6. EndoVII cleavage of the 260bp *BRCA1* C/C homozygote (C28), C/T heterozygote (C25 & C26) and T/T homozygote (C23) hybridised to a C/C probe with a sense FAM label. The identity of the mismatch base pair if present can be seen to the right of the ABI trace. The arrow indicates the mismatch specific cleavage peak.

equal amounts. This reduction in heteroduplex translates into the reduction of the mismatch specific cleavage fragment seen in figure 5.6 for the heterozygotes C25 and C26.

5.2.1.4. Endo VII Cleavage of 281bp *BRCA1* amplicon.

BRCA1-281F and BRCA1-281R primers were used to amplify a second amplicon from exon11 of the *BRCA1* gene placing the polymorphic site in the middle of the amplicon. Sense probe was generated using the above primers with a fluorescent tag using C•G/C•G homozygote as a template. This was mixed with a homologous C•G/C•G (C28) sample, a homologous T•A/T•A (C23) and two C•G/T•A heterozygotes (C25 & C26). The homozygote T/T and heterozygotes both displayed a cleavage fragment approximately 150-155 bases in length (Fig 5.7). The polymorphism occurs at nucleotide position 149; this would generate an expected cleavage fragment size of 151-155 nucleotides. This cleavage fragment is absent in the homozygote CC sample that upon hybridisation would form homoduplex DNA. Similar to the cleavage of the 260bp *BRCA1* amplicon there is a quantitative difference between the cleavage peaks seen for the homozygote T•T and the heterozygotes. The reduction in heteroduplex in the heterozygote sample results in the reduction in the specific peak seen in the ABI traces.

5.2.2. *IGF2 Apa*l exon 7

The second region used to generate EndoVII substrates was from human insulin like growth factor *IGF2* exon 7 genbank X07868. This was the same region as used for MutS studies and contains an A→ G polymorphism. The identity of the alleles can be determined using restriction digest.

5.2.2.1. Endo VII Cleavage of a Synthetic *IGF2 Apa*l duplex

As interpreting initial studies based on the cleavage of the *IGF2* site proved difficult a synthetic 50bp duplex spanning the polymorphic site was created. Collections of duplexes representing all the possible heteroduplexes

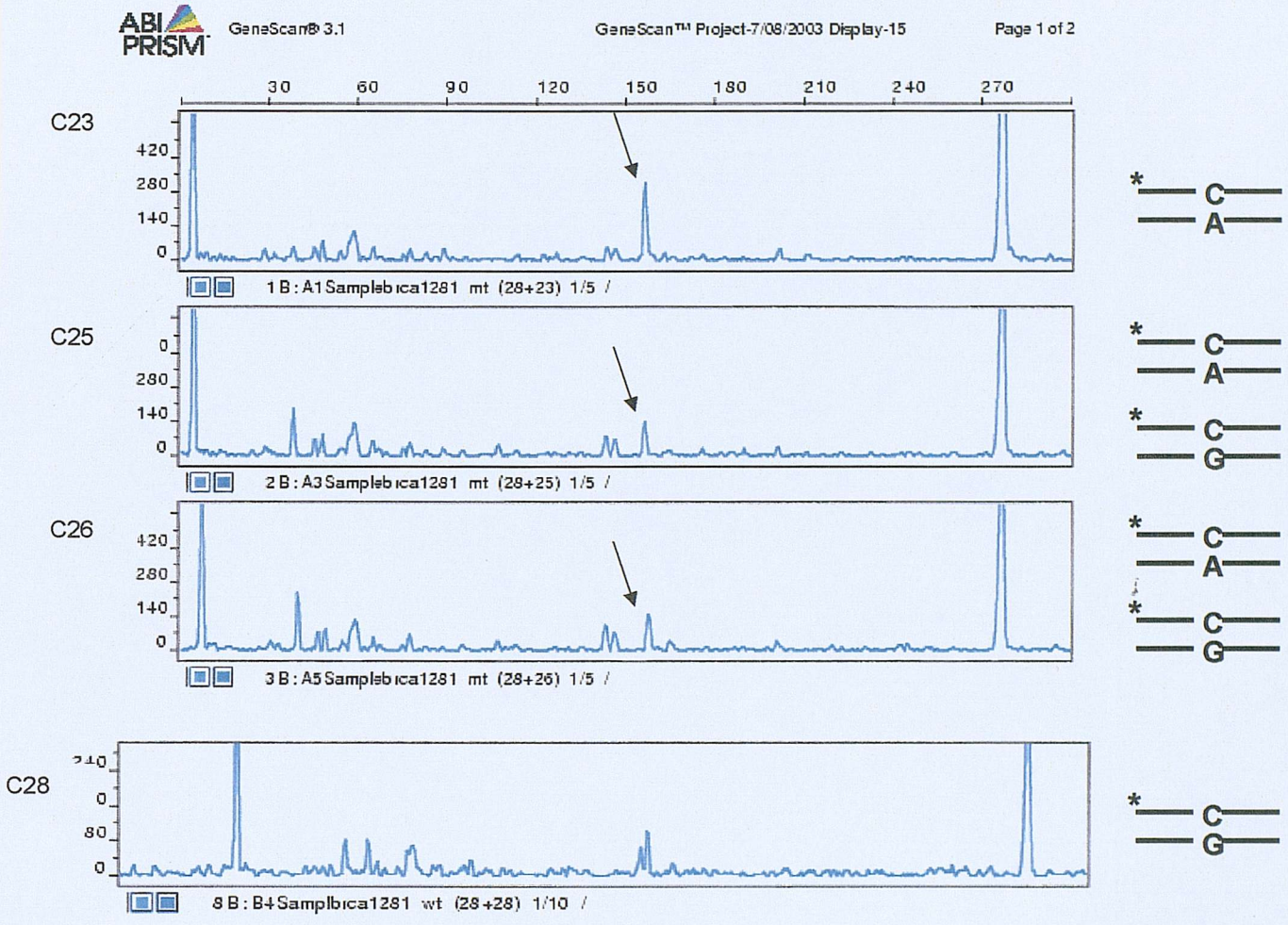


Figure 5.7. EndoVII cleavage of the 281bp *BRCA1* C/C homozygote (C28), C/T heterozygote (C25 & C26) and T/T homozygote (C23) hybridised to a C/C probe with a sense FAM label. The identity of the mismatch base pair if present can be seen to the right of the ABI trace. The arrow indicates the mismatch specific cleavage peak.

or homoduplexes that would be formed upon probe and sample hybridisation were created. Probe generated using the A•T/A•T homozygote as the PCR template can generate duplex DNA containing an A*•T or A*•C base pairing at the polymorphic site based upon sense strand labelling. The DNA species that would be observed with antisense labelling would be an A•T* or G•T* mismatched duplex. In contrast probe generated from a G•C/G•C homozygote upon hybridisation with test DNA can form G*•C or G*•T sense labelled duplex DNA or G•C* and A•C* if detection is based upon an antisense fluorescent tag.

Figure 5.8 shows an example of the cleavage pattern produced by EndoVII cleavage of G*•T and A*•C heteroduplex and the corresponding G*•C and A*•T homoduplex with a 5' FAM sense label. Comparison of the A*•C mismatch duplex and the corresponding A*•T homoduplex failed to produce a specific cleavage fragment. A similar scenario occurred for the G*•T mismatch duplex and the G*•C homoduplex. Although the presence of a mismatch did not produce a specific cleavage fragment the identity of the base did affect Endo VII cleavage. Sense strand with an adenine at the mismatch site produced a significantly larger cleavage peak at the position indicated by the arrow than the corresponding strand with a guanine. As the rest of the strand sequence is identical this difference can only be explained by the difference of this one base between the sense strands.

Analysis of the antisense cleavage (Fig 5.9) showed a significant quantity of non-specific background cleavage similar to the sense strand cleavage. Comparison of the cleavage patterns showed a cleavage peak that was present in the A•C* mismatch duplex that did not occur in the homoduplex control as indicated by the arrow. A similar peak can be seen in the T•G* mismatch duplex though to a lesser extent. This peak is in close proximity to two background peaks that are present in both the mismatch duplexes and the homoduplexes.

5.2.2.2. EndoVII Cleavage of 120bp *IGF2 Apa*I PCR Amplicon

Based upon the cleavage of the short synthetic *IGF2* duplexes, EndoVII nicks the antisense strand in a mismatch specific manner. The A•C

