

# **Variation in the Matrix Metalloproteinase-3 Gene in Relation to Atherosclerosis**

By

Seyyare Beyzade

Human Genetics Division  
Southampton General Hospital,  
Tremona Road,  
Southampton SO16 6YD

Submitted for the degree of Doctor of Philosophy at the University of  
Southampton in October 2002

## Abstract

Atherosclerosis is a multifactorial progressive disease. It is the underlying cause of coronary heart disease (CHD) and strokes. Though its cause is unknown, the prevailing hypothesis is that atherosclerotic lesions are initiated by an altered endothelial lining of the arterial walls. Factors that can influence changes in the inner lining of the arterial wall include oxidised low-density lipoproteins (OxLDL), high blood pressure, homocysteine, toxins and viral infections. Various designs of heritability studies of both intermediate traits and CHD have demonstrated that genetic susceptibility is also important.

Matrix metalloproteinase-3, (MMP-3 or stromelysin-1) is a member of the matrix metalloproteinase (MMPs) family, capable of degrading extracellular macromolecules. MMPs have been implicated in a number of pathological conditions such as rheumatoid arthritis, tumour invasion and metastasis, and atherosclerosis. It has been shown that MMP-3 is expressed in coronary atherosclerotic plaques, particularly in regions prone to rupture, making it a good candidate as a genetic factor for atherosclerosis. In 1996, a common variant was identified in the promoter region of the MMP-3 gene, located at 1612bp upstream from the start of transcription. This 5A/6A polymorphism has since been shown to be associated with atherosclerosis in a number of independent studies.

In this study further investigations of the MMP-3 gene were undertaken in relation to atherosclerosis. Using the mutation scanning technique dideoxy fingerprinting (ddF), the promoter and coding regions of the MMP-3 gene were screened for unknown polymorphisms. Six new polymorphisms were identified using this method and confirmed by sequencing. Relative to transcription start site these were -1986 T/C; -1612 5A/6A; -1346 A/C; -709 A/G; -376G/C; and in exon 2, 802 A/G and 952 A/G. All six variants are single nucleotide polymorphisms (SNPs). Genotyping of these SNPs in unrelated individuals showed that they were in strong linkage disequilibrium with each other and with the previously identified 5A/6A polymorphism. The two commonest haplotypes were T-5A-A-A-G-A-A and C-6A-C-G-C-G-G. In a cohort of 913 patients with angiographically documented coronary artery disease (CAD), the 5A



allele-containing genotypes were over-represented in patients with a history of MI ( $p=0.03$ ). Furthermore, an association between the number of coronary arteries  $>50\%$  stenosis and the 6A/6A genotype was observed ( $p=0.01$ ).

To determine whether these polymorphisms affect MMP-3 promoter activity, transient transfection studies were performed using reporter genes for the two alleles of each individual polymorphism. Small but statistically significant differences of promoter activity were observed for each individual SNP. However, relative to the two commonest haplotypes, the common allele of three SNPs belonging to one haplotype conferred higher activity, whereas the common allele of the other two conferred lower activity.

These results suggest that individuals carrying the transcriptionally more active 5A allele are predisposed to the development of unstable plaques and CHD, whereas those carrying the transcriptionally less active 6A allele are more likely to develop stable but highly stenotic plaques.

## **Acknowledgements**

I would like to thank the British Heart Foundation for their support in this research under the PG98/192 grant.

I also express my gratitude to my mentors Dr Shu Ye and Prof Ian Day for all their help, encouragement, assistance and comments during my laboratory work and write up of this thesis. I would also like to thank all the staff in the Human Genetics Division for their help and guidance during my study. Especial acknowledgement is owed to Patricia Briggs for her help and support.

Finally I would like to thank my parents for their emotional support.

## Contents

|                      |     |
|----------------------|-----|
| Title page           | I   |
| Abstract             | II  |
| Acknowledgements     | IV  |
| Contents             | V   |
| List of Figures      | XII |
| List of Tables       | XIV |
| List of Abbreviation | XVI |

### **Chapter 1: General Introduction: -**

|   |    |
|---|----|
| 1.1 Atherosclerosis                           | 1  |
| 1.1.1 Coronary Heart Disease                  | 3  |
| 1.1.1.1 Angina Pectoris                       | 4  |
| 1.1.1.2 Acute Myocardial Infarction           | 4  |
| 1.1.2 Progression of Atherosclerosis          | 5  |
| 1.1.2.1 Response to Injury Theory             | 5  |
| 1.1.2.2 Benditt and Benditt Monoclonal Theory | 7  |
| 1.1.2.3 Lipoproteins                          | 7  |
| 1.1.3 Macrophages and Smooth Muscle Cells     | 8  |
| 1.1.3.1 Macrophages                           | 8  |
| 1.1.3.2 Smooth Muscle Cells                   | 8  |
| 1.1.4 Classical and Newer Risk Factors        | 10 |
| 1.1.4.1 High Cholesterol                      | 10 |
| 1.1.4.2 Clinical Diabetes Mellitus            | 11 |
| 1.1.4.3 Hypertension                          | 11 |
| 1.1.4.4 Smoking                               | 12 |
| 1.1.4.5 Family History                        | 12 |
| 1.1.5 Heritability                            | 12 |
| 1.1.6 Genetic Risk Factors and Candidate Gene | 13 |
| 1.1.6.1 Angiotensin-Converting Enzyme         | 15 |
| 1.1.6.2 Apolipoprotein E                      | 15 |
| 1.1.6.3 Paradoxonase                          | 16 |

|  |    |
|--|----|
| 1.1.6.4 Factor V   | 16 |
| 1.2 Matrix Metalloproteinases (MMPs)                     | 18 |
| 1.2.1 The Basic Structural domains of MMPs               | 19 |
| 1.2.2 The Regulation of MMPs                             | 22 |
| 1.2.2.1 Transcriptional Regulation                       | 22 |
| 1.2.2.2 Activation of Latent MMPs                        | 24 |
| 1.2.2.3 Inhibition of MMPs activity                      | 25 |
| 1.2.3 MMPs Role in Biological and Pathological Processes | 29 |
| 1.2.3.1 Tumour Invasion and Metastasis                   | 30 |
| 1.2.3.2 Rheumatoid Arthritis                             | 31 |
| 1.2.3.3 Coronary Heart Disease                           | 32 |
| 1.2.3.4 Aneurysm   | 32 |
| 1.2.4 MMP-3 (Stromelysin-1), a Member of theMMP Family   | 35 |
| 1.3 Hypothesis   | 42 |
| 1.3.1 Plan   | 42 |

## **Chapter2: Materials and Methods**

|   |    |
|---|----|
| 2.1 Materials   | 43 |
| 2.1.1 Chemicals   | 43 |
| 2.1.2 Enzymes   | 43 |
| 2.1.3 Water   | 43 |
| 2.1.4 Sterilization   | 43 |
| 2.1.5 Bacterial Growth Medium   | 44 |
| 2.1.5.1 Luria-Bertani (LB) Medium   | 44 |
| 2.1.5.2 LB Agar   | 44 |
| 2.1.5.3 Ampicillin  | 44 |
| 2.1.6 Statins   | 44 |
| 2.1.7 PMA (phorbol 12-myristate 13-acetate, a protein kinase C activator) | 44 |
| 2.1.8 Oligonucleotide Primers   | 45 |
| 2.1.9 Plasmids  | 45 |
| 2.1.10 Purification Kits  | 45 |
| 2.1.11 Buffers and Solutions  | 45 |
| 2.2 Methods   | 49 |

|            |   |    |
|------------|---|----|
| 2.2.1      | Mutation Scanning of the MMP-3 Gene                   | 49 |
| 2.2.1.1    | Polymerase Chain Reaction                             | 49 |
| 2.2.1.2    | Gel electrophoresis and purification of PCR products  | 51 |
| 2.2.1.3    | Dideoxy termination reaction                          | 53 |
| 2.2.1.4    | Non-denaturing polyacrylamide gel electrophoresis     | 56 |
| 2.2.2      | Radioactive, manual sequencing                        | 57 |
| 2.2.3      | Fluorescent, automated sequencing                     | 58 |
| 2.2.3.1    | Principal of the procedure                            | 58 |
| 2.2.3.2    | Amplification of DNA fragments                        | 59 |
| 2.2.3.3    | Purification of PCR products                          | 59 |
| 2.2.3.4    | Sequencing reaction                                   | 60 |
| 2.2.3.5    | Purification of extension products                    | 60 |
| 2.2.3.6    | Electrophoresis on the ABI PRISM 310 genetic analyser | 61 |
| 2.2.4      | Allelic association study                             | 61 |
| 2.2.4.1    | Subjects and DNA preparation                          | 61 |
| 2.2.4.2    | PCR amplification of genomic DNA                      | 62 |
| 2.2.4.3    | Restriction enzyme digest                             | 63 |
| 2.2.4.4    | Electrophoresis                                       | 64 |
| 2.2.4.5    | Statistical analysis                                  | 65 |
| 2.2.5      | Transient reporter gene expression                    | 65 |
| 2.2.5.1    | Reporter gene constructs for individual polymorphisms | 66 |
| 2.2.5.1.1  | Restriction endonuclease digest of pGL3-vectors       | 67 |
| 2.2.5.1.2  | Purification of pGL3-promoter vector                  | 68 |
| 2.2.5.1.3  | Preparation of double stranded oligonucleotides       | 68 |
| 2.2.5.1.4  | Phosphorylation of double stranded oligonucleotides   | 69 |
| 2.2.5.1.5  | Ligation of plasmid vector and insert DNA             | 69 |
| 2.2.5.1.6  | Growth and preservation of <i>E.coli</i>              | 70 |
| 2.2.5.1.7  | Preparation of competent <i>E.coli</i>                | 70 |
| 2.2.5.1.8  | Transformation of competent <i>E.coli</i>             | 71 |
| 2.2.5.1.9  | Analysis of transformants                             | 71 |
| 2.2.5.1.10 | PCR amplification of transformants                    | 71 |
| 2.2.5.1.11 | Restriction endonuclease digestion of transformants   | 72 |
| 2.2.5.2    | Large scale preparation of plasmid DNA                | 72 |

|  |    |
|--|----|
| 2.2.5.3 Mammalian cell culture, transfection and reporter assay  | 74 |
| 2.2.5.3.1 Propagation and storage of mammalian cells             | 74 |
| 2.2.5.3.2 Transient transfection of MALU and RAW cells           | 74 |
| 2.2.5.3.3 Reporter assay   | 76 |
| 2.2.6 Electrophoretic mobility shift assay (EMSA)                | 76 |
| 2.2.6.1 Preparation of probes                                    | 77 |
| 2.2.6.2 Isolation of nuclear protein extract from cultured cells | 78 |
| 2.2.6.3 Protein-DNA binding and electrophoresis                  | 79 |
| 2.2.7 Association study  | 79 |
| 2.2.7.1 Statistical analysis                                     | 80 |

### **Chapter 3: Mutation Scanning of the MMP-3 Gene**

|   |     |
|---|-----|
| 3.1 Introduction-A Comparison of Different Mutation Detection Methods | 81  |
| 3.1.1 Single Stranded Conformation Polymorphism (SSCP)                | 83  |
| 3.1.2 Heteroduplex Analysis (HA)                                      | 84  |
| 3.1.3 Mutation Detection Enhancement (MDE)                            | 85  |
| 3.1.4 Denaturing Gradient Gel Electrophoresis (DGGE)                  | 86  |
| 3.1.5 Temperature Gradient Gel Electrophoresis (TGGE)                 | 87  |
| 3.1.6 Base Excision Sequence Scanning (BESS)                          | 88  |
| 3.1.7 Ribonuclease Cleavage of Mismatch RNase                         | 89  |
| 3.1.8 Chemical Cleavage of Mismatch (CCM)                             | 90  |
| 3.1.9 Denaturing High-Performance Liquid Chromatography (DHPLC)       | 91  |
| 3.1.10 Melt MADGE   | 92  |
| 3.1.11 Direct Sequencing (DS)   | 93  |
| 3.1.12 Automated Sequencing   | 94  |
| 3.1.13 DNA Chip   | 95  |
| 3.1.14 Southern Blotting  | 95  |
| 3.1.15 Dideoxy Fingerprinting (ddF)                                   | 96  |
| 3.2 Results   | 99  |
| 3.2.1 ddF Results   | 99  |
| 3.2.1.1 Polymorphism detected using primer MMP3-ddF2r                 | 102 |
| 3.2.1.2 Polymorphism detected using primer MMP3-ddF5f                 | 103 |
| 3.2.1.3 Polymorphism detected using primer MMP3-ddF8r                 | 104 |

|  |     |
|--|-----|
| 3.2.1.4 Polymorphism detected using primer MMP3-ddF10r | 104 |
| 3.2.1.5 Polymorphism detected using primer MMP3-ddF13f | 105 |
| 3.2.1.6 Polymorphism detected using primer MMP3-ddF14r | 106 |
| 3.2.2 Sequencing results                               | 107 |
| 3.2.2.1 Polymorphism detected using MMP3-ddF2(r)       | 107 |
| 3.2.2.2 Polymorphism detected using MMP3-ddF5(f)       | 109 |
| 3.2.2.3 Polymorphism detected using MMP3-ddF8(r)       | 110 |
| 3.2.2.4 Polymorphism detected using MMP3-ddF10(r)      | 111 |
| 3.2.2.5 Polymorphism detected using MMP3-ddF13(f)      | 112 |
| 3.2.2.6 Polymorphism detected using MMP3-ddF14(r)      | 112 |
| 3.3 Discussion   | 115 |

#### **Chapter 4: Linkage Disequilibrium of Markers in MMP-3**

|   |     |
|---|-----|
| 4.1 Introduction                        | 123 |
| 4.1.1 Random genetic drift              | 124 |
| 4.1.2 Natural selection                 | 124 |
| 4.1.3 Variable rates of mutation        | 124 |
| 4.1.4 Recombination                     | 126 |
| 4.1.5 Demographic factors               | 126 |
| 4.1.6 Gene conversion                   | 127 |
| 4.2 Linkage disequilibrium measurements | 127 |
| 4.3 Study design                        | 131 |
| 4.4 Results                             | 132 |
| 4.5 Discussion                          | 143 |

#### **Chapter 5: Association Studies of MMP Polymorphisms in Relation to CAD**

|  |     |
|--|-----|
| 5.1 Introduction                         | 146 |
| 5.1.1 Association studies                | 150 |
| 5.1.1.1 Cohort studies                   | 150 |
| 5.1.1.2 Case-control studies             | 151 |
| 5.1.2 Linkage analysis                   | 153 |
| 5.1.2.1 Haplotype relative risk test     | 153 |
| 5.1.2.2 Transmission disequilibrium test | 154 |

|       |                |     |
|-------|----------------|-----|
| 5.2   | Results        | 155 |
| 5.2.1 | Study subjects | 155 |
| 5.2.2 | Data analysis  | 156 |
| 5.3   | Discussion     | 162 |

## **Chapter 6: Functional Analysis of the MMP-3 Promoter and Protein-DNA**

### **Interaction Assays**

|         |   |     |
|---------|---|-----|
| 6.1     | Introduction                                  | 165 |
| 6.2     | Transfection                                  | 168 |
| 6.2.1   | Chemical reagents                             | 168 |
| 6.2.1.1 | DEAE-Dextran                                  | 168 |
| 6.2.1.2 | Calcium phosphate co-precipitation            | 169 |
| 6.2.2   | Cationic lipids                               | 169 |
| 6.2.2.1 | Liposome mediated transfection                | 169 |
| 6.2.3   | Physical                                      | 170 |
| 6.2.3.1 | Electroporation                               | 170 |
| 6.2.3.2 | Direct micro-injection                        | 170 |
| 6.3     | Reporter gene assay                           | 171 |
| 6.3.1   | CAT reporter assay system                     | 172 |
| 6.3.2   | $\beta$ -galactosidase reporter assay system  | 172 |
| 6.3.3   | Firefly luciferase reporter assay system      | 173 |
| 6.4     | Protein interaction assay                     | 173 |
| 6.4.1   | EMSA  | 174 |
| 6.4.2   | Footprint analysis of DNA-protein complexes   | 174 |
| 6.4.2.1 | DNAse I footprinting                          | 175 |
| 6.4.2.2 | Methylation interference assay                | 175 |
| 6.4.3   | Supershift gel mobility shift assay           | 176 |
| 6.4.4   | Sequence-specific DNA affinity chromatography | 176 |
| 6.5     | Results                                       | 182 |
| 6.6     | Discussion                                    | 190 |

## **Chapter 7: Overview and Possible Further Studies**

|     |          |     |
|-----|----------|-----|
| 7.1 | Overview | 193 |
|-----|----------|-----|



|                   |   |             |
|-------------------|---|-------------|
| 7.2               | Further work  | 199         |
| 7.2.1             | Electrophoretic mobility shift assay and DNA I footprinting | 199         |
| 7.2.2             | Affinity Chromatography                                     | 199         |
| 7.2.3             | Functional studies  | 200         |
| 7.2.4             | Association studies   | 202         |
| 7.2.5             | Conclusion  | 203         |
| <b>References</b> |   | <b>204</b>  |
| <b>Appendix</b>   |   | <b>i-ix</b> |

## List of Figures

### Chapter 1

|  |    |
|--|----|
| Figure 1.1: Diagram comparing two types of atherosclerotic lesions                       | 2  |
| Figure 1.2: Diagram of heart and coronary arteries                                       | 3  |
| Figure 1.3: Response to injury theory  | 6  |
| Figure 1.4: Schematic diagram of basic structural domains of MMPs                        | 21 |
| Figure 1.5: Ribbon diagram of the complex of TIMP-1 and the<br>catalytic domain of MMP-3 | 27 |
| Figure 1.6: Flow chart representing the different regulatory events<br>of MMPs           | 28 |
| Figure 1.7: Schematic diagram of the MMP-3 gene  | 36 |

### Chapter 2

### Chapter 3

|  |     |
|--|-----|
| Figure 3.1: Flow chart showing the steps involved in ddF           | 82  |
| Figure 3.2: Schematic diagram of MMP-3 gene variants               | 100 |
| Figure 3.3: Autoradiograph of ddF for the 5A/6A polymorphism       | 101 |
| Figure 3.4: Autoradiograph of ddF using MMP3-ddF2                  | 102 |
| Figure 3.5: Autoradiograph of ddF using MMP3-ddF5                  | 103 |
| Figure 3.6: Autoradiograph of ddF using MMP3-ddF8                  | 104 |
| Figure 3.7: Autoradiograph of ddF using MMP3-ddF10                 | 104 |
| Figure 3.8: Autoradiograph of ddF using MMP3-ddF13                 | 105 |
| Figure 3.9: Autoradiograph of ddF using MMP3-ddF14                 | 106 |
| Figure 3.10: Autoradiograph of direct sequencing using MMP3-ddF2   | 107 |
| Figure 3.11: Chromatograph of A-1986/G polymorphism                | 108 |
| Figure 3.12: Autoradiograph of direct sequencing using MMP3-ddF5   | 109 |
| Figure 3.13: Chromatograph of A-709/G polymorphism                 | 110 |
| Figure 3.14: Chromatograph of G-376/C polymorphism                 | 111 |
| Figure 3.15: Chromatograph of G+802/A polymorphism                 | 113 |
| Figure 3.16: Chromatograph of G+952/A polymorphism                 | 114 |
| Figure 3.17: MMP-3 sequence data from human, rabbit, rat and mouse | 118 |

## **Chapter 4**

|   |     |
|---|-----|
| Figure 4.1: Bst NI digest for the MMP-3 promoter polymorphism A-1986/G          | 132 |
| Figure 4.2: Xmn I digest for the 5A/6A MMP-3 promoter polymorphism              | 133 |
| Figure 4.3: Hpa II digest for the MMP-3 promoter polymorphism A-1346/C          | 134 |
| Figure 4.4: Eco R1 digest for MMP-3 promoter polymorphism A-709/G               | 135 |
| Figure 4.5: Pfl I digest for the MMP-3 promoter polymorphism C-376/G            | 136 |
| Figure 4.6: Taq <sup>α</sup> I digest for the MMP-3 exon 2 polymorphism +802 bp | 137 |
| Figure 4.7: Psh AI digest for the MMP-3 exon 2 polymorphism +952 bp             | 138 |

## **Chapter 5**

## **Chapter 6**

|   |     |
|---|-----|
| Figure 6.1: Circular map of the pRL-TK vector   | 178 |
| Figure 6.2: Circular map of the pGL3 vectors  | 179 |
| Figure 6.3: Flow chart presenting the method used to produce reporter gene constructs | 180 |
| Figure 6.4: Flow chart showing the general scheme used for transient transfection     | 181 |
| Figure 6.5: Schematic diagram representing the two approaches                         | 183 |
| Figure 6.6: Schematic diagram of reporter gene constructs used                        | 184 |
| Figure 6.7: Transient reporter gene assays in transfected MALU cells                  | 187 |
| Figure 6.8: DNA electrophoretic mobility shift assay                                  | 189 |

## **Chapter 7**

|  |     |
|--|-----|
| Figure 7.1: Diagram showing the progression of atherosclerosis | 195 |
| Figure 7.2: Diagram representing stable and unstable plaques   | 196 |
| Figure 7.3: Flow chart illustration further studies            | 201 |

## List of Tables

### Chapter 1

|   |    |
|---|----|
| Table 1: Candidate genes for coronary vascular diseases                       | 14 |
| Table 1.2: Members of the MMP family  | 18 |
| Table 1.3: Biological and pathological processes in which MMPs are implicated | 34 |
| Table 1.4: Factors specific to MMP-3 synthesis                                | 37 |
| Table 1.5: Normal and pathological cells and tissues associated with MMP-3    | 38 |

### Chapter 2

|   |    |
|---|----|
| Table 2.1: Conditions of PCR in mutation scanning                   | 50 |
| Table 2.2: PCR primer sequences used in mutation scanning           | 50 |
| Table 2.3: Primers used in ddF reaction                             | 55 |
| Table 2.4: Specimen requirements                                    | 58 |
| Table 2.5: PCR primers utilised in association study                | 63 |
| Table 2.6: Enzyme and digest conditions used                        | 64 |
| Table 2.7: Oligonucleotides used to create reporter gene constructs | 67 |
| Table 2.8: Oligonucleotide probes used for EMSA                     | 77 |

### Chapter 3

|  |     |
|--|-----|
| Table 3.1: Starting materials for mutation detection, there advantages and disadvantages | 98  |
| Table 3.2: Qualitative data analysis of ddF autoradiographs                              | 101 |
| Table 3.3: Three MMP-3 promoter variants showing linkage disequilibrium                  | 117 |
| Table 3.4: Haplotypes expressed by the seven polymorphism in linkage disequilibrium      | 117 |
| Table 3.5: Transcription factor binding sites identified in the two haplotypes           | 119 |

### Chapter 4:

|   |     |
|---|-----|
| Table 4.1: Haplotype frequencies                        | 140 |
| Table 4.2: Linkage disequilibrium between polymorphisms | 141 |
| Table 4.3: Linkage disequilibrium between polymorphisms | 142 |

**Chapter 5:**

|   |     |
|---|-----|
| Table 5.1: Characteristics of the subjects  | 156 |
| Table 5.2: Relationship between polymorphisms and number of >50% stenotic<br>Coronary arteries          | 159 |
| Table 5.3: Analysis of individual polymorphisms in relation to previous MI                              | 160 |
| Table 5.4: Stepwise logistic regression analysis of allele and haplotypes in<br>Relation to previous MI | 161 |

**Chapter 6:**

|   |     |
|---|-----|
| Table 6.1: Transcription factors  | 167 |
| Table 6.2: Transcriptional activity between each biallelic polymorphism | 186 |
| Table 6.3: Reporter assays on MALU cells                                | 188 |

**Chapter 7:**

## List of Abbreviations

|       |   |
|-------|---|
| A     | Adenosine   |
| aa    | Amino acid  |
| ACE   | Angiotensin-converting enzyme                     |
| ADAMs | A Disintergrine And Metalloproteinases            |
| AP-1  | Activator protein-1                               |
| ApoE  | Apolipoprotein E                                  |
| APS   | Ammonium persulphate                              |
| AMI   | Acute myocardial infarction                       |
| AMP   | Adenosine monophosphate                           |
| APMA  | Aminophenyl mercuric acid                         |
| ARMS  | Amplification refractory mutation system          |
| ATP   | Adenosine triphosphate                            |
| BESS  | Base excision sequence scanning                   |
| BMP 1 | Bone morphogenic protein 1                        |
| BMI   | Body mass index                                   |
| C     | Cytosine  |
| CAD   | Coronary artery disease                           |
| CCM   | Chemical cleavage of mismatch                     |
| CD    | Coronary disease                                  |
| CHD   | Coronary heart disease                            |
| CTP   | Cytosine triphosphate                             |
| CVD   | Cardiovascular disease                            |
| ddA   | Dideoxy adenosine triphosphate                    |
| ddC   | Dideoxy cytosine triphosphate                     |
| ddF   | Dideoxy fingerprinting                            |
| ddG   | Dideoxy guanine triphosphate                      |
| ddT   | Dideoxy thymine triphosphate                      |
| DGGE  | Denaturing gradient gel electrophoresis           |
| DHPLC | Denaturing high-performance liquid chromatography |
| DMS   | Dimethylsulphate                                  |
| DMSO  | Dimethylsulphoxide                                |

|               |  |
|---------------|--|
| DNA           | Deoxyribonucleic acid                            |
| DS            | Direct sequencing                                |
| ds            | Double stranded                                  |
| DZ            | Dizygote   |
| E             | Glutamic acid                                    |
| ECM           | Extracellular matrix                             |
| <i>E.coli</i> | <i>Escherichia coli</i>                          |
| e.g.          | For example                                      |
| EGF           | Epidermal growth factor                          |
| EMSA          | Electrophoretic mobility shift assay             |
| G             | Guanine  |
| GAWTS         | Genomic amplification with transcript sequencing |
| HA            | Heteroduplex analysis                            |
| HDL           | High density lipoprotein                         |
| HRR           | Haplotype relative risk                          |
| IDDM          | Insulin-dependant diabetes mellitus              |
| K             | Lysine   |
| LD            | Linkage disequilibrium                           |
| LDL           | Low-density lipoprotein                          |
| i.e.          | That is  |
| IL-1          | Interleukin 1                                    |
| IMT           | Intima-media thickness                           |
| MADGE         | Microplate array diagonal gel electrophoresis    |
| MBS           | Molecular biology shortcuts                      |
| MDE           | Mutation Detection Enhancement                   |
| MI            | Myocardial infarction                            |
| MMP           | Matrix metalloproteinase                         |
| MTHFR         | Methylenetetrahydrofolate reductase              |
| MZ            | Monozygote                                       |
| NIDDM         | Non Insulin-dependant diabetes mellitus          |
| OxLDL         | Oxidised low-density lipoprotein                 |
| PAGE          | Polyacrylamide gel electrophoresis               |
| PCR           | Polymerase chain reaction                        |

|               |   |
|---------------|---|
| PDGF          | Platelet-derived growth factor                                  |
| PEA3          | Polyomavirus enhancer A-binding protein-3                       |
| PON           | Paraoxonase   |
| RNA           | Ribonucleic acid  |
| (mRNA)        | Messenger ribonucleic acid                                      |
| RFLP          | Restriction fragment length polymorphism                        |
| SDS           | Sodium dodecylsulfate   |
| ss            | single stranded   |
| SNP           | Single nucleotide polymorphism                                  |
| SPRE          | Stromelysin-1 platelet derived growth factor-responsive element |
| SPBP          | SPRE-binding protein  |
| SSCP          | Single stranded conformation polymorphism                       |
| STR           | Short tandem repeats  |
| T             | Thymine   |
| t-PA          | Tissue-type plasminogen activator                               |
| TAE           | Tris acetate EDTA buffer  |
| TBE           | Tris borate EDTA buffer   |
| TDT           | Transmission disequilibrium test                                |
| TEMED         | NNN,N'Tetramethylethylenediamine                                |
| TF            | Transcription factor  |
| TFBS          | Transcription factor binding site                               |
| TGGE          | Temperature gradient gel electrophoresis                        |
| TIMPs         | Tissue inhibitors of metalloproteinases                         |
| TNF- $\alpha$ | Tumour necrosis factor-alpha                                    |
| TRE           | Tumour promoter-responsive element                              |
| u-PA          | Urokinase plasminogen activator                                 |
| UK            | United Kingdom  |
| U.V.          | Ultra-violet light  |



## **Chapter 1:**

### **Introduction**

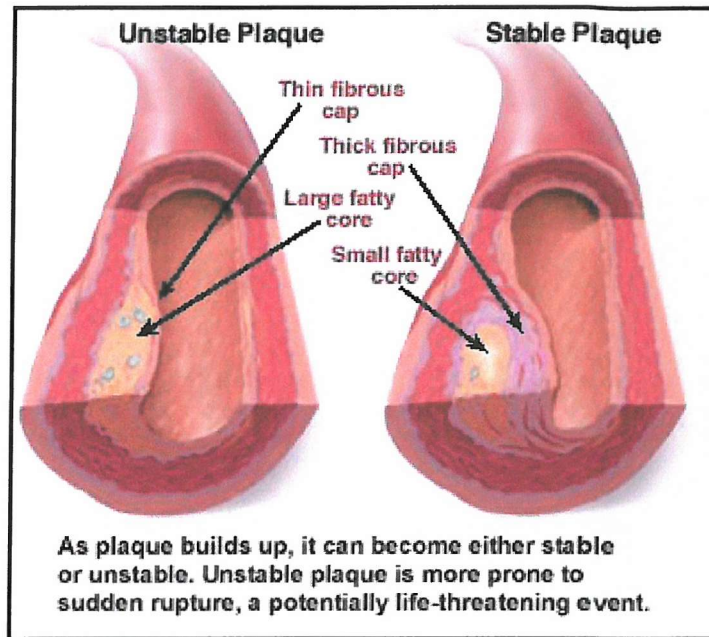
#### **1.1 Atherosclerosis**

Atherosclerosis is a multifactorial disease caused by genetic and environmental factors, resulting in the hardening of medium to large sized arteries. It underlies the majority of myocardial infarction, stroke and peripheral vascular disease, which are the leading causes of mortality and morbidity in the industrialized countries.

Atherosclerosis is a type of arteriosclerosis affecting the major arteries of the body. The name is derived from the Greek words athero (meaning gruel or paste) and sclerosis (hardening) and was coined by Marchand in 1904 <sup>1</sup>. It is an inflammatory condition that can be defined essentially as 'a variable combination of changes of the intima of arteries, consisting of the focal accumulation of lipids, complex carbohydrates, blood products, fibrous tissue and calcium deposits and is associated with medial changes' <sup>1</sup>. The structure and composition of an advanced atherosclerotic lesion represents the results of three fundamental biological processes:

- 1) the accumulation of intimal smooth muscle cells, macrophages and T-lymphocytes;
- 2) the production of large amounts of connective tissue matrix, such as collagen, elastic fibres and proteoglycans by proliferating smooth muscle cells;
- 3) and the accumulation of lipids, cholesteryl esters and free cholesterol within the cell and in the surrounding tissue <sup>1</sup>.

The amount of tissue formed in the lesion by each of the above named processes varies greatly. Lesions can therefore be dense and predominantly fibrous (sclerotic) or they may contain large amounts of lipid and necrotic debris (Figure 1.1). Atherosclerotic lesions demonstrate combinations of each of these characteristics. The distribution of lipid and connective tissue determines whether the lesion is stable or unstable. Unstable plaques are at a higher risk of rupture and thrombosis, where as stable plaques are associated with higher grade stenosis <sup>2,3</sup>.



**Figure 1.1:** - Diagram comparing the two types of atherosclerotic lesions: stable and unstable. The unstable plaque has a well preserved lumen as the plaque grows initially outwards, having a large lipid core and a thin fibrous cap. In contrast, the stable plaque has a relatively thick fibrous cap that protects the lipid core from contacting with blood. This diagram was taken directly from Atherosclerosis Heartcentre online.

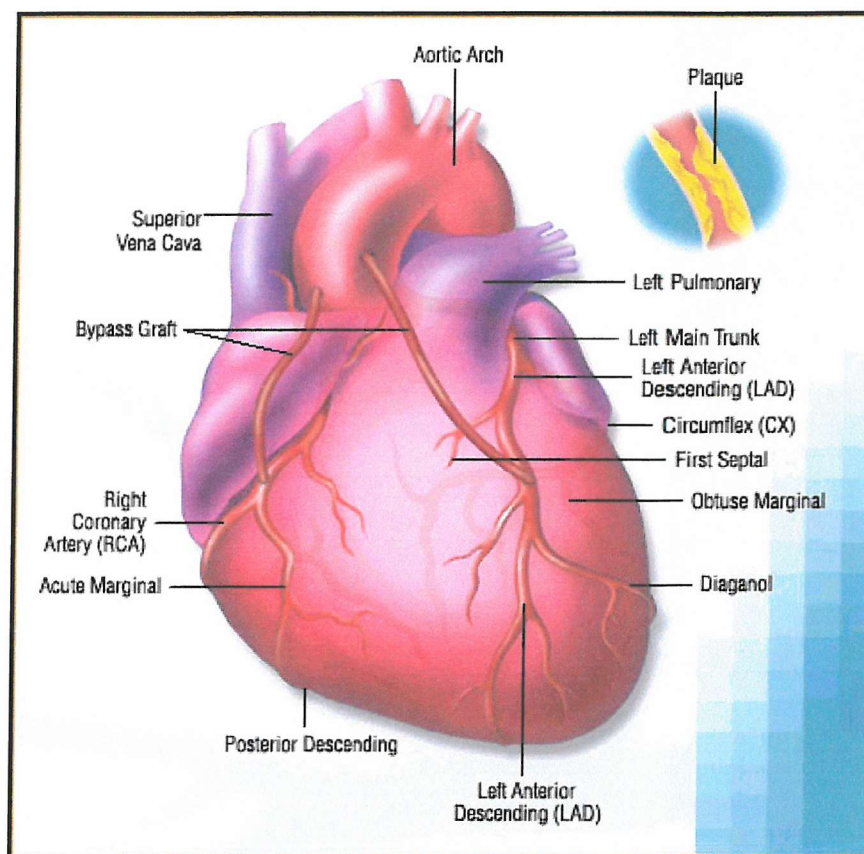
The circulation of blood is essential for life, as it transports chemical substances to and from the individual cells that comprise the entire human body. These chemical substances include oxygen, carbon dioxide, proteins and nitrogenous waste, carbohydrates, fats and various hormones that control the metabolic processes of the body. The heart and blood vessels are the means by which blood is transported around the body. The heart muscle has incredible endurance, such that it is able to contract billion of times in an average lifetime. However, like other parts of the human body, it is vulnerable and can break down. Heart disease has been known since A.D. 1500, but it was only with the introduction of autopsies that it was discovered that the heart could be chronically diseased and still allow life to continue for many years. Autopsies were introduced slowly; due to prejudices of ancient and medieval time and were finally permitted in the early part of the sixteenth century. Even so, it was not until the mid twentieth century, with the introduction of classification of the different forms of heart disease, that medical advances were achieved. There are many causes and types of heart disease, some caused by genetic inheritance or infection while others are due to

poor diets and lifestyles. Nonetheless, there are only a few important ones that are listed below, some of which are uncommon.

- 1) Congenital heart disease
- 2) Rheumatic heart disease (a consequence of rheumatic fever)
- 3) Hypertensive heart disease (secondary to high blood pressure)
- 4) Coronary heart disease (caused by the narrowing and obstruction of blood flow to the heart muscle)

### 1.1.1 Coronary Heart Disease

Coronary heart disease (CHD) is the most common cause of cardiac disability and death in the world and is the most common cause of sudden death.



**Figure 1.2:** - Diagram of the heart and the coronary arteries that supply it. Taken directly from the web site [www.guidant.com](http://www.guidant.com). In the top right hand, side of the picture is a diagram of an atherosclerotic coronary artery.

CHD is caused by the gradual narrowing (stenosis) of the coronary arteries that supply blood to the heart muscle. Stenosis occurs due to the build up of fatty plaques, atherosclerosis, which generally peak in the fourth and fifth decades of life. The consequence of this is angina pectoris (a choking chest pain), coronary thrombosis (the clotting of blood in one of the coronary arteries) and eventually death (infarction) of part of the heart muscle.

#### **1.1.1.1 Angina Pectoris**

Angina pectoris is a pain in the centre of the chest, often described as tightness or dullness, which can spread up to the jaw, neck and shoulders and radiate down the arms. It occurs when the demand for blood by the heart exceeds the supply from the coronary arteries and is generally caused by coronary artery atheromas. These attacks are brought on by factors that increase the workload of the heart such as physical activity, exposure to cold, emotional upset or the consumption of a heavy meal. The more severe the reduction of coronary flow reserve to a portion of the myocardium, the greater the severity of angina. Stable angina is the most common form of angina pectoris. Patients that suffer this form of angina may have established stenosis of the coronary arteries and can be managed, without immediate threat to life, with cardiac medication. Attacks are caused by coronary artery spasms around the atheromatous plaques. In unstable angina, there is an increase in the severity of the disease as a result of progress of stenosis. The pains may occur more frequently and be of longer duration than with stable angina, and often occur at rest. The pathogenesis of this form of angina is similar to that of myocardial infarction (MI) and may progress to MI or sudden death. The cause of unstable angina is thought to be an acute change in the coronary circulation.

#### **1.1.1.2 Acute Myocardial Infarction**

Acute myocardial infarction (AMI) is the death of cardiac tissue as a result of cardiac ischaemia. It is usually caused by partial or total occlusion of the coronary vessel(s) supplying the region of cardiac muscle threatened by infarction. It can also be caused by hypotension (low blood pressure), hypoxia (low oxygen), or cardiac arrest and also but rarely cardiac arrest spasms (CA spasms). The mortality of AMI is approximately

30-35% with 50% of these deaths occurring within the first hour of onset from ventricular fibrillation <sup>4</sup>. The more tissue infarction suffered by a patient the greater the chances that he or she will suffer another. The more coronary arteries affected by atherosclerotic plaques, the greater the probability of an AMI.

### **1.1.2 Progression of Atherosclerosis**

Only in man does atherosclerosis have widespread clinical significance. Lesions have been observed in large mammals that live for a long time, such as elephants <sup>5</sup>. They are generally uncommon in birds, with the exception of turkeys, where spontaneous aortic ruptures are quite common with hypertension <sup>6</sup>. Atherosclerotic plaques have been observed in Egyptian mummies from ~1580 BC and other ancient civilizations <sup>7,8</sup>. However, only in the twentieth century has cardiovascular disease been of epidemic proportion. This increase is not just due to unhealthy lifestyles, more individuals are living to advanced ages when heart disease is more prevalent.

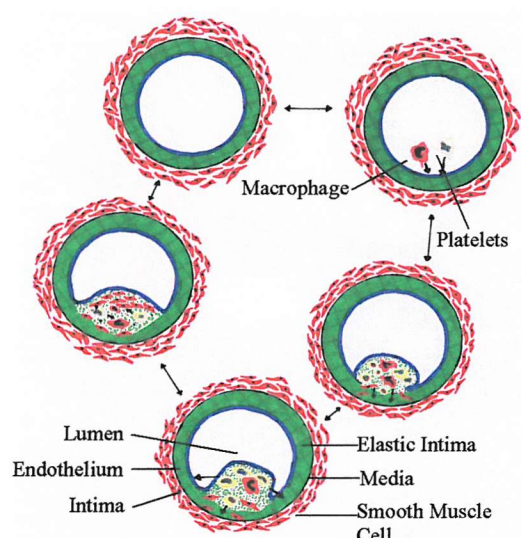
Atherosclerosis begins in infancy and early childhood <sup>9</sup>. Denoted “fatty streak”, it is recognised as being slightly yellowish in colour and longitudinal in shape in the innermost layer of the artery wall, the tunica intima. The fatty streak is essentially an increase in the number of intimal macrophages and the appearance of macrophages filled with lipid droplets (foam cells) intermingled with T-lymphocytes. Exactly how atherosclerosis begins or what causes it is not yet understood, however a number of theories have been advanced.

#### **1.1.2.1 Response to Injury Hypothesis**

Response to injury is one such theory. In this theory, pathogenesis is thought to be related to endothelial injury, with increases in lipid permeability, smooth muscle proliferation, macrophage emigration and proliferation of the intima <sup>10</sup>. Figure 1.3 below gives a brief outline of this idea. In essence, it proposes that the lesions are initiated as a result of injury or by an alteration in the endothelial lining <sup>10</sup>. Viral infection, homocysteine, toxin or oxidised low-density lipoproteins (OxLDL) can all cause such changes. This injury prompts an increase in lipoproteins and specific glycoproteins. Monocytes and T-lymphocytes attach to these molecules and migrate



between the endothelial cells, a response initiated by growth factor regulatory molecules and chemoattractants, which are thought to be released by the endothelial cells, leukocytes and the underlying smooth muscle cells. Monocytes then move further into the artery wall becoming macrophages, some ingesting LDL and forming foam cells. At the same time, smooth muscle cells move out of the medial layer and into the intima layer. All these components, (foam cells, macrophages, T-lymphocytes, lipid droplets and the smooth muscle cells (SMC)) interact to form a fatty layer of tissue, commonly termed the 'fatty streak' lesion<sup>9</sup>. As the lesion progresses, SMCs synthesise connective tissue, specifically a fibrous cap of collagen. A necrotic centre is formed in the lesions, filled with cellular debris and large deposits of external lipids, essentially cholesteryl esters and free cholesterol calcium deposits and blood components<sup>10</sup>.



**Figure 1.3:** - Response to injury hypothesis: Injury of the endothelium leads to an increase in lipid permeability, smooth muscle proliferation and migration of macrophages. The accumulation of these cells results in a fatty streak formation, which will progress to the more mature atheroma as illustrated here in the diagram.

Initially atherosclerotic lesions grow outward for a considerable part of their formation, although the more advanced lesions and fibrous plaques do encroach into the lumen effecting blood flow. The lesions clinically manifest themselves as atherosclerotic plaques in mid-late adulthood. It is the rupture of these atherosclerotic plaques, which cause thrombus formation, and complete occlusion of the coronary arteries. This then

leads to AMI. In the early hours of onset can be dissolved by thrombolytic agents streptokinase and TPA (total plasminogen activator), and hence rescue the threatened myocardium.

#### **1.1.2.2 Benditt and Benditt Monoclonal Theory**

Another theory 'Benditt and Benditt Monoclonal Theory of Smooth Muscle Cell Proliferation' suggests that smooth muscle proliferation initiates the plaque formation. This idea is similar to the pathogenesis of neoplastic tumours.

#### **1.1.2.3 Lipoproteins**

Lipoprotein oxidation has also been proposed to be the primary cause of atherogenesis. When excess amounts of lipoproteins, in particular the cholesterol low-density lipoproteins (LDL), occur in the blood, they produce toxins that damage the endothelial lining of the arterial wall. The lipoproteins get trapped within the artery wall where they accumulate and become oxidised. When oxidised they have a number of pathogenic effects, a few are listed below <sup>11</sup>: -

- 1) When oxidised lipoproteins become cytotoxic. In this form, they promote the secretion of cytokines by endothelial cells and kill smooth muscle cells and macrophages.
- 2) Oxidised lipoproteins are taken into macrophage cells by scavenger receptors. The macrophages then fill up with lipids to become foam cells.
- 3) Lipoproteins affect the migration of leukocytes.
- 4) Some oxidised lipoproteins are antigenic and so can cause immune reactions in the lesions.

Lipoprotein a, a type of lipoprotein, is postulated to have an influence on the progression of atherosclerosis, by interacting with the endothelium and also promoting thrombotic complications of atherosclerosis. Studies have shown that this protein blocks fibrinolysis in vitro. This results in the deposition of large amounts of fibrin and platelets in the blood vessel wall, which promotes thrombus accumulation. It inhibits

the production of plasmin. One of the functions of plasmin is to activate transforming growth factor (TGF), a factor that functions by blocking SMC growth <sup>12,13</sup>.

### **1.1.3 Macrophage and Smooth Muscle Cells**

Macrophages and SMCs are thought to play a major role in the progression of atherosclerosis. An assembly of macrophages and SMCs are known to be embedded in a complex extracellular matrix, which comprises the greater part of intermediate lesions. The fibrous cap in advanced lesions are also made up of SMCs, collagen fibrils, proteoglycans, macrophages and T-lymphocytes <sup>14</sup>.

#### **1.1.3.1 Macrophages**

Macrophages are derived from haemopoietic stem cells in bone marrow. They circulate in the blood as monocytes before migrating through the endothelium into connective tissue, where they either attach themselves to matrix fibres or stay motile. While there, they can undergo a limited number of mitotic divisions before they eventually die. Their primary function is phagocytic, they ingest particles such as bacteria, OxLDL and other foreign bodies. They can dispose of dead or moribund cells prior to tissue regeneration. They also secrete a number of biologically active materials, such as essential growth factors, chemotactic agents and matrix metalloproteinases (MMPs), that participate in inflammation, repair and regeneration of damage tissue as well as the maintenance of normal cell proliferation and differentiation in healthy tissue throughout the body. Macrophages are thus diverse in function and their activities are thought to be determined largely by the tissue they reside in <sup>15-17</sup>.

#### **1.1.3.2 Smooth Muscle Cells**

Smooth muscle, also known as non-striated or involuntary muscle, is found in all systems of the body. They are typically found in the walls of tubular structure and hollow viscera where they operate by regulating the diameter, for example in blood vessels and branches of the bronchia in the lungs. SMCs are derived from the medial in blood vessels. Originally their functional activity was only thought to be contraction.



SMCs however, are now known to also synthesize a number of extracellular substances such as elastins, collagens and proteoglycans, which form a reticular layer that bridges the gap between adjacent cells, thus providing mechanical continuity throughout the smooth muscle layer. These fibres interweave with the adjacent tissue<sup>18</sup>. SMCs and fibroblasts are the principal cells that produce connective tissue components in the body. SMCs are, therefore, thought to play a major role in the development of atherosclerotic lesion. This is because proliferation of these cells in the intima layer of the arterial wall is thought to help form the intermediate and advanced lesions of atherosclerosis. A small number of SMCs have been observed in plaques prone to rupture, i.e., those characterised by thin fibrous caps and a large lipid core,<sup>10,17,19</sup>. The contribution of SMCs to the maintenance and repair of the extracellular matrix is thought to be critical to plaque stability. It is postulated that the lack of SMCs in these unstable plaques is due to the production of cytokines by both macrophages and T-lymphocytes. These cytokines include fas ligand gamma interferon, tumour necrosis factor and interleukin-1<sup>10,17</sup>.

Atherosclerotic plaques are white-yellow lesions and can reach up to 1.5cms in diameter. Clinical effects are manifested principally in medium sized muscular arteries such as the coronary, carotid, basilar and vertebral arteries supplying the cerebrum and the superficial femoral arteries in the lower extremities. Larger arteries such as the aorta and the iliac may also be involved. The advanced lesions can either (1) partially or totally occlude the lumen, obstructing blood flow; or (2) weaken the vessel walls, causing cracks or fissures in the lesions, leading to thrombosis or internal blood coagulation and aneurysms where the internal wall is locally dilated. Aneurysmal dilation usually occurs in larger arteries such as the aorta. Sometimes the plaques can break away from the walls, are washed downstream and get lodged in a small artery causing an embolism. If these small arteries are situated in the brain or heart, this can lead to strokes or MI respectively. It is the unstable plaques that are subject to rupture and thrombosis and represent the principal cause of MI and sudden ischaemic death.

#### **1.1.4 Classical and Newer Risk Factors**

Epidemiological studies in England, Europe and the USA found significant associations of morbidity and mortality from complications of atherosclerosis, and have led to the recognition of a number of risk factors: -

- 1) Increased concentration of blood cholesterol
- 2) Clinical diabetes mellitus
- 3) Hypertension
- 4) Obesity
- 5) Smoking
- 6) Age
- 7) Gender (male)
- 8) Family history
- 9) Stress
- 10) Lack of exercise
- 11) Low dietary levels of vitamin E

##### **1.1.4.1 High Cholesterol**

Increased cholesterol levels have for many years been known to influence the progression of atherosclerosis. There are several types of cholesterol, however the main offender is LDL. Macrophage cells ingest oxidised LDL (OxLDL) and become foam cells. OxLDL induces monocyte chemotaxis and promotes the secretion of cytokine by endothelial cells. It also inhibits the release of the potent vasodilator nitric oxide and increases the risk of blood clotting (hypercoagulable state) by preventing the activation of protein C and up-regulating the procoagulant known as tissue factor<sup>20,21</sup>.

Atherosclerotic plaques have increased levels of OxLDL and patients suffering from this progressive disorder have increased levels of antibodies to LDL. In the UK, the blood cholesterol levels are treated in order to produce a cholesterol level less than 5 mmol/l. However studies, particularly those involving Japanese patients, have shown that progression is decreased when the cholesterol levels are much lower and ideally patients should have as low a cholesterol level as possible.

#### **1.1.4.2 Clinical Diabetes Mellitus**

Both types of diabetes mellitus, i.e. Type I (insulin-dependent) and Type II (non-insulin dependent), diabetes promote atherosclerosis and thus cardiovascular diseases. Insulin replacement therapy for patients suffering from Type I diabetes mellitus was initially thought to be complete and adequate therapy in the same way as thyroxine provides for hyperthyroidism. However, over time it was realised that insulin-treated patients' life expectancy was still considerably reduced.

Much research has been carried out in this area, all of which has shown that the development of atherosclerosis is increased in patients who suffer from diabetes <sup>22,23</sup>. They are twice as likely to suffer from strokes and three to five times likely to have an MI. Diabetes is a particularly strong cardiovascular risk in women as they lose their premenopausal protection from coronary artery disease <sup>4</sup>. This increased cardiovascular risk is thought to be partly due to the effect of diabetes on hypertension, dyslipidaemia, hyperfibrinogenaemia and so forth, but may also be related to the direct effects of hyperglycaemia <sup>4,23</sup>.

#### **1.1.4.3 Hypertension**

Hypertension (high blood pressure) is associated with most vascular diseases as it is so common <sup>24</sup>. It accelerates the progression of atherosclerosis. The increase in blood pressure increases the shearing forces on the endothelial lining, which in turn promotes atherogenesis. It is also thought to increase death in patients with coronary heart disease through its effect on the kidney. Waste products and poor filtration of endotoxins can lead to the damage of the endothelial lining and thus promotes the atherosclerotic state. This may also worsen a hypertensive state by actively retaining fluid and sodium. Hypertension can also cause kidney damage, which in turn causes intracranial complications. It can also cause left ventricular hypertrophy as a result of increased systemic vascular resistance. The left ventricle hypertrophy is to compensate this increase in pressure <sup>4</sup>.

#### **1.1.4.4 Smoking**

Tobacco smoke is a major risk factor of coronary heart disease and cardiac arrest. Much research has been carried out examining the effects of smoking on human health<sup>25,26</sup>. Results have shown that smoking cigarettes in particular, but also pipes and cigars if the smoke is inhaled, increases the risk of atherosclerosis by 50% and advances the natural history by ten years<sup>27,28</sup>. Smoking is thought to be the strongest risk factor for premature atherosclerosis. Smoking massively increases and aggravates the progression of complications of peripheral arterial disease and doubles the risk of MI and strokes in patients with high blood pressure. There are several means by which smoking promotes the progression of atherosclerosis. Smoking directly damages the endothelium; increases vascular tone and platelet activation may promote the oxidation of LDL and decreases levels of free protein S<sup>4</sup>.

#### **1.1.4.5 Family History**

Coronary heart disease is often found in several members of the same family and so a family history can indicate possible genetic factors involved. Clinical trials have suggested that family history is perhaps the strongest risk factor of cardiovascular disease. This disease however, is so prevalent due to other risk factors, that it is difficult to treat family history alone as an independent risk factor.

Any one of these risk factors listed above can alone increase the progression of atherosclerosis. However, a combination of these risk factors significantly increases the susceptibility and progression of atherosclerosis.

#### **1.1.5 Heritability**

Heritability is the measure of a character/trait that is inherited rather than attributable to a non-heritable factor. In other words heritability ( $h^2$ ) is the ratio of the additive genetic variance ( $V_A$ ) to the total phenotypic variance ( $V_P$ ):

$$h^2 = V_A / V_P$$

It varies between 0 (purely environmental) and 1 (purely genetic).

Heritability is a method used to study genetic diseases, in particular complex diseases that have a strong genetic component but no simple Mendelian traits, such as insulin-dependent diabetes mellitus (IDDM) and ischaemic heart disease. Such diseases are caused by a combination of genetic and environmental factors and are termed incomplete penetrance. Comparisons of concordance in monozygotic (identical) and dizygotic (fraternal) twins can indicate a role of genetic factors in the causation of certain complex traits <sup>29</sup>.

#### **1.1.6 Genetic Risk Factors and Candidate Gene**

Relatives of an individual effected by a given disease or trait, have a greater probability of also being effected, than someone in the general population. Studies using families or sibs (nuclear families) are, therefore, typically used to find the genetic determinants of complex diseases. There are two types of studies used to determine these genes, linkage analysis and association studies. Both have played important roles in finding the mutations and therefore the genes that cause complex diseases <sup>30</sup>.

Linkage analysis localises putative trait genes and considers at the gene or DNA sequence at specific loci to be inherited together due to the physical proximity. Lod scores are used to measure the likelihood of genetic linkage between two loci. It is a robust technique for localizing disease genes.

A large number of genes (at least sixty) have been implicated in cardiovascular diseases (CVD) and atherosclerosis. Angiotensin-converting enzyme (ACE), apolipoproteins, paraoxonases (PON) and Factor V are but a few. Table 1.1 below represents a sample of the genes that have been explored as candidates.

**Table 1.1: Candidate Genes for Coronary Vascular Diseases**<sup>31</sup>

| GENE   | ABBREVIATION | CHROMOSOME |
|--|--------------|------------|
| $\alpha_{2A}$ adrenergic receptor                | ADRA2A       | 10q        |
| Aldosterone synthase                             | CYP11B2      | 8q         |
| Angiotensin-I Converting enzyme                  | ACE          | 17q        |
| Angiotensin-II type 1 receptor                   | AGTR1        | 3q         |
| Apolipoprotein (a)                               | LPA          | 6q         |
| Apolipoprotein B                                 | APOB         | 2p         |
| $\beta_1$ adrenergic receptor                    | ADRB1        | 10q        |
| Cholesteryl ester transfer protein               | CETP         | 16q        |
| Colony stimulating factor 1                      | CSF1         | 5q         |
| Early growth response protein 1                  | EGR1         | 5q         |
| Endothelin 1                                     | EDN1         | 6p         |
| Endothelin converting enzyme                     | ECE          | 1p         |
| Endothelin A receptor                            | EDNRA        | 4          |
| Endothelin B receptor                            | EDNRB        | 13q        |
| Fibrinogen $\alpha$                              | FGA          | 4q         |
| Fibrinogen $\beta$                               | FGB          | 4q         |
| Fibrinogen $\gamma$                              | FGG          | 4q         |
| Granulocyte-macrophage colony-stimulating factor | CSF2         | 5q         |
| Hepatocyte growth factor                         | HGF          | 7q         |
| Insulin-like growth factor 1                     | IGF1         | 12q        |
| Intercellular circulating adhesion molecule 1    | ICAM1        | 19q        |
| Interleukin 1- $\alpha$                          | IL1A         | 2q         |
| Interleukin 6                                    | IL6          | 7p         |
| Macrophage-scavenger receptor                    | MSR1         | 8p         |
| Matrix Gla protein                               | MGP          | 12p        |
| Microsomal triglyceride transfer protein         | MTP          | 4q         |
| NO synthase endothelial                          | NOS3         | 7q         |
| NO synthase inducible                            | NOS2A        | 17q        |
| P-selectin                                       | SELP         | 1q         |
| Paraoxonase                                      | PON1         | 7q21-q22   |
| Platelet-derived growth factor A                 | PDGFA        | 7p         |
| Platelet-derived growth factor A-receptor A      | PDGFRA       | 4q         |
| Scavenger receptor                               | CD36         | 7q         |
| Thrombin receptor                                | TR           | 5q         |
| Transforming growth factor $\beta$ 1             | TGFB1        | 19q        |
| Tumour necrosis factor $\alpha$                  | TNF          | 6p         |
| Vascular cell adhesion molecule-1                | VCAM1        | 1p         |
| Vitamin D receptor                               | VDR          | 12q        |

#### **1.1.6.1 Angiotensin-Converting Enzyme**

The ACE gene is located on chromosome 17q23 and encodes two isoenzymes (somatic ACE and germinal ACE isoenzymes). Somatic ACE is expressed in several tissues, including, vascular endothelium and renal epithelium cells. ACE, also known as kininase II plays an important role in blood pressure regulation. A number of studies have therefore looked at the ACE gene and its possible contribution into the development of coronary heart disease (CHD). Tiret et al (1992), identified an insertion (I)/deletion (D) polymorphism of 250bp in intron 16 of the ACE gene<sup>32</sup>. In a cohort study of patients with a history of MI and control individuals, Cambien et al (1992), observed a significant higher frequency of the D/D genotype in the patient group compared to the control group ( $p=0.007$ ). This was particularly evident in individuals with a low body mass index (BMI) and low plasma levels of ApoB ( $p=0.0001$ )<sup>33</sup>. Subsequent studies have shown a significant association of increased risk of myocardial infarction with D/D genotype as compared to the I/I genotype. Pfeffer et al (1992) showed that by administering an ACE inhibitor, the risk of heart failure and myocardial infarction (MI) was reduced<sup>34</sup>.

#### **1.1.6.2 Apolipoprotein E**

Apolipoprotein E (APOE) is a protein that is associated with the metabolism of triacylglycerol-rich lipoproteins and HDL (High-density lipoproteins). It also mediates the transport of cholesterol, phospholipids and their fatty acids to the liver and other parts of the body, thus functioning as a ligand for LDL receptors and LDL receptor related proteins. Three major isoforms of the human APOE gene have been identified by isoelectric focusing, APO\*E2, APO\*E3 and APO\*E4. These are encoded for by three alleles epsilon 2, 3 and 4, which combine to form six possible genotypes. The alleles differ in amino acid sequence at two sites, residue 112 and residue 158. At these sites APO\*E2, \*E3 and \*E4 comprise cysteine/cysteine, cysteine/arginine and arginine/arginine respectively<sup>35,36</sup>. These amino acid substitutions have profound functional consequences at both the cellular and molecular level. APO\*E3 is the most common alleles found in the human population and is therefore thought to be the 'normal' or wild-type isoform<sup>37</sup>. The APO\*E2 and APO\*E4 alleles are known to be dysfunctional. APO\*E2 is associated with the genetic disorder Type II

hyperlipoproteinemia and through this has been found to have both an increased and decreased effect on the risk of atherosclerosis<sup>38</sup>. APO\*E4 has been associated with a greater risk of CHD<sup>39</sup>. Patients with this allele were found to develop early calcification of the arteries and had an increase body mass index and increased blood levels of cholesterol and triglycerides<sup>40</sup>. A promoter polymorphism G-219/T upstream from the transcription start site has also been identified and is associated with MI. More specifically individuals carrying the -219/T allele are thought to have an increased risk of MI<sup>41</sup>. This polymorphism was also found to be independent of the presence of APOE isoforms E2/E3/E4. The E4 isoform is also associated with the risk of other diseases such as Alzheimer's, impaired cognitive function and reduced neurite outgrowth.

#### **1.1.6.3 Paradoxonase**

Paradoxonase (PON1) is located on chromosome 7q21-q22. It is a serum enzyme whose precise function is still poorly understood. PON1 has been hypothesised to prevent the formation of atherosclerotic plaques by destroying proinflammatory molecules and hence reducing the risk of CHD<sup>42,43</sup>. A common polymorphism has been identified at codon 192, which has been associated with an increased risk of CHD. Recently two genes similar to PON1 have been identified. Linked to PON1 on chromosome 7, they were named PON2 and PON3. Studies have identified common polymorphisms in these genes, which have been found to be associated with CHD both alone and combined with the variant in the PON1 gene located at codon 192. One such common variant is that found in the PON2 gene at codon 311 (Cys→Ser; PON2\*S).

#### **1.1.6.4 Factor V**

Factor V is well known for its involved in the extrinsic and intrinsic pathways of blood coagulation. In this pathway, it functions as an essential cofactor for the Xa-catalysed activation of prothrombin to the clotting enzyme thrombin<sup>44</sup>. However, Factor V also plays an important role in the anticoagulation system and has been shown to activate protein C. Patients with activated protein C resistance (APC resistance), are prone to thrombosis. A deficiency in a co-factor for activated protein C was identified as the cause for this tendency to thrombus. Factor V is thought to be this cofactor<sup>45</sup>. Bertina



et al (1994) identified a mutation in the Factor V gene sequence, which is believed to cause a deficiency in the activation of protein C and consequently cause an increase risk of thrombosis in patients. Located at position +1691, it occurs as a single nucleotide substitution from guanine to adenine and causes an amino acid change from arginine (CGA) to glutamine (CAA). This mutation is referred to as FV Leiden <sup>46</sup>. An increase risk of thrombosis can lead to a number of clinical disorders, such as myocardial infarction, strokes, pulmonary embolism and deep vein thrombosis.

Coronary disease is believed to be polygenic and the genes in the table above account for only a small percentage of the genes involved.

It is estimated that known genetic risk factors only account for about 50% of genetic liability of CHD, suggesting that other susceptibility or disease modifying genes are yet to be identified. Matrix metalloproteinases (MMPs) are a group of at least 26 matrix-degrading enzymes. They play a central role in degradation of the extracellular matrix during a number of biological processes. These include: morphogenesis, embryonic implantation and development, tissue resorption and wound healing. Enhanced activity of these proteins has been observed in a number of pathological conditions such as rheumatoid arthritis, tumour invasion and metastasis and atherosclerosis. It is believed that MMPs contribute to connective tissue dysfunction and the eventual plaque rupture that leads to thrombosis. The MMP-3 gene is an important member of the MMP family of proteinases and plays a key role in arterial wall remodelling. This enzyme and its gene have been the focus of considerable attention in the study of the pathogenesis of atherosclerosis and CHD <sup>31,47,48</sup>.

The purpose of this study is to extend further the analysis of the MMP-3 genomic region, with particular respect to genetic epidemiology and in vitro functional studies.

## 1.2 Matrix Metalloproteinases (MMPs)

MMPs are a family of tightly controlled enzymes that were originally described as degraders of extracellular matrix macromolecules (ECM) and basement membrane components including: Types II, IV, and IX collagen, proteoglycans, lamins, fibronectin, gelatins and elastins <sup>48</sup>. Each MMP has its own substrate preference and together all components of the extracellular matrix can be degraded by this class of enzymes. In addition, the MMPs are also able to influence many cellular functions such as cell migration and activate enzymes in their own family. At present over 22 members have been identified, which are listed in Table 1.2 <sup>48-50</sup>; -

**Table 1.2: Members of the MMP Family**

| MMP    | Common Name(s)                   |
|--------|----------------------------------|
| MMP-1  | Collagenase-1                    |
| MMP-2  | Gelatinase A                     |
| MMP-3  | Stromelysin-1                    |
| MMP-7  | Matrilysin                       |
| MMP-8  | Collagenase-2                    |
| MMP-9  | Gelatinase B                     |
| MMP-10 | Stromelysin-2                    |
| MMP-11 | Stromelysin-3                    |
| MMP-12 | Macrophage metalloelastinase     |
| MMP-13 | Collagenase-3                    |
| MMP-14 | MT1-MMP                          |
| MMP-15 | MT2-MMP                          |
| MMP-16 | MT3-MMP                          |
| MMP-17 | MT4-MMP                          |
| MMP-18 | Collagenase-4 ( <i>Xenopus</i> ) |
| MMP-19 | RASI-1                           |
| MMP-20 | Enamelysin                       |
| MMP-21 | XMMP ( <i>Xenopus</i> )          |
| MMP-22 | CMMP (chicken)                   |
| MMP-23 |                                  |
| MMP-24 | MT5-MMP                          |
| MMP-25 | MT6-MMP                          |
| MMP-26 | Endometase, Matrilysin-2         |
| MMP-27 |                                  |
| MMP-28 | Epilysin                         |

Table adapted from Sternlicht, M.D. and Werb, Z. 2001 <sup>51</sup>

The MMPs may be classified into the following groups:

- 1) Collagenases (MMP-1 [Interstitial collagenase], MMP-8 and MMP-13)
- 2) Gelatinases (MMP-2 [gelatinase A] and MMP-9 [gelatinase B])
- 3) Stromelysins (MMP-3, MMP-10, MMP-11,)
- 4) Membrane-Type MMPs (MT-MMP, MMP-14, -MMP-15, MMP-16 and MMP-17)
- 5) Other (MMP-7, MMP-12, MMP-19, MMP-20 etc)

The interstitial collagenase (MMP-1) was the first MMP to be discovered in 1962<sup>52</sup>. It was described as an enzyme that dissolved the tadpole tails in metamorphosis. Other MMPs were recognized through homology searches. Almost all collagenase and stromelysin genes are located in a gene cluster along the long arm of chromosome 11<sup>52,53</sup>. Stromelysin 3 (MMP-11) is an exception, its gene lies on chromosome 20. The two gelatinases MMP-2 and MMP-9 have been mapped to chromosome 16 and 20 respectively and MT-MMP-1 (MMP-14) gene resides on chromosome 14.