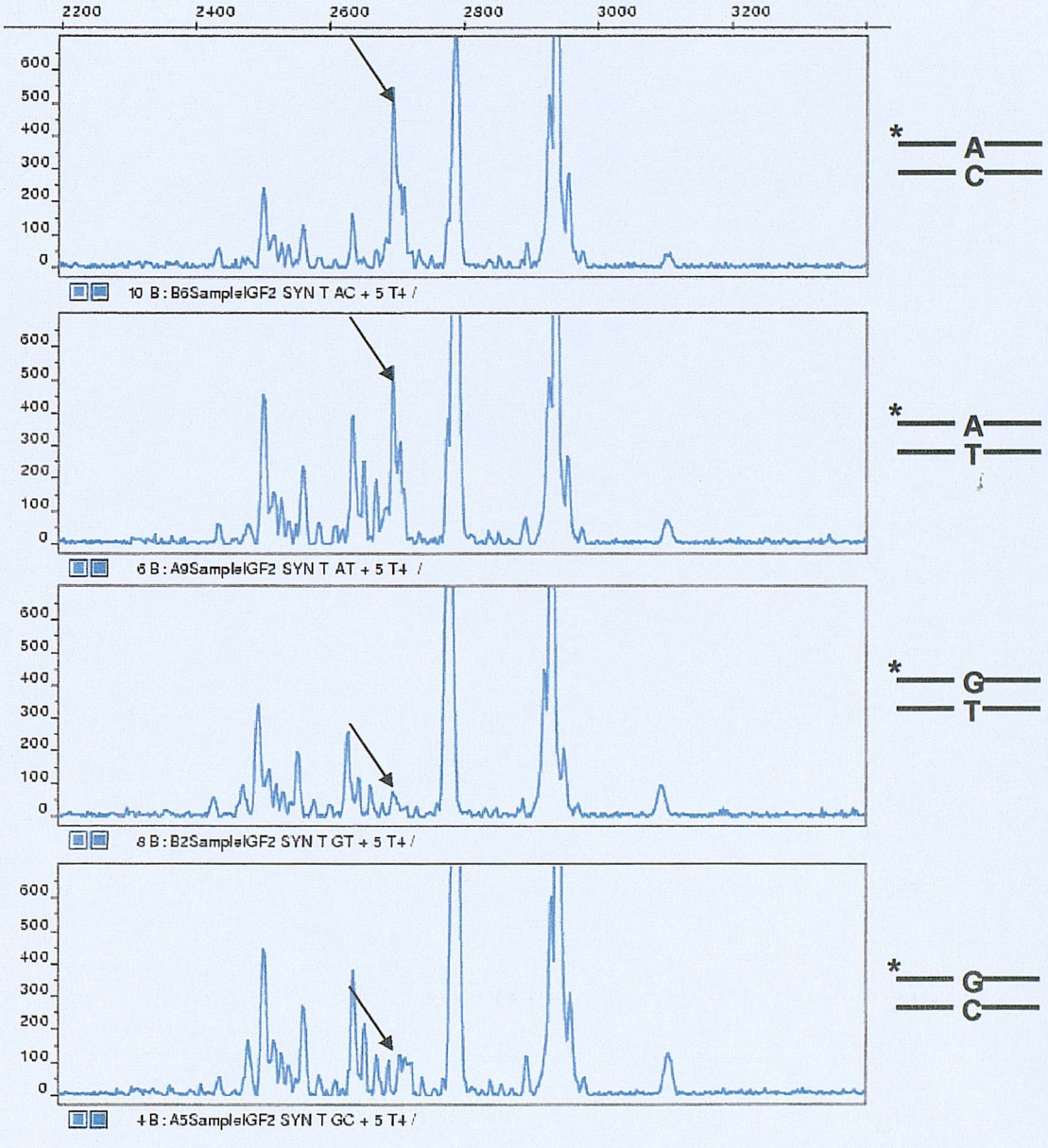


Figure 5.8. Annealing a pair of 50mer oligonucleotides created a 50bp synthetic duplex of the *IGF2 Apa*I polymorphic site. Depending on the sequence of these oligonucleotides a series of homoduplexes or heteroduplexes were formed that differed at the polymorphic site. These represented the potential mismatches or correct base pairing that could occur upon hybridising a probe to sample DNA PCR amplified from the *IGF2 Apa*I region. Duplexes containing an A•C or G•T mismatch and the corresponding homoduplexes were treated with EndoVII. The sense strand was FAM labelled.

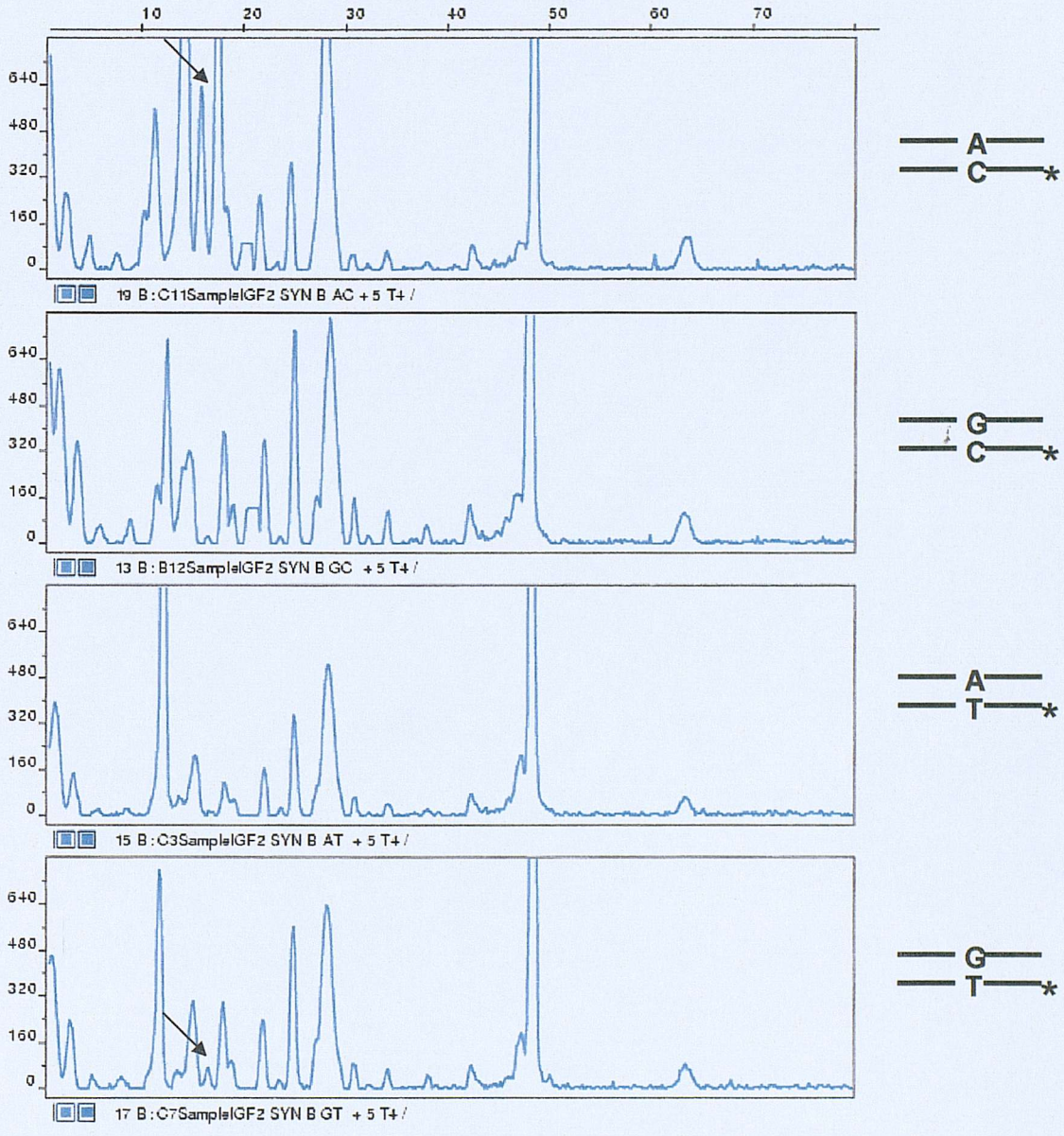
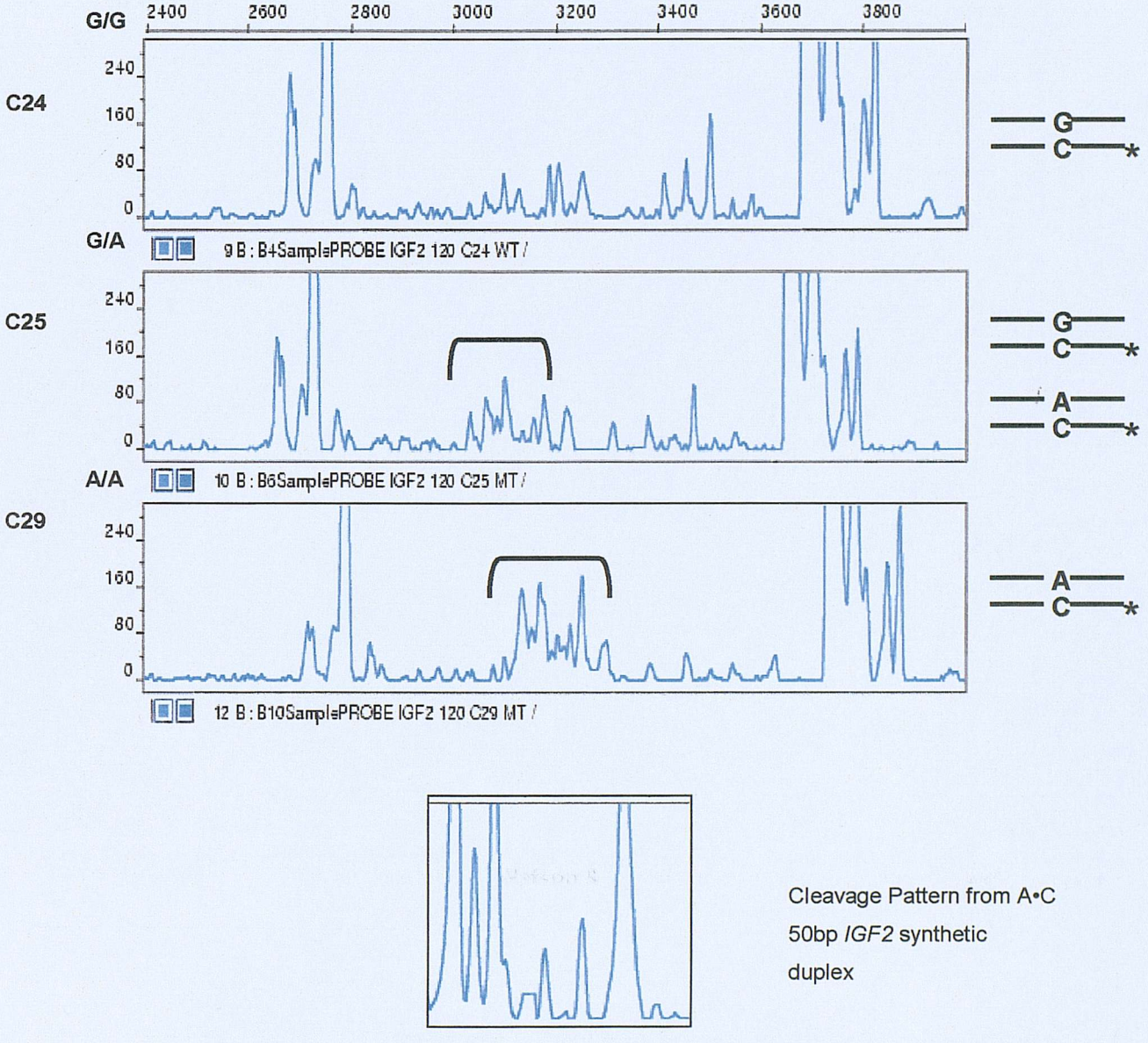


Figure 5.9. EndoVII cleavage of *IGF2 ApaI* synthetic duplexes with antisense FAM label. Specific cleavage fragments indicated by an arrow.

mismatch is more efficiently cleaved than the T•G. To generate the antisense-labelled A•C mismatch using PCR amplicons, the probe needs to be amplified using the G•C/G•C homozygote as the template.

In house control DNA was used to generate test amplicons and the genotype was determined using restriction digestion. An A•T/A•T homozygote (C29), G•C/G•C homozygote (C24) and G•C/A•T heterozygote (C25) were hybridised to antisense labelled probe and treated with EndoVII. The resulting cleavage reaction was analysed using ABI-310 genetic analyser (Figure 5.10). Successful cleavage at the mismatch site would produce a cleavage fragment approximately 52-56 bases in length. As with the synthetic duplex, the region surrounding the location of the mismatch displayed high levels of background cleavage. Comparison of the cleavage patterns between the A•C* synthetic duplex and the A•T/A•T homozygote of the area surrounding the location of the mispaired base pair display close similarities. The A•C* mismatch in the synthetic duplex was characterised by an extra peak in between two other strong cleavage peaks. The corresponding G•C* duplex lacked the central peak but had the two other flanking peaks, although the size of these peaks indicated that the cleavage was not as strong as in the A•C* heteroduplex. C29 displayed a very similar arrangement of peaks, with two flanking peaks and a smaller central peak although the definition of the peaks was not as clear as seen with the synthetic duplex. The heterozygote sample also displayed this cleavage pattern, although the overall size of the peaks was smaller. This is as expected due to the reduction in heteroduplex in the sample. The wildtype homozygote displayed a similar reduction in the two flanking peaks as observed using the synthetic G•C duplex and the absence of the central peak. At position 3200 is located a pair of background cleavage peaks that are found in the synthetic duplexes and the three PCR amplified samples. In the A•C mismatch duplex the left hand peak is smaller than the right hand peak whilst in the G•C homoduplex the two peaks are roughly equal in height. This difference in peak height between the heteroduplex and homoduplex transferred to the PCR amplified samples as well. The C24 sample displays peaks of similar height whilst C25 and C29, which contain heteroduplex DNA, display a difference in height between these two peaks. The identification of mismatch specific cleavage is based upon a



Cleavage Pattern from A•C 50bp *IGF2* synthetic duplex

Figure 5.10. Endo VII cleavage pattern of a 120bp *IGF2* ApaI region. G•C probe was hybridised to PCR amplicons generated from a G/G homozygote (WT, control C24), G/A heterozygote (MT, C25) and an A/A homozygote (MT, C29). The identity of the base pairing at the polymorphic site can be seen to the right of the ABI trace whilst to the left is the identity of the two alleles in the sample. The cleavage pattern for the synthetic A•C IGF2 duplex is located below C29 ABI trace.

weak cleavage fragment and the surrounding background cleavage.

5.2.3. *LDLR* exon 3

A PCR amplified region of exon 3 from the human low density lipoprotein receptor gene (L00337) was used as a substrate for Endo VII cleavage. Within this region there are six heterozygous mutations comprising of five base substitutions, which includes both transitions and transversions. The remaining mutation is a single base deletion. This would provide suitable test for Endo VII. As part of another study within the group a collection of samples from a population study have been characterised and the identity of the mutations they contain if any determined. This collection of known mutations provided a suitable number of samples to test the high throughput capabilities of Endo VII cleavage with MADGE. The identities of the mutations are listed in Table 5.1. The resulting mismatches from heteroduplex formation depend upon the labelled strand of the probe and are listed in table 5.2.

5.2.3.1. Endo VII Cleavage of *LDLR* exon 3 Amplicon

Probe with the wildtype sequence was generated using fluorescently 5' end labelled primers (LDLRex3-F & LDLRex3-R). The probe was designed to have either a FAM label located on the sense strand or a HEX label on the anti-sense strand. Samples known to contain one of the six mutations were used to test Endo VII cleavage. A homozygote wildtype was used as a control to enable the discrimination of specific cleavage and background cleavage. Test DNA was generated using identical primers as used for probe generation except for the removal of the fluorescent tags.

5.2.3.2. Detection of *LDLR* exon 3 Mutations using the ABI 310 Genetic Analyser

Figure 5.12 & 5.13 display the ABI 310 traces from each of the six mutations and a wildtype homozygote. Figure 5.12 shows the cleavage

LDLR exon 3

GCCTCAGTGGGTCTTTCCTTTGAGTGACAGTTCAATCCTGTCTCTTCTGTAGTGTCTGTC
ACCTGCAAATCCGGGGACTTCAGCTGTGGGGGCCGTGTCAACCGCTGCATTCCTCAGTT
CTGGAGGTGC**GA**TGGCCAAGTGGACTGCGACAACGGCTCAGAC**G**AGCAAGGCTGTC**GT**
AAGTGTGGCCCTGCCTTTGCTATTGAGCCTATCTGAGTCCTGG

PCR Product = 220bp

Table 5.3

Mutation	Base Pair Change	Position in Amplicon
D69G	A → G	131
C68Y	G → A	130
313+1	G → T	176
E80K	G → A	163
FsE80	Del G	163
W66G	T → G	121

Table 5.4

	Sense Strand Labelled			Antisense Strand Labelled		
Mutation	Hmz (wildtype)	Htz	Hmz (mutant)	Hmz (wildtype)	Htz	Hmz (mutant)
D69G	A*•T (100%)	A*•T (50%) A*•C (50%)	A*•C (100%)	A•T* (100%)	A•T* (50%) G•T* (50%)	G•T* (100%)
C68Y	G*•C (100%)	G*C (50%) G*T (50%)	G*•T (100%)	G•C* (100%)	G•C* (50%) A•C* (50%)	A•C* (100%)
313+1	G*•C (100%)	G*•C (50%) G*•A (50%)	G*•A (100%)	G•C* (100%)	G•C* (50%) T•C* (50%)	T•C* (100%)
E80K	G*•C (100%)	G*C (50%) G*•T (50%)	G*•T (100%)	G•C* (100%)	G•C* (50%) A•C* (50%	A•C* (100%)
W66G	T*•A (100%)	T*•A (50%) T*•C (50%)	T*•C (100%)	T•A* (100%)	T•A* (50%) G•A* (50%)	G•A* (100%)

Figure 5.11. *LDLR* exon 3 region amplicon that acted as the sequence for the wildtype probe. Table 5.3 lists the 6 mutations found within this site. Table 5.4 gives the identities of the visible DNA species formed upon annealing of mutant DNA with wildtype probe that is labelled on the 5' end of the sense or antisense strand and there relative proportion to one another.

fragments generated by EndoVII cleavage of the sense strand. ABI traces of the sense strand failed to display any peaks present in the mutant samples that were not present in the control sample save for a single fragment indicated by the arrow. This was present in all the mutants, apart from this; samples containing mutations had the same background cleavage as the wildtype homozygote. Endo VII therefore failed to introduce a mismatch specific nick upon the top strand of heteroduplex DNA.

In contrast analysis of the cleavage pattern of the bottom strand (fig 5.13) demonstrated that Endo VII had introduced mismatch specific nicks into the various heteroduplexes generated by the presence of a mutant allele. All contain a peak at the far right (approx pos 5200) of the trace that corresponds to DNA that has not been nicked by the Endo VII. Comparison of the mutants with the control allows the elimination of the small peaks at position 4000, and the peaks below 3200 as mismatch specific because they are present in all traces. This indicates that they are due to non-specific cleavage and can be classed as background. Data from the ABI indicates that Endo VII nicks the bottom strand in a mismatch specific manner in the presence of 5 out of the 6 mutations. Of the 5, 4 generate a strong cleavage fragment. These were 313+1, E80K, Fse80 and W66G whilst C68Y displayed a weak cleavage peak. These peaks were not observed in the control sample and undigested amplicon so are the result of the presence of Endo VII. Based upon Endo VII cleaving 2-6bp on the 3' side of the mismatch the size of the cleavage fragment for the mutations are 38→43 for 313+1, 51→56 for E80K and Fse80, 83→88 for D69G, 84→89 for C68Y and 93→98 for W66G. Comparison with an internal standard showed that the size of the fragments corresponded to the expected size of the mismatch specific cleavage and corresponded to a nick being introduced in the location of the given mismatch.

5.2.3.3. D69G

The D69G sample was resequenced (from the amplicon used for the meltMADGE) to reconfirm sample identity and the presence of the A → G transition. Sequencing confirmed the presence of the mutation. A D69G mutant probe was generated with a HEX group on the 5' end of the bottom

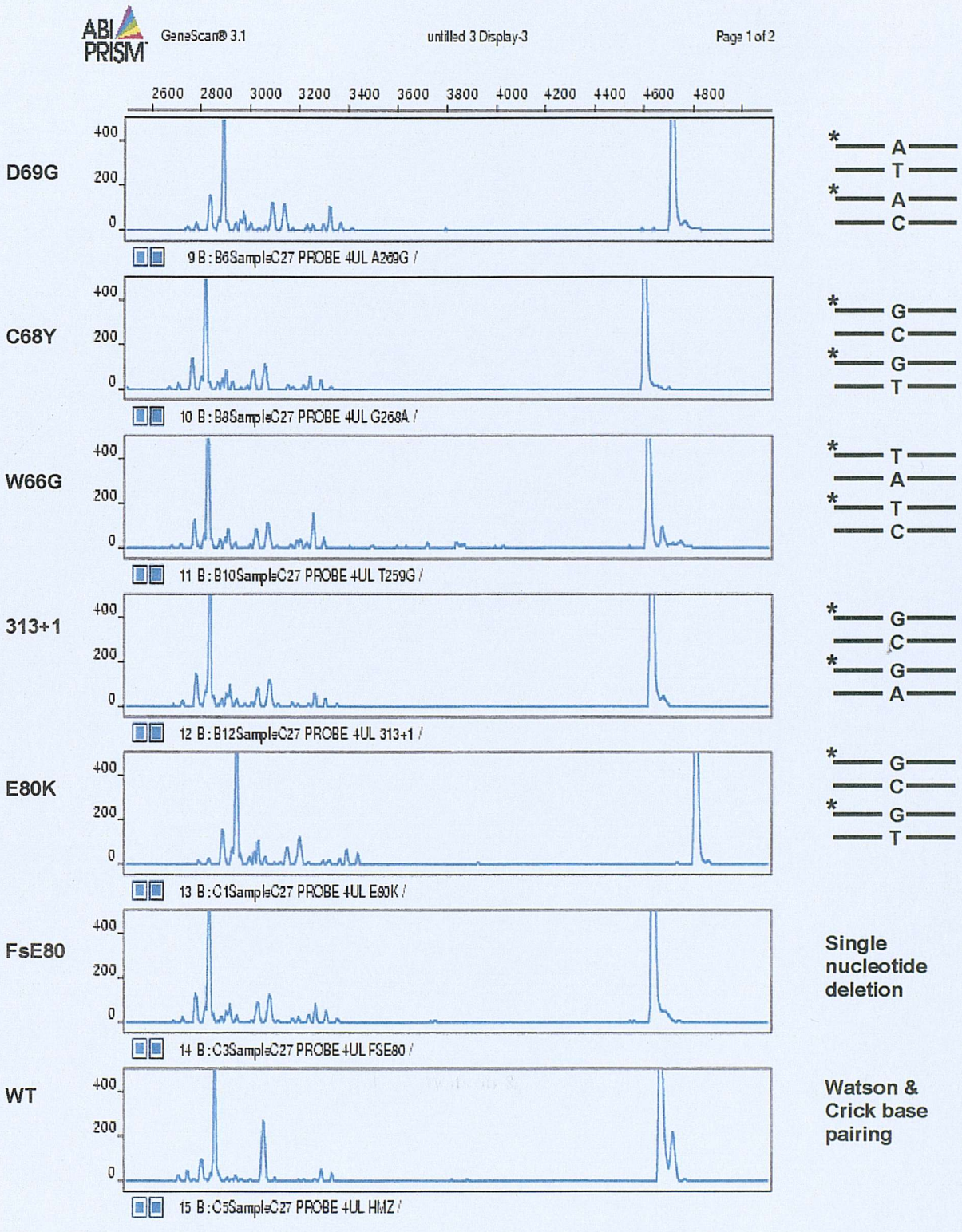


Figure 5.12. Endo VII analysis of 220bp section of the *LDLR* gene from exon 3 known to contain 6 rare heterozygous mutations. DNA samples known to contain one of the mutations and a sample without any of the mutations were amplified using PCR. The resulting amplicons were hybridised to a FAM sense labelled wildtype probe. The resulting EndoVII cleavage failed to produce any specific cleavage fragments for any of the mutations.