### 1.2.1 The Basic Structural Domains of MMPs

The primary structure of MMPs comprises of five domains. These include a signal peptide, a propeptide, a catalytic domain, a hinge region and a haemopexin/fibronectin-like COOH terminal domain<sup>49,50</sup>. Two additional domains (i.e. a fibronectin domain and a transmembrane domain) also exist, and are specific to certain classes of MMPs.

The signal peptide domain usually consists of 17-20 residues rich in hydrophobic amino acids. This section acts as a signal for secretion into the endoplasmic reticulum. With the exception of MMP-17 all the MMPs possess this signal peptide.

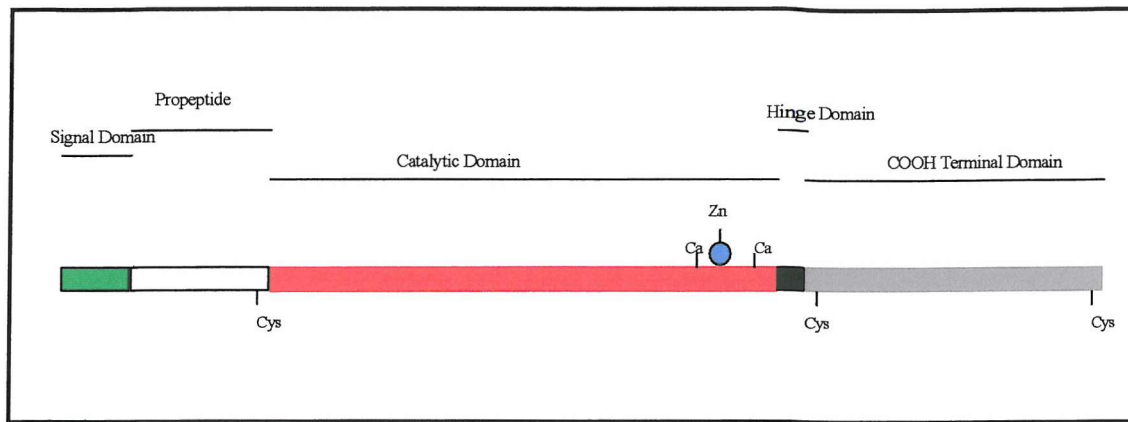
The propeptide region is generally 80 amino acids (aa) in length and possesses an N-Terminal of hydrophobic residues. There is a highly conserved sequence, PRCXXP, near the C-Terminal of this domain. This conserved region has a cysteine residue incorporated in it. Known as the “cysteine switch” it binds to a catalytic zinc atom in the catalytic domain and maintains the latency of the pro-enzyme<sup>54,55</sup>.

The catalytic domain is typically 160-170 residues in length. It contains a  $\text{Zn}^{2+}$  ion and 2-3 calcium ions at the active site, which are thought to be important in the stability and expression of enzyme activity. The C-Terminal end of the catalytic domain also contains a highly conserved zinc-binding motif HEXGHXXGXXHS/T and a conserved methionine. The methionine forms a unique 'Met-turn' structure. This domain is made up of five-stranded  $\beta$ -sheets, three  $\alpha$ -sheets and a bridge loop <sup>56,57</sup>.

The hinge region, also known as the proline-rich linker peptide, is between 0-75 residues in length and connects the catalytic domain to the haemopexin domain. This section of the MMP domain structure, typically possesses about 16 residues, which include a number of proline residues. MMP-9 is known to have the longest hinge region, while it is absent in MMP-7 <sup>52</sup>.

The haemopexin or fibronectin-like COOH terminal domain consists of approximately 200 residues with four repeats that resemble haemopexin or fibronectin, thus the name. All MMPs with the exception of MMP-7 have this structure. The terminal domain has a cysteine residue positioned at both ends. It is not thought to be essential for catalytic activity, but is believed to be involved in controlling substrate specificity for each enzyme. It is particularly important/essential for the collagenases to cleave triple helical interstitial collagens. The haemopexin domain is also required for the cell surface activation of pro-MMP2 by MT1-MMP <sup>58,59</sup>. Some molecules are also postulated to bind to this region such as the MMP inhibitors, tissue inhibitors of metalloproteinases (TIMPs). Figure 1.4 below represents a rough outline of the general structural features of MMPs.

In addition to the five common domain regions, the gelatinases (MMP-2 and MMP-9) contain a fibronectin insert in the catalytic domain sector. This insert comprises three-fibronectin type II repeats, which is postulated to facilitate the binding of the gelatinases to their substrate. The MT-MMPs also have an additional transmembrane domain, which gives this class of enzyme an insoluble characteristic, enabling them to anchor to the cell membrane <sup>49,50</sup>.



**Figure 1.4:** - Schematic diagram of the basic structural domains in MMPs.

MMPs are classified not only according to their structure but also to their substrate preference. As already mentioned they are involved in the degradation of ECM and basement membrane components. There is an overlap in substrate specificity; nearly all MMPs cleave gelatin, fibronectin and Types IV and V collagens to some extent. A few MMPs acting in tandem therefore have the potential to catalyse the complete degradation of the protein components of the basement membrane and extracellular matrix (ECM). In general however, collagenases degrade Type I–III collagens. They initiate the unwinding of the helix structure of the collagens. Gelatinases degrade Type IV collagens and finish the breaking down of collagens Type I–III initiated by the collagenases. Stromelysins degrade a wide range of ECM macromolecules and are also thought to activate other members of the family. The substrate specificity of MT-MMPs is less well known but it is suggested that they degrade the Lamin B proteins 49,50,52

MMPs are highly conserved and are approximately 70% similar in sequence. The promoter regions of the MMP genes have common features specifically cis-elements, that are important in gene regulation (see section 1.2.2). These include one or more copies of the activator protein-1 binding site and the polyomavirus enhancer A-binding site. Similarities are also present in the coding region of the MMP genes. In general, both collagenases and stromelysins comprise of ten exons and nine introns and extend 8-12 Kb in length. The gelatinases are significantly larger in length (26-27 Kb) and

contain three additional exons. These additional exons encode the three fibronectin-like inserts.

With the exception of MMP-2, all other MMPs are expressed at low levels in normal adult tissues, but these levels are increased during normal and pathological remodelling processes.

### **1.2.2 The Regulation of MMPs**

MMPs are generally not stored in cells (except MMP-8), there are a few exceptions such as macrophages and neutrophils. They are only synthesized and secreted into the matrix when functionally required for physiological or pathological tissue remodelling. MMPs are secreted by a number of different cell types such as SMC, macrophages, stromal cells, fibroblasts, keratinocytes and so forth. Although their regulation is not fully understood, the general consensus is that their expression is tightly controlled by three main mechanisms <sup>48</sup>:

- 1) The regulation of transcription
- 2) Activation of latent MMPs
- 3) Inhibition of MMP activity

MMPs are typically regulated in the same manner. Much research has been carried out to try and identify the exact mechanisms involved in MMP gene regulation. The following is a general out-line of the mechanisms involved in MMP regulation, which will be centred on the regulation of the MMP-3 gene. Figure 1.6 below, gives a schematic representation of the basic factors involved in MMP regulation.

#### **1.2.2.1 Transcriptional Regulation**

Synthesis and secretion of MMPs are initiated primarily by certain signals in the cells and are stopped or reduced by a decrease in the cell signal or the presence of a negative signal. These regulatory signals are the product of a number of inductor and suppressor agents. They work almost exclusively at the transcription level and act via the promoter region of the MMP genes. Gene expression of most MMPs at this initial step

is controlled by growth factors, cytokines, hormones and cellular regulators. Some chemical reagents have also been shown to induce MMP expression such as phorbol esters, cyclic AMP and actin stress fibre-disrupting drugs. Physical stress, heat shock, UV irradiation and oncogenic transformation can also initiate the transcription of MMP genes. A number of suppressor agents have also been identified that down-regulate the expression of MMP genes. These include transforming growth factor  $\beta$ , retinoic acid, oestrogen, progesterone and glucocorticoids <sup>52</sup>.

Though there are many factors that control the expression of the MMP genes, they are cell type specific. Most of these agents themselves are only secreted from the surrounding cells when they are needed. During an inflammatory response, certain cytokines are released from endothelial cells, SMC etc. These cytokine, such as interleukin-1 (IL-1), IL-6 and tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) trigger certain signal pathways, which leads to the transcription of MMPs <sup>50,60</sup>. Induction and suppression of the MMP genes by agents such as growth factors and cytokines can cause a 20-50 fold change in the mRNA and protein levels <sup>50</sup>.

Transcription of MMPs is initiated by the binding of transcription factors at specific sites in the promoter region of the gene. For example, an activator protein-1 (AP-1) binding site, two polyomavirus enhancer A-binding protein-3 (PEA3) binding sites and stromelysin-1 platelet derived growth factor-responsive element (SPRE) have been found in the MMP-3 gene promoter. The AP-1 site in MMP-3 is situated at -70-64bp from the start of transcription. This site is also known as the tumour promoter-responsive element (TRE). Though the exact mechanisms involved in activating AP-1 is unknown, it is thought that c-Jun and c-Fos transcription factors bind together by a leucine zipper mechanism to form a heterodimer. Treatment with phorbol esters increase the levels of Fos and Jun in cells. The effect of some growth factors and cytokines, such as Interleukin (IL-1) and epidermal growth factor (EGF), is also mediated by AP-1. Other proteins components such as Jun related proteins and Fos related protein Fra-1, also bind with c-Jun and c-Fos to form an activator protein (AP-1) that binds to this site <sup>48,61-63</sup>.

The two PEA3 binding sites are palindromically arranged –208bp from the start of transcription in MMP-3. They are bound by Ets transcription factors and are activated by c-Ets, Ha-Ras, v-Src and v-Mos<sup>63</sup>.

SPRE is positioned at –1573bp from the start of transcription and is required for efficient response to PDGF and other mitogens. The SPRE-binding protein (SPBP) is the transcription factor thought to bind at this site to enhance transcription. Studies have suggested that it interacts with c-Jun, forming a heterodimer to transactivate the SPRE site. Thus, cross-talk is thought to occur between different enhancer elements in the human MMP-3 gene<sup>62</sup>.

#### 1.2.2.2 Activation of Latent MMPs

MMPs are generally secreted from cells as inactive zymogens. The latent enzymes can be activated by a number of non-proteolytic agents and proteases. The non-proteolytic agents identified *in vitro* include: mercurial compounds such as HgCl and 4-aminophenylmercuric acid (APMA), thiol-modified agents such as HOCl, iodoacetamide, N-ethylmaleimide (NEM) and oxidized glutathione (GSSG) and denaturants such as urea, sodium dodecyl sulfate (SDS) and NaSCN. Several proteinases have also been identified as activators of latent MMPs; these include plasmin, furin and MMP-3<sup>52</sup>.

Plasmin is considered to be a potent activator of most MMPs. Its activation from plasminogen to plasmin is by two regulatory enzymes, uPA and tPA. Plasmin has a similar specificity to trypsin. It is a serine protease and has an active site catalytic triad, his-57, asp-102 and ser-195. Pro-MMP activation is believed to occur by the uPA/plasmin system<sup>50,64,65</sup>.

Activation of MMP zymogens involves several steps and begins with the disruption of a highly conserved unpaired cysteine residue in the N-terminal propeptide domain. This cysteine residue is coordinated with a zinc atom at the active centre. Termed the 'cysteine switch' both chemical and proteolytic enzymes can cause the disruption of this bond, causing the cysteine to fold back, thus breaking the cysteine-zinc bond<sup>66,67</sup>. During MMP activation, proteolytic cleavage, ahead of the cysteine residue, takes place



and is then followed by removal of the cysteine containing propeptide. MMPs at this point are thought to be partially active.

MMP activation can occur through intracellular, extracellular and cell surface-mediated mechanisms. The mechanisms involved in MMP activation are complex. Not all MMPs can be activated by the same agents and mechanisms. A so-called 'bait region', a short sequence found in the middle subunit of the propeptide domain in MMPs, dictates which proteinases can trigger the activation of the pro-enzyme. MMP-11 is activated intracellularly in the golgi network. Activation is mediated by the intracellular serine protease furin <sup>68</sup>. Both the MMP-1 and MMP-3 can be activated in an extracellular pathway involving the uPA/plasmin system <sup>67,69,70</sup>. Activation of MMP-3 in this manner can then lead to the activation of pro-MMP-9 in what is thought to be a protease-protease cascade. MMP, MT-1 MMP is able to activate MMP-2 and MMP-13 on the cell surface. Cell surface activation of MMP-2 involves the formation of a complex with TIMP-2 and MT-1 MMP. Cell-mediated activation of MMP-2 was first described in normal cells in response to concanavalin A treatment <sup>71</sup>. The activation complex is still unclear but it is postulated that pro-MMP-2 binds to MT1-MMP via the tissue inhibitor TIMP-2 <sup>59,72</sup>. TIMP-2 is thought to make a bridge between the two MMPs, thus forming a ternary complex. An alternate cell surface mediated activation of pro-MMP-2 is by way of the uPA/plasmin systems.

MMPs once activated, can activate other proenzymes in its family, thus forming a positive feedback loop. It should be noted that if MMPs are involved in the activation at any step of pro-MMPs, then inhibition by TIMP present in the surrounding medium could prevent activation of the latent enzyme.

### **1.2.3.3 Inhibition of MMP Activity**

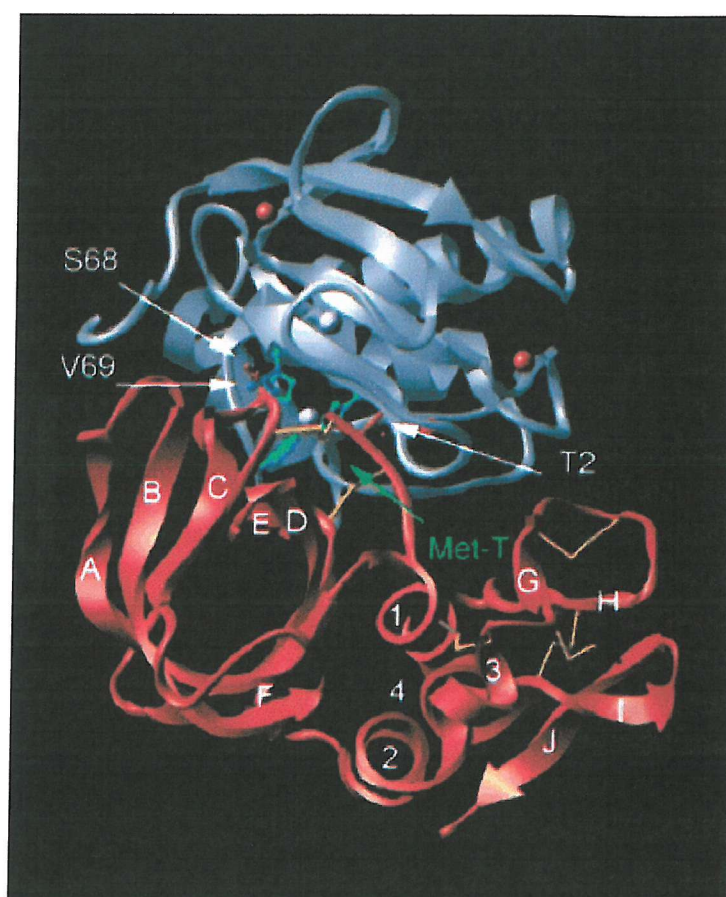
Prior to full activation, MMPs may be inhibited by, tissue inhibitors of metalloproteinases (TIMPs) and  $\alpha_2$ -macroglobulin.

$\alpha_2$ -macroglobulin is a plasma proteinase inhibitor consisting of four identical subunits of 180kDa. It inactivates MMPs by first cleaving one or more bonds comprising the

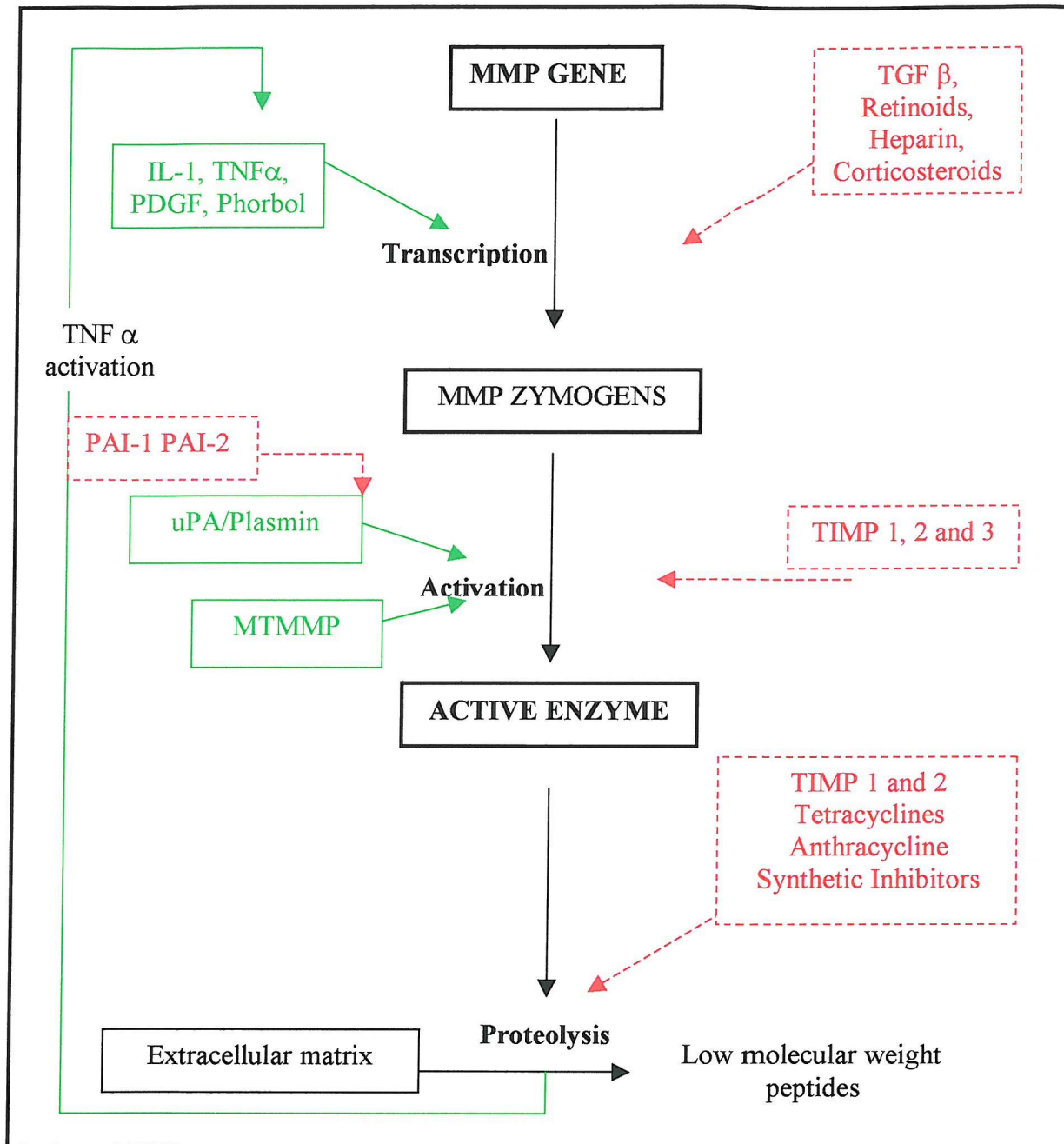
bait region. This triggers a series of events beginning with the initiation of a conformational change and ending with the entrapment of the proteinases. Trapping the proteinases does not fully block the enzyme's active site and so small substrate molecules can still be catalysed. The catalysis of larger substrate molecules is completely inhibited due to steric hindrance<sup>50</sup>. The  $\alpha_2$ -macroglobulin/MMP complexes are removed by scavenger receptor-mediated endocytosis,  $\alpha_2$ -macroglobulin inhibition is therefore irreversible and results in the complete removal of MMPs.  $\alpha_2$ -macroglobulin is abundant in plasma protein and is therefore thought to represent the major inhibitors of MMPs in tissue fluid.

TIMPs play an important role in regulating the activities of MMPs. In so doing, they control the rate of ECM degradation. They are the major endogenous regulators of MMP activation in tissue. Four TIMP have been identified to date (TIMP-1, -2, -3 and 4). They are thought to be approximately 40% identical in DNA sequence<sup>73,74</sup>. Their expression is regulated by growth factors and cytokines. TIMPs inhibit MMPs by forming a 1:1 stoichiometric complex. Inhibition by TIMPs is reversible. Disulphide bonds formed by cysteine residues in the N-terminal and C-terminals identified in both TIMP-1 and TIMP-2 are thought to be the active site at which inhibition occurs. NMR<sup>75</sup>, and X-ray crystallographic studies<sup>76</sup> were used to determine the exact interaction between MMP-3 and TIMP-2 and TIMP-1 respectively. Observations from these studies indicate that the main site at which inhibition of MMPs occur, is around the residues where the disulphide bonds form in the N-terminal domain of the TIMP molecule. With respect to inhibition of MMP-3 by TIMP-1, two cysteine residues Cys<sup>1</sup> and Cys<sup>70</sup> are thought to play critical role. The N-terminal domain is not the only site at which TIMPs can block MMP function. Studies have shown that latent forms of MMP-9 and MMP-2 are prevented from activation by the binding of the C-terminal domains of TIMP-1 and TIMP-2 respectively to their C-terminal, haemolytic domain<sup>73</sup>. They are therefore, able to inhibit more than one MMP at a given time. TIMPs are substrate specific and do not all inhibit the same MMPs. With respect to MMP-3 both TIMP-1, a 29-30kDa glycoprotein and TIMP-2, 22kDa, contain sites for MMP-3 in their N-terminal domain. TIMP-1 however is the main inhibitor (Figure 1.5), as MMP-3 binds more readily to TIMP-1 than TIMP-2<sup>52</sup>.

TIMPs are not just MMP inhibitors they have other roles also. They are able to aid the activation of some MMPs. Small doses of TIMP-2 in cells expressing MT1-MMP can enhance MMP-2 activation <sup>59,72</sup>. TIMP-1 is able to stimulate fibroblast cells to produce MMP-1 <sup>77</sup>. Both TIMP-1 and 2 can affect the number of cell division of a range of cell types. There is data to show that TIMP-3 can induce apoptosis <sup>78</sup>. When over expressed these inhibitors can lead to the reduction of tumour genesis and metastasis <sup>79</sup>.



**Figure 1.5:** - Ribbon diagram of the complex of TIMP-1 and the catalytic domain of MMP-3. The image was prepared from Brookhaven Protein DNA Bank entry (1UEA) using ribbons and taken directly from a review paper by Nigase H. and Woessner J. F <sup>80</sup>. MMP-3 is shown in silver and TIMP-1 in red. The 3 red spheres represent the calcium ion in MMP-3. The 3 histidines that ligate the catalytic zinc ion are shown in blue. The yellow bands represent the disulfide bonds in TIMP-1.



**Figure 1.6:** -Flow chart representing the different regulatory event of MMPs. Positive regulation is represented in green and suppressors and inhibitors of MMP regulation is in red.

Extracellular matrix is a complex and dynamic meshwork of proteins and proteoglycans. Its main function is to provide structural support to organisms, however, it also influences many biological processes. Basic processes such as cell proliferation and differentiation, cell adhesion and migration and tissue morphogenesis are all affected by the ECM. The ECM is therefore considered to be of great importance in organisms and thus any agent that can modify it has the potential to affect a variety of normal and pathological processes. More specifically it is believed that the degradation of the ECM regulates its assembly and is therefore essential in both biological and pathological processes. The major enzymes that degrade the ECM are MMPs, the adamalysin-related membrane proteinases that comprise of disintegrin and metalloproteinases domains (ADAMs), bone morphogenetic protein 1 (BMP1) family of metalloproteinases and the tissue serine proteinases such as thrombin, tissue-type plasminogen activator (tPA), urokinase type plasminogen activator (uPA) and plasmin

81

### 1.2.3 MMPs Role in Biological and Pathological Processes

There are a number of biological and pathological processes that MMPs are thought to play an important role in, some of which are listed below in Table 1.3.

Evidence for these processes have arisen typically from *in situ* hybridisation and immunohistochemistry techniques. *In situ* hybridisation showed the presence of MMP mRNAs at the sites of tissue remodelling and immunohistochemistry was used to demonstrate the presence of active MMP proteins together with their specific degradation products at these sites<sup>82-85</sup>.

As mentioned earlier MMPs play an important role in ECM remodelling and in this way is associated in a wide range of biological processes. Embryonic development, remodelling, tissue repair/wound healing and morphogenesis, are just a few of the everyday processes MMPs are thought to participate in. Cell migration is an essential part of development and morphogenesis. In order to migrate, cells need to change from an adhesive phenotype to a non-adhesive phenotype. MMPs are able to facilitate this process through several means: (1) They can activate cytoskeletal motor function thus enabling cells to move, (2) Initiate traction by modulating adhesive sites and cell-

surface adhesive molecules, (3) Remove physical barriers by degradation of the ECM and (4) Stimulate the secretion of chemoattractants that guide migration. Cell proliferation and apoptosis can be initiated by changing the ECM microenvironment another mechanism related with MMP function<sup>86</sup>.

MMPs are tightly regulated (see 1.2.3), their expression is minimal in normal biological processes. When MMPs are expressed in high doses the surrounding tissue is damaged, this results in a variety of destructive diseases such as arthritis, tumour invasion and metastasis and atherosclerosis. The increased expression of MMPs in such destructive diseases is thought to be influenced by genetic variants in the promoter regions of the MMP genes.

#### **1.2.3.1 Tumour Invasion and Metastasis**

Cancers such as carcinomas and sarcomas, are malignant tumours that arise from the abnormal and uncontrolled division of cells that leads to the invasion and destruction of the surrounding tissue. The development of cancer is thought to occur primarily via a change in one or more genes in the affected cells. There are, however, a number of other causative factors that are believed to induce the development of cancer. Smoking has been associated with the development of lung cancer and radiation is thought to initiate some bone sarcomas and leukaemias. There are also some known viruses that have been shown to cause tumours. The major factors in the morbidity and mortality of cancers are the ability of malignant tumours to invade normal tissue and spread to distant sites in the body. The spread of cancer cells, metastasis, occurs via (1) the blood stream, (2) the lymphatic system or (3) across body cavities, which lead to secondary tumours at sites distant from the original tumours. 'Tumour invasion and metastasis involves the attachment of tumour cells to the basement membrane, degradation of the local connective tissue, followed by penetration and migration through proteolysed stroma'. MMPs possess proteolytic activity against the ECM and are therefore thought to be major contributors of the destruction of the connective tissue.

Most of the 26 members so far identified in the MMP family, have been associated with the development of cancer. The gelatinases, MMP-2 and MMP-9, in particular are

thought to play key roles in tumour progression. These two MMPs can degrade type IV collagen, which is the major structural protein of the basement membrane. Degradation of this protein facilitates the spread of tumour cells to other sites of the body. MMP-1, the most ubiquitously expressed collagenase, has been associated with a wide variety of advanced cancers. Recently a common polymorphism was identified in the promoter region of the MMP-1 gene. In the form of an insertion/deletion, this variant is located at position -1607bp relative to the start of transcription and gives rise to an additional guanine base (G)<sup>87</sup>. In the general population, the genotype frequency of this common variant was found to be 30% homozygous for the 1G allele, 30% homozygous for the 2G allele and 40% heterozygous. However *in vitro* experiments on cultured tumour cells, the occurrence of the 2G allele was found to be much higher at 62%<sup>87</sup>. Studies have also shown that patients with this 2G allele have an increased expression of the MMP-1 protein<sup>87</sup>. Individual MMPs tend to be expressed at higher levels in tumour cells than normal cells, the levels increasing as the tumour advances. This expression pattern thus supports the role of MMPs in tumour invasion and metastasis<sup>88-90</sup>.

#### **1.2.3.2 Rheumatoid Arthritis**

Rheumatoid arthritis is an autoimmune disease of connective tissue. It affects a number of joints in the body and is characterised by swelling, warmth, redness over the overlying skin, pain and restriction in motion of the effected areas. It causes inflammation of the synovial lining and is also associated with cartilage destruction. Monocytes/macrophages play an important role in the pathogenesis of rheumatoid arthritis due to their association with immunoregulatory function. During an inflammatory response, they can produce MMPs and protease inhibitors and can also secrete catabolic cytokines such as IL-1 and TNF- $\alpha$ , which can subsequently induce MMP gene expression by the surrounding synovial fibroblast cells. In this way, monocytes and macrophages have the potential to directly and indirectly modulate the ECM. Proteinase functions, such as the MMPs, degrade the ECM, thus causing rheumatoid arthritis. A number of MMPs have been demonstrated to be associated with rheumatoid arthritis. Higher levels of MMP-1, MMP-3 and MMP-9 have been detected in the serum and synovial fluid of patients with this inflammatory disease. Therefore these MMPs are thought to participate in joint destruction<sup>91-94</sup>.

### 1.2.3.3 Coronary Heart Disease

MMPs are thought to play a crucial role in the development and progression of atherosclerosis. It is thought that they facilitate the migration and proliferation of vascular smooth muscle cells and are associated with the weakening of the connective tissue in the atherosclerotic plaque. There are several lines of evidence to suggest this: Interstitial collagenase (MMP-1), gelatinase B (MMP-9) and stromelysin (MMP-3) are all expressed at high levels in macrophages, SMC, lymphocytes and the endothelium of atherosclerotic plaques<sup>83,84,95</sup>. A number of polymorphisms have been identified in the promoter region of several MMPs which influence their transcriptional activity. Studies have shown that alleles which cause increased MMP expression are more prevalent in patients with atherosclerosis and its consequence, heart disease. Recently, a number of promoter polymorphisms have been identified in the MMP-9 gene. One of these variants, a single base change from cytosine (C) to thymine (T) –1562 bp from the start of transcription were found to cause an increase in transcriptional activity in the T allele. A cohort study of coronary atherosclerotic patients revealed patients with one or two copies of the T allele had an increased risk of stenosis in all three coronary arteries<sup>96</sup>. A common functional variation in the promoter region of the MMP-12 gene was found to have an effect on the luminal dimensions of coronary arteries in patients with diabetes<sup>97,98</sup>. This A-82/G polymorphism is located near the AP-1 cis-element and influences the binding of the AP-1 protein. The A allele showed higher affinity with transcription factor AP-1<sup>97</sup>. Another MMP common promoter polymorphism associated with CHD is the 5A/6A found –1612 bp from the start of transcription in the MMP-3 gene. In this case, it is the 5A alleles that possess the higher functional activity and this is discussed in more detail later<sup>99</sup>.

### 1.2.3.4 Aneurysms

An aneurysm is a bulge or a balloon that forms out of the walls of arteries. They form in areas where the artery wall has weakened, often due to the accumulation of fatty plaques but may also be an inherited condition or a complication of high blood pressure or of other disease that weaken the vessel wall. Over time the elasticity of the vessel walls are lost and the aneurysm may tear or burst (rupture). Variation in MMP expression, are thought to contribute to the susceptibility of aneurysms. Collagen



degradation is required in the pathogenesis of abdominal aortic plaques. MMP-13 is a potent member of its family as, like MMP-3, it has a broad substrate specificity. It is expressed by SMC of vascular tissue. Large quantities of MMP-13 have been localised to abdominal aortic aneurysms and have therefore been postulated to play a prominent role in its pathogenesis <sup>100</sup>. Over expression of MMP-3 and TIMP-3 have recently been observed in abdominal aneurysms <sup>101</sup>. A recent study investigated MMP promoter polymorphisms as a risk factor for the development of coronary aneurysms. Several polymorphisms in different MMPs were examined in a large cohort study, namely MMP-2 (C-1306/T), MMP-3 (5A/6A), MMP-9 (CA repeat) and MMP-12 (A-82/G). No association was found between coronary aneurysms and polymorphisms in the MMP-2, MMP-9 and MMP-12. The data indicated a possible association between the 5A allele of the MMP-3 gene 5A/6A common polymorphism, and coronary aneurysm

<sup>102</sup>.

**Table 1.3****Biological and pathological processes in which MMPs are implicated.**

| BIOLOGICAL PROCESSES           | PATHOLOGICAL PROCESSES |
|--------------------------------|------------------------|
| Development                    | Tissue Destruction     |
| Blastocyst implantation        | Cancer invasion        |
| Embryonic development          | Cancer metastasis      |
| Enamel maturation              | Decubitus ulcer        |
| Growth plate cartilage removal | Gastric ulcer          |
| Nerve growth                   | Periodontal disease    |
| Nerve outgrowth                | Osteoarthritis         |
| Skeletal, bone growth          | Rheumatoid arthritis   |
| Reproduction                   | Fibrotic Disease       |
| Cervical dilatation            | Atherosclerosis        |
| Endometrial Cycle              | Fibrotic lung disease  |
| Graafian follicle rupture      | Liver cirrhosis        |
| Luteolysis                     | Multiple sclerosis     |
| Postpartum uterine involution  | Otosclerosis           |
| Mammary gland morphogenesis    |                        |
| Mammary gland involution       |                        |
| Rupture of fetal membranes     |                        |
| Maintenance                    | Weakening of matrix    |
| Angiogenesis                   | Aortic aneurysm        |
| Apoptosis                      | Dilated cardiomyopathy |
| Hair follicle cycle            | Epidermolysis bullosa  |
| Macrophage function            |                        |
| Nerve regeneration             |                        |
| Neutrophil function            |                        |
| Remodelling of bone            |                        |
| Wound healing                  |                        |

(Adapted from 'Matrix Metalloproteinases' by William C Parks and Robert P Mecham)  
52

MMP-3 is an important member of its family because it has a broad specificity. It is not only expressed in many biological processes but also in many pathological ones too (see Table 1.3). It is because it has an effect on so many processes, including CHD, that it is the focus of this study in relation to atherosclerosis.

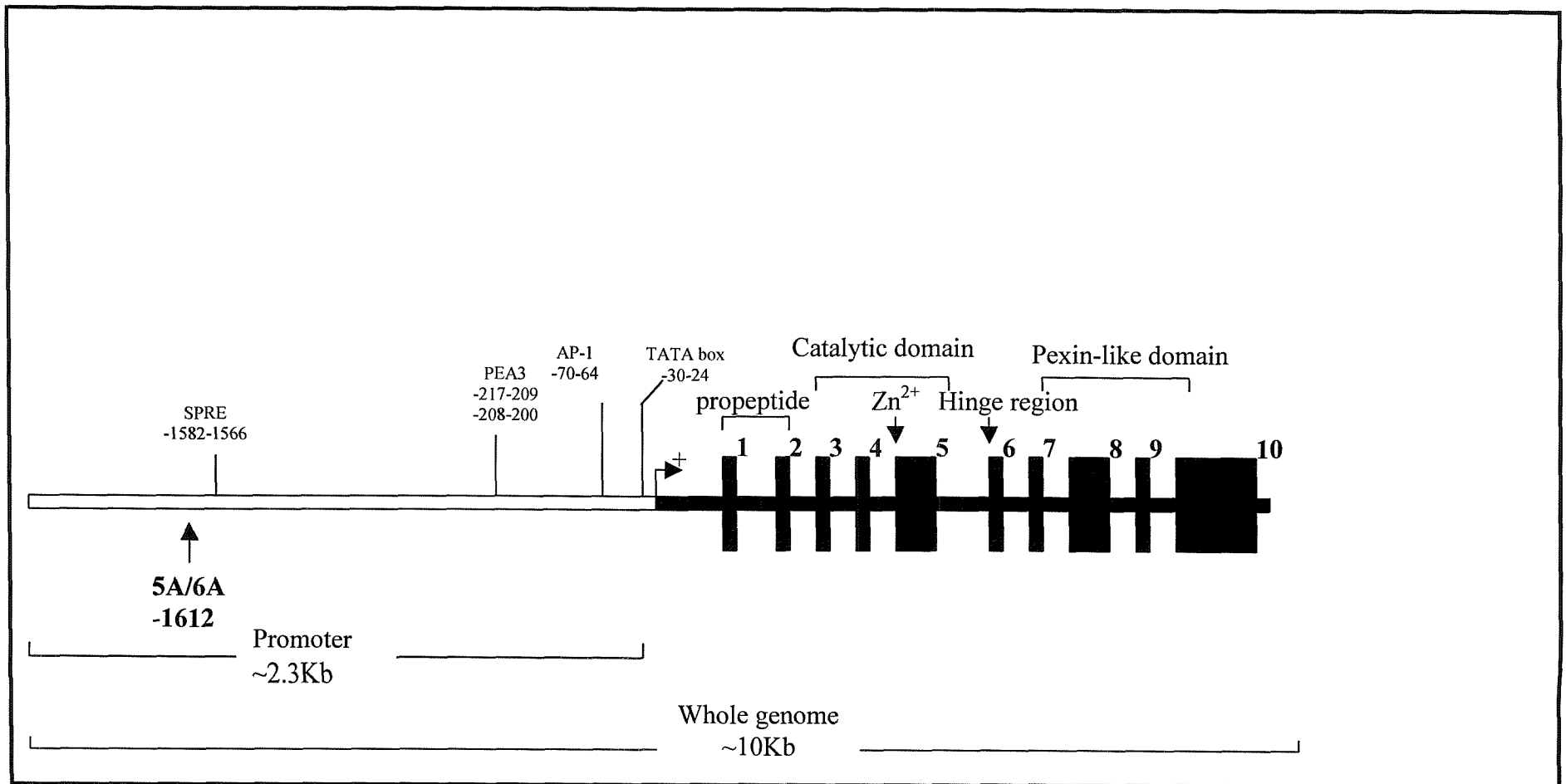
#### 1.2.4 MMP-3 (Stromelysin-1), a Member of the MMP Family

MMP-3 is a proteoglycanase and has been classified as a member of the stromelysin group of the MMP family. It is considered to be an important member of its family as it has a broad specificity and can activate other MMPs.

MMP-3 was first recognised in the mid seventies in both rabbit synovial fibroblasts and extracts of human articular cartilage as a noncollagenolytic metalloprotease<sup>103,104</sup>. It was isolated several years later from condition medium of rabbit bone explants. Using such techniques as somatic cell hybridisation, *in situ* hybridisation and linkage analysis the MMP-3 gene was mapped to chromosome 11q22.3. The MMP-3 gene in fact lies within a cluster containing several other MMP genes in the long arm of chromosome 11, which includes MMP-8, MMP-10, MMP-1, MMP-12, MMP-7 and MMP-13<sup>53</sup>. It is approximately 10 Kb in length, comprising of a promoter region and 10 exons (see Figure 1.7 below).

cDNA clones of human, rabbit, rat and mouse were used to deduce the primary structure of MMP-3. Human proMMP-3 consists of a propeptide region of 82 amino acids (aa), a catalytic domain 165 aa, a hinge region 25 aa and a C-Terminal domain 188 aa residues. The pro- and the catalytic domains of MMP-3 have separated folding units. The pro-domain is made up of three  $\alpha$  helices with an extended peptide region around the cysteine switch sequence. The bait region is located in front of the second  $\alpha$  helix. In contrast, the catalytic domain consists of five-stranded  $\beta$ -sheet, three  $\alpha$  helices and connecting loops. The catalytic domain contains three calcium ions and two zinc molecules. One of the zinc molecules is located in the active site and interacts with the side chains His<sup>201</sup>, His<sup>205</sup> and His<sup>211</sup>. The other is a structural zinc molecule, and thought to play a crucial role in keeping MMP-3 active<sup>52</sup>.

MMP-3 is not stored in normal human cells and is only synthesised when required. Transcriptional activation of MMP-3 has been discussed earlier in this chapter, but a list of activators and inhibitors of MMP-3 transcription is given in Table 1.4.



**Figure 1.7:** - Schematic diagram of the MMP-3 gene.

**Table 1.4: Factors Specific to MMP-3 Synthesis**

| Activators   | Suppressors                        |
|--|------------------------------------|
| Cytokines and Growth Factors                               | Calmodulin                         |
| Basic fibroblast growth factor                             | cAMP                               |
| Epidermal growth factor                                    | EIA-F                              |
| Interleukin 1 $\alpha$ , $\beta$                           | Glucocorticoid                     |
| Interleukin 6  | High glucose                       |
| Interleukin 10   | Interferon $\gamma$                |
| Nerve growth factor  | Interleukin 4                      |
| Platelet-derived growth factor                             | Phosphate citrate                  |
| Relaxin  | Progesterone                       |
| Transforming growth factor $\alpha$                        | Retinoic acid                      |
| Tumour necrosis factor $\alpha$                            | SV40 T-antigen                     |
|  | Tetracycline                       |
| Cell surface activators                                    | Transforming growth factor $\beta$ |
| Anti- $\alpha$ 5 $\beta$ 1 integrin antibody               |                                    |
| Calcium ionophore A23187                                   |                                    |
| Concanavilin A   |                                    |
| Crystals: calcium pyrophosphate<br>hydroxyapatite<br>urate |                                    |
| EMMPRIN (basigin/M6 antigen)                               |                                    |
| Fibronectin fragments                                      |                                    |
| RGD peptides   |                                    |
| SPARC (osteonectin/BM 40)                                  |                                    |
| Staphylococcal endotoxin A                                 |                                    |
| Chemical agents  |                                    |
| Cycloheximide  |                                    |
| Cytochalasin B   |                                    |
| Linoleic acid hydroperoxide                                |                                    |
| Lipopolysaccharide   |                                    |
| N-6-aminohexyl)-5-chloro-1-nephtalene-sulfonamide<br>(W-7) |                                    |
| Phorbol ester  |                                    |
| Prostaglandin E <sub>2</sub>                               |                                    |
| Trifluoperazine  |                                    |
| Other stimuli  |                                    |
| Cellular aging   |                                    |
| Heat shock   |                                    |
| Mechanical injury  |                                    |
| Oncogene products  |                                    |
| UV irradiation   |                                    |
| Viral transformation                                       |                                    |

Adapted from Matrix Metalloproteinases, Academic Press, edited by William C Parks and Robert P Mecham <sup>52</sup>.

Part of MMP-3 function is to degrade ECM substrates, which include collagens III, IV, V, IX, X, and XI, lamins, elastin, entactin, fibronectin, fibrin, fibrillin, fibulin, link protein, osteonectin, tenascin, vitronectin and ECM proteoglycans<sup>90</sup>. MMP-3 is also able to release cell surface molecules such as E-cadherin, L-selectin, herarin-binding EGF-like growth factor and TNF- $\alpha$ . It can also activate members of its family, such as MMP-1, MMP-7, MMP-8, MMP-9 and MMP-13, as well as proteinases not in its family. It is also able to inactivate several serine proteinases inhibitors. MMP-3 is therefore believed to play a key role in connective tissue remodelling in biological processes and in the pathogenesis of a number of diseases<sup>52,99,105</sup>. Table 1.5 gives a list of some normal and pathological cells and tissues MMP-3 has been associated with.

**Table 1.5**

**Normal and pathological cells and tissues associated with MMP-3**

| Normal Tissue               | Pathological Tissue                         |
|-----------------------------|---|
| Capillary endothelial cells | Aneurysmal abnormal aorta                   |
| Cartilage                   | Atherosclerosis                             |
| Chondrocytes                | Basal cell carcinoma                        |
| Dermal papilla cells        | Bronchial carcinoma                         |
| Endometrial stroma          | Colorectal cancer                           |
| Fibroblasts                 | Esophageal squamous carcinoma               |
| Growth plate                | Gastrintestinal ulcers                      |
| Hepatic lipocytes           | Head and neck carcinoma                     |
| Macrophages                 | Herniated invertibral disc                  |
| Mammary glands              | Lung squamous carcinoma                     |
| Osteoblasts                 | Osteoarthritic cartilage                    |
| Placenta                    | Respiratory epithelium cells after wounding |
| Serum                       | Rheumatoid synovium                         |
| Smooth muscle cells         | Serum (elevated in Rheumatoid arthritis)    |

Adapted from Matrix Metalloproteinases, Academic Press, edited by William C Parks and Robert P Mecham<sup>52</sup>.

The exact role of MMP-3 in many biological and pathological process is not fully understood, however elevated levels of MMP-3 have been detected in the cornea, bone marrow, spleen, brain and spinal cord, skeletal muscle, the heart, liver, kidney and lung. Its extensive substrate specificity and presence in a wide range of tissues only emphasises its importance. MMP-3 is therefore postulated to play key roles in development and morphogenesis, tissue resorption, reproduction, wound healing,

arthritis, tumour invasion and metastasis and atherosclerosis. Though the specific role of MMP-3 is still poorly understood it is believed to take part in the development and morphogenesis of bone, tissue resorption, reproduction, mammary glands, lung rudiments, salivary glands and hair follicle formation<sup>52,90</sup>.

MMPs facilitate cell migration and tissue remodelling, part of the processes involved in wound healing. A series of experiments, namely *in situ* hybridisation and immunohistochemistry, has demonstrated that MMP-3 is produced by basal keratinocyte close to the wound edge of chronic ulcers and dermal wounds. It has been postulated that this site of MMP-3 production is the point where the epidermis proliferates. Elevated levels of MMP-3 have also been detected in dermal fibroblast, thus suggesting that MMP-3 may have specific roles in wound healing<sup>83,106,107</sup>.

There is extensive evidence that over-expression of MMPs correlates with tumour invasion and metastasis. It has also become clear that the type of MMP expressed in tumours and the surrounding tissue depends on the tumour cells. MMP-3 is just one of the MMPs that have been associated with promoting the development and progression of cancer in particular carcinomas. There are several types of carcinomas that MMP-3 is associated with, these are listed in Table 1.3 above. Breast cancer is just one of example. It is ubiquitously expressed in the stromal cells throughout the mammary glands and its levels are highest during ECM remodelling and alveolar regression<sup>108,109</sup>. MMP-3 expression in mammary epithelial cells initiates a cascade of events. It up-regulates more MMP-3 and other members of it family. It also induces cell proliferation, an altered stroma, increased expression of tenascin-C and angiogenesis. All these activities are characteristics of cancer progression<sup>110</sup>. Sternlicht *et al* (1999) demonstrated that induction of an altered stromal environment by MMP-3 results in promotion of a phenotypic conversion and malignant transformation of mammary epithelial cells<sup>90</sup>. Once the tumour is initiated, it becomes independent of MMP-3 expression.

MMP-3 is associated with a number of other destructive degenerative diseases. Degenerative changes in the intervertebral disc in the lower back is a common musculoskeletal disability, characterised by the degradation of the disc matrix. The



aging process, heavy physical activity, infection and smoking are factors that have all been associated with this disability. Genetic factors are also thought to play a part. A potent member of its family MMP-3 activity has been postulated to play a role in the degeneration of intervertebral disc. Studies have shown expression of MMP-3 during both mechanical loading and inflammation<sup>111,112</sup>. These stresses are thought to induce MMP-3 synthesis. A recent study investigated the 5A/6A promoter polymorphism in association with accelerated degenerative of intervertebral disc. A cohort of 103 young and old Japanese individuals were studied. The data indicated that the 5A allele was a possible risk factor for this disability<sup>113</sup>. As an important regulator of the ECM and an important mediator of tissue injury, MMP-3 is also thought to be involved in the inflammatory condition primary sclerosising cholangitis. The 5A/6A MMP-3 gene promoter polymorphism was investigated by as a possible risk factor. A cohort of case and control individuals were investigated. The data showed a possible association for the 5A allele and is thought to be involved in the disease susceptibility and progression<sup>114</sup>.

MMP-3 is believed to play a crucial role in atherogenesis and consequently CHD. There are several lines of evidence to support this. Experiments using such techniques as mRNA *in situ* hybridisation, immunocytochemistry and SDS-page zymography, have shown an increased expression of MMP-3 in vessels containing coronary atherosclerotic plaques, when compared to non atherosclerotic arteries<sup>83</sup>. MMP-3 has been demonstrated to be extensively expressed in smooth muscle cells and macrophage foam cells, particularly in regions prone to rupture such as the plaque cap. These findings led to the initial hypothesis that MMP-3 initiates platelet activation and coagulation and might therefore be involved in plaque rupture<sup>84,95,106</sup>.

More recently, using single stranded conformation polymorphism (SSCP), a common variant was identified in the promoter region of the MMP-3 gene. This variant is located -1612bp upstream from the start site of transcription and gives rise to one allele with a 6 adenosine (6A) run and the other allele with a 5 adenosine (5A) run<sup>105</sup>.

In a study of 354 healthy individuals, the frequency of this 5A/6A variant was found to be 0.51/0.49 respectively. In a cohort of 72 patients with CHD defined by angiography, the 6A/6A genotype was found to be associated with a greater progression of coronary



atherosclerosis over a period of three years<sup>105,115</sup>. *In vitro* assays of promoter activity showed that the 5A allele had a 2 fold higher promoter activity than the 6A allele. Subsequently DNA-protein interaction assays were performed and two putative transcription factors were found to bind to this region. One of these transcription factors was observed to have a higher affinity to the 6A allele, the other was observed to have similar interactions with both alleles<sup>99</sup>. It was therefore proposed that the transcription factor with the higher affinity to the 6A allele is likely to be a transcription repressor, which would explain the results gained by promoter functional analysis<sup>99</sup>. Thus it was postulated that a decrease in MMP-3 expression results in an increase in matrix deposition and therefore a more rapid growth of the fibrotic atherosclerotic plaque. It therefore follows that an increase in MMP-3 expression will result in a decrease in matrix deposition and therefore a more lipid based atherosclerotic plaque. A case-control study of acute myocardial infarction revealed that the frequency of the 5A allele was significantly higher in patients than in controls. Thus, patients with the 5A allele are genetically more susceptible to plaque rupture<sup>116</sup>.

The difference in functional activity of the two alleles of the 5A/6A promoter polymorphism has also been associated with a number of other diseases, namely acute lung injury, degeneration of the intervertebral disc and primary sclerosis cholangitis<sup>113,114,117</sup> as discussed earlier.

In summary, MMP-3 is involved in arterial wall remodelling. It is believed that MMP-3 contributes to connective tissue dysfunction and the eventual plaque rupture that leads to thrombosis<sup>118</sup>. MMP-3 has already been demonstrated to be associated with CHD and atherosclerosis. The 5A/6A variant has been examined in many studies, which have shown its association with this disease. MMP-3 also shows an increased expression in coronary atherosclerotic plaques and over-expression in smooth muscle cells and macrophage foam cells. Together these factors make the MMP-3 gene a good candidate for the study of atherosclerosis.

Based on these findings, this study is to extend further the analysis of the MMP-3 genomic region, with particular respect to genetic epidemiology and *in vitro* functional studies.

### 1.3 Hypothesis: -

Sequence variants (in addition to the 5A/6A polymorphism) in the MMP-3 gene may influence the susceptibility and outcome of coronary atherosclerosis through differential effects on MMP-3 expression or activity.

#### 1.3.1 Plan: -

1. To search for sequence variants in a 2.5 Kb promoter sequence and the coding region of the MMP-3 gene using the mutation scanning technique Dideoxy fingerprinting.
2. To determine the extent of linkage disequilibrium between the sequence variants identified in a collection of unrelated subjects.
3. To examine the sequence variants identified in the MMP-3 gene in cohorts of patients with coronary heart disease.
4. To perform functional assays, to determine whether the sequence variants identified influence the expression of the MMP-3 gene or the properties of the gene product. For promoter variants, reporter gene assays will be used to determine if the regulation of transcription is affected by any variants identified in this region. In addition, the electrophoretic mobility shift assay (EMSA) will be carried out to investigate whether there are nuclear proteins binding differentially to the different alleles. For variants in the coding region, the different alleles will be expressed *in vitro*, followed by functional analysis of the different forms of MMP-3 protein.

## **Chapter 2**

### **Materials and Methods**

#### **2.1 Materials**

##### **2.1.1 Chemicals**

Chemicals were generally purchased from Sigma Aldrich Co Ltd, Poole, UK and Merck Ltd Hunter Boulevard (BDH) Lutterworth, UK. SeaKem LE agarose was bought from Flowgen Instruments, Staffordshire, UK. Adenosine 5'-Triphosphate was acquired from Amersham Life Sciences, Inc, Buckinghamshire, UK.

Radioactive isotopes were acquired from Amersham or ICN Biochemicals Ltd., Thames UK.

##### **2.1.2 Enzymes**

All restriction endonucleases were purchased from New England BioLabs Inc, Hitchin, UK, as were Calf intestinal alkaline phosphatase, T4 polynucleotide kinase and T4 DNA ligase. Taq DNA polymerase was obtained from GIBCO BRL Life Technologies and Thermo Sequenase DNA polymerase were procured from Amersham Life Sciences, Inc.

##### **2.1.3 Water**

All solutions were made with water purified by reverse osmosis through a Purite DC8 deioniser (UHQ). The Purite DC8 yields water of high quality and was therefore suitable to use for solutions utilised in DNA manipulation investigations as well as tissue culture following sterilization.

##### **2.1.4 Sterilization**

Sterilization of solutions and water was achieved by autoclaving at 121°C for 15 min

## **2.1.5 Bacterial Growth Medium**

### **2.1.5.1 Luria-Bertani (LB) Medium**

20 g of LB broth base powder (Lennox L Broth Base) was suspended in 1 L of deionised water and heated to boiling using a hotplate magnetic stirrer. The solution was autoclaved for 15 min at 121°C and stored at room temperature until required.

### **2.1.5.2 LB Agar**

32 g of LB agar powder (Lennox L Agar) was suspended in deionised water and heated to boiling using a hotplate magnetic stirrer. The solution was autoclaved for 15 min at 121°C and stored at room temperature until required.

### **2.1.5.3 Ampicillin**

A stock solution of 100 mg/ml ampicillin was made up using sterilized UHQ water. This was stored in aliquots at -20°C. The working concentration of ampicillin routinely used was 100 µg/ml.

## **2.1.6 Statins**

5 mM stock solutions of both Pravastatin and Fluvastatin were prepared using sterilized UHQ water. Stock solutions were stored at -20°C until required.

### **2.1.7 PMA (phorbol 12-myristate 13-acetate, a protein kinase C activator)**

This was stored at 20°C in darkness as a 2 mM stock solution. Stock solution was prepared with DMSO (1 mg PMA + 0.81 ml DMSO)

### **2.1.8 Oligonucleotide Primers**

All oligonucleotide primers utilized in this study were purchased from MWG-Biotech AG.

### **2.1.9 Plasmids**

Plasmids employed in this study are as follows:

pGL3-Promoter vector (Promega)

pGL3-Basic vector (Promega)

pGL3-Control vector (Promega)

pRL-TK vector (Promega)

pGEM®-T-Easy Vector (Promega)

### **2.1.10 Purification Kits**

A number of purification kits were utilized in this study. These are listed below with the name of manufacturer. The protocol performed for each kit was as recommended by the manufacturer.

Wizard® DNA Clean-up System kit (Promega)

Wizard® PCR prep DNA purification system (Promega)

Mobius™ 1000 Plasmid Kits (Novagen)

HiSpeed™ Plasmid Purification Handbook (Qiagen)

Microspin G-25 column (Pharmacia)

### **2.1.11 Buffers and Solutions**

#### **10 x TAE Buffer**

|                     |         |
|---------------------|---------|
| Tris                | 121 g   |
| Glacial acetic acid | 28.5 ml |
| NaEDTA              | 9.3 g   |
| dH <sub>2</sub> O   | 2.5 l   |

**TE Buffer**

10 mM Tris-HCl pH 8.0

1 mM EDTA

**10 x Medium Salt Buffer**

4 M NaCl 125 µl

1 M Tris-HCl 100 µl

1 M MgCl<sub>2</sub> 100 µl

1 M DTT 10 µl

dH<sub>2</sub>O 665 µl

**4 M NaCl**

NaCl 23.376 g

dH<sub>2</sub>O 90 ml

**0.1 M DTT**

DTT 0.154 g

dH<sub>2</sub>O 1 ml

**10 x ELB (erythrocyte lysis buffer)**

K(CO<sub>3</sub>)<sub>2</sub> 5.05 g

NH<sub>4</sub>Cl 41 g

0.5 M EDTA pH8.0 2 ml

**NLB (nucleic lysis buffer)**

1.0 M Tris pH 8.0 5 ml

5 M NaCl 40 ml

0.5 M EDTA pH 8.0 2 ml

**10 x PCR Buffer**

200 mM Tris-HCL pH 8.4

500 mM KCl

**1 x T4 Polynucleotide Kinase Buffer**

70 mM Tris-HCL

10 mM MgCl<sub>2</sub>

5 mM DTT

**1 x T4 DNA Ligase Buffer**

50 mM Tris-HCL

10 mM MgCl<sub>2</sub>

10 mM DTT

1 mM ATP

25 µg/ml BSA

**1 x NEBuffer 2**

50 mM NaCl

10 mM Tris-HCl

10 mM MgCl<sub>2</sub>

1 mM DTT

**1 x NEBuffer 3**

100 mM NaCl

50 mM Tris-HCl

10 mM MgCl<sub>2</sub>

1 mM DTT

**1 x NEBuffer 4**

50 mM potassium acetate

20 mM Tris acetate

10 mM magnesium acetate

1 mM DTT

**Stop Solution**

95% formamide

500 mM EDTA

0.05% Bromophenol blue

0.05% Xylene cyanol

**MADGE Dye**

98% deionised formamide    9.8 ml,

10 mM EDTA (pH 8.0)        200 µl,

0.0025% Xylene cyanol ff    1.5 g,

0.025% Bromophenol blue    1.5 g

**6 x Blue/Orange Loading Dye**

10% Ficoll®

10 mM Tris-HCl pH 7.5

50 mM EDTA pH 8.0

0.025% Bromophenol blue    1.5 g

0.0025% Xylene cyanol ff

0.4% orange G



## **2.2 Methods**

### **2.2.1 Mutation Scanning of the MMP-3 Gene**

To search for sequence variants in the MMP-3 gene, dideoxy fingerprinting (ddF) assays were performed. In these assays, the target sequences were PCR amplified (see 2.2.1.1), then the amplicons were purified (see 2.2.1.2) and subjected to dideoxy termination reactions (see 2.2.1.3), followed by non-denaturing polyacrylamide gel electrophoresis (see 2.2.1.4). The assays were carried out using DNA samples from 10 male patients with coronary heart disease and 10 healthy male subjects.

#### **2.2.1.1 Polymerase Chain Reaction**

The polymerase chain reaction (PCR) is an *in vitro* technique used to amplify target DNA sequence. Two oligonucleotides are used as primers for a series of synthetic reactions that are catalysed by a DNA polymerase. The oligonucleotide primers are generally designed to be an exact complement of the template DNA flanking the segment of DNA to be amplified and are typically between 20-30 nucleotides in length<sup>119</sup>. PCR is a cycling process, the quantity of target DNA sequence doubles with each cycle. The template DNA is first separated by thermal denaturation in the presence of a large molar excess of each oligonucleotide primer and the four dNTP's. The samples are then cooled to a temperature that allows the primers to anneal specifically to their target regions. The DNA polymerase then acts to extend the primers. The cycle of denaturing, annealing and extension is normally repeated 20-40 times.

In this study, the promoter and coding regions of the MMP-3 gene were amplified in 5 separate PCR reactions:

1. The promoter region (approximately 2.5 kilo bases (Kb) in length);
2. Exons 1 through to 4 (approximately 1.5 Kb in length);
3. Exon 5 and 6 (639 base pairs (bp));
4. Exons 7 and 8 (926 bp);
5. Exons 9 and 10 (approximately 1.7 Kb)

Typically, the PCR reactions were carried out in a total volume of 25  $\mu$ l, containing 20 ng genomic DNA, 10 pmol of each primer, 200  $\mu$ M dNTP, 1-3 mM  $MgCl_2$ , 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.05% (v/v) W1 (Gibco BRL), 1 unit Taq polymerase (Gibco BRL) and sterile deionised water ( $dH_2O$ ). The solution was overlaid with 25  $\mu$ l of mineral oil to prevent evaporation and incubated at 95°C for 3 min, followed by 30 cycles at 95°C for 1 min, 50°C for 1 min and 72°C for 3 min. The PCR products were stored at 4°C until they were purified.

The magnesium chloride concentration, annealing temperature, and sequences of primers used in each of the five PCR reactions are described in Tables 2.1 and 2.2.

**Table 2.1: Conditions of PCR in Mutation Scanning**

| Target sequence | Primers    | % G/C | Product length | Magnesium concentration | Annealing temperature |
|-----------------|------------|-------|----------------|-------------------------|-----------------------|
| Promoter region | MMP3-ddF1  | 27    |                |                         |                       |
|                 | MMP3-ddF10 | 33    | ~2.5 Kb        | 2.5 mM                  | 50°C                  |
| Exon 1-4        | MMP3-ddF11 | 45    |                |                         |                       |
|                 | MMP3-ddF18 | 40    | ~1.59 Kb       | 2.5 mM                  | 50°C                  |
| Exon 5-6        | MMP3-ddF19 | 50    |                |                         |                       |
|                 | MMP3-ddF22 | 45    | 639 bp         | 1.0 mM                  | 45°C                  |
| Exon 7-8        | MMP3-ddF23 | 40    |                |                         |                       |
|                 | MMP3-ddF26 | 55    | 926 bp         | 2.5 mM                  | 50°C                  |
| Exon 9-10       | MMP3-ddF27 | 50    |                |                         |                       |
|                 | MMP3-ddF30 | 45    | ~1.7 Kb        | 1.0 mM                  | 50°C                  |

**Table 2.2: PCR Primer Sequences used in Mutation Scanning**

| Primer       | Sequence                         |
|--------------|----------------------------------|
| MMP3 – ddF1  | 5'-AAT GTC TGC ATG ATT TTT-'3    |
| MMP3 – ddF10 | 5'-CTT TAC TTA GCT CTA TGT-'3    |
| MMP3 – ddF11 | 5'-CCA AAC AAA CAC TGT CAC TC-'3 |
| MMP3 – ddF18 | 5'-CAT TTC TTG AGC ATT CTT GG-'3 |
| MMP3 – ddF19 | 5'-CAA TGC ACA GCA GAC ATG TG-'3 |
| MMP3 – ddF22 | 5'-CTC TAT CAA GTT TCT GCG TG-'3 |
| MMP3 – ddF23 | 5'-CTG ATT TCT ATC TGT CAT CC-'3 |
| MMP3 – ddF26 | 5'-AGA GAG AAG CAG GCC TAA GG-'3 |
| MMP3 – ddF27 | 5'-ATG TTC GGT AAC ATG GCC TC-'3 |
| MMP3 – ddF30 | 5'-ATG AGA ACA ACT TGC CCT TG-'3 |

### 2.2.1.2 Gel Electrophoresis and Purification of PCR Products

Gel electrophoresis is one of the most important biological techniques used today for both the characterisation and purification of nucleic acids. Nucleic acids are negatively charged molecules. In the presence of an electric field such molecules migrate towards the positive electrode. The rate at which this occurs depends on the size and shape of the nucleic acid. Thus different sizes and forms of nucleic acid will migrate at different rates, which provides a basis for their separation. Separation of mixtures of nucleic acids enables the specific study, purification and recovery of target molecules<sup>120</sup>.