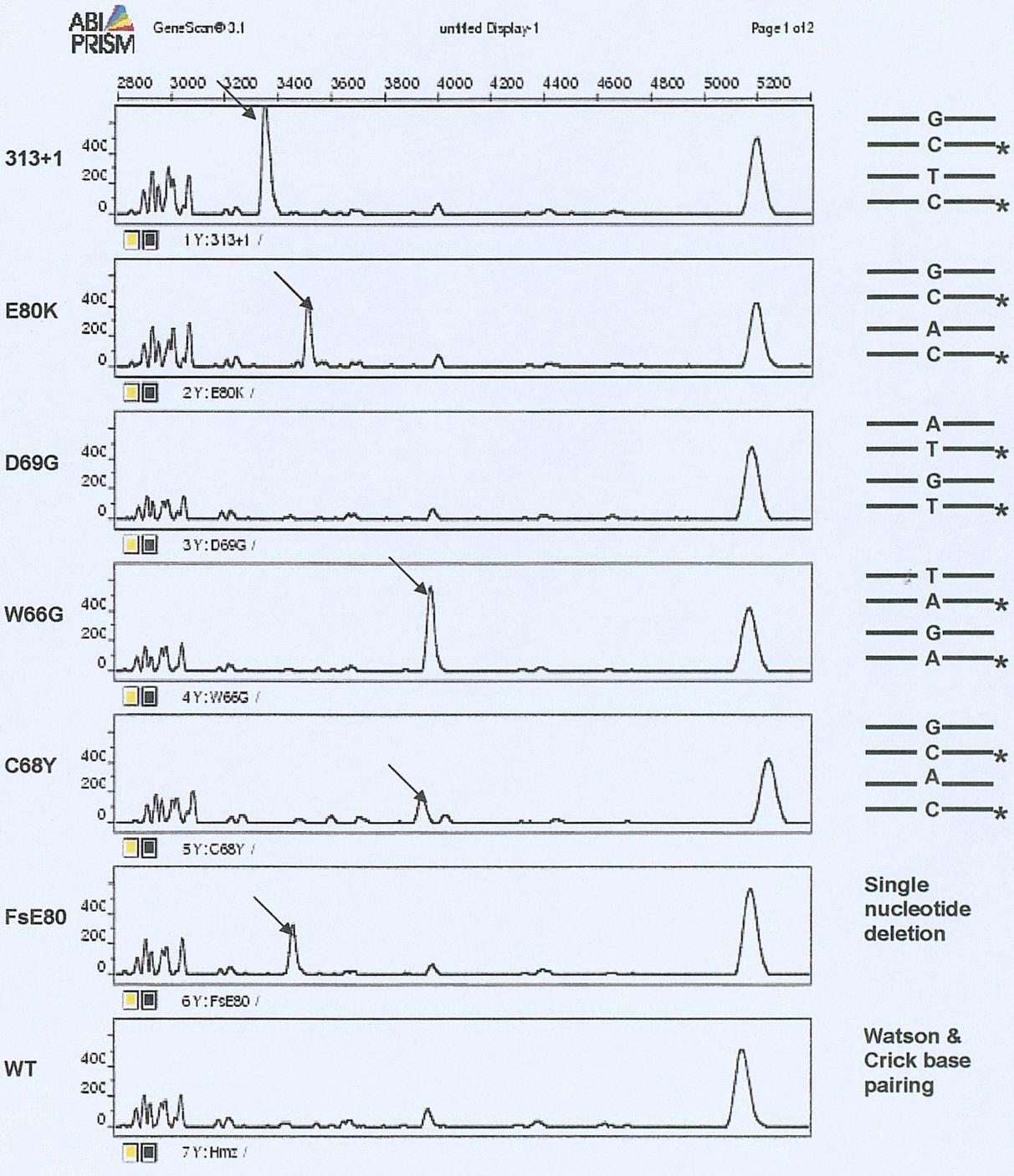


Figure 5.13. EndoVII analysis of 220bp section of the *LDLR* gene from exon 3 known to contain 6 rare heterozygous mutations. DNA samples known to contain one of the mutations and a sample without any of the mutations were amplified using PCR. The resulting amplicons were hybridised to a HEX labelled antisense wildtype probe. The resulting EndoVII cleavage produced specific cleavage fragments for 5 of the 6 mutations as indicated by the arrows. The types of duplex formed by probe hybridisation can be seen to the right of the trace

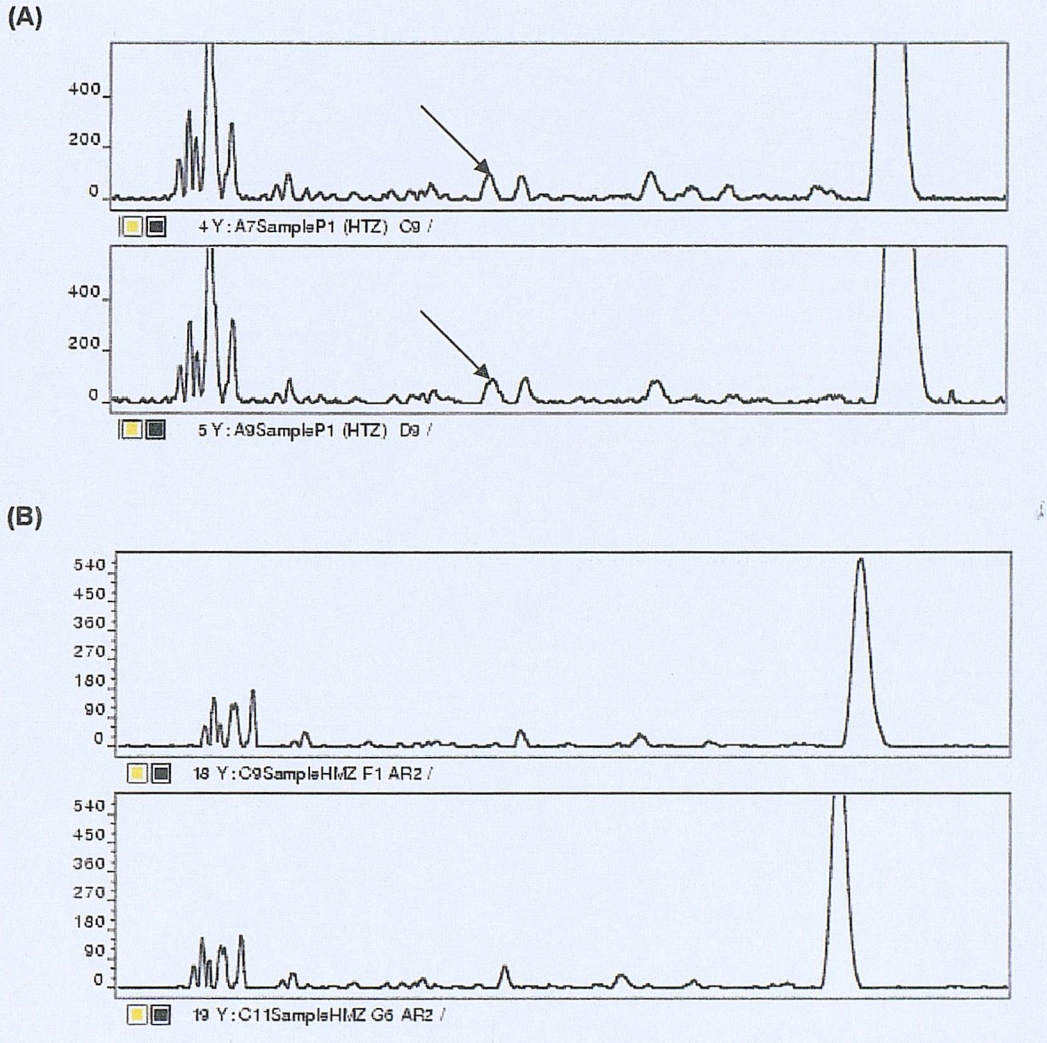
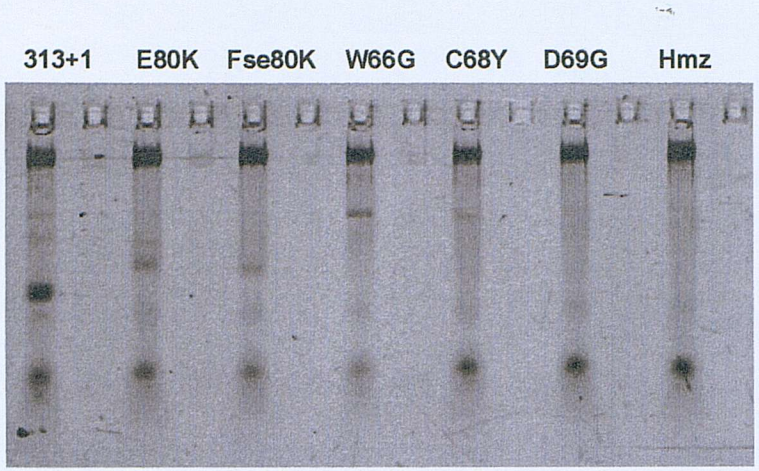


Figure 5.14. Endo VII cleavage of 220bp region of the *LDLR* exon 3 gene. (A) Homozygote wildtype hybridised with an antisense HEX labelled mutant probe generated from D69G sample. The species formed are an A•T* homoduplex and A•C* heteroduplex. (B) Homozygote wildtype hybridised with an antisense HEX labelled wildtype probe. The species formed was an A•T* homoduplex. Specific cleavage fragment indicated by the arrow.

strand. When hybridised to wild type homozygote this would generate A•T* labelled homoduplex and A•C* heteroduplex rather than the A•T*/G•T* species formed between the hybridisation of labelled wildtype probe and unlabelled D69G sample DNA. This was subjected to Endo VII cleavage and analysed on the ABI-310 (Fig 5.14). In contrast to the trace that used wildtype homozygote as the probe, the presence of the A•C* mismatch has produced a weak specific cleavage peak that is not present in the D69G sample hybridised with wildtype probe.