In this study, the PCR amplicons were separated from the by-products of amplification by agarose gel electrophoresis. This was performed using a horizontal submerged slab gel electrophoresis system. The two ends of a horizontal perspex tray were sealed with tape to form a gel mold. 1.2 g of agarose was mixed with 100 ml 1 x TBE buffer and then heated to boiling in a microwave oven. The agarose was allowed to cool to approximately 50°C and 6 µl of the nucleic acid binding dye ethidium bromide (0.6 µg/ml) was added to the agarose solution before pouring into the open mold (ethidium bromide is the most widely used nucleic acid dye). Positively charged, it neutralises a portion of the negative charges on the DNA or RNA molecules. Binding of the ethidium bromide molecule to nucleic acids is by intercalation between adjacent bases in the DNA and RNA strands<sup>120</sup>. The molten agarose was poured gently to avoid the formation of air bubbles. A rigid perspex comb was inserted into the gel to create loading wells. The gel was allowed to set for 30 min. Once set the comb was removed and the gel completely submerged in 1 x TBE electrophoresis buffer.

Each sample was PCR amplified five times to ensure there was sufficient of each sample for purification. These replicates were pooled together and each sample was mixed with 5 µl of 6 x Blue/Orange loading dye prior to loading. A total volume of 100 µl per sample was loaded per well of the agarose gel. Electrophoresis was performed at a constant 100 volts (v) for 1 hr.

On completion of electrophoresis, the agarose gel was placed onto a Dual-Intensity Transilluminator and viewed under UV-light. When ethidium bromide binds to DNA

or RNA strands, the molecule is able to absorb UV wavelengths of light, which is re-emitted as energy in the visible spectrum (560 nm). This is because the purines and pyrimidine bases in the DNA or RNA molecules absorb the energy and then transmit it to the bound ethidium bromide, which in turn converts the energy and generates a strong fluorescence. The DNA and RNA molecules illuminate as red-orange bands on a dark brown-black background. In this way the target, molecules can be easily selected and examined. Nucleic acids absorb UV radiation strongly at 254 nm, which causes substantial damage to DNA and RNA molecules producing base alterations, breaks and cross-links. The damage is drastically reduced at longer wavelengths, so for purification purposes the gel was irradiated at a wavelength of 302 nm.

The appropriately sized DNA fragments were excised from the gel using a sharp sterilised blade. Each DNA sample was extracted from the agarose by centrifugation through glass wool. This was achieved with the aid of two microcentrifuge tubes 0.5 ml and 1.5 ml in size per sample, some glass wool and a microcentrifuge. The bottom of the smaller microcentrifuge tube was cut using a pair of scissors. Using a clean pair of tweezers a few strands of glass wool was placed into the smaller tube and pushed to the lower end forming a tight pack through which the DNA can be filtered. The gel slice was then placed on top of the glass wool in 0.5 ml microcentrifuge tube. The tube was sealed and 0.5 ml tube was then placed into larger 1.5 ml tube. Each assembly was then placed into the microcentrifuge and the samples were centrifuged through the glass wool at full speed for 20 min. The filtrates were collected in the large microcentrifuge tubes ready for purification. Purification was performed using a standard ethanol precipitation procedure<sup>121</sup>. The exact procedure is as follows: Approximately 300 µl of DNA solution was collected per sample. 30 µl 3 M NaOAc pH 5.2 (1/10<sup>th</sup> of the filtrate) and 900 µl 100% ethanol (x 3 total sample volume) was added to each sample. The samples were vortex briefly and the incubated on dry ice for 20 min to allow DNA precipitation. At the end of the incubation period the sample were centrifuged once more for 20 min at  $14 \times 1,000 \text{ revs/min}^{-1}$ . The DNA collected at the bottom of the microcentrifuge tubes in the form of a pellet. The supernatant was carefully removed ensuring the pellet was not disturbed. The DNA pellet was then cleaned with 70% ethanol by centrifuging once more in a microcentrifuge for 10 min at  $14 \times 1,000 \text{ revs/min}^{-1}$ . Again the supernatant was removed from each sample and the pellet was left to dry at room temperature (approx. 25 C) for 30 min. When dry the

DNA pellets were resuspended in 20  $\mu$ l dH<sub>2</sub>O (sterile) and were stored at -20°C until required.

### **2.2.1.3 Dideoxy Termination Reactions**

ddF is a hybrid technique between SSCP and Sanger's dideoxy sequencing. In ddF, one of the four standard dideoxy-sequencing reactions is performed to generate a ladder of bands, which are then resolved by electrophoresis on a non-denaturing polyacrylamide gel <sup>122</sup>. Non-denaturing gels gives the advantage of not just separating nucleotide sequences by nucleotide length but also by the conformation of the secondary structure. Autoradiography gives a fingerprint, in which genetic variants can be identified from by either a gain or loss of a dideoxy termination segment and/or by an alteration in the mobility of at least one of the termination segments that contain the mutation.

In this study, the dideoxy termination reactions were performed using a Thermo Sequenase cycle sequencing kit (Amersham Pharmacia Biotech). Cycle sequencing uses repeated cycle of thermal denaturing, annealing and extension/termination. In this way it increases signal levels and therefore decreases the amount of template required. These characteristic were important for this application as the volume of DNA per sample was small and the concentrations were unknown and varied between samples. The method used was based on that recommended by the manufacturer entitled 3'-dNTP internal label cycle sequencing. This involves two steps, a labelling reaction, followed by a chain-termination step, which uses dideoxynucleotides. In the labelling reaction, the primer is extended using four deoxynucleotide triphosphates, one of which is radioactively labelled. In the termination step, ddNTP is introduced to the reaction mix. This molecule lacks the hydroxyl group that attaches to the 3'carbon of the sugar component present in a normal dNTP. It is this 3'carbon molecule that participates in the phosphodiester bond formation, however without this hydroxyl group, this reaction cannot take place. Consequently, any further elongation is blocked, and the chain terminated.

Because only one terminator is required for performing the ddF reaction, in this study, the protocol in the sequencing kit manual (Amersham Pharmacia Biotech) was scaled

down to one quarter of the volume. In the labelling reaction, 1  $\mu$ l of purified PCR product was mixed with 3.125  $\mu$ l of label mix which comprised of 0.5 pmol/ $\mu$ l primer (see Table 2.3), 7-deaza-GTP, dTTP, dCTP, [ $\alpha$ -<sup>33</sup>P] dATP, reaction buffer, thermosequenase DNA polymerase and water, in a 0.5 ml eppendorf tube. 3  $\mu$ l of paraffin oil was used to cover the reaction mix to prevent both contamination and evaporation. Samples were then subjected to 60 cycles at 95°C for 15 sec, 50°C for 15 sec and 72°C for 15 sec. In the termination reaction, 3.5  $\mu$ l of labelling reaction was transferred into a termination tube containing 4  $\mu$ l of the terminator ddGTP. Again, paraffin oil was used as an overlay and the samples were subjected to 50 cycles at 95°C for 30 sec and 72°C for 90 sec. Subsequently, 4  $\mu$ l of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added to each sample to terminate the reaction.

**Table 2.3: Primers used in ddF Reactions**

| Primer ID    | Primer sequence                         | Orientation | Location |
|--------------|---|-------------|----------|
| MMP3 – ddF1  | 5'-AAT GTC TGC ATG ATT TTT-3            | Forward     | Promoter |
| MMP3 – ddF2  | 5'-CTT TGT AAA TTT TCT CTG-             | Reverse     | Promoter |
| MMP3 – ddF3  | -TAT CAG GCT TTC CTC TAA-3              | Forward     | Promoter |
| MMP3 – ddF4  | 5'-ACA CAG TTG ATA GCC ACT-3            | Reverse     | Promoter |
| MMP3 – ddF5  | 5'-CTT AAA TGT GAT GTA TAG-3            | Forward     | Promoter |
| MMP3 – ddF6  | 5'-ATT TTC TCT TCA TGC TCT-3            | Reverse     | Promoter |
| MMP3 – ddF7  | 5'-AGA AAA GAA AGG AAA GGA-3            | Forward     | Promoter |
| MMP3 – ddF8  | 5'-GCA AAT AAA AGA TGT AAG-3            | Reverse     | Promoter |
| MMP3 – ddF9  | 5'-GAC TAT AGC TAT GTA TGT-3            | Forward     | Promoter |
| MMP3 – ddF10 | 5'-CTT TAC TTA GCT CTA TGT-3            | Reverse     | Promoter |
| MMP3 – ddF11 | 5'-CCA AAC AAA CAC TGT CAC TC-3         | Forward     | Exon 1   |
| MMP3 – ddF12 | 5'-TTT CAG CTT ACT CTG GAA GCT-3        | Reverse     | Exon 1   |
| MMP3 – ddF13 | 5'-ATT AAG AAG TGA GCA ACT GC-3         | Forward     | Exon 2   |
| MMP3 – ddF14 | 5'-AAG CAG ATG CTA TTT TAT CC-3         | Reverse     | Exon 2   |
| MMP3 – ddF15 | 5'-AAT GGT TCA GAG GGG TTT TC-3         | Forward     | Exon 3   |
| MMP3 – ddF16 | 5'-AAC ACC AGA TAG ATG AAT GG-3         | Reverse     | Exon 3   |
| MMP3 – ddF17 | 5'-TGT TGT ACT GGA AAA GGC TC-3         | Forward     | Exon 4   |
| MMP3 – ddF18 | 5'-CAT TTC TTG AGC ATT CTT GG-3         | Reverse     | Exon 4   |
| MMP3 – ddF19 | 5'-CAA TGC ACA GCA GAC ATG TG-3         | Forward     | Exon 5   |
| MMP3 – ddF20 | 5'-TCC AAA CAT TCT CAC GGT GC-3         | Reverse     | Exon 5   |
| MMP3 – ddF21 | 5'-TGG AAA GCA TAC AAA TGG GC-3         | Forward     | Exon 6   |
| MMP3 – ddF22 | 5'-CTC TAT CAA GTT TCT GCG TG-3         | Reverse     | Exon 6   |
| MMP3 – ddF23 | 5'-CTG ATT TCT ATC TGT CAT CC-3         | Forward     | Exon 7   |
| MMP3 – ddF24 | 5'-ACT GTA TTA GCC AGA ACT TG-3         | Reverse     | Exon 7   |
| MMP3 – ddF25 | 5'-TGG GCT ACT ACT TTA CCA TG-3         | Forward     | Exon 8   |
| MMP3 – ddF26 | 5'-AGA GAG AAG CAG GCC TAA GG-3         | Reverse     | Exon 8   |
| MMP3 – ddF27 | 5'-ATG TTC GGT AAC ATG GCC TC-3         | Forward     | Exon 9   |
| MMP3 – ddF28 | 5'-CAA TGC TCC TCT TCC CTC TG-3         | Reverse     | Exon 9   |
| MMP3 – ddF29 | 5'-CTG AGT TTT GTA TAT TGT CAA ATG TC-3 | Forward     | Exon 10  |
| MMP3 – ddF30 | 5'-ATG AGA ACA ACT TGC CCT TG-3         | Reverse     | Exon 10  |

#### 2.2.1.4 Non-Denaturing Polyacrylamide Gel Electrophoresis

The above samples were resolved by electrophoresis on a 6% non-denaturing polyacrylamide gel. The procedure was as follows:

The two glass plates used for casting the sequencing gel were first cleaned thoroughly using 1% neutrocon followed by deionised water and finally methanol. The inner surface of the smaller glass plate was treated with repelcote V solution. This produces a hydrophobic surface that prevented the sequencing gel from sticking to the plate when it was removed. The two plates were held apart by plasticard spacers (~0.3 mm thick) and a gel mix (19.5 ml 36% bis/acrylamide mix 19:1, 10 ml 1 x TBE buffer pH 8.3, 70.5 ml dH<sub>2</sub>O, 262.5 µl 20% Ammonium persulphate (APS) and 71 µl NNN'N tetramethylethylenediamine (TEMED)) was poured into the sandwich to produce a gel. A comb was gently slid into the open end of the gel. Bulldog clips were used to hold the two plates together tightly. The gel was allowed to polymerise for 2 hr before it was assembled onto the gel tank apparatus, together with 1 x TBE electrophoresis buffer. The gel was pre-run at a constant 2000 watts (w), for 30 min to allow the buffer and gel to equilibrate to the same temperature. The samples were denatured by heating at 72°C for 2 min prior to loading onto the sequencing gel. Prior to loading, the comb was removed from the top of the gel and the square cut-out wells produced by the comb were flushed out with buffer. 5 µl per sample was loaded onto the sequencing gel and the samples were electrophoresed at a constant 2000 v and a constant temperature of 20°C for 3 hr 30 min, or until the bromophenol blue dye had reached the bottom of the sequencing gel. No loading dye was required when loading sequenced samples, as the stop solution used was equivalent.

At the end of electrophoresis, prising glass plates apart dismantled the gel sandwich. The gel was transferred from the larger glass plate to 3 MM Whatman filter paper and placed in a vacuum drier at 80°C for 1 hr 30 min. Once dry, the gel was set up for an autoradiograph. This was performed in complete darkness as hypersensitive film was used (BIOMAX MR from Kodak Scientific Imaging film). Hypersensitive film was chosen to increase the detection of  $\beta$ -particle emission. The gel was placed in a film cassette, an X-ray film was placed over it and the cassette was left at room temperature



for three to five days before developing the film. Once the films were developed the DNA ladder bands (the fingerprint) was analysed.

### **2.2.2 Radioactive, Manual Sequencing**

When a sequence variant was identified, the nature of the variation was defined by sequencing. As in the ddF assays, the Thermo Sequenase cycle sequencing kit (Amersham Pharmacia biotech) was used and the 3'-dNTP internal label cycle sequencing protocol in the kit was followed. For this application no scaling down was necessary. The dNTP label used was [ $\alpha$   $^{33}\text{P}$ ] dATP or [ $\alpha$   $^{33}\text{P}$ ] dCTP, according to the A/C richness in the target sequences. [ $\alpha$   $^{33}\text{P}$ ] dNTP was preferred over [ $\alpha$   $^{32}\text{P}$ ] dNTP because scatter was reduced and bands appear sharper on an autoradiograph when using the former. Four termination reactions were set up for each sample, ddA, ddC, ddG and ddT, the reactions were performed in 0.5 ml eppendorf tubes. The cycle programmes were the same as for ddF, i.e. the labelling reaction was performed in 60 cycles at 95°C for 15 sec, 50°C for 15 sec, and 72°C for 15 sec, and the termination reaction was run in 50 cycles at 95°C for 30 sec and 72°C for 90 sec. Stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cynol FF) was employed to stop the reactions and the samples were stored at -20°C until they could be electrophoresed.

The sequenced samples were resolved by electrophoresis on a 6% denaturing polyacrylamide gel, using the following procedure: The two gel plates were cleaned in the same manner as with the non-denaturing polyacrylamide gel, to ensure that the inner surfaces of the two plates were dust free. Repelcote V was applied to one of the plates and a gel mix (10 ml 30% acrylamide/bisacrylamide 19:1(w/w), 5 ml 10 x TBE buffer pH 8.3, 21.25 g Urea, ~20 ml dH<sub>2</sub>O, 37.5  $\mu\text{l}$  of 25% APS and TEMED) was poured and assembled as previously mentioned. The gel was pre-run at a constant 30 watts for about one hour before loading, to bring the temperature up to approximately 50°C. 600 ml 1 x TBE pH 8.3 was the buffer used. The sequenced reactions were heated to 75°C in a thermal cycler for 5 min before loading 5  $\mu\text{l}$  of each reaction onto the gel. The samples were electrophoresed at a constant 30 w and a constant

temperature of 50°C for varying lengths of times depending on the position of the putative variants.

On completion of electrophoresis, the gel was dismantled and then transferred from the glass plate to 3MM Whatman filter paper and dried using a vacuum drier at 80°C for 1 hr 30 min. Once dry the gel was set up for an autoradiograph.

### 2.2.3 Fluorescent, Automated Sequencing

The assays were performed using the ABI 310 Genetic Analyser (PE Applied Biosystems Warrington, UK) and a BigDye™ Terminator Cycle Sequencing Kit (PE Applied Biosystems Warrington, UK). This protocol involved a series of steps, including: (1) amplification of DNA samples; (2) PCR purification; (3) cycle sequencing reaction, (4) isopropanol precipitation, (5) electrophoresis on the ABI 310 genetic analyser, and (6) analysis.

#### 2.2.3.1 Principal of the Procedure

Fluorescently labelled dyes are attached to ACGT extension products in the DNA sequence reaction. The products of sequencing are purified and then electrophoresed through the capillary of the ABI 310 Genetic Analyser. A current aids the products through the capillary and as they pass through the detector window in the capillary coating, a laser excites the fluorescent dye labels. The CCD camera collects the fluorescence emitted from the dyes. The Prism® 310 Data collection software interprets the raw data and presents it in the form of a chromatograph.

The recommended amounts of purified DNA required for optimal conditions is given in the Table 2.4 below:

**Table: 2.4 Specimens Requirements**

| <b>DNA</b>                 | <b>QUANTITY</b> |
|----------------------------|-----------------|
| <b>Single-stranded DNA</b> | 50-100 ng       |
| <b>Double-stranded DNA</b> | 200-500 ng      |
| <b>PCR product DNA</b>     | 30-90 ng        |
| <b>BAC DNA</b>             | 0.5-1.0 µg      |

### **2.2.3.2 Amplification of DNA Fragments**

In this study, four DNA samples were chosen for sequencing and therefore PCR amplified. These samples included one wild type homozygote, 2 heterozygotes and one mutant homozygote. The target sequences were the promoter region (approximately 2.5 kilo bases (Kb) in length) and the sequence from exons 1 to 4 (approximately 1.5 Kb in length). The primers and PCR conditions used to amplify these regions in the ddF assays were used here.

### **2.2.3.3 Purification of PCR Products**

Purification of PCR amplicons was performed using the Wizard® PCR Preps DNA Purification System from Promega, as follows:

Five sets of amplifications were performed for each DNA sample and the PCR products were pooled together in a clean, labelled 1.5ml microcentrifuge tube, to give a total volume of 100 µl. 100 µl of Direct Purification Buffer was added to each sample. They were vortexed briefly, before adding 1 ml of resin to each sample. Each sample was again vortexed briefly three times over a 1-minute period. One Wizard® Minicolumn was prepared for each resin/DNA. This involved attaching a syringe barrel to each minicolumn and then inserting the tip of the minicolumn/syringe barrel assembly into the vacuum manifold. Each resin/DNA mix was then pipetted into the syringe barrel; a vacuum was applied until the resin/DNA mix was drawn into the minicolumn. 2 ml 80% isopropanol was then pipetted into each syringe barrel and the vacuum was applied once more. When the solution had been drawn through the vacuum was allowed to continue for a further 30 seconds to dry the resin. The syringe barrel was then removed. Each minicolumn was then transferred to 1.5 ml microcentrifuge tubes. They were centrifuged at 10,000 x g for a period of 2 minutes; this was to remove any residual isopropanol. The minicolumns were transferred to fresh microcentrifuge tubes. 50 µl of distilled water was then added to each of the minicolumns and allowed to incubate for 1 minute before they were centrifuged for 20 seconds at 10,000 x g to elute the bound DNA. The purified DNA samples were stored at -20°C until required.

#### 2.2.3.4 Sequencing Reaction

For each reaction the following reagents were mixed in a 0.2 ml tube:

|                                     |      |
|-------------------------------------|------|
| Terminator Ready Reaction Mix       | 8 µl |
| Purified PCR product DNA (30-90 ng) | 5 µl |
| Primer (3.2 pmol)                   | 2 µl |
| Deionised water                     | 5 µl |

The samples were centrifuged briefly, placed on a thermal cycler (DNA engine Tetrad, PTC-225 Peltier Thermal Cycler, MJ Research), and subjected to the following thermal cycling conditions: 96°C for 10 sec, 25 cycles at 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min.

#### 2.2.3.5 Purification of Extension Products

PE Applied Biosystems suggested several purification methods for purifying the extension products. These include (1) spin column, (2) isopropanol precipitation, (3) ethanol precipitation, and (4) ethanol/sodium acetate precipitation. In this study, isopropanol precipitation was performed, and the procedure was as follows:

The entire content of each extension reaction was transferred into new 1.5 ml microcentrifuge tubes using a pipette. 80 µl 75% isopropanol was then mixed into each reaction. The tubes were sealed and then vortexed briefly. Each sample was incubated at room temperature 25°C for a period of 15 minutes; this was to allow the extension products to precipitate. The tubes were then placed into the micro centrifuge, their orientations were noted and the samples were spun at maximum speed for 20 minutes. The supernatants from each sample were carefully removed and discarded. The pellets were washed, this was achieved by adding 250 µl 75% isopropanol to each pellet, briefly vortexing and centrifuging, taking particular note that each tube was placed in the same orientation as before.

#### **2.2.3.6 Electrophoresis on the ABI PRISM 310 Genetic Analyser**

Each sample pellet was resuspended in 12-25 µl of Template Suppression Reagent (TSR, supplied with the polymer). The samples were then vortexed and centrifuged briefly before heating the DNA at 95°C for 2 min to denature. Each sample was then chilled on ice. The samples were again vortexed and spun place on ice.

ABI 310 Genetic Analyser was set up for sequencing. This involved attaching a capillary (61 cm, 50 µm i.d. capillary) and a sequencing polymer (310 POP-6). A sample sheet was prepared which was then transferred to an injection list for sequencing. The samples were then loaded onto the tray in the ABI 310 Genetic Analyser. Sequencing was performed at a constant 50°C.

#### **2.2.4 Allelic Association Study**

The polymorphisms identified in the promoter and coding regions of the MMP-3 gene were independently genotyped using a simple assay protocol. The purpose of this was to investigate if linkage disequilibrium existed in any form between the variants. Subjects were PCR amplified in the target regions, digested using the appropriate restriction endonucleases and the products were fractionated on a polyacrylamide MADGE gel.

##### **2.2.4.1 Subjects and DNA Preparation**

The duplicated DNA samples from 32 unrelated individuals were analysed in this study. The DNA samples were isolated from peripheral blood using a salting-out method (7 ng/µl)<sup>123</sup>.

Several methods for DNA extraction have been developed over the years, the most established being the deproteinizing cell digests technique. This technique however is quite taxing and involved the use of hazardous organic solvents (phenol and isochochloroform). Though other non-toxic extraction procedures have been developed, they involve either extensive dialysis or require the use of filters, which can be quite expensive. For this study, an adaptation of the salting-out DNA extraction technique

was employed. It is a rapid, safe and inexpensive method, originally developed by S.A. Miller et al in 1988. It was carried out as follows: Whole blood samples were mixed on a rotary mixer for a period of 1 hr at 4°C. 3 volumes of x 10 ELB (erythrocyte lysis buffer) was added to each sample and incubated on ice for 15 min. They were then centrifuged at 1500 rpm for 15 min at 4°C. The supernatants were poured off and the pellets were washed in ELB twice. Pellets were transferred to fresh conical tubes and resuspended in x 2 NLB (nucleic lysis buffer). The cell lysates were then digested overnight, in a shaking incubator at 37°C, in the presence of protease K solution and 10% SDS. Saturated NaCl (~6 M) was added, and the samples were shaken vigorously and then centrifuged at 3000 rpm for 15 min to precipitate any protein and the supernatant transferred to a fresh tube. The DNA in the supernatant was precipitated using 2 volumes of 100% ethanol and incubated at 70°C for at least 30 minutes. The samples were centrifuged and the pellets were washed with 1 ml of 70% ethanol and left to air dry. The DNA was dissolved in 300 – 500 µl of TE buffer and stored at -20°C.

The quantitation of the DNA samples was performed using the software programme 'Molecular Dynamics DNA Quantification'. The fluorescent dye PicoGreen was used to illuminate the DNA samples for scanning. In brief, stock PicoGreen solution was diluted 1:200 in TE buffer or dH<sub>2</sub>O. 50 µl of working PicoGreen was pipetted into the required number of wells in a clear flat-bottom, microplate. 2 µl of each unknown DNA sample or calibration standards (concentration 36, 18, 9, 4.5 ng/µl) were then added to each well followed by 50 µl of dH<sub>2</sub>O. Samples were mixed gently by pipetting slowly to avoid air bubble formation, before incubating at room temperature in the dark. Samples were then scanned on the Flourimager 595 (Molecular Dynamics) and quantitated using the Molecular Dynamics ImageQuant™ software.

#### **2.2.4.2 PCR Amplification of Genomic DNA**

To execute this allelic association study, 6 separate PCR reactions were performed on each sample, one for each variant, using five oligonucleotide primer pairs described in Table 2.5. The primer pairs used to amplify each section of template DNA were MMP3-ddF1/ddF2, MMP3-Xmn/SLPROM5, MMP3-ddF5b/ddF6b MMP3-

P4F/MMP3-P4R and MMP3-ddF13/ddF14. The magnesium and annealing condition were optimised for each oligonucleotide primer pair. A 25µl PCR reaction was performed, typically comprising of 7 ng/µl DNA (dried at 80°C), 1 U *Taq* DNA polymerase, 10 pmol each primer, 0.05% W1, 1-3 mM MgCl<sub>2</sub>, 200 µM dNTP, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, and dH<sub>2</sub>O in a microtitre plate with a mineral oil overlay. Thermal cycling was carried out on an MJ Research PTC-225 (Peltier Thermal Cycler) using the following cycling programme: 95°C for 2 min, 35 cycles at 95°C for 30 sec, 50/55°C for 30 sec, and 72°C for 30 sec, followed by 72°C for 10 min.

**Table 2.5: PCR Primers Utilized in Association Study**

| SNP           | Primer sequence  |
|---------------|--|
| G (-1986)/A   | MMP3 – ddF1f: 5'-AAT GTC TGC ATG ATT TTT-3' (forward primer)<br>MMP3 – ddF2r: 5'-CTT TGT AAA TTT TCT CTG-3' (reverse primer)               |
| 5A (-1612) 6A | MMP3-Xmnbf: 5'-TTT CAA TCA GGA CAA GAC GAA GTT T-3' (forward primer)<br>SLPROM5r: 5'-GAT TAC AGA CAT GGG TCA CG-3' (reverse primer)        |
| C (-1346) A   | MMP3-5bf: 5'-AAA AGT CTT TAG ACA TAA TC-3' (forward primer)<br>MMP3-6br: 5'-CTG AAG ATG CCC ACA CCG-3' (reverse primer)                    |
| C (-709) G    | MMP3-709F: 5'-CCG GTA AGC AAT GTA ATT CAT TTG A -3' (forward primer)<br>MMP3-709R: 5'-CCC TTT GGG TTG GTT TTA ATT CTG -3' (reverse primer) |
| C (-376) G    | MMP3-P4F: 5'-CAT CTT TTA TTT GCT TTG ACA ATG T-3' (forward primer)<br>MMP3-P4R: 5'-CTA GAC AAT TTA TGT TTG TCC-3' (reverse primer)         |
| Exon 2        | MMP3 – ddF13f: 5'-ATT AAG AAG TGA GCA ACT GC-3' (forward primer)<br>MMP3 –ddF14r: 5'-AAG CAG ATG CTA TTT TAT CC-3' (reverse primer)        |

#### 2.2.4.3 Restriction Enzyme Digestion

There is a wide range of restriction enzymes that cleave DNA at specific target sites. These sites are pre-dominantly 4-8 bp in length. If a mutation occurs in the restriction site the enzyme is unable to cleave the site, this can be detected by the alteration in



restriction fragment seen by electrophoresis. In this way these enzymes have been utilised as tools in the direct detection of known and unknown mutations.

Restriction endonucleases were used here to determine the genotypes in the 32 subjects described above (see 2.1.4.1). The PCR products were digested with the appropriate restriction endonuclease, one per polymorphism. For each digest reaction 15 µl of PCR product was mixed with 5 µl of a mix containing 3 units of restriction enzyme, 1 x buffer (appropriate buffer provided with each enzyme) and dH<sub>2</sub>O. Paraffin-oil was used as an overlay to prevent evaporation and contamination. The digests were incubated overnight (12-16 hr) on an MJ Research PTC-225 (Peltier Thermal Cycler) at the enzyme's recommended temperature. Table 2.6 below gives the restriction enzyme and the digestion temperature for each variant.

**Table 2.6: Enzyme and Digest Conditions Used**

| Polymorphism                     | Restriction enzyme      | Restriction site                   | Temperature |
|----------------------------------|-------------------------|------------------------------------|-------------|
| Promoter region<br>G (-1986)/A   | <i>BstNI</i>            | 5'CC↓AGG'3<br>3'GGA↑CC'5           | 60°C        |
| Promoter region<br>5A (-1612) 6A | <i>Xmn I</i>            | 5'GAANN↓NNTTC'3<br>3'CTTNN↑NNAAG'5 | 37°C        |
| Promoter region<br>C (-1346) A   | <i>Hpa II</i>           | 5'C↓CGG'3<br>3'GGC↑C'5             | 37°C        |
| Promoter region<br>A (-709) G    | <i>EcoR I</i>           | 5'G↓AATTC'3<br>3'CTTAA↑G'5         | 37°C        |
| Promoter region<br>C (-376) G    | <i>Pfl-F1</i>           | 5'GACN↓NNGTC'3<br>3'CTGNN↑NCAG'5   | 37°C        |
| Exon 2<br>G (56679) A            | <i>Taq<sup>α</sup>1</i> | 5'T↓CGA'3<br>3'AGC↑T'5             | 65°C        |
| Exon 2<br><u>G (56524) A</u>     | <i>PshA1</i>            | 5'GACNN↓NNGTC'3<br>3'CTGNN↑NNCAG'5 | 37°C        |

#### 2.2.4.4 Electrophoresis

The products of digestion were separated by electrophoresis on a 10% horizontal polyacrylamide MADGE gel (16.6 ml 30% bis/acrylamide 19:1, 5 ml 10 x TAE pH 8.1, 28.4 ml dH<sub>2</sub>O). 5 µl of each digest was mixed with 2 µl of MADGE dye (9.8 ml 98% deionised formamide, 200 µl 10 mM EDTA (pH 8.0), 1.5 g 0.0025% Xylene



cyanol ff, 1.5 g 0.025% Bromophenol blue). A total of 5 µl per sample was loaded into the wells and the fragments were electrophoresed at a constant 120 volts for 20 min using 600 ml 1 x TAE electrophoresis buffer. At the end of electrophoresis the gel was stained with 1:10,000 Vistra Green for 15 min before scanning on the Flourimager 595 (Molecular Dynamics).

Vistra Green is a staining dye for ds DNA. It is a superior substitute for ethidium bromide, as it is more sensitive and bands are sharper. In contrast to the ethidium bromide, which binds to nucleotides by intercalating between adjacent nucleic acids, Vistra Green is thought to bind non-covalently to the backbone of the DNA molecules.

#### **2.2.4.5 Statistical Analysis**

The “Estimating Haplotype-frequencies (EH)” programme (<ftp://linkage.rockefeller.edu/software/eh>) was used to determine if linkage disequilibrium, between the common polymorphisms identified in the MMP-3 gene was present. In this programme the haplotype frequencies are estimated with allelic association (H1) and without (H0). This programme also calculates the log-likelihood, chi-squared and the number of degrees of freedom under the hypotheses H1 and H0.

#### **2.2.5 Transient Reporter Gene Expression**

The possible effects, of the promoter variants identified, on transcriptional activity were investigated, both as individual polymorphisms and together as a collective using a reporter gene assay.

A number of reporter genes and assays have been developed for the analysis of transcriptional regulation, such as CAT,  $\beta$ -galactosidase, firefly luciferase and growth hormone. For this study, luciferase pGL3-reporter vectors from Promega and a Dual luciferase assay system also from Promega, were employed. Four different pGL3-vectors are available from Promega<sup>124</sup>:

- 1) pGL3-Basic vector: This lacks eukaryotic promoter and enhancer elements.
- 2) pGL3-Promoter vector: This contains an SV40 promoter upstream of the luc + reporter gene.
- 3) pGL3-Enhancer vector: This contains an SV40 enhancer downstream from the luc+ reporter gene.
- 4) pGL3-Control vector: This contains both SV40 promoter and enhancer sequences.
- 5) pRL-TK-vector: This contains the herpes simplex virus thiamine kinase promoter upstream from the Rluc+ reporter gene/

Three of the above mentioned vectors were used in this study. The pGL3- Promoter vector was utilized to analyse individual polymorphisms, the pGL3-Control vector was used as one of the controls and the pRL-TK vector was used to standardise the uptake of DNA between the different samples.

#### **2.2.5.1 Reporter Gene Constructs for Individual Polymorphisms**

Two strategies were employed to create the reporter gene constructs for each individual promoter variant. The first was to try and clone 2-3 concatenated copies of each double stranded (ds) DNA sequence encompassing each variant into the pGL3-Promoter vector. These ds DNA sequences were in the form of ds oligonucleotide probe 31-mer in length with *Bgl* II and *Bam* H1 cutting site overhang sequences on the ends. In this approach all copies of the DNA insert were required to be in the same orientation. The second strategy required the cloning of a longer ds oligonucleotide probe 56-mer in length into a pGL3-Promoter vector. This longer oligonucleotide encompassed 2 copies of the original concatenated oligonucleotide probe. Hence only one insert, in the correct orientation for the construct, would be required.

For each promoter polymorphism, complimentary single stranded oligonucleotides were ordered from MWG-Biotech AG. These corresponded to the MMP-3 promoter sequences encompassing the polymorphic site and are shown below in Table 2.7.

**Table 2.7: Oligonucleotides Used to Create Reporter Gene Constructs**

| Oligonucleotide | Sequence (31/56-mer)  | Polymorphism                    |
|-----------------|---|---------------------------------|
| FS-MMP3-2bF     | 5'-GAT CCG GGA AAC ATC TTC CAG GGA AGT CGT A-3'   | -1986 G allele<br>upper strand  |
| FS-MMP3-2bR     | 5'-GAT CTA CGA CTT CCC TGG AAG ATG TTT CCC G-3'   | -1986 G allele<br>lower strand  |
| FS-MMP3-2F      | 5'-GAT CCG GGA AAC ATC TTT CAG GGA AGT CGT<br>AGA TCC GGG AAA CAT CTT TCA GGG AAG TCG TA-3' | -1986 A allele<br>upper strand  |
| FS-MMP3-2R      | 5'-GAT CTA CGA CTT CCC TGA AAG ATG TTT CCC GGA<br>TCT ACG ACT TCC CTG AAA GAT GTT TCC CG-3' | -1986 A allele<br>lower strand  |
| FS-MMP3-3F      | 5'-GAT CAG ACA TGG TTT TTC CCC CCA TCA AAG<br>AGA CAT GGT TTT TCC CCC CAT CAA AA-3'         | -1612 5A allele<br>upper strand |
| FS-MMP3-3R      | 5'-GAT CTT TGA TGG GGG GAA AAA CCA TGT CTC TTT<br>GAT GGG GGG AAA AAC CAT GTC TG-3'         | -1612 5A allele<br>lower strand |
| FS-MMP3-3bF     | 5'-GAT CAG ACA TGG TTT TTT CCC CCA TCA AAG AGA<br>CAT GGT TTT TTC CCC CAT CAA AA-3'         | -1612 6A allele<br>upper strand |
| FS-MMP3-3bR     | 5'-GAT CTT TGA TGG GGG GAA AAAA CCA TGT CTC<br>TTT GAT GGG GGG AAA AAC CAT GTC TG-3'        | -1612 6A allele<br>lower strand |
| FS-MMP3-5F      | 5'-GAT CCA AGT GGC TAT CAA CTG TGT GGG CAT<br>AAG TGG CTA TCA ACT GTG TGG GCA TA-3'         | -1346 A allele<br>upper strand  |
| FS-MMP3-5R      | 5'-GAT CTA TGC CCA CAC AGT TGA TAG CCA CTT ATG<br>CCC ACA CAG TTG ATA GCC ACT TG-3'         | -1346 A allele<br>lower strand  |
| FS-MMP3-5bF     | 5'-GAT CCA AGT GGC TAT CAC CTG TGT GGG CAT<br>AAG TGG CTA TCA CCT GTG TGG GCA TA-3'         | -1346 C allele<br>upper strand  |
| FS-MMP3-5bR     | 5'-GAT CTA TGC CCA CAC AGG TGA TAG CCA CTT<br>ATG CCC ACA CAG GTG ATA GCC ACT TG-3'         | -1346 C allele<br>lower strand  |
| FS-MMP3-7F      | 5'-GAT CCG TAA TTC ATT TCA ATT CTA CAA CTA GTA<br>ATT CAT TTC AAT TCT ACA ACT AA-3'         | -709 A allele<br>upper strand   |
| FS-MMP3-7R      | 5'-GAT CTT AGT TGT AGA ATT GAA ATG AAT TAC TAG<br>TTG TAG AAT TGA AAT GAA TTA CG-3'         | -709 A allele<br>lower strand   |
| FS-MMP3-7bF     | 5'-GAT CCG TAA TTC ATT TCA GTT CTA CAA CTA GTA<br>ATT CAT TTC AGT TCT ACA ACT AA-3'         | -709 G allele<br>upper strand   |
| FS-MMP3-7bR     | 5'-GAT CTT AGT TGT AGA ACT GAA ATG AAT TAC TAG<br>TTG TAG AAC TGA AAT GAA TTA CG-3'         | -709 G allele<br>lower strand   |
| FS-MMP3-9F      | 5'-GAT CCC TTT TAT AAT GTC TTT CTT AAA ATA CTT<br>TTA TAA TGT CTT TCT TAA AAT AA-3'         | -376 C allele<br>upper strand   |
| FS-MMP3-9R      | 5'-GAT CTT ATT TTA AGA AAG ACA TTA TAA AAG TAT<br>TTT AAG AAA GAC AAT ATA AAA GG-3'         | -376 C allele<br>lower strand   |
| FS-MMP3-9bF     | 5'-5'-GAT CCC TTT TAT AAT GTG TTT CTT AAA ATA<br>CTT TTA TAA TGT GTT TCT TAA AAT AA-3'      | -376 G allele<br>upper strand   |
| FS-MMP3-9bR     | 5'-GAT CTT ATT TTA AGA AAC ACA TTA TAA AAG TAT<br>TTT AAG AAA CAC AAT ATA AAA GG-3'         | -376 G allele<br>lower strand   |

#### 2.2.5.1.1 Restriction Endonuclease Digest of pGL3-Vectors

The restriction endonuclease *Bgl* II was used to digest the pGL3-vector in preparation for the ligation reaction. Enzyme and buffer were used as recommended by the manufacturer. The digest reaction typically contained 5-10 µg of vector, 2 units of the enzyme *Bgl* II per sample and 1 unit of calf intestinal alkaline phosphatase (CIAP) per 5 µg of DNA in a total reaction volume of 20-50 µl. The digest reaction was incubated at 37°C for 3-6 hrs. Care was taken to make certain enzyme specificity was not

affected by ensuring the concentration of glycerol in the reaction never exceeded 5% (v/v). CIAP was added to the digest reaction to dephosphorylate the 5'-end terminus of the vector DNA. Dephosphorylation, inhibits recircularisation of the vector, a favourable reaction. The digest reaction was stopped by DNA purification.

#### **2.2.5.1.2 Purification of pGL3-Promoter Vector**

The plasmid vector was purified from the digest products using the Wizard® DNA Clean-up System kit from Promega. The pores in the minicolumns of this kit were too small to purify the ~5 Kb vector and therefore minicolumns from an alternative Promega purification kit (Wizard® PCR prep DNA purification system) were utilized. The purification was performed using the protocol as described and recommended by the manufacturer's own Technical Bulletin in the kit. In brief deionised water was first heated to 65-70°C using a heat block. A Wizard® minicolumn was then assembled by attaching the syringe barrel to the Luer-Lok® extension of the minicolumn and inserting the assembly into a vacuum manifold. 1 ml of Wizard® DNA Clean Up Resin was mixed into the vector/digest sample, before pipetting into the syringe barrel of the Wizard® minicolumn assembly. A vacuum was applied to draw the solution through the minicolumn. 2 ml of 80% isopropanol was then added to the syringe barrel to wash the minicolumn. Again a vacuum was applied to draw the solution through the minicolumn. Once this was achieved the vacuum was applied for a further 30 sec to dry the resin. The syringe barrel was removed and discarded and the minicolumn was transferred to a fresh microcentrifuge tube. The minicolumn was centrifuged at 10,000 x g for 2 min to remove any residual isopropanol. The minicolumn was again transferred to a fresh microcentrifuge tube and 50 µl of pre-warmed dH<sub>2</sub>O was administered. After a minute the minicolumn was centrifuged once more to elute the bound plasmid DNA.

#### **2.2.5.1.3 Preparation of Double Stranded Oligonucleotides**

For each promoter polymorphism, two pairs of complimentary single stranded oligonucleotide probes were ordered (see Table 2.7). These were annealed together (one for each polymorphic allele) using a simple assay protocol. 40 ng of each oligonucleotide pair were mixed together with 1 x medium salt buffer (500 mM NaCl,

100 mM Tris-HCl, 100 mM MgCl<sub>2</sub>, 10 mM DTT, dH<sub>2</sub>O) to give a total reaction volume of 20 µl. The reaction mix was then heated at a constant 95°C for 10 min. This was achieved by using a glass beaker containing 200 ml of water, a thermometer and a hot-plate magnetic stirrer. The samples were then allowed to cool slowly to room temperature by removing from the hot plate and resting on the workbench.

#### 2.2.5.1.4 Phosphorylation of Double Stranded Oligonucleotides

Each ds oligonucleotide was phosphorylated using T4 polynucleotide kinase (T4 PNK) for cloning. T4 PNK, catalyses the transfer of the γ-phosphor from ATP to the 5'hydroxyl terminus of polynucleotides. Approximately 1µl of each ds oligonucleotide was mixed with T4 PNK, 1 x T4 PNK buffer, 1 µM ATP and dH<sub>2</sub>O in a total reaction volume of 25 µl. The samples were incubated at 37°C for 30 min in a Hybaid Omnigene thermal cycler.

#### 2.2.5.1.5 Ligation of Plasmid Vector and Insert DNA

T4 DNA ligase was utilized to covalently join vector and insert DNA together. The concentrations of vector and ds oligonucleotides were estimated together with their molecular weight standards. Three vectors: insert DNA ratios (3:1, 1:1 and 1:3 respectively) were calculated and routinely employed. A simple equation was used for this calculation:

$$\frac{\text{Ng of vector} \times \text{Kb size insert}}{\text{Kb size of vector}} \times \text{molar ratio of} \frac{\text{insert}}{\text{vector}}$$

A typical ligation reaction comprised of 100 ng vector, 1.8 ng/0.6 ng/0.19 ng insert DNA or control, 1 Weiss unit T4 DNA ligase, 1 x T4 DNA ligase buffer and dH<sub>2</sub>O in a total reaction volume of 10 µl. The ligation reactions were performed in a water-bath at 16°C overnight (16 hr).

#### **2.2.5.1.6 Growth and Preservation of *E.coli*.**

The *E.coli*. strain used for transformation in this study was the JM109 (Promega). This was chosen as it is a *recA*-host and thus prevents undesirable restriction of cloned DNA and recombination with the host chromosomal DNA. This strain is known to grow well and can be used efficiently in transformation experiments by a variety of techniques.

*E.coli*. JM109 was routinely grown aerobically at 37°C on LB agar plated and kept as working stocks on these plates at 4°C. Liquid cultures of *E.coli*. were grown by inoculating LB medium with one freshly colony and incubating overnight (16 hr) at 37°C in a platform shaker at 225 rpm.

#### **2.2.5.1.7 Preparation of Competent *E.coli*.**

A fresh colony of *E.coli*. (JM109) was inoculated in 25 ml of LB medium and incubated overnight in a platform shaker at 225 rpm and at a constant temperature of 37°C. 1 ml of overnight culture was used to inoculate 100 ml LB medium (1:100 dilution) and allowed to grow at 37°C in a platform shaker at 225 rpm until an optical density reading (OD) of 0.4-0.6 was reached (approx. 2-3 hr). Centrifuging at 4,000 rpm at 4°C for 10 min collected cells. The supernatants were removed and discarded and the cell pellets were resuspended in 10 ml ice-cold 0.1 M CaCl<sub>2</sub>. The suspension was incubated on ice for 30 min. The cells were collected once more by centrifuging at 3,000 rpm at 4°C for 5 min. Again the supernatants were discarded and the cell pellets suspended in 4 ml 0.1 M CaCl<sub>2</sub> before incubating on ice once more for a further 2 hr. The cells at this point were ready for transformation and could also be prepared for storage.

To freeze, 140 µl DMSO (dimethylsulphoxide) was added per 4 ml of cell and gently mixed. Cells were incubated on ice for 15 min, before another 140 µl DMSO was added once more and gently mixed into the cells. Cells were then dispensed into aliquots in microcentrifuge tubes and liquid nitrogen was used to snap freeze sample before storing at -70°C.

#### **2.2.5.1.8 Transformation of Competent *E.coli*.**

LB/agar plates containing the antibiotic ampicillin (100 µg/ml) were prepared for each sample and control, prior to performing the transformation. Transformation was achieved as follows: 2 µl of each ligation reaction (~20 ng of DNA) was mixed gently with 100 µl of competent *E.coli* cells and incubated on ice for 30 min. The samples were then heat-shocked in a water bath at 42°C for a period of 2 min. and then placed on ice for a further 2 min. 300 µl LB medium was then added to each tube. The tubes were covered with foil and cells were allowed to recover by incubating at 37°C for 45 min in a platform shaker at ~150 rpm. 200 µl of each transformation mix was plated onto LB/ampicillin plates and incubated overnight at 37°C. Transformants were selected by their to ampicillin resistance.

#### **2.2.5.1.9 Analysis of Transformants**

In order to identify the clone of interest, several methods have been developed to screen transformants. The most popular of these methods is to isolate small-scale plasmid DNA preparations (minipreps) which can then either be sequenced or digested by restriction endonuclease enzymes of the desired clone, followed by fractionation of the products by gel electrophoresis. The method employed here was simple and less time consuming. Recombinants were screened by a straightforward PCR amplification across the point of DNA insertion, followed by a restriction digest reaction and gel electrophoresis. The length of the digest fragments identified recombinants.

#### **2.2.5.1.10 PCR Amplification of Transformants**

Transformed colonies were carefully picked from the LB/agar plates and suspended in 100 µl dH<sub>2</sub>O. They were PCR amplified using primers that annealed onto the vector sequence flanking the point of DNA insertion. Recombination could be examined by the size of the PCR products.

The programme 'Primer 3' (<http://www.genome.wi.mit.edu>) was used to find suitable primers for this amplification. As with all PCR amplifications, the magnesium and annealing conditions were first optimised for the oligonucleotide primer pair. A 25 µl

PCR reaction was performed, typically comprising of 3 µl resuspended colony, 1 unit *Taq* DNA polymerase, 10 pmol primer, 0.05% W1, 1.5 mM MgCl<sub>2</sub>, 200 µM dNTP, 5 M Betaine, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.05% (v/v) and dH<sub>2</sub>O in a microtitre plate with a mineral oil overlay. Thermal cycling was carry out on an MJ Research PTC-225 (Peltier Thermal Cycler) using the following cycling programme: 95°C for 2 min, then 30 cycles at 95°C for 30 sec, 57°C for 30 sec, and 72°C for 30 sec, followed by 72°C for 10 min.

The amplicons were separated on a 5% polyacrylamide gel, stained with ethidium bromide and then scanned on the Flourimager 595 (Molecular Dynamics). In those DNA fragments of correct size, the orientation of the DNA insert in the vector was examined by restriction endonuclease digestion.

#### **2.2.5.1.11 Restriction Endonuclease Digestion of Transformants**

The PCR products containing possible inserts were digested with the restriction endonuclease, Bgl II. The enzyme and its buffer were used as recommended by the manufacturer. For each digest reaction 15 µl of PCR product was mixed with 5 µl of standard digest mix containing 3 units of restriction enzyme 1 x NEB 3 (buffer provided with each enzyme) and dH<sub>2</sub>O. Paraffin-oil was used as an overlay to prevent evaporation and contamination of each reaction sample. The digests were incubated overnight (12-16 hr) at 37°C, on an MJ Research PTC-225 (Peltier Thermal Cycler).

The products of digestion were then fractionated on a 10% polyacrylamide gel, details as stated in section 2.2.4.4 of this chapter.

#### **2.2.5.2 Large Scale Preparation of Plasmid DNA**

Transformants with DNA inserts in the correct orientation were cultured and purified for transient transfection studies. Plasmid purification kits were employed for the purification of plasmid DNA. Two kits were utilized during this study; these were the Mobius™ 1000 Plasmid Kits (Novagen) and the HiSpeed™ Plasmid Purification Handbook (Qiagen). The purifications were performed using the manufacturers



recommended standard protocols. The procedures for both were similar, however the Novagen's kit was more effective and so will be described here.

Novagen's Mobius™ 1000 and UltraMobius™ 1000 standard protocol:

Transformed colonies were transferred into 20 ml universal tubes containing 5 ml liquid LB cultures containing 100 µg/ml ampicillin and incubated at 37°C for 6-8 hrs in a platform shaker 300 rpm. 0.5 ml of this culture was then transferred to a sterile 500 ml conical flask containing 100 ml LB medium supplemented with 100 µg/ml ampicillin and incubated at 37°C for overnight (12-16 hrs) in a platform shaker 300 rpm. The cultures were harvested by centrifuging at 5,000 x g 4°C for 10 min. Pellets were resuspended in 8 ml of Bacterial Resuspension Buffer by pipetting gently until the pellets were completely resuspended. 8 ml of Bacterial Lysis Buffer (1) was added to the samples and mixed in by swirling gentle for a few min until the lysate appeared clear and viscous. The lysate was incubated for 5 min at room temperature before 8ml of chilled Mobius™ Neutralizing Buffer (2) was added. Samples were mixed gently by swirling until they formed a flocculent precipitate. The samples were then incubated on ice for 5 min, during this time 10 ml of Mobius™ Equilibration Buffer (A) was applied to the reservoir of the Mobius™1000 columns to verify column flow is by gravity. Samples were centrifuged at 10,000 x g 4°C for 2 min. The supernatants from each sample were gently decanted into a Clearspin™ Filter unit. The supernatants were allowed to flow through the unit by gravity. The filtrates were then transferred to the column reservoirs of the Mobius™1000 columns and the clarifying lysate was once more allowed to flow slowly through the column by gravity. The plasmid DNA was eluted from the columns with 5ml of Mobius™ Elution Buffer (C). 3.5 ml Isopropanol was added to each sample to precipitate the plasmid DNA, samples were mixed gently by inverting tubes several times before centrifuging at 15,000 x g for 20 min at room temperature. The supernatants were removed and discarded and the pellets were washed with 3 ml 70% ethanol and centrifuged at 15,000 x g for 10 min at room temperature. Again, the supernatants were removed and discarded, each tube was inverted onto a clean paper towel to remove residual ethanol and the pellets were left to air dry. 600 µl TE buffer was used to resuspend each pellet. At the end of purification DNA concentration for each sample was measured on the BECKMAN DU-7000 spectrophotometer in mg/ml. Samples were stored at -20°C until required.

### **2.2.5.3 Mammalian Cell Culture, Transfection and Reporter Assay**

Two mammalian cell lines were used in this study. These were MALU and RAW cells, both of which are murine cell lines. MALU cells are derived from murine alveolar macrophages and were obtained from the Sir William Dunn School of Pathology Sciences cell bank. RAW cells were derived from the tumour induced by Adelson murine leukaemia virus from the organism *Mus musculus* (mouse).

#### **2.2.5.3.1 Propagation and Storage of Mammalian Cells**

Both cell lines were grown in RPMI 1640 supplemented by 4 mM L-glutamine, 10% foetal calf serum (FCS), penicillin and streptomycin. The cultures were maintained at 37°C in an atmosphere of 5% CO<sub>2</sub>. The cells were fed every 3-4 days and split once a week. All cell culture work was performed under sterile conditions and in a tissue culture hood in order to maximise asepsis.

For long term storage cells were suspended in RPMI 1640 containing 10% dimethyl sulphoxide (DMSO) and maintained in liquid nitrogen.

#### **2.2.5.3.2 Transient Transfection of MALU and RAW Cell Lines**

Transfection of the mammalian cells was carried out by electroporation. This is a quick and simple method that is widely used for introducing cloned DNA into a variety of cell types. Its development was based upon the original observations by Zimmerman *et al* (1983), who found that high-voltage electrical pulses could induce cell plasma membranes to fuse. It is thought that short electrical impulses create pores momentarily in the plasma membrane, which then allows the cell to take up endogenous DNA and other molecules from its surrounding environment. The efficiency at which the cell takes up DNA through electroporation varies greatly between trials and it is for this reason co-transfection was employed here.

The day before electroporation was performed the cultured mammalian cells were split and RPMI 1640 with L-glutamine and 10% FCS was added. The cell were then allowed to grow at 37°C in an atmosphere of 5% CO<sub>2</sub>.

For transfection, the cells were transferred to sterile 50 ml conical tubes and centrifuged for 5 min at 1000 rpm. The supernatants were removed and the pellets were washed by resuspending in RPMI 1640 and centrifuging for 5 min at 1000 rpm. The wash was repeated. Again, the supernatants were removed and the pellets were pooled together into 2 ml RPMI 1640. A haemocytometer and cover slip were set up. 1  $\mu$ l of these cells was diluted in 99  $\mu$ l RPMI 1640 and gently mixed, this was then loaded onto both sides of the haemocytometer and the cells were observed and counted with a low power microscope. The cells were resuspended in RPMI 1640 to give a concentration of  $2.8 \times 10^7$  cells/ml. An aliquot of  $1.4 \times 10^7$  cells/0.5 ml was placed into a 0.4 cm cuvette.

Plasmid reporter gene constructs (described in sections 2.1.5.1 and 2.1.5.2) were mixed with plasmid pRL-TK (which serves as a reference for transfection efficiency) at ratio of 30:1. The mixture was then added to the cuvette containing  $1.4 \times 10^7$  cells/0.5 ml described above. The samples were mixed gently and left at room temperature for 5 min. Each sample was then electroporated in a “Gene Pulse® II” electroporation system (Bio Rad) at 300 V (0.3kv) and 1050  $\mu$ F (capacitator). The cells were incubated on ice for 15 min and then transferred into a well in a 6-well culture plate containing 7 ml RPMI 1640 supplemented with 10% FCS. They were cultured for 24 hr at 37°C in an atmosphere of 5% CO<sub>2</sub>.

The day after electroporation cells were harvested from each well and transferred into 15 ml conical tubes and washed with phosphate buffered saline (PBS) by centrifuging at 1000 rpm for 5 min. The supernatant was decanted until no drops were left on the cell pellets. Each pellet was lysed with 0.3 ml 1 x PLB (Passive Lysis Buffer from Promega’s “Dual-Luciferase Assay Kit”).

The effects of certain stimuli on the transcriptional activity for the whole promoter region were also investigated. The drugs chosen were PMA (phorbol 12-myristate 13-acetate, a protein kinase C activator) and the statins including pravastatin (hydrophilic) and fluvastatin (lipophilic). For these studies the chemicals were added into the culture medium of transfected cells. PMA was administered at a concentration of 1  $\mu$ M (0.5

$\mu\text{l/ml}$  of 2 mM stock solution). Pravastatin was administered at a concentration of 0.1  $\mu\text{M}$ . Fluvastatin was administered at a concentration 0.4  $\mu\text{M}$ .

#### **2.2.5.3.3 Reporter Assay**

20  $\mu\text{l}$  of each sample was placed into a well in a 96 well microplate and the luciferase activities measured using the Dual-Luciferase Reporter Assay System (Promega) on the *rosy anthos* Lucy 1 microplate luminometer. In this assay, the activities of firefly and Renilla luciferases are measured sequentially from a single sample. The firefly luciferase reporter is measured first by adding Luciferase Assay Reagent II (LAR II) to generate a “glow-type” luminescent signal. After quantifying the firefly luminescence, this reaction is quenched, and the Renilla luciferase reaction is initiated by simultaneously adding Stop & Glo Reagent to the same tube. The Stop & Glo Reagent also produces a “glow-type” signal from the Renilla luciferase, which decays slowly over the course of the measurement. The luciferase levels are presented in arbitrary units after standardising the firefly luciferase levels against Renilla luciferase levels.

#### **2.2.6 Electrophoretic Mobility Shift Assay (EMSA)**

The electrophoretic mobility shift assay is commonly used to detect DNA-protein interactions. The method relies on the ability of a protein to bind to a radiolabelled DNA fragment (probe) *in vitro*. The products are separated from their unbound counterparts by electrophoresis on a non-denaturing polyacrylamide gel. The DNA-protein complexes can be identified from the unbound molecules by the extent of retardation in the mobility of the complex through the gel.

### 2.2.6.1 Preparation of Probes

For each probe, two complimentary single stranded oligonucleotides were ordered from MWG-Biotech AG (Table 2.8).

**Table 2.8: Oligonucleotide probes used for EMSA**

|             | Sequence (25-mer)                        | Alleles                        |
|-------------|--|--------------------------------|
| GT-MMP3-2F  | 5'-GGG AAA CAT CTT TCA GGG AAG TCG T-'3  | -1986 A allele<br>upper strand |
| GT-MMP3-2R  | 5'-ACG ACT TCC CTG AAA GAT GTT TCC C-'3  | -1986 A allele<br>lower strand |
| GT-MMP3-2bF | 5'-GGG AAA CAT CTT CCA GGG AAG TCG T-'3  | -1986 G allele<br>upper strand |
| GT-MMP3-2bR | 5'-ACG ACT TCC CTG GAA CAT GTT TCC C-'3  | -1986 G allele<br>lower strand |
| GT-MMP3-5F  | 5'-AAG TGG CTA TCA ACT GTG TGG GCA T-'3  | -1346 A allele<br>upper strand |
| GT-MMP3-5R  | 5'-ATG CCC ACA CAG TTG ATA GCC ACT T-'3  | -1346 A allele<br>lower strand |
| GT-MMP3-5bF | 5'-AAG TGG CTA TCA CCT GTG TGG GCA T-'3  | -1346 C allele<br>upper strand |
| GT-MMP3-5bR | 5'-ATG CCC ACAA CAG GTG ATA GCC ACT T-'3 | -1346 C allele<br>lower strand |

The oligonucleotides were suspended in dH<sub>2</sub>O at a concentration of 1 µg/µl. They were annealed as described in section 2.2.5.1.3 of this chapter. Double stranded oligonucleotide probes were labelled with the radioactive isotope [ $\gamma$ -<sup>32</sup>P]ATP by mixing 30pmol double stranded oligonucleotide, 3 µl [ $\gamma$ -<sup>32</sup>P]ATP, 5 units T4 DNA polynucleotide kinase, 2.5 µl 10 x T4 kinase buffer (500 mM Tris-HCl pH 7.6, 100 mM MgCl<sub>2</sub>, 50 mM DTT, 1 mM spermidine, 1 mM EDTA) and dH<sub>2</sub>O to give a total volume of 25 µl, and then incubating the solution at 37°C for 30 min. Following incubation a further 25 µl of dH<sub>2</sub>O was added to the sample before the labelled probe was purified from unincorporated [ $\gamma$ -<sup>32</sup>P]ATP and other molecules. Purification was achieved using a kit "Microspin G-25 column" (Pharmacia) as per manufacturers instructions. The radioactivities of the probes were then measured using a scintillation counter. The activity was typically found to be between 3 x 10<sup>6</sup> counts per minute (cpm) and 5 x 10<sup>6</sup> cpm.

### 2.2.6.2 Isolation of Nuclear Protein Extract From Cultured Cells

Nuclear protein extracts were prepared from cultured bronchial fibroblast cells. All solutions were first cooled on ice. With the exception of PBS all the solutions were supplemented with protease inhibitors including 0.7 µg/ml leupeptin, 16.7 µg/ml aprotinin, 0.5 mM phenylmethanesulfonyl fluoride and 0.33 µg/ml 2-mercaptoethanol were utilized.

The procedure was performed as follows: The cells were first cultured to a high density and harvested into the culture medium and collected by centrifuging at 1000 rpm for 5 min. The cell pellets were then washed with PBS and centrifuged once more at 1000 rpm for a further 5 min at 4°C. Cells were resuspended in 1 ml of ice-cold PBS and then transferred to microtubes before centrifuging at 9,000 x g 4°C for 1 min. The supernatant was removed and the cells were resuspended in 60 µl of ice-cold hypotonic buffer (10 mM Tris pH 7.3, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>) and centrifuged at 9,000 x g 4°C for 30 sec. The supernatant was removed once more and the cells were dissolved in lysis buffer (hypotonic buffer + 0.4% NP-40). The cells were then incubated on ice for 10 min before centrifuging at 9,000 x g 4°C for 1 min. The supernatant was removed again and the pellets were washed in 1 ml of ice-cold 0.02 mM KCl buffer (20 mM Tris pH 7.3, 21.75% glycerol, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20 mM KCl). They were then centrifuged at 9,000 x g 4°C for 1 min. The supernatant was removed and the pellets were resuspended in 15 µl of ice-cold 0.02 mM KCl buffer. 60 µl of ice-cold 0.6 mM KCl buffer (20 mM Tris pH 7.3, 21.75% glycerol, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 600 mM KCl) was then added drop-by-drop. The samples were incubated at 4°C for 30 min with continuous agitation. Again the samples were centrifuged at 9,000 x g 4°C for 15 min. The supernatant (nuclear protein extract) was transferred into a fresh microcentrifuge tube. The protein extract concentration was estimated by measuring the absorbance of a 1:40 dilution of the extract at 230 nm and 260 nm and calculated using a following formula: Protein concentration (µg/ml) =  $183 \times A_{230} - 75.8 \times A_{260}$ . The protein extracts were then divided into 10 µl-15 µl aliquots, snap-frozen in liquid nitrogen and stored at -70°C.

### **2.2.6.3 Protein-DNA Binding and Electrophoresis**

For the binding reaction, the following solutions were mixed together: 1 µl 10mM DTT, 1 µl 10 mM EDTA, 3.6 µl 0.1 M Hepes (pH 7.9), 3.2 µl 50% Ficoll, 2 µl 1 mg/ml poly(dI-dC) and 3 µg crude nuclear extracts (this was added last). Sterile dH<sub>2</sub>O was added to the solution to give a final volume of 16 µl. The binding mix was incubated on ice for 10 min. 1 µl of unlabelled oligonucleotide competitors (or 0.1 M NaCl) followed by 3 µl of radiolabelled probe (20,000 cpm/µl) was then added to the sample. It was mixed and incubated at room temperature for 20 min.

A 10% polyacrylamide gel was prepared prior to performing the binding reaction. The gel was approximately 1.5 mm in thickness with an acrylamide: bis acrylamide ration of 80:1 respectively and contained 0.25 x TBE buffer. The gel was pre-run for 1hr at 100 v, 4°C in 0.25 x TBE electrophoresis buffer. 2 µl of loading buffer with tracking dyes was mixed with the control samples only, as the dye may interfere with DNA-protein interaction, before they were fractionated by gel electrophoresis. Electrophoresis was performed at 200 v 4°C for 2 hrs. The gel was then dried onto 3 MM Whatman filter paper using a vacuum drier at 80°C for 1 hr 30 min. Once dry, the gel was autoradiographed. This was performed in complete darkness as hypersensitive film was used (BIOMAX MR from Kodak Scientific Imaging film). The gel was placed in a film cassette, an X-ray film was placed over it and the cassette was placed at -80°C overnight.

### **2.2.7 Association Studies**

Association studies were performed to investigate relations between the MMP-3 gene polymorphisms and the risk and progression of atherosclerosis. The subjects of these studies were 964 patients with coronary atherosclerosis defined by angiography and 310 individuals with a negative coronary angiographic finding. The DNA samples were extracted from peripheral blood from these subjects, using the salting out method and were quantitated using PicoGreen as described earlier in this chapter. MMP-3 genotypes were determined using the methods described in detail in sections from 2.2.4.2 to 2.2.4.4 of this chapter.

### **2.2.7.1 Statistical Analysis**

Analysis of data was carried out using SPSS for Windows (SPSS Inc. Chicago, IL, USA). The  $\chi^2$  test was used to examine deviations of genotype distribution from the Hardy-Weinberg equilibrium and to determine whether there was any significant difference in allele or genotype frequencies between patient and control subjects. A p value of <0.05 was considered significant. Logistic regression analysis was applied to study possible effects of individual polymorphisms on the risk of MI. Ordinal regression analysis was used to study possible effects of individual polymorphisms on the extent of coronary atherosclerosis measured by the number of coronary arteries with >50% stenosis. Stepwise logistic regression was used to study possible effects of haplotypes on the risk of MI or extent of coronary atherosclerosis. All analysis were carried out firstly without and then with adjustment for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes.



## Chapter 3

### Mutation Scanning of the MMP-3 Gene

#### 3.1 Introduction

##### A Comparison of Different Mutation Detection Methods

Most mutation detection techniques rely on comparisons between small fragments of PCR amplified DNA, ranging from sequencing, non-gel based capillary techniques to all gel based techniques. In general, sequencing is used to define a mutation once it has been proven to be present in an amplicon using a scanning method which does not identify the base change(s).

DNA sequencing can be used to detect mutations and defines the location and precise nature of the change. Direct sequencing analyses PCR products without prior subcloning into sequencing vectors such as M13 bacteriophage. If it were more rapid, accurate and efficient, it would be an ideal mutation detection technique. Computer software, improved reagents including fluorescent technology and a transition from slab gel towards capillary electrophoresis have enhanced this process and therefore, DNA samples can be screened at a much faster rate. Automated sequencing could become the primary technique used for mutation detection, however it is technically demanding and requires specialised equipment which can limit its availability in some laboratories <sup>125</sup>. For this reason, several screening methods for the detection of sequence heterogeneity have been developed. No single method is suitable for all situations and it is influenced by the expected nature of the mutation, size and structure of the locus in question, the degree of sensitivity required and resources available. All the techniques established so far have difficulty in guaranteeing 100% effectiveness, some are, however more efficient than others <sup>126</sup>.

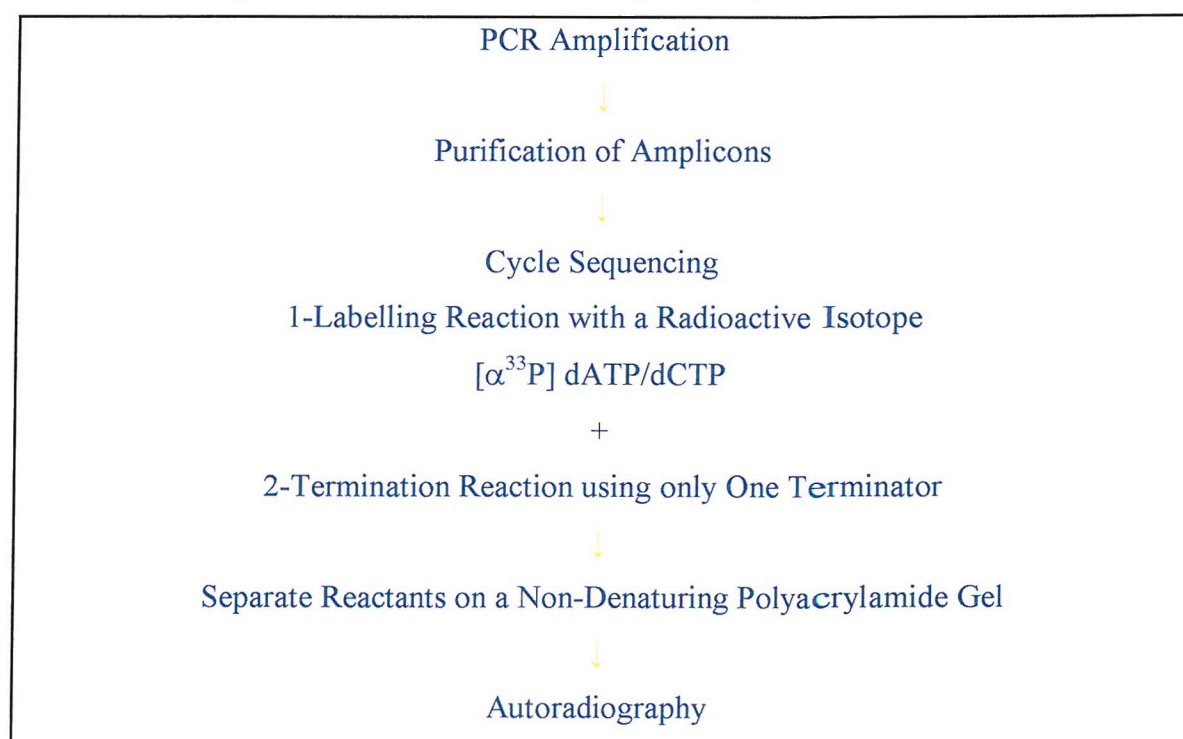
The screening technique used in this study was developed in 1992 by Sarkar *et al* <sup>122</sup> and is known as dideoxy fingerprinting (ddF). It was suggested to be more sensitive than single strand conformation polymorphism (SSCP), which has been widely used in mutation detection, and was the technique used to identify the 5A/6A polymorphism in the MMP-3 promoter gene sequence <sup>105</sup>. It is a hybrid technique between SSCP and Sanger's dideoxy sequencing. In ddF, performing one of four standard dideoxy-

sequencing reactions generates a ladder of bands. The products are resolved by electrophoresis on a non-denaturing polyacrylamide gel <sup>126</sup>. Autoradiography gives a fingerprint from which genetic variants can be identified by either a gain or loss of a dideoxy termination segment and/or by an alteration in the mobility of at least one of the termination segments that contain the mutation.

There are several reasons why ddF was chosen for scanning genetic variation in the MMP-3 gene: -

1. It is suggested to be a highly sensitive technique capable of detecting close to 100% single base changes <sup>122</sup>.
2. Its efficiency is independent of the length of amplified PCR products, allowing large sections of the gene to be amplified, prior to ddF reactions using internal primers.
3. It not only detects a mutation but also provides information about the location of the sequence change.
4. It uses one ddF terminator; thus it is quicker and cheaper than direct sequencing in screening mutations.

**Figure 3.1: - Flow chart showing the steps involved in ddF**



The choice of mutation scanning techniques is arbitrary, but some of the merits and limitations of each are summarised below.

### 3.1.1 Single Strand Conformation Polymorphism (SSCP)

SSCP was first developed in 1989 by Orita *et al* <sup>127</sup>. It is simple and one of the most commonly used techniques. It relies on the idea that each single stranded DNA and RNA nucleotide sequence folds in a particular way, in the event of a mutation the secondary conformation of the nucleotide will differ from its wild type due to altered intrastrand base pairing. When the samples are electrophoresed on a non-denaturing polyacrylamide gel, those samples containing variants will migrate at a different rate from the wild type. The motility difference can be seen as new band fragments on an autoradiograph.

SSCP is a PCR-based detection method. DNA samples need to be amplified first to ensure there is sufficient DNA for detection. To ensure efficiency of SSCP, PCR conditions, such as the  $Mg^{2+}$  concentration and thermo cycling conditions, must be optimised for each primer pair. DMSO and non-ionic detergents have also been found to be useful, for high yielding PCR products.

There are many variations to this technique, however the general format is that the DNA samples are labelled with a radioactive isotope such as  $\alpha$ -<sup>32</sup>P/  $\alpha$ -<sup>33</sup>P [dCTP or dATP], though other labelling methods can be used <sup>128,129</sup>. At the end of PCR, a loading buffer containing, formamide and tracking dyes is added and the samples are denatured by heating in a thermal cycler at 75°C for a few minutes before resolving the DNA fragments on a non-denaturing gel. The conditions of electrophoresis are important for SSCP sensitivity, including the temperature of the gel, the current the gel is electrophoresed at and the ionic gradient.

Detection rate of this method depends both on the size of the DNA fragment and the conditions used throughout the procedure. It is generally agreed that for DNA fragments between 100–200bp the detection rate is between 70-95%. The sensitivity of this method decreases when the size of the DNA template used is increased. For fragments of >300bp the detection rate drops to 50% <sup>126,130</sup>.

### **Advantages**

- No special equipment required.
- It is quick to do.
- The mutant bands are electrophoresed at a different rate from the wild-type.
- Non-isotopic detection methods available.
- Cost effective

### **Disadvantages**

- Size of DNA fragments used is limited.
- Position of mutation is unknown.
- Not all mutations are detected using this method.
- It is very sensitive to the conditions used.
- Gels can be difficult to interpret.

#### **3.1.2 Heteroduplex Analysis (HA)**

This technique was also developed in the late 80s'. It is a simple method that rests on the theory that there are conformational differences between double-stranded (ds) heteroduplex DNA and homoduplex DNA. Ds wild type DNA (homoduplex) and mutant DNA (homoduplex), will migrate on a gel differently from the mutant carrying heteroduplex<sup>131</sup>.

HA is a PCR-based technique. The general protocol is as follows: DNA strands are labelled during a standard PCR. Labelling is usually by incorporating  $\alpha$ -<sup>32</sup>P or  $\alpha$ -<sup>33</sup>P [dCTP/dATP] although other labelling modes can be applied. The PCR products are then diluted 2-fold with 1 x TBE buffer. The samples are denatured by heating to 95°C for about 5min and then left to cool slowly, to allow the reannealing of the DNA strand. A loading buffer is then added to the sample before carrying out electrophoresis on a non-denaturing polyacrylamide gel. An autoradiograph will show the mismatch heteroduplex molecules from the homoduplex bands<sup>126,132</sup>.

The detection rate of this technique is thought to be between 80-100% for DNA fragments of 193-800bp in length (1), though this relies greatly on the conditions used. Optimising the conditions improves the detection rate.

### **Advantages**

- Simple and quick to perform.
- No specialised equipment is required.
- Efficient.
- Non-isotopic detection available.
- Heteroduplex bands have an altered mobility to homoduplex bands so can be analysed easily.
- Close to 100% sensitivity.
- Cheap to run.

### **Disadvantages**

- The precise location of the mutation is unknown.
- The size of DNA fragments used is limited.
- Optimisation of conditions is required for a high detection rate.
- Not all mutations are detectable.
- A difficult technique to use with unknown mutations.
- Relies on the formation of heteroduplexes for homozygous detection.
- Interpretation of gels can be difficult.

### **3.1.3 Mutation Detection Enhancement (MDE)**

This is a gel matrix, similar to polyacrylamide, but has a higher sensitivity to conformational differences in DNA. It can separate DNA fragments on the basis of size and conformation. It was developed to increase mutation detection sensitivity by SSCP and HA. MDE has been shown to increase the probability of detecting sequence differences from 15% on polyacrylamide gels to 80% on the MDE gels.

### **Advantages**

- High sensitivity.
- Can be used for standard sequencing and on general electrophoresis apparatus.
- Quick to prepare.
- Cheap.
- Many staining methods are applicable for detecting mutations, including, ethidium bromide, SYBR Green I, silver staining as well as radioactive isotopes.

### **Disadvantages**

- Precise location of mutation is unknown.
- Fragment size is limited.

### **3.1.4 Denaturing Gradient Gel Electrophoresis (DGGE)**

This technique bases itself on the supposition that a single base change in a ds DNA nucleotide sequence will alter its melting properties slightly. This discrimination will therefore cause the fragments to move differently from its wild type, in a denaturing polyacrylamide gel. The ds DNA molecules are electrophoresed through an increasing chemical denaturant (formamide or urea) gradient. This causes the molecules to dissociate or melt. The segments that melt first are those with a lower melting point. A single point mutation, such as a substitution from A to C, will increase the DNA fragments' melting point. Thus the fragment will migrate further than its wild type form. Electrophoresis is driven from a low to a high concentration gradient <sup>133</sup>.

There are a number of variations to this process, but the conventional procedure used is as follows: -

A DNA molecule is amplified by PCR. One of the primers for PCR would have a 30-50bp GC-clamp; this is to ensure that the amplified DNA segments have a high melting point and thus a low dissociation temperature. They are then electrophoresed through a denaturing gel. This technique is adaptable to many different types of DNA labelling

methods and so does not require radioactivity. It can be applied to ds DNA molecules of lengths between 100-600bp with a high detection rate.<sup>126</sup>.

### **Advantages**

- Close to 100% efficient in detecting all mutations.
- Labelled probes are not necessary.
- Few steps to perform.
- Mutant form separates from its wild type.
- DNA modifications such as methylation can be detected (5).

### **Disadvantages**

- Special equipment is needed to regulate the temperature and for pouring gradient gel.
- GC-clamp expensive and problematic.
- Gel pouring can be cumbersome.
- Calls for the use of the toxin formamide.
- Fragment size is limited.
- The exact position of the mutation is unknown.
- Computer programme or preliminary experiments are required for sequence analysis.

### **3.1.5 Temperature Gradient Gel Electrophoresis (TGGE)**

Like DGGE, this method of detection works on the theory that a single base change in a ds DNA, RNA or protein molecule will have altered properties from its wild type form. It works on temperature- dependant motility controlled by an electric field. It can be used to detect conformational differences and sequence variations in nucleic acids and proteins. DNA samples are first amplified by PCR and then electrophoresed through a denaturing gel. The equipment used is a horizontal gel apparatus with an ancillary device attached to it (an electrical metal plate). Heating the plate at one end and cooling it at the other by two thermosetting baths produces the temperature gradient. The gradient can vary between 10°C and 80°C and can be run both perpendicularly or parallel to electrophoresis. For ds DNA/RNA molecules analysis is generally carried

out by hybridising with radioactive isotopes. Silver staining can be used to study protein molecules <sup>134,135</sup>.

### **Advantages**

- Mutation detection sensitivity is close to 100%.
- Isotopic probes are not necessary.
- Short protocol.
- Mutant forms separate from its wild type counterpart.
- DNA modifications can be detected.

### **Disadvantages**

- Specialised equipment is required.
- The location of the mutation is not known.
- Fragment size is limited.
- GC-rich clamp is expensive and can cause problems during PCR.
- Fragment size is limited.

### **3.1.6 Base Excision Sequence Scanning (Bess)**

Bess T Scan is an automated enzymatic cleavage screening method. It requires a special kit “BESS T Scan™ Mutation Detection and Localisation Kit (Epicentre Technologies) together with the LI-COR IR<sup>2</sup> Automated DNA Sequencer.” The general protocol involves amplification of the DNA samples using PCR. PCR is performed using one or two labelled primers in the presence of limited quantities of dUTP. On PCR completion the DNA, now with uracil integrated into its sequence, is enzymatically cleaved. Cleavage occurs at the sites of deoxyuridine. The fragments produced are then loaded onto a 6% denaturing “Long Ranger™ gel” (FMC CORP. and the fragments are electrophoresed on a LI-COR IR<sup>2</sup> Automated DNA Sequencer. The automated sequencer detects mutations and a DNA fingerprint <sup>136</sup>.

### **Advantages**

- Detects nature and location of mutation.
- Technique is quick and easy to perform.



- It is automated, so there is no time loss in analysing data.
- It is automated, making it a relatively fast mutation screening technique.

### **Disadvantages**

- Special equipment is required.
- Fragment size is limited (189-440bp).
- Conditions (excision enzyme reaction) need to be optimised to attain efficiency of this technique.

### **3.1.7 Ribonuclease Cleavage of Mismatch RNase**

This method uses heteroduplex mismatch to screen for mutations. The wild-type RNA fragments are radioactively labelled and these are hybridised with potential mutant molecules. RNA-DNA heteroduplex molecules are generated by PCR. The samples are then treated with RNase A, which cleaves the ssRNA molecules at the point of mismatch. The fragments are then separated by gel electrophoresis followed by autoradiography. A cleavage band of a given size indicates the presence and location of the mutation. Its sensitivity to mutation detection varies from 60-80%, and so this method has generally been disregarded and replaced with chemical cleavage (CCM), see below for details <sup>126</sup>.

### **Advantages**

- Location of mutation is known.
- Fragment length is unlimited.

### **Disadvantages**

- Detection rate varies between 60-80%.
- Cloning or PCR is required to prepare RNA probe.
- Difficult to end label probe.
- Samples need to be manipulated after RNase incubation.

### 3.1.8 Chemical Cleavage of Mismatch (CCM)

CCM, detects and locates mismatches in heteroduplex DNA molecules. The heteroduplex molecule generally constitutes a radiolabelled wild-type DNA molecule and a potential mutant DNA or RNA molecule. It is formed by first amplifying wild type and potential mutant DNA samples separately by PCR. The wild-type amplicons are electrophoresed on an agarose gel, the bands are excised and purified and then end-labelled using a radioactive isotope. After a few more purification steps to remove any excess solvents, the now labelled wild-type DNA fragments are mixed together with the mutant PCR products. A process of boiling and reannealing produces the heteroduplex molecule. The mismatched bases are then chemically modified using the Maxam-Gilbert's sequencing approach. Osmium tetroxide is used to modify mispaired thymine and hydroxylamine is used for mismatched cytosine bases. Labelled DNA is then cleaved at the sites of modification using piperidine. The cleaved products are then separated on a denaturing PAGE gel and analysed via autoradiography. Labelling the mutant DNA strand of the heteroduplex molecule will enable detection of mismatches in adenosine and guanosine bases. It should be noted that labelling only the wild-type DNA will give a detection rate of >95%. 100% detection can be achieved by labelling both wild-type and mutant strands in the heteroduplex molecules<sup>126</sup>.

#### Advantages

- Very sensitive technique: 95% effective when just the wild type DNA is labelled and 100% when both the wild type and mutant forms are labelled.
- Detects exact position of mutation.
- Reagents used are readily available and cheap to buy.
- Fragment size is unlimited; PCR products of up to 1.7Kb in length have been used successfully.
- Fluorescent probes can be applied.

#### Disadvantages

- Protocol is lengthy and laborious.
- Requires the use of toxic chemicals (Osmium tetroxide and Hydroxylamine).

- Limited unlabelled forms.

### **3.1.9 Denaturing High-Performance Liquid Chromatography (DHPLC)**

This screening method detects variants in reannealed DNA stands (heteroduplexes). It identifies single nucleotide substitutions and insertion/deletions directly from PCR amplified DNA without the aid of DNA sequencing. DHPLC requires specialised automated equipment, which is made up of: “An online degasser (such as SD-200), two high-pressure pumps, an electronic pressure module, a dynamic mixer, column oven with air-enforced heating and Peltier cooling, an automated sample injector and a computer programme to control the process and analyse the data.” It is performed in the following manner: DNA templates are amplified using a standard PCR. The PCR products are then loaded on to the column of the DHPLC instrument using an automatic sample injector. The products are denatured at a temperature of 95°C for 3min and then cooled down slowly to a temperature of approximately 65°C to generate reannealing of heteroduplex molecules in a thermal cycler. Heteroduplex molecules form only when a variant is present in the heterozygous state. The samples pass down the temperature gradient column and are then eluted from it, usually with a linear acetonitrile gradient. The start and end point of the gradient depends on the size of the PCR products. An UV absorbance detector then detects the eluted DNA fragments. The data is analysed and processed and the results are expressed in the form of a chromatograph<sup>137,138</sup>.

#### **Advantages**

- Highly sensitive method, detecting close to 100% of variants.
- Though not fully examined, fragment size is not limited; segments up to 1500bp have been screened successfully.
- It is a fast technique; DNA can be screened for mutations in 5-7min.
- High capacity for screening large DNA banks can screen 96 samples in 12hr.
- No gel set up.

#### **Disadvantages**

- Expensive specialised equipment is required.

- Column temperatures need to be optimised following PCR amplification.

### **3.1.10 Melt MADGE**

This is a new technique still in the process of development. Its theory is based on the same idea as DGGE and TGGE; In the event of a single base change in the nucleotide sequence the dsDNA molecules' melting properties will be altered from its wild type form. Melt-MADGE does not use a chemical gradient in its analysis of duplex molecules but relies solely on a temperature gradient. The procedure is very simple to perform. It consists of a standard PCR in which one primer is tagged with a 30-50bp GC rich clamp. Once PCR has been completed, a 15% sucrose solution containing a dye is added to each sample. The samples are loaded dry onto a denaturing urea/formaldehyde gel. The gel is then covered by carefully sliding a glass plate over it. Two elastic bands are used to hold the plates together and the gel is placed onto the temporal thermal ramp electrophoresis apparatus. More than one gel can be electrophoresed simultaneously, up to 10 or 12. The samples are electrophoresed through an increasing temperature gradient over a set period of time <sup>139</sup>.

#### **Advantages**

- Close to 100% efficient in detecting mutations.
- Its protocol is quick to perform.
- Unique in that it can scan large numbers of genomic DNA samples in one procedure, 1000 samples can be scanned in 30min.
- Basic DNA staining techniques can be applied, such as ethidium bromide, however SYBR Gold gives better definitions of bands.

#### **Disadvantages**

- GC-rich-clamp is expensive and can cause problems during PCR.
- Optimised conditions are required for high sensitivity.
- Specialised equipment required though not particularly expensive. The electrophoresis apparatus can also be used with other gel band methods such as DGGE.
- Precise nature of mutation is unknown.

- Can be difficult to interpret gels.
- Fragment size is limited, 100-500bp (300bp fragments are regarded as the best size for optimal sensitivity)

### 3.1.11 Direct Sequencing (DS)

Direct sequencing (DS) defines the exact location and nature of a mutation. It is usually required as the final step for most mutation scanning techniques. There are two standard DNA sequencing methods: Maxam-Gilbert's chemical reaction and Sanger's enzymatic chain-termination method. Originally these techniques required library construction, screening and sub cloning, in order to obtain sufficient and pure template for each reaction. Now direct sequencing can be achieved using PCR products of both genomic and DNA-RNA templates.

Sanger's sequencing method is generally preferred over Maxam-Gilbert's; mainly due to the toxic chemicals involved in the latter's protocol. Maxam-Gilbert's sequencing relies on nucleotide-specific chemical cleavage of DNA and is performed in the same way as CCM (see 3.1.8). Sanger's enzymatic chain-termination method is executed using either radiolabelled isotopes or fluorescent probes to detect the variants. These can be integrated into the nucleotide sequence by using either end-labelled primers or by using an isotopic dNTP in the second PCR reaction. Direct sequencing involves several steps. Ds DNA is first amplified by PCR, however ssDNA templates are required for sequencing. There are two modes generally used to generate ss nucleotides:

I) Asymmetric PCR, PCR products are reamplified, in this second reaction one primer is in excess of the other or may be absent.

II) A second PCR is performed with a 5'biotin labelled primer. The biotinylated ssDNA fragments is then captured by streptavidin or avidin bound to a solid support such as a magnetic bead. The ss nucleotides are then sequenced using all four dNTPs and a single ddNTP terminator. Increasing the concentration of ddNTP will result in more frequent chain elongation terminates. The fragments are separated on a denaturing

(urea) gel by electrophoresis. The products can be analysed through autoradiography  
140,141

### **Advantages**

- Displays not only the mutation but also the location, and precise nature of the change.

### **Disadvantages**

- Time consuming.
- Laborious to carry out.
- It lacks efficiency.
- Difficult to compare samples.
- Susceptible to unmanageable numbers of false positives and uncertainties where sequencing is not of perfect quality.

### **3.1.12 Automated Sequencing**

This is a modern version of the original sequencing method Sanger's dideoxy sequencing. Automated sequencing is performed with the aid of a sequencing machine such as the ABI PRISM®377 DNA Sequencer. Such instruments automatically analyse DNA molecules with any of four different fluorescent dyes. All four termination reactions are loaded into one gel lane in the sequencing machine. Fluorescently labelled DNA fragments are separated on the polyacrylamide gel, and are electrophoresed past a laser. The fluorescence is detected and signals are sent to the computer, processed and interpreted. A colour profile or a chromatogram, which displays peaks and corresponding base calls, is then produced by the computer. The software generally uses four different colours for each peak, which represent each base.

### **Advantages**

- Fast system for detecting mutations.
- Highly sensitive technique detecting close to 100% of all mutations.
- Provides information of the location and nature of mutation.
- Efficiency is independent of the length of fragment analysed.

### **Disadvantages**

- Requires expensive specialised equipment.

### **3.1.13 DNA Chip**

This method of DNA analysis is a hybrid between computer technology and molecular genetics. The DNA chip consists of a non-porous solid support, such as glass, onto which arrays, consisting of in the region of 400,000 different oligonucleotide probes, have been bound. The process is highly automated, in particular probe binding and data reading. Samples are analysed using a fluorescence detector, such as a phosphorimager. DNA chips were created to study genetic variation in DNA and in RNA expression. Different permutations of this technology are commercially available from an increasing number of companies. This technique is generally performed as follows: the DNA is extracted from the cells, the region of interest is amplified and fluorescently labelled, a different label being used for each of the four possible nucleotide bases. The tagged fragments are then hybridised to the DNA chip. A laser scans the chip and the pattern of fluorescence is used to determine the gene sequences of the hybridised targets <sup>140</sup>.

### **Advantages**

- High sensitivity to genetic variants.
- Rapid analysis.
- High through-put
- Cuts out human error, standard DNA analysis requires human intervention at several stages. DNA chip does not.

### **Disadvantages**

- Chips are costly.

### **3.1.14 Southern Blotting**

This is one of the oldest methods for detecting mutations. Its development represented a turning point for molecular biology. This mode of screening can detect large gene

alterations, point mutations and polymorphisms altering a restriction site. DNA fragments are produced by restriction endonuclease digestion. The fragments are separated according to size by electrophoresis on an agarose gel. The separated DNA fragments are transferred from the gel onto a sheet of nitrocellulose or derivative nylon sheet which overlays the gel. This process is known as blotting. The transfer is accomplished by capillary action, achieved by passing a high salt buffer through the gel from a buffer saturated wick. The buffer flows from the wick, through the gel, the nitrocellulose sheet and onto an absorbent paper layer placed on the top of the nylon filter. A fixation step follows which immobilises the DNA fragments on the nitrocellulose membrane. A hybridisation step follows, the now inert DNA fragments are incubated in the presence of radiolabelled single-stranded DNA probes, which contain the complimentary sequence of the molecules on the membrane. Any excess probes are washed away leaving only tagged molecules that have formed hybrids with the target DNA. The results can be observed by autoradiographing the blot. There are a number of variations to this technique: Dot and slot blots allow samples to be spotted directly onto a nitrocellulose membrane without the need for separation. RNA fragments can be analysed using northern blotting. Colony and plaque lifts, in which fragments can be lifted from agar colonies or phage plates respectively as DNA spots on a nitrocellulose membrane<sup>126,141</sup>.

### **Advantages**

- No detailed knowledge of the gene sequence is necessary.
- Fast method for screening mutations

### **Disadvantages**

- Single base changes can only be identified if located at restriction sites.
- Blotting can be messy.

The list above consists of brief descriptions of the most commonly used and efficient mutation detection techniques.



### 3.1.15 Dideoxy Fingerprinting (ddF)

ddF is a new screening method, a variation of SSCP <sup>122</sup>. This technique is a blend of SSCP and Sanger's dideoxy sequencing. It is simple to perform: DNA samples are first amplified using the standard PCR protocol. The products are labelled using a radiolabelled isotope ( $[\alpha-^{33}\text{P}] \text{ dATP/dCTP}$ ), when the labelling reaction is completed the samples are sequenced using one of the four standard dideoxy-sequencing reactions. Sequencing is achieved using any of the sequencing kits commercially available (T7 Sequenase version 2.0 DNA polymerase or Thermo Sequenase cycle sequencing kit both from Amersham Pharmacia Biotech). The sequenced reactants are then separated on a non-denaturing polyacrylamide gel by electrophoresis. An autoradiograph produces a fingerprint and the genetic variants can be analysed.

#### Advantages

- High sensitivity, close to 100%.
- Fragment size is unlimited.
- It sites the mutation and gives information of its location.
- No special equipment is required.
- Unlike direct sequencing, only one terminator is required, making it more cost efficient.

#### Disadvantages

- Precise nature of the mutation is unknown and so direct sequencing needs to be performed once a variant is observed.
- ddF is sensitive to the conditions used.

Mutation analysis can be carried out using both mRNA and genomic DNA. It is therefore important to consider not only the technique but also the material one wishes to use. There are several factors that need to be considered when choosing the type of template to use. Both mRNA and genomic DNA have advantages and disadvantages. These are presented in Table 3.1 below: -

**Table 3.1: - Starting materials for mutation detection: Their advantages and disadvantages.**

|             | ADVANTAGES  | DISADVANTAGES  |
|-------------|---|--|
| GENOMIC DNA | <ul style="list-style-type: none"> <li>• Easily obtainable from blood.</li> <li>• Can screen for mutation in the promoter, intron and exon gene sequence regions.</li> <li>• Both alleles are equally represented in genomic DNA.</li> </ul>  | <ul style="list-style-type: none"> <li>• Gene structure and genomic sequence information are required for its use.</li> <li>• Can only analyse small segments of the coding region at one time.</li> <li>• Requires a lot of PCR reactions.</li> </ul>   |
| mRNA        | <ul style="list-style-type: none"> <li>• Long segments of peptide coding regions can be analysed at a given time.</li> <li>• The use of mRNA does not require prior knowledge of the gene sequence.</li> <li>• Very little PCR reaction needs to be performed.</li> <li>• Prime material to use for X-linked traits.</li> </ul> | <ul style="list-style-type: none"> <li>• Difficult to obtain mRNA samples.</li> <li>• Only one allele is present in autosomal loci.</li> <li>• Only the exon region of the gene is represented in mRNA fragments, the promoter and intron regions are spliced out and cannot therefore be screened.</li> </ul> |

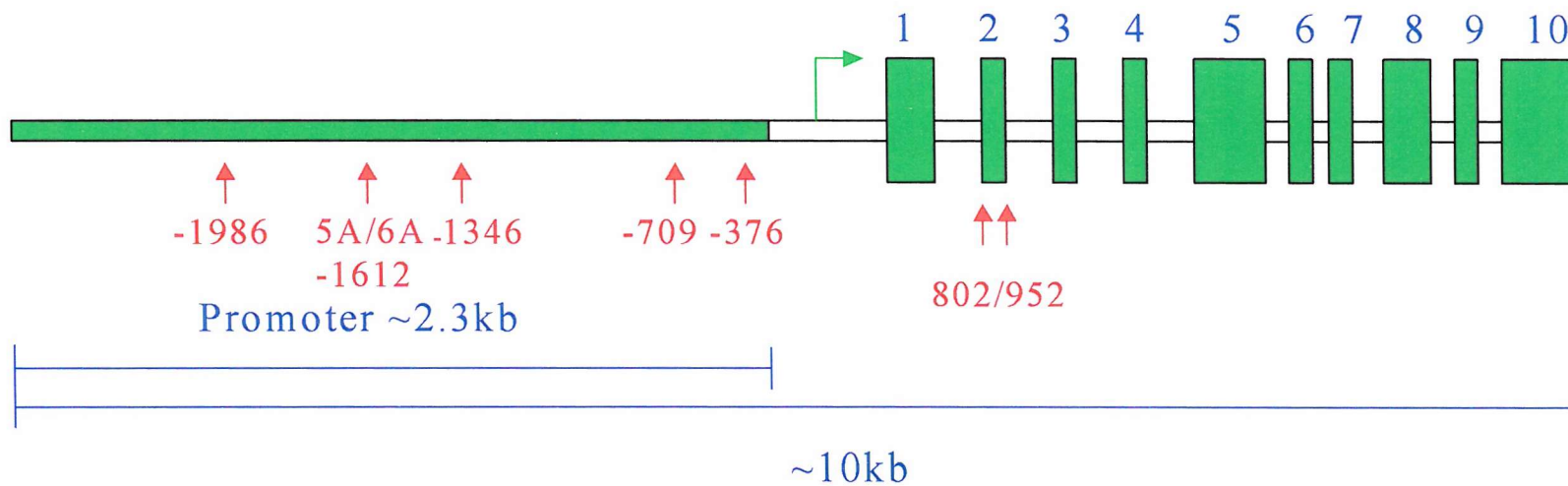
## **3.2 RESULTS**

The resultant autoradiographs of ddF were initially analysed qualitatively by eye for the occurrence of variants. Variants were identified by the presence or absence of bands present in the wild type (wt) samples or by an alteration in the electrophoretic mobility of the bands between the 20 samples screened.

The sequences portrayed on the autoradiographs were matched against the published sequences of the MMP3 gene and its promoter region, to identify the exact regions screened in the gene. Wild-type homozygote, mutant homozygote and heterozygote forms were then chosen for conformation of the polymorphisms by sequencing.

### **3.2.1 ddF Results**

The results of ddF are represented in Figure 3.2 and Table 3.2 below. The former is a schematic diagram of the MMP-3 genome with the polymorphisms identified across the gene. Table 3.2 gives the types of polymorphism and the suspected position of each variant identified.

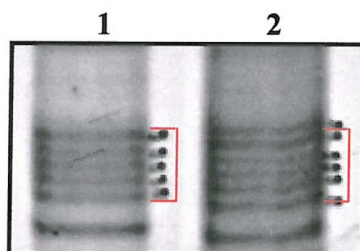


**Figure 3.2:** - Schematic diagram of MMP-3 gene variants identified through ddF followed by sequencing. Variants are given, in red, relative to the transcription start site

**Table 3.2: Qualitative data analysis of ddF autoradiographs.**

| PRIMER     | LENGTH OF SEQUENCE SCREENED | SAMPLES GENERATING ANALYSABLE RESULTS | NUMBER OF VARIANTS | POSITION OF VARIANTS | VARIANT CHANGE                          |
|------------|-----------------------------|---------------------------------------|--------------------|----------------------|---|
| MMP3-ddF2  | 254bp                       | 13/20                                 | 1                  | -1986bp              | Substitution<br>A→G                     |
| MMP3-ddF3  | 419bp                       | 20/20                                 | 1                  | -1612bp              | 5A/6A<br>polymorphism                   |
| MMP3-ddF5  | 236bp                       | 13/20                                 | 1                  | Near the<br>primer   | Substitution/<br>insertion/<br>deletion |
| MMP3-ddF7  | 249bp                       | 14/20                                 | -                  | ?                    | -                                       |
| MMP3-ddF8  | 239bp                       | 13/20                                 | 1                  | ?                    | Substitution                            |
| MMP3-ddF9  | 222bp                       | 14/20                                 | -                  | -                    | -                                       |
| MMP3-ddF10 | 245bp                       | 15/20                                 | 1                  | ?                    | Substitution                            |
| MMP3-ddF13 | 332bp<br>(Exon2<br>242bp)   | 17/20                                 | 2                  | 802, ?               | Substitution.<br>C→T<br>? Insertion     |
| MMP3-ddF14 | 191bp<br>(Exon2<br>242bp)   | 15/20                                 | 1                  | 950-955bp            | Deletion or<br>Substitution?            |

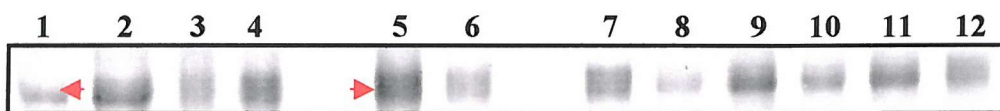
To ensure that the ddF technique was successful, the first region of the MMP-3 promoter to be screened was that which contained the previously identified polymorphism known as the 5A-1612/6A polymorphism. DdF identified this mutation clearly, which acts as a useful positive control for our study. Figure 3.3 below is a scanned section from the autoradiograph of this ddF reaction.



**Figure 3.3:** - Autoradiograph of ddF, sample 1, homozygous for the 5A allele and sample 2, homozygous for the 6A allele. Red brackets indicate the site at which the mutation occurs.

### 3.2.1.1. Polymorphism detected using MMP3-ddF2r

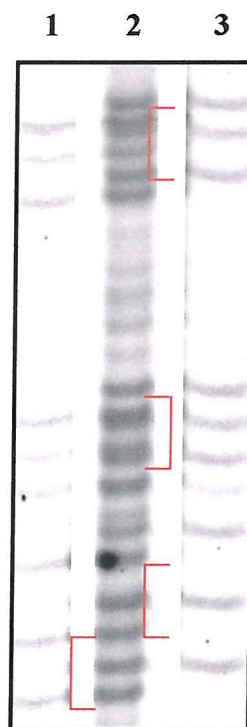
Using the primer MMP3-ddF2r, a single nucleotide polymorphism (SNP) at position – 1986bp upstream from the start of transcription was detected. As shown in Figure 3.4 although the bands in this region of the ddF autoradiograph are not sharp and rather faint, they differ in thickness between samples, indicating the presence of a sequence variation. For example, the band in sample 1 is significantly thinner than that in sample 5, suggesting that sample 1 is a homozygote whereas sample 5 is a heterozygote (it appears to contain more than one band in this region).



**Figure 3.4:** - Autoradiograph of ddF using primer MMP3-ddF2r. Arrows indicate a difference in band thickness between samples 1 and 5.

### 3.2.1.2. Polymorphism detected using MMP3-ddF5f

The MMP3-ddF5 oligonucleotide primer was used to screen the promoter region from -1365bp upstream from the start of transcription. The ddF result from this region showed a variation in mobility of the DNA sequences between samples. The exact type of mutation present however was unclear. All three alleles were clearly represented on the autoradiograph, as shown below in Figure 3.5.

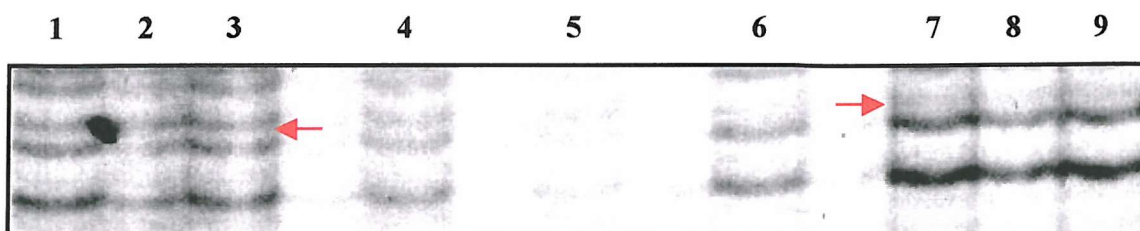


**Figure 3.5:** - Autoradiograph of ddF using primer MMP3-ddF5f from 3 samples each with a variation in DNA sequence by an alteration in mobility, indicated by the red brackets. Sample 2 contains all bands present in samples 1 and 3, suggesting that the former is a heterozygote whereas the latter are the two different types of heterozygote.



### 3.2.1.3. Polymorphism detected using MMP3-ddF8r

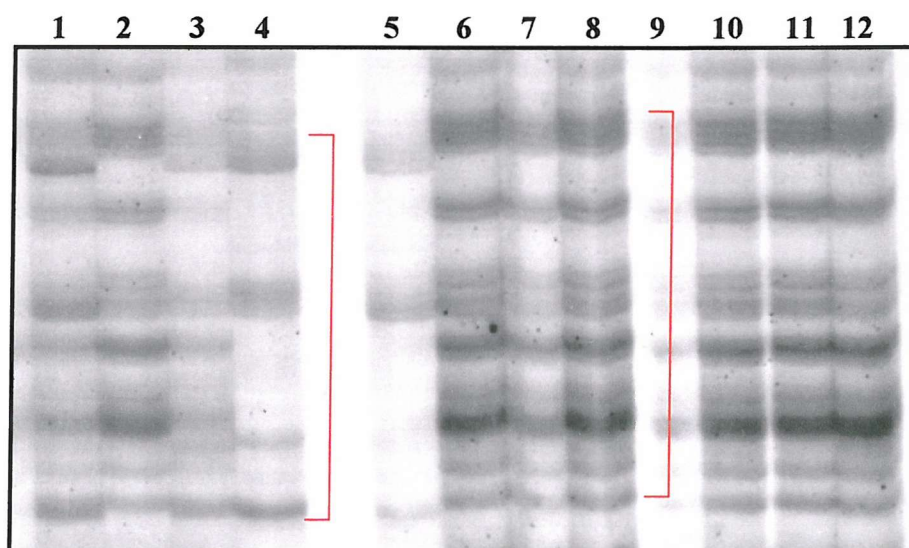
Using the reverse primer MMP3-ddF8r which begins at position -409bp, a variation was observed between samples in this region. As shown in Figure 3.6, an extra band is present in samples 1-5, suggesting that these samples contain a different DNA sequence from the other samples in the autoradiograph.



**Figure 3.6:** - Autoradiograph of ddF using primer MMP3-ddF8r. An extra band is present in samples 1-5 which is absent in samples 6-9. Arrows indicate the position of the bands presence/absence.

### 3.2.1.4. Polymorphism detected using MMP3-ddF10r

The resultant autoradiograph of ddF for this region of the promoter was quite blurred, however differences in the pattern of the DNA fingerprints between samples was evident as seen in Figure 3.7 below. There are three different patterns, pattern 1 is seen for samples 1 and 3, pattern 2 is seen for sample 4, and pattern 3 is seen for all other samples. These three patterns likely represent three different genotypes.

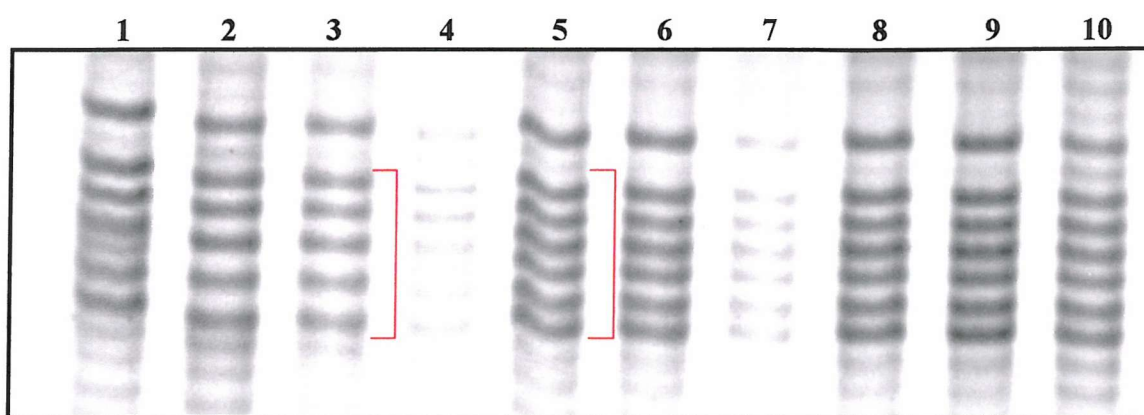


**Figure 3.7:** - Autoradiograph of ddF using primer MMP3-ddF10r, variation in the DNA sequences is present between samples and denoted by the brackets. Samples 2,6-12 have more bands present than samples 1,3 and 5.



### 3.2.1.5. Polymorphism detected using MMP3-ddF13f

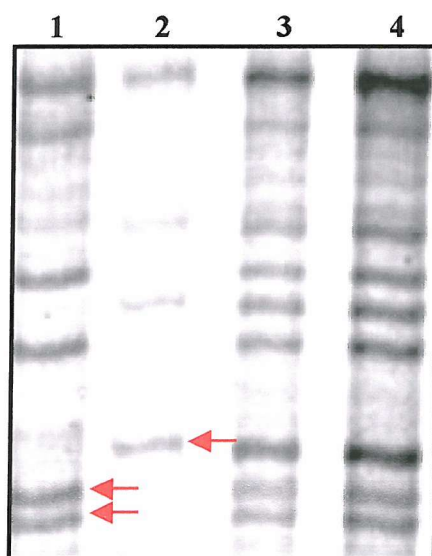
MMP3-ddF3 was the forward nested oligonucleotide primer used to screen exon 2 for polymorphisms. A variation was found amongst the 20 samples screened (Figures 3.8). On examination an extra band was identified in the adenosine runs. When matched against the published sequence, the position of this SNP was found to occur at 802, a single nucleotide substitution from G→A.



**Figure 3.8:** - Autoradiograph of ddF using primer MMP3-ddF13f. In samples 2-4 there is a run of 5 bands and in the other samples a run of 6 bands is represented. The brackets indicate two samples with varying adenosine bases present.

### 3.2.1.6. Polymorphism detected using MMP3-ddF14r

The reverse oligonucleotide primer used to screen exon 2 for the presence of a polymorphism was MMP3-ddF14. As shown in Figure 3.9, a variation was detected using this primer. The arrows indicate that a band is present in sample 2, but not in sample 1 whereas two bands are present in sample 1 but not in sample 2. Samples 3 and 4, on the other hand, contain all three bands mentioned above, suggesting that samples 3 and 4 are heterozygote whereas samples 1 and 2 are two different types of homozygotes. The altered sequence pattern further up the autoradiograph is caused by this SNP. The precise position of this variant could not be determined but the region is between 950 bp and 955 bp.



**Figure 3.9:** - Autoradiograph of ddF using primer MMP3-ddF14r. Arrows indicate the position of the SNP in each homozygote genotype.

### 3.2.2 Sequencing Results:

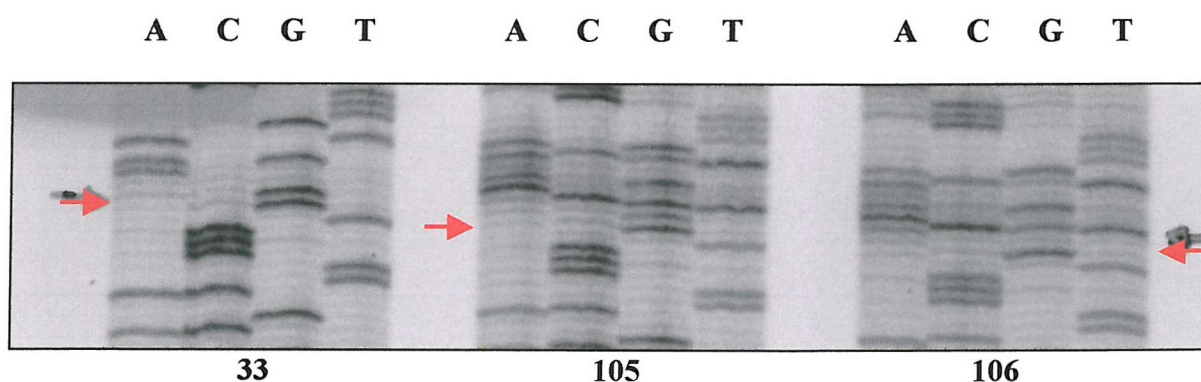
Following the ddF analysis, DNA sequencing was carried out to confirm the putative polymorphisms identified and to determine the nature of the polymorphisms.

#### 3.2.2.1 Polymorphisms identified using MMP3-ddF2(r)

Manual sequencing was performed on the following three samples:

|      |              |
|------|--------------|
| - 33 | homozygote   |
| 105  | heterozygote |
| 106  | homozygote   |

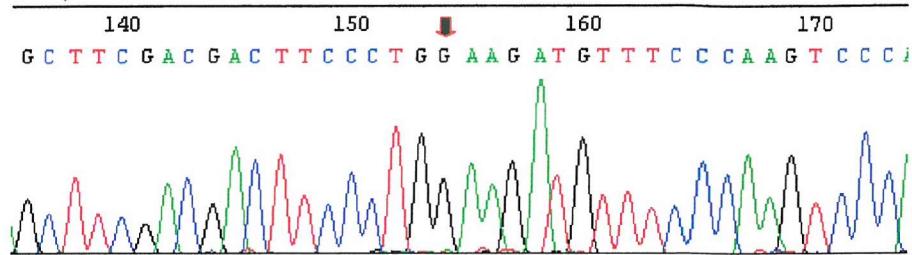
The results of manual sequencing are presented in Figure 3.10. There appears to be a base substitution from A→G at position -1986 from the start of transcription, as indicated by arrows in Figure 3.10. The results, however, are not clear and therefore automated sequencing was subsequently performed to confirm the base change. The results of automated sequencing are present in Figure 3.11, which clearly showed the A to G substitution at position -1986.



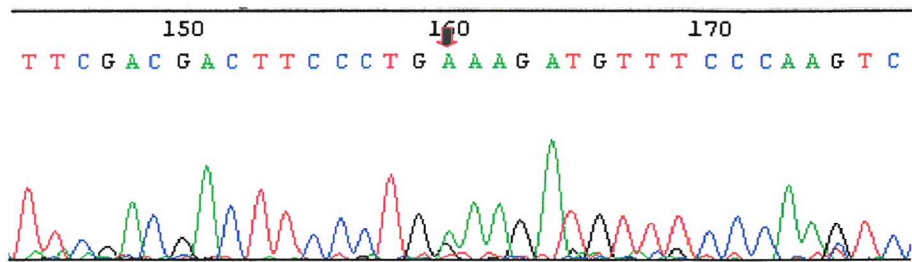
**Figure 3.10:** - Direct sequencing results for the A (-1986) G SNP identified in the promoter region of the MMP-3 gene. The red arrows signify the point at which the SNP occurs.

### MMP3 Promoter

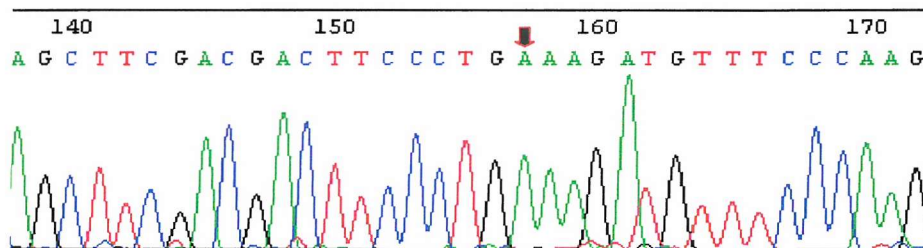
2B-G(-1986)



3B-G/A (-1986)



4B-A(-1986)



**Figure 3.11:** - Section of the chromatograph of three DNA sample showing the different genotypes of the A (-1986) G SNP. The arrows indicate the exact site of the SNP; sample 3B is heterozygote for the polymorphism.

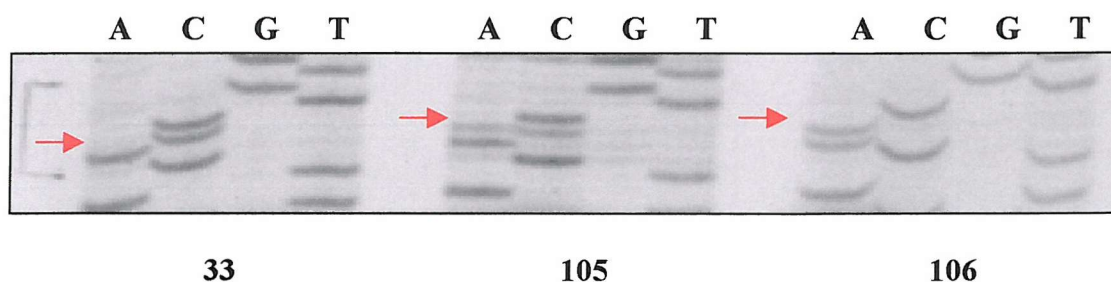
### 3.3.2.2 Polymorphism identified using MMP3-ddF5

Five of the 20 samples screened by ddF, using MMP3-ddF5, showed a variation in mobility of one of its alleles. The exact nature of this variation was difficult to determined using ddF alone, however it was clear that the mutation was close to the primer (MMP3-ddF5) and for this reason an alternative primer was chosen for sequencing. This primer was called MMP3-ddF5b and was 19bp upstream from the original MMP3-ddF5 primer.

Three samples were chosen for sequencing:

|     |              |
|-----|--------------|
| 33  | homozygote   |
| 105 | heterozygote |
| 106 | homozygote   |

As shown in Figure 3.12, DNA sequencing revealed that this variation is due to an A to C substitution 56bp downstream from the primer.



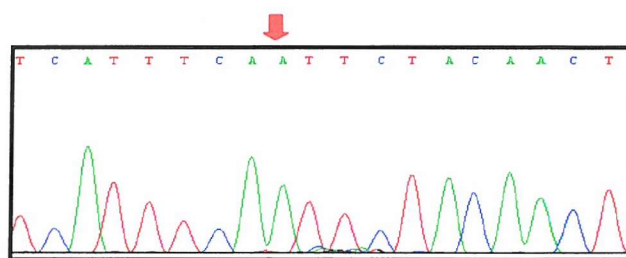
**Figure 3.12:** - Direct sequencing results for the A (-1346)C SNP identified in the promoter region of the MMP-3 gene. The red arrows point at the exact position of the SNP sample 33 = C/C (ATCACCT) genotype, 105 = A/C, 106 = A/A genotype (ATCAACT).



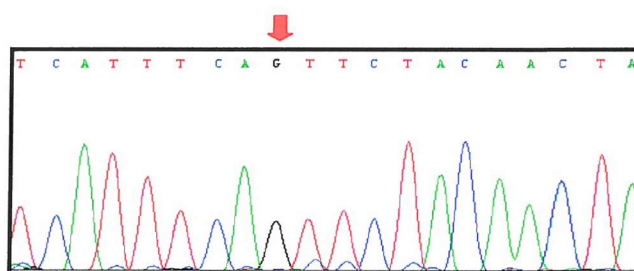
### 3.2.2.3 Polymorphism identified using MMP3-ddF8

Automated sequencing was carried out using the reverse primer MMP3-ddF8. Sections of the chromatographs encompassing the polymorphic site are shown in Figure 3.13. The sequencing results indicate that this polymorphism was caused by a substitution at position -709bp from an A to G.

5A-A -709



6A-G -709

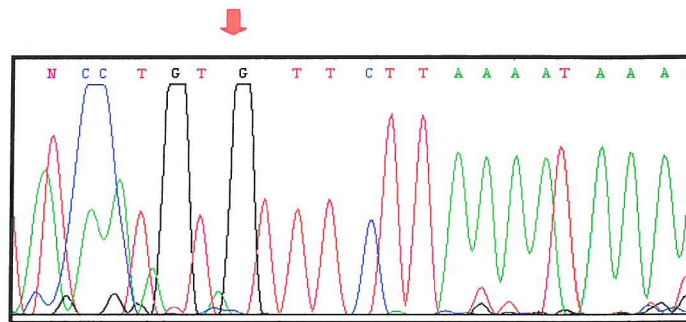


**Figure 3.13:** - Section of the chromatograph of two DNA samples showing the different homozygote genotypes of the A (-709)/G SNP. The arrows indicate the exact site of the SNP.

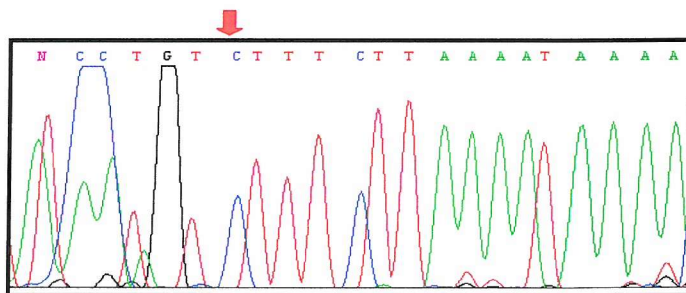
#### 3.2.2.4 Polymorphism identified using MMP3-ddF10

The reverse primer MMP3-ddF10 was used to identify the exact nature of the variant identified in the tail end of the promoter sequence. Sections of the chromatographs where the polymorphic site is situated are given in Figure 3.14. The sequencing results indicate that this polymorphism was caused to a G to C -376bp from the start of transcription.

5A-G-376



6A-C -376



**Figure 3.14:** - Section of the chromatograph of two DNA samples showing the homozygote genotypes of the G (-376)/C biallelic polymorphism. The arrows indicate the exact site of the SNP.



#### **3.2.2.5. Polymorphism identified using MMP3-ddF13**

Automated sequencing was carried out using the forward primer MMP3-ddF13. Sections of the chromatographs encompassing the polymorphic site are shown in Figure 3.15. The sequencing results indicate that this polymorphism was due to an A to G substitution at nucleotide position 802 and amino acid codon 10 in exon 2 of the MMP3 gene. In the A allele, this codon specifies a lysine (K). In the G allele, this codon specifies a glutamic acid (E). Glutamic acid is an organic acid, which has a negative side chain, while lysine is an organic base, which has a hydrophilic side chain. Therefore, the change from lysine to glutamic acid may have an effect on protein function.

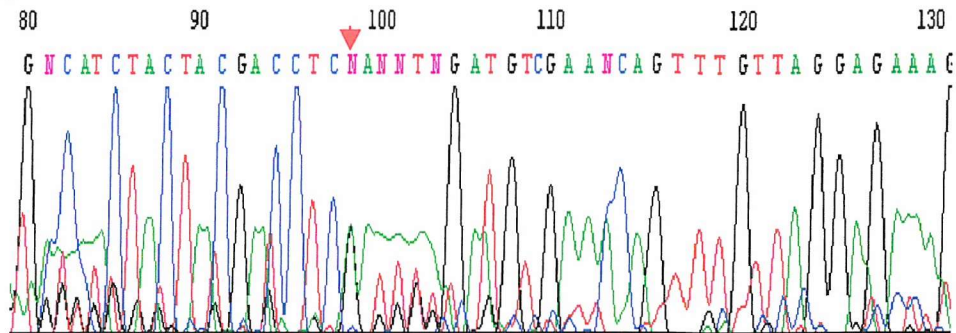
#### **3.2.2.6 Polymorphism identified using MMP3-ddF14**

Automated sequencing was also used to confirm the second polymorphism identified in exon 2. The results chromatographs (Figure 3.16) indicated this was an A to G substitution at nucleotide position 952 and amino acid codon 53 in exon 2 of the MMP-3 gene. The new codon however specifies the same amino acid (aspartic acid (D)) and thus it is a “silent mutation” which is unlikely to have an effect.

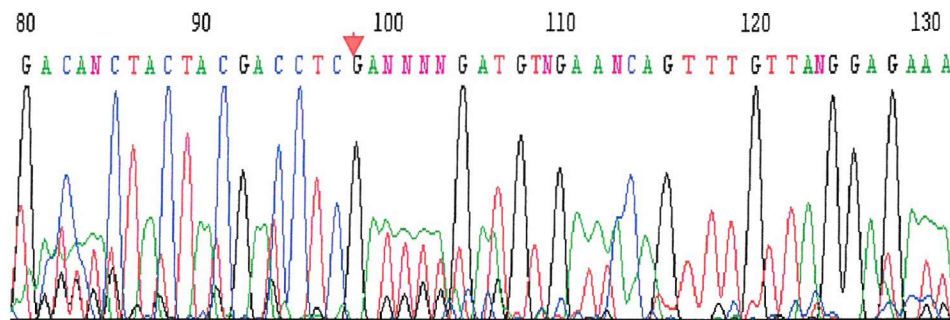


## MMP3 Exon 2

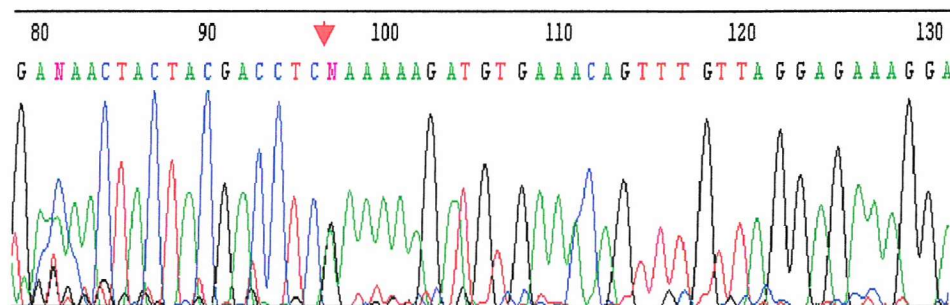
1B-ddF13



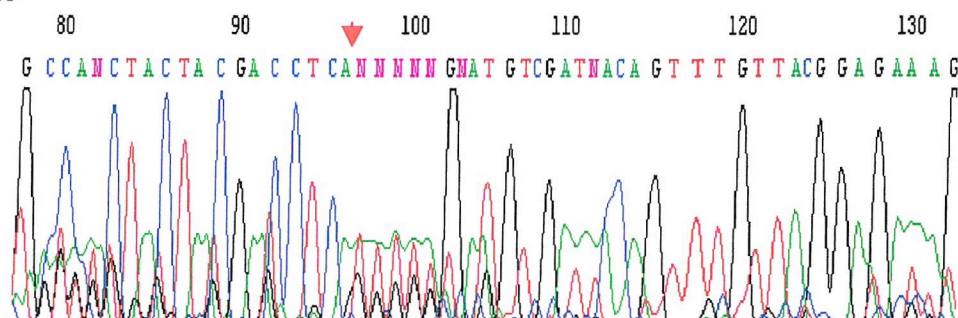
2B-ddF13



3B-ddF13



4B-ddF13

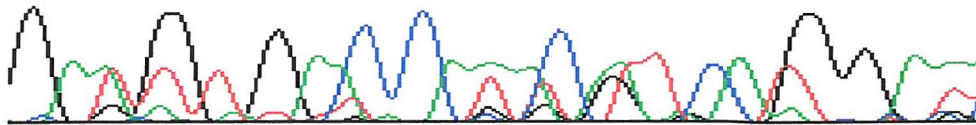


**Figure 3.15:** - Section of the chromatographs of four DNA samples showing the different genotypes of the G (802) A SNP. The arrows indicate the exact site of the SNP, sample 1B and 3B represent the heterozygotes for the polymorphism.

# MMP3 Exon 2

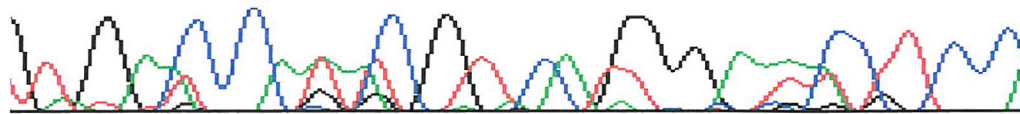
1B-ddF14

150 160 170  
G A N G T G A C C A N A C N T C A G G G A A



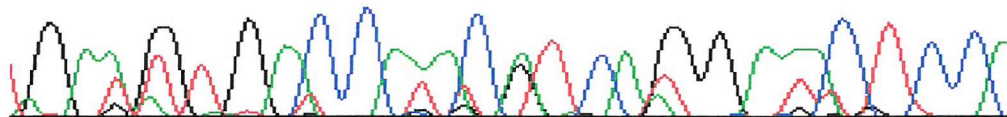
2B-ddF14

140 150 160  
T G A C C A N N C G T C A G G G A N C C T C C



3B-ddF14

150 160  
G A A G T G A C C A A C N T C A G G G A A C T C C A



4B-ddF14

140 150 160  
A G T G A C C A A A A N T C A G G G A A C T C



**Figure 3.16:** - Section of the chromatographs of four DNA samples showing the different genotypes of the G (952) A SNP. The arrows indicate the exact site of the SNP, sample B and 3B represent the heterozygotes for the polymorphism.

### 3.3. Discussion

Single nucleotide polymorphisms (SNPs) are the most common type of polymorphism and can have significant effects on the susceptibility of a disease or drug response<sup>142</sup>. In this study, MMP-3, a candidate gene for CHD, was systematically screened across the coding and regulatory regions for sequence variants. A modified version of the dideoxy fingerprinting (ddF) technique was employed to identify putative mutations which were then confirmed by direct sequencing.

DdF was first described by Sarkar *et al* (1992)<sup>122</sup>; their method entailed genomic amplification with transcript sequencing (GAWTS), in which they incorporated dideoxy CTP (ddC) to generate a ladder (fingerprint) of bands. This protocol involves four steps in which genomic DNA is amplified using a standard PCR protocol, T7 RNA polymerase then transcribes the amplicons to an RNA template. The transcripts are subsequently sequenced by reverse transcription using nested end labelled sequencing primers, which anneal to the ssRNA template<sup>119</sup>. A more recent development of this technique, is to PCR amplify the DNA target and then perform sequencing on the DNA, rather than RNA templates to generate their ladder of bands. DNA fragments of ~300-500bp in length are sequenced and the PCR primers are used for sequencing primers<sup>143,144</sup>. In contrast longer PCRs were applied in this study and internal primers were then employed for sequencing. The longer PCRs performed, produced amplicons encompassing the whole promoter region, exons 1-4, exons 5-6, exons 7-8 and exons 9-10. Consequently, only five PCRs were required instead of 20 or more, thus improving efficiency. ddG was the terminator recommended for this technique<sup>145</sup> however, in this study, the percentage of GC in each region was taken into account choosing the terminators. Both the promoter and coding regions had a greater percentage of As and Ts and so the terminators used here were ddA and ddT to increase the probability of detecting.