5.2.4. Detection of Mutations using Short tracks

For the LDLR amplicon, EndoVII only specifically nicks the bottom strand, therefore for mutation scanning using short track electrophoresis the probe was generated with only a single fluorophore located on the 5' terminal of the bottom strand. It was observed that the choice of fluorophore plays an important role in the detection of cleavage fragments. Comparisons of HEX and FAM detection by the Molecular Dynamics 595-FluorImager indicated that FAM gave the best signal. Cleavage reactions of the six mutations and a wildtype homozygote run on short tracks produced specific cleavage fragments for the five mutations successfully identified using the ABI 310 (Fig 5.15). The 313+1, E80K, FsE80k, mutations produced clearly visible cleavage fragments that were not present in any of the other samples. The W66G and C68Y cleavage product fell on a non-specific background cleavage fragment. Due to the short tracks this cleavage fragment has migrated to almost the same point as the specific cleavage fragment. In the case of W66G the increase in intensity at this point is clearly visible and translates as an 8 fold increase in intensity over the corresponding peak seen in the wildtype homozygous sample. The C68Y mutation is harder to identify due to the poorer efficiency of EndoVII cleavage. Phoretix analysis determined the intensity of this band to be three fold higher than the background band seen in the other samples excluding W66G. The increases in band intensity therefore act as an indicator to the presence of a mismatch. Traces generated by phoretix show that the C68Y peak although appears to



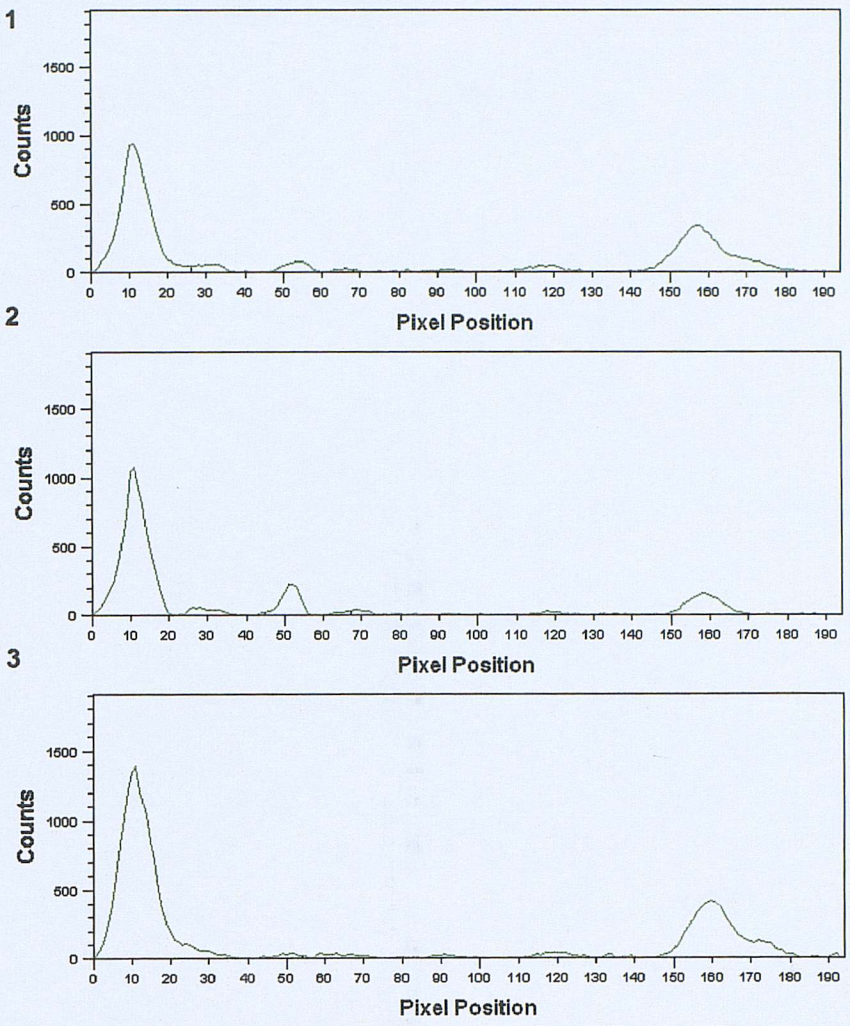


Figure 5.15. EndoVII analysis of 220bp section of the LDLR gene from exon 3 known to contain 6 rare heterozygous mutations using short track electrophoresis. DNA samples known to contain one of the mutations and a sample without any of the mutations were amplified using PCR. The resulting amplicons were hybridised to a FAM labelled antisense wildtype probe. The resulting EndoVII cleavage produced specific cleavage fragments for 5 of the 6 mutations. Phoretix traces for (1) C68Y, (2) W66G, (3) wildtype control (hmz) show the increase in fragment in the C68Y sample in comparison to the wildtype sample and the difference in location of the W66G cleavage fragment and the C68Y/wildtype fragment.

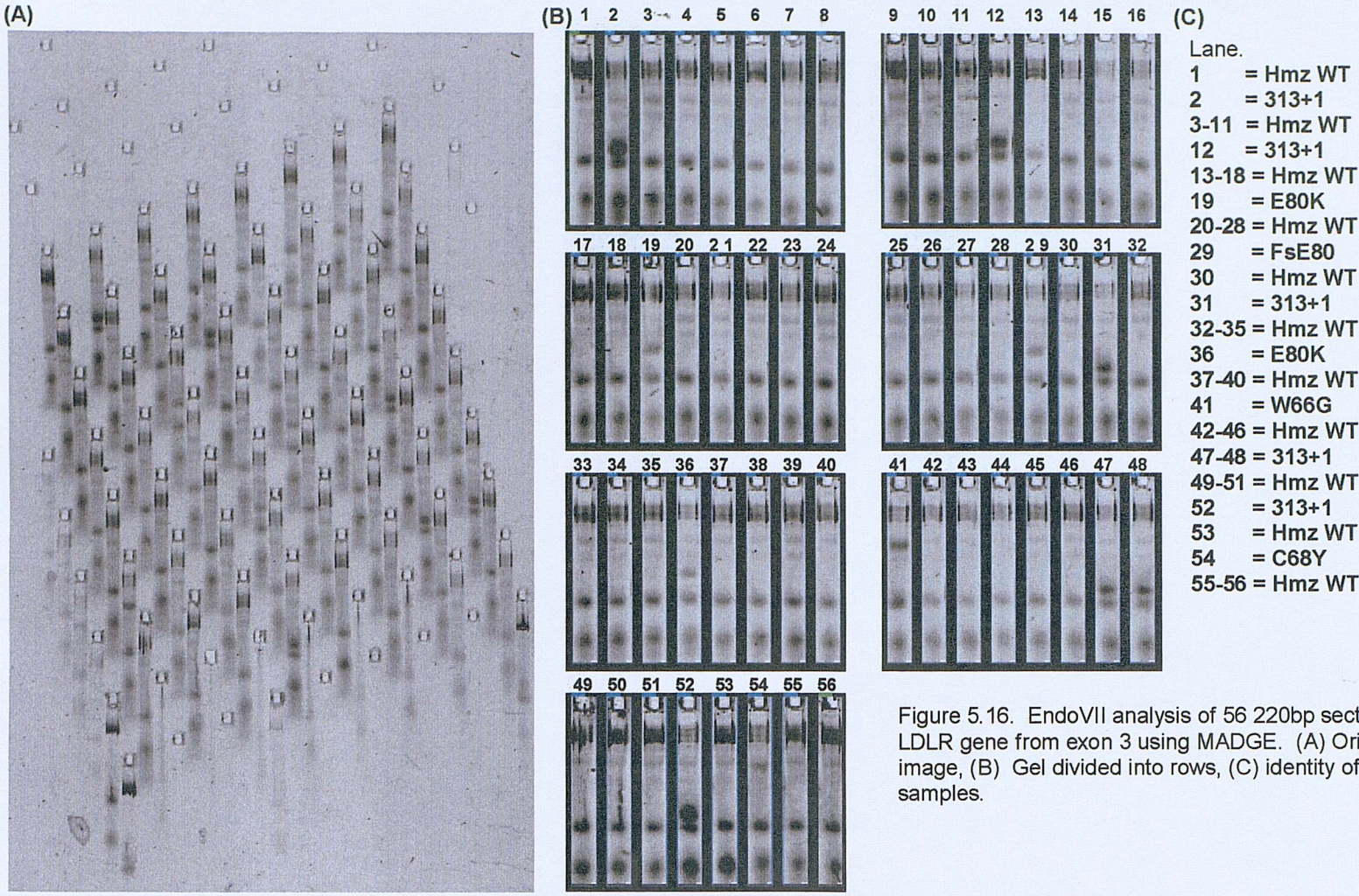


Figure 5.16. EndoVII analysis of 56 220bp section of the LDLR gene from exon 3 using MADGE. (A) Original gel image, (B) Gel divided into rows, (C) identity of the samples.

be at the same point as the non-specific cleavage and the W66G cleavage fragment in the gel image is in fact slightly to the right of these peaks.

5.2.4.1. Endo VII Assay of 56 Amplicons from LDLR exon 3 gene

To test the robustness of Endo VII cleavage a collection of 56 *LDLR* exon 3 amplicons were subjected to Endo VII cleavage with resulting cleavage fragments run on a denaturing MADGE gel (Fig 5.16). The 56 amplicons contained 47 wildtype homozygotes and 11 known heterozygotes, six 313+1 mutants, two E80K mutants and a single FsE80, W66G and C68Y. From the resulting gel 10 of the 11 mutations could be determined by eye from the gel image. The detection of C68Y required further analysis using phoretix. By calculating the proportion of band intensity as a fraction of the total intensity in the track it was observed that the non-specific band seen in the homologous wildtype samples ranged from 1-3%. For C68Y this value increased to 6%.

5.3. Discussion

The results presented in this chapter investigate the suitability of EndoVII for mutation detection and its application to high throughput mutation detection. In total three different regions were studied with varying levels of success to examine Endo VII specificity. Previous studies on the use of Endo VII in mutation detection have screened the mouse β-globin promoter (Youil et al., 1996;Babon *et al.*, 1999), *MLH1* & *MSH2* (Otway *et al.*, 2000) *p53* (Inganas et al., 2000), *BRCA1* (Andrulis et al., 2002) and *FBN1* gene (Youil et al., 2000). Here we used Endo VII to detect known polymorphisms in the region of the *BRCA1*, *IGF2* and mutations in the *LDLR* genes to investigate the mutation detection of Endo VII and to develop EndoVII-MADGE mutation detection assay.

The identification of mutations was based upon cleavage peaks that were absent in control homoduplex reactions. It was also necessary on occasions to compare cleavage patterns with an undigested PCR amplicon to rule out PCR artefacts as cleavage peaks. Two of the three regions were

analysed using capillary electrophoresis using an ABI 310, whilst the *LDLR* exon 3 region was initially analysed using the ABI 310, a larger sample number containing a collection of mutants was successfully analysed and detected using MADGE.

5.3.1. Cleavage Depends on Mismatch

In support with published work (Pottmeyer and Kemper, 1992;Solaro et al., 1993) the efficiency of Endo VII cleavage was dependent on mismatch type and the sequence context. The successful detection of mispaired bases was influenced by both these factors. A third factor was the strand specific nicking of Endo VII. Endo VII reportedly nicks DNA in a nick and counter-nick reaction (Pottmeyer *et al.*, 1992). This means using a probe labelled at each end, upon hybridisation with a heterozgote two heteroduplex species would be formed each labelled on a separate strand. This provides two mispair bases for Endo VII recognition, therefore if one gives a poor cleavage the other should compensate for it. This however supposes that Endo VII will nick both strands specifically, if it does not then the specificity of the technique is reduced as the failure to nick a strand in particular the one containing the stronger mismatch could result in missing mutations. Previous studies (Youil et al., 1995;Youil *et al.*, 1996;Del, Jr. et al., 1998) involving Endo VII have relied upon Endo VII nicking either both strands or a single strand. In all cases if a mutation was missed in the sense strand then it was detected by cleavage of the antisense strand. In this project EndoVII was observed to only nick a single strand in a mismatch specific manner. Both the *IGF2* ApaI amplicon and the *LDLR* exon 3 displayed mismatch specific nicks on the antisense strand and non-specific nicks on the sense strand. This scenario was reversed in the case of the *BRCA1* amplicon. This meant that only a single heteroduplex species was undergoing specific cleavage, the identity of which would determine the specificity of an EndoVII assay. If the recognised heteroduplex contains a G•T mismatch rather than the A•C then there is a reduced likelihood of identifying the mutation. In respect to this project it was observed that, in all gene regions tested, one strand was specifically nicked but the other strand, even if it contained one of the better recognised

mismatches, was not nicked in a mismatch specific manner. i.e. C*•A mismatch generated by hybridising antisense labelled T/T probe with C/C for the *BRCA1* amplicon. It is unclear if this is a common phenomenon or a result of the assay conditions or sequence content. The nicking of the heteroduplexes either on the sense or antisense strand rules out a defect in the Endo VII protein. If only one of the active sites was nicking the DNA we would expect to see all nicking occurring on only one of the strands and no background cleavage on the other strand.