In this study, cycle sequencing was employed for ddF because the concentration of DNA template used was unknown and the supply was limited. Be that as it may, the sample yield was still weak and low, this made it difficult to differentiate true polymorphisms and bands that were just electrophoretic artefacts or secondary sequence bands produced during the sequencing reaction (ghost bands).

A total sequence length of 5034bp was screened for sequence variants in this investigation, from which 7 polymorphisms were detected. These include the previously identified 5A/6A variant, which was used as a control to ensure that the method employed worked, and a further six novel variants. The six new polymorphisms were all SNPs. SNPs are classically defined as single base pair substitutions in genomic DNA that exist in individuals of some populations. There are two subtypes: transitions and transversions<sup>146</sup>. Transitions are substitutions of a pyrimidine (C or T) by a pyrimidine, or a purine (A or G) by a purine. Transversions are substitutions of a pyrimidine by a purine or a purine by a pyrimidine. There are therefore two possible substitutions for transversions and one for transitions. One would therefore expect transversions to be twice as frequent as transitions, however it is transitions that are favoured in the human genome. This higher frequency is primarily caused by spontaneous errors made during DNA replication and by repair mechanisms in the cells. DNA replication is a complex process. Its fidelity depend on three factors: (1) DNA polymerases, (2) Exonucleolytic proofreading and (3) Mismatch repair. Replicative DNA polymerase duplicates the DNA sequence of a given genome and while these enzyme are very precise, their observed error rate is estimated as  $\sim 10^{-5}$ <sup>147</sup>. Most of these errors are corrected through the proofreading and mismatch mechanisms of the polymerase enzyme, however some are still missed. DNA replication errors are estimated to occur at a range of one per  $10^{-9}$ - $10^{-11}$  nucleotides synthesised, and these errors are thought to favour transitions over transversions<sup>146</sup>. Another explanation for this high frequency is thought to be due, in part, to the instability of cytosine residues so that 5-methylcytosine spontaneously deaminates and becomes thymine<sup>148</sup>. Of the SNPs identified in this study, four (A-1986/G, A-1346/C A-709/G and G-376/C) occurred in the promoter region and the other two were located in exon 2 of the coding region (A802/G and A952/G). Four out of the six SNPs are transitions, which is consistent with the general consensus (see Table 3.3).

The frequency at which these SNPs occur in the MMP3 gene is 1 in every 719bp, which corresponds to the general estimation for the human genome (1 in 250-1,000bp)<sup>149,150</sup>. On examination, these polymorphisms including the previously identified 5A/6A polymorphism appeared to be in linkage disequilibrium. Of the 20 samples screened by ddF, the same samples were found to possess the sequence variants identified (Table 3.3). This is not unexpected since the five promoter variants lie very close together within a 1590bp-sequence length. Together with the exon variants, this cluster of SNPs lies within

a 3kb sequence. The probability of recombination occurring between these SNPs is small. In one centimorgan (cM) of the human genome, there is a 1% probability of recombination during meiosis and typically 1cM is equivalent to the physical distance of approximately 1 megabase (Mb) of DNA genome<sup>146</sup>, however e.g. in human history (2,000yrs Europe), there have not been sufficient generations, (only 51,000) for linkage equilibrium to have been reached between such closely spaced SNPs.

**Table 3.3: - The MMP-3 promoter variants showing linkage disequilibrium**

| ALLELES                | SAMPLE/TRACK<br>NUMBERS | -1986 | -1612     | -1346 | -709 | -376 | 802 | 952 |
|------------------------|-------------------------|-------|-----------|-------|------|------|-----|-----|
| Homozygous<br>(wt)     | 1, 2, 14                | T     | 5A        | A     | A    | G    | A   | A   |
| Heterozygous           | 4, 12, 13, 15           | T/C   | 5A/6<br>A | A/C   | A/G  | G/C  | G/A | G/A |
| Homozygous<br>(mutant) | 5, 6, 8, 9, 10, 11, 16  | C     | 6A        | C     | G    | C    | G   | G   |

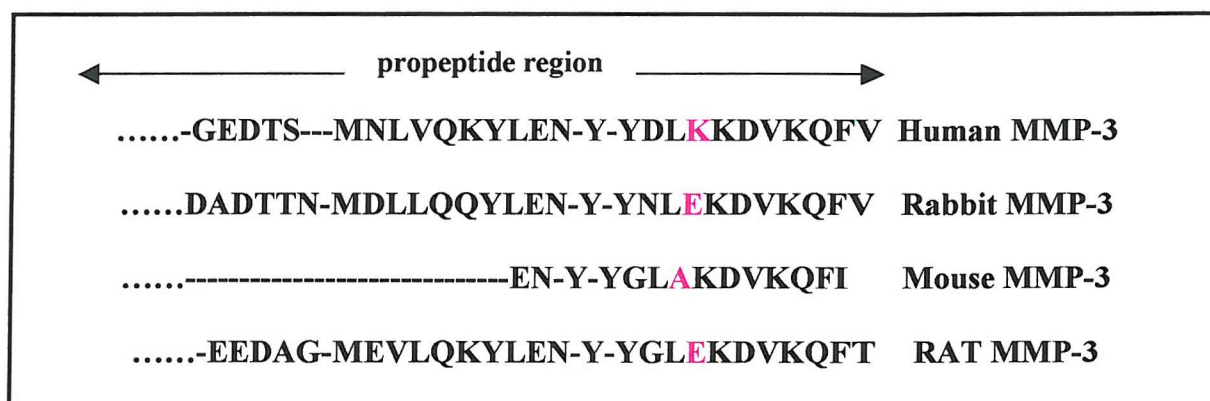
If linkage does exist between these polymorphisms then these variants do not occur separately and it is probable that they arose from the same ancestral chromosome. It is universally acknowledged that alleles at two tightly linked loci will show association only if they mark shared ancestral chromosomes. If the six confirmed polymorphisms in the MMP-3 gene are in linkage disequilibrium with the 5A/6A polymorphism, only two haplotypes will exist, which are represented in the Table 3.4 below.

**Table 3.4: - Haplotypes expressed by the seven polymorphisms in linkage disequilibrium.**

| Alleles                    | PROMOTER |       |       |      | EXON 2 |     |     |
|----------------------------|----------|-------|-------|------|--------|-----|-----|
|                            | -1986    | -1612 | -1346 | -709 | -376   | 802 | 952 |
| <b>Homozygote<br/>(5A)</b> | T        | 5A    | A     | A    | G      | A   | A   |
| <b>Heterozygote</b>        | C/T      | 5A/6A | C/A   | A/G  | G/C    | G/A | G/A |
| <b>Homozygote<br/>(6A)</b> | C        | 6A    | C     | G    | C      | G   | G   |



Of the two polymorphisms identified in exon 2 of the coding region, only one was found to create a change in the amino acid sequence. Arising as a single base change from A to G at position 802 relative to the start of transcription, it converts the organic base amino acid lysine (K) to glutamic acid (E), an organic acid. It is a conservative change and any effect it may cause is likely to be small. Further investigation showed that this amino acid is situated in the propeptide domain of the gene and lies close to the cleavage site of three exogenous proteinases, namely, plasma kallikrein (PKK), chymotrysin (CT) and human neutrophil elastase (HNE) <sup>50</sup>. These proteinases cleave short basic sequences on the surface of the molecules, which initiate a second autolytic cleavage, starting the cascade of events that leads to the activation of the MMP-3 precursor. Whether this codon change does effect MMP-3 activation remains obscure, however, data-mining studies suggest that this is unlikely. Comparing the protein sequences in human, rabbit, rat and mouse revealed that this codon is not conserved in evolution, where the amino acid K was found in human, E arose in both the rabbit and rat protein sequences and A (alanine) in mouse.



**Figure 3.17:** - MMP-3 sequence data from human, rabbit, rat and mouse <sup>50</sup>

To investigate whether the five promoter variants were located at certain consensus regulatory elements, both the wt promoter sequence and the sequence containing the variant were analysed using an in-silico method against a transcription factor binding site database (TFSEARCH: Searching Transcription Factor Binding Sites RWCP Parallel Application TPC Laboratories). A number of potential transcription factor binding sites were found to be abolished or created by the sequence variations. The transcription factor binding sites (TFBS) identified in the SNP regions of the promoter sequence are presented in Table 3.5 below:

**Table 3.5: -Transcription factor binding sites identified in the two haplotypes.**

| Promoter region                    | 5A allele                  |                           |                              |                          |       | 6A allele |         |                                |       |                              |
|------------------------------------|----------------------------|---------------------------|------------------------------|--------------------------|-------|-----------|---------|--------------------------------|-------|------------------------------|
|                                    | T-1986                     | 5A-1612                   | A-1346                       | A-709                    | G-376 | C-1986    | 6A-1612 | C-1346                         | G-709 | C-376                        |
| Transcription factor binding sites | <b>STATX</b><br>←TTCAGGGAA | <b>ADRI*</b><br>←TCCCCC   | <b>c-Myb*</b><br>ATCAACTGTG→ | <b>CdxA</b><br>→CAATTCTA |       |           |         | <b>DeltaE*</b><br>TATCACCTGTG→ |       | <b>SRY</b><br>TGTGTTTC       |
|                                    |                            | <b>NF-κB</b><br>←TGGTTTTT |                              |                          |       |           |         | <b>cap*</b><br>TCACCTGT→       |       | <b>Oct-1</b><br>TTATAATGTGTT |
|                                    |                            |                           |                              |                          |       |           |         | <b>MyoD</b><br>←TATCACCTGTGT   |       |                              |
|                                    |                            |                           |                              |                          |       |           |         | <b>Sn*</b><br>←TATCACCTGTGTG   |       |                              |

Eleven transcription factor binding sites were either abolished or produced by the presence of the five promoter variants. Five of these potential TFBS are present in the 5A allele-containing haplotypes but absent in the 6A allele-containing haplotype (STAT, c-Rel, ADR1 c-Myb and CdxA). Six other potential TFBS were found in the 6A allele containing haplotype (DeltaE, cap, MyoD Sn SRY and Oct-1). Some of these TFBS are not thought to affect the transcription of the MMP-3 gene, as the transcription factors (TF) for these sites are not found in humans: ADR1, for example, is a metabolic regulator found in yeast *Saccharomyces cerevisiae* and some fungi, Sn is a TF found in snails. The others are highly conserved TF with human homologs and a few are addressed briefly below. It should be noted however that none of the sites identified using this database have been demonstrated to regulate MMP-3 expression.

Oct-1 is a regulatory element that is important in tissue and cell specific transcription. It is highly conserved in evolution and is ubiquitously expressed in all cells. Oct-1 is a member of the POU family of transcription factors and is involved in the activation of snRNA promoter and some mRNA promoters. The consensus sequence for Oct-1 is identified in the C-376 allele but not in the G-376 allele.

The proto-oncogene protein c-myb is a transcriptional activator that specifically recognizes the sequence 5`yaac(g/t)g`3. It plays an important role in the control of proliferation and differentiation of haematopoietic progenitor cells and is also highly conserved with homologs in avian (chick), mouse and humans. Its binding site is abolished in the C-allele of the A-1346/C polymorphism.

The binding site for c-Rel, is abolished in the presence of the 6A polymorphism (-1612bp). C-Rel is a proto-oncogene protein and a member of the Rel family of transcriptional activators. It is highly conserved, with vertebrate homologs in mouse, rat, and chick as well as in humans. It is postulated that c-Rel, may play a role in differentiation and lymphopoiesis.

SRY is a human transcription factor protein that has homologs in other eukaryotes. It operates as a sex-determining region Y gene product and is also known as a testis determining factor.



The binding site for the transcription factor MyoD is also created in C-1346 allele. MyoD is a transcription factor found in mice. It is however highly conserved with homologs in drosophilae, clawed frog and humans. MyoD transcription factor sites, act as transcription activators of muscle specific genes and promote cell cycle exit by inducing CDKN1A. It is also postulated that they are involved in several lineages including myogenesis, osteogenesis and neurogenesis. There are three human homologs for MyoD, these are MEF1, MyoD1 and Myf-3.

Several factors have been characterised as suppressors of the MMP-3 gene. Transforming growth factor  $\beta$  is one of these. Studies on rat fibroblast cells revealed that TGF- $\beta$  represses both the constitutive and cytokine-induced expression of MMP-3 in a c-Fos-dependant mechanism. It involves a specific 10 bp TGF- $\beta$  inhibitory element (TIE) at -700bp relative to the transcription start site<sup>151,152</sup>. The consensus sequence for TIE<sub>-709</sub> GNNTTGGtGa and is believed to be unrelated to the AP-1 site. The A-709/G polymorphism identified here may therefore explain previous studies showing a 2-fold increase in promoter activity in the 5A/5A genotype when compared to the 6A/6A genotype<sup>99,105</sup>. Though this consensus sequence could not be matched with the human MMP-3 gene sequence, it is possible that this specific variant causes a change in a repressor sequence motif. The interstitial collagenase, fibroblast-type collagenase or MMP-1 has the same element present in its promoter region at -246 bp from the transcription start site<sup>50</sup>. A number of other TGF- $\beta$ 1-responsive proteins, namely u-Pa, elastase, proliferin and c-Myc have also been characterised as down-regulators of MMP-1 expression. Binding of Fos protein complexes at this site in someway counteracts the positive regulation occurring downstream at the AP-1 site.

A large number of SNPs have been identified but not confirmed by the Human Genome Project and the SNP Consortium. They are readily available in a number of public domain databases. Five SNPs were found in the databases, two of which correlate with the findings in this study and the remaining three are not confirmed (our laboratory has found that a number of putative SNPs in other genes retrieved from these databases are false signals). Current SNP detection methods are not ideal. Those used in this study, as already mentioned, presented difficulties when reading autorads, however those used by the Human Genome Project and the SNP Consortium also failed to identify four of

the six SNP identified in this study. For this reason it is possible that further SNPs may exist in the promoter and coding regions of the gene that have not yet been identified.

In summary, six novel polymorphisms have been identified and defined, four of which are located the promoter region (-1986bp, -1346bp, -709bp and -376bp) and two in exon 2 (802bp and 952bp) of the MMP3 gene. These SNPs appear to be in linkage disequilibrium. In silico analyses suggests that some of these SNPs might create or abolish transcription factor binding sites, but their importance is yet unclear. Equally, not all TFBS are included in the database and thus it is possible that these polymorphisms could affect certain TFBS that were not presented in the database and could not be identified.

## Chapter 4:

### Linkage Disequilibrium of Markers in MMP-3

#### 4.1 Introduction

Linkage disequilibrium (LD) is best described as a non-random association between particular allele at two closely linked loci. This kind of association is seen with some, but not all, closely linked loci. Alleles at two tightly linked loci will show association only if they mark shared ancestral chromosomes<sup>146</sup>. Therefore, significant LD is only seen if the particular alleles/chromosomes being investigated in a population originate from the same ancestral chromosome.

LD is highest for loci that are physically close in the genome because recombination between them is unlikely. Such closely linked alleles will therefore pass together through the family and thus be inherited through generation and population. Closely linked alleles are known as a haplotype and are useful in predicting the molecular pathology of an associated mutant gene due to LD. This concept is not new, it dates back to the beginning of the last century, (1917)<sup>153</sup>.

Over the past 10 years, geneticists have shown a huge surge of interest in the extent and distribution of LD. LD plays a major role in gene mapping. It can be used to localise disease genes to 1 to 2 cM regions of a chromosome (fine mapping) and can also be used to further refine the location of the disease gene through population based association studies<sup>154,155</sup>. Studying LD can increase our knowledge of the history and origins of the human genome<sup>156</sup>. It can also give us a greater understanding of homologous gene conversion and recombination events that occur in the human genome.

There are several factors that can affect LD and inter-loci physical distance. These are genetic drift, natural selection, demographic factors, variable rates of mutation, recombination and gene conversion: -

#### **4.1.1 Random Genetic Drift**

There is the tendency for gene frequencies to change by chance in the genetic composition of small isolated inbreeding populations. Genetic drift is only significant in small populations of stable size. In such instances, LD of haplotype frequencies are preserved. This is because they have a simpler population history, fewer founders and less population admixture. Such small isolated populations have been very useful in mapping genes for rare monogenic disorders. It is thought by some that the same advantages would hold for common diseases and for this reason many scientists are using such populations in their search to identify loci that cause complex diseases<sup>157,158</sup>.

#### **4.1.2 Natural Selection**

Natural selection is the process by which inherited genetic variation will result in differences between individuals with regards to their survival rate and their ability to reproduce successfully<sup>146</sup>. In other words, it is the fitness of an individual/population to adapt best to their environment. Natural selection is measured solely by their genetic make-up. Those with weaker genes (less fit) are prone to illness, cannot adapt to their environment, eventually dying out. Natural selection can work with or against certain sequences. Its action can force alleles that are in linkage to a higher or lower frequency and in so doing will ultimately, strengthen LD between two loci at a particular genomic region<sup>159</sup>.

#### **4.1.3 Variable Rates of Mutation**

Mutation is the process which produces a heritable change in DNA. It can result in a positive change to DNA by creating new and improved function or can result in disease, that can be passed on through generations by natural selection.

Point mutations are the most common form of mutations that occur in the human genome. These are changes that involve loss, duplication or alterations in small segments of DNA, which are often in the form of a single nucleotide base change. Single nucleotide substitutions are a common type of point mutation. Their frequency

reflects the fidelity of DNA replication and the efficiency of natural DNA repair mechanisms. Most of these mutations occur spontaneously and are generally unexplained. It is estimated that the rate of a single nucleotide substitution is in the order of one for every  $10^9$  to  $10^{10}$  bases replicated. The rate of mutation is markedly increased by exposure to environmental mutagenic agents, such as ultraviolet light and chemical agents.

The most frequent point mutations are depurination and deamination. In brief, depurination involves the removal of a purine base (adenine or guanine) from the deoxyribose molecule by a hydroxyl group. It is estimated that 5000 purine bases are lost each day from the DNA of each human cell. Cytosine is deaminated into uracil at a rate of 100 bases per human cell. Both these events can be recognised and removed by the depurination/deamination repair pathways<sup>146</sup>.

CpG islands are another form of point mutation. Regularly transcribed genes are marked by the presence of CpG islands. These are short DNA sequences approximately 1-2Kb in length, which consist of a large number of dinucleotide CpG<sup>146</sup>. A proportion of the cytosine residues that reside in CpG sequences are converted into 5-methyl-cytosine. This form of methylation only occurs when the complementary strand is already methylated by way of an enzyme. CpG methylation is inherited through generations. It is thought to play an important part in gene regulation and is generally related to repression of transcription. It is also believed to be involved in certain gene repressive mechanisms such as X-chromosome inactivation and imprinting genomes with their parental origins<sup>160</sup>. Methylation is lost over time by spontaneous deamination to give thymine. The DNA repair mechanisms that correct errors due to mismatch during replication, spontaneous chemical attack or damage to the DNA sequence, cannot recognise this type of mutation.

The human genome also has other mutational hotspots in its sequence. These include mononucleotide repeat regions, DNA-polymerase  $\alpha$ -arrest sites and/or other rare occurring motifs<sup>161</sup>.

#### 4.1.4 Recombination

LD is considered to be unstable over long periods of time. With successive generations, LD between two loci is typically decreased due to meiotic recombination, also known as independent assortment or crossing over. Thus, alleles are rearranged at different loci along the DNA sequence. There are certain areas in the chromosomes where the frequency of meiotic crossing over is elevated (recombination hotspots). These specific regions are thought to be initiation sites for recombination. Linked alleles in closer proximity to one another are less likely to be separated by recombination; this is because one crossover is generally considered to occur every 1 Morgan (cM) per individual chromatid strand.

#### 4.1.5 Demographic Factors

Demography is the study of different populations on a national, regional or local basis with regards to age, sex, migration patterns and survival rates. Different populations and communities can affect LD according to the age of the population, their size and recent expansion, migration and admixture, inbreeding and stability. Thus, populations with different demographic histories will often exhibit different patterns of LD. Inbreeding, for example, leads to low diversity levels and to increased LD. This was clearly shown by a recent study of eight families <sup>162</sup>. The investigation found homozygous chromosomal segments of distances greater than 10cM. Such distances are far greater than one would expect to find in randomly mating populations. Recent admixture of populations with different allele frequency will lead to an initial increase in LD between two unlinked loci <sup>163</sup>. This linkage will decay rapidly with random mating. It will however, still cause problems when fine mapping disease loci and marker alleles in population based association studies.

There have only been a few population comparisons of LD undertaken to date and in general, these studies have found that the level of LD is higher in newer populations such as in Finland, than those in older more established populations such as in Africa <sup>164-166</sup>. LD levels in relatively isolated populations such as Finnish and Sardinian populations, were not substantially greater when compared to more cosmopolitan populations such as those found in the UK and USA <sup>166</sup>.

#### **4.1.6 Gene Conversion**

Gene conversion is the replacement of one allele at a given loci by a copy of another allele, present elsewhere in the cell either on the homologous chromosome or another copy on the same chromosome. Very little is known about gene conversion. It is thought to arise due to a rare error in the DNA repair mechanisms that occurs during the recombination step in meiotic prophase of the cell cycle. Gene conversion events can change the pattern of LD and thus LD will be lost over time.

The age of an allele has important implications in human genetics because it is a major factor determining the extent of LD. New human alleles have a higher allelic frequency due to genetic drift and selection. It is in these newer alleles that one expects, therefore, to find high levels of LD. The ancestral human allele would be expected to show negligible levels of LD, due to multiple recombination events over many thousands of years <sup>146,167</sup>. Some scientists believe that common disease susceptibility alleles are probably ancient, predating the formation of modern human populations.

Leonard Kruglyak used computer simulations to estimate the extent of LD between SNPs in the genome. He based his calculations around several premises, 1) the human population remained small ( $N = 10,000$ ) 2) All SNPs arose once and from a single mutation and 3) all SNPs are neutral to selection. From his findings he concluded that LD is unlikely to extend further than 3Kb in a general population <sup>168</sup>. Though there has not yet been any real experimental evidence to refute this, it is the general consensus that LD can extend up to 500Kb.

#### **4.2 Linkage Disequilibrium Measurements**

There are several methods employed in calculating LD. The most commonly used method to measure LD to date was developed almost 40 years ago by Richard Lewontin <sup>169</sup>. Termed  $D$ , this statistic measure is employed to find the difference between two quantities, a pair of diallelic loci, A and B:

$$D = P_{11} - p_1 q_1$$

Where:  $P_{11}$  = the number of chromosomes alleles  $A_1$  and  $B_1$  occur together  
 $p_1$  = the frequency of  $A_1$   
 $q_1$  = the frequency of  $B_1$

Gametic association is another statistical measurement for  $D$ , similar to that of Richard Lewontin, but takes into account all four possible alleles found for a pair of diallelic loci<sup>160</sup>:

$$D = ad - bc$$

Where  $a, b, c, d$  = the haplotype frequencies  $A_1B_1, A_1B_2, A_2B_1$  and  $A_2B_2$  respectively.

If  $D$  differs significantly from 0 then LD is said to exist. However, if two alleles are found to be in linkage, they are not necessarily in association. This is because the degree of linkage between two given loci are dependant on recombination rate ( $r$ ), also know as the recombination fraction ( $\theta$ ) and time in generations. If we consider the two loci  $A$  and  $B$ , if the  $A_2$  allele first appeared as a mutation on a chromosome which was carrying the  $B_1$  allele, then the initial association seen between these two alleles would gradually decrease by recombination. This is because for each generation,  $D$  is decreased by a factor  $r-1$ . It is estimated that when  $r=0.01$ , then  $D$  will be halved with every 70 generations or every 2000 years ( $0.99^{70}=0.49$ ). Without natural selection, LD should only be seen with loci in very close proximity or recent mutations and thus continued LD is caused by selection for or against particular haplotype.

$D$  is dependent on the allele frequencies in the population. Its maximum value is given as  $D_{\max} = \min(p_1 q_2, p_2 q_1)$  and the minimum value is given measured by the formula  $D_{\min} = \max(-p_1 q_2, -p_2 q_1)$  where  $p_1 = f(A_1)$ ,  $p_2 = 1 - p_1 = f(A_2)$ ,  $q_1 = f(B_1)$  and  $q_2 = 1 - q_1 = f(B_2)$ <sup>169</sup>.

Lewontin also scaled  $D$  as  $D' = D/D_{\max}$



Hill and Robertson developed another formula for estimating  $D$ <sup>170</sup>, known as the standard disequilibrium coefficient. The value is given as  $R$  or  $\Delta$ :

$$\Delta = \frac{D}{\sqrt{(p_1 p_2 q_1 q_2)}}$$

or

$$\Delta = \frac{ad-bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

This value  $R$  or  $\Delta$  is equal to  $\sqrt{\chi^2}/N$ . Here the statistical value of  $\chi^2$  can be obtained from the  $\chi^2$  table of haplotype frequencies and  $N$  represents the total number of haplotypes in the population. By using this formula, the significance of  $R/\Delta$  can be tested. Significance can also be estimated using other methods such as that devised by Zaykin *et al*, this method employed permutation to evaluate the significance of LD<sup>171</sup>.  $R/\Delta$  is often squared and in doing so, arbitrary signs introduced when the marker alleles are labelled, are removed.

Another common scaling of  $D$  is Levin's population attributable risk,  $\delta$  and is given by  $D/(q_1 P_{22})$  (1953). Here  $p_1$  represents the population frequency of a disease allele,  $B_1$  and  $P_{22}$  is the chromosome frequency of the marker allele  $A_2$  and the normal,  $B_2$  allele<sup>172,173</sup>.

Though  $D$  and  $R$  are the traditional methods used to measure LD, several studies have shown that, more often than not, their estimated values are unreliable because they depend on allele frequencies<sup>174,175</sup>. These studies compared different methods of statistical measurements of LD in the analysis of two loci. For estimations of physical distance they found that  $D'$  and  $\delta$  gave more reliable results.

The majority of DNA sequence variants are in the form of SNPs. Several SNPs have been identified and characterised in this study. In this chapter, strength of LD between

these SNPs will be investigated. Following this, a series of functional studies will be performed on the five promoter polymorphisms to identify any possible effects they may have on transcriptional regulation, before performing a series of case-control studies to see if any of these SNPs have a bearing on atherosclerosis.

Conventional SNP genotyping methods involved restriction fragment length polymorphism (RFLP). However with the advent of PCR in the 1980's, a number of methods have been developed that have improved SNP genotyping. The most commonly used methods are PCR based, for example, restriction enzyme (RE), amplification refractory mutation system, (ARMS) and mini- and micro-sequencing.

Micro array diagonal gel electrophoresis (MADGE) is one method that can be used to separate products of SNP genotyping that involves target amplification of each SNP by PCR. MADGE is an agarose/polyacrylamide gel with standard wells arranged in the format of industry standard 96, 192, 384 or 768 well, microtitre plates that have been rotated anticlockwise by 71.6 degrees relative to the electrode <sup>139</sup>. It enables samples to be loaded in close proximity to each other and allows for subject separation to occur for genotypes to be scored.

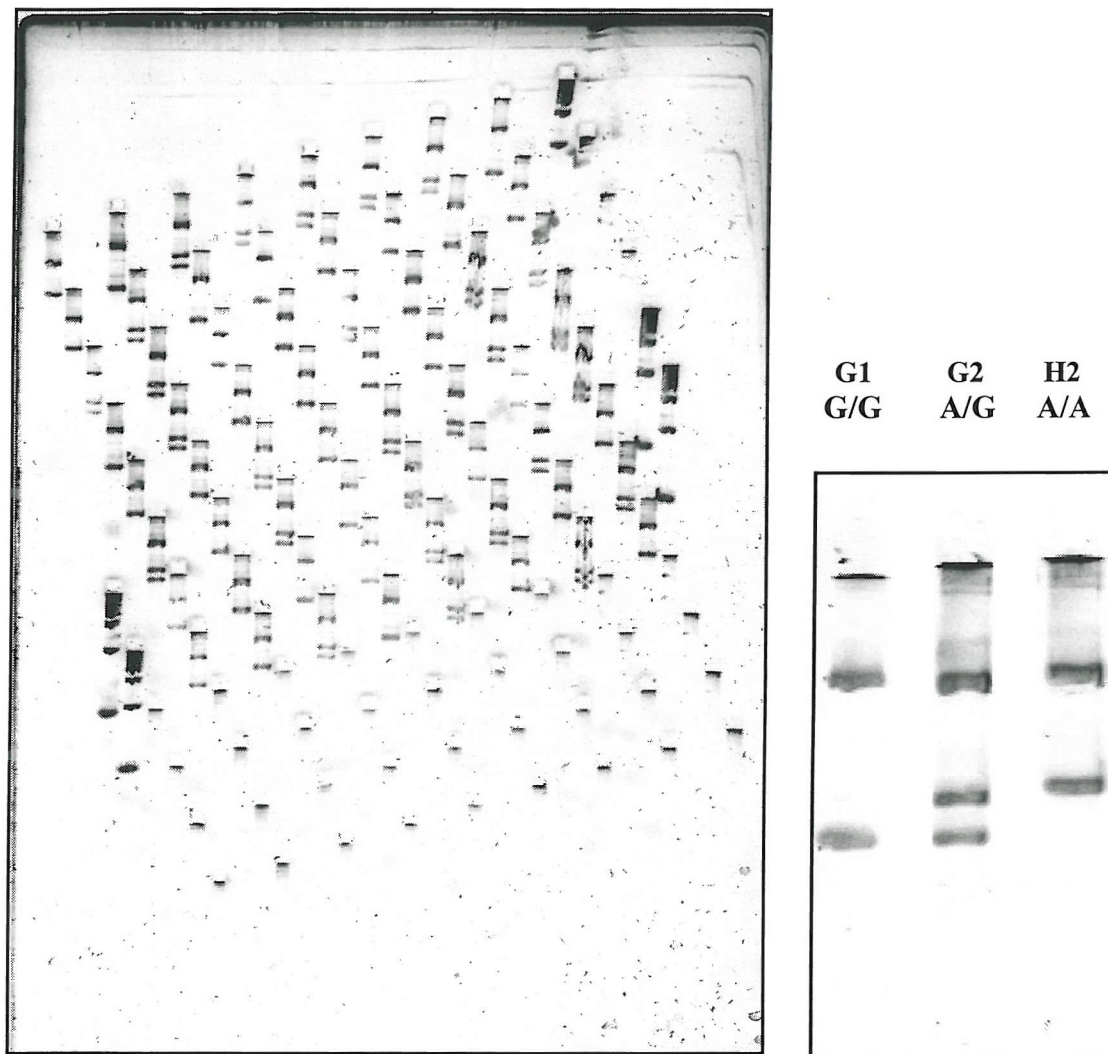
### 4.3 Study Design

Two sets of DNA samples were employed for this investigation. The first consisted of thirty-two DNA samples from unrelated individuals, duplicated to give a total number of 64 samples. The small sample set was used in an initial pilot study. The second set was a larger cohort of 1205 unrelated individuals of whom 913 were patients with CHD defined by angiography and 292 were controls (see Chapter 5 for details). Both sample sets were recruited at Southampton General Hospital. Each subject was PCR amplified. The amplicons were digested with an appropriate restriction enzyme that cuts one of the two alleles and the fragments separated on a polyacrylamide MADGE gel. DNA fragments were stained with the fluorescent dye vistra green and then viewed using a Flourimager 595.

Restriction sites at the point of each polymorphic site were identified using the website Cutter at <http://www.medken.gu.se/cutter>. The sequences flanking, and including the variant sites in both alleles, were entered as input sequences. Only sites that were present in one allele, but absent for the other were used. In cases where the SNPs proved to be troublesome, that is, were not found to either destroy or create a restriction site, then the introduction of a base change to produce a restriction site was executed.

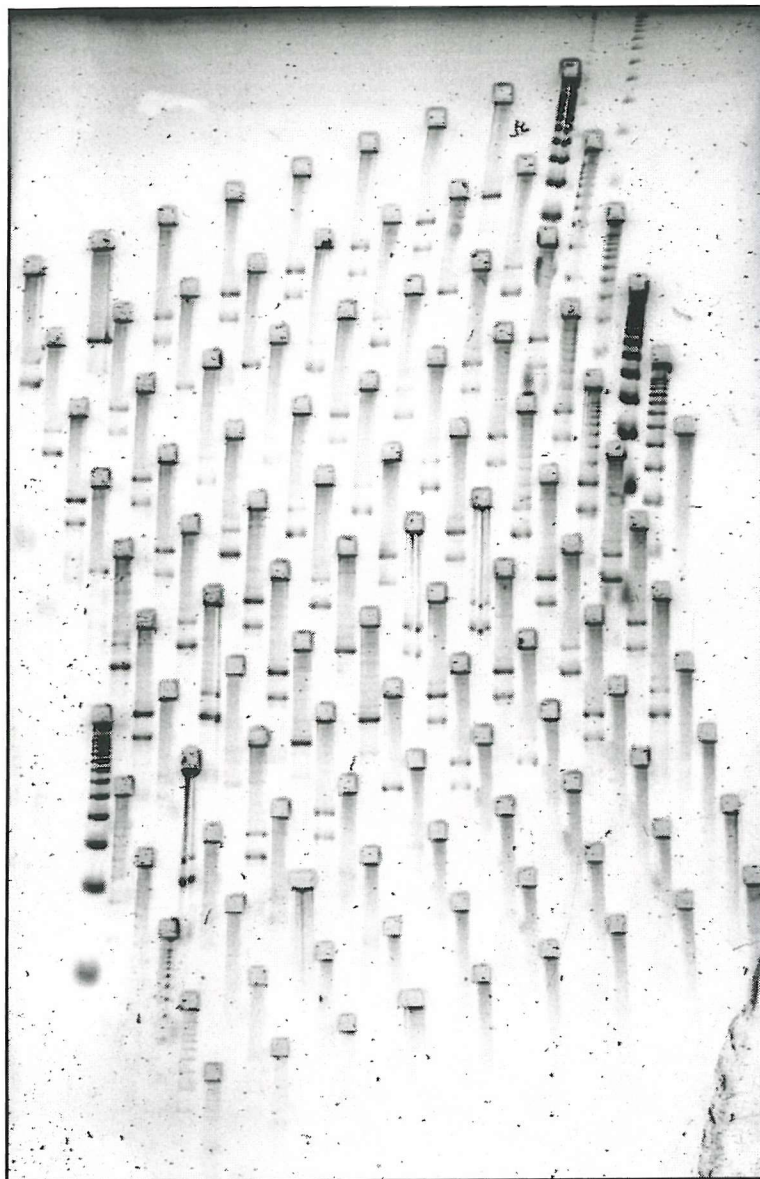
#### 4.4 Results

Representative results for the seven polymorphisms are shown in Figures 4.1-4.7. The DNA samples analysed were stored in an array format (A-H (1-12)) and were therefore loaded onto the gel in the sample way. Rows A-H were loaded from right to left and columns 1-12 from top to bottom.

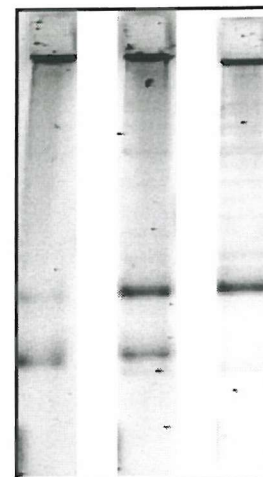


**Figure 4.1: BstN1 digest for the MMP-3 promoter polymorphism A-1986/G**

A 505 bp fragment was produced during amplification. Digestion with BstN1 cut this fragment in two places for the G/G genotype (three fragments of length 302 bp, 181 bp and 22 bp) and once for the A/A genotype (two fragments of length 302 bp and 203 bp).

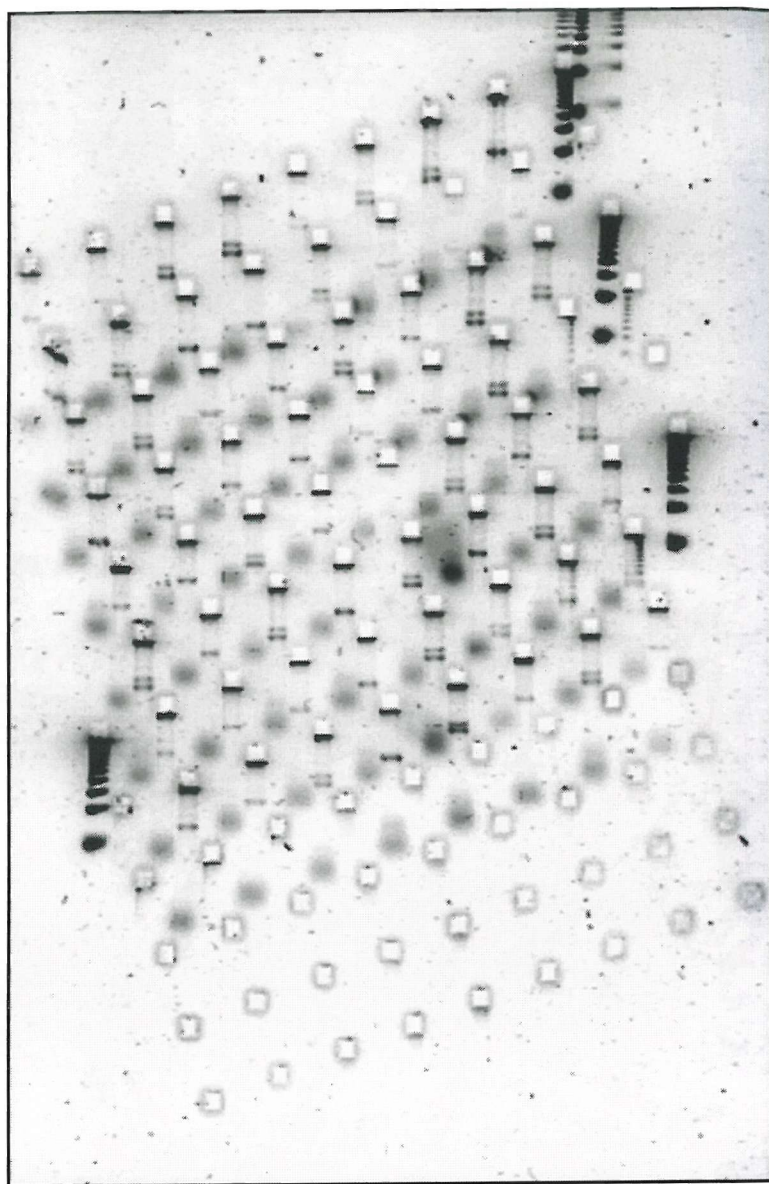


|    |       |    |
|----|-------|----|
| E4 | C7    | E7 |
| 5A | 5A/6A | 6A |

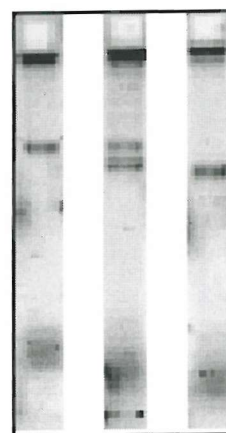


**Figure 4.2: Xmn1 digest for the 5A/6A MMP-3 promoter polymorphism**

A 118 bp fragment was produced during amplification. Digestion with Xmn 1 cut this fragment at the site of the 5A/5A genotype (two fragments of length 97 bp, and 21 bp) while leaving the 6A/6A genotype uncut. The primers used for PCR amplification mask the 21 bp fragments and so only the larger DNA fragments can be used for scoring the genotypes.



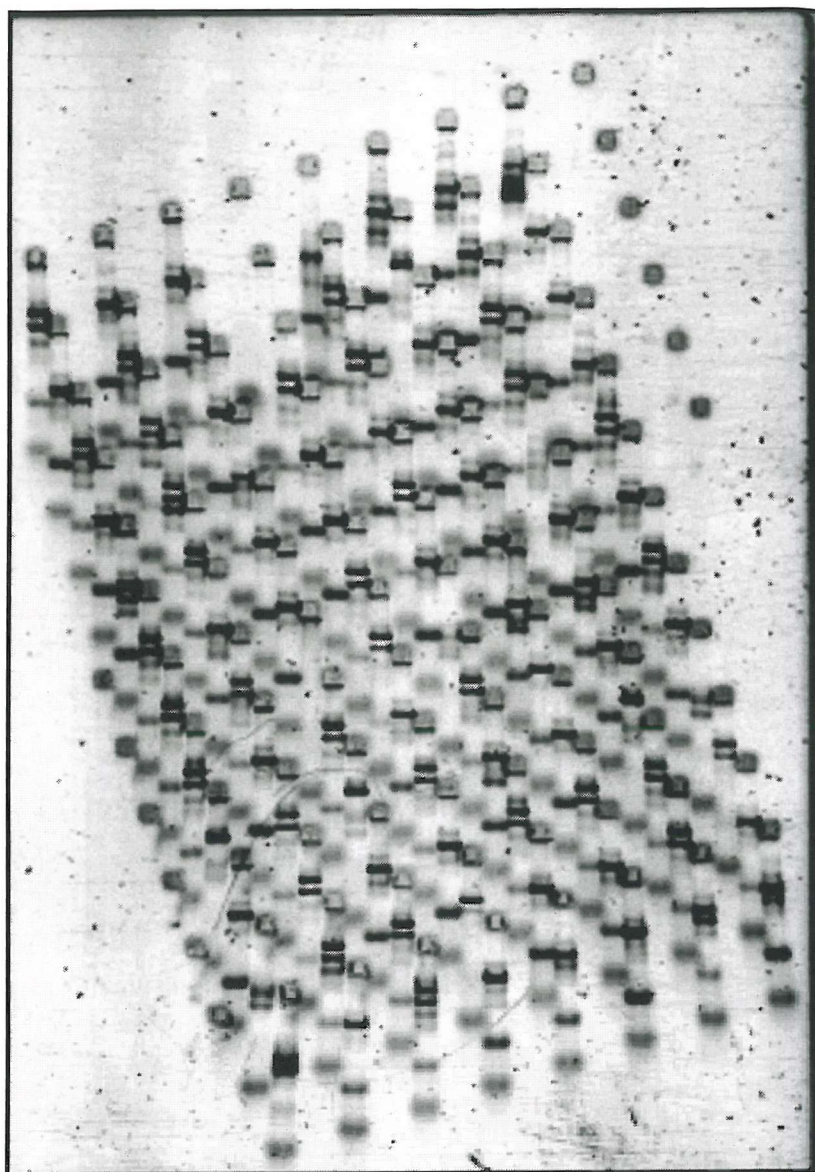
|     |     |     |
|-----|-----|-----|
| E3  | A3  | E2  |
| A/A | A/C | C/C |



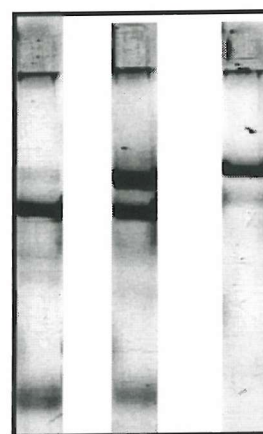
**Figure 4.3: Hpa II digest for the MMP-3 promoter polymorphism A-1346/C**

A 95 bp fragment was produced during amplification. Digestion with Hpa II cuts this fragment in one place for the C/C genotypes (two fragments of length 77 bp, and 18 bp) and leaves the A/A genotypes uncut. The lower band of 18 bp cannot be seen in this gel as it is masked by the primers used for amplification, therefore only the upper band can be used to score the genotyping.



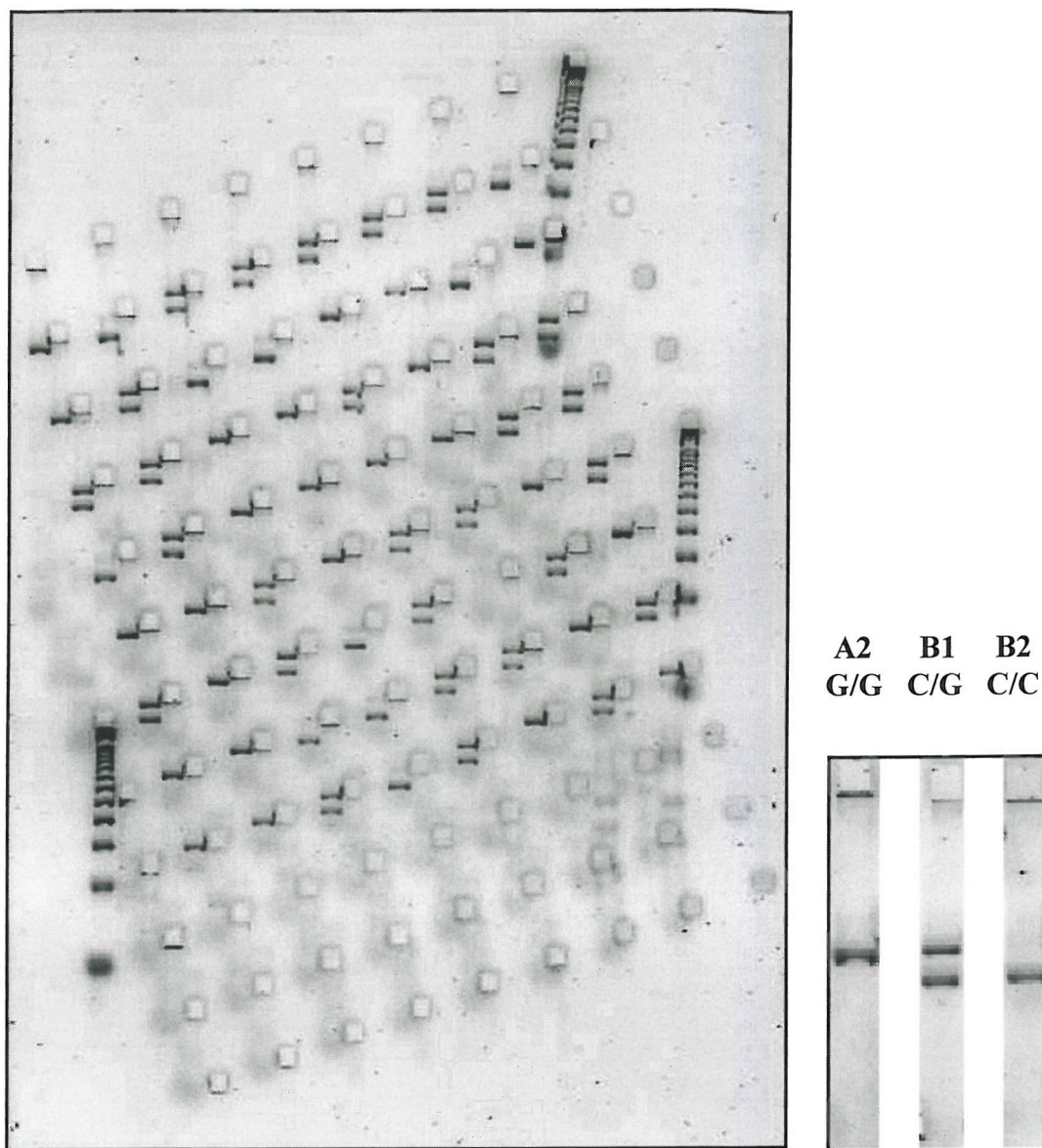


G12 G5 E11  
G/G A/G A/A



**Figure 4.4: EcoR1 digest for MMP-3 promoter polymorphism A-709/G**

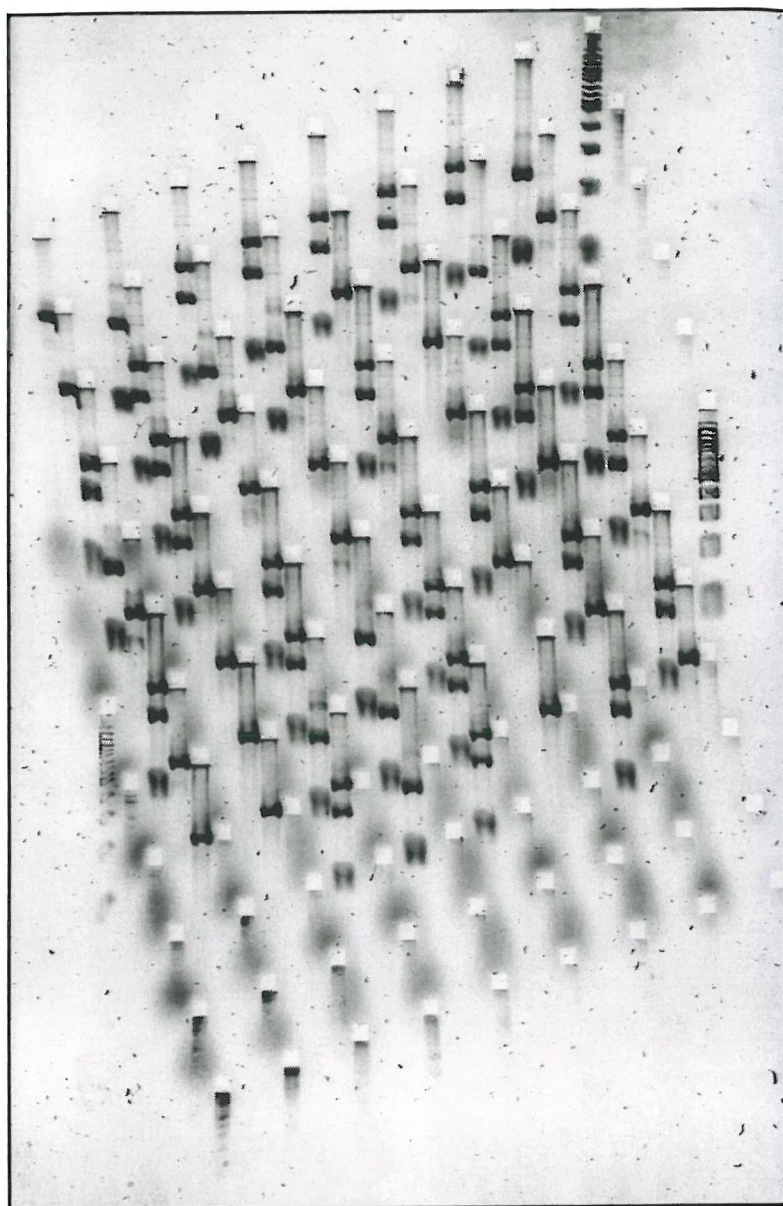
A 107 bp fragment was produced by PCR amplification. Restriction enzyme digest with EcoR 1 cut this fragment in two. Subjects with the G/G genotype, are represented by two bands of length 84 bp and 23 bp. Subjects with the A/A genotype remain uncut with a fragment length of 107 bp. The lower band of 23 bp cannot be seen in this gel as it is masked by the primers used in PCR amplification. Therefore, only the upper bands were used to score the genotypes. Samples containing two band fragments represent subject heterozygote for this polymorphism.



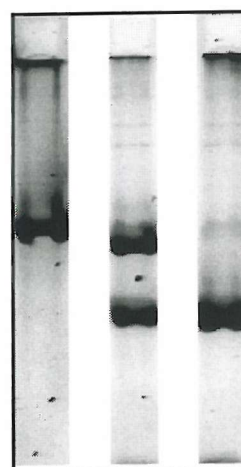
**Figure 4.5: Pfl fI digest for the MMP-3 promoter polymorphism C-376/G**

PCR amplification produced fragments of 125 bp in length. Restriction enzyme digestion with PflF1 cut the fragments containing the C/C genotype (two fragments of length 105 bp, and 20 bp) leaving the G/G genotype uncut. The lower band of 20 bp cannot be seen in this gel as it is masked by the primers used for amplification therefore only the upper band can be used to score the genotyping.



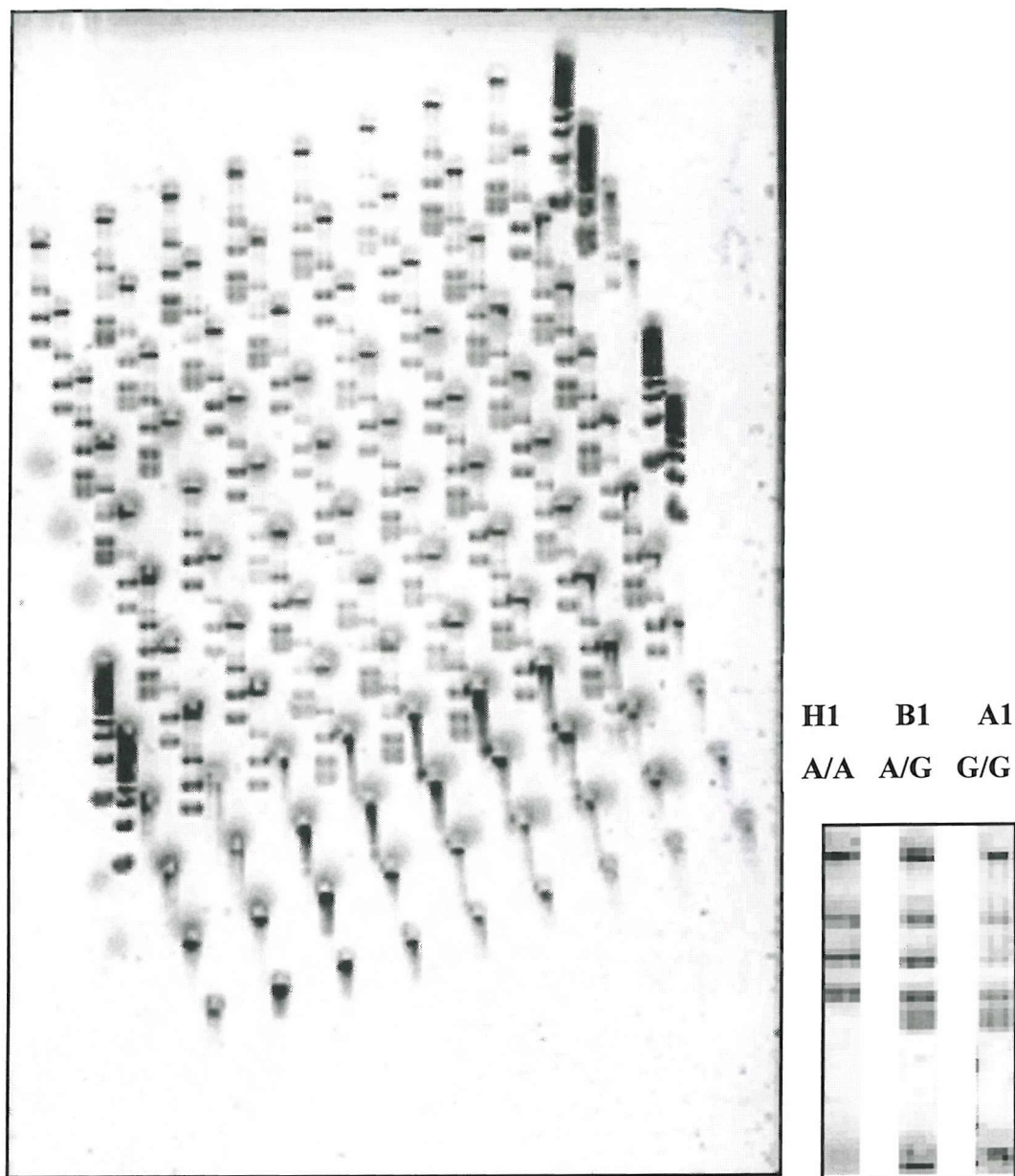


|     |     |     |
|-----|-----|-----|
| G11 | B1  | A1  |
| A/A | A/G | G/G |



**Figure 4.6: Taq<sup>1</sup> digest for the MMP-3 exon 2 polymorphism +802 bp**

A 413 bp fragment was produced during amplification of exon two. Digestion with Taq<sup>1</sup> cut this fragment in one place for the G/G genotype (two fragments of length 293 bp, and 120 bp) leaving the A/A genotype uncut.



**Figure 4.7: PshA1 digest for the MMP-3 exon 2 polymorphism +952 bp**

A 413 bp fragment was produced during amplification of exon two. Digestion with PshA1 cut this fragment in two places for the G/G genotype (three fragments of length 158 bp, 138 bp and 117 bp) and one place for the A/A genotype (two fragments of length 255 bp and 158 bp).

The seven bi-allelic SNPs were found to be in strong linkage disequilibrium. The frequency of a given haplotype was calculated using the equation below: -

$$\frac{\text{observed number of a given haplotype}}{\text{observed number of all haplotypes}}$$

In the smaller sample set, which was genotyped for the A-1986/G, 5A/6A, A-1346/C, C-376/G, exon 2 A+802/G and A+952/G polymorphisms, only two haplotypes were observed. These were C-6A-C-C-G-G and T-5A-A-G-A-A. Out of the 32 samples, 10 were found to be homozygous for the T-5A-A-G-A-A haplotype, 28 were heterozygous and 26 were homozygous for the C-6A-C-C-G-G haplotype. Therefore, the frequencies were found to be 0.375 and 0.625 respectively for the former and latter haplotypes.

This sample size was too small to make a definitive comment about allele frequency and therefore a larger cohort of subjects were subsequently genotyped. Six of the seven polymorphisms in the MMP-3 gene were studied in this large cohort. This was because the amount of DNA from the large cohort was limited, and because the A+952/G polymorphism in exon 2 does not cause an amino acid change, it would be unlikely to be biologically important and therefore was not studied in this cohort.

In the large cohort, 42 possible haplotypes were found. The estimated frequencies are given in Table 4.1. The most common haplotypes were T-5A-A-A-G-A and C-6A-C-A-C-G

**Table 4.1: - Haplotype frequencies**

| Haplotype           | All subjects<br>(n=1764 chromosomes) | Those with previous MI<br>(n=266 chromosomes) |
|---------------------|--------------------------------------|---|
| T-5A-A-G-G-A        | 7                                    | 0.0030329                                     |
| T-5A-A-G-G-G        | 3                                    | 0.0012998                                     |
| T-5A-A-G-C-G        | 3                                    | 0.0012998                                     |
| <b>T-5A-A-A-G-A</b> | <b>952</b>                           | <b>0.4124783</b>                              |
| T-5A-A-A-G-G        | 67                                   | 0.0290294                                     |
| T-5A-A-A-C-A        | 14                                   | 0.0060658                                     |
| T-5A-A-A-C-G        | 20                                   | 0.0086655                                     |
| T-5A-C-G-G-A        | 1                                    | 0.0004332                                     |
| T-5A-C-G-C-A        | 1                                    | 0.0004332                                     |
| T-5A-C-G-C-G        | 1                                    | 0.0004332                                     |
| T-5A-C-A-G-A        | 57                                   | 0.0246967                                     |
| T-5A-C-A-G-G        | 5                                    | 0.0021663                                     |
| T-5A-C-A-C-A        | 1                                    | 0.0004332                                     |
| T-5A-C-A-C-G        | 42                                   | 0.0181975                                     |
| T-6A-A-G-G-A        | 21                                   | 0.0091000                                     |
| T-6A-A-G-G-G        | 1                                    | 0.0004332                                     |
| T-6A-A-A-G-A        | 14                                   | 0.0066580                                     |
| T-6A-A-A-G-G        | 2                                    | 0.0008665                                     |
| T-6A-A-A-C-A        | 1                                    | 0.0004332                                     |
| T-6A-C-A-G-A        | 1                                    | 0.0004332                                     |
| C-5A-A-G-C-G        | 2                                    | 0.0008665                                     |
| C-5A-A-A-G-A        | 1                                    | 0.0004332                                     |
| C-5A-A-A-C-A        | 1                                    | 0.0004332                                     |
| C-5A-A-A-C-G        | 1                                    | 0.0004332                                     |
| C-5A-C-G-C-G        | 5                                    | 0.0021663                                     |
| C-5A-C-A-G-A        | 1                                    | 0.0004332                                     |
| C-5A-C-A-C-G        | 25                                   | 0.0108318                                     |
| C-6A-A-G-G-A        | 5                                    | 0.0021663                                     |
| C-6A-A-G-C-A        | 5                                    | 0.0021663                                     |
| C-6A-A-G-C-G        | 34                                   | 0.0147313                                     |
| C-6A-A-A-G-A        | 38                                   | 0.0164644                                     |
| C-6A-A-A-G-G        | 1                                    | 0.0004332                                     |
| C-6A-A-A-C-A        | 18                                   | 0.0077989                                     |
| C-6A-A-A-C-G        | 31                                   | 0.0134315                                     |
| C-6A-C-G-G-A        | 2                                    | 0.0008665                                     |
| C-6A-C-G-G-G        | 3                                    | 0.0012998                                     |
| C-6A-C-G-C-A        | 20                                   | 0.0086655                                     |
| C-6A-C-G-C-G        | 386                                  | 0.1672443                                     |
| C-6A-C-A-G-A        | 4                                    | 0.0017331                                     |
| C-6A-C-A-G-G        | 6                                    | 0.0025996                                     |
| C-6A-C-A-C-A        | 14                                   | 0.0060658                                     |
| <b>C-6A-C-A-C-G</b> | <b>491</b>                           | <b>0.2127383</b>                              |



LD between the polymorphisms was examined using the disequilibrium coefficients  $D'$  and  $\delta$ , the results are given in Tables 4.2.

**Table 4.2: Linkage disequilibrium between polymorphisms**

|             | -1986 T>C | -1612 5A>6A | -1346 A>C | -709 A>G | -376 G>C | +802 A>G |
|-------------|-----------|-------------|-----------|----------|----------|----------|
| -1986 T>C   |           | 0.9794      | 0.9939    | 0.9545   | 0.9910   | 0.9721   |
| -1612 5A>6A | 0.9851    |             | 0.9782    | 0.9339   | 0.9533   | 0.9497   |
| -1346 A>C   | 0.9147    | 0.9052      |           | 0.9255   | 0.9871   | 0.9485   |
| -709 A>G    | 0.6500    | 0.6316      | 0.6062    |          | 0.9585   | 0.9964   |
| -376 G>C    | 0.9957    | 0.9825      | 0.9385    | 0.9799   |          | 0.9706   |
| +802 A>G    | 0.9864    | 0.9752      | 0.9320    | 0.9657   | 0.9858   |          |

Figures shown are disequilibrium coefficient  $D'$  (above diagonal) and  $\delta$  (below diagonal), respectively

Visual inspection of this table reveals a number of interesting features. The first is that there does not seem to be a strong relationship between LD and physical distance. Alleles in close proximity to one another are not as strongly linked as those with a greater distance between them. For example, T-1986/C is more strongly linked with G-376/C than -1612 5A/6A. These findings are consistent with previous observations that disequilibrium and physical distance do not correlate significantly when distance is less than 60 Kb<sup>176</sup>.

The 5A/6A polymorphism was previously shown to be associated with progression of atherosclerosis. Table 4.3 shows which alleles in the other polymorphisms were linked with the 5A allele and which alleles were linked with the 6A allele.

**Table 4.3: Linkage disequilibrium between polymorphisms**

|              |          | -1612 5A>6A |             |             |       |
|--------------|----------|-------------|-------------|-------------|-------|
| Polymorphism | Genotype | 5A/5A       | 5A/6A       | 6A/6A       | Total |
| -1986 T>C    | T/T      | 208(95.0%)  | 11 (5.0%)   | -           | 219   |
|              | T/C      | 4 (0.9%)    | 429 (97.9%) | 5 (1.1%)    | 438   |
|              | C/C      | 2 (1.1%)    | 1 (0.6%)    | 176 (98.3%) | 179   |
| -1346 A>C    | A/A      | 186 (86.1%) | 25 (11.6%)  | 5 (2.3%)    | 216   |
|              | A/C      | 6 (1.5%)    | 358 (89.9%) | 34 (8.5%)   | 398   |
|              | C/C      | -           | 12 (9.4%)   | 115 (90.6%) | 127   |
| -709 A>G     | A/A      | 197 (99.5%) | 1 (0.5%)    | -           | 198   |
|              | A/G      | 10 (2.1%)   | 458 (97.7%) | 1 (0.2%)    | 469   |
|              | G/G      | 1 (0.5%)    | 2 (0.9%)    | 213 (98.6%) | 216   |
| -376 G>C     | G/G      | 179 (95.2%) | 9 (4.8%)    | -           | 188   |
|              | G/C      | 2 (0.6%)    | 345 (98.3%) | 4 (1.1%)    | 351   |
|              | C/C      | 2 (1.3%)    | 9 (6.0%)    | 140 (92.7%) | 151   |
| +802 A>G     | A/A      | 143 (94.1%) | 8 (5.3%)    | 1 (0.7%)    | 152   |
|              | A/G      | 4 (1.4%)    | 276 (96.8%) | 5 (1.8%)    | 285   |
|              | G/G      | 6 (4.5%)    | 12 (9.1%)   | 114 (86.4%) | 132   |

#### 4.4 Discussion

In this chapter, the strength of LD between each polymorphic site was investigated. Analysis of allelic association and thus LD is a method used to find a significant association between alleles at two or more loci. If significant association is found between such alleles, they are said to be allelically associated and LD generally causes this. This theory stems from the idea that the moment a new disease causing mutation occurs in a gene, the disease allele is associated with the alleles of other polymorphic loci in the region. More over, the disease and marker alleles are in complete linkage disequilibrium and it is therefore assumed that even after several generations the two would be strongly associated. What should also be considered is that common polymorphisms of complex traits are very old and patterns of association are highly complex as they are influenced by recombination, mutation and evolutionary factors.

LD in the sense of allelic association, underlies all types of mapping. This is because LD mapping uses unrelated individuals as opposed to the well characterised pedigrees employed by linkage analysis. Genes/chromosomes of unrelated individuals will be more distantly related to one another and to their founder than the individuals used in a traditional pedigree. There are more opportunities for recombination to occur. LD mapping is therefore considered to be best suited for fine-scale mapping of human disease loci <sup>154,174,177</sup> as well as historical and evolutionary influences <sup>167</sup>.

The initial investigation, which used a small sample set, showed complete linkage disequilibrium between all 6 SNPs. The frequencies of the two haplotypes represented were: T-5A-A-G-A-A, 37.5%; C-6A-C-C-G-G, 62.5%. Such an association suggests that these tightly linked loci may share the same ancestral chromosome. The strong LD observed is not surprising since the SNPs are located within a region of only 2.9Kb. Small sample sizes are known to provide LD estimations that are biased upwards. This is because values of LD must always be positive. The rate of decay of LD with increasing distance can therefore be underestimated in such small sample sizes.

The possibility that some uncommon haplotypes were not represented in these samples could not be precluded and so to ensure accuracy in our LD estimation a larger cohort was used. The two haplotypes defined by these SNPs however, were thought to

represent the two most common haplotypes in the population. This was observed in the larger sample set investigated. These two haplotypes were expressed at frequencies of 41% for the T-5A-A-A-G-A haplotype, and 21% for the C-6A-C-A-C-G haplotype. The C-6A-C-G-C-G haplotype was also highly represented in the large sample set at a frequency of 16.7%. The other 39 possible haplotypes had much smaller frequencies most of which were less than 1%. These higher frequencies could be the result of natural selection, but this cannot be concluded on this data alone. What is clear from these high frequencies is that these markers are likely to be very old.

The A-709/G polymorphism had a lower degree of LD compared to the other variants. This is because this variant had a lower allele frequency. One allele was represented at a frequency of ~75% while the other occurred at a frequency of ~25%. The ratio between the alleles of all the other polymorphisms were ~50/50. Different allele frequencies affect the disequilibrium coefficient, and thus a lower LD figure was observed.

The extent of LD in the human population is of considerable controversy and a number of theories have been brought forward ranging from 3Kb in the general population, up to 10Kb <sup>168</sup>. The general consensus however is: 1) That there is a strong relationship between LD and distance, 2) LD around a variant that has been in existence for over 5000 generations is unlikely to occur beyond markers a few kb apart, 3) LD between younger alleles will extend much further. The observations in this study have shown strong LD between all alleles. Different strengths were observed between alleles. Alleles next to each other were found to exhibit slightly less LD when compared to alleles not as physically close. This was higher than expected as the cluster of SNPs investigated here lie within a distance of 2.9 Kb with respect to T-1986/C to A+952/G. There are other studies that have found similar results <sup>164,178</sup>. There are many explanations for this such as recombination, mutational hotspots and demography. It is unlikely however, that recombination is one of these. This is because the distance between the alleles is so small. While most of the polymorphic sites are in the form of transition mutations, the decrease in LD due to CpG islands is also unlikely. MMP-3 is not C/G rich and the number of C/Gs and A/Ts is approximately 50:50. Moreover the MMP-3 gene is not a gene that is highly transcribed and chromosome 11 is not one of the chromosomes in the human genome that has a high density of CpG islands <sup>179</sup>. So,



what has caused this decrease in LD over short distances? There are a number of possible explanations including, genetic drift, natural selection and demographic factors such as admixture. One interesting theory is the idea that gene conversion occurs more frequently than originally thought. Our knowledge of gene conversion event in the human genome is very little. It is believed that gene conversion and crossing over are alternative outcomes of a common recombination mechanism as its effect is very similar to that of crossing over. Despite this, gene conversion is not generally considered, and for markers megabases apart, its contribution to the overall level of genetic change is considered to be negligible. Thus, genetic map based estimates of recombination rates are estimated as crossing over events alone. While this may give accurate predictions of LD between markers megabases apart, it does not take into account the possibility that gene conversion can cause a decrease in LD at closely linked markers. Przeworski and Wall recently showed that recombination events are two-thirds more likely to be caused by gene conversion than crossing over<sup>180</sup>.

It would be interesting to see the allele frequencies of these SNPs in other populations. Different populations may show a variation in allele frequencies. It may also give information on the history of the polymorphisms. Only a few population studies of LD have been carried out. So far, such studies have shown that LD varies between different populations, although no major difference in the extent of LD among European populations was found for common disease susceptibility alleles. This is not surprising, given that common polymorphisms are thought to arise from ancestral chromosomes<sup>164-166</sup>.

## **Chapter 5**

### **Association Studies of MMP Polymorphisms in Relation to CAD**

#### **5.1 Introduction**

Many important diseases are not single gene disorders and therefore families cannot be traced and classified in a simple, autosomal dominant, autosomal recessive or X-linked manner. Common illnesses that affect western society such as heart disease, stroke, hypertension, rheumatoid arthritis and diabetes are more complicated. While twin studies have shown that these common diseases have a strong genetic component, as they tend to run in families, they are determined by the interaction of several genes at different loci each with an additive effect. Furthermore epidemiological studies have provided strong evidence to suggest that environmental factors such as smoking, alcohol and diet play a major role in their pathogenesis. These common diseases thus have multifactorial aetiologies that are caused by a combined effect of multiple genetic and environmental factors. It is theorized that these risk factors have a small effect individually, but when a sufficient number occur together, a threshold is crossed and the disease occurs. This notion is known as the liability-threshold model.

Identifying susceptibility genes for multifactorial diseases is significantly more complex than for single-gene disorders. The main reason for this is the lack of a simple relationship between genotype and phenotype. Not all people with a given susceptibility gene will get the disease, while people who do not carry the given susceptibility gene can also get the disease. Examining the complex molecular processes underlying the pathogenesis of the given disease, increases our understanding and directs us to possible candidate genes. Linkage studies are utilized to identify loci for the complex disorder and to search for evidence of a major gene effect.

The key events that take place in the development of atherosclerotic lesions include endothelium injury, monocyte recruitment, macrophage development, foam cell formation, lipid deposition, platelet activation, smooth muscle cell migration and proliferation, and the synthesis and degradation of intracellular matrix and fibrous tissue.

The proposal that MMP-3 takes part in connective tissue remodelling during plaque rupture and development of atheromas has come from a number of studies: Firstly, Henney *et al* (1991) by way of *in situ* mRNA hybridisation and immunocytochemistry, demonstrated that MMP-3 was present in coronary atheromas. More specifically MMP-3 was found to be more prevalent in regions prone to rupture such as the shoulders of the atherosclerotic plaque<sup>83</sup>. Another group devised a method which directly detects localised MMP enzymatic activity in frozen tissue sections. This technique, *in situ* zymography, revealed gelatinolytic and caseinolytic activity in sections of atherosclerotic tissue, which were not present in normal tissue, sections<sup>84</sup>. These findings were supported by several subsequent studies<sup>95,106,107</sup>.

The 5A/6A polymorphism was later identified in the promoter sequence of the human MMP-3 gene. As mentioned earlier, this variant has one allele with a 5 adenosine (5A) run and the other has a 6 adenosine (6A) run<sup>105</sup>. Functional studies on the promoter strength showed a 2-fold increase in promoter activity in the 5A allele than in the 6A allele in both cultured fibroblasts and smooth muscle cells. Band shift assay attributed this decrease in promoter activity of the 6A allele to preferential binding of a putative transcript repressor protein. The 5A allele through its increase promoter activity is thus postulated to contribute to weakening of the cap leading to plaque rupture and AMI, while carriers of the 6A allele are more likely to accumulate ECM due to a decrease in degradation resulting from a decline in MMP-3 activity. This would thus lead to more rapid development and progression of atherosclerotic plaques<sup>84,99</sup>.

Several studies have since provided evidence of a correlation between the 6A allele and the progression of atherosclerosis. The first was a cohort study on 72 patients with CHD defined by angiography. The patients were part of the St. Thomas Atherosclerosis Regression Study (STARS). The data indicated that patient homozygous for the 6A allele had more rapid progression of global and focal atherosclerotic stenosis over a 3-year study period. This was more apparent in patients with a baseline stenosis <20% and in patients with higher LDL cholesterol concentrations<sup>105</sup>. The Regression Growth Evaluation Statin Study (REGRESS) investigated whether this polymorphism influenced the risk of restenosis and the progression of CAD. They also examined any changes in effect in patients treated with the lipid lowering drug, pravastin. The results showed heterozygote and 6A

homozygote patients had more clinical events than 5A homozygote patients. These events were reduced in heterozygote and 6A homozygote patients treated with pravastatin, when compared to patients not treated with the drug. No changes were observed in 5A homozygote patients. The incidence of restenosis after angioplasty was found to be greater in both the heterozygote and 6A homozygote patients when compared to the 5A genotypes. Restenosis was substantially reduced in both 5A/6A and 6A genotyped patient treated with pravastatin and found to be unchanged in 5A homozygote patients <sup>181</sup>. These results were substantiated by similar data from the Lipid Coronary Angiography Trail (LOCAT) <sup>115</sup>. Like the REGRESS study, patients who had had previous coronary bypass surgery and had low HDL cholesterol were observed over a 32-month period. One group of patients were treated with a slow-release gemfibrozil (Lipid SR) and compared to a second group that had no treatment. Little or no restenosis was observed in patients who were homozygous for the 5A allele. Patients with one or more 6A alleles was found to have significant increase in luminal narrowing, averaging 1.79%. Their data also indicated that progression of atherosclerosis was reduced in the gemfibrozil treated groups. However the effects were found to be similar between the three genotypes <sup>115</sup>. These results suggest that the 6A allele is a genetic risk factor for rapid progression of atherosclerosis.

More recently, an intracoronary ultrasound study was performed on patients who underwent conventional balloon coronary angioplasty; one group of patients had a Palmaz-Schatz stent implantation, while the other was without a stent. Genotypes for the patients were determined for the 5A/6A polymorphism. After 6 months quantitative computer assisted angiography was performed on each patient to estimate restenosis. The results found an increased degree of restenosis in patients without stent who were homozygote for the 6A allele, compared to those heterozygous and homozygous for the 5A allele. In the stented patients, restenosis was not found to occur in any of the MMP-3 genotypes <sup>182</sup>.

Several studies have used common carotid geometry to investigate the relationship between the 5A/6A genetic variant of MMP-3 and the pathogenesis of atherosclerosis. Diameters of the common carotid artery and/or the intima-media thickness (IMT) were measured in patients by non-invasive ultrasonography. The first of these studies investigated the diameter, wall shear stress and IMT in forty-two healthy male subjects.

Their investigations showed subjects homozygous for the 6A genotype had a significantly larger arterial lumen and IMT, while wall shear stress was found to be significantly lower when compared to the other genotypes (5A/6A and 5A/5A)<sup>183</sup>. In another study, high-resolution ultrasonography was used to quantify the extent of atherosclerosis by measuring IMT in a group of patients with CAD defined by angiography. The data in this study again showed a significant increase in IMT for patients with the 6A/6A genotype compared to the other two<sup>184</sup>. In a further study, B-mode carotid ultrasound was employed to measure IMT of the carotid artery in order to identify association between atherosclerosis and functional promoter variants of three genes. A cohort of 87-mixed race subjects was used. With respect to MMP-3, subjects with a 6A/6A genotype were found to have an 8% greater mean IMT than the combined result of the other two MMP-3 genotypes<sup>185</sup>. These findings suggest that the 6A allele, associated with a decrease in promoter activity, predisposes to arterial wall thickening and arterial wall enlargement and is therefore likely to be involved in progression of atherosclerotic plaques.

Further to these investigation Terashima *et al* performed a case-control study to examine the 5A/6A polymorphism in relation to AMI. The 5A allele was found to be significantly more frequent in AMI patients than in control subjects<sup>116</sup>. Their data indicated that the 5A allele predisposes to AMI. This is likely to result from over degradation of matrix proteins in atherosclerotic plaques in 5A allele carriers since the 5A allele, has a higher promoter activity.

There is strong evidence linking the 5A/6A polymorphism to both the progression of atherosclerotic plaques and AMI. A further six common polymorphisms have been identified in the MMP-3 gene during the course of this investigation. Variations in promoter activity between the different alleles have also been observed. This study looks further into the relationship between MMP-3 polymorphic alleles and possible haplotypes and previous MI and the progression of atherosclerosis.

There are two main approaches used to localise genetic markers that influence susceptibility to a given disease. These are association studies and linkage analysis. Brief descriptions of the methodology of each technique are given below:

### 5.1.1 Association Studies

Association studies or mapping are used to analyse the relationship between alleles and a disease trait. It relies on the theory that if a mutation increases disease susceptibility, then one would expect it to be more frequent in affected individuals (cases) than among unaffected individuals (controls). The idea behind association mapping, is that markers close to the disease causing allele, may also have allele frequency differences between case and control individuals if there is strong LD between them. In a randomly mating population, LD decays quite rapidly with distance, therefore if association between a marker loci is found it implies that the marker is tightly linked to a disease causing mutation. Studies have demonstrated that association mapping is potentially a very powerful strategy for identifying the loci that contribute to complex diseases<sup>154,186,187</sup>. Association studies can detect genes with a small effect better than linkage studies, as long as the marker and disease causing allele are in close proximity. The standard methods for such a study are either a cohort or a case-control association.

#### 5.1.1.1 Cohort Studies

In a cohort study, patients who presently have a certain condition and/or are receiving a given treatment are followed over time and compared with another (control) group who are not affected by the condition under investigation. The health status of both the patients and control subjects are observed and recorded in the same way during the course of the study.

There are two main calculations performed in a cohort study, the incidence rate and relative risk. The incidence rate is the measure of the disease, which is given as the subjects who develop the disease over a specific time period. Association between the given disease and the exposure of variant under study is in the form of relative risk. This is the ratio of the incidence rate of patients to that of the control subjects.

There are several benefits in performing such a study: (1) Complete information on each subject's exposure can be made, (2) It provides a clear temporal sequence of exposure and disease, (3) The data from cohort studies are usually easy to interpret, (4) This data can be used to calculate both incidence rate (absolute risk) and relative risk.

Cohort studies are not suitable for the study of rare diseases, as a large number of subjects are generally required. It can be quite expensive to perform and take a long time. Another disadvantage is that things can change over the course of the study i.e. subjects may die, move away or develop other conditions.

#### **5.1.1.2 Case-Control Studies**

This technique compares the frequency of a particular allele in a set of patients with the disease or condition under study, and a control group of individuals without the disease or trait under study. The simplest protocol involves genotyping a set of markers in a sample of cases and unrelated controls. Though many have shown that association studies are more powerful than linkage studies for identifying genetic determinants for complex diseases, it is fraught by a host of potential biases. In order to maximise the power of detection and to avoid biases, certain criteria needs to be considered. These include: how the case subjects are defined and whether they meet certain criteria for the given phenotype is critical. Case subjects need to be as genetically homologous as possible, as varied associations present in different ethnic groups could mask each other. Control subject must be free of the disease under study and must match the case subjects with respect to demographic and environmental factors. Sample size is also important, as the power of study is positively related and the rate of type I error is negatively related to the sample size.

The validity of association mapping depends upon the appropriate selection of patients and controls. Case-control association studies require certain properties to hold. Most importantly, they require random mating in the population. If population stratification is present or has occurred in the recent past, then this could give rise to association between alleles that is not due to LD. In the past, some population based case control studies have given false results due to population stratification. Different ethnic groups may carry alleles not present in other ethnic groups. One of the best-known examples in which population structure has affected association studies, is in the investigations made on NIDDM in the Pima and Papago Indian tribes of Arizona<sup>188</sup>. The findings in this study indicated that there was a strong negative association between diabetes and the haplotype at immunoglobulin G loci. It was found that the subjects used had recent

European ancestry and the average proportion of European ancestry was higher in the control participants than in the affected subjects. The European individuals had a higher frequency of the haplotypes in question, regardless of the phenotype. Stratification of the data according to reported ancestry showed that the protective effect of the given haplotype disappeared. A mixture of subjects of different ethnicities can therefore create an artificial stratified population when used in a case-control study. In order to avoid false association due to population stratification, the control group needs to match the case subjects for comparative factors such as ethnicity and geographical origin and also with respect to certain characteristics (age, body mass index (BMI), sex). Hence, some believe that this method is unreliable because of the difficulties encountered in selecting a good control group. Other methods (family-based tests) of analysis have been developed which use internal controls to avoid false results derived from population stratification. These are the haplotype relative risk (HRR) test, and the transmission disequilibrium test (TDT), (see 5.1.2.1-2 below for details). While these methods avoid population structure problems, they are more time-consuming and it is expensive to assemble family-based samples. For late onset diseases such as atherosclerosis, these methods may also be impractical.

The frequency of alleles in the case and control groups can be compared using the  $\chi^2$  test. The odds ratio is also used to measure association between the candidate gene and the disease. It is also used to approximate the relative risk of a disease for a genotype. If there is no effect of genotype on relative risk, then the odds ratio will be 1.

The main advantages of case-control studies are: (1) They are well suited to the study of rare diseases or disease with long latency between exposure and manifestation, (2) They can be performed quickly and are relatively inexpensive to perform, (3) They can be used to study multiple potential causes of disease, (4) There is no risk to patients. The difficulties include (1) Choice of control subjects may be difficult, (2) Incidence rate of disease cannot be determined, (3) Records on exposure and past history is based on interviews and may be subject to bias.



### 5.1.2 Linkage Analysis

Linkage analysis or mapping is a method used to locate disease genes to a particular DNA fragment in related individuals (generally siblings or extended pedigrees). It involves the use of multiple SNPs or microsatellite markers to search for evidence of an ancestral haplotype “a series of allele found at linked loci on a single (paternal or maternal) chromosome”<sup>146</sup> over represented in affected individuals. The theory behind this method is, that an allele found more frequently in case than in control sample groups, will be in LD with and in close proximity to, the functional allele causing the disease investigated. Thus, linkage is presented when the alleles of a marker co-segregate (are inherited together) with a disease within the family more often than by chance. This method has been used successfully in localising many Mendelian disease genes. Now it is used to localise genes underlying susceptibility to complex diseases and has been used successfully to identify single-gene sub-forms of some multifactorial diseases such as breast cancer and Alzheimer’s disease.

Linkage analysis in humans requires complex methods for estimating the outcome of linkage because families are small and pedigree structures rarely ideal. LOD scores are the statistical method employed to measure the likelihood of genetic linkage between loci in linkage analysis studies. They are the logarithm (base 10) of an odds ratio and is calculated using computer programs from the pedigree data. The end result is a table or graph of LODs at various recombination fractions. Statistical significance of linkage requires a LOD score of +3.0 or greater. It is a robust technique for localizing disease genes.

#### 5.1.2.1 Haplotype Relative Risk Test

The haplotype relative risk (HRR) was first described by Falk and Rubinstein (1987)<sup>189</sup>. They believed that the best place to get an ethnically matched control for a case-control study is in the family of the case. The HRR test looks at affected individuals and their parents. All three are typed for a marker and where one allele is thought to be associated with the disease, the control is based on the two alleles that the offspring has not inherited from their parents. The results of such a test are analysed as if they were a conventional case-control study. While this test ensures that the case and control

individuals are from the same population, it does not completely eliminate the possibility of stratification, but just reduces the chances. In the presence of stratification the HRR test may produce false positive results. Parents are required for this study, which in late onset diseases may cause a problem.

#### **5.1.2.2 Transmission Disequilibrium Test**

The transmission disequilibrium test is a modification of the HRR method and was first described by Spielman *et al* in 1993<sup>190</sup>. It analyses families where one or more of the offspring are affected and at least one of the parents is heterozygous for the trait in question. One or two marker alleles of each heterozygous parent are transmitted to each affected offspring. The frequency of the putative diseased marker is then compared with the transmitted and non-transmitted parent alleles. The Chi<sup>2</sup> Test is used to find any significant association between the frequencies. The TDT tests for linkage in the presence of LD. It eliminates stratification effects completely. Disadvantages of this technique are that it only uses heterozygous parents and therefore, in the absence of stratification, is less powerful than HRR which uses both homozygous and heterozygous parents. Parents are required for this study, which as mentioned before, in late onset diseases may cause a problem<sup>146</sup>.

Earlier, the strength of LD, that is, the allelic association between the SNP identified in the MMP-3 gene was measured. Here again a series of LD studies was performed. In this instance however, the linkage also known as association of each SNP to the susceptibility of CHD was looked at, rather than the strength of LD between the SNP.

SNPs have been postulated to be useful tools for identifying complex disease genes through association. This is because SNPs have been estimated to occur on average every 250-1000 bp and have a low mutation rate making them ideal markers for genetic analysis. The aim of this study was therefore to see if there was any association between variants identified in the MMP-3 gene and the susceptibility of CHD.

## 5.2 Results

### 5.2.1 Study Subjects

The subjects used for this study were recruited from individuals referred for diagnostic and interventional coronary arteriography to the Wessex Cardiothoracic Unit between August 1997 to March 1998. These DNA samples were originally collected and have been used for other studies on CHD and approved by the local regional ethics committee, and all subjects gave consent<sup>191</sup>.

The total number of subjects included in this study were 1205 men and women, aged >50 years. 913 were of patients with >50% stenosis in at least one coronary artery according to angiography, and 292 were of subjects who showed symptoms of heart disease (e.g. chest pains) but angiographic results did not show severe stenosis (i.e. <50% stenosis). Patient histories utilized in the study were taken directly from their clinical notes.

All subjects were Caucasians from the Wessex region. Lifestyles and medical histories were taken for each participant during recruitment. The characteristics of Caucasians with CHD and of the control subjects are listed in Table 5.1 below. There was no significant difference in ages between groups. The coronary risk factors examined i.e. smoking, hypercholesterolaemia, hypertension and diabetes, were significantly pronounced in the patient group.

**Table 5.1**  
**Characteristics of the subjects**

|                       | Cases           | Controls        | p-value |
|-----------------------|-----------------|-----------------|---------|
| Age (mean $\pm$ SD)   | 63.6 $\pm$ 10.2 | 60.2 $\pm$ 12.0 | <0.001  |
| Gender                |                 |                 |         |
| Male                  | 692 (73.4%)     | 138 (46.5%)     | <0.001  |
| Female                | 251 (26.6%)     | 159 (53.5%)     |         |
| Smoking               |                 |                 |         |
| Current smokers       | 156 (16.5%)     | 46 (15.5%)      | <0.001  |
| Ex-smokers            | 517 (54.8%)     | 128 (43.1%)     |         |
| Non-smokers           | 270 (28.6%)     | 123 (41.4%)     |         |
| Hypercholesterolaemia |                 |                 |         |
| Yes                   | 741 (78.6%)     | 189 (63.6%)     | <0.001  |
| No                    | 202 (21.4%)     | 108 (36.4%)     |         |
| Hypertension          |                 |                 |         |
| Yes                   | 344 (36.5%)     | 76 (25.6%)      | =0.001  |
| No                    | 599 (63.5%)     | 221 (74.4%)     |         |
| Diabetes              |                 |                 |         |
| Yes                   | 113 (12.0%)     | 19 (6.4%)       | =0.006  |
| No                    | 830 (88.0%)     | 278 (93.6%)     |         |
| Family history of CAD |                 |                 |         |
| Yes                   | 523 (55.5%)     | 117 (39.4%)     | <0.001  |
| No                    | 420 (44.5%)     | 180 (60.6%)     |         |

In this study, two groups of samples were used, one group containing individuals with >50% stenosis in at least one coronary artery, the second containing individuals without >50% stenosis.

The subjects chosen for this study were recruited from one clinic and there is therefore a potential selection bias. One way to avoid such potential bias, is to study population-based, prospective collections. However, an advantage of this study is the clear diagnosis based on angiography which would be impractical for population-based sampling. Several risk factors for coronary heart disease have been listed in Table 5.1 above. This could potentially affect the results of this study, to ensure that the MMP-3 genotypes are investigated as an independent risk factor the data analysed will be adjusted for risk factors.

### 5.2.2 Data Analysis

Analysis of data was carried out with SPSS for Windows (SPSS Inc. Chicago, IL, USA). The HWE program was used to examine deviations of genotype distribution from the Hardy-Weinberg equilibrium and the  $\chi^2$  test was used to determine whether there was any significant difference in allele or genotype frequencies between patient and control subjects. A p value of <0.05 was considered significant. The logistic

regression analysis was applied to study possible effects of individual polymorphisms on the risk of MI. Ordinal regression analysis was used to study possible effects of individual polymorphisms on the extent of coronary atherosclerosis measured by the number of coronary arteries with >50% stenosis. Stepwise logistic regression was used to study possible effects of haplotypes on the risk of MI or extent of coronary atherosclerosis. All analysis were carried out firstly without and then with adjustment for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes.

All 1205 subjects were genotyped for six of the MMP-3 polymorphisms (T-1986/C, 5A-1612/6A, A-1346/C, A-709/G, G-376/C and A+809/G) by restriction enzyme digestion. Haplotype and allele frequencies of the different polymorphisms of the MMP-3 gene are summarized in Chapter 4 (Table 4.1 and 4.3). The distribution of genotypes is consistent with the distribution predicted by the Hardy Weinburg equilibrium.

Data from previous studies have suggested that individuals with the MMP-3 gene 6A/6A genotype had a more rapid progression of coronary atherosclerosis: It was therefore hypothesised that variations in this gene could influence the extent of coronary atherosclerosis in CAD patients. To test this hypothesis, the 5A/6A polymorphism and the newly identified polymorphisms were analysed in a cohort of Caucasian subjects with angiographically documented CAD. The data showed that patients without a history of MI had a higher frequency of the 6A/6A genotype and a decreased number of 5A/5A genotypes, with increasing number of coronary arteries with >50% stenosis (odds ratio = 1.52,  $p=0.008$  in ordinal logistic regression analyses, Table 5.2). This association remain significant after adjustment for covariates including age, gender, smoking, hypercholesterolaemia, hypertension and diabetes. No statistically significant association was detected between the other polymorphisms in the MMP-3 gene and the extent of coronary atherosclerosis.

Analyses were then carried out to test the hypothesis that variations in the MMP-3 gene could influence the risk of MI and CAD patients. The results for the analysis of individual polymorphisms in relation to previous MI are represented in Table 5.3. The data indicated that the 5A allele of the 5A/6A polymorphism was associated with risk of MI as its frequency was observed to be higher in CAD patients with an MI history,

compared to CAD patients without such a history. Individuals with the 5A/5A genotype had an over 2 fold higher risk of MI compared with individuals with the 6A/6A genotype (odds ratio 2.02, 95% CI 1.14, 3.06,  $p=0.016$ ). Those with the 5A/6A genotype had a 1.78 fold higher risk of MI compared with individuals with the 6A/6A genotype (odds ratio 1.78 95% CI 1.05, 3.00,  $p=0.032$ ) (Table 5.3). This association remained after adjustment for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes. No statistically significant association was detected between MI risk and the other polymorphisms studied.

Logistic regression analysis of haplotypes showed statistical significance for two. These were T-5A-A-A-G-A 1.36 (95% CI, 1.02, 1.80),  $p=0.033$  and T-5A-A-A-C-G 4.04 (95% CI, 1.29, 12.67),  $p=0.017$  without adjustment for covariants. The association changed only slightly after adjustments for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes (Table 5.4). After adjustment for these covariants, the C-5A-C-G-C-G haplotype was also associated with risk of MI (odds ratio 8.31, 95% CI, 1.08, 63.76,  $p=0.04$ ).

**Table 5.2: Relationship between polymorphisms and number of >50% stenotic coronary arteries**

| Polymorphism | Genotype | Number of stenotic vessels |             |             |             | Odds ratio (95% CI)        | * Adjusted Odds ratio (95% CI) |
|--------------|----------|----------------------------|-------------|-------------|-------------|----------------------------|--------------------------------|
|              |          | 0                          | 1           | 2           | 3           |                            |                                |
| -1986 T>C    | C/C      | 54 (20.5%)                 | 54 (22.2%)  | 40 (20.3%)  | 66 (24.2%)  | 1.24 (0.89, 1.71), p=0.198 | 1.17 (0.78, 1.74), p=0.448     |
|              | C/T      | 133 (50.6%)                | 122 (50.2%) | 110 (55.8%) | 139 (50.9%) | 1.15 (0.88, 1.51), p=0.305 | 1.3 (0.74, 1.43), p=0.874      |
|              | T/T      | 76 (28.9%)                 | 67 (27.6%)  | 47 (23.9%)  | 68 (24.9%)  | 1                          | 1                              |
| -1612 5A>6A  | 6A/6A    | 56 (19.3%)                 | 58 (22.6%)  | 47 (22.3%)  | 89 (28.5%)  | 1.52 (1.11, 2.07), p=0.008 | 1.53 (1.12, 2.10), p=0.009     |
|              | 5A/6A    | 151 (52.1%)                | 133 (51.8%) | 115 (54.5%) | 151 (48.4%) | 1.13 (0.87, 1.46), p=0.374 | 1.09 (0.83, 1.43), p=0.517     |
|              | 5A/5A    | 83 (28.6%)                 | 66 (25.7%)  | 49 (23.2%)  | 72 (23.1%)  | 1                          | 1                              |
| -1346 A>C    | C/C      | 49 (21.1%)                 | 50 (22.9%)  | 40 (23.4%)  | 45 (18.7%)  | 1.06 (0.75, 1.48), p=0.757 | 0.77 (0.51, 1.18), p=0.233     |
|              | A/C      | 103 (44.4%)                | 107 (49.1%) | 90 (52.6%)  | 124 (51.5%) | 1.25 (0.94, 1.66), p=0.113 | 0.94 (0.65, 1.34), p=0.733     |
|              | A/A      | 80 (34.5%)                 | 61 (28.0%)  | 43 (25.1%)  | 72 (29.9%)  | 1                          | 1                              |
| -709 A>G     | A/A      | 13 (4.5%)                  | 8 (2.8%)    | 12 (5.0%)   | 19 (5.0%)   | 1.67 (0.82, 1.67), p=0.387 | 0.65 (0.32, 1.30), p=0.221     |
|              | A/G      | 88 (30.6%)                 | 96 (33.2%)  | 76 (31.8%)  | 136 (35.7%) | 1.08 (0.80, 1.45), p=0.624 | 0.83 (0.41, 1.68), p=0.599     |
|              | G/G      | 187 (64.9%)                | 185 (64.0%) | 151 (63.2%) | 226 (59.3%) | 1                          | 1                              |
| -376 G>C     | C/C      | 48 (22.4%)                 | 45 (22.1%)  | 32 (20.3%)  | 56 (25.8%)  | 0.74 (0.43, 1.26), p=0.264 | 1.15 (0.74, 1.79), p=0.521     |
|              | G/C      | 105 (49.1%)                | 101 (49.5%) | 89 (56.3%)  | 103 (47.5%) | 0.92 (0.53, 1.58), p=0.762 | 1.08 (0.64, 1.37), p=0.682     |
|              | G/G      | 61 (28.5%)                 | 58 (28.4%)  | 37 (23.4%)  | 58 (26.7%)  | 1                          | 1                              |
| +802 A>G     | G/G      | 39 (21.9%)                 | 41 (24.8%)  | 27 (21.4%)  | 47 (26.0%)  | 0.91 (0.62, 1.35), p=0.644 | 0.82 (0.50, 1.34), p=0.431     |
|              | A/G      | 90 (50.6%)                 | 85 (51.5%)  | 68 (54.0%)  | 85 (47.0%)  | 0.88 (0.62, 1.24), p=0.473 | 0.75 (0.50, 1.14), p=0.180     |
|              | A/A      | 49 (27.5%)                 | 39 (23.6%)  | 31 (24.6%)  | 49 (27.1%)  | 1                          | 1                              |

\* Adjusted for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes



**Table 5.3: Analysis of individual polymorphisms in relation to previous MI**

| Polymorphism |                | Previous MI | Non-MI       | Unadjusted odds ratio<br>(95% CI) | *Adjusted odds ratio<br>(95% CI), p-value |
|--------------|----------------|-------------|--------------|-----------------------------------|---|
| -1986 T>C    | T/T genotype   | 37 (28.2%)  | 182 (25.5%)  | 1.60 (0.91, 2.92), p=0.103        | 1.66 (0.92, 2.99), p=0.095                |
|              | C/T genotype   | 74 (56.5%)  | 371 (52.0%)  | 1.63 (0.94, 2.70), p=0.083        | 1.59 (0.93, 2.71), p=0.090                |
|              | C/C genotype   | 20 (15.3%)  | 160 (22.4%)  | 1                                 | 1   |
| -1612 5A>6A  | 5A/5A genotype | 39 (29.5%)  | 187 (24.0%)  | 2.02 (1.14, 3.60), p=0.016        | 2.04 (1.14, 3.65), p=0.016                |
|              | 5A/6A genotype | 73 (55.3%)  | 399 (51.2%)  | 1.78 (1.05, 3.00), p=0.032        | 1.76 (1.04, 3.00), p=0.035                |
|              | 6A/6A genotype | 20 (15.2%)  | 194 (24.9%)  | 1                                 | 1   |
| -1346 A>C    | A/A genotype   | 41 (34.5%)  | 176 (27.9%)  | 1.77 (0.94, 3.35), p=0.079        | 1.82 (0.96, 3.47), p=0.068                |
|              | A/C genotype   | 63 (52.9%)  | 340 (54.0%)  | 1.41 (0.77, 2.57), p=0.265        | 1.49 (0.81, 2.73), p=0.202                |
|              | C/C genotype   | 15 (12.6%)  | 114 (18.1%)  | 1                                 | 1   |
| -709 A>G     | A/A genotype   | 87 (66.4%)  | 475 (61.13%) | 1.25 (0.47, 3.27), p=0.656        | 1.22 (0.46, 3.23), p=0.688                |
|              | A/G genotype   | 39 (29.8%)  | 269 (34.6%)  | 0.99 (0.36, 2.67), p=0.978        | 0.97 (0.36, 2.65), p=0.954                |
|              | G/G genotype   | 5 (3.8%)    | 34 (4.4%)    | 1                                 | 1   |
| -376 G>C     | G/G genotype   | 35 (29.4%)  | 153 (26.3%)  | 1.52 (0.84, 2.76), p=0.168        | 1.55 (0.84, 2.83), p=0.159                |
|              | G/C genotype   | 64 (53.8%)  | 293 (50.6%)  | 1.45 (0.85, 2.50), p=0.177        | 1.49 (0.86, 2.58), p=0.156                |
|              | C/C genotype   | 20 (16.8%)  | 133 (23.0%)  | 1                                 | 1   |
| +802 A>G     | A/A genotype   | 33 (31.7%)  | 119 (25.2%)  | 1.52 (0.83, 2.78), p=0.176        | 1.55 (0.84, 2.89), p=0.164                |
|              | A/G genotype   | 50 (48.1%)  | 238 (50.4%)  | 1.15 (0.66, 2.01), p=0.622        | 1.19 (0.67, 2.10), p=0.552                |
|              | G/G genotype   | 21 (20.2%)  | 115 (24.4%)  | 1                                 | 1   |

\* Adjusted for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes



**Table 5.4**

**Stepwise logistic regression analysis of alleles and haplotypes in relation to previous MI**

| Haplotypes       | Haplotype frequency |        | Unadjusted odds ratio (95% CI) | *Adjusted odds ratio (95% CI) |
|------------------|---------------------|--------|--------------------------------|-------------------------------|
|                  | Previous MI         | Non MI |                                |                               |
| T-5A-A-A-G-A     | 0.466               | 0.370  | 1.36 (1.02, 1.80), p=0.033     | 1.36 (1.02, 1.81), p=0.034    |
| T-5A-A-A-C-G     | 0.019               | 0.005  | 4.04 (1.29, 12.67), p=0.017    | 4.28 (1.34, 13.8), p=0.014    |
| C-5A-C-G-C-G     | 0.008               | 0.001  | n.s.                           | 8.31 (1.08, 63.76), p=0.04    |
| other haplotypes |                     |        | n.s.                           | ns                            |

\* Adjusted for age, gender, smoking, hypercholesterolaemia, hypertension and diabetes

### 5.3 Discussion

In this present study, an association between the 5A/6A polymorphism in the MMP-3 gene and previous MI was observed. More specifically the 5A allele was significantly more frequent in CAD patients with previous MI than in those without an MI history. These results thus indicate that the 5A allele, which has increased promoter activity, is likely to predispose CAD patients to plaque rupture. These findings support those previously made by Terashima *et al* 1999<sup>116</sup>.

It has also been hypothesised that a genetically determined reduction in the MMP-3 promoter activity, namely the 6A allele is associated with the progression of atherosclerosis<sup>99,182,185</sup>. This hypothesis was also substantiated in this study. The 6A/6A genotype was observed to be associated with increased number of coronary arteries with >50% stenosis.

MMP-3 has a broad substrate specificity. It can degrade many of the extracellular membrane components and also activate other members of its own family such as the collagenases and gelatinases<sup>50</sup>. Its expression is primarily regulated at the level of transcription in response to various stimuli. It has also been demonstrated that variation in its promoter sequence does indeed effect the expression of its gene product (Chapter 6). Plaque progression and plaque rupture are distinct processes involving different pathological mechanisms. Alteration in the ECM of arterial walls is characteristic to atherosclerosis. There is also strong evidence both in this study and investigations made previously, that an imbalance of MMP-3 expression in the arterial wall in either direction can have adverse effects on atherosclerotic lesions. While a reduced MMP-3 activity leads to the accumulation of ECM and the progression of atherosclerosis,<sup>105</sup> an increase in its activity at crucial plaque sites can induce rupture of fibrous plaques and acute thrombosis<sup>107</sup>.

Haplotype analysis investigating association between the disease loci and multiple markers, is postulated to be a powerful tool for the use of SNPs in complex disease. In this study, analysis of different haplotypes in relation to previous MI found three that showed association to CHD. These were T-5A-A-A-G-A, T-5A-A-A-C-G and C-5A-C-G-C-G. The three haplotypes all contain the 5A allele, which supports the analyses

of individual polymorphisms that the 5A/6A polymorphism, but not the other polymorphisms studied, is associated with risk of MI in CAD patients.

The T-5A-A-A-G-A haplotype was found to increase plaque vulnerability by 1.36-fold. This haplotype represented one of the most common haplotypes in this study, thus suggesting that with respect to MI, this haplotype is the most important one to consider for further investigations. The data however also showed a 4-fold and 8-fold increase in disease susceptibility with the T-5A-A-A-C-G and C-5A-C-G-C-G haplotypes. In the Caucasian samples investigated in this study, the frequencies of this haplotypes were found to be low. These rare haplotypes are thus unlikely to contribute to the development of MI in a large proportion of the patients.

These findings support the concept that matrix accumulation is increased in individuals carrying the 6A allele of the MMP-3 gene, which has a lower transcriptional activity; and individuals carrying the 5A allele of the MMP-3 gene, which is transcriptionally more active, are more prone to an increase in matrix degradation and atherosclerotic plaque instability. It therefore follows that the 6A allele predisposes individuals to the development of more fibrous plaques, which characteristically cause a higher-grade stenosis, and carriers of the 5A allele are more likely to develop the more vulnerable lipid rich plaque prone to rupture, causing MI. The data also shows that individuals heterozygous for the 5A/6A polymorphism have a similar phenotype to individuals homozygous for the 5A allele, thus suggesting that this allele is more dominant over the recessive 6A allele.

The large increase in MI susceptibility expressed by the T-5A-A-A-C-G and C-5A-C-G-C-G haplotypes is rather interesting and difficult to explain on data produced in this study. The -376/C and +802/G alleles do not alone cause susceptibility, as there are other haplotypes in which they occur that have not shown any significance to the risk of MI. Functional studies performed in this study found that the -376/G allele was transcriptionally more active than the -376/C allele. Thus, the presence of the -376/C allele in these two haplotypes is likely to be due to LD with another disease causing allele (e.g. the 5A allele), rather than due to a functional effect from the -376G>C itself. It is also likely that the +802 codon region might affect activation of the MMP-3 protein as it is located close to the cleavage site of three proteinases. As already

discussed in Chapter 3, this codon is not evolutionarily conserved and so it is also unlikely to cause an effect. Further analysis may still be required to rule this concept out completely. A simple in vitro study on MMP-3 protein activation can be performed on the two protein isoforms where both are subjected to the same proteinases, specific to MMP-3 and compared.

MMP-3 is involved in a number of biological activities and as an important mediator of matrix remodelling; it has been implicated in a number of destructive diseases. The 5A/6A polymorphism has stimulated a great deal of interest in recent years. Data from several studies have found evidence of its involvement in a number of degenerative and inflammatory diseases<sup>90,113,114,192</sup>. The data of this present study indicate that the effect of the MMP-3 variation is attributable to the 5A/6A polymorphism, rather than the other variants in this gene, may also be relevant to the understanding of the genetic basis of these diseases.

In conclusion, the results of this study show that the 5A/6A genetic variant is a major genetic determinant for the risk of atherosclerosis and MI. Three haplotypes have also been observed to increase the risk of MI. Though this was a large study of patient and control subjects, further investigations are still warranted to verify the impact of both alleles of this gene polymorphism and the associated haplotypes on patients with these clinical events in different ethnic populations.

## **Chapter 6**

### **Functional analysis of the MMP-3 promoter polymorphisms and Protein-DNA interaction assays**

#### **6.1 Introduction**

There are several processes involved in the conversion of genetic information in DNA to produce the biologically active protein (gene expression). These are, (1) Transcription, (2) RNA processing, (3) RNA transport, (4) Translation, (5) mRNA degradation and (6) Protein activation. In theory, gene expression can be regulated at any one of these steps. Though there are cases where post-transcriptional control have been observed, it is the process of transcription, the initial step, in which gene regulation is primarily achieved.

Gene transcription is controlled by two fundamental types of components, specific short DNA sequences in the promoter region (cis-element) and the gene regulatory proteins that recognise and bind to them (transcription factors). There are four classes of cis-elements. These are promoter elements, enhancer sequences, response elements and silencers. Response elements are only found in some genes and silencers as the name suggests, are negative regulatory elements. Promoter and enhancer elements are found in all genes. Enhancer sequences increase RNA polymerase activity by binding transcription factors, while the promoter elements include the short sequence elements of the TATA box and CCAAT box. These regulatory sequences are predominantly located upstream (5') from the start of transcription. Some elements are known to occur downstream relative to the transcription start site and have been found to exist within the genes themselves.

Transcription factors are typically homo or hetero-oligomeric. It is believed that just over 5% of our genes encodes for transcription factors. Binding of transcription factors to DNA regulatory sequences initiates transcription. But it is the combinations of enhancer sequences, together with the strength of the promoter elements within the DNA sequence of the gene and the interactions between the activator and repressor transcription factors, that exert control on eukaryotic gene expression. Biophysical studies have recently shown that transcription factor assembly is often subject to

regulation. This regulation is thought to determine the rate at which transcription initiation occurs via its effect on the affinity of the transcription factor for the specific DNA sequence. There are a number of different factors that can modulate transcription factor assembly. These include post-transcriptional modification, the binding of small molecules, DNA binding and the interaction of other proteins. The order proteins assemble onto these promoter sequences, and the number and type of transcription factors involved, are specific to the gene. Gene regulatory proteins can activate or repress the transcription of a gene. They can function alone, however most act as part of a complex composed of several polypeptides each with distinct functions. In this way, these proteins can participate in more than one regulatory complex and may act as an activator of transcription in one and as a repressor in the other. A wide range of regulatory pathways controls transcription factors and in general, different cell types express characteristic transcription factors relative to their function. The expression of metabolic enzymes for example, is generally regulated by metabolites. The initial activation of transcription promotes the access to DNA and facilitates the recruitment of RNA polymerase enzyme to the start of transcription.

There are three RNA polymerase complexes in eukaryote that play an important role in the transcription of RNA. In general, RNA polymerase I transcribes ribosomal RNAs. RNA polymerase II transcribes the protein encoding gene, mRNAs and some nuclear RNAs (snRNA) and RNA polymerase III transcribes tRNA genes. RNA polymerase II is the protein that assembles just in front of the start site of transcription. It transcribes the gene and in simple organisms, no other proteins are required for this initial step in gene expression. Complex organisms require other proteins, i.e. transcription factors, that assemble onto the promoter. Some of these bind to the polymerase itself to aid the binding of RNA polymerase II to the gene sequence. The order these proteins assemble onto the promoter and the type of proteins involved, is specific to the gene. Control of gene expression is important in multicellular organisms. Activating the right gene in the right cell at the right time is essential for both development and cell and tissue regulation. Polymorphisms in the promoter region of a gene can delete or introduce cis-elements, thus effecting transcription factor binding and consequently transcription initiation. This in turn can cause defects in the cell and surrounding tissue which can lead to dysfunction and disease.

**Table 6.1: -Transcription Factors**

| Transcription Factors | Sequence Motifs   | Notes   |
|-----------------------|---|---|
| c-Fos                 | TGA <sup>C</sup> /G T <sup>C</sup> /A A                                 | These proteins bind to the AP-1 site              |
| c-Jun                 |   |   |
| c-Ets                 | G <sup>G</sup> /C <sup>A</sup> /C GGA <sup>A</sup> /T G <sup>T</sup> /C | Predominately associated with B and T cells       |
| c-Myb                 | T <sup>T</sup> /A AAC <sup>G</sup> /T G                                 | Haematopoietic cell-specific transcription factor |
| MyoD                  | CAACTGAC  | Associated with myoblasts and myotubes            |
| GATA                  | T <sup>T</sup> /A GATA  | Erythroid cell specific transcription factor      |
| NF- (kappa) B         | GGGA <sup>A</sup> /C TN <sup>T</sup> /C CC                              | Predominately associated with B and T cells       |
| c-Myc                 | CACGTG  | Oncovirus   |

Table 6.1 A selection of transcription factors.

The term gene expression commonly refers to the entire process whereby the information encoded in a particular gene is decoded into a particular protein. Theoretically, each of these regulatory steps could lead to differential gene expression. Analysis of the various steps of gene expression in various prokaryotic and eukaryotic genes has lead to the conclusion that the decision to initiate transcription is the most important controlling point in determining whether or not most genes are expressed, and how much of the encoded mRNAs, and consequently proteins, are produced.

Analysis of the regulation of gene expression in animal cells is one of the focal points of present-day molecular biology. Gene regulation can be assayed by introducing manipulated genes into embryonic stem cells and cultured cells, a procedure called transfection. There are two types of transformation techniques, transient and stable. Rapid analysis can be achieved using transient transfection. Here the DNA of interest is not integrated into the host genome and inherited as progeny cell, but is maintained

for several days and so is used to characterise *in vivo* responses. Stable transfection requires exogenous DNA to be integrated into the host genome, giving rise to stable transfected cells. This form of transfection is long-term and its purpose is to isolate and propagate cloned individuals containing the transfected DNA. Stable transfection occurs at a relatively low frequency and these cells require a selectable marker, such as neomycin, dihydrofolate reductase, hygromycin B phosphotransferase or XGPRT.

## **6.2 Transfection**

The approaches for introducing DNA into cells is diverse, from direct micro-injection of DNA into the fertilized eggs of *Xenopus*, or the embryos of mouse or *Drosophila*, to the less taxing methods of introducing foreign DNA into cultured animal cells<sup>193-197</sup>. There are a number of methods for introducing DNA into cells, however viral infection is the most efficient method of getting foreign DNA into a large number of cells. These techniques require the formation of recombinant viruses. DEAX-Dextran, calcium phosphate co-precipitation, liposome mediated transfection and electroporation are a few of the established transfection methods mentioned briefly below: -

### **6.2.1 Chemical Reagents**

#### **6.2.1.1 DEAE-Dextran**

DEAE-Dextran (Diethylaminoethyl-dextran) was one of the first chemical methods employed for the transfer of endogenous DNA into cultured mammalian cells<sup>194</sup>. Two distinct methods have been developed for DEAE-Dextran mediated transfection. In the standard protocol, DEAE-Dextran is mixed into the transfer medium in which the DNA is present<sup>198</sup>. The second protocol involves pretreating the cell with DEAE-Dextran, removing the solution and washing the cells before integrating the DNA<sup>199</sup>. DEAE-Dextran is a polymeric cation, which therefore associates tightly with negatively charged DNA and in this way carries the DNA into the cell by endocytosis.

This is a simple and inexpensive method and has a high efficiency of transfer. DEAE-Dextran however, is quite toxic to cells and therefore requires careful optimisation for



individual cell lines for both DEAE-Dextran concentration and exposure time. It is suitable for transfection in both adherent and suspended cell types.

### **6.2.1.2 Calcium Phosphate Co-Precipitation**

This is another widely used method for introducing DNA into mammalian cells which has been generally accepted after the application by Graham and Van der Eb (1973)<sup>195</sup>. It involves slowly mixing DNA directly with calcium chloride and a HEPES-buffered phosphate solution to form a fine precipitate that can then be dispersed over the cultured cells. Successful transfer is dependent on the co-precipitation of DNA with calcium phosphate, which can be achieved when calcium phosphate is made fresh in the presence of the DNA. The cells phagocytose calcium phosphate/DNA precipitate granules and in this way, the DNA of interest is integrated into cells. It is routinely used for both transient and stable transfection in a range of cell types<sup>200,201</sup>.

This is a simple and inexpensive technique useful for both transient assays and stable transformations. It has a high efficiency of transfer in some cell types but can be toxic to others and so requires careful optimisation of reagent concentrations and exposure time.

Transfection efficiencies can be increased in some cell types when using either DEAE-Dextran or calcium phosphate precipitation mediated transfection by treating the cultured cells with one of several chemicals after primary exposure to cell. Glycerol, DMSO (dimethylsulphoxide), chloroquine and sodium butyrate are some of the most effective and routinely used agents for this purpose.

## **6.2.2 Cationic Lipids**

### **6.2.2.1 Liposome Mediated Transfection**

Cationic liposome-mediated transfection, also known as lipofection, was first reported in 1987<sup>196</sup>. DNA in solution spontaneously forms a complex with the liposomes. The cationic head of the lipid compound is positively charged and so associated with the

negatively charged phosphate groups on the DNA molecules. Transfection is thought to occur by the fusion of the lipid/DNA complex with the plasma membrane of the cultured cells.

Cationic liposome-mediated transfection is more efficient than DEAX-dextran and calcium phosphate co-precipitation. It is highly reproducible and can deliver DNA to some difficult cell types. Liposomes are easy to use and have been utilised successfully in *in vivo* application. Though this requires no specialised equipment, liposome reagents are rather expensive.

### **6.2.3 Physical**

#### **6.2.3.1 Electroporation**

Electroporation can be used for both transient and stable assays. It uses electrical impulses that cause temporary relaxation of the selective permeability properties of the plasma membrane. This allows exchange of molecules including DNA, across the membrane <sup>202</sup>. It has been used successfully for both transient and stable transfection experiments. Transformation efficiency is improved by treating the cells with colcemid before electroporation <sup>203</sup>. The theory is that this drug causes the cells to arrest in metaphase and either the nuclear membrane is lacking or unusually permeable.

It is a rapid and simple method of introducing cloned genes into most cell types and is the preferred method for cell lines that are difficult to transfect. Transfection efficiency between samples however, varies greatly and so needs to be standardised between samples which can be achieved by co-transfection. It requires specialised equipment, namely an electroporator apparatus.

#### **6.2.3.2 Direct micro-injection**

As its name suggests, microinjection delivers nucleic acids directly into the nucleus or cytoplasm of cells through a fine needle <sup>193</sup>. This method is primarily used to transfer

DNA into established cell lines such as embryonic stem cells. These carry the integrated copies of DNA to create transgenic organisms.

This is a very efficient technique and unlike other transfection methods, DNA is not exposed to cellular compartments such as low pH endosomes. It cannot however be used to introduce DNA to a large number of cells. It requires special equipment, which is costly and the technique is labour intensive.

### 6.3 Reporter Gene Assays

Gene expression is detected through either reporter genes or by RNA analysis. Analysis of RNA can be achieved via a number of different methods such as northern blots, RNase protection, or S-1 hybridisation. These methods are more direct than the alternative, which measures protein expression. They can, however be quite tiresome and some cannot be quantitated. Consequently more assays have been developed to measure reporter genes/proteins rather than mRNA. Reporter genes are purposely designed such that they are not endogenous to the cell or found naturally in human cells and can be coupled with a sensitive assay system for the gene product. They are easy to use and versatile<sup>204</sup>.

A reporter assay system was the method employed in this study. In such an assay, the sequence of interest is cloned into the vector immediately upstream from the reporter gene. The vectors are designed in such a way that the expression is controlled almost entirely by the cloned DNA fragment. Expression of the reporter gene occurs when the construct is transfected into the cultured animal cells. Gorman *et al* first initiated the reporter gene concept in 1982 using the bacterial chloramphenicol acetyltransferase (CAT) gene and assay<sup>205</sup>. Since then several reporter genes and assays have been developed for the analysis of transcriptional regulation such as  $\beta$ -galactosidase, firefly luciferase and growth hormone<sup>204</sup>. Brief descriptions of some of the popular reporter genes are given below.

Reporter gene assays have played a significant role in the study of eukaryotic gene expression and regulation. Reporter gene assays are most frequently used for the analysis of cis-acting transcription elements<sup>206</sup>. When applied in this form the

character and function of the promoter or enhancer elements for study can be identified by the transcriptional activity expressed by the reporter gene. The promoter is cloned into the vector upstream of the reporter gene, while the enhancer elements are cloned either up or downstream from the reporter gene. The construct is then transfected into the cell culture or germ cell and using reporter gene technology, the expression of the reporter can be analysed. Control gene vectors are used to normalise for differences in transfection efficiency or cell lysate recovery between treatments and transfection experiments. These vectors are co-transfected with the experimental constructs and are typically driven by a strong promoter. The reporter genes in the control vectors differ from those in the experimental vectors so they can be assayed individually. Commonly used controls vectors are those containing genes for  $\beta$ -galactosidase or *Renilla* luciferase.

### **6.3.1 CAT (chloramphenicol acetyltransferase) Reporter Assay Systems**

The CAT gene is one of the most widely used reporter gene for gene expression assays in mammalian cells, because the gene is not found in any eukaryotes. Therefore, there is no background CAT activity to effect assay results.

The CAT gene is derived from transposon 9 of *E.coli*.<sup>207</sup> It encodes a trimeric protein, which comprises of three identical subunits of 25 kDa in length<sup>208</sup>. The CAT enzyme catalyzes the transfer of the acetyl group from acetyl-CoA to the substrate chloramphenicol. The level of expression is generally quantitated by incubating the cell lysate with [<sup>14</sup>C] chloramphenicol. When the product is formed, it is separated by using either thin layer chromatography (TLC) or organic extracts<sup>209</sup>.

### **6.3.2 $\beta$ -galactosidase Reporter Assay Systems**

$\beta$ -galactosidase is an enzyme encoded by the Lac Z gene found in *E.coli*<sup>210</sup>. It is a tetrameric enzyme approximately 500,000 Da in size and catalyzes the hydrolysis of lactose into glucose and galactose. A number of specialized substrates can be assayed with this protein to quantitate enzyme activity. The levels of expression can be measured using a spectrophotometer, fluorometer or luminometer.

The  $\beta$  galactosidase reporter gene is a favoured choice in histochemical studies as its expression can be easily assayed *in situ* using histochemical staining. One drawback in using this reporter gene is that there are certain mammalian cells that have endogenous lysosomal  $\beta$  galactosidase activity. It is therefore more frequently used as a control vector for normalizing transfection efficiency in co-transfection experiments with another reporter.

### 6.3.3 Firefly Luciferase Reporter Assay Systems

In recent years, the luciferase enzyme has become one of the most popular reporter genes. Derived from the luc gene of the firefly *Photinus pyralis* it encodes an enzyme of 60.7 kDa in size<sup>211-213</sup>. It functions by catalysing D-luciferin and ATP in the presence of oxygen and magnesium ( $Mg^{2+}$ ), which causes the emission of light. This reaction can be quantitated with the aid of a luminometer, which measures light output. The total amount of light measured at a given time period is proportional to the amount of luciferase activity in the sample.

The luciferase protein has a shorter half-life in transfected mammalian cells than other reporter proteins such as CAT and is therefore more suited for transient transfection assays. Another advantage it has over other reporter gene assay systems, is that results can be obtained in minutes rather than hours or days. Luciferase has the advantage of being very sensitive, as the catalytic reaction of the oxidation of luciferin with the emission of yellow-green light can be detected easily at low levels.

### 6.4. Protein Interaction Assays

Transcriptional assays were initially developed to identify the roles of specific factors and enzymes involved in transcription initiation. They are now more commonly used to identify the location/presence of the proteins and their binding sites.

#### **6.4.1 EMSA (Electrophoretic-mobility shift assay)**

EMSA (Electrophoretic-mobility shift assay) also referred to as the gel-shift assay, is a technique that determines the binding interactions between the DNA and the DNA-binding proteins. In this assay, the electrophoretic mobilities of radiolabelled DNA fragments are determined in the presence or absence of a sequence-specific DNA-binding protein. EMSA relies on the effect of the bound protein on the migration of the DNA molecule in an electric field. Protein binding generally reduces the mobility of a DNA fragment, causing a shift in the location of the fragment band, which is detected by autoradiography<sup>214-216</sup>.

The exact location of the DNA-protein interaction cannot be determined using the mobility shift assay alone. Therefore, further experiments need to be executed. Such techniques include DNase I footprinting assay or methylation interference.

#### **6.4.2 Footprinting Analysis of DNA-Protein Complexes**

A number of footprinting experiments have been developed that can be used to find out to which sites proteins bind when they interact with nucleic acids. This method is based around the idea that nucleic acids are protected from indiscriminate cleavage by enzymes and chemical agents. Chemical modification of a DNA molecule can inhibit the binding of sequence specific DNA binding proteins as it can change the properties of the nucleotides in the sequence (interference). Alternatively, if a sequence specific protein is bound to the DNA molecule, it can protect the binding site from both chemical and nuclease modification (protection). These effects of interference and protection can be analysed by separating the fragments through a denaturing polyacrylamide gel followed by autoradiography or phosphor-image analysis. Autoradiography produces a ladder of bands. The region where no bands occur is called the “footprint” of the protein. The DNA is protected from the cleavage agent by the protein bound to it<sup>217</sup>.

#### 6.4.2.1 DNase I Footprinting

DNase I is the favoured nuclease, although other enzymes such as nuclease P1 have also been employed in protection analysis<sup>218,219</sup>. Cis-acting DNA control sites can be further analysed by employing deoxyribonuclease (DNase) I footprinting. In this technique, the cis-acting element is protected from DNase I digestion, by the binding of a transcription factor. This method is the least toxic of the protection/interference footprinting assays and was first used by Schmitz and Galas in 1978<sup>220</sup>.

DNase I is a 37 kDa enzyme. It has comparatively little sequence preference and therefore cleaves along the entire length of the DNA backbone. DNase I can function in a wide range of temperatures, pH and salt conditions<sup>120</sup>.

The DNA sequence and protein are mixed in a similar method to the mobility shift assay, however only one end of one strand of the DNA fragment is required for radioactive labelling. The DNA fragment is then digested with a small amount of DNase I, which cleaves DNA at phosphodiester bonds. After digestion the bound proteins are removed and the DNA fragments are separated through a polyacrylamide gel by electrophoresis<sup>216</sup>. This produces a ladder of bands representing the products of DNase I digestion. The binding of a transcription factor protein prevents digestion of the DNA fragment from occurring and these fragments will therefore be absent from the footprint. Low occupancy of a site may result in small changes in the intensity of the bands on the gel instead of a footprint<sup>120</sup>.

#### 6.4.2.2 Methylation Interference Assay

This technique also characterises the DNA-protein interaction, but, unlike the protection assays mentioned above, the DNA is modified before it is incubated with binding protein. The alkylating agent dimethylsulphate (DMS), first used by Mirzabekov, Gilbert *et al*, is the most commonly used chemical agent for this type of study<sup>221</sup>. The chemicals used as modifying agents leave the backbone of the DNA intact and instead cause changes in individual nucleotide bases.

DMS is the most commonly used chemical reagent and functions by alkylating the purines. Modification is then followed by cleavage in hot piperidine. This enzyme is Guanine (G)-specific <sup>216</sup>. Ethylnitrosourea (ENU) is another commonly used agent in interference assays and it works by alkylating the phosphates on the DNA backbone <sup>222</sup>.

This assay has the advantage that insufficient binding does not affect the results. Unfortunately, recovery of the DNA-protein complex and free DNA bands from the acrylamide gel is difficult and low yielding <sup>216</sup>.

### **6.4.3 Supershift Gel Mobility Shift Assay**

This is an extension of EMSA. Double stranded oligonucleotide probes are prepared and N-labelled in the same way as the EMSA technique. In the supershift assay, however, the probes are incubated not only with nuclear extracts, but also with an antibody with a specific transcription factor. The mobility of the complex containing the probe, transcription factor and antibody is further retarded compared with a complex without the antibody.

### **6.4.4 Sequence-Specific DNA Affinity Chromatography**

This technique is usually performed as a final step in the process of determination and isolation of transcription factors. It relies on the sequence-specific protein to recognise and bind to the target sequence, with a high affinity, in the presence of non-specific DNA <sup>216</sup>. It requires the synthesis and hybridisation of long DNA strands comprising of multiple copies of the transcription factor-binding site. These molecules are then coupled covalently with a solid support, creating a sequence-specific affinity column. Partially purified extract containing the protein (transcription factor) in question together with excess of soluble, competitive DNA is applied to the column in a low-salt buffer (100 mM KCl). Proteins that do not bind to the sequence are washed away with the low-salt buffer. Proteins with a low affinity to the binding site can also be removed with the application of a higher concentration of the same buffer (300 mM KC). The transcription factor that has a high affinity for this binding site can then be eluted from the column by the simple application of a high salt buffer (1 M KCl).



For this study, a Dual-luciferase reporter assay system from Promega was utilized. This assay system allows the measurement of two luciferase reporter enzymes simultaneously. The two types are firefly luciferase and *Renilla* luciferase. The latter is a 36 kDa monomeric protein cloned from the anthozoan coelenterate '*Renilla reniformis*', the sea pansy. The firefly and *Renilla* luciferases arise from two distinct evolutionary origins and therefore enzyme and substrate requirements are different. Thus, it is possible to discriminate between their respective bioluminescent reactions

223

In this study, the firefly luciferase luminescence was used to measure the rate of transcription and the *Renilla* luciferase was used to standardise the uptake of DNA between the different samples.

Constructs were created using one of Promega's pGL3 vectors, containing the firefly luc gene. There are four different pGL3-vectors with a combination of modifications that provides a greater flexibility in performing genetic manipulation (Figure 6.2)<sup>124</sup>; -

- 1) pGL3-Basic vector: This lacks eukaryotic promoter and enhancer elements.
- 2) pGL3-Promoter vector: This contains an SV40 promoter upstream of the luc+ reporter gene.
- 3) pGL3-Enhancer vector: This contains an SV40 enhancer downstream from the luc+ reporter gene.
- 4) pGL3-Control vector: This contains both SV40 promoter and enhancer sequences.

The control reporter gene vector was one of the pRL family of vectors, i.e. pRL-TK (Figure 6.1). This vector contains the herpes simplex virus thiamine kinase promoter, upstream from the Rluc gene.