The generation of a probe using either of the homozygotes as templates in both the *IGF2 Apa*I, *BRCA1* regions or in the case of the D69G mutation in *LDLR* ex3 region showed that the identity of the mismatch had an effect on the efficiency of mismatch detection. In all three cases the presence of an A•C mismatch produced a more visible cleavage fragment than the alternative G•T mismatch. Both the *IGF2* and *BRCA1* G•T mismatch specific cleavage fragments were detectable whilst the *LDLR* exon 3 G•T was undetectable and it was only on the use of D69G to produce an A•C mismatch that the presence of the mutation could be detected using the ABI 310. Although the G•T mismatches here displayed relatively poor cleavage signals previous studies (Youil *et al.*, 1995) indicate that Endo VII recognizes G•T efficiently enough for mutation detection.

5.3.2. Mismatch Cleavage Sequence Context

The same base mispairing at different positions displayed different cleavage efficiencies. For example the E80K and C68Y mutations in the *LDLR* exon 3 amplicon. Although both generate an A•C* mismatch through a G→ A transition, they displayed varying cleavage efficiencies, with E80K far more readily detectable than C68Y. The *IGF2 Apa*I region amplicon, which contained a G→ A transition, like the E80K and C68Y mutations, displayed one of the weakest cleavage peaks. This clearly indicates that local nucleotide sequence influences Endo VII recognition of mispaired bases.

All three regions displayed a level of background cleavage that depended upon the sequence of the fragment. The *IGF2* region displayed greater non-specific background when compared to the *LDLR* amplicon. This

background cleavage was highly reproducible between fragments of the same sequence. It has previously been suggested that successful cleavage of heteroduplex DNA can lead to a reduction in background cleavage fragments when compared to a control (Babon et al., 1999). This effect is evident in the cleavage of the *IGF2* amplicon. Of the three gene regions studied the G→ A polymorphism in the *IGF2* amplicon was the most difficult to identify. The cleavage of the synthetic *IGF2* site confirmed that Endo VII recognised and cleaved the mismatch specifically. However the relative efficiency of this cleavage was weak and located in a number of background cleavage peaks. Unlike a distinct single cleavage fragment as seen for the *BRCA1* or *LDLR* exon 3 regions, identification of a base change was based upon the pattern of a number of cleavage fragments in and around the mismatch site. As previously reported (Otway *et al.*, 2000), the presence of the mismatch was seen to increase the intensity of the background peaks flanking the mismatch (Fig 5.9).

To date it has not been determined if there are specific sequence motifs that induce background cleavage. It has been suggested that deviations in local nucleotide structure and their effect on DNA conformation can influence the conformation of DNA at other sites in the DNA molecule (Otway et al., 2000). The cleavage of the two regions of the *BRCA1* gene where the change in mismatch location influenced the background cleavage supports this view (Fig 5.6 & 5.7). The cleavage pattern for the 260bp amplicon differed from the 281bp amplicon. The 260bp amplicon shares 32 bases to the left of the polymorphism and 132 bases to the right with the 280bp. The majority of the differences can therefore be attributed to the sequences not common between the two amplicons. Situated to the right of the mismatch cleavage fragment in the 260bp amplicon is a background peak that is absent in the 281bp amplicon although both contain the sequence. Similarly, to the left of the polymorphic site in the 281bp amplicon is a background cleavage fragment absent in the 260bp amplicon. One explanation for the difference in the latter case could be that as the background cleavage is approximately 10 bases form the main cleavage fragment, in the 260bp amplicon this would place the fragment in the vicinity of the primer band, which could mask the background peak.

The relative position of the mismatch and the sequence context could explain the absence of the background cleavage in the 281bp amplicon seen in the 260bp amplicon. Whereas the 260bp only has 31bp downstream of the mismatch, the 281bp amplicon has 149bp. This may help stabilise the sequence distortion inducing the non-specific background cleavage. In the 260bp amplicon a strong background cleavage fragment is absent when the location of the mismatch is moved and there is an increase of sequence to the left of the mismatch.

5.3.3. High Throughput Identification of Mutations

Although offering greater resolution and more accurate size calling, the ABI 310 does not offer the throughput capabilities of MADGE. Fragment detection utilising the ABI 310 allowed Endo VII specificity and sensitivity to be examined whilst validating that fragments thought to be the result of Endo VII specific cleavage were the size expected.

Due to the availability of a collection of DNA samples with known genotypes, from which to generate test PCR samples and the presence of a number of heterozygous mutations, the *LDLR* amplicon provided a suitable site with which to test both the mismatch specific cleavage of Endo VII and the detection of mutations using MADGE. The reduced resolution and increased background noise associated with electrophoresis using short tracks did not prevent the identification of the 5 *LDLR* mutations identified using the ABI 310 based upon WT probe hybridisation (Fig 5.13 & 5.15).

Following this 56 *LDLR* amplicons were treated to Endo VII cleavage and the products run on a denaturing MADGE gel. 100% of the mutations present were detected. The location and intensity of the band depended upon the identity of the mutation. The strongest cleavage fragment generated by the T•C mismatch from mutation 313+1 was easily detectable whereas the weakest specific cleavage fragment for the C68Y mutation required subsequent analysis to identify. Furthermore there were no false positives.

In terms of population studies the need to purify all the test samples would be labour intensive and time consuming. In respect to the *LDLR* ex3 amplicon it was observed that only probe purification was necessary for Endo

VII cleavage reactions. The presence of PCR reagents unincorporated primers and other PCR artefacts in the test sample did not prevent successful identification of specific mismatch cleavage fragments or adversely affect the activity of Endo VII. Probe purification was necessary to improve the signal to noise ratio. However as Endo VII has a poor specificity for ssDNA (Pottmeyer *et al.*, 1992) and the PCR reagents do not inhibit Endo VII, purification of the probe may be avoided if the quality of the PCR is good enough.

In its present format using wildtype probe this assay would not be able to detect the *LDLR* D69G mutant. It is unclear if this is a sensitivity or specificity issue. As a cleavage fragment was seen when a probe was generated using the D69G mutant forming an A•C mismatch it demonstrates Endo VII is capable of recognising a mispaired base pair at that position and within that sequence context. Unfortunately a D69G homozygote mutant was unavailable which could have been used to hybridise to a wild type probe thus generating a sample containing increased levels of heteroduplex DNA. The presence of a cleavage fragment would have indicated that the lack of cleavage seen in the heterozygote sample was an Endo VII sensitivity issue. The absence of a specific cleavage fragment would indicate an Endo VII specificity issue.

The D69G was the only mutation, which failed to give a specific cleavage fragment using wild type probe. As yet it is not known whether this is a common or rare event; however it can be assumed based upon this work and previous work on Endo VII mutation detection that the majority of mutations would produce a specific nick either on the sense or antisense strand. Although not displaying 100% efficiency the simplicity, throughput capabilities and cost effectiveness of Endo VII-MADGE still make it an attractive mutation scanning technique.

CHAPTER 6
GENERAL CONCLUSIONS

6.1. Project Summary

The primary goal of this project was the investigation of protein-based approaches to heteroduplex recognition for high throughput detection of unknown mutations. Both MutS and Endo VII were produced in house providing a cost effective source of protein. Following initial characterisation of the proteins the mutation detection capabilities of both MutS and Endo VII were tested using a variety of known SNPs and mutations. MutS and Endo VII offered different approaches to mutation detection, MutS offered mobility shift and Endo VII cleavage pattern. In terms of mutation detection Endo VII displayed the greatest potential for mutation detection.

The failure to produce a robust mutation detection assay utilising the mismatch binding properties of MutS is due to specificity of MutS rather than the MADGE format. MutS has displayed some specificity for mismatch DNA but this specificity is too little to be used successfully in this format. Unlike Endo VII, which displayed a significant preference for all the mismatch base pairs, not all mismatch pairs were recognised by MutS. The most poorly recognised mismatch pair for Endo VII showed a greater quantitative difference than the most strongly recognised mismatch pair for MutS. In respect to MutS the application to large number of samples indicated that the discrimination seen was too inconsistent and produced a large number of false results. This unreliability makes MutS unsuitable for high throughput scanning in this format.

The detection of all the base alterations tested clearly indicates that the accuracy of Endo VII is acceptable for mutation detection. The combination of EndoVII cleavage using fluorescently labelled probes and MADGE as the detection platform proved successful in the detection of 11 known mutations in a relatively small sample size of 56 amplicons. The final gel image was available approximately 35 minutes after the commencement of the electrophoretic run and the FluorImager proved effective in detecting the specific cleavage fragments signals. In comparison in this time frame the ABI-310 would have been able to screen between 1 and 2 samples. Previous mutation detection using Endo VII (Babon et al., 1999) has indicated

electrophoresis as the most time consuming part of enzymatic mutation detection.

The mutations at different locations provided distinct cleavage fragments at different positions relative to one another. The inclusion of a DNA ladder would give sizing information and would give an approximate location for the mutation. Although only 56 samples were analysed, the MADGE gel could potentially hold up to 96 samples, coupled with the capacity to run 10 gels in a single tank offers the potential to run just under 1000 samples in a single electrophoresis run. Phoretix offers the ability to analyse the samples quickly.

Coupled with the increased throughput of MADGE is the simplicity of the Endo VII assay. Potentially each reaction can be performed in a single tube. Test PCR can be amplified in a 96 well tetrad plate. To this the probe can be added and the formation of heteroduplex DNA using a pre programmed PCR tetrad machine. Endo VII reaction buffer and Endo VII can subsequently be added to this and the plate placed at the reaction temperature. The final step would be the addition of the stop solution and the denaturing of the DNA, which once again can be performed on the PCR tetrad. As the well array on MADGE is compatible with 96 well arrays the samples can be loaded quickly using a multi channel pipette or alternatively the 96 pinned gel loader.