In atherogenesis, MMP-3 is largely expressed by macrophages, so macrophage cell lines were utilized in this study. It therefore followed that electroporation was the method employed for the transfection as macrophage cells are difficult to manipulate and therefore, electroporation, is probably the easiest and cheapest method to achieve success.

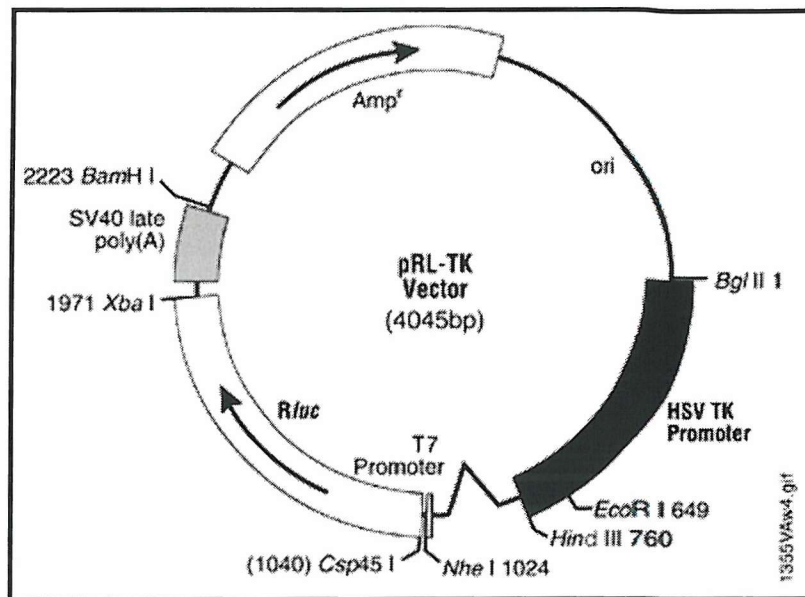
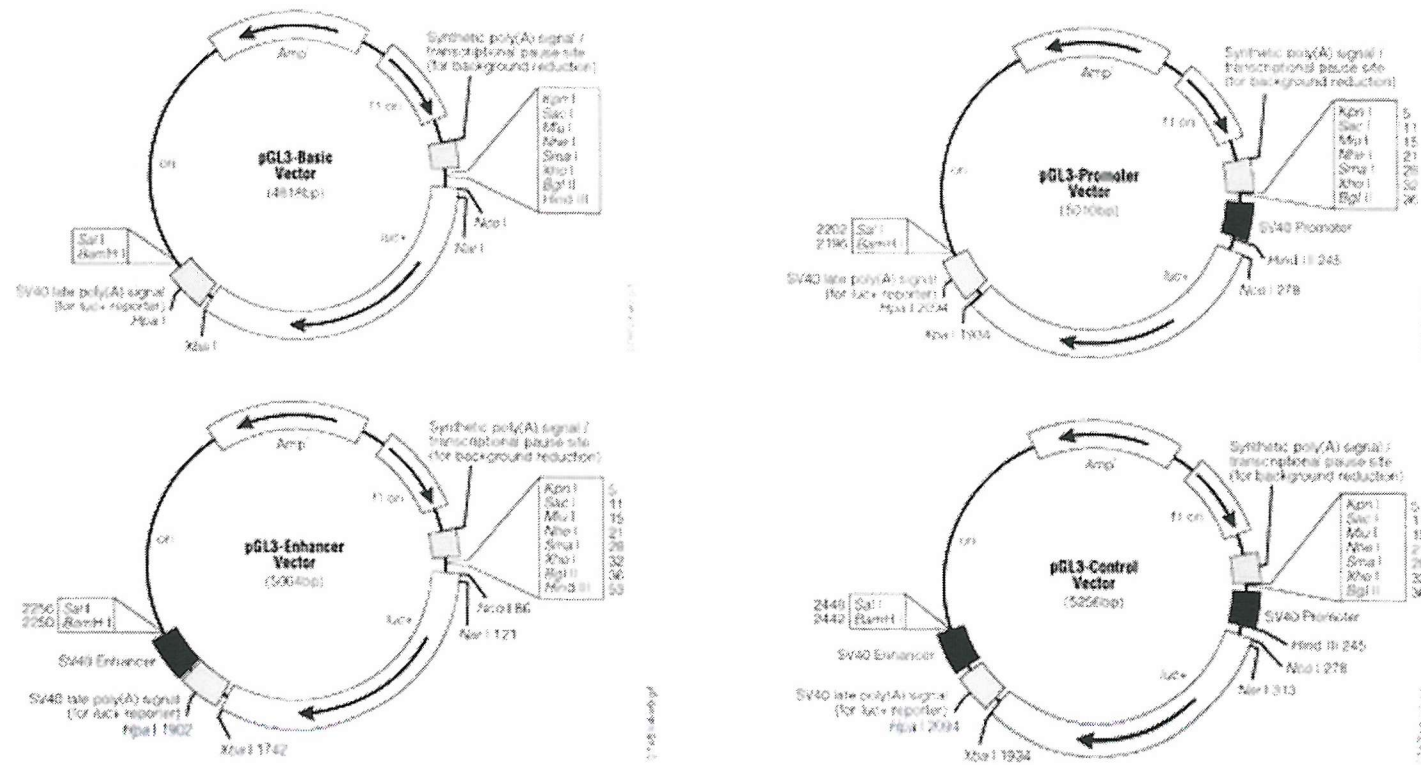
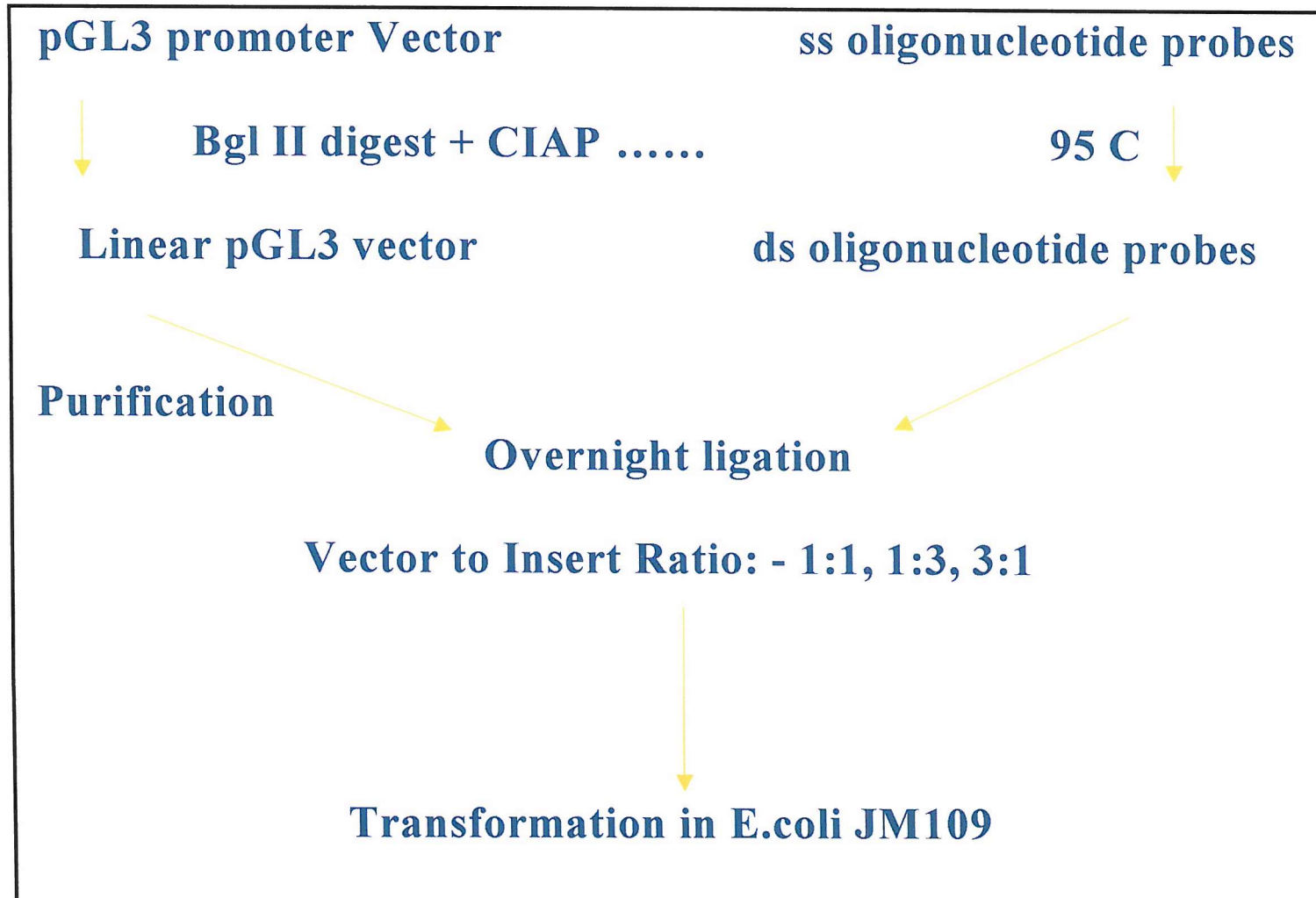


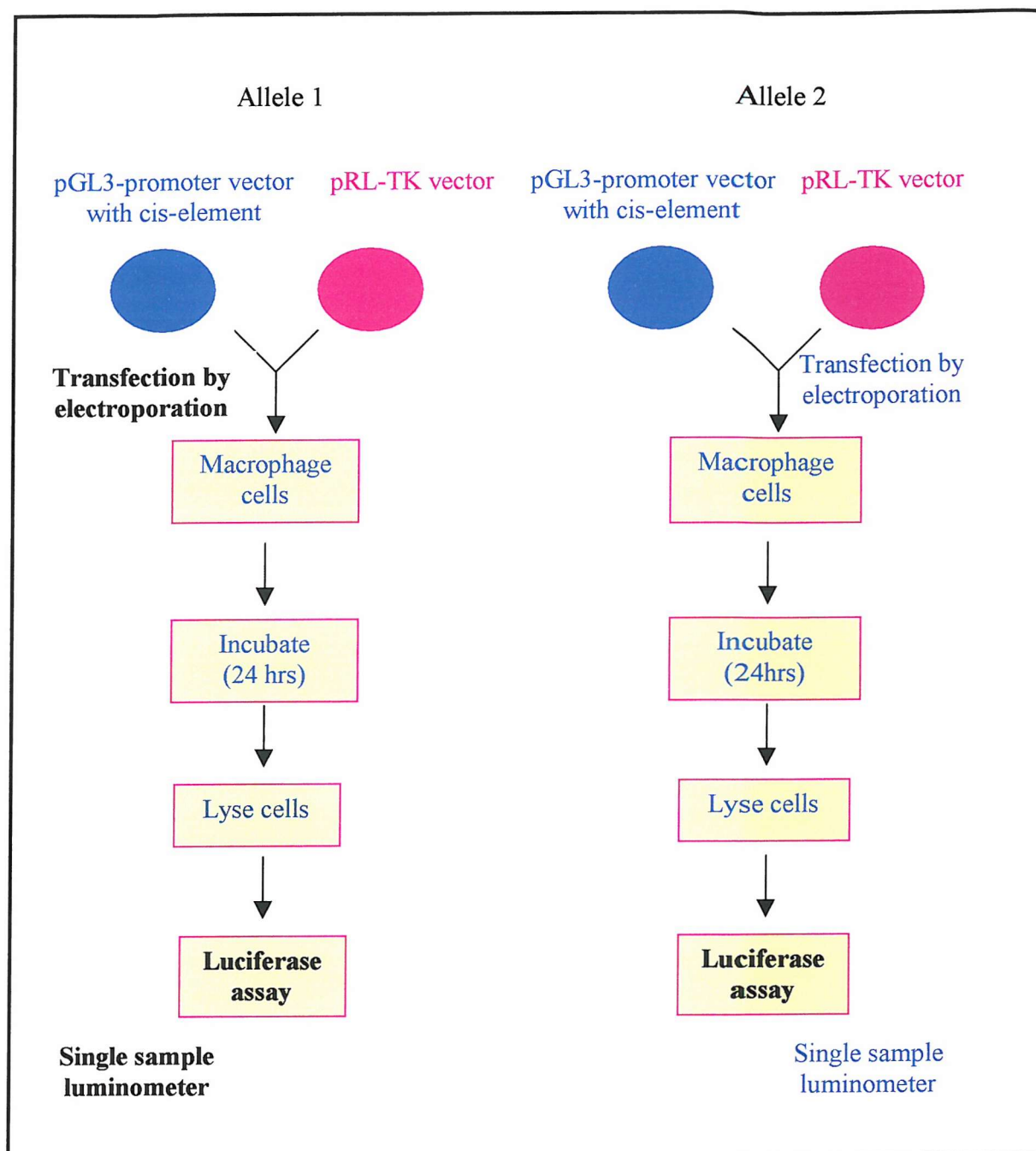
Figure 6.1: - Circular map of the pRL-TK vector<sup>124</sup>



**Figure 6.2: - Circular map of the pGL3 Vectors**<sup>124</sup>



**Figure 6.3:** - Flow chart illustrating the method used to produce reporter gene constructs



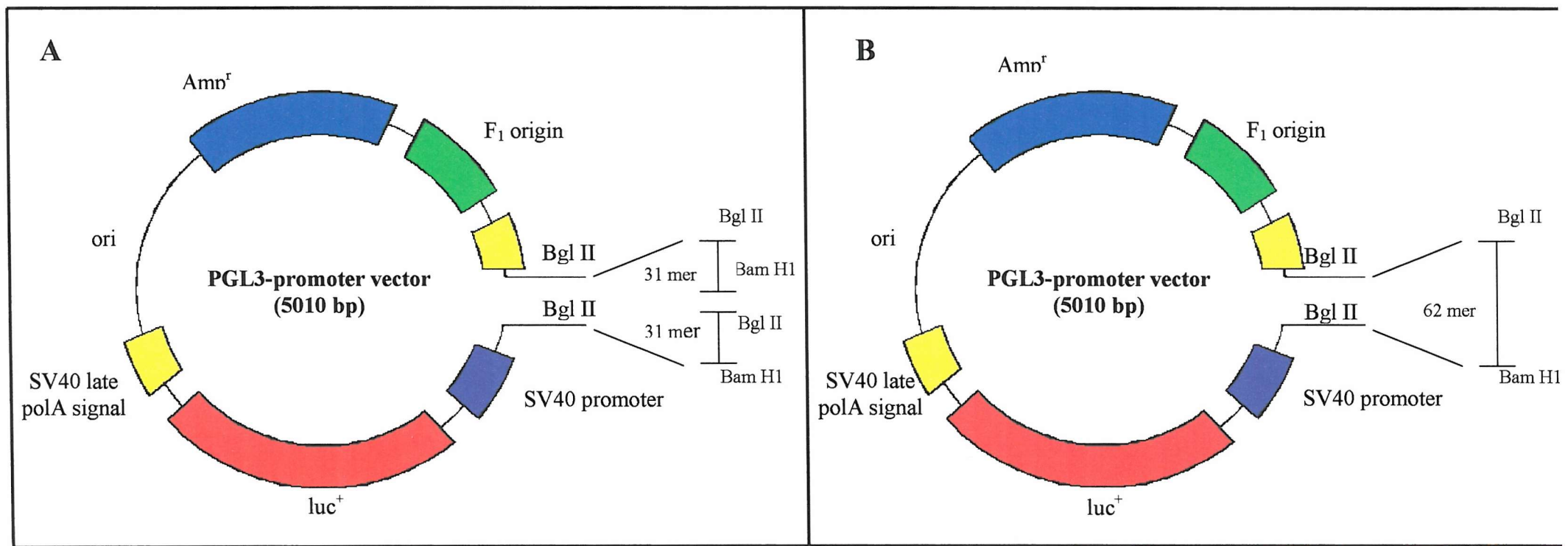
**Figure 6.4:** - Flow chart showing the general scheme used for transient transfection.

## 6.5 Results

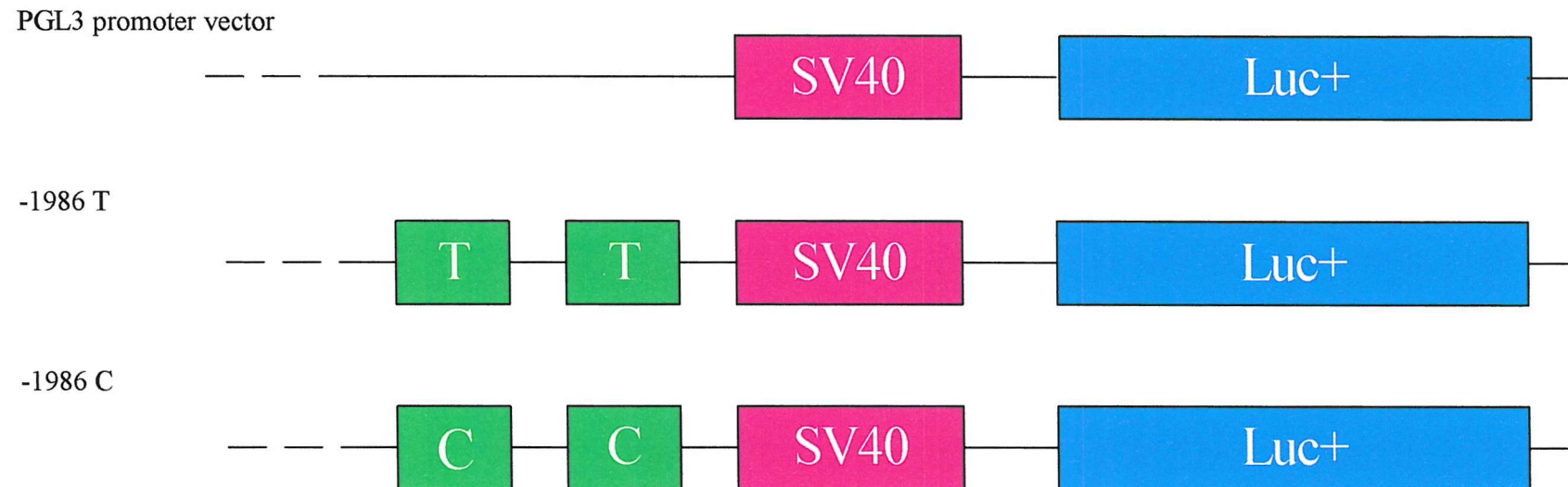
To distinguish functional variants from their non-functional counterpart, a series of reporter gene constructs were generated for each individual polymorphism at positions –1986, –1612, –1346, –709 and –376. Reporter gene constructs for the whole MMP-3 promoter region, encompassing all five polymorphisms, were also prepared to determine the overall effect on MMP-3 promoter activity. The differences in allelic expression were measured by performing a series of mammalian cell, transient transfection experiments using the constructs.

The initial strategy for producing each SNP insert was to insert two copies of a 31 mer DNA fragment, encompassing each allele of the SNP of interest. Though this design has been used successfully by various groups<sup>52,96,98,224</sup>, here it was difficult to reproduce. The DNA fragments, when inserted, favoured the wrong orientation. For this reason a second approach was formulated whereby a longer 62 mer DNA fragment consisting of two copies of the original probe was inserted into the vector (Figures 6.5 - 6.6). Using this method there was a 50:50 chance that the DNA fragment would be inserted in the correct orientation. The DNA sequences were cloned immediately upstream from a SV40 minimal promoter and the gene encoding luciferase in a pGL3-promoter vector.

The two most common haplotypes were investigated for their effect on promoter activity. Only one copy of each allele encompassing the whole MMP-3 promoter was cloned upstream from the gene encoding luciferase in a pGL3-basic vector, to produce a construct (Figure 6.8). The DNA construct covered the region from –2309 to +53 of either the T-5A-A-A-G-A-A (pGL3-basic 5A) or C-6A-C-G-C-G-G (pGL3-basic 6A) haplotype.



**Figure 6.5:** -Schematic diagrams representing the two approaches used to produce the reporter gene constructs: A, was the original strategy used, B, characterizes the revised strategy



**Figure 6.6:** - Schematic diagram of reporter gene constructs used to measure allelic expression in transient transfection experiments, each containing 2 concatenated 31 bp DNA sequences inserted as one unit and corresponding to alleles – T or C of the SNP at the -1986 polymorphic site.



Constructs were transfected into cultured macrophage cells by electroporation. After an incubation period of 24 hrs, the luciferase levels were measured and standardised to the levels of *Renilla* luciferase.

The levels of luciferase between the alleles at each polymorphic site were compared with each other and the control vectors (pGL3-control and pGL3-promoter vector) without inserts. Consistent differences in MMP-3 promoter activity were found to exist between the alleles at each polymorphic site in six independent sets of experiments. The results of these experiments are clearly represented in Table 6.1 and Figure 6.7.

The T allele of the T-1986/C variant expressed a ~1.4-fold increase in promoter activity than the C allele.

The 5A allele of the 5A-1612/6A variant expressed a ~1.3-fold increase in promoter activity than the 6A allele.

The A allele of the A-1346/C variant expressed a ~1.4-fold increase in promoter activity than the C allele.

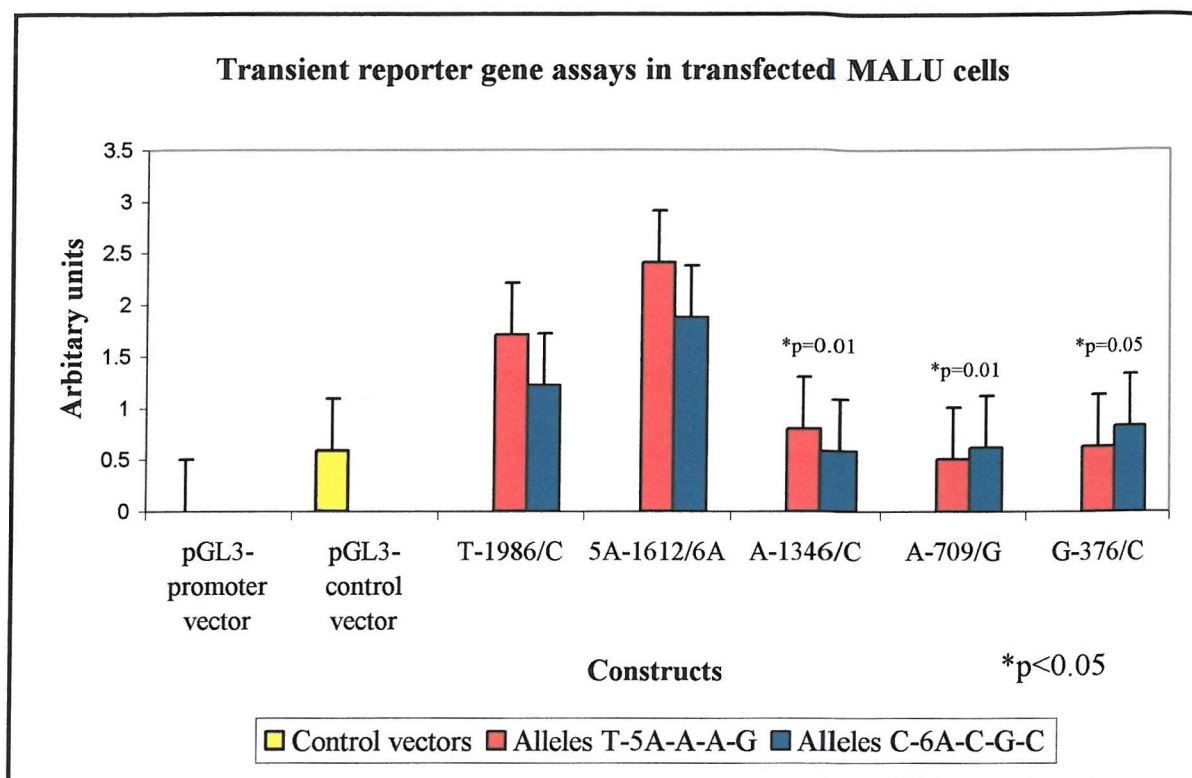
The A alleles of the A-709/G variant expressed a ~1.2-fold decrease in promoter activity than the G allele.

The G alleles of the G-376/C variant expressed a ~1.3-fold decrease in promoter activity than the C allele.

Cumulatively the expression levels between the two haplotypes were found to be higher in the one containing the 5A variant (T-5A-A-A-G), this is consistent with previous studies.

**Table 6.2:- Data represents the difference in transcriptional activity between each biallelic polymorphism**

|                             |                             |                            | SNP 1    |        | SNP 2   |         | SNP 3  |        | SNP 4  |       | SNP 5  |         |
|-----------------------------|-----------------------------|----------------------------|----------|--------|---------|---------|--------|--------|--------|-------|--------|---------|
|                             | pGL3-<br>promoter<br>vector | pGL3-<br>control<br>vector | T-1986/C | C-1986 | 5A-1612 | 6A-1612 | A-1346 | C-1346 | A-709  | G-709 | G-376  | C-376/G |
| Assay 1                     | 0                           | 0.67                       | 0.781    | 0.575  | 1.887   | 1.626   | 0.823  | 0.6599 | 0.577  | 0.701 | 0.592  | 0.724   |
| Assay2                      | 0                           | 0.48                       | 0.812    | 0.538  | 2.012   | 1.928   | 0.816  | 0.648  | 0.642  | 0.745 | 0.445  | 0.526   |
| Assay 3                     | 0                           | 0.46                       | 0.75     | 0.505  | 2.808   | 1.854   | 0.806  | 0.494  | 0.458  | 0.634 | 0.599  | 0.773   |
| Assay 4                     | 0                           | 0.44                       | 0.68     | 0.54   | 2.939   | 2.1     | 0.789  | 0.487  | 0.468  | 0.565 | 0.48   | 0.67    |
| Assay 5                     | 0                           | 0.731                      | 3.332    | 2.726  |         |         | 0.798  | 0.647  | 0.431  | 0.542 | 0.965  | 1.337   |
| Assay 6                     | 0                           | 0.692                      | 3.9      | 2.323  |         |         | 0.755  | 0.571  | 0.492  | 0.545 | 0.707  | 0.97    |
| Mean value                  | 0                           | 0.5788                     | 1.70916  | 1.2011 | 2.4115  | 1.877   | 0.7978 | 0.5844 | 0.5113 | 0.622 | 0.6313 | 0.8333  |
| p-values<br>(paired t-test) | -                           | -                          | 0.072    |        | 0.087   |         | 0.01   |        | 0.01   |       | 0.05   |         |



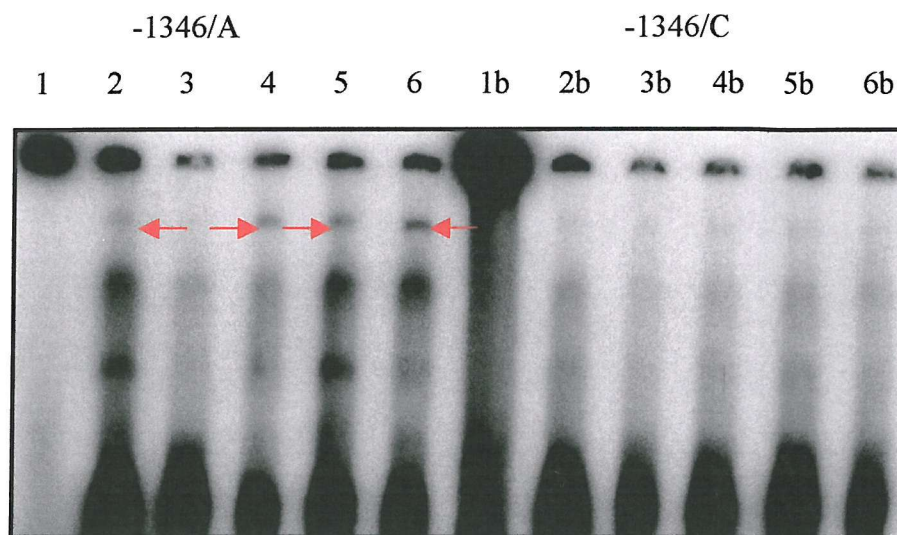
**Figure 6.7:** - Firefly luciferase levels (standardised against *Renilla* luciferase levels) in macrophage MALU cells transfected with different constructs. Data shown are the mean values from six independent experiments.

This allelic-specific effect on promoter activity was also observed in transient transfection experiments with reporter gene constructs containing the whole MMP-3 promoter region (~2.3 Kb). These experiments were performed in two types of mammalian cells, MALU and RAW, both of which are murine macrophage cells (Tables 6.3 and 6.4 Figure 6.9 and 6.10). Reporter gene expression in the T-5A-A-A-G haplotype was between 1.2-1.3-fold greater than in the C-6A-C-G-G-C haplotype. Cultured RAW cells were treated with a protein kinase C activator (PMA) and two types of HMG-CoA reductase inhibitors (statins - pravastatin (hydrophilic) and fluvastatin (lipophilic).) following transfection to determine if they would affect the level MMP-3 of gene expression between the two promoter haplotypes. PMA was used at a concentration of 1  $\mu$ M, while the statin agents were administered at two different concentrations 1  $\mu$ M and 25  $\mu$ M. Each experiment was repeated several times, and Tables 6.3-6.5 and Figures 6.9-6.11 show the results of these experiments.

The DNA mobility shift or gel retardation assay was used in this study to determine whether there were nuclear proteins differentially binding to the different alleles of the promoter polymorphisms identified in this chapter. Dr Shoali Zhang performed this technique. So far, EMSA have been carried out on two (T-1986/C and A-1346/C) of the promoter polymorphisms.

In these assays, the two oligonucleotide probes corresponded to the sequences –1998 bp to –1974 bp and –1358 bp to –1334 bp of each allele for both T-1986/C and A-1346/G polymorphisms in the MMP-3 promoter respectively. Each probe was labelled with  $^{32}\text{P}$  and allowed to interact with crude nuclear extracts prepared from cultured bronchial fibroblasts.

Of the two SNPs so far investigated, only one variant A-1346/C, showed the presence of a DNA-protein complex. The DNA-protein complex (indicated by arrows) was detected in the –1346/A probe but not the –1346/C allele (Figure 6.8). The presence of a specific competitor (unlabelled –1346/A probe) abolished this band (Track 3, Figure 6.8). The presence of unlabelled –1346/C probe or non-specific competitors did not affect this band (Tracks 4, 5, 6 Figure 6.8)



**Figure 6.8:** - DNA electrophoretic mobility shift assay. Each probe was prepared by annealing two complimentary single stranded oligonucleotides (25-mer). The double stranded probes were then labelled with the radioactive isotope [ $\gamma$ - $^{32}$ P]ATP. The radiolabelled probes were then added to a binding mix containing crude nuclear extracts prepared from cultured bronchial fibroblast cells.

Tracks: 1, 1b: probe only (-1346/A or -1346/C respectively) without protein extracts  
 2, 2b: probe + protein extracts  
 3, 3b: probe + protein extracts + 50x cold unlabelled -1346/A probe  
 4, 4b: probe + protein extracts + 50x cold unlabelled -1346/C probe  
 5, 6, 5b, 6b: probe + protein extracts + non-specific competitors

DNA-protein complexes are indicated by red arrows. Only the -1346/A probe showed evidence of the presence of such a complex (Tracks 2, 4, 5 and 6). This complex was not detected (Track 3) when a specific competitor (unlabelled -1346/A probe) was included in the assay.

## 6.6 Discussion

Gene expression is primarily regulated at the transcription level<sup>52</sup>. There are several cis-elements in the MMP-3 promoter that have been shown to play an important role in its regulation. These include the AP-1 site (-70-64 bp), which interacts with transcription factors of the c-Fos and c-Jun families<sup>225,226</sup>; two palindromically arranged PEA3 binding sites at -208 bp from the start of transcription which interact with the Ets family of transcription factors<sup>227</sup> and SPRE at -1573 bp also thought to interact with c-Jun family of transcription factors<sup>228</sup>. The mechanisms involved in MMP-3 regulation are still not fully characterised and it is possible that there are more functionally important cis-elements not yet identified<sup>52</sup>.

It is estimated that a SNP occurs every 1 in 250-1,000bp. The majority of these SNPs are likely to be functionally neutral, but some may exert allele-specific effects on the regulation of the gene or the protein function, thus causing differences between biological traits and susceptibility to disease<sup>142</sup>. It is however, difficult to separate functional SNPs from their non-functional counterparts by database analysis alone. To investigate whether the SNPs identified in the MMP-3 gene promoter have an effect on MMP-3 promoter activity, transient transfection experiments were performed. Experiments using constructs containing individual polymorphic sites were first carried out and the results indicated that the five promoter polymorphisms all have some influence on MMP-3 promoter activity. Further functional studies are still required using reporter gene constructs containing the 2.3 Kb MMP-3 promoters representing the most common haplotypes. Although the effects of these polymorphisms on MMP-3 promoter activity appear to be small, they may still have a contribution to the development of atherosclerosis. This is because, atherosclerosis is a slow process spanning several decades of life and thus a small excess or deficit of MMP expression can still be biologically significant. The magnitude of effect on promoter activity found here is comparable to that found for polymorphisms in some other cardiovascular candidate genes (please see further discussion below).

Several functional variants have been identified in a number of MMP genes (MMP-1, MMP-2, MMP-9 and MMP-12) as well as other candidate genes for CHD (Matrix Gla). In general, the differences in transcriptional activity between the alleles were similar to

those demonstrated here. Transient transfection experiments of the common promoter polymorphism C-1306/T in the MMP-2 gene consistently showed ~1.4-2-fold increase in the C allele when compared to the T allele. This difference was found to occur due to the abolishment of an Sp1 binding site<sup>224</sup>. A common functional polymorphism was also identified in the MMP-12 promoter. Here the polymorphism is located just outside the AP-1 binding site, a transition A→G at position -82 bp. The A allele had a higher transcriptional activity to that of the G allele. EMSA experiments further demonstrated that this might be due to the A allele having a stronger binding affinity of the AP-1 complex<sup>97</sup>. Several polymorphisms have been identified in the promoter of the Matrix Gla Protein gene, though the effects on transcriptional activity for most of these SNPs were minimal, although the promoter region with the -138/C SNP was found to consistently express 20% less activity in rat muscular smooth muscle activity and 50% less in human fibroblast cells<sup>229</sup>. Rutter *et al* in 1998 performed a series of experiments on the MMP-1 promoter polymorphism -1607 bp to investigate the effects on transcriptional activity. This SNP is caused by the insertion of a guanine nucleotide and created an Ets family consensus sequence 5'-GGA-3'. Their result showed that the 2G allele was transcriptional more active than the 1G allele. Furthermore the 2G homozygotes are more frequent in tumour cell lines derived from melanomas and breast cancers than in the normal population<sup>87</sup>.

A DNA-protein complex was detected as binding to the A allele, but not to the C allele of the -1346 bp polymorphic site. Transfection assays performed in this study (Chapter 6) revealed a 1.4-fold increase in transcriptional activity in the A allele compared to the C allele. It is therefore likely that the protein interacting with this DNA sequence is a transcription activator. Further to this, a consensus sequence for the transcription activator c-myc was identified in the promoter sequence at the -1346/A polymorphic site, which was abolished at the -1346/C polymorphic site using the TFSEARCH: Searching Transcription Factor Binding Sites RWCP Parallel Application TPC Laboratories database (Chapter 3).

No DNA-protein complexes were detected as binding to either the T or C probe for the -1986 bp polymorphism, although transfection experiments had previously showed a difference of 1.4-fold in transcriptional activity between the two alleles.

Unfortunately, there was insufficient time to complete the transcription factor investigations and further EMSA analysis is required for the two remaining promoter polymorphisms (-709 bp and -376 bp). DNase I footprinting or Methylation interference assays may also be carried out in the future, to identify the specific protein binding sequences. Any DNA-binding proteins identified can then be purified from nuclear extracts by column chromatography.

The importance of regulatory proteins has already been discussed earlier in this chapter. A clearer understanding in the expression of the MMP-3 gene will be achieved by identifying the gene regulatory proteins involved in its transcription. The abolishment or creation of transcription factor binding sites caused by common polymorphisms is likely to have functional consequence. Studies have consistently shown that an alteration in the rate of transcription and thus gene expression results in the susceptibility to destructive diseases.

In relation to MMPs, an increase in gene expression causes an imbalance between the synthesis and degradation of the ECM (extracellular matrix) thus aiding the progression of diseases, such as tumour invasion and metastasis, and atherosclerosis. The progression of atherosclerosis is encouraged by the migration of monocyte/macrophages through the endothelial lining and its underlying basement membrane and by the migration of SMC. Extracellular barriers need to be broken down for migration to occur. The clinical implications of increased MMP activity have also been recognized. The weakening of the fibrous cap in atherosclerotic lesions causes thrombosis and MI. Identifying the proteins involved in gene expression may lead to a better understanding of the underlying disease processes so that prevention, drugs or therapy can be developed.



## **Chapter 7**

### **Overview and Possible Further Studies**

#### **7.1 Overview**

During this study a number of novel polymorphisms were identified in the MMP-3 gene, which were all found to be in strong LD (Chapters 3 and 4). The two most common haplotypes showed significant differences in transcriptional activity (Chapter 6), which are associated with the extent of stenosis in the coronary arteries and the risk of MI. The previously identified 5A/6A polymorphism in particular was found to be the variant of significance. The 5A allele with increased transcriptional activity was associated with increased risk of MI, a clinical event that is often caused by rupture of lipid-rich, unstable atherosclerotic plaques. The 6A allele with a lower transcriptional activity was associated with an increase in risk of stenosis in all three coronary arteries (Chapter 5).

This chapter will discuss the findings of the investigations presented here, and will compare similar studies performed on other MMPs (i.e. MMP-1, -2, -9 and -12) in relation to atherosclerosis and CHD, as well as put forward possible further studies.

In this study, fifteen different PCR primer pairs, encompassing the promoter sequence (~2.3Kb) and the 10-exon regions, were designed to assess the extent and nature of sequence variants in the human MMP-3 gene (Chapter 3). Amplicons were generated from a group of 20 unrelated individuals and the PCR products were analysed using the mutation scanning technique ddF. The nature and location of variants were determined by sequencing. By using this approach a total of six new sequence variants were identified in the MMP-3 gene. All variants were in the form of single nucleotide substitutions, comprising of four transitions and two transversions. Four variants were located in the promoter region and two were located in the exon 2 of the coding region.

Of the two coding region variant, one was found to be a synonymous substitution. The other at position A+802/G causes a conservative amino acid change from lysine to glutamic acid. Data-mining studies suggest that this amino acid is unlikely to have an affect on MMP-3 function, as the codon is not evolutionary conserved.

Restriction endonuclease digests were performed on a group of unrelated individuals to determine the extent of LD between the variants. Strong LD was found to exist.

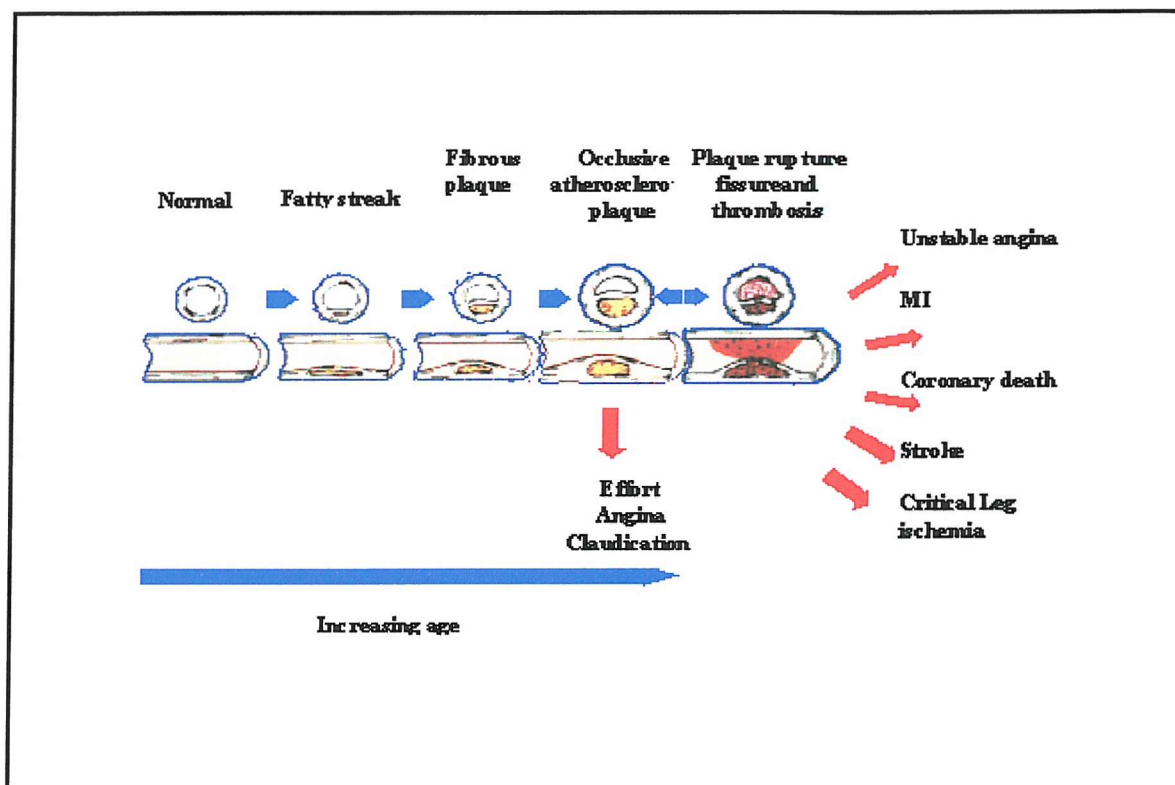
In silico analysis was used to determine whether the promoter variants created or abolished any potential cis-elements. Several sites were identified, though their functional importance is unclear.

Since the effect of sequence variants on gene regulation may be important in determining disease risk, several reporter gene constructs were generated to measure differences in allelic expression. The variants were found to produce allelic differences between 1.1-fold to 2-fold. While these differences were only moderate, they may still be biologically important, since atherosclerosis is a slow process, taking place over many years of life, and thus, small changes in MMP expression/activity might lead to a significant cumulative effect.

Three haplotypes, T-5A-A-A-G-A, T-5A-A-A-C-G and C-5A-C-G-C-G, were associated with a higher risk of MI. All three haplotypes contain the 5A allele but they contain either allele of the other polymorphisms. Haplotypes T-5A-A-A-C-G and C-5A-C-G-C-G showed a much higher effect on disease susceptibility than the T-5A-A-A-G-A haplotype. The T-5A-A-A-C-G and C-5A-C-G-C-G haplotypes, confers a 4.04 fold and an 8.31-fold increase in disease susceptibility and the T-5A-A-A-G-A haplotype confers a 1.36-fold increase in disease susceptibility. These former two haplotypes however, are less frequent in the cohort studied. Natural selection could be one of the reasons for its low frequency in this population study; alternatively, demographic factors may play a part. The UK is one of the countries in which admixture is of a high percentage and it may therefore be interesting to look further into the population structure of the Wessex area. Even though the T-5A-A-A-C-G and C-5A-C-G-C-G haplotypes represent such a low proportion of the population, investigations into the effects of this haplotype may further our understanding of the mechanisms involved in CHD. These data further indicate an association between the 5A allele and risk of MI.

In summary, the major findings of this study are that the 5A/6A polymorphism is a major genetic determinant for both the arterial stenosis due to atherosclerosis and risk

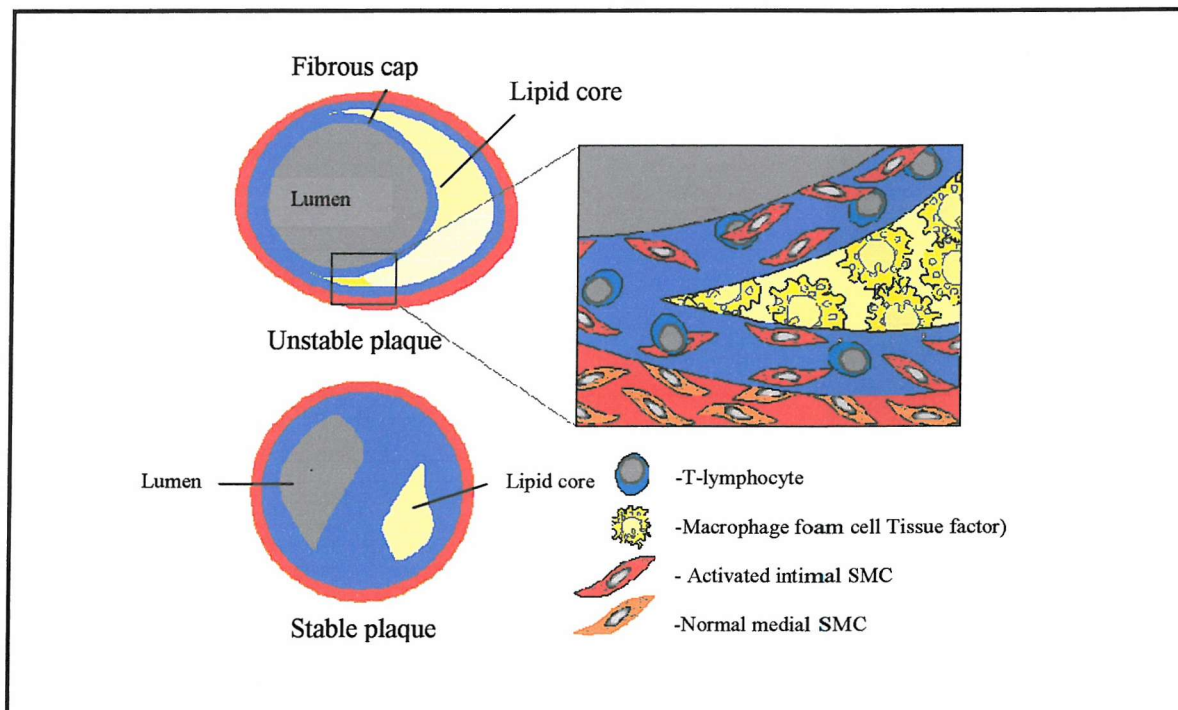
of MI. The 5A allele, increases promoter activity and may therefore lead to increased gene expression. An increase in MMP-3 may contribute to plaque rupture either by directly degrading the matrix proteins in the plaque cap or through activation of other MMPs such as the collagenases MMP-1 and MMP-13 or gelatinases MMP-9 and MMP-2 that are thought to be involved in plaque rupture, or both. Carriers of this allele are more likely to suffer from MI. The 6A allele has a lower promoter activity and may consequently lead to less MMP-3 expression, less degradation and an increased accumulation of ECM. Carriers of the 6A allele are therefore more likely to have a more severe progression of atherosclerosis i.e. a more fibrous plaque, arterial wall thickening and a smaller lumen. The data present in this study is consistent with previous investigations performed on MMP-3 and CHD.



**Figure 7.1:** - Diagram showing the progression of atherosclerosis, adapted from, Lipids Online. [www.lipidsonline.org](http://www.lipidsonline.org)

Atherosclerosis is a multifactorial progressive disease and is the underlying cause of CHD. The main characteristics of atherosclerosis are the accumulation of subendothelial deposits of lipids, fibrous material, SMC and macrophages, calcium and necrotic debris<sup>3</sup>. Atherosclerosis begins in infancy and early childhood<sup>9</sup>. However, it

is the advanced stages of the disease that are the most life threatening (Figure 7.1). The amount of tissue that forms in these lesions varies greatly between a more lipid to a more fibrous consistency, resulting in an unstable or stable plaque respectively (Figure 7.2). Stable plaques are those that develop a thicker fibrous cap. It is this form of lesion that is primarily the cause of hardening and narrowing of the arteries (stenosis) leading to stable angina. The 6A allele is associated with a higher-grade stenosis. Unstable plaques have a more fatty core and a thinner fibrous cap and are more prone to rupture. The cause of plaque rupture is unknown but is thought to be attributable to high blood pressure, increased blood flow, due to stress or exercise. Unstable plaques that rupture may lead to embolisms or thrombosis, which if presented in one of the coronary arteries can trigger MI. The 5A allele is associated with unstable plaques. Susceptibility to coronary heart disease and myocardial infarction is influenced by the interactions of both genetic and environmental factors. Known genetic factors do not account for all the genetic susceptibility to CHD, indicating that other genetic factors are yet to be identified.



**Figure 7.2** adapted from Molecular Bases of the Acute Coronary Syndromes Libby 1995<sup>232</sup>

The results of this study support the notion that sequence variants in MMP genes might influence MMP expression and matrix remodelling. There has been evidence in literature to support this notion:

Atherosclerotic plaques typically contain a large quantity of macrophages, which express MMP-3 as well as several other MMPs (MMP-1, -2, -9, -12 -13 and MT-MMP) <sup>233</sup>. MMP activity is thus thought to contribute to both the progression of atherosclerosis and plaque rupture leading to thrombus formation. The skeletal strength of atherosclerotic plaques is dependant on the interstitial collagens, and therefore over-degradation of these proteins may cause plaque instability and rupture. MMP-1 is known to mediate the initial step of interstitial collagens degradation, while complete degradation also requires other family members including the stromelysins (MMP-3) and the gelatinases (MMP-2 and MMP-9) <sup>83-85,234</sup>. It has been found that both MMP-1 and MMP-13 are expressed more in unstable lipid plaques than the more stable fibrous plaques. MMP-13 mRNA is generally not found in human tissue except in areas of chronic inflammation and extensive matrix remodelling, and absent in normal arteries <sup>235</sup>.

MMP-9 is another member of the MMP family that is over expressed in unstable plaques. Like MMP-3, MMP-9 has a broad specificity. Its substrates include gelatins, Type IV collagens and it also has proteolytic activity against proteoglycan core proteins and elastins <sup>50</sup>. Regulated primarily at the level of transcription, it is postulated that common promoter variants in MMP-9 may affect gene expression, and in so doing influence connective tissue degradation in atherosclerotic lesions. A total number of 10 sequence variants have been identified in the MMP-9 gene. Unlike the sequence variants identified in the MMP-3 gene, these variants were evenly distributed throughout the gene sequence. Four were located in the promoter, five in the coding region and one in the 3' untranslated region of the gene <sup>98</sup>. Allelic association studies identified strong linkage disequilibrium between these variants. A polymorphism arising from a C to T substitution at position -1562 relative to the transcription start site has been shown to have an allelic effect on MMP-9 promoter activity. The T allele has been found to have higher transcriptional activity over the C allele, likely due to its abolishment of a transcription repressor protein-binding site. Data from a study of patients with coronary disease showed that the T allele which is transcriptionally more active, was more prevalent in patients which had >50% stenosis in the three coronary arteries <sup>96</sup>. These results are noteworthy as the 6A allele of the MMP-3 gene, which is transcriptionally less active, has also been associated with a higher-grade stenosis. These differences are likely to be due to different roles of the MMPs in the

pathogenesis of atherosclerosis. It should be noted that MMP-9 plays an important role in smooth muscle cell migration and proliferation during atherogenesis, which may explain the association between the -1562/T allele and greater extent of atherosclerosis.

Other MMPs with functional variants thought to influence CAD susceptibility include MMP-2, MMP-12 and MT1-MMP. The gelatinase MMP-2 that possesses proteolytic activity against Type IV collagen, a major basement membrane component, is expressed in the atherosclerotic plaque. Recently, a functional polymorphism was identified in the promoter region located C-1306/T relative to the start of transcription. The T allele abolishes the cis-element, Sp1 binding site and in so doing causes a reduction in MMP-2 synthesis<sup>224</sup>. No direct association has as yet been made for this common polymorphism with CAD, however MMP-2 has been implicated in a number of destructive diseases including atherosclerosis and this polymorphism presents itself as a strong candidate. MMP-12 has a broad substrate specificity; it not only digests elastin but can also degrade basement membrane components. It is secreted by activated macrophages and so is likely to be involved in the events leading up to macrophage penetration of the inner layers of arterial walls during atherosclerosis. A promoter polymorphism A-82/G in the MMP-12 gene, which is located close to the AP-1 transcription factor-binding site, has been identified<sup>97</sup>. This variant affects the binding affinity of the AP-1 protein complex and the A allele has been associated with a higher transcriptional activity. Preliminary studies have shown that this polymorphism influences coronary artery luminal diameter in diabetic patients. Expressions of MT1-MMP by both SMC and macrophages have also been observed in atherosclerotic lesions<sup>236</sup>. MT1-MMP is a major activator of pro-MMP-2, and both MMP-2 and TIMP-2 have been demonstrated in plaque development<sup>237</sup>. It is speculated that activation of MT1-MMP may contribute to the enhanced degradation of connective tissue matrix in atherosclerotic plaques.

The findings of this study of variation in the MMP-3 gene, together with data from studies of the other MMP genes discussed above, all support the notion that genetic variation affecting the expression and/or activity of MMPs may influence the development and progression of atherosclerosis, through effects on vascular matrix turnover during atherogenesis.

## 7.2 Further Work

### 7.2.1 Electrophoretic Mobility Shift Assay and DNase I Footprinting

It is possible that the promoter polymorphisms identified are located in regions that are bound by transcription factors and the sequence changes may affect binding of these factors. To investigate whether this is the case, the EMSA can be performed to study all these polymorphisms. The exact location of the DNA-protein interaction cannot be determined using the mobility shift assay and therefore a DNase I footprinting assay would need to follow. For this technique, the DNA and protein are mixed in a similar method to the mobility shift assay. However, only one end of one strand of the DNA fragment is required for radioactive labelling and the DNA fragment is then treated with a small amount of deoxyribonuclease (DNase I). After digestion, the bound proteins are removed and the DNA fragments are separated on a polyacrylamide gel by electrophoresis<sup>216</sup>. This would produce a ladder of bands representing the products of DNase I digestion. The binding of a transcription factor protein would prevent digestion of the DNA fragment from occurring.

### 7.2.2 Affinity Chromatography

Any protein complexes identified will need to be defined. A popular method for accomplishing such a task is affinity chromatography (see Chapter 6 for details). Regulatory elements that influence MMP-3 activity are not completely defined. A number of factors have been shown to influence MMP-3 expression, but the underlying mechanisms are not completely understood. Several studies have shown that changes in MMP mRNA levels are influenced by chemical agents, cytokines, neurohormones, changes in basement membrane adhesion and cytoskeletal architecture<sup>50, 238, 239, 67</sup>. MMP activity is tightly controlled at three levels: transcription, activation and inhibition/deactivation. Cytokines such as TNF and IL-1 have been demonstrated to increase MMP transcription in a number of cell types. These cytokines are also influential factors in the progression of atherosclerosis. Known as “prime mover”, cytokines act on a broad range of molecules and cells including other cytokines, neutrophils, monocytes and endothelial cells. They can induce the procoagulant changes in endothelium, and are also involved in the expression of adhesion molecules,



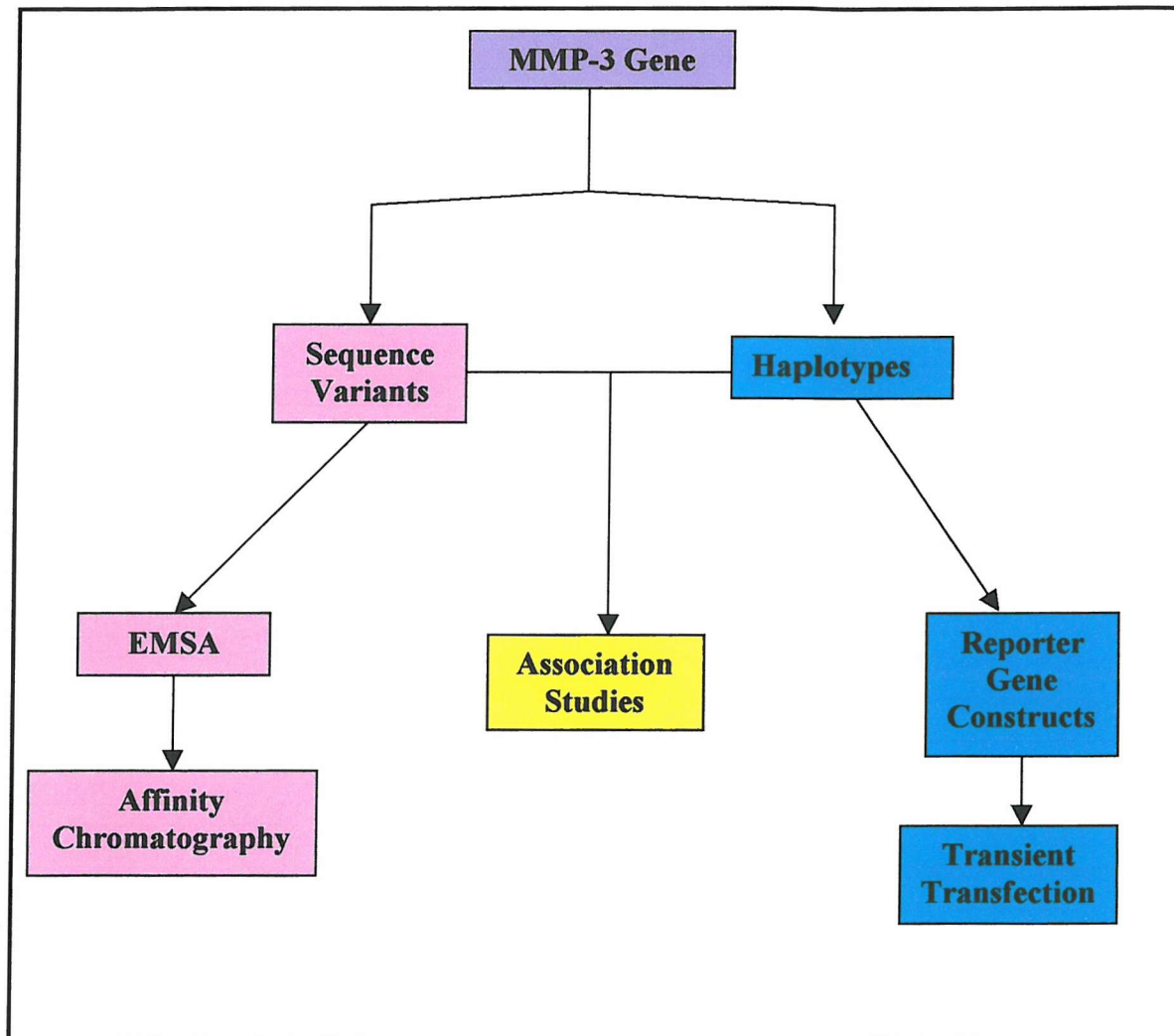
which draw monocytes and neutrophils into the developing lesion. Increased levels of both TNF and IL-1 have been observed in atherosclerotic plaques. A number of other cytokines have also been associated with atherosclerosis IL-8, IL-6 M-CSF, GM-CSF and monocyte chemoattractant protein <sup>240</sup>. As discussed earlier in Chapter 1, the promoter regions of MMPs have common regulatory sequences, however, there are also several elements that differ between different MMPs. It therefore follows that extracellular stimuli such as cytokines as well as other agents are likely to cause different levels of MMP expression. Another important control point for MMP activity is proteolytic processing of the pro-enzyme. MMPs are primarily secreted as inactive zymogens or pro-MMPs and activation occurs after they are secreted into the ECM. Activation is a complicated process involving several steps including proteolytic cleavage by a serine protease, followed by either autolytic cleavage or cleavage by another MMP. MMP-3 is the key member in its family that is involved in this important regulatory step, the overall activation of MMPs and it therefore follows that the expression and activation of MMP-3 is crucial for the complete activation of MMPs <sup>67</sup>.

### 7.2.3 Functional Studies

In this study, experiments were performed to investigate whether the SNPs identified in the MMP-3 promoter region have an effect on MMP-3 promoter activity. This was achieved by performing transient transfection experiments for individual polymorphisms. The combined effect of all five polymorphisms on promoter activity however has not been investigated and thus still needs to be carry out. This can be achieved by first making reporter gene constructs of the most common haplotypes (T-5A-A-A-G-A and C-6A-C-A-C-G) and then by performing a series of transient transfection studies in mammalian cells for each individual haplotype.

Although having lower frequencies in the population, the T-5A-A-A-C-G and C-5A-C-G-C-G haplotypes are associated with very high risk of MI and therefore need to be addressed further. Functional studies need to be performed to investigate whether these haplotypes affect MMP-3 promoter activity, and if so, to what extent. This can be achieved by first making reporter gene constructs of each haplotype, which can then be used for transient transfection studies.





**Figure 7.3:** -Flow chart reflecting possible further studies for the investigation of MMP-3 and its influences on the progression of atherosclerosis.

#### 7.2.4 Association Studies

The association studies performed here need to be replicated. While strong association between the 5A alleles and MI and the 6A allele and arterial stenosis was observed and consistent with previous finding, the results still need to be clarified in further studies, such as other case-control association studies or a family-base studies.

The MMP-3 gene, located on the long arm of chromosome 11, is situated in a cluster of six other MMPs, namely MMP-8, -10, -1, -12, -7 and MMP-13. It is possible that the 5A/6A polymorphism acts as a marker for a disease susceptible variant in one of the other MMP genes included in this cluster. To investigate this theory functional studies would first have to be performed on all promoter polymorphisms in this gene cluster. All variants would need to be genotyped in large cohorts of individuals. Stepwise regression analysis can be performed on the data, which would give the sequence variants that are associated with disease risk. The effects of haplotype on disease risk can also be analysed in this way.

The notion that genes lying on the same chromosome may be co-regulated is not new. One of the best-known examples of linked genes is the imprinting genes insulin-like growth factor 2 (Igf2) and H19. Recent studies on the transcriptional regulation of these neighbouring genes found that there is competition between the promoters for a shared enhancer element <sup>241</sup>. MMPs are known to play an important role in the progression of atherosclerosis and risk of MI by mediating the degradation of the ECM. While the expression of different MMPs in atherosclerotic lesions is well established, the relationship between the enzymatic activity of family member and disease risk is not. MMPs share a number of gene regulatory proteins, which can either activate or repress transcription. MMPs may react to some stimuli in a different way. An example of this is the regulatory protein TNF- $\beta$ , which stimulates the gelatinases, MMP-2 and MMP-9 but represses MMP-1 and MMP-3 synthesis. The final steps of MMP activation may require the expression an activation of another member of its family. It would therefore be beneficial to our understanding of this complex disease to investigate these destructive enzymes in this way.

### 7.2.5 Conclusion

There are several advantages for identifying susceptibility genes. These include: (1) Improving our understanding of disease mechanisms in both the pathophysiological and gene-environment interactions, (2) Reclassifying diseases on the basis of aetiology, thus improving predictions regarding prognosis and treatment response, (3) Developing new and improved treatments, (4) Providing preventative measures for individuals at a high risk to the disease.

The data in this study indicate that variation in the MMP3 gene influences the extent of coronary atherosclerosis and risk of myocardial infarction. These influences are attributable to the 5A/6A polymorphism, with the 6A/6A genotype being associated with the extent of atherosclerosis and the 5A allele-containing genotypes being associated with myocardial infarction risk. Thus the effect of the 5A/6A polymorphism may contribute to patient-to-patient variability in atherosclerotic plaque composition. These findings not only are relevant to the understanding of the pathogenesis of coronary artery disease, but also provide a target for future development of predictive, preventive and therapeutic measures.