The requirement for any mutation screening technique comprises a mixture of a number of factors including specificity, sensitivity, throughput, cost effectiveness, reliability or ease of use. By combining the specificity and sensitivity of Endo VII cleavage of fluorescently end labelled amplicons with denaturing MADGE a method has been created that is capable of examining thousands of amplicons for mutations. As such EndoVII-MADGE offers a cost effective alternative to present scanning techniques. In its present format it would be most effective in scanning large population samples for rare or unknown mutations in relatively short amplicons <250bp, prior to identification of any mutations by sequencing. Unlike SSCP conditions for different amplicons are universal therefore no optimisation is required. It is not technically demanding like DGGE or CCM and in respect to CCM far less hazardous. Analysis and data interpretation is straightforward with the use of

phoretix software. The weaknesses of EndoVII-MADGE are that cleavage efficiency will vary with mismatch type and sequence context. Secondly the background cleavage may obscure weak specific cleavage products. The size of amplicon scanned will be limited due to the track lengths found in MADGE gels and localisation information will not be as accurate as compared to other detection platforms such as the ABI 310.

6.2. Future Work

EndoVII-MADGE as a mutation detection assay is still at an early stage and requires further optimisation and method validation. Future work can be broadly put in two categories, that involving Endo VII cleavage and that involving using MADGE for mutation scanning.

Although Endo VII is capable of cleaving all possible mismatches it is apparent from this study and previous studies that sequence context played a role in the efficiency of mismatch cleavage of Endo VII. The role of sequence context and the sequences that induce non-specific background cleavage could form part of future study. Prior to mutation scanning this would allow any sequences that could reduce mismatch cleavage to be identified.

It has been reported (Golz et al., 1998) that the replacement of Tris/HCl buffer with a phosphate buffer increased the selectivity of Endo VII for mismatches that had previously been poorly recognised. A more extensive investigation of reaction conditions could help determine more favourable reaction conditions increasing the efficiency of Endo VII cleavage. Linked to this is the addition of a biological or chemical agent to influence the activity of Endo VII. For example destabilising the mismatch may increase Endo VII recognition of the mismatch site over correctly paired base pairs and encourage greater cleavage at the mismatch site.

In terms of mutation scanning the larger the fragment that can be scanned the more efficient the technique. Firstly it reduces the number of PCR reactions required to scan a region of interest, secondly enables more samples to be scanned. A 220bp fragment was successfully scanned in this project therefore future work needs to determine the size of fragment that can be scanned successfully. As amplicon size increases this will lead to the

increase in background cleavage fragments. With increased length the likelihood of the region containing a SNP increases, which could mask rare mutations. Therefore although larger amplicons can increase method efficiency it may not be entirely beneficial in respect to the detection of rare mutations. To date the largest number of mutations detected simultaneously in a single 1.3kb fragment is 5 (Inganas et al., 2000). The number of mutations in a single fragment detected by MADGE has yet to be determined.

As with any electrophoretic method the movement of fragments is not linear through the gel. Smaller fragments migrate faster through the gel matrix than larger fragments. The long track lengths utilised in conventional gel electrophoresis means that the gels could be run long enough to clearly resolve the cleavage fragments. The short tracks of MADGE act to condense this pattern reducing the resolution of the gel. This potentially means that mutations close to the 5' end labelled strand of the DNA could be missed. The detection of such mutations would be influenced by amplicon length and gel percentage and requires further investigation.

As demonstrated in this project, Endo VII nicked the gene regions tested either on the sense or anti sense in a sequence specific manner. The high throughput capabilities mean that it is possible to run sense and antisense labelled reactions on separate gels. However to reduce the workload a probe needs to be generated with both 5' ends tagged with different fluorophores. It therefore needs to be determined which is the best pair of fluorophores to use when imaging with the Molecular Dynamics FluorImager-595. The fluorophores need to give a good signal so as to identify weakly cleaved mismatches and display very little or no cross over with the other fluorophore to prevent background contamination that could result in the miscalling of mutations.

As yet this technique has only been used to detect known mutations for method validation purposes. Ultimately future work will involve the implementation of this technique to the high throughput identification of unknown mutations.

REFERENCES

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

Andrulis,I.L., Anton-Culver,H., Beck,J., Bove,B., Boyd,J., Buys,S., Godwin,A.K., Hopper,J.L., Li,F., Neuhausen,S.L., Ozcelik,H., Peel,D., Santella,R.M., Southey,M.C., van Orsouw,N.J., Venter,D.J., Vijg,J., and Whittemore,A.S. (2002). Comparison of DNA- and RNA-based methods for detection of truncating BRCA1 mutations. *Hum. Mutat.,* **20,** 65-73.

Avery,O.T., MacLeod,C.M., and McCarty,M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.,* **79,** 137-157.

Babon,J.J., McKenzie,M., and Cotton,R.G. (1999). Mutation detection using fluorescent enzyme mismatch cleavage with T4 endonuclease VII. *Electrophoresis,* **20,** 1162-1170.

Barabino,S.M. and Keller,W. (1999). Last but not least: regulated poly(A) tail formation. *Cell,* **99,** 9-11.

Barnes,W.M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U. S. A,* **91,** 2216-2220.

Bhattacharyya,A., Murchie,A.I., von Kitzing,E., Diekmann,S., Kemper,B., and Lilley,D.M. (1991). Model for the interaction of DNA junctions and resolving enzymes. *J. Mol. Biol.,* **221,** 1191-1207.

Birkenbihl,R.P. and Kemper,B. (1998a). Endonuclease VII has two DNA-binding sites each composed from one N- and one C-terminus provided by different subunits of the protein dimer. *EMBO J.,* **17,** 4527-4534.

Birkenbihl,R.P. and Kemper,B. (1998b). Localization and characterization of the dimerization domain of holliday structure resolving endonuclease VII of phage T4. *J. Mol. Biol.,* **280,** 73-83.

Birkenbihl,R.P. and Kemper,B. (2002). High affinity of endonuclease VII for the Holliday structure containing one nick ensures productive resolution. *J. Mol. Biol.,* **321,** 21-28.

Biswas,I., Obmolova,G., Takahashi,M., Herr,A., Newman,M.A., Yang,W., and Hsieh,P. (2001). Disruption of the helix-u-turn-helix motif of MutS protein: loss of subunit dimerization, mismatch binding and ATP hydrolysis. *J. Mol. Biol.,* **305,** 805-816.

Bjornson,K.P., Allen,D.J., and Modrich,P. (2000). Modulation of MutS ATP hydrolysis by DNA cofactors. *Biochemistry,* **39,** 3176-3183.

Blackwell,L.J., Martik,D., Bjornson,K.P., Bjornson,E.S., and Modrich,P. (1998). Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. *J. Biol. Chem.,* **273,** 32055-32062.

Boorstein,R.J., Hilbert,T.P., Cunningham,R.P., and Teebor,G.W. (1990). Formation and stability of repairable pyrimidine photohydrates in DNA. *Biochemistry,* **29,** 10455-10460.

Brookes,A.J. (1999). The essence of SNPs. *Gene,* **234,** 177-186.

Brown,J., Brown,T., and Fox,K.R. (2001). Affinity of mismatch-binding protein MutS for heteroduplexes containing different mismatches. *Biochem. J.,* **354,** 627-633.

Brown,T. and Hunter,W.N. (1997). Non-watson-Crick Base Associations in DNA and RNA Revelaed by Single Crystal X-Ray Diffraction Methods: Mismatches, Modified Bases, and Nonduplex DNA. *Biopolymers,* **44,** 91-103.

Cheng,K.C., Cahill,D.S., Kasai,H., Nishimura,S., and Loeb,L.A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. *J. Biol. Chem.,* **267,** 166-172.

Cotton Richard,G.H. (1997). *Mutation detection.* Oxford ; New York : Oxford University Press.

Cotton,R.G., Rodrigues,N.R., and Campbell,R.D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. U. S. A,* **85,** 4397-4401.

Day,I.N. and Humphries,S.E. (1994). Electrophoresis for genotyping: microtiter array diagonal gel electrophoresis on horizontal polyacrylamide gels, hydrolink, or agarose. *Anal. Biochem.,* **222,** 389-395.

Day,I.N., O'Dell,S.D., Cash,I.D., Humphries,S.E., and Weavind,G.P. (1995). Electrophoresis for genotyping: temporal thermal gradient gel electrophoresis for profiling of oligonucleotide dissociation. *Nucleic Acids Res.,* **23,** 2404-2412.

Day,I.N., Spanakis,E., Palamand,D., Weavind,G.P., and O'Dell,S.D. (1998). Microplate-array diagonal-gel electrophoresis (MADGE) and melt-MADGE: tools for molecular-genetic epidemiology. *Trends Biotechnol.,* **16,** 287-290.

Del,T.B., Jr., Poff,H.E., III, Novotny,M.A., Cartledge,D.M., Walker,R.I., Earl,C.D., and Bailey,A.L. (1998). Automated fluorescent analysis procedure for enzymatic mutation detection. *Clin. Chem.,* **44,** 731-739.

Eckert,K.A. and Hile,S.E. (1998). Alkylation-induced frameshift mutagenesis during in vitro DNA synthesis by DNA polymerases alpha and beta. *Mutat. Res.,* **422,** 255-269.

Ellis,L.A., Taylor,G.R., Banks,R., and Baumberg,S. (1994). MutS binding protects heteroduplex DNA from exonuclease digestion in vitro: a simple method for detecting mutations. *Nucleic Acids Res.,* **22,** 2710-2711.

Fazakerley,G.V., Quignard,E., Woisard,A., Guschlbauer,W., van der Marel,G.A., van Boom,J.H., Jones,M., and Radman,M. (1986). Structures of mismatched base pairs in DNA and their recognition by the Escherichia coli mismatch repair system. *EMBO J.,* **5,** 3697-3703.

Feng,G. and Winkler,M.E. (1995). Single-step purifications of His6-MutH, His6-MutL and His6-MutS repair proteins of escherichia coli K-12. *Biotechniques,* **19,** 956-965.

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

Friedberg, Walker, and Siede (1995). *DNA repair and mutagenesis.* ASM Press, Washington, D.C.

Galio,L., Bouquet,C., and Brooks,P. (1999). ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. *Nucleic Acids Res.,* **27,** 2325-2331.

Gaunt,T.R., Hinks,L.J., Rassoulian,H., and Day,I.N. (2003). Manual 768 or 384 well microplate gel 'dry' electrophoresis for PCR checking and SNP genotyping. *Nucleic Acids Res.,* **31,** e48.

Giraud-Panis,M.J., Duckett,D.R., and Lilley,D.M. (1995). The modular character of a DNA junction-resolving enzyme: a zinc-binding motif in bacteriophage T4 endonuclease VII. *J. Mol. Biol.,* **252,** 596-610.

Giraud-Panis,M.J. and Lilley,D.M. (1996). T4 endonuclease VII. Importance of a histidine-aspartate cluster within the zinc-binding domain. *J. Biol. Chem.,* **271,** 33148-33155.

Giraud-Panis,M.J. and Lilley,D.M. (1997). Near-simultaneous DNA cleavage by the subunits of the junction-resolving enzyme T4 endonuclease VII. *EMBO J.,* **16,** 2528-2534.