## Reference

- 1 Hahnemann Symposium. Atherosclerotic Vascular Disease. (1967).
- 2 Libby,P. Lesion versus lumen. *Nat. Med.* **1**, 17-18 (1995).
- 3 Libby,P. Changing concepts of atherogenesis. *J. Intern. Med.* **247**, 349-358 (2000).
- 4 Camm,A.J. Clinical Medicine. Bailliere Tindall, (1994).
- 5 McCullagh,K.G. Arteriosclerosis in the African elephant: Part 2. Medial sclerosis. *Atherosclerosis* **21**, 37-59 (1975).
- 6 Krista,L.M. & McQuire,J.A. Atherosclerosis in coronary, aortic, and sciatic arteries from wild male turkeys (*Meleagris gallopavo silvestris*). *Am. J. Vet. Res.* **49**, 1582-1588 (1988).
- 7 Zimmerman,M.R. *et al.* The paleopathology of an Aleutian mummy. *Arch. Pathol. Lab Med.* **105**, 638-641 (1981).
- 8 Magee,R. Arterial disease in antiquity. *Med. J. Aust.* **169**, 663-666 (1998).
- 9 Stary,H.C. *et al.* A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Arterioscler. Thromb. Vasc. Biol.* **15**, 1512-1531 (1995).
- 10 Ross,R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* **362**, 801-809 (1993).
- 11 Nielsen,L.B. Atherogenicity of lipoprotein(a) and oxidized low density lipoprotein: insight from in vivo studies of arterial wall influx, degradation and efflux. *Atherosclerosis* **143**, 229-243 (1999).
- 12 Scanu,A.M., Lawn,R.M. & Berg,K. Lipoprotein(a) and atherosclerosis. *Ann. Intern. Med.* **115**, 209-218 (1991).
- 13 Scanu,A.M. Lp(a) as a marker for coronary heart disease risk. *Clin. Cardiol.* **14**, I35-I39 (1991).
- 14 Ross,R. Heart Diseases. W.B. Saunders Company, (1998).
- 15 Gray's Anatomy. Churchill Livingstone, (1995).
- 16 Libby,P. & Clinton,S.K. The Role of Macrophages in Atherogenesis. *Curr. Opin. Lipidol.* 355-363 (1993).
- 17 Ross,R. Atherosclerosis is an inflammatory disease. *Am. Heart J.* **138**, S419-S420 (1999).
- 18 Basic Histology. Appleton and Lange, (1989).

- 19 Bonin,L.R. *et al.* Generation and characterization of human smooth muscle cell lines derived from atherosclerotic plaque. *Arterioscler. Thromb. Vasc. Biol.* **19**, 575-587 (1999).
- 20 Hoff,H.F. & O'Neil,J.A. Oxidation of LDL: role in atherogenesis. *Klin. Wochenschr.* **69**, 1032-1038 (1991).
- 21 Witztum,J.L. & Steinberg,D. Role of oxidized low density lipoprotein in atherogenesis. *J. Clin. Invest* **88**, 1785-1792 (1991).
- 22 Nathan,D.M., Meigs,J. & Singer,D.E. The epidemiology of cardiovascular disease in type 2 diabetes mellitus: how sweet it is ... or is it? *Lancet* **350 Suppl 1**, SI4-SI9 (1997).
- 23 Bierman,E.L. George Lyman Duff Memorial Lecture. Atherogenesis in diabetes. *Arterioscler. Thromb.* **12**, 647-656 (1992).
- 24 MacMahon,S. *et al.* Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* **335**, 765-774 (1990).
- 25 Prescott,E., Hippe,M., Schnohr,P., Hein,H.O. & Vestbo,J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. *BMJ* **316**, 1043-1047 (1998).
- 26 Wilhelmsen,L. Coronary heart disease: epidemiology of smoking and intervention studies of smoking. *Am. Heart J.* **115**, 242-249 (1988).
- 27 Bolego,C., Poli,A. & Paoletti,R. Smoking and gender. *Cardiovasc. Res.* **53**, 568-576 (2002).
- 28 Manson,J.E. *et al.* The primary prevention of myocardial infarction. *N. Engl. J. Med.* **326**, 1406-1416 (1992).
- 29 Gustavo Maroni. Molecular and Genetic Analysis of Human Traits. Blackwell Science, Massachusetts (2001).
- 30 Day,I.N. & Humphries,S. Genetics of Complex Diseases. BIOS Scientific Publishers, Oxford (1997).
- 31 Cambien,F. *et al.* Sequence diversity in 36 candidate genes for cardiovascular disorders. *Am. J. Hum. Genet.* **65**, 183-191 (1999).
- 32 Tiret,L. *et al.* Evidence, from combined segregation and linkage analysis, that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. *Am. J. Hum. Genet.* **51**, 197-205 (1992).
- 33 Cambien,F. *et al.* Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature* **359**, 641-644 (1992).

- 34 Pfeffer, M.A. *et al.* Effect of captopril on mortality and morbidity in patients with left ventricular dysfunction after myocardial infarction. Results of the survival and ventricular enlargement trial. The SAVE Investigators. *N. Engl. J. Med.* **327**, 669-677 (1992).
- 35 Weisgraber, K.H., Rall, S.C., Jr. & Mahley, R.W. Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. *J. Biol. Chem.* **256**, 9077-9083 (1981).
- 36 Rall, S.C., Jr., Weisgraber, K.H., Innerarity, T.L. & Mahley, R.W. Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects. *Proc. Natl. Acad. Sci. U. S. A* **79**, 4696-4700 (1982).
- 37 Smit, M. *et al.* Genetic heterogeneity in familial dysbetalipoproteinemia. The E2(lys146- ---gln) variant results in a dominant mode of inheritance. *J. Lipid Res.* **31**, 45-53 (1990).
- 38 Breslow, J.L. *et al.* Studies of familial type III hyperlipoproteinemia using as a genetic marker the apoE phenotype E2/2. *J. Lipid Res.* **23**, 1224-1235 (1982).
- 39 Wilson, P.W., Schaefer, E.J., Larson, M.G. & Ordovas, J.M. Apolipoprotein E alleles and risk of coronary disease. A meta-analysis. *Arterioscler. Thromb. Vasc. Biol.* **16**, 1250-1255 (1996).
- 40 Dallongeville, J., Lussier-Cacan, S. & Davignon, J. Modulation of plasma triglyceride levels by apoE phenotype: a meta- analysis. *J. Lipid Res.* **33**, 447-454 (1992).
- 41 Lambert, J.C. *et al.* Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations-the ECTIM study. *Hum. Mol. Genet.* **9**, 57-61 (2000).
- 42 Sanghera, D.K., Aston, C.E., Saha, N. & Kamboh, M.I. DNA polymorphisms in two paraoxonase genes (PON1 and PON2) are associated with the risk of coronary heart disease. *Am. J. Hum. Genet.* **62**, 36-44 (1998).
- 43 Heinecke, J.W. & Lusis, A.J. Paraoxonase-gene polymorphisms associated with coronary heart disease: support for the oxidative damage hypothesis? *Am. J. Hum. Genet.* **62**, 20-24 (1998).
- 44 Kane, W.H., Ichinose, A., Hagen, F.S. & Davie, E.W. Cloning of cDNAs coding for the heavy chain region and connecting region of human factor V, a blood coagulation factor with four types of internal repeats. *Biochemistry* **26**, 6508-6514 (1987).
- 45 Dahlback, B., Carlsson, M. & Svensson, P.J. Familial thrombophilia due to a previously unrecognized mechanism characterized by poor anticoagulant response to activated protein C: prediction of a cofactor to activated protein C. *Proc. Natl. Acad. Sci. U. S. A* **90**, 1004-1008 (1993).

- 46 Bertina,R.M. *et al.* Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* **369**, 64-67 (1994).
- 47 Ye,S. & Henney,A.M. Vascular matrix proteins and their remodelling in atherosclerosis an introduction to vascular biology. Cambridge University Press, Cambridge (1996).
- 48 Woessner,J.F. Matrix metalloproteinases and their inhibitors in connective tissue remodeling. *FASEB J.* **5**, 2145-2154 (1991).
- 49 Ye,S., Humphries,S. & Henney,A. Matrix metalloproteinases: implication in vascular matrix remodelling during atherogenesis. *Clin. Sci. (Colch.)* **94**, 103-110 (1998).
- 50 Birkedal-Hansen,H. *et al.* Matrix metalloproteinases: a review. *Crit Rev. Oral Biol. Med.* **4**, 197-250 (1993).
- 51 Sternlicht,M.D. & Werb,Z. How matrix metalloproteinases regulate cell behavior. *Annu. Rev. Cell Dev. Biol.* **17**, 463-516 (2001).
- 52 Parks,W.C. & Mecham,R.P. Matrix Metalloproteinases. Academic Press, USA (1998).
- 53 Pendas,A.M., Santamaria,I., Alvarez,M.V., Pritchard,M. & Lopez-Otin,C. Fine physical mapping of the human matrix metalloproteinase genes clustered on chromosome 11q22.3. *Genomics* **37**, 266-268 (1996).
- 54 Van Wart,H.E. & Birkedal-Hansen,H. The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proc. Natl. Acad. Sci. U. S. A* **87**, 5578-5582 (1990).
- 55 Becker,J.W. *et al.* Stromelysin-1: three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Protein Sci.* **4**, 1966-1976 (1995).
- 56 Bode,W., Gomis-Ruth,F.X. & Stockler,W. Astacins, serralyins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett.* **331**, 134-140 (1993).
- 57 Dhanaraj,V. *et al.* X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure.* **4**, 375-386 (1996).
- 58 Murphy,G. *et al.* The C-terminal domain of 72 kDa gelatinase A is not required for catalysis, but is essential for membrane activation and modulates interactions with tissue inhibitors of metalloproteinases. *Biochem. J.* **283** ( Pt 3), 637-641 (1992).

- 59 Strongin,A.Y. *et al.* Mechanism of cell surface activation of 72-kDa type IV collagenase. Isolation of the activated form of the membrane metalloprotease. *J. Biol. Chem.* **270**, 5331-5338 (1995).
- 60 Kusano,K. *et al.* Regulation of matrix metalloproteinases (MMP-2, -3, -9, and -13) by interleukin-1 and interleukin-6 in mouse calvaria: association of MMP induction with bone resorption. *Endocrinology* **139**, 1338-1345 (1998).
- 61 Quinones,S., Saus,J., Otani,Y., Harris,E.D. & Kurkinen,M. Transcriptional regulation of human stromelysin. *J. Biol. Chem* **264**, 8339-8344 (1989).
- 62 Kirstein,M. *et al.* Cross-talk between different enhancer elements during mitogenic induction of the human stromelysin-1 gene. *J. Biol. Chem* **271**, 18231-18236 (1996).
- 63 Gutman,A. & Wasylyk,B. The collagenase gene promoter contains a TPA and oncogene-responsive unit encompassing the PEA3 and AP-1 binding sites. *EMBO J.* **9**, 2241-2246 (1990).
- 64 Ramos-DeSimone,N. *et al.* Activation of matrix metalloproteinase-9 (MMP-9) via a converging plasmin/stromelysin-1 cascade enhances tumor cell invasion. *J. Biol. Chem.* **274**, 13066-13076 (1999).
- 65 Mazzieri,R. *et al.* Control of type IV collagenase activity by components of the urokinase- plasmin system: a regulatory mechanism with cell-bound reactants. *EMBO J.* **16**, 2319-2332 (1997).
- 66 Springman,E.B., Angleton,E.L., Birkedal-Hansen,H. & Van Wart,H.E. Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. *Proc. Natl. Acad. Sci. U. S. A* **87**, 364-368 (1990).
- 67 Nagase,H. Activation mechanisms of matrix metalloproteinases. *Biol. Chem.* **378**, 151-160 (1997).
- 68 Pei,D. & Weiss,S.J. Furin-dependent intracellular activation of the human stromelysin-3 zymogen. *Nature* **375**, 244-247 (1995).
- 69 HE,C.S. *et al.* Tissue cooperation in a proteolytic cascade activating human interstitial collagenase. *Proc. Natl. Acad. Sci. U. S. A* **86**, 2632-2636 (1989).
- 70 Nagase,H., Enghild,J.J., Suzuki,K. & Salvesen,G. Stepwise activation mechanisms of the precursor of matrix metalloproteinase 3 (stromelysin) by proteinases and (4- aminophenyl)mercuric acetate. *Biochemistry* **29**, 5783-5789 (1990).
- 71 Overall,C.M. & Sodek,J. Concanavalin A produces a matrix-degradative phenotype in human fibroblasts. Induction and endogenous activation of collagenase, 72-kDa gelatinase, and Pump-1 is accompanied by the suppression of the tissue inhibitor of matrix metalloproteinases. *J. Biol. Chem.* **265**, 21141-21151 (1990).



- 72 Kinoshita,T. *et al.* TIMP-2 promotes activation of progelatinase A by membrane-type 1 matrix metalloproteinase immobilized on agarose beads. *J. Biol. Chem.* **273**, 16098-16103 (1998).
- 73 Murphy,G. & Willenbrock,F. Tissue inhibitors of matrix metalloendopeptidases. *Methods Enzymol.* **248**, 496-510 (1995).
- 74 Greene,J. *et al.* Molecular cloning and characterization of human tissue inhibitor of metalloproteinase 4. *J. Biol. Chem.* **271**, 30375-30380 (1996).
- 75 Williamson,R.A., Carr,M.D., Frenkiel,T.A., Feeney,J. & Freedman,R.B. Mapping the binding site for matrix metalloproteinase on the N-terminal domain of the tissue inhibitor of metalloproteinases-2 by NMR chemical shift perturbation. *Biochemistry* **36**, 13882-13889 (1997).
- 76 Gomis-Ruth,F.X. *et al.* Crystal structure of an oligomer of proteolytic zymogens: detailed conformational analysis of the bovine ternary complex and implications for their activation. *J. Mol. Biol.* **269**, 861-880 (1997).
- 77 Clark,I.M., Powell,L.K. & Cawston,T.E. Tissue inhibitor of metalloproteinases (TIMP-1) stimulates the secretion of collagenase from human skin fibroblasts. *Biochem. Biophys. Res. Commun.* **203**, 874-880 (1994).
- 78 Smith,M.R., Kung,H., Durum,S.K., Colburn,N.H. & Sun,Y. TIMP-3 induces cell death by stabilizing TNF-alpha receptors on the surface of human colon carcinoma cells. *Cytokine* **9**, 770-780 (1997).
- 79 Gomez,D.E., Alonso,D.F., Yoshiji,H. & Thorgeirsson,U.P. Tissue inhibitors of metalloproteinases: structure, regulation and biological functions. *Eur. J. Cell Biol.* **74**, 111-122 (1997).
- 80 Nagase,H. & Woessner,J.F., Jr. Matrix metalloproteinases. *J. Biol. Chem.* **274**, 21491-21494 (1999).
- 81 Werb,Z. ECM and cell surface proteolysis: regulating cellular ecology. *Cell* **91**, 439-442 (1997).
- 82 Nikkari,S.T. *et al.* Interstitial collagenase (MMP-1) expression in human carotid atherosclerosis. *Circulation* **92**, 1393-1398 (1995).
- 83 Henney,A.M. *et al.* Localization of stromelysin gene expression in atherosclerotic plaques by in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A* **88**, 8154-8158 (1991).
- 84 Galis,Z.S., Sukhova,G.K., Lark,M.W. & Libby,P. Increased expression of matrix metalloproteinases and matrix degrading activity in vulnerable regions of human atherosclerotic plaques. *J. Clin. Invest* **94**, 2493-2503 (1994).
- 85 Brown,D.L., Hibbs,M.S., Kearney,M., Loushin,C. & Isner,J.M. Identification of 92-kD gelatinase in human coronary atherosclerotic lesions. Association of active enzyme synthesis with unstable angina. *Circulation* **91**, 2125-2131 (1995).

- 86 Vu,T.H. & Werb,Z. Matrix metalloproteinases: effectors of development and normal physiology. *Genes Dev.* **14**, 2123-2133 (2000).
- 87 Rutter,J.L. *et al.* A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription. *Cancer Res.* **58**, 5321-5325 (1998).
- 88 Kanamori,Y. *et al.* Correlation between expression of the matrix metalloproteinase-1 gene in ovarian cancers and an insertion/deletion polymorphism in its promoter region. *Cancer Res.* **59**, 4225-4227 (1999).
- 89 Lochter,A., Sternlicht,M.D., Werb,Z. & Bissell,M.J. The significance of matrix metalloproteinases during early stages of tumor progression. *Ann. N. Y. Acad. Sci.* **857**, 180-193 (1998).
- 90 Sternlicht,M.D. *et al.* The stromal proteinase MMP3/stromelysin-1 promotes mammary carcinogenesis. *Cell* **98**, 137-146 (1999).
- 91 Konttinen,Y.T. *et al.* Analysis of 16 different matrix metalloproteinases (MMP-1 to MMP-20) in the synovial membrane: different profiles in trauma and rheumatoid arthritis. *Ann. Rheum. Dis.* **58**, 691-697 (1999).
- 92 Evanson,J.M., Jeffrey,J.J. & Krane,S.M. Human collagenase: identification and characterization of an enzyme from rheumatoid synovium in culture. *Science* **158**, 499-502 (1967).
- 93 Case,J.P., Lafyatis,R., Remmers,E.F., Kumkumian,G.K. & Wilder,R.L. Transin/stromelysin expression in rheumatoid synovium. A transformation-associated metalloproteinase secreted by phenotypically invasive synoviocytes. *Am. J. Pathol.* **135**, 1055-1064 (1989).
- 94 Okada,Y., Takeuchi,N., Tomita,K., Nakanishi,I. & Nagase,H. Immunolocalization of matrix metalloproteinase 3 (stromelysin) in rheumatoid synovioblasts (B cells): correlation with rheumatoid arthritis. *Ann. Rheum. Dis.* **48**, 645-653 (1989).
- 95 Shah,P.K. *et al.* Human monocyte-derived macrophages induce collagen breakdown in fibrous caps of atherosclerotic plaques. Potential role of matrix-degrading metalloproteinases and implications for plaque rupture. *Circulation* **92**, 1565-1569 (1995).
- 96 Zhang,B. *et al.* Functional polymorphism in the regulatory region of gelatinase B gene in relation to severity of coronary atherosclerosis. *Circulation* **99**, 1788-1794 (1999).
- 97 Jormsjo,S. *et al.* Allele-specific regulation of matrix metalloproteinase-12 gene activity is associated with coronary artery luminal dimensions in diabetic patients with manifest coronary artery disease. *Circ. Res.* **86**, 998-1003 (2000).
- 98 Zhang,B. *et al.* Genetic variation at the matrix metalloproteinase-9 locus on chromosome 20q12.2-13.1. *Hum. Genet.* **105**, 418-423 (1999).

- 99 Ye,S. *et al.* Progression of coronary atherosclerosis is associated with a common genetic variant of the human stromelysin-1 promoter which results in reduced gene expression. *J. Biol. Chem* **271**, 13055-13060 (1996).
- 100 Mao,D., Lee,J.K., VanVickle,S.J. & Thompson,R.W. Expression of collagenase-3 (MMP-13) in human abdominal aortic aneurysms and vascular smooth muscle cells in culture. *Biochem. Biophys. Res. Commun.* **261**, 904-910 (1999).
- 101 Carrell,T.W., Burnand,K.G., Wells,G.M., Clements,J.M. & Smith,A. Stromelysin-1 (matrix metalloproteinase-3) and tissue inhibitor of metalloproteinase-3 are overexpressed in the wall of abdominal aortic aneurysms. *Circulation* **105**, 477-482 (2002).
- 102 Lamblin,N. *et al.* Polymorphisms in the promoter regions of MMP-2, MMP-3, MMP-9 and MMP-12 genes as determinants of aneurysmal coronary artery disease. *J. Am. Coll. Cardiol.* **40**, 43-48 (2002).
- 103 Sapolsky,A.I., Howell,D.S. & Woessner,J.F., Jr. Neutral proteases and cathepsin D in human articular cartilage. *J. Clin. Invest* **53**, 1044-1053 (1974).
- 104 Werb,Z. & Reynolds,J.J. Stimulation by endocytosis of the secretion of collagenase and neutral proteinase from rabbit synovial fibroblasts. *J. Exp. Med.* **140**, 1482-1497 (1974).
- 105 Ye,S., Watts,G.F., Mandalia,S., Humphries,S.E. & Henney,A.M. Preliminary report: genetic variation in the human stromelysin promoter is associated with progression of coronary atherosclerosis. *Br. Heart J.* **73**, 209-215 (1995).
- 106 Galis,Z.S., Sukhova,G.K., Kranzhofer,R., Clark,S. & Libby,P. Macrophage foam cells from experimental atheroma constitutively produce matrix-degrading proteinases. *Proc. Natl. Acad. Sci. U. S. A* **92**, 402-406 (1995).
- 107 Libby,P. *et al.* Macrophages and atherosclerotic plaque stability. *Curr. Opin. Lipidol.* **7**, 330-335 (1996).
- 108 Talhouk,R.S., Bissell,M.J. & Werb,Z. Coordinated expression of extracellular matrix-degrading proteinases and their inhibitors regulates mammary epithelial function during involution. *J. Cell Biol.* **118**, 1271-1282 (1992).
- 109 Witty,J.P., Wright,J.H. & Matrisian,L.M. Matrix metalloproteinases are expressed during ductal and alveolar mammary morphogenesis, and misregulation of stromelysin-1 in transgenic mice induces unscheduled alveolar development. *Mol. Biol. Cell* **6**, 1287-1303 (1995).
- 110 Lochter,A. *et al.* Matrix metalloproteinase stromelysin-1 triggers a cascade of molecular alterations that leads to stable epithelial-to-mesenchymal conversion and a premalignant phenotype in mammary epithelial cells. *J. Cell Biol.* **139**, 1861-1872 (1997).
- 111 Handa,T. *et al.* Effects of hydrostatic pressure on matrix synthesis and matrix metalloproteinase production in the human lumbar intervertebral disc. *Spine* **22**, 1085-1091 (1997).

- 112 Ito,A. *et al.* Degradation of interleukin 1beta by matrix metalloproteinases. *J. Biol. Chem.* **271**, 14657-14660 (1996).
- 113 Takahashi,M. *et al.* The association of degeneration of the intervertebral disc with 5a/6a polymorphism in the promoter of the human matrix metalloproteinase-3 gene. *J. Bone Joint Surg. Br.* **83**, 491-495 (2001).
- 114 Satsangi,J. *et al.* A functional polymorphism of the stromelysin gene (MMP-3) influences susceptibility to primary sclerosing cholangitis. *Gastroenterology* **121**, 124-130 (2001).
- 115 Humphries,S.E. *et al.* The 5A/6A polymorphism in the promoter of the stromelysin-1 (MMP-3) gene predicts progression of angiographically determined coronary artery disease in men in the LOCAT gemfibrozil study. Lipid Coronary Angiography Trial. *Atherosclerosis* **139**, 49-56 (1998).
- 116 Terashima,M. *et al.* Stromelysin promoter 5A/6A polymorphism is associated with acute myocardial infarction. *Circulation* **99**, 2717-2719 (1999).
- 117 Warner,R.L. *et al.* Role of stromelysin 1 and gelatinase B in experimental acute lung injury. *Am. J. Respir. Cell Mol. Biol.* **24**, 537-544 (2001).
- 118 Newby,A.C., Southgate,K.M. & Davies,M. Extracellular matrix degrading metalloproteinases in the pathogenesis of arteriosclerosis. *Basic Res. Cardiol.* **89 Suppl 1**, 59-70 (1994).
- 119 PCR Protocols: A Guide to Methods and Applications. Academic Press Inc., San Diego (1990).
- 120 Robin Martin. Gel Electrophoresis:Nucleic Acids. Graham, J. M. 1996. Bios Scientific Publishers Ltd. The INTRODUCTION TO BIOTECHNIQUES. Graaham, J. M and Billington, D.  
Ref Type: Serial (Book,Monograph)
- 121 Sambrook,J., Fritsch,E.F. & and T.Maniatis. Molecular Cloning, A Laboratory Manual. Second edition. 1989. Cold Spring Harbor Laboratory Press.  
Ref Type: Serial (Book,Monograph)
- 122 Sarkar,G., Yoon,H.S. & Sommer,S.S. Dideoxy fingerprinting (ddE): a rapid and efficient screen for the presence of mutations. *Genomics* **13**, 441-443 (1992).
- 123 Miller,S.A., Dykes,D.D. & Polesky,H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).
- 124 Protocols and Applications Guide. Fisher Scientific, (1996).
- 125 Grompe,M. The rapid detection of unknown mutations in nucleic acids. *Nat. Genet.* **5**, 111-117 (1993).
- 126 Cotton,R.G. Mutation Detection. Oxford University Press, Oxford (1997).

- 127 Orita,M., Iwahana,H., Kanazawa,H., Hayashi,K. & Sekiya,T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. U. S. A* **86**, 2766-2770 (1989).
- 128 Whittall,R. *et al.* Utilities for high throughput use of the single strand conformational polymorphism method: screening of 791 patients with familial hypercholesterolaemia for mutations in exon 3 of the low density lipoprotein receptor gene. *J. Med. Genet.* **32**, 509-515 (1995).
- 129 Sekiya,T. Detection of mutant sequences by single-strand conformation polymorphism analysis. *Mutat. Res.* **288**, 79-83 (1993).
- 130 Fan,E., Levin,D.B., Glickman,B.W. & Logan,D.M. Limitations in the use of SSCP analysis. *Mutat. Res.* **288**, 85-92 (1993).
- 131 White,M.B., Carvalho,M., Derse,D., O'Brien,S.J. & Dean,M. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301-306 (1992).
- 132 Keen,J., Lester,D., Inglehearn,C., Curtis,A. & Bhattacharya,S. Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* **7**, 5 (1991).
- 133 Myers,R.M., Maniatis,T. & Lerman,L.S. Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol.* **155**, 501-527 (1987).
- 134 Riesner,D. *et al.* Temperature-gradient gel electrophoresis of nucleic acids: analysis of conformational transitions, sequence variations, and protein-nucleic acid interactions. *Electrophoresis* **10**, 377-389 (1989).
- 135 Rosenbaum,V. & Riesner,D. Temperature-gradient gel electrophoresis. Thermodynamic analysis of nucleic acids and proteins in purified form and in cellular extracts. *Biophys. Chem* **26**, 235-246 (1987).
- 136 Hawkins,G.A. & Hoffman,L.M. Base excision sequence scanning. *Nat. Biotechnol.* **15**, 803-804 (1997).
- 137 Underhill,P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996-1005 (1997).
- 138 O'Donovan,M.C. *et al.* Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. *Genomics* **52**, 44-49 (1998).
- 139 Day,I.N., O'Dell,S.D., Spanakis,E. & Weavind,G.P. Microplate array diagonal gel electrophoresis (MADGE), CpG-PCR and temporal thermal ramp-MADGE (Melt-MADGE) for single nucleotide analyses in populations. *Genet. Anal.* **14**, 197-204 (1999).
- 140 Cotton,R.G. Slowly but surely towards better scanning for mutations. *Trends Genet.* **13**, 43-46 (1997).

- 141 Laboratory Protocols for Mutation Detection. Oxford University Press, Oxford (1996).
- 142 Brookes,A.J. The essence of SNPs. *Gene* **234**, 177-186 (1999).
- 143 Stratakis,C.A. *et al.* Dideoxyfingerprinting (ddF) analysis of the type X collagen gene (COL10A1) and identification of a novel mutation (S671P) in a kindred with Schmid metaphyseal chondrodysplasia. *Biochem. Mol. Med.* **59**, 112-117 (1996).
- 144 Felmlee,T.A. *et al.* Genotypic detection of Mycobacterium tuberculosis rifampin resistance: comparison of single-strand conformation polymorphism and dideoxy fingerprinting. *J. Clin. Microbiol.* **33**, 1617-1623 (1995).
- 145 Liu,Q., Feng,J. & Sommer,S.S. Bi-directional dideoxy fingerprinting (Bi-ddF): a rapid method for quantitative detection of mutations in genomic regions of 300-600 bp. *Hum. Mol. Genet.* **5**, 107-114 (1996).
- 146 Strachan,T. & Read,A.P. Human Molecular Genetics. Bios Scientific Publishers Ltd, Oxford (1996).
- 147 Kunkel,T.A. DNA replication fidelity. *J. Biol. Chem.* **267**, 18251-18254 (1992).
- 148 Holliday,R. & Grigg,G.W. DNA methylation and mutation. *Mutat. Res.* **285**, 61-67 (1993).
- 149 Collins,F.S., Brooks,L.D. & Chakravarti,A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229-1231 (1998).
- 150 SNP attack on complex traits. *Nat. Genet.* **20**, 217-218 (1998).
- 151 Edwards,D.R. *et al.* Transforming growth factor beta modulates the expression of collagenase and metalloproteinase inhibitor. *EMBO J.* **6**, 1899-1904 (1987).
- 152 Kerr,L.D., Miller,D.B. & Matrisian,L.M. TGF-beta 1 inhibition of transin/stromelysin gene expression is mediated through a Fos binding sequence. *Cell* **61**, 267-278 (1990).
- 153 Jennings,H.S. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independant, with special relation to the effect of linkage. *Genetics* **2**, 97-154 (1917).
- 154 Risch,N. & Merikangas,K. The future of genetic studies of complex human diseases. *Science* **273**, 1516-1517 (1996).
- 155 Horikawa,Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* **26**, 163-175 (2000).
- 156 Tishkoff,S.A. *et al.* Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**, 1380-1387 (1996).

- 157 de la,C.A. & Wright,F.A. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl. Acad. Sci. U. S. A* **95**, 12416-12423 (1998).
- 158 Peltonen,L. Positional cloning of disease genes: advantages of genetic isolates. *Hum. Hered.* **50**, 66-75 (2000).
- 159 Terwilliger,J.D. & Weiss,K.M. Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.* **9**, 578-594 (1998).
- 160 Read,A. Medical Genetics:An illustrated Outline. Gower Medical Publishings, (1989).
- 161 Templeton,A.R. *et al.* Recombinational and mutational hotspots within the human lipoprotein lipase gene. *Am. J. Hum. Genet.* **66**, 69-83 (2000).
- 162 Broman,K.W. & Weber,J.L. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* **65**, 1493-1500 (1999).
- 163 Pritchard,J.K. & Rosenberg,N.A. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* **65**, 220-228 (1999).
- 164 Dunning,A.M. *et al.* The extent of linkage disequilibrium in four populations with distinct demographic histories. *Am. J. Hum. Genet.* **67**, 1544-1554 (2000).
- 165 Taillon-Miller,P. *et al.* Juxtaposed regions of extensive and minimal linkage disequilibrium in human Xq25 and Xq28. *Nat. Genet.* **25**, 324-328 (2000).
- 166 Eaves,I.A. *et al.* The genetically isolated populations of Finland and sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **25**, 320-323 (2000).
- 167 Cargill,M. *et al.* Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **22**, 231-238 (1999).
- 168 Kruglyak,L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**, 139-144 (1999).
- 169 Lewontin,R.C. The interactions of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49-67 (1964).
- 170 Hill,W.G. & Robertson,A. The effects of inbreeding at loci with heterozygote advantage. *Genetics* **60**, 615-628 (1968).
- 171 Zaykin,D., Zhivotovsky,L. & Weir,B.S. Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* **96**, 169-178 (1995).
- 172 Levin,M.L. The occurrence of lung cancer in man. *Int. Contra Cancrum* **19**, 531-541 (1953).

- 173 Levin,M.L. & Bertell,R. RE: "simple estimation of population attributable risk from case- control studies". *Am. J. Epidemiol.* **108**, 78-79 (1978).
- 174 Devlin,B. & Risch,N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**, 311-322 (1995).
- 175 Guo,S.W. Linkage disequilibrium measures for fine-scale mapping: a comparison. *Hum. Hered.* **47**, 301-314 (1997).
- 176 Jorde,L.B. Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* **56**, 11-14 (1995).
- 177 Collins,A., Lonjou,C. & Morton,N.E. Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci. U. S. A* **96**, 15173-15177 (1999).
- 178 Abecasis,G.R. *et al.* Extent and distribution of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68**, 191-197 (2001).
- 179 Craig,J.M. & Bickmore,W.A. The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.* **7**, 376-382 (1994).
- 180 Przeworski,M. & Wall,J.D. Why is there so little intragenic linkage disequilibrium in humans? *Genet. Res.* **77**, 143-151 (2001).
- 181 de Maat,M.P. *et al.* Effect of the stromelysin-1 promoter on efficacy of pravastatin in coronary atherosclerosis and restenosis. *Am. J. Cardiol.* **83**, 852-856 (1999).
- 182 Humphries,S. *et al.* The 5A6A polymorphism in the promoter of the stromelysin-1 (MMP3) gene as a risk factor for restenosis. *Eur. Heart J.* **23**, 721-725 (2002).
- 183 Gnasso,A. *et al.* Genetic variation in human stromelysin gene promoter and common carotid geometry in healthy male subjects. *Arterioscler. Thromb. Vasc. Biol.* **20**, 1600-1605 (2000).
- 184 Rauramaa,R. *et al.* Stromelysin-1 and interleukin-6 gene promoter polymorphisms are determinants of asymptomatic carotid artery atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **20**, 2657-2662 (2000).
- 185 Rundek,T. *et al.* Carotid intima-media thickness is associated with allelic variants of stromelysin-1, interleukin-6, and hepatic lipase genes: the Northern Manhattan Prospective Cohort Study. *Stroke* **33**, 1420-1423 (2002).
- 186 Greenberg,D.A. Linkage analysis of "necessary" disease loci versus "susceptibility" loci. *Am. J. Hum. Genet.* **52**, 135-143 (1993).
- 187 Hodge,S.E. What association analysis can and cannot tell us about the genetics of complex disease. *Am. J. Med. Genet.* **54**, 318-323 (1994).
- 188 Knowler,W.C., Williams,R.C., Pettitt,D.J. & Steinberg,A.G. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am. J. Hum. Genet.* **43**, 520-526 (1988).



- 189 Falk,C.T. & Rubinstein,P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51** ( Pt 3), 227-233 (1987).
- 190 Spielman,R.S., McGinnis,R.E. & Ewens,W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506-516 (1993).
- 191 Wong,Y.K., Dawkins,K.D. & Ward,M.E. Circulating Chlamydia pneumoniae DNA as a predictor of coronary artery disease. *J. Am. Coll. Cardiol.* **34**, 1435-1439 (1999).
- 192 van Meurs,J. *et al.* Active matrix metalloproteinases are present in cartilage during immune complex-mediated arthritis: a pivotal role for stromelysin-1 in cartilage destruction. *J. Immunol.* **163**, 5633-5639 (1999).
- 193 Capecchi,M.R. High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell* **22**, 479-488 (1980).
- 194 Vaheri,A. & Pagano,J.S. Infectious poliovirus RNA: a sensitive method of assay. *Virology* **27**, 434-436 (1965).
- 195 Graham,F.L. & van der Eb,A.J. Transformation of rat cells by DNA of human adenovirus 5. *Virology* **54**, 536-539 (1973).
- 196 Felgner,P.L. *et al.* Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. U. S. A* **84**, 7413-7417 (1987).
- 197 Ye,G.N., Daniell,H. & Sanford,J.C. Optimization of delivery of foreign DNA into higher-plant chloroplasts. *Plant Mol. Biol.* **15**, 809-819 (1990).
- 198 McCutchan,J.H. & Pagano,J.S. Enhancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethylaminoethyl-dextran. *J. Natl. Cancer Inst.* **41**, 351-357 (1968).
- 199 al Moslih,M.I. & Dubes,G.R. The kinetics of DEAE-dextran-induced cell sensitization to transfection. *J. Gen. Virol.* **18**, 189-193 (1973).
- 200 Wigler,M. *et al.* Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. *Cell* **11**, 223-232 (1977).
- 201 Botchan,M., Stringer,J., Mitchison,T. & Sambrook,J. Integration and excision of SV40 DNA from the chromosome of a transformed cell. *Cell* **20**, 143-152 (1980).
- 202 Potter,H. Electroporation in biology: methods, applications, and instrumentation. *Anal. Biochem.* **174**, 361-373 (1988).
- 203 Potter,H., Weir,L. & Leder,P. Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci. U. S. A* **81**, 7161-7165 (1984).

- 204 Wood,K.V. Marker proteins for gene expression. *Curr. Opin. Biotechnol.* **6**, 50-58 (1995).
- 205 Gorman,C.M., Moffat,L.F. & Howard,B.H. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell Biol.* **2**, 1044-1051 (1982).
- 206 Rosenthal,N. Identification of regulatory elements of cloned genes with functional assays. *Methods Enzymol.* **152**, 704-720 (1987).
- 207 Alton,N.K. & Vapnek,D. Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. *Nature* **282**, 864-869 (1979).
- 208 Leslie,A.G., Moody,P.C. & Shaw,W.V. Structure of chloramphenicol acetyltransferase at 1.75-A resolution. *Proc. Natl. Acad. Sci. U. S. A* **85**, 4133-4137 (1988).
- 209 Seed,B. & Sheen,J.Y. A simple phase-extraction assay for chloramphenicol acyltransferase activity. *Gene* **67**, 271-277 (1988).
- 210 Hall,C.V., Jacob,P.E., Ringold,G.M. & Lee,F. Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. *J. Mol. Appl. Genet.* **2**, 101-109 (1983).
- 211 de Wet,J.R., Wood,K.V., Helinski,D.R. & DeLuca,M. Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A* **82**, 7870-7873 (1985).
- 212 de Wet,J.R., Wood,K.V., DeLuca,M., Helinski,D.R. & Subramani,S. Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell Biol.* **7**, 725-737 (1987).
- 213 Thompson,J.F., Hayes,L.S. & Lloyd,D.B. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* **103**, 171-177 (1991).
- 214 Fried,M. & Crothers,D.M. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
- 215 Garner,M.M. & Revzin,A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res.* **9**, 3047-3060 (1981).
- 216 **Transcription Factors** A Practicle Approach. Oxford University Press Inc., Oxford (1999).
- 217 Wissmann,A. & Hillen,W. DNA contacts probed by modification protection and interference studies. *Methods Enzymol.* **208**, 365-379 (1991).

- 218 Spiro,C., Bazett-Jones,D.P., Wu,X. & McMurray,C.T. DNA structure determines protein binding and transcriptional efficiency of the proenkephalin cAMP-responsive enhancer. *J. Biol. Chem.* **270**, 27702-27710 (1995).
- 219 Sigman,D.S. Chemical nucleases. *Biochemistry* **29**, 9097-9105 (1990).
- 220 Galas,D.J. & Schmitz,A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
- 221 Kolchinsky,A.M., Mirzabekov,A.D., Gilbert,W. & Li,L. Preferential protection of the minor groove of non-operator DNA by lac repressor against methylation by dimethyl sulphate. *Nucleic Acids Res.* **3**, 11-18 (1976).
- 222 Siebenlist,U. & Gilbert,W. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. *Proc. Natl. Acad. Sci. U. S. A* **77**, 122-126 (1980).
- 223 Matthews,J.C., Hori,K. & Cormier,M.J. Purification and properties of Renilla reniformis luciferase. *Biochemistry* **16**, 85-91 (1977).
- 224 Price,S.J., Greaves,D.R. & Watkins,H. Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation. *J. Biol. Chem.* **276**, 7549-7558 (2001).
- 225 Kerr,L.D., Holt,J.T. & Matrisian,L.M. Growth factors regulate transin gene expression by c-fos-dependent and c-fos-independent pathways. *Science* **242**, 1424-1427 (1988).
- 226 Curran,T. & Franza,B.R., Jr. Fos and Jun: the AP-1 connection. *Cell* **55**, 395-397 (1988).
- 227 Wasylyk,C., Gutman,A., Nicholson,R. & Wasylyk,B. The c-Ets oncoprotein activates the stromelysin promoter through the same elements as several non-nuclear oncoproteins. *EMBO J.* **10**, 1127-1134 (1991).
- 228 Sanz,L. *et al.* Zeta PKC plays a critical role during stromelysin promoter activation by platelet-derived growth factor through a novel palindromic element. *J. Biol. Chem.* **269**, 10044-10049 (1994).
- 229 Herrmann,S.M. *et al.* Polymorphisms of the human matrix gla protein (MGP) gene, vascular calcification, and myocardial infarction. *Arterioscler. Thromb. Vasc. Biol.* **20**, 2386-2393 (2000).
- 230 Guijarro,C. *et al.* 3-Hydroxy-3-methylglutaryl coenzyme a reductase and isoprenylation inhibitors induce apoptosis of vascular smooth muscle cells in culture. *Circ. Res.* **83**, 490-500 (1998).
- 231 Fukumoto,Y. *et al.* Statins alter smooth muscle cell accumulation and collagen content in established atheroma of watanabe heritable hyperlipidemic rabbits. *Circulation* **103**, 993-999 (2001).

- 232 Libby,P. Molecular bases of the acute coronary syndromes. *Circulation* **91**, 2844-2850 (1995).
- 233 Pasterkamp,G. *et al.* Atherosclerotic arterial remodeling and the localization of macrophages and matrix metalloproteases 1, 2 and 9 in the human coronary artery. *Atherosclerosis* **150**, 245-253 (2000).
- 234 Li,Z. *et al.* Increased expression of 72-kd type IV collagenase (MMP-2) in human aortic atherosclerotic lesions. *Am. J. Pathol.* **148**, 121-128 (1996).
- 235 Sukhova,G.K. *et al.* Evidence for increased collagenolysis by interstitial collagenases-1 and -3 in vulnerable human atheromatous plaques. *Circulation* **99**, 2503-2509 (1999).
- 236 Rajavashisth,T.B. *et al.* Membrane type 1 matrix metalloproteinase expression in human atherosclerotic plaques: evidence for activation by proinflammatory mediators. *Circulation* **99**, 3103-3109 (1999).
- 237 Hernandez-Barrantes,S., Bernardo,M., Toth,M. & Fridman,R. Regulation of membrane type-matrix metalloproteinases. *Semin. Cancer Biol.* **12**, 131-138 (2002).
- 238 Coker,M.L. *et al.* Matrix metalloproteinase expression and activity in isolated myocytes after neurohormonal stimulation. *Am. J. Physiol Heart Circ. Physiol* **281**, H543-H551 (2001).
- 239 Ries,C. & Petrides,P.E. Cytokine regulation of matrix metalloproteinase activity and its regulatory dysfunction in disease. *Biol. Chem. Hoppe Seyler* **376**, 345-355 (1995).
- 240 Clinton,S.K. & Libby,P. Cytokines and growth factors in atherogenesis. *Arch. Pathol. Lab Med.* **116**, 1292-1300 (1992).
- 241 Leighton,P.A., Saam,J.R., Ingram,R.S., Stewart,C.L. & Tilghman,S.M. An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev.* **9**, 2079-2089 (1995).

## Appendix

### MMP-3 Promoter Region: -

LOCUS HSU43511 2376 bp DNA linear PRI 03-NOV-2000

DEFINITION Homo sapiens stromelysin-1 gene, promoter region.

ACCESSION U43511

VERSION U43511.2 GI:11093513

KEYWORDS Human stromelysin-1 promoter; metalloproteinase; MMP-3.

SOURCE Homo sapiens.

ORGANISM [Homo sapiens](#)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 2376)

AUTHORS Kirstein,M., Sanz,L., Quinones,S., Moscat,J., Diaz-Meco,M.T. and  
Saus,J.

TITLE Cross-talk between different enhancer elements during mitogenic  
induction of the human stromelysin-1 gene

JOURNAL J. Biol. Chem. 271 (30), 18231-18236 (1996)

MEDLINE [96279378](#)

PUBMED [8663478](#)

REFERENCE 2 (bases 1 to 2376)

AUTHORS Saus,J.

TITLE Direct Submission

JOURNAL Submitted (15-DEC-1995) Molecular Cytology, Instituto de  
Investigaciones Citologicas, Amadeo de Saboya, 4, Valencia 46010,  
Spain

REFERENCE 3 (bases 1 to 2376)

AUTHORS Saus,J.

TITLE Direct Submission

JOURNAL Submitted (03-NOV-2000) Molecular Cytology, Instituto de  
Investigaciones Citologicas, Amadeo de Saboya, 4, Valencia 46010,  
Spain

REMARK Sequence update by submitter  
COMMENT On Nov 3, 2000 this sequence version replaced gi:[1155344](#).

FEATURES Location/Qualifiers

source 1..2376  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="XXB1"  
/cell\_type="lymphocyte"  
/tissue\_type="blood"

[promoter](#) 1..2310  
/evidence=experimental

[TATA\\_signal](#) 2281..2287

[mRNA](#) 2311..>2376  
/product="stromelysin-1"

BASE COUNT 801 a 439 c 436 g 700 t

ORIGIN

```
1 tctagaaatg tctgcatgat ttttggattt ttgactttt aatttacctg tttgacattt
61 gctatgagcc ttctactcat aactaatata ttatttagtt ctctaagtaa tttttggtta
121 cctactatat atcagatacc atgctaagta ctaggaatac agaatcaaat gaggcatggt
181 ccataccctc aagtagctta cattagaatg agagagacag ataaaccatt tcactacagt
241 tcagtgtgga aaatagagta gcagaggcag gtacaaggta ccattgaaca atgattaatg
301 actcttctct ggacttggga aacatcttcc agggaagtcg tcgaagctgt tttaaaatat
361 agcaaacttt tgtatttagt tcaggaacag catggcccat ttgccaatc acatcttaac
421 agtttgaaaa gcaaacatat tatctatcag gctttcctct aaactttaaa tatgttttat
481 aagttataac tccagagaaa atttacaaag gataaacctt aatatagaag gaattagagc
541 tgccacagct tctacacttt taacctctca atattttatc tgttgggctc cactgtttct
601 tcctggaatt cacatcactg ccaccactct gttctccttg tcctcatatc aatgtggcca
661 aatattttcc ctgtatttca atcaggacaa gacatggttt tttccccca tcaaaggaat
721 ggagaaccat agaatactag ttttaaaatg tctttaggcc aggtgccgtg acccatgtct
781 gtaatcctag cactttgaga ggttgaggca ggagaatcac ttgatcccag agctcgaaac
841 cagcctgggc aacatagtga aacctctgtc tctatTTTTT aaataaaatt tgaaaaagtc
```



901 tttagacata atctagtcta aaaatgaagg cttaaagtgt atgtatagcc ccttgccaag  
 961 tggctatcac ctgtgtgggc atcttcagtc atagggatct tattgccaca gagaaatccc  
 1021 tttaaactta ttgggtaaaa tctctccaat gtttattaag aaacacacaa aaaataaagc  
 1081 aaagaagaaa atgcaaaaga gttataaatg agaggaagca aaatgggcac ttattaaagg  
 1141 tctaataaat gcacatttgt atccatcatt ctactgagtt cttactccca agatgttctt  
 1201 cccttttagca aacaaataag caagtcagca aagaaagaaa gaacaaacaa aatgtgggtga  
 1261 tcagggaagc attgaggaga tggatgggtg caggtggcaa gaggactata aaagtittac  
 1321 aaaatgtctt cctctgaata tgtttagagt cttgcattca agcatttatt atacaccaat  
 1381 aatgtgagca acactttact tgacaaagaa acagaaaaga aaggaaagga agaaaacaga  
 1441 agagcatgaa gagaaaattt aggatggatt ctgtttctca acttcaaagc atctgctaatt  
 1501 ttgaatttag ggaggagggg aaaaggttga aagagaataa gacatgtgta gaagacaagg  
 1561 acagagagaa tttcagtccg gtaagcaatg taattcattt caattctaca actatttatg  
 1621 gagcagctac gtgggcccat caccattaa taaattggtt acagaattaa aaccaacca  
 1681 aagggaatat acttccttct tttcacaga cctctttgt tctattctgc ccatgaggtt  
 1741 ttcctcctca agaaccagca aatccaagca cagtcaatag caggcattac aaatcagatt  
 1801 cagaaaaata aatcaccctt tctaaatttc ttctagatat tatcttttat gttttgagta  
 1861 taattgtata tagtatagac tatagctatg tatgtacact ttccacttac atcttttatt  
 1921 tgcttttata atgtctttct taaaataaaa ctgcttttag aagtcttgca caattctgat  
 1981 ttttaccaag tcaacctact tcttctctca aaaggacaaa cataaattgt ctagtgaatt  
 2041 ccagtcaatt ttccagaag aaaaaaatg ctccagtttt ctctcttacc aagacaggaa  
 2101 gcacttcctg gagattaatc actgtgttgc cttgcaaaat tgggaagggt gagagaaatt  
 2161 agtaaagtag gttgtatcat cctacttga atttggaatg ttgggaaatg gtctgtctgc  
 2221 catttgatg aaagcaagga tgagtcaagc tgcgggtgat ccaacaaac actgtcactc  
 2281 tttaaaagct gcgtccccga ggttggaact acaaggagge aggcaagaca gcaaggcata  
 2341 gagacaacat agagctaagt aaagccagtg gaaatg

## The coding region of the MMP-3 gene

■ = Exon regions of the MMP-3 gene in reverse order, from exon 10 to exon 1 respectively

■ = Intron regions of the MMP-3 gene

### Genomic chromosome 11:

```
49921 tcaacaatta agccagctgt tactcttcaa agtgtgtgtc actttctttg catttgggtc
49981 aaactccaac tgtgaagatc cagtaaagaa ataaaagaac cctgcaaata cagacaaggg
50041 aaatagtaag atatttttcc aaataaagac atttgacaat atacaaaact cagaatcata
50101 gagaactaaa aatagaaagt gtttatagag ctttacttta tgtcagggtc tatgccaagc
50161 ttgttacatg aatttatattt catttatctt tgcaacaacc ctatgaggtt gtatgtctaa
50221 ctccatttac agattgagca agttgaggtt gagacagatt tcgatgcttc ctcaagggtt
50281 ctcagcaagg aagaatggag acagcactcg gtgccagctc tgtctgaaaa cacagccaga
50341 ttccagggtt cagggttatt ctgcttccga tcagataaat tctccacttg cttggaaact
50401 ctaatcacct atttctttct tccccaaaaa tctctctccc ttttccttgc attgcagcct
50461 aggaagcaca gcaactatct caaaacacca ggggaccctt tagtgctctg caaacatggt
50521 gatcagggtt cctttcaata aagatcatca gcctccactt ccttaccttg agtagaaaac
50581 aaaatctttt ttttttttta aactttgggc acatggaaaag gtaagtattt gcaaatgact
50641 ggcatgggca atgactgaca actcaggaaa gacagacaaa aaatctccct gggaattagt
50701 agcagcaata agatagggtg gaggagaagg ccagctaacc aacattaaac ctggccacac
50761 cacttgatgg gatttttatg ctaacatggt aatctgaaag tcgagcttga ctccacatgt
50821 tagctcgaga tgtaagtgc actggagcac ttgattttta tatgcatgta taagataaag
50881 tagcgagatt tgtgtaatta tctttatggt tccaaacaag ttattttgtg aagtagttct
50941 ataatgagta caaatgtagg cgaatcaaaa ttttgaggag ggaactggcc ctaagaaaca
51001 cttgttatta aaaagtttca atgctctctt tccctctgat atcataaaat gtttcaagtg
51061 ctctatggcc aacacagtaa gtatcctctt accaaattct tcaaaaacag catcaatctt
51121 tgagtcaatc cctggaaagt cttcagctat ttgcttggga aagcctggct ccatggaatt
51181 tctcttctca tcaaattctg accaagtaaa agaaacagct cagcattgca tagaggccat
```



51241 gttaccgaac atcttagatc atgttaggtt taatgtggaa ttttactaaa ctttataaat  
51301 actgcaaata aatgccaagc atatttggga ctatagaaat atgtgacttt aatggtactg  
51361 taaagagttc tgaaagccat atctccatcc ggaacagggg ccgcatcctg ctgtatgtag  
51421 cacatgctgt ttgtcatcat atttcttaag cctggacaga aagaacctta atcatctttc  
51481 tgtcacaaag acatgaaagt gtgaacatat atataaagtg aattgaatca acatctaaaa  
51541 tgcaatttta attccctacc acaattggcc catatatgcc tgctgtcccc tccccigcac  
51601 atctaaatct tctctggtgg cctcctggga aaagggatta ggataagaag caaagcacag  
51661 taagtgtgtt tttagcccat tggtctggata tcagatggtc ggctcgatg tctcaaagtc  
51721 tggccaaatg acccaggaat caacctgtct gtgagtgttg ggttgatatcc ataaactgga  
51781 ctgctattga tcatcatctc tttaatgagc catcattatc aggtagaggt gacaagtggg  
51841 tgaggttaga agtagccctt gaaaatcata tagtagtggt aggaagacaa aacaaacaca  
51901 caatgagaca gtgacaaaca cacagtgaga cagttagcca aactaaaaga caatgggaag  
51961 ggaacgctaa attgcggtct gggttctcag aacattttcc caatgggtct aaagtccacg  
52021 agggaaagct gcatggaggt gatgagactt gagccaggac ttgaaggatg ggcaagattt  
52081 gaatagggtg aggagagaag agggggctcc aggcaggtgg ggtgacagca ccaagggaag  
52141 caccagccaa ctgcacttgg gactgaaagc aggagagcct aaggtgctgc tgttttaag  
52201 taagaacttt cccacagttg tctgaagtcc catccttctt actaagagag aagcaggcct  
52261 aaggttgggg tgggggattt ccttagtaaa agaaactaat ttccttaatt aggtccata  
52321 caaagtcatt tctcttgcac ctcacctcca gtatttgtcc tctacaaaga aatatgtttt  
52381 gttcttttcc ttatcagaaa tggtctgcat gattttcttc acggttggag ggaaacctag  
52441 ggtgtggatg cctcttgggt atccagctcg tacctcattt cctctgatgg ccagaattg  
52501 atttctgtt aaatataaat aattcagagt catattgaat tataacagtt aagcccttcc  
52561 gcttttagaaa tacactttag catctttgat tttcatggta aagtagtagc ccaaatttta  
52621 gatttagtca ccagcttaca gtgggctttt cacaaaaatt tgcagactga atgaatgaac  
52681 taaggtcaaa tgagagaatt tacaactctg ggtaatttat cttcttgtct ttatttttat  
52741 catctatgaa atataaatac aatgatatac gccctgggta tttcttcatt accgttgact  
52801 actgtattag ccagaacttg tagtagggaa aattaaagtg ttgttcaatt tcttatgaga  
52861 gggtttactt tggcaagtta aacaaatgat gatatagtaa ttacctttaa aatgaaaac  
52921 gaggtccttg ctagtaactt catatgcggc atccacgcct gaaggaagag atggccaaaa  
52981 tgaagagatc aatgcaatt caggttcaag cttcctgagg gatttgcgcc aaaagtgcct

53041 aaaatatatg taaaaagaaa tgtaaattga aaaacaatct ttcaccttta gaatatatttc  
53101 ctcaccgtct tgcctcacc agcttcccc attctctcat cttctaggct ggatgacaga  
53161 tagaaatcag gtatatggcc attagaaaat ttctttcctt tttcttcaga agactgtagt  
53221 gagtctcccg tcataggttt ccagagtaat gtcttctatg aaagaaaaat tattctatcg  
53281 tgcaaatgaa ttagcccagc ataagcttta attacctgca atgtatgcaa aacagtgcgt  
53341 gtgctatacc ctcttccaaa agctaggccg tgtttgtata cgtgtatgtg ctgtgtaaat  
53401 atattcttgt atgtgtgtag aaagattctg gaaggatgca caaaaagtct tgactctgga  
53461 tgagtacagg gagaactggg aactaaaatt ttttaatttg ttctcatttg tgctctttga  
53521 atgtttgtca tgtgcatatt gatccatcag taagtaaatt acttgattaa atgtaatggt  
53581 gatcctgatg tcatttttaa aggaagtaaa aataaaatta ttattcgaaa gatctactca  
53641 tgccttttga aaagtaactg tgacttaaga catttggaag cagcaattag aagctaagaa  
53701 ccagatcagc ctgtgccttc aaaaatgaac ctgacaattt taggaaaatg tgggaaaagg  
53761 cagcaccaga tctaaatgat tccaagctca gaagaaacgt ttttgtttta aatgtctaag  
53821 aatctctatc aagtttctgc gtgttttctt ttctaaacat tgtagattaa aatggtagga  
53881 aaatataatt tttcagcctg acctgtcttt aaagatcagg atttctcccc tcagagtgtc  
53941 gacagcatca aaggacaaag caggatcaca gttggctggc gtcccagggt ctggaggggac  
54001 aggttccgtg ggtaccaggg gggctctcagg ggagtcaggg ggaggctcta aagggaacat  
54061 taggggaaat gtgatacgtt tcaatatatg cccatttgta tgctttccaa acattctcac  
54121 ggtgctttga actaaatcat aaatacttta tgtagcattt gaaagctctt ctgaagacca  
54181 tcattgcaaa actacaatgg ctgattttcc cagtgtcact caccatagag ggactgaatg  
54241 ccatttatat catcttgaga caggcggaac cgagtcaggt ctgtgagtga gtgatagagt  
54301 gggtagatca aagcttcagt gttggctgag tgaaagagac ccaggagtg gccaatcca  
54361 tgagcagcaa cgagaaataa attggtccct atttaagaaa tcgagaacaa tagttactta  
54421 ttttttaaat acatgtgcat tacacaggtc tgctgtgcat tgtattggtt tttacttctt  
54481 ggaaccacac agggcttttt acaccttggt tgttggttaa ttgaacaatt gactaattac  
54541 acacttaatg gaaagctaag tgtattttta aactcatacc tgtaatcaat ctttgagatt  
54601 gatttcccag ttacatccac ttgccccca tgcccccaa cccaccccc gcctcacaag  
54661 catattaact ccagaagaac cagatccctt ttcttgtagt caggaatgga atgaaccac  
54721 ttgcttagca tatcagatat ttacctcat ttgtcaatgc aaagctaaac actgttttaa  
54781 tatatttaca tttgccattt gcagtctcag ttaaaggaaa ctatattcat gtttaattgat



54841 aatgtctcag cctcatcatc atgaagtaca tattgagact aaaaacatga ataatgatag  
54901 atgtcatagg atttgcaaaa ttggtttaaa acacagtatg cccaatcatg tctagcattt  
54961 gtcacacaac agatttagaa tacataaaca ctggctttat tcaaaatfff ttccttcata  
55021 tttttaaaaa caccataat gctaaaattg ggaggcattg ggtaattatg atcaatatag  
55081 taagtaaate taccttactt ttgctagtat tacttgtgta acttgtgatg ctaggaacaa  
55141 tctcatttga gagtttgcag gaaggtttct ttacaagaat tcgttaactc aaccaatcaa  
55201 gaattctate caggactaaa ctacatttg tacttctac agaggtttac attttcatta  
55261 taaatcccag ttaaacctca gacatagtat taattgtttt cctatfffff tagaagtagt  
55321 gttaattgtt ttagctgatt tttttgaaaa ctttgatgt cattcccaa aacatacagt  
55381 tatgaactga gaaacattct tctacctaaa cagtttgaac gttgtatttg ggatttagtc  
55441 tttcccttca aaacaagccc ataaaaattc ttttaacct aattactaa aataatgcct  
55501 cagccagaaa aaaaaaaaaa aaaaaacccc cagttcctga attttaatcc ctaattccgg  
55561 gtttgaaaat taaaaattta atgggagaaa tttccttatt tccccccaa accacttttt  
55621 attttaaatc cggtgcaata tactaccct taaattttgg tatttaataa taaaaatft  
55681 tccttaataa gggttcccc caaatccca aaaacccgtt cttaaaaaa annnaaant  
55741 tattncnatn tentatattg tgcttaccaa tataggaaaa ctgttaaagc atctgaaaaa  
55801 tattttcaat aataattaaa acaaaaatat attctgctga tctctgatcc actggaaaat  
55861 ccagaacatc tcatttcttg agcattcttg gttattaaca gatttctaca tataaaaaaa  
55921 ttaatgccaa aatcattctg agagatgtgt aactaacctg ttgtatcctt tgtccattgt  
55981 tcatcatcat caaagtgggc atctccatta atccctggcc caggggcata ggcattggcc  
56041 aaaacatttc caggtccatc aaaagggtaa aagtctccat gttctagtag gaaaaaatt  
56101 tattggaaag atatgtaaca aggatccctt ttgagccttt tccagtacaa caacaccaga  
56161 tagatgaatg gatacatttc atgtgttttt tgttttgctt tttttttttt ttttttttta  
56221 cctctaactg caaaagagat cattatatca gcctctcctt catacagcct ggagaatgtg  
56281 agtggagtca cctcttccca gactttcaga gctttctcaa cagcagaatc aacagcatct  
56341 tttggcaaat ctggtgtata attcacaatc ctgtaggaga aaaattgaag cagatgctat  
56401 tttctctat gatatagttt ataaaattcc aatcttatgt gaaaaccct ctgaaccatt  
56461 acctgtatgt aaggtgggtt ttcctccact tcgggatgcc aggaaagggt ctgaagtgc  
56521 caacgtcagg aactccacac ctgggcttgc gcatcacctc cagagtgtcg gagtccagct  
56581 tccccgtcac ctccaatcca aggaacttct gcatttctcg gattttttta acaacaggac

56641 cactgtcctt tctcctaaca aactgtttca catctttttc gaggtcgtag tagttttcta  
 56701 gatattttcta acagaataag tatagttttt aggtcactct tacaattttt actatagcta  
 56761 tctattttata gtaagttttt gcagttgctc acttcttaat ctatttggag tattttctcta  
 56821 gcttgctgaa ataatgccaa attttataat atgatactag caacacaaat atttagctaa  
 56881 aattacgttg cattaaaaaa aaaaactgaa gctgttccaa agcaagatat gggtcatcca  
 56941 ctctccactt gttactcatg tgccttaagc cccagtgtg ctactgctat ggtgttgtgt  
 57001 gatcttaaga aaacacctaa tttttctggc ctttgtttct tcctacataa gaggagaagc  
 57061 ccaaaggtg tgataatgag accttttcca cttccaacac tctataactc tctattcttg  
 57121 aactgttcct acacagtaag ctcaagcatt ctatgtgggt tttaagacac acagaatggt  
 57181 ttcagcttac tctggaagct cccacctgg ccaggtcagt tagtgtaat tacctgaaca  
 57241 aggttcatgc tgggtgcctc accccttgca gctccatcca atggataggc tgagcaaact  
 57301 gccacgcaca gcaacagtag gattggaaga ctcttcattt ccactggctt tacttagctc  
 57361 tatgttgtct ctatgccttg ctgtcttgcc tgcctccttg taggtccaac ctcgggagcg  
 57421 cagcttttaa agagtgacag tgtttgttg gatcacccgc agcttgactc atccttgctt  
 57481 tcatccaaat ggcagcagga ccatttccaa acattccaaa ttcaaagtag gatgatacaa  
 57541 cctactttac taattttctt caaccttccc aattttgcaa ggcaacacag tgattaatct

Table of the SAS control DNA duplicated sample

|                |                |                          |
|----------------|----------------|--------------------------|
| <b>A1 – B2</b> | <b>D1 – D6</b> | <b>G1 – F2</b>           |
| <b>A2 – F4</b> | <b>D2 – C6</b> | <b>G2 – G3</b>           |
| <b>A3 – F8</b> | <b>D3 – G4</b> | <b>G3 – G2</b>           |
| <b>A4 – C7</b> | <b>D4 – B7</b> | <b>G4 – D3</b>           |
| <b>A5 – B4</b> | <b>D5 – F6</b> | <b>G5 – E3</b>           |
| <b>A6 – C3</b> | <b>D6 – D1</b> | <b>G6 – G8 – H1 – H2</b> |
| <b>A7 – B1</b> | <b>D7 – F5</b> | <b>G7 – E5</b>           |
| <b>A8 – H8</b> | <b>D8 – B8</b> | <b>G8 – G6 – H1 – H2</b> |
| <b>B1 – A7</b> | <b>E1 – B6</b> | <b>H1 – G6 – G8 – H2</b> |
| <b>B2 – A1</b> | <b>E2 – E7</b> | <b>H2 – G6 – G8 – H1</b> |
| <b>B3 – C5</b> | <b>E3 – G5</b> | <b>H3 – C1</b>           |
| <b>B4 – A5</b> | <b>E4 – H5</b> | <b>H4 – F7</b>           |
| <b>B5 – C8</b> | <b>E5 – G7</b> | <b>H5 – E4</b>           |
| <b>B6 – E1</b> | <b>E6 – E8</b> | <b>H6 – F1</b>           |
| <b>B7 – D4</b> | <b>E7 – E2</b> | <b>H7 – F3</b>           |
| <b>B8 – D8</b> | <b>E8 – E6</b> | <b>H8 – A8</b>           |
| <b>C1 – H3</b> | <b>F1 – H6</b> |                          |
| <b>C2 – C4</b> | <b>F2 – G1</b> |                          |
| <b>C3 – A6</b> | <b>F3 – H7</b> |                          |
| <b>C4 – C2</b> | <b>F4 – A2</b> |                          |
| <b>C5 – B3</b> | <b>F5 – D7</b> |                          |
| <b>C6 – D2</b> | <b>F6 – D5</b> |                          |
| <b>C7 – A4</b> | <b>F7 – H4</b> |                          |
| <b>C8 – B5</b> | <b>F8 – A3</b> |                          |