Golz,S., Birkenbihl,R.P., and Kemper,B. (1995). Improved large-scale preparation of phage T4 endonuclease VII overexpressed in E. coli. *DNA Res.,* **2,** 277-284.

Golz,S., Christoph,A., Birkenkamp-Demtroder,K., and Kemper,B. (1997). Identification of amino acids of endonuclease VII essential for binding and cleavage of cruciform DNA. *Eur. J. Biochem.,* **245,** 573-580.

Gradia,S., Acharya,S., and Fishel,R. (2000). The role of mismatched nucleotides in activating the hMSH2-hMSH6 molecular switch. *J. Biol. Chem.,* **275,** 3922-3930.

Greger,B. and Kemper,B. (1998). An apyrimidinic site kinks DNA and triggers incision by endonuclease VII of phage T4. *Nucleic Acids Res.,* **26,** 4432-4438.

Griffiths,A.J.F. (1999). *Modern genetic analysis.* New York ; Basingstoke : W. H. Freeman : Macmillan.

Haber,L.T. and Walker,G.C. (1991). Altering the conserved nucleotide binding motif in the Salmonella typhimurium MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *EMBO J.,* **10,** 2707-2715.

Hansen,L.L., Justesen,J., and Kruse,T.A. (1996). Sensitive and fast mutation detection by solid phase chemical cleavage. *Hum. Mutat.,* **7,** 256-263.

Hsu,I.C., Yang,Q., Kahng,M.W., and Xu,J.F. (1994). Detection of DNA point mutations with DNA mismatch repair enzymes. *Carcinogenesis,* **15,** 1657-1662.

Imlay,J.A. and Linn,S. (1988). DNA damage and oxygen radical toxicity. *Science,* **240,** 1302-1309.

Inganas,M., Byding,S., Eckersten,A., Eriksson,S., Hultman,T., Jorsback,A., Lofman,E., Sabounchi,F., Kressner,U., Lindmark,G., and Tooke,N. (2000). Enzymatic mutation detection in the P53 gene. *Clin. Chem.,* **46,** 1562-1573.

Jones,M., Wagner,R., and Radman,M. (1987). Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics,* **115,** 605-610.

Joshi,A., Sen,S., and Rao,B.J. (2000). ATP-hydrolysis-dependent conformational switch modulates the stability of MutS-mismatch complexes. *Nucleic Acids Res.,* **28,** 853-861.

Karran,P. and Lindahl,T. (1980). Hypoxanthine in deoxyribonucleic acid: generation by heat-induced hydrolysis of adenine residues and release in free form by a deoxyribonucleic acid glycosylase from calf thymus. *Biochemistry,* **19,** 6005-6011.

Kato,R., Kataoka,M., Kamikubo,H., and Kuramitsu,S. (2001). Direct observation of three conformations of MutS protein regulated by adenine nucleotides. *J. Mol. Biol.,* **309,** 227-238.

Kemper,B. and Brown,D.T. (1976). Function of gene 49 of bacteriophage T4. II. Analysis of intracellular development and the structure of very fast-sedimenting DNA. *J. Virol.,* **18,** 1000-1015.

Kirk,B.W., Feinsod,M., Favis,R., Kliman,R.M., and Barany,F. (2002). Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Res.,* **30,** 3295-3311.

Kramer,B., Kramer,W., and Fritz,H.J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell,* **38,** 879-887.

Lai,E., Riley,J., Purvis,I., and Roses,A. (1998). A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics,* **54,** 31-38.

Lamers,M.H., Perrakis,A., Enzlin,J.H., Winterwerp,H.H., de Wind,N., and Sixma,T.K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. *Nature,* **407,** 711-717.

Latchman,D.S. (2002). *Gene regulation : a eukaryotic perspective.* Cheltenham : Nelson Thornes.

Lindahl,T. (1993). Instability and decay of the primary structure of DNA. *Nature,* **362,** 709-715.

Lindahl,T. and Nyberg,B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry,* **11,** 3610-3618.

Lishanski,A., Ostrander,E.A., and Rine,J. (1994). Mutation detection by mismatch binding protein, MutS, in amplified DNA: application to the cystic fibrosis gene. *Proc. Natl. Acad. Sci. U. S. A,* **91,** 2674-2678.

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

Lu,A.L. and Hsu,I.C. (1992). Detection of single DNA base mutations with mismatch repair enzymes. *Genomics,* **14,** 249-255.

Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F., and Richmond,T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature,* **389,** 251-260.

Mizuuchi,K., Kemper,B., Hays,J., and Weisberg,R.A. (1982). T4 endonuclease VII cleaves holliday structures. *Cell,* **29,** 357-365.

Myers,R.M., Fischer,S.G., Lerman,L.S., and Maniatis,T. (1985a). Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.,* **13,** 3131-3145.

Myers,R.M., Larin,Z., and Maniatis,T. (1985b). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science,* **230,** 1242-1246.

Nachman,M.W. and Crowell,S.L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics,* **156,** 297-304.

Nickerson,D.A., Taylor,S.L., Weiss,K.M., Clark,A.G., Hutchinson,R.G., Stengard,J., Salomaa,V., Vartiainen,E., Boerwinkle,E., and Sing,C.F. (1998).

DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.,* **19,** 233-240.

Obmolova,G., Ban,C., Hsieh,P., and Yang,W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature,* **407,** 703-710.

Oleykowski,C.A., Bronson Mullins,C.R., Godwin,A.K., and Yeung,A.T. (1998). Mutation detection using a novel plant endonuclease. *Nucleic Acids Res.,* **26,** 4597-4602.

Orita,M., Suzuki,Y., Sekiya,T., and Hayashi,K. (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics,* **5,** 874-879.

Otway,R., Tetlow,N., Hornby,J., and Kohonen-Corish,M. (2000). Evaluation of enzymatic mutation detectiontrade mark in hereditary nonpolyposis colorectal cancer. *Hum. Mutat.,* **16,** 61-67.

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

Pestova,T.V., Kolupaeva,V.G., Lomakin,I.B., Pilipenko,E.V., Shatsky,I.N., Agol,V.I., and Hellen,C.U. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A,* **98,** 7029-7036.

Pestova,T.V., Lomakin,I.B., Lee,J.H., Choi,S.K., Dever,T.E., and Hellen,C.U. (2000). The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature,* **403,** 332-335.

Pohler,J.R., Giraud-Panis,M.J., and Lilley,D.M. (1996). T4 endonuclease VII selects and alters the structure of the four-way DNA junction; binding of a resolution-defective mutant enzyme. *J. Mol. Biol.,* **260,** 678-696.

Pottmeyer,S. and Kemper,B. (1992). T4 endonuclease VII resolves cruciform DNA with nick and counter-nick and its activity is directed by local nucleotide sequence. *J. Mol. Biol.,* **223,** 607-615.

Raaijmakers,H., Vix,O., Toro,I., Golz,S., Kemper,B., and Suck,D. (1999). X-ray structure of T4 endonuclease VII: a DNA junction resolvase with a novel fold and unusual domain-swapped dimer architecture. *EMBO J.,* **18,** 1447-1458.

Riley,P.A. (1994). Free radicals in biology: oxidative stress and the effects of ionizing radiation. *Int. J. Radiat. Biol.,* **65,** 27-33.

Rossetti,S., Englisch,S., Bresin,E., Pignatti,P.F., and Turco,A.E. (1997). Detection of mutations in human genes by a new rapid method: cleavage fragment length polymorphism analysis (CFLPA). *Mol. Cell Probes,* **11,** 155-160.

Sachidanandam,R., Weissman,D., Schmidt,S.C., Kakol,J.M., Stein,L.D., Marth,G., Sherry,S., Mullikin,J.C., Mortimore,B.J., Willey,D.L., Hunt,S.E., Cole,C.G., Coggill,P.C., Rice,C.M., Ning,Z., Rogers,J., Bentley,D.R., Kwok,P.Y., Mardis,E.R., Yeh,R.T., Schultz,B., Cook,L., Davenport,R., Dante,M., Fulton,L., Hillier,L., Waterston,R.H., McPherson,J.D., Gilman,B., Schaffner,S., Van Etten,W.J., Reich,D., Higgins,J., Daly,M.J., Blumenstiel,B., Baldwin,J., Stange-Thomann,N., Zody,M.C., Linton,L., Lander,E.S., and Altshuler,D. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* **409,** 928-933.

Sixma,T.K. (2001). DNA mismatch repair: MutS structures bound to mismatches. *Curr. Opin. Struct. Biol.,* **11,** 47-52.

Solaro,P.C., Birkenkamp,K., Pfeiffer,P., and Kemper,B. (1993). Endonuclease VII of phage T4 triggers mismatch correction in vitro. *J. Mol. Biol.,* **230,** 868-877.

Strachan and Read (1999). *Human molecular genetics.* Bios, Oxford.

Stryer (1988). *Biochemistry.* W.H. Freeman, New York.

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

Surdi,G.A., Yaar,R., and Smith,C.L. (1999). Discrimination of DNA duplexes with matched and mismatched tandem repeats by T4 endonuclease VII. *Genet. Anal.,* **14,** 177-179.

Tachiki,H., Kato,R., Masui,R., Hasegawa,K., Itakura,H., Fukuyama,K., and Kuramitsu,S. (1998). Domain organization and functional analysis of Thermus thermophilus MutS protein. *Nucleic Acids Res.,* **26,** 4153-4159.

Taylor,G.R. (1999). Enzymatic and chemical cleavage methods. *Electrophoresis,* **20,** 1125-1130.

Vairapandi,M. and Duker,N.J. (1994). Excision of ultraviolet-induced photoproducts of 5-methylcytosine from DNA. *Mutat. Res.,* **315,** 85-94.

Watson,J.D. and Crick,F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* **171,** 737-738.

Wood,R.D., Mitchell,M., Sgouros,J., and Lindahl,T. (2001). Human DNA repair genes. *Science,* **291,** 1284-1289.

Wu,T.H. and Marinus,M.G. (1999). Deletion mutation analysis of the mutS gene in Escherichia coli. *J. Biol. Chem.,* **274,** 5948-5952.

Yamamoto,A., Schofield,M.J., Biswas,I., and Hsieh,P. (2000). Requirement for Phe36 for DNA binding and mismatch repair by Escherichia coli MutS protein. *Nucleic Acids Res.,* **28,** 3564-3569.

Yang,B., Wen,X., Kodali,N.S., Oleykowski,C.A., Miller,C.G., Kulinski,J., Besack,D., Yeung,J.A., Kowalski,D., and Yeung,A.T. (2000). Purification, cloning, and characterization of the CEL I nuclease. *Biochemistry,* **39,** 3533-3541.

Youil,R., Kemper,B., and 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., and 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.

Youil,R., Toner,T.J., Bull,E., Bailey,A.L., Earl,C.D., Dietz,H.C., and Montgomery,R.A. (2000). Enzymatic mutation detection (EMD) of novel mutations (R565X and R1523X) in the FBN1 gene of patients with Marfan syndrome using T4 endonuclease VII. *Hum. Mutat.,* **16,** 92-93